UNIVERSITY OF CALIFORNIA
RIVERSIDE

A Study of Genes in DNA Methylation and Transcriptional Gene Silencing in *Arabidopsis*

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

by

So Youn Won

August 2013

Dissertation Committee:
Dr. Xuemei Chen, Chairperson
Dr. Patricia S. Springer
Dr. Shou-Wei Ding

The Dissertation of So Youn Won is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I am deeply grateful to many people without whom this work would not have been possible and to whom I am greatly indebted.

First of all, I would like to express my sincere gratitude to my advisor, Dr. Xuemei Chen. Without her encouragement, support and academic guidance, I would not have been able to overcome hard times during my graduate study and finish my dissertation. She has always given me inspiration on how to become a scientist, mentor, leader and advisor. It has been a great privilege for me to learn a lot of things from her.

I would also like to appreciate the other members of my dissertation committee: Drs. Patricia Springer and Shou-Wei Ding. They have always provided insightful comments and constructive criticisms on my research and helped me stay on the right track throughout my graduate career.

I want to thank my fabulous lab-mates in Dr. Chen's lab. They have provided suggestions, advices and supports to accomplish my Ph.D research. In addition to the science aspect, they have also provided friendship and funny moments. I have shared the best and worst moments of my Ph.D career with them. I could not have imagined better fellow colleagues.

The text of Chapter 1 of this dissertation, in full, is a reprint of the material as appeared in *Silence* 3:6 in 2012. Dr. Xuemei Chen directed and supervised the research that forms the basis for Chapter 1 of the dissertation. Dr. Shengben Li, Dr. Binglian

Zheng, Yuanyuan Zhao, Dongming Li, Xin Zhao, Huilan Yi, Lei Gao and Thanh Theresa Dinh provided technical expertise.

Last but not least, I would like to thank my dear family. Specially, none of this would have been possible without the love, concern and support of my parents. They have become a source of driving force to finish the most challenging journey of my life.

So Youn Won

Riverside, California

August 2013

ABSTRACT OF THE DISSERTATION

A Study of Genes in DNA Methylation and Transcriptional Gene Silencing in *Arabidopsis*

by

So Youn Won

Doctor of Philosophy, Graduate Program in Plant Biology
University of California, Riverside, August 2013
Dr. Xuemei Chen, Chairperson

Cytosine methylation is a key epigenetic mark that mediates diverse biological phenomena by regulating chromatin states. In plants, it is predominantly found in transposons and repeats, and contributes to genome defense through transcriptional gene silencing (TGS). In my thesis research, two genes and their molecular mechanisms in DNA methylation and TGS are uncovered in the model plant *Arabidopsis thaliana*. First, the *LUCH* line containing a luciferase (LUC) transgene is characterized as a suitable system for forward genetic screens that aim to isolate genes in TGS. RNA-directed DNA methylation (RdDM), a mechanism that establishes DNA methylation, and active demethylation influence LUC expression by affecting its DNA methylation levels. The moderate LUC expression in *LUCH* enables genetic screens to isolate both enhancers and suppressors of TGS. Second, TATA-binding proteins-associated factor6 (TAF6) is identified as a positive factor of RdDM and TGS. *TAF6* is required for the TGS of the *LUCH* reporter and several endogenous RdDM targets. Genome-wide methylation assays

have uncovered *TAF6*-dependent loci for proper DNA methylation. TAF6 is well known as a transcription factor of RNA polymerase II (Pol II), however, I find that a *TAF6* mutation compromises the transcription by a Pol II homolog Pol V. These findings indicate that Pol V adopts the Pol II regulatory mechanism to generate long non-coding RNAs that guide RdDM. Third, a heat shock protein 20 (HSP20) is uncovered as a negative regulator of DNA methylation and TGS. *HSP20* prevents the silencing of TGS reporters as well as of several endogenously methylated targets. Mutations in *HSP20* result in hypermethylation and mislocalization of the RdDM effector protein ARGONAUTE4 in the nucleus. Moreover, HSP20 is found to interact with three methyl CpG-binding domain (MBD) proteins that may link cytosine methylation to histone modifications. Mutation in *MBD5* or *MBD6* also hypermethylated several *HSP20*-dependent loci. Based on these findings, I propose that *HSP20* acts with MBDs to recognize loci that undergo DNA methylation and antagonize TGS at these loci. While HSP20-MBDs likely act downstream of DNA methylation, they also affect the DNA methylation status through direct or indirect effects, such as AGO4 recruitment or histone modifications.

Table of Contents

List of Figures

INTRODUCTION

CHAPTER 1

CHAPTER 2

CHAPTER 3

List of Tables

CHAPTER 1

CHAPTER 2

CHAPTER 3

INTRODUCTION

Metabolism and function of small RNAs in plants

ABSTRACT

Small regulatory RNAs 20-30 nucleotides in length guide RNA-mediated silencing processes in eukaryotes in a sequence-specific manner. Based on their biogenesis and precursors, plant small RNAs are classified as miRNAs or siRNAs. Both types of small RNAs are processed from longer precursors by RNase III enzymes and are stabilized by 2'-*O*-methylation at the 3' terminal ribose. Mature small RNAs are incorporated into and guide the catalytic action of ARGONAUTE proteins. Small RNAs repress their targets at the posttranscriptional level via mRNA cleavage or translational inhibition or at the transcriptional level via heterochromatin formation. Small RNAs may also originate from exogenous sources such as viruses and transgenes. Consistent with the crucial roles of small RNAs in regulating many biological processes, the homeostasis of these molecules is precisely maintained through the regulation of their biogenesis and turnover.

INTRODUCTION

Ribonucleic acid (RNA) is a major macromolecule that executes biological events in living organisms. According to the central dogma of gene expression, RNA serves as an intermediate in the flow of genetic information from DNA to protein in the form of messenger RNA (mRNA). Additionally, there are different types of non-protein-coding RNAs (ncRNAs), whose classification is based primarily on their molecular functions but also reflects differences in size and accumulation. Several classes of ncRNAs are abundant and perform housekeeping duties: ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) are involved in protein synthesis, small nuclear RNAs (snRNAs) are involved in pre-mRNA splicing, and small nucleolar RNAs (snoRNAs) guide the modification of other RNAs. The length of ncRNAs is another important criterion for their classification. Long noncoding RNAs (lncRNAs) refer to ncRNAs that are longer than the arbitrary size of 200 nucleotides (nt) [1]. While it remains unclear whether the vast number of intergenic lncRNAs detected in microarray or RNA-seq experiments are transcriptional noise or functional RNAs, there has been increasing evidence of their regulatory functions in gene expression, especially in animals (reviewed in [2,3]). Small RNAs 20-30 nt in length, the focus of this chapter, have come to be recognized as an important class within the broad spectrum of ncRNAs because their regulatory functions are critical for biological processes.

Small RNAs are a central component of RNA-mediated silencing in all eukaryotes and regulate genes in a sequence-specific manner through the recognition of complementary nucleic acid sequences. Small RNAs are known to act via two main

mechanisms. In post-transcriptional gene silencing (PTGS), small RNAs guide the cleavage or translational inhibition of their target mRNAs [4-8]. In transcriptional gene silencing (TGS), small RNAs guide DNA or histone methylation, resulting in heterochromatin formation [9-11].

In plants, small RNAs are classified as microRNAs (miRNAs) or small interfering RNAs (siRNAs) based on their precursors and biogenesis [12]. miRNAs derive from longer RNA precursors containing a stem-loop or hairpin structure with partial base-pair mismatch in the stem region. While the mature miRNA is the single most abundant species generated from a precursor, the passenger strand (miRNA*) is the second most abundant species and may also be found *in vivo*. In contrast, siRNAs derive from longer double-stranded RNAs (dsRNAs) that exhibit nearly perfect sequence complementarity. Typically, multiple siRNA species are generated from a single precursor. Despite the differences in precursors and biogenesis that distinguish the different classes of small RNAs, however, it is important to emphasize that all small RNAs function as sequence-specific guides in target regulation.

Research over the past two decades has significantly improved the understanding of small RNAs and their regulatory mechanisms. Since the initial discovery of a miRNA, *lin-4*, from genetic screens of the nematode *Caenorhabditis elegans* in 1993, hundreds of thousands of small RNAs have been identified, particularly with the aid of next-generation sequencing technology [13,14]. Along with the improved understanding of its critical regulatory functions for numerous biological processes, small RNA-mediated gene silencing is also recognized as a powerful research tool in biology. The use of small

RNAs to knock down selected genes permits the dissection of the molecular functions of those genes and related pathways. Small RNA-based gene silencing has also been used for crop improvement and fighting human diseases (reviewed in [15,16]). The awarding of the 2006 Nobel Prize in Physiology and Medicine for the discovery of RNA interference (RNAi), a homology-based gene silencing phenomenon conferred by small RNAs, further exemplifies the significance of small RNA-mediated gene regulation.

In the present chapter, small RNAs in the model plant *Arabidopsis thaliana* will be discussed in terms of their biogenesis, their molecular mechanisms for target repression, and their biological functions. Major differences with animal small RNAs will also be introduced. Finally, mechanisms that contribute to small RNA homeostasis such as degradation will be discussed.

miRNAs

Biogenesis of miRNAs

miRNAs are small regulatory RNAs 20-22 nt in length that act in a sequence-specific manner primarily through PTGS (reviewed in [17]). Their biogenesis involves the following steps: transcription of the miRNA precursor, cleavage to yield the mature precursor, stabilization by methylation, nuclear export and incorporation into effector proteins.

miRNA precursors that give rise to mature miRNAs are encoded by *MIR* genes, which are located in intergenic regions. *MIR* genes are individual gene units with their own promoters and terminators and are transcribed by RNA polymerase II (Pol II)

[18,19]. Accordingly, *MIR* gene promoters harbor *cis*-acting elements for transcription by Pol II. As with protein-coding genes, the expression of *MIR* genes is subject to regulation, with Pol II transcription affected by spatiotemporal inputs specific to particular developmental stages and organs. In addition to these endogenous signals, exogenous cues from the environment, such as biotic and abiotic stresses, also affect transcription. Thus, the transcription of *MIR* genes and, ultimately, miRNA abundance, is governed by regulatory frameworks that respond to various signals. Mediator, a multi-protein complex, serves as a general transcription factor and is thought to integrate various signals to promote the recruitment of Pol II to promoters [20]. Mediator is required for the transcription of *MIR* genes in *Arabidopsis* [21]. After and/or during transcription, miRNA precursors are capped and polyadenylated at their 5' and 3' ends, respectively, and introns are spliced out in a manner similar to the processing of Pol II-transcribed pre-mRNA. These *MIR* gene transcripts, or primary miRNAs (pri-miRNAs), form hairpin structures with partial base-pair mismatch in the stem regions and are subsequently processed by small RNA biogenesis enzymes.

Through the successive action of Dicer-like (DCL) RNase III enzymes in the nucleus, a pri-miRNA is processed into a precursor miRNAs (pre-miRNA), which is in turn processed into a mature miRNA/miRNA* duplex. This duplex contains both the guide strand, the functional miRNA species that promotes PTGS, and the miRNA* passenger strand, which is eventually degraded. DCL1, an RNase III enzyme that specifically cleaves dsRNA, is responsible for the processing of most miRNAs [19,22-25]. However, the *Arabidopsis* genome encodes four DCL homologs, and several

evolutionarily young miRNAs are processed by DCL4 instead of DCL1 [24,26,27]. The first dicing step generates the pre-miRNA from the pri-miRNA: DCL1 cleaves the stem approximately 15 nt away from the base of the stem and generates a 2-nt 3' overhang. The second dicing step by DCL1 cleaves the newly formed pre-miRNA at a position closer to the terminal loop, generating a 20-22 nt miRNA/miRNA* duplex with 2-nt 3' overhangs [23,25]. During this process, the DCL1 enzyme is aided by DAWDLE (DDL), a forkhead-associated (FHA) domain protein; SERRATE (SE), a zinc finger protein; and HYPONASTIC LEAVES1 (HYL1), an RNA binding protein [28-32]. It has been proposed that DDL facilitates the recognition of pri-miRNAs by DCL1, while SE and HYL1 may improve the accuracy and efficiency of the dicing activity of DCL1 [32-34].

Following the release of the miRNA/miRNA* duplex from the pre-miRNA, HUA ENHANCER (HEN1) methylates both ends of the duplex [35]. Specifically, HEN1 deposits a single methyl group at the 2'-OH position of the 3' terminal ribose. As discussed in section 5 below, HEN1-mediated methylation enhances the stability of miRNAs. That miRNAs are generated in the nucleus while miRNA-directed PTGS occurs in the cytoplasm indicates that miRNAs are exported from the nucleus to the cytoplasm. HASTY (HST), an Exportin-5 (Exp-5) homolog, has been implicated in the nuclear export of miRNAs [36], but it is unknown whether HEN1-mediated methylation precedes or follows nuclear export.

In the cytoplasm, miRNAs are loaded into ARGONAUTE (AGO) effector proteins (reviewed in [37,38]), and small RNA-mediated repression reflects the functions of these two key players: guidance by small RNAs and the catalytic activity of AGO-

containing protein complexes. The *Arabidopsis* genome encodes ten AGO homologs. Among these, AGO1 functions as the major effector protein for miRNA-mediated PTGS and binds most miRNAs [39,40]. AGO7 specifically binds miR390, while AGO10 exhibits a binding preference for miR166/165 over other miRNA species [41-43].

Molecular mechanism of miRNAs

miRNAs repress the expression of targets via PTGS, which is associated with two modes of repressive action: mRNA cleavage and translational inhibition [4-8]. In miRNA-pathway-compromised mutants, these changes can be assessed through the detection of the mRNA transcript levels or protein abundance of miRNA targets. In the case of mRNA cleavage, target mRNAs are sliced at the center of the sequence bound by the miRNA. The cleaved products, particularly the 3' fragments, can be detected in wild-type plants. When miRNAs that regulate their targets via mRNA cleavage are disrupted, the levels of both target mRNAs and the corresponding protein products increase. In contrast, there is a disproportionate accumulation of target protein relative to that of target mRNAs when miRNAs that regulate their targets by translational inhibition are disrupted. mRNA cleavage and translational inhibition may occur in parallel. For instance, a fraction of the mRNA pool targeted by a single miRNA may be repressed by cleavage while the remaining fraction is regulated by translational inhibition.

In plants, miRNA-guided cleavage has been observed for most miRNAs and is considered to be a widespread regulatory mechanism of plant miRNAs. The endonucleolytic activity of AGO1 cleaves (or slices) the phosphodiester bond linking two

nucleotides in the target mRNA that correspond to the 10th and 11th nucleotides from the 5' terminal end of the miRNA [39,40]. The newly exposed 5' and 3' fragments are subsequently degraded by the exosome with 3'-5' exonuclease activity and EXONUCLEASE4 (XRN4) with 5'-3' exonuclease activity, respectively [44-47]. The degradation of the 5' fragment is further accelerated by template-independent oligo-uridylation [44]. Uridine tails are attached at the 3' end of the 5' fragment, which promotes decapping activity at the 5' end [44,48]. The 5' fragment is thus rendered susceptible to 5'-3' degradation, and translation of the cleavage fragment is prevented.

miRNA-directed translational inhibition is less commonly observed in plants than transcript cleavage, and there are two main explanations for this. While miRNAs in animals require perfect base pairing with the target mRNA only in the seed region, which corresponds to the 2nd to 7th nucleotides from the 5' end of the miRNA, plant miRNAs require much more extensive complementarity with the target mRNA for PTGS [49,50]. This difference may underlie the predominance of distinct repressive mechanisms in the two kingdoms, i.e., translational inhibition in animals and transcript cleavage in plants. Alternatively, the perceived predominance of miRNA-directed transcript cleavage in plants may reflect technical limitations. Although monitoring the effects of miRNAs on their targets is facilitated by their sequence complementarity, high quality antibodies for the proteins corresponding to the targeted mRNAs are necessary to assess the occurrence or extent of translational inhibition. Thus, the technical challenge of producing high quality antibodies may contribute to the less frequent observation of translational inhibition in plants.

The earliest reports of miRNA-directed translational inhibition in plants were the findings that *APETALA2* (*AP2*) and *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE3* (*SPL3*) are translationally repressed by miR172 and miR156/157, respectively [6-8]. Subsequent studies identified additional miRNAs that regulate their target mRNAs via translational inhibition. Moreover, several players in miRNA-directed translational inhibition have been identified from forward genetic screens in *Arabidopsis*. Mutations disrupting the microtubule-severing enzyme KATANIN1 (KTN1), the P body component VARICOSE (VCS)/Ge-1, the GW-repeat protein SUO and the ER membrane protein ALTERED MERISTEM PROGRAM1 (AMP1) compromise miRNA-mediated translational inhibition of exogenous reporter constructs and endogenous miRNA targets [51-53]. Notably, these studies show that miRNAs, such as miR156/157, miR164, miR165/166, miR172 and miR398, which are known to guide target transcript cleavage, also inhibit the translation of target mRNAs. The fact that genetic mutations (in *KTN1*, *VCS*, *SUO* and *AMP1*) can uncouple the transcript cleavage and translational inhibition activities of these miRNAs suggests that the two repressive modes of action are independent and occur in parallel to regulate target transcripts. The molecular mechanism of translational inhibition and the events that follow remain unclear. However, two studies in zebrafish and fruit fly suggest that translational inhibition primarily affects the initiation step rather than elongation or termination and that the subsequent stimulation of mRNA deadenylation and decay occurs in a miRNA-cleavage-independent manner [54,55].

Biological function of miRNAs

Because miRNAs repress or silence their target mRNAs, studies of the targets of miRNAs have also been critical for understanding miRNA function. According to the miRBase miRNA database (www.mirbase.org, Release 19) [56], the numbers of mature miRNAs and precursors in *Arabidopsis* are 338 and 299, respectively. Among known miRNA targets, transcription factors are the most highly represented functional group (reviewed in [57]). By recognizing *cis*-acting elements at the promoters of their target genes, transcription factors can systematically activate or repress genes belonging to a particular regulatory or functional network. The regulation of the transcription of downstream genes by transcription factors is affected by both endogenous and exogenous signals (reviewed in [58]). miRNAs regulating transcription factors therefore provide an additional layer of regulation for specific biological processes.

Although miRNA-mediated PTGS affects a wide variety of biological phenomena, its role in development is particularly well established, and a large number of miRNA-targeted transcription factors are implicated in developmental processes (reviewed in [17]). Interactions between miRNAs and their targets have been reported in a wide range of developmental contexts, such as embryogenesis, cell differentiation, pattern formation, phase transition and hormone signaling. Loss- or gain-of-function mutations in *MIR* genes or their targets often result in specific developmental phenotypes that are informative about their functions. Expressing miRNA-resistant targets under their endogenous promoters also affects plant morphology, demonstrating that miRNA-mediated PTGS is a critical regulatory component of developmental programs. Mutations

disrupting miRNA biogenesis genes and *AGO1* consistently result in pleiotropic developmental defects (reviewed in [59]). For example, the null *dcl1* allele is embryonic lethal, and the morphological phenotype of hypomorphic *dcl1* is similar to those of null *hyl1*, *hen1* and *hst* alleles. Mutations disrupting *AGO1* also yield phenotypes similar to those of mutants with disrupted *DCL* function. Taken together, the developmental defects of miRNA pathway mutants further establish the vital functions of miRNAs in development.

Other miRNAs affect the gene regulatory networks that govern responses to environmental cues (reviewed in [60]). Although less is known about miRNAs involved in stress responses compared to miRNAs involved in developmental processes, deep sequencing of small RNA populations has increased the number of identified miRNAs that are specifically expressed under certain environmental conditions, such as biotic and abiotic stresses [61]. For example, stress-related hormones such as abscisic acid can activate or repress the expression of certain miRNAs. In turn, stress-responsive miRNAs may target genes involved in detoxification or enhancing resistance. Because stress response signals also impact the developmental network, these changes may be critical for the ability of plants to alter their developmental program under harsh external conditions and to resume the normal program when the stress condition is removed.

Autoregulation of the miRNA pathway

Several self-feedback mechanisms are known to regulate the miRNA pathway. Two critical components of the miRNA pathway, *DCL1* and *AGO1*, are themselves

targets of miRNA-mediated repression [5,62]. Whereas DCL1 protein catalyzes miRNA biogenesis, there are three possible fates for *DCL1* transcripts: they may be translated into DCL1 protein, recognized as a cleavage target of miRNA-mediated PTGS or processed as a pri-miRNA. When miRNA levels are high, miR162-directed cleavage of *DCL1* transcripts by an AGO1-containing complex is likely favored [5]. Alternatively, the fold-back RNA structure within the *DCL1* transcript may recruit the miRNA biogenesis machinery and be diced by DCL1 to generate miR838 [24]. Thus, high miRNA levels or high DCL1 protein levels lead to a decrease in *DCL1* mRNA abundance and consequently reduced *DCL1* expression and activity. *AGO1* mRNA contains a miR168 binding site, which similarly permits feedback regulation of *AGO1* [62]. In this manner, the autoregulation of critical enzymes ensures the balanced dynamics of the miRNA pathway.

miRNAs in animals

As in plants, miRNAs in animals represent an essential regulatory module of gene expression (reviewed in [63,64]). Although miRNAs and miRNA targets are not well conserved between the two kingdoms, the general principles of miRNA biogenesis and function are held in common: stem-loop-containing precursors are processed into mature miRNAs, and target repression is accomplished by miRNA-directed AGO function. Nevertheless, there are specific characteristics of animal miRNAs that are not observed in plants. First, a large fraction of miRNA genes in animals are clustered together and generate polycistronic precursors. Moreover, miRNA genes may reside within the

transcriptional units of other genes and consequently depend on the transcription of these genes for their own expression. Following the transcription of miRNA genes in animals, the processing of the precursors involves two RNase III enzymes, Drosha and Dicer. These enzymes perform two independent dicing events, in contrast to DCL1 in plants, which performs both dicing steps. Animal pri-miRNAs are initially processed into pre-miRNAs by Drosha within the nucleus, and the pre-miRNAs are subsequently transported to the cytoplasm by Exp-5. Further processing of pre-miRNAs to generate the miRNA/miRNA* is performed by Dicer in the cytoplasm. In contrast to the methylation of plant miRNAs by HEN1, the miRNA/miRNA* duplex is not methylated in animals. Stable incorporation of the mature miRNA into Ago1 subsequently directs target repression.

In animals, miRNA-directed target repression generally occurs through translational inhibition and RNA decay. Transcript cleavage is not widely observed for animal miRNAs, which may reflect the relatively low complementarity between animal miRNAs and their targets (reviewed in [49,65]). miRNA binding sites are typically located in the 3' untranslated region (3' UTR) of target transcripts, with single mRNA molecules bound by multiple miRNAs for silencing. As previously described, miRNA-mediated PTGS in plants requires high sequence complementarity at target binding sites, which are generally found within coding sequences. In animals, perfect complementarity of the miRNA and its target within the seed region alone (the 2nd to 7th nucleotides from the 5' end of the miRNA) is sufficient for target recognition. As a result, a single miRNA generally has a large number of mRNA targets. Thus, miRNA-mediated silencing in

animals involves a less stringent hybridization requirement between miRNAs and their targets, and target transcript cleavage is uncommon in animals.

Evolution of *MIR* genes

High-throughput deep sequencing of small RNA populations has led to the discovery of many miRNA species. The miRNA database (www.mirbase.org, Release 19) [56] currently lists 338 mature miRNAs in *A. thaliana*, 2042 in human and 368 in *C. elegans* in addition to the thousands of mature miRNAs from 190 other species. Despite the large number of identified miRNAs in both plants and animals, the poor conservation of miRNA sequences between the two kingdoms indicates that miRNA families in plants and animals evolved independently [27].

There are two major hypotheses regarding the evolution of miRNAs. The first proposes that *MIR* genes evolved from inverted duplications of miRNA targets [66]. In this model, the duplication of protein-coding genes in a head-to-head or tail-to-tail orientation yields stem-loop structures whose stem regions exhibit extensive base-pair complementarity and are processed by the siRNA pathway rather than the miRNA pathway [24,67]. Small RNAs from these young *MIR* genes regulate their homologous targets. Over evolutionary time, both *MIR* genes and their targets may undergo duplication and accumulate mutations. Some of the double-stranded precursors eventually acquire the characteristic hairpin structure of miRNA precursors and are processed by the miRNA pathway rather than the siRNA pathway. Over time, the targets come to be regulated by a limited number of specific small RNAs. According to this

model, recently evolved *MIR* genes are expected to have a higher degree of sequence similarity with their targets, which has been observed for *MIR161* and *MIR163*.

Because the precursors of many young miRNAs do not match any other sequences in the genome, the duplication hypothesis cannot explain the genesis of all of the young miRNAs in plants. The random hairpin theory was proposed in part to address this shortcoming [68,69]. Organisms produce a large number of hairpin structures that could potentially generate foldback precursors of small RNAs. For example, the *A. thaliana* genome contains more than 130,000 imperfect inverted repeats. The random hairpin theory proposes that *MIR* genes can evolve when the following conditions are met: a DNA segment that generates a foldback structure retains a transcriptional unit; by chance, a small RNA produced from the structure targets a protein-coding gene; and the resulting regulatory relationship confers an evolutionary advantage.

The classification of a miRNA as old or young is based on its degree of conservation among different species [27]. While ancient miRNAs are conserved among animal or plant species of great evolutionary distance, young miRNAs are specific to a species or genus. miRNA genes may be duplicated as a result of gene duplication, whole segment duplication or genome duplication, thereby giving rise to a miRNA family [70]. The miRNAs produced by these families are considered old miRNAs, and their abundance is high probably because they are encoded by multiple genes. The processing of the precursors of old miRNAs by the DCL1-mediated biogenesis pathway is more precise than the processing of young miRNA precursors. Having multiple *MIR* genes within each conserved family may result in a complex relationship among the individual

miRNA family members and their targets [67]. In contrast, non-conserved miRNAs tend to be encoded by a single locus and are characterized by their short evolution times. In addition to being weakly expressed, young miRNAs typically regulate few, if any, genes, and the processing of their precursors is less precise.

Heterochromatic siRNAs

siRNAs are small RNAs 21-24 nt in length that are generated from long dsRNAs or single-stranded RNAs (ssRNAs) that produce longer and more perfect hairpin structures compared to miRNA precursors. While only one miRNA is produced from a single precursor, siRNA precursors generate multiple siRNAs. siRNAs trigger TGS or PTGS and direct the enzymatic action of their effector proteins through sequence-specific interactions with their targets.

Plants have two major families of endogenous siRNAs: heterochromatic siRNAs (hc-siRNAs) and trans-acting siRNAs (ta-siRNAs). siRNAs may also derive from exogenous sources, such as viruses and transgenes.

Biogenesis of hc-siRNAs

Hc-siRNAs are small RNAs 21-24 nt in length that derive from heterochromatic regions, including repeats, transposons and intergenic regions (reviewed in [71,72]). They comprise the most abundant and diverse small RNA family: approximately 80 percent of all small RNAs are hc-siRNAs, with tens of thousands of unique hc-siRNAs present in wild-type *Arabidopsis* [9,11]. Hc-siRNAs mediate TGS of heterochromatin by guiding

DNA methylation and histone modification in a sequence-specific manner. In *Arabidopsis*, the biogenesis of hc-siRNAs involves the transcription of primary precursors, the conversion of the precursors to dsRNAs, further maturation by dicing, methylation and association with AGO proteins.

That the sequences of hc-siRNAs often map to repeats and transposons implies that hc-siRNA precursors are transcribed from these regions. In *Arabidopsis*, this transcription is probably performed by the plant-specific DNA-dependent RNA polymerase Pol IV (reviewed in [73]). Although Pol IV is evolutionarily derived from Pol II, the largest subunit of Pol IV, NUCLEAR RNA POLYMERASE D1 (NRPD1), is distinct from that of Pol II and confers the specific catalytic activity of Pol IV [74,75]. Pol IV-dependent transcripts have not been detected experimentally, but more than 90 percent of hc-siRNAs are lost in the *nrpd1* mutant [9,11]. The current model suggests that Pol IV produces long ssRNAs from regions that spawn siRNAs. The chromatin remodeling factor CLASSY1 (CLSY1) may promote hc-siRNA biogenesis through Pol IV [76], while the homeodomain protein SAWADEE HOMEODOMAIN HOMOLOG1 (SHH1) has been implicated in the recruitment of Pol IV to regions that produce hc-siRNAs [77].

siRNAs are typically generated from the cleavage of long dsRNAs into smaller fragments of a precise length. Hc-siRNAs are produced in this manner, following the conversion of Pol IV-dependent ssRNA transcripts into dsRNAs. Of the six RNA-dependent RNA polymerase (RDR) homologs encoded in the *Arabidopsis* genome, RDR2 is responsible for this conversion [10]. RDR2 physically interacts with Pol IV and

converts Pol IV-transcribed ssRNAs into long dsRNAs with perfect sequence complementarity [77]. Thereafter, the siRNA precursors are diced into 24-nt small RNAs with 2-nt overhangs by DCL3 and methylated at the 2'-OH group of the 3' terminal nucleotide by HEN1 [19,35,40]. A mutation disrupting DCL3 function can be compensated by other DCL homologs: in *dcl3*, DCL2 and DCL4 generate hc-siRNAs 22 nt and 21 nt in length, respectively [78]. As previously mentioned, the siRNA precursors processed by DCL3 yield multiple siRNAs rather than a single species from a locus.

After processing, mature hc-siRNAs are loaded into AGO effector proteins belonging to the AGO4 clade [79]. Although hc-siRNAs are synthesized in the nucleus, their abundance is ten times greater in the cytoplasm than in the nucleus [80]. It has been reported that hc-siRNAs are incorporated into AGO4 in the cytoplasm with the assistance of HEAT-SHOCK PROTEIN90 (HSP90) and then transported to the nucleus. Four of the ten AGO homologs in *Arabidopsis*, AGO4, AGO6, AGO8 and AGO9, belong to the AGO4 clade. All of these except AGO8 function in hc-siRNA-mediated TGS, and AGO4 is the major binding partner of hc-siRNAs [81]. AGO6 and AGO9 have been reported to function as hc-siRNA effector proteins in specific cell types and organs [81-84]. AGO8 may be a pseudogene [37].

Molecular mechanism of hc-siRNAs

As the name implies, hc-siRNAs are generated from heterochromatin, and they guide cytosine methylation at the site of their transcription. In *Arabidopsis*, hc-siRNAs play a major role in determining the methylated targets of the RNA-directed DNA

methylation (RdDM) pathway (reviewed in [71,72]). In RdDM, two classes of ncRNAs and a methyl transferase enzyme are required for target selection and catalytic activity, respectively. A number of subsidiary players have also been implicated in RdDM.

In addition to hc-siRNAs, another type of ncRNA is also generated from RdDM target regions and helps direct cytosine methylation [85]. The biogenesis of these long ncRNAs requires a second plant-specific RNA polymerase, Pol V (reviewed in [73]). Like Pol IV, Pol V is evolved from Pol II, and its largest subunit, NUCLEAR RNA POLYMERASE E1 (NRPE1), is distinct from those of Pol II and Pol IV [74,75]. The recruitment of Pol V to its targets in the genome is facilitated by the DDR complex, whose major components are DEFECTIVE IN MERISTEM SILENCING3 (DMS3), a structural maintenance of chromosome (SMC) domain protein; DEFECTIVE IN RNA-DIRECTED DNA METHYLATION1 (DRD1), an SNF2-like chromatin remodeling protein; and REQUIRED FOR DNA METHYLATION1 (RDM1), a single-stranded methyl DNA-binding protein [86]. At a subset of the methylated targets of RdDM, Pol II is responsible for synthesizing the long ncRNA rather than Pol V [87].

AGO4-loaded hc-siRNAs are recruited to their targets through two distinct interactions: physical contact between AGO4 and WG/GW motif-containing proteins and the binding of hc-siRNAs to Pol V-dependent ncRNAs. The WG/GW motif is an AGO hook motif found in AGO4-binding proteins, and the interaction between AGO4 and WG/GW motif-containing proteins governs downstream molecular events (reviewed in [88]). It has been proposed that hc-siRNAs recognize the nascent Pol V-dependent ncRNA through base-pair complementarity and guide silencing of the target DNA in a

sequence-specific manner [89]. NRPE1 contains a WG/GW motif in its C-terminal region and is known to physically interact with AGO4 [90]. Thus, an hc-siRNA-containing AGO4 protein may be shuttled to the target region through the interaction with NRPE1 while the hc-siRNA recognizes the nascent, long ncRNA generated by Pol V. Another WG/GW motif-containing protein, SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5L)/KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR1 (KTF1), interacts with both Pol V-dependent ncRNA and AGO4 [91]. Thus, SPT5L/KTF1 may stabilize the hybridization of AGO4-loaded hc-siRNAs and Pol V-dependent ncRNAs by bridging AGO4 and Pol V-dependent ncRNAs.

These complex interactions among proteins and ncRNAs must ultimately provide a stable foundation for the recruitment of a methyltransferase enzyme to the hc-siRNA targets. It has been proposed that the binding of AGO4-loaded hc-siRNAs to Pol V-dependent ncRNAs is followed by the release of AGO4. The INVOLVED IN DE NOVO2 (IDN2)/ REQUIRED FOR DNA METHYLATION12 (RDM12)-containing complex is critical for the consolidation and integration of factors required for the downstream methylation event [92-94]. In addition to IDN2/RDM12, the complex contains an IDN2/RDM12 paralog, either FACTOR OF DNA METHYLATION1 (FDM1)/IDN2-LIKE1 (IDNL1)/IDN2 PARALOG1 (IDP1) or FDM2/IDNL2/IDP2. The protein domains of IDN2/RDM12 and its two paralogs in *Arabidopsis* occur in the following order from the N-terminus: a zinc finger for RNA and/or DNA binding, an XS domain for dsRNA recognition and a coiled-coil domain and XH domain for protein dimerization. The zinc finger domain may bind the methylated target DNA or function as

a second RNA binding motif alongside the XS domain. XS domains bind dsRNAs with 5' overhangs; in the context of RdDM, the XS domain may stabilize the duplex formed by hc-siRNAs and Pol V-dependent ncRNAs. IDN2/RDM12 dimerizes with FDM1/IDNL1/IDP1 or FDM2/IDNL2/IDP2 through the coiled-coil and XH domains in an antiparallel manner, permitting the recruitment of two distinct hc-siRNAs in a single IDN2/RDM12-containing complex. The IDN2/RDM12-containing complex may anchor the dsRNA duplex formed by hc-siRNAs and long ncRNAs to the target DNA. While the recruitment of DOMAIN REARRANGED METHYLTRANSFERASE2 (DRM2) to the target remains unclear, the RdDM effector protein RDM1 has been proposed to mediate this recruitment based on its ability to physically interact with AGO4 and DRM2 [95]. Methylation by DRM2 involves the deposition of a methyl group to the fifth carbon of cytosine residues [96]. Considering these complex interactions in aggregate, hc-siRNAs ultimately guide DRM2, which is responsible for the methylation of the target regions.

After *de novo* methylation of the siRNA target regions by RdDM, cytosine methylation is maintained by different methyltransferase enzymes based on the sequence context of the methylated cytosines (reviewed in [72,97]). The three possible sequence contexts of cytosine are the symmetric CG and CHG contexts and the asymmetric CHH context, where H stands for A, C or T. Following DNA replication, fully methylated symmetric cytosines are hemi-methylated: the template DNA strand maintains the methylated cytosine, while the newly synthesized DNA strand is unmethylated. Using the methylated cytosine in the template as a methylation cue, CG and CHG cytosines in the nascent DNA are methylated by METHYLTRANSFERASE1 (MET1) and

CHROMOMETHYLTRANSFERASE3 (CMT3), respectively [98,99]. The enzymatic activity of MET1 is facilitated by the chromatin remodeling factor DECREASED IN DNA METHYLATION1 (DDM1), and there is some degree of cross-talk between CMT3 and another repressive chromatin modifier, SUPPRESSOR OF VARIEGATION3-9 HOMOLOGUE4/KRYPTONITE (SUVH4/KYP), which mediates histone H3K9 methylation [100-102]. In contrast to CG and CHG residues, CHH residues require hc-siRNAs for methylation maintenance after DNA replication.

Biological function of hc-siRNAs

Hc-siRNAs and the RdDM pathway determine the methylation landscape in the genome. Genome-wide analyses in *Arabidopsis* have revealed a high degree of overlap among regions containing transposons and repeats, cytosine methylation and hc-siRNAs in terms of their distribution and abundance [103,104]. By guiding methylation at repeats and transposons, hc-siRNAs are critical for the maintenance of genome integrity and gene expression regulation.

Transposons and repeats occupy large portions of the *Arabidopsis* genome (reviewed in [105,106]). Some of these elements have the ability to jump to other regions or to amplify themselves, which may disrupt functional genes or be detrimental to the organization of the host genome. However, there are several defense mechanisms that protect the genome from the movement or amplification of transposons and repeats. For example, epigenetic modifications such as cytosine methylation help silence and immobilize repeats and transposons. Reduced cytosine methylation and derepression of

the expression of transposons and repeats have been observed in loss-of-function RdDM pathway mutants. Similarly, loss of MET1 and DDM1 function leads to a reduction in cytosine methylation and induces amplification and mobilization of some transposons [107-109].

Some repeats and transposons are located in intergenic regions, particularly in the promoters of protein-coding genes, and generate 24-nt small RNAs. These mobile elements are regulated by RdDM, and their methylation level affects the expression of nearby genes. As one example, two tandemly arranged repeats are found within the promoter of the gene encoding the homeobox transcription factor FLOWERING WAGENINGEN (FWA) and are highly methylated in most tissue types in wild-type *Arabidopsis* [110]. Accordingly, *FWA* is transcriptionally silenced in these tissues. However, *FWA* is actively transcribed in the endosperm, where an active demethylation mechanism removes methylated cytosines from the repeats of the maternal *FWA* promoter [111]. In one epigenetic *fwa* mutant, the repeats in the promoter region are hypomethylated, resulting in the transcriptional activation of *FWA* and late flowering.

Hc-siRNAs and reproductive growth

During the reproductive growth stage of *Arabidopsis*, the embryo and endosperm are characterized by opposing hc-siRNA-mediated DNA methylation programs. This difference in DNA methylation is also observed between the gametes and their supporting cells (reviewed in [71,72,112]). The female gametophyte contains the egg cell, the central cell and five accessory cells, while the male gametophyte contains two sperm

cells and one enlarged vegetative cell. The process of double fertilization in angiosperms involves the fertilization of both the egg cell and the central cell by the two sperm cells, thereby producing the embryo (zygote) and the endosperm, respectively. As companion cells, the central cell and the vegetative cell support the development of their adjacent cells, the egg cell and the sperm cells, respectively. Similarly, the endosperm supports the development of the zygote. The gametes and zygote exhibit a sharp contrast with their respective companion cells and the endosperm in terms of hc-siRNA biogenesis and DNA methylation. While the nursing cells lose CG DNA methylation and exhibit increased expression of transposons and siRNAs, the gametes and zygote maintain their CG methylation and other repressive marks at repeats and transposons [113,114]. It has been proposed that the decrease in CG methylation in companion cells and the increased transcription from transposons and transcribed RNAs enlarge the siRNA pools [115,116]. Subsequently, transposon-specific siRNAs may be transported from the companion cells and endosperm to the gametes and zygote to enhance transposon silencing through cytosine methylation. Whereas a germ line is established during the early stages of animal embryogenesis, the differentiation of germ cells from somatic stem cells occurs late in the plant life cycle. In effect, the changes in DNA methylation in the companion cells, whose genetic material is not transferred to the next generation, may help overcome problems resulting from the delayed establishment of the germ cells in plants and ensure the integrity of the parental genomes transferred to the offspring. DDM1 and MET1 are repressed in the vegetative cell and the central cell, respectively, resulting in a global decrease in cytosine methylation [115,117]. In the companion cells of both gametes,

DEMETER (DME), an active demethylase enzyme, further reduces the level of methylation through the demethylation of methylated cytosines [113,114,118]. Transcripts are generated from the demethylated transposons and made double-stranded. The resulting precursor dsRNAs are further processed into 21- and 24-nt small RNAs in the vegetative and central cells, respectively. Expression of a GFP transgene and an artificial transgene-targeting miRNA in the sperm and the vegetative cell, respectively, leads to the suppression of the GFP expressed in the sperm [115]. A similar outcome is also observed in the egg cell and central cell, indicating that siRNAs produced in one cell can move into an adjacent cell and induce silencing during reproductive growth in *Arabidopsis* [114].

piRNAs in animals

A specialized class of small RNAs known as piRNAs is enriched in animal germ line cells and is also present in somatic cells (reviewed in [71,119-121]). piRNAs are 25-30 nt in length and are incorporated into an AGO protein belonging to the PIWI clade. These small RNAs guide heterochromatin formation in a manner similar to hc-siRNAs. Unlike small RNAs in plants, however, piRNA biogenesis is Dicer-independent and entails a 'ping-pong' mechanism of primary biogenesis and amplification; whether one or both of these two processes occurs depends on the cell type. *Drosophila* germ cells require three members of the PIWI protein subfamily for piRNA-mediated genome protection: AGO3, Aubergine (AUB) and PIWI. Transposon fragments or relics aggregate into large clusters that generate piRNAs, and the transcription of these piRNA

clusters generates long ssRNAs that are antisense to the transposons. During primary processing, these transcripts are processed into antisense piRNAs. These antisense piRNAs exhibit a uridine bias at their 5' ends and bind AUB or PIWI. During the amplification phase, sense RNAs from transcribed transposons are recognized and cleaved by AUB-loaded antisense piRNAs, generating sense piRNAs. Transposon-specific sense piRNAs exhibit an adenosine bias at the 10^{th} nucleotide from the 5' terminus and are incorporated into AGO3. Finally, AGO3-loaded sense piRNAs trigger the biogenesis of antisense piRNAs from long piRNA cluster transcripts, and the ping-pong cycle is reiterated. As sense RNAs from transposons are consumed during the ping-pong cyle, the ping-pong pathway promotes the post-transcriptional silencing of targeted transposons. piRNA-guided slicing requires the endonucleolytic activity of the PIWI-family proteins. Furthermore, piRNAs are stabilized by HEN1-mediated methylation. In addition to the repression of transposons at the posttranscriptional level via transcript slicing, piRNAs also guide the deposition of repressive epigenetic marks at homologous chromatin to induce transcriptional silencing.

ta-siRNAs

siRNAs such as hc-siRNAs function at their origin or at homologous regions. Thus, hc-siRNAs act *in cis*, as their sources coincide with their targets. In contrast, trans-acting siRNAs (ta-siRNAs) function at loci distinct from the site of their biogenesis. As their name indicates, ta-siRNAs act *in trans*, and their regulatory mechanism is similar to that of miRNAs.

Biogenesis of ta-siRNAs

Arabidopsis ta-siRNAs are small regulatory RNAs 21 nt in length whose precursors are generated from *TAS* loci (reviewed in [122]). Their biogenesis is clearly distinct from that of other small RNAs: transcripts generated from *TAS* loci are targets of miRNA-directed cleavage, and the cleavage products serve as the sources for the biogenesis of secondary siRNAs. The biogenesis of the secondary siRNAs is similar to that of hc-siRNAs, following the conversion of the single-stranded cleavage products into dsRNA precursors.

There are four groups of *TAS* genes in *Arabidopsis*. *TAS1* and *TAS3* are each encoded by three isoforms: *TAS1a*, *TAS1b* and *TAS1c* for *TAS1* and *TAS3a*, *TAS3b* and *TAS3c* for *TAS3*. *TAS2* and *TAS4* are transcribed from single loci. *TAS* transcripts are non-protein-coding RNAs and contain RNA sequences that are recognized and regulated by miRNAs. miR173 targets both *TAS1* and *TAS2*, and miR390 and miR828 target *TAS3* and *TAS4*, respectively [24,123,124]. *TAS* transcripts are subject to miRNA-directed cleavage carried out by miRNA-bound AGO1 or AGO7. The stabilization of the cleavage products involves SUPPRESSOR OF GENE SILENCING3 (SGS3), and the ssRNA products are made double-stranded by RDR6 [124-126]. The resulting long dsRNA precursors are processed into 21-nt small RNAs by DCL4 in a precisely phased manner: DCL4 successively cleaves the dsRNA precursor beginning at one end and generating multiple small RNAs at 21-nt intervals [124,127,128]. Although diverse small RNAs are generated from a single dsRNA precursor, only some of the small RNAs are stable and

incorporated into AGO1 [39]. Like miRNAs, ta-siRNAs are methylated by HEN1, and their downstream function is similar to that of miRNAs.

Although hundreds of miRNAs have been identified in *Arabidopsis*, only three miRNAs (miR173, miR390 and miR828) are known to initiate the biogenesis of ta-siRNAs. Several studies have revealed specific factors that influence the initiation of secondary siRNA biogenesis: a specialized AGO protein, the length of the miRNAs, the structure of the miRNA/miRNA* duplex, the position of the miRNA binding site within the target RNA and the degree of sequence complementarity between the miRNAs and their targets. Among the ten AGO homologs in *Arabidopsis*, AGO7 is the only family member that can generate ta-siRNAs from *TAS3* [41]. AGO7 exclusively binds miR390, which has two target sites in the *TAS3* transcript. The 3' miRNA target site has nearly perfect complementarity with miR390, and AGO7-mediated cleavage at this site triggers ta-siRNA biogenesis. However, ta-siRNA biogenesis at the *TAS3* locus also requires the 5' target site of miR390 in the *TAS3* transcript, as described by the two-hit trigger model. Although the 5' miR390 target site is resistant to AGO7 cleavage, AGO7 must be recruited to both the 5' and 3' target sites to initiate ta-siRNA biogenesis from *TAS3* transcripts.

miRNA length is another determinant of ta-siRNA biogenesis [129,130]. While the majority of miRNAs are 21 nt in length, ta-siRNA-generating miRNAs are 22 nt in length with the exception of miR390, which is 21 nt in length. In a transient expression study in *Nicotiana benthamiana*, artificially engineered miRNAs (miR173, miR472 and miR828) 21 or 22 nt in length were tested for their ability to trigger secondary siRNA

biogenesis from a co-infiltrated target construct. Only the 22 nt forms of the miRNAs successfully triggered secondary siRNA biogenesis.

The asymmetric structure of the miRNA/miRNA* duplex has also been found to affect the initiation of ta-siRNA biogenesis [131]. In a transient system similar to that described in the preceding paragraph, four artificial miR173/miR173* duplexes were examined: 22/21-, 21/22- and 21/21-nt duplexes with asymmetric bulges along with a symmetric 21/21-nt duplex. The experiment showed that duplexes with asymmetric bulges could generate secondary siRNAs, regardless of their length. High-throughput sequencing techniques have identified more pairs of miRNAs and mRNAs that generate secondary siRNAs. When miRNAs were found to induce the production of secondary siRNAs from target mRNAs, the miRNA/miRNA* duplexes were found to contain a 22-nt strand and a 21-nt strand (i.e., either 21/22- or 22/21-nt duplexes). Additionally, the miRNA/miRNA* duplexes tended to be asymmetrically bulged in terms of their structure. It has been proposed that the AGO1-containing RNA-induced silencing complex (RISC) induces either target repression or secondary small RNA biogenesis based on the structure of the miRNA/miRNA* duplex.

Two additional factors that influence the initiation of ta-siRNA biogenesis are the location of the miRNA-binding site in the target transcript and the degree of sequence complementarity between the miRNA and its target site [132]. In a study in which ta-siRNAs were generated in plants from a synthetic GFP reporter by miR173, the efficiency of ta-siRNA generation was maximal when the miR173-binding site was located immediately after the stop codon. When premature stop codons were introduced

further upstream of the miR173-binding site, the abundance of ta-siRNAs decreased while the abundance of longer transcripts from the transgene increased. These observations suggest a link between translation by ribosomes and *TAS* precursor processing. Finally, reduced complementarity at the 3' end of synthetic miR173 has been shown to abolish the generation of ta-siRNAs, while reduced complementarity at the 5' end has a less detrimental effect.

Molecular mechanism and biological function of ta-siRNAs

As observed for miRNAs, ta-siRNAs regulate the expression of their target genes *in trans* [125,126]. Furthermore, ta-siRNAs are 21 nt in length, associate with AGO1 and direct PTGS of their targets [39]. *TAS1* ta-siRNAs target several uncharacterized genes and multiple mRNAs encoding pentatricopeptide repeat (PPR) proteins. *TAS2* ta-siRNAs similarly target *PPR* mRNAs [123]. *PPR* genes are commonly found in eukaryotes and appear to have undergone a rapid expansion in plants; the *Arabidopsis* genome, for example, encodes 448 *PPR* genes [133]. Some PPR proteins bind RNA and are predicted to regulate gene expression through RNA processing, editing, stability and translation in mitochondria and chloroplasts. Although *PPR* genes are targeted by *TAS1* and *TAS2* ta-siRNAs, the biological relevance of these regulatory interactions remains unclear. Interestingly, *PPR* genes are targeted by both ta-siRNAs and miRNAs, and some transcripts contain multiple small RNA-binding sites [134,135]. Considering the large number of *PPR* genes in plants, ta-siRNAs and small RNAs may serve to dampen the detrimental effects caused by the rapid expansion of gene families.

TAS3 ta-siRNAs modulate auxin signaling networks by targeting *AUXIN RESPONSE FACTOR2* (*ARF2*), *ARF3* and *ARF4* and are thus referred to as tasiR-ARF [123,136]. Auxin is a major plant hormone and is involved in every phase of plant development (reviewed in [137]). Although auxin may affect growth and development through numerous mechanisms, the mechanism that is best understood is auxin-mediated regulation of gene expression through the ARF and Aux/IAA proteins. In the basal condition, ARF proteins are bound and repressed by Aux/IAA proteins, and auxin-responsive genes are not expressed. When the level of auxin is increased, Aux/IAA is degraded by the ubiquitin-mediated proteasome pathway, and ARF proteins are released from Aux/IAA repression. ARFs recognize auxin-responsive elements in the promoters of downstream genes and activate their expression. Two of the diverse developmental processes affected by tasiR-ARF are phase transition and leaf pattern formation. When the tasiR-ARF binding site in *ARF3* is mutated to make *ARF3* resistant to tasiR-ARF, juvenile plants enter the adult phase prematurely, which is similarly observed in ta-siRNA biogenesis mutants such as *ago7*, *sgs3* and *rdr6* [125,136]. Thus, tasiR-ARF suppresses the juvenile-to-adult phase transition. In terms of leaf development, tasiR-ARF is expressed in the adaxial (upper) leaf region, and its movement to the abaxial (lower) region generates a concentration gradient of tasiR-ARF (reviewed in [138]). *ARF3* is expressed throughout the leaf primordia, and *ARF4* RNA is detected in abaxial leaf tissue [139,140]. Due to the higher concentration of tasiR-ARF in the adaxial region, ARF activity is higher in or restricted to the abaxial leaf region. Thus, the pattern of ARF

activity across the adaxial and abaxial regions contrasts that of tasiR-ARF accumulation, and these distinct gradients are critical for polarized leaf pattern formation.

Lastly, *TAS4* ta-siRNAs are predicted to repress genes encoding MYB transcription factors [24]. However, the *TAS4* ta-siRNAs were the last to be identified, likely owing to their low abundance, and their function is currently unknown.

Exogenous siRNAs

In addition to endogenously produced small RNAs, plants also contain small RNAs that derive from exogenous sources. In fact, small RNAs were first detected in plants that were infected with viruses and plants harboring transgenes [141]. This pioneering discovery revealed the first clues that small RNAs play an important role in the repression of viruses and transgenes and revolutionized the field of RNA silencing.

Viral siRNAs (viRNAs)

Plants have adopted a small RNA-mediated repression mechanism to combat viral infection (reviewed in [142]). After infection, plant DCL enzymes generate primary viRNAs from viral dsRNAs, which are produced by viral RDR during replication, by intramolecular hybridization or by convergent transcription. Primary viRNAs elicit the biogenesis of secondary viRNAs in a manner similar to that of ta-siRNA biogenesis: viral target RNAs are cleaved, the cleavage products are made double-stranded by plant RDRs, and DCL enzymes cleave the newly generated double-stranded precursors. Amplified viRNAs are incorporated into AGO proteins and repress the virus through PTGS.

Multiple DCLs, RDRs and AGOs in host plants have redundant functions, work in tandem and/or perform a specialized function to defend plants against viral infection. The activities of these proteins also depend on the type of viral infection.

In response to the antiviral defense of the host plant, viruses have also developed mechanisms to counteract the host response. Numerous viruses encode viral suppressors of RNA silencing (VSRs) that oppose the repressive action of viRNA-mediated silencing in the host (reviewed in [143]). Specifically, VSRs intercept viral dsRNAs or silencing factors generated by the host plant. P19 from *Cymbidium ringspot virus* and P21 from *Beet yellows virus* sequester short dsRNAs, effectively disrupting RISC assembly with viRNAs in the host [144,145]. Other VSRs are capable of binding AGO proteins: 2b from *Cucumber mosaic virus*, P0 from *Beet polerovirus*, P1 from *Sweet potato mild mottle virus* and P38 from *Turnip crinkle virus*. 2b inhibits the slicing activity of AGO1 in pre-assembled RISC [146]. P0 contains a minimal F-box motif that may induce AGO1 degradation through ubiquitin-dependent proteolysis [147]. P1 and P38 contain an AGO-hook GW/WG motif and may therefore compete with endogenous AGO-binding proteins in plants [148-150]. Other components of the plant silencing machinery are also targeted by viruses. For example, the binding of V2 from *Tomato yellow leaf curl virus* to SGS3 compromises RDR6-mediated secondary viRNA biogenesis [151]. Additionally, HC-Pro from *Zucchini yellow mosaic virus* disrupts the methylation of small RNAs by HEN1 [152].

siRNAs from transgenes

Early studies in which plants were transformed with sense transgenes revealed the suppression of both transgenes and endogenous homologous genes in several transgenic lines, and the silencing phenomenon was termed co-suppression [153,154]. Subsequent studies revealed that transgene-specific small RNAs accumulate in silenced plants and that proteins required for small RNA biogenesis and action are also involved in transgene silencing. In cases of transgene silencing, the ssRNA transcripts generated from the transgene are recognized by the plant machinery as aberrant and made double-stranded by RDR6 [155]. The dsRNA subsequently triggers downstream events, including primary and secondary siRNA biogenesis. As a result, both the transgene and endogenous homologous genes are targeted for silencing.

Small RNA turnover

Consistent with the critical roles of small RNAs in diverse biological processes, the abundance of small RNAs is also precisely regulated. Disrupting the homeostasis of small RNAs detrimentally affects developmental and metabolic processes. Because the abundance of small RNAs is affected by both internal and external signals, the balanced expression of small RNAs requires a precise regulatory mechanism. In plants, small RNA biogenesis and turnover are the critical phases for regulating the dynamics of small RNA populations (reviewed in [156]).

The methylation of small RNAs during biogenesis is crucial for their stabilization [35]. Small RNAs in *Arabidopsis* are methylated at the 2'-OH of the 3' terminal ribose by

HEN1. In *hen1* mutants, the abundance of small RNAs is dramatically reduced, and the residual small RNAs are tailed or trimmed. High-throughput sequencing of the small RNA population in *hen1* further revealed that the residual small RNAs are identical at their 5' ends but heterogeneous at their 3' ends. Specifically, the small RNAs were found to have oligonucleotide tails 1 to 7 nt in length, with a predominant enrichment of uridine among the four nucleotides [157,158]. Furthermore, truncation from the 3' terminus was observed for both intact and uridylated small RNAs in the *hen1* mutant. Thus, HEN1-mediated methylation at the 3' end of small RNAs ultimately inhibits their degradation.

In *Arabidopsis*, the SMALL RNA-DEGRADING NUCLEASE (SDN) family of 3'-5' exonucleases is responsible for small RNA degradation [159]. When multiple *SDN* genes are simultaneously knocked down, increased miRNA levels and pleiotropic developmental defects are observed. SDN1 specifically degrades 17- to 27-nt single-stranded small RNAs, and its activity is partially inhibited by 2'-*O*-methylation at the 3' end of small RNAs. Based on these observations, the removal of the 3' most nucleotide by SDNs may be rate-limiting and probably requires other assistant proteins or the combined activity of multiple SDNs.

From both forward and reverse genetic studies of the *hen1* mutant, HEN1 SUPPRESSOR1 (HESO1) was found to poly-uridylate small RNAs in *hen1* mutants [158,160]. In contrast to the protective function of methylation, uridine tails at the 3' end of small RNAs make miRNAs more susceptible to 3'-5' exonuclease activity. Consistent with the hypothesis that a defect in uridylation activity should rescue the loss of methylation in *hen1* (i.e., unmethylated small RNAs that do not undergo uridylation

should be less susceptible to 3'-5' exonuclease activity), a mutation in *HESO1* suppresses the morphological defects of *hen1* mutants. Compared to *hen1*, miRNA levels are increased in *hen1 heso1*. However, tailed and trimmed miRNAs are still observed in the double mutant. Based on *in vitro* analysis, HESO1 has terminal nucleotidyl transferase activity with a preference for uridine substrates, and HESO1 function is completely inhibited by 2'-*O*-methylation at the 3' end of small RNA substrates. High-throughput small RNA data for the *hen1 heso1* double mutant reveal shorter uridine tails in the double mutant compared to *hen1*, which further suggests that HESO1 is partially responsible for uridylation in the *hen1* mutant.

In addition to HEN1, SDN exonucleases and HESO1, long RNA molecules may influence the rate of degradation of specific small RNAs. In a technical analysis of target mimicry, a short tandem target mimic (STTM), composed of two short sequences mimicking small RNA-binding sites tandemly arrayed with an optimal spacing between them, was found to reduce the abundance of miRNAs whose binding sites were mimicked by the STTM [161]. Interestingly, the reduction in miRNA abundance was dependent on SDN1 and SDN2 activity. Similarly, although less effectively, other artificial target mimicry transgenes led to reductions in the levels of cognate miRNAs. This suggests that target transcripts, especially those that cannot be cleaved by miRNAs, impact the stability of miRNAs. This raises the question of whether such targets exist naturally.

In *Arabidopsis*, miR399 recognizes two target RNAs: the mRNA transcript corresponding to the E2 ubiquitin conjugation enzyme PHOSPHATASE (PHO2) and the

INDUCED BY PHOSPHATASE STARVATION1 (*IPS1*) ncRNA [162,163]. miR399 is induced under phosphate (Pi) starvation conditions and represses the activity of PHO2 through mRNA cleavage as an adaptive response that alters the metabolism of Pi in plants. In general, signaling cascades triggered by a certain event or treatment are eventually attenuated, and the recovery of steady expression levels facilitates the response to a prolonged stimulus. In a similar way, *PHO2* is temporarily silenced by miR399 under Pi starvation conditions but eventually achieves a steady level of activity, which is mediated by target mimicry. Long *IPS1* ncRNAs are also induced by Pi deficiency and sequester miR399 from *PHO2* mRNAs. Unlike *PHO2* mRNA, which is subject to miRNA-directed cleavage, *IPS1* ncRNAs are bound but not sliced by miR399 due to a mismatch at the cleavage site. Although *IPS1* ncRNAs do not alter the *in vivo* abundance of miR399, they suppress the effect of miR399 on *PHO2*.

A recent study identified many *IPS1*-like intergenic long ncRNAs that can pair with other miRNAs. Overexpression of some of the long ncRNAs led to a decrease in the abundance of the cognate miRNAs, raising the possibility that long ncRNAs regulate the stability of specific miRNAs *in vivo* [164].

Future directions

While the overall framework of miRNA biogenesis is relatively well established, many aspects of the regulation of miRNA biogenesis remain unclear. The abundance of mature miRNAs is regulated by Pol II-mediated transcriptional regulation and during the processing of pri-miRNAs to mature miRNAs. However, it is also possible that the

processing of miRNA precursors is also directly affected by the endogenous or exogenous signals integrated by Pol II. Furthermore, unique factors may differentially regulate certain miRNA species during the process of miRNA maturation. In terms of the activities directed by miRNAs, the molecular mechanisms of mRNA cleavage and translational inhibition require further study. It has been proposed that the extent of sequence complementarity between miRNAs and their targets dictates the mode of repression by miRNAs. However, this is unlikely because miRNAs with a high degree of sequence complementarity to their targets have also been shown to act via translational repression. In fact, the two modes of action may occur simultaneously for a given miRNA-target pair. The degree of miRNA-target complementarity that is required for translational repression has not been experimentally determined. If less extensive base pairing is sufficient to induce translational inhibition as observed in animals, the current views of the regulatory networks between miRNAs and their targets would need to be reevaluated. The translational repression activity of plant miRNAs also needs to be dissected at the mechanistic level. For instance, it is unknown how and at what step (e.g., ribosome loading, elongation or termination) miRNAs inhibit protein synthesis carried out by ribosomes.

Through intensive genetic studies, the key players in hc-siRNA biogenesis and DNA methylation have been identified. Additionally, high-throughput methylome analysis has provided a wealth of information about targets methylated by RdDM at the nucleotide level. Nevertheless, major aspects of hc-siRNA biogenesis and cytosine methylation are not understood or require further experimental evidence. For example,

although Pol IV is essential for the biogenesis of hc-siRNAs, Pol IV-dependent transcripts have not yet been detected. How Pol IV recognizes, and is recruited to, the promoters of these transcripts is also not known. The recruitment of DRM2 to target regions is known to be mediated by small RNAs and Pol V-dependent transcripts, but the underlying molecular mechanism remains to be elucidated. Along with cytosine methylation, there are other epigenetic modifications that undoubtedly contribute to TGS, including histone modification, histone variants, chromatin condensation and higher order chromatin structures. Future studies will need to establish the relationships between these different types of modification and address how cross-talk among them governs the epigenetic landscape.

Although factors that favor ta-siRNA biogenesis have been uncovered, the biological function of ta-siRNAs and their targets remain enigmatic. Particularly, *PPR* genes are abundant in the *Arabidopsis* genome, but the underlying cause of the rapid expansion of this gene family is unclear, as is the functional relevance of the regulation of *PPR* genes by ta-siRNAs.

Mature small RNAs are loaded into AGO effector proteins to direct silencing activity, and the association with AGO proteins may protect small RNAs from harmful enzymatic activity, such as degradation by SDN1 or uridylation by HESO1. The molecular mechanism by which small RNAs are dislodged from AGO proteins and subsequently degraded is not well characterized. Another possibility is that small RNAs may be degraded while they are associated with AGO proteins. Both uridylation and 3' truncation mechanisms that affect small RNAs warrant further study, not only in terms of

the underlying molecular events but also with respect to whether and how these mechanisms are orchestrated in tandem. The fact that loss of HESO1 function reduces but does not eliminate the uridylation of miRNAs suggests the existence of other enzymes with overlapping functions. Moreover, there may be regulatory factors that determine the rate of degradation and sequester or degrade specific small RNAs in response to a signal or stress. Lastly, recent findings about small RNA turnover induced by small RNA target mimics challenge the current understanding of SDN exonuclease activity. Further study is required to address how SDN enzymes, which specifically degrade single-stranded small RNAs, may also be involved in the degradation of sequestered or bound small RNAs.

FIGURES

Figure 1 miRNA biogenesis and silencing mechanism.

Pri-miRNA transcripts are generated from *MIR* gene loci by RNA polymerase II and processed into pre-miRNAs by DCL1 with the assistance of HYL1, SE, DDL, CPL1, MOS2 and TGH. Pre-miRNAs are further processed into miRNA/miRNA* duplexes by DCL1. Both strands of the miRNA/miRNA* duplex are methylated by HEN1 either before or after HST-mediated transport to the cytoplasm. Mature miRNAs are loaded into AGO1 and guide target repression via mRNA cleavage or translational inhibition.

Figure 1

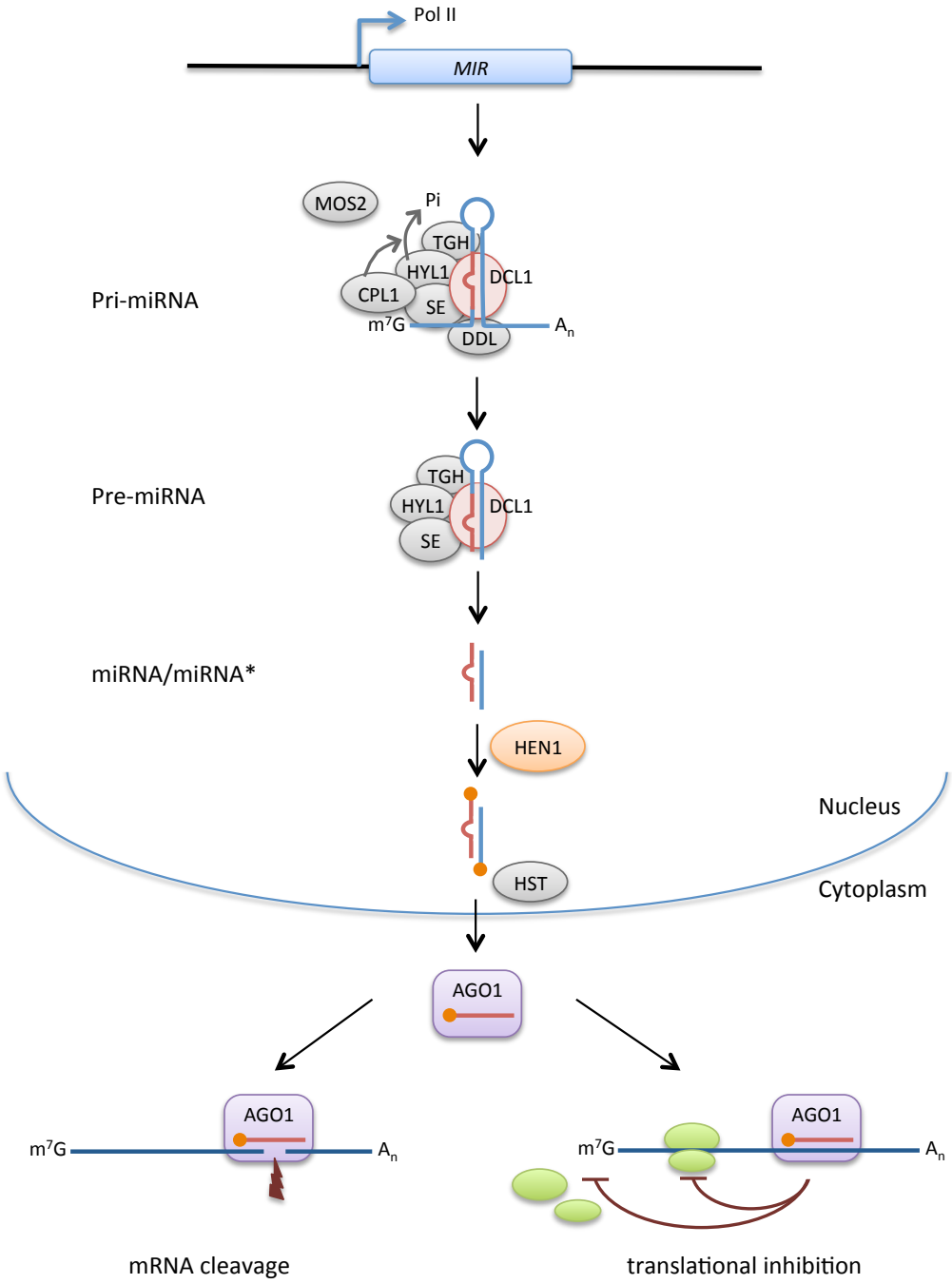

Pol II
MIR
MOS2
Pi
TGH
HYL1
DCL1
CPL1
SE
m7G
An
DDL
Pri-miRNA
TGH
HYL1
DCL1
SE
Pre-miRNA
miRNA/miRNA*
HEN1
Nucleus
HST
Cytoplasm
AGO1
AGO1
m7G
An
AGO1
m7G
An
mRNA cleavage
translational inhibition

Figure 2 Hc-siRNA biogenesis and silencing mechanism.

RNA polymerase IV generates long non-coding RNAs (lncRNAs) from target regions with the aid of CLSY1 and SHH1. These lncRNAs are made double-stranded by RDR2. DCL3 dices the double-stranded RNAs into 24-nt siRNAs, which are subsequently methylated by HEN1 and transported to the cytoplasm. siRNAs are loaded into AGO4 with the assistance of HSP90 then transported into the nucleus. AGO4-loaded siRNAs recognize the nascent transcripts generated by RNA polymerase V, and the DRM2 methyltransferase is recruited to the target. These interactions confer sequence-specific target methylation.

Figure 2

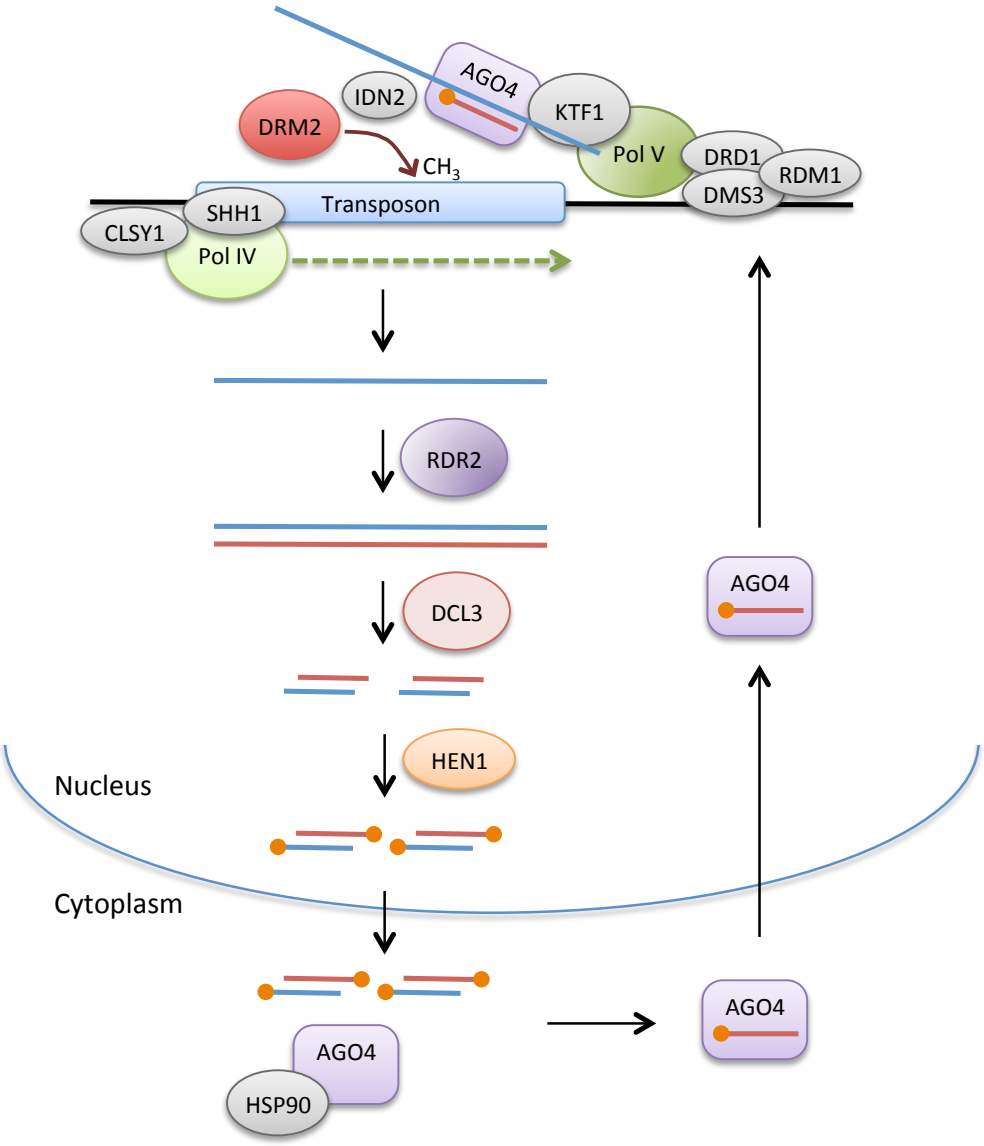
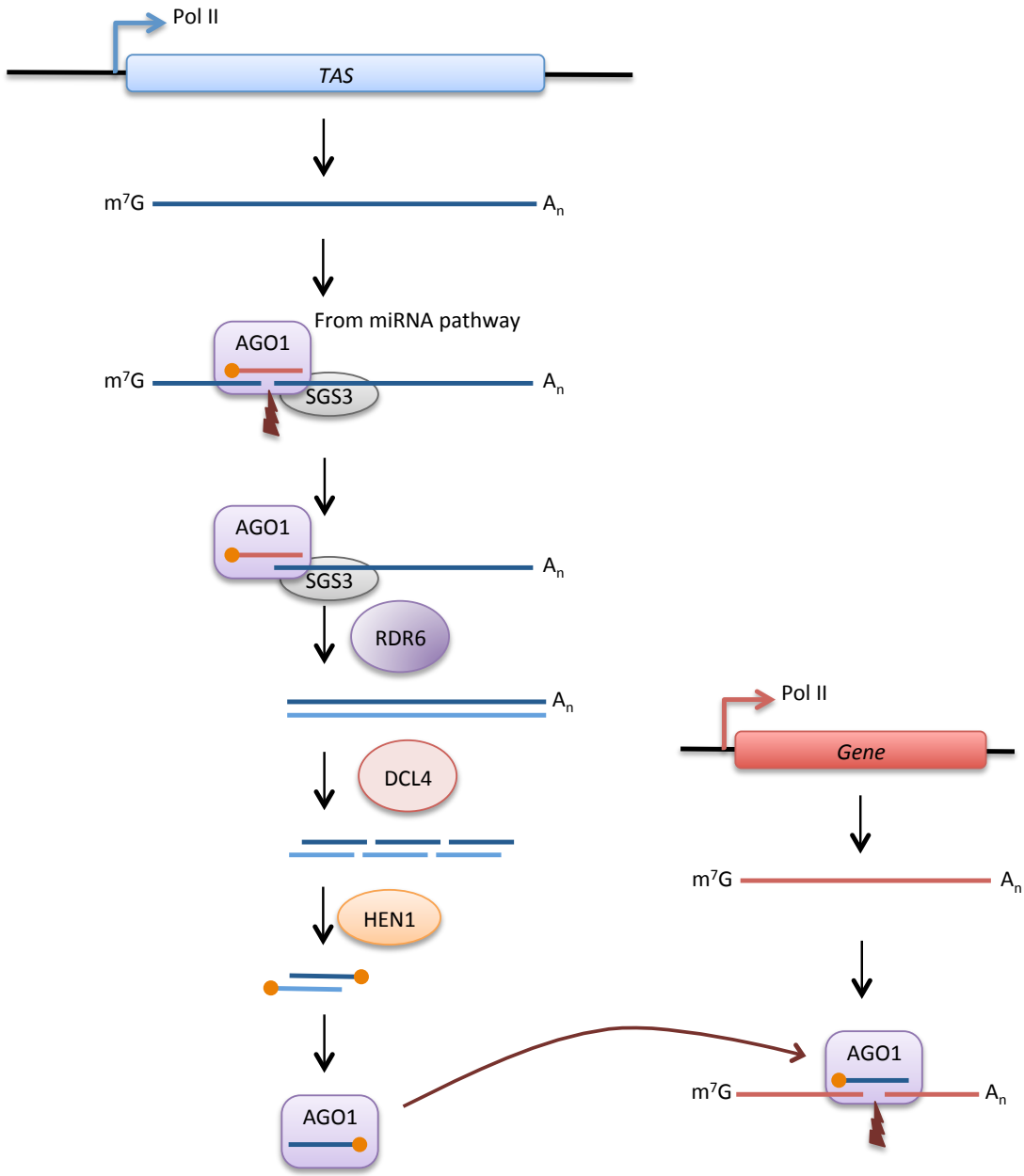

AGO4
IDN2
KTF1
DRM2
Pol V
DRD1
RDM1
DMS3
CH3
Transposon
SHH1
CLSY1
Pol IV
RDR2
DCL3
HEN1
Nucleus
Cytoplasm
AGO4
AGO4
HSP90

Figure 3 ta-siRNA biogenesis and silencing mechanism.

TAS transcripts are generated by Pol II and cleaved by miRNA-associated AGO1. The cleaved 3' fragments are protected from degradation by miRNA-containing complexes, RISC and SGS3 then made double-stranded by RDR6. The double-stranded RNAs are processed into 21-nt siRNAs by DCL4 and methylated by HEN1. ta-siRNAs are loaded into AGO1 and regulate their targets in the same manner as miRNA-mediated target repression.

Figure 3


Pol II
TAS
m7G
An
From miRNA pathway
AGO1
SGS3
RDR6
DCL4
HEN1
Pol II
Gene

REFERENCES

1. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, et al. (2007) RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. Science 316: 1484-1488.

2. Lee JT (2012) Epigenetic Regulation by Long Noncoding RNAs. Science 338: 1435-1439.

3. Mercer TR, Mattick JS (2013) Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol 20: 300-307.

4. Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA. Science 297: 2053-2056.

5. Xie Z, Kasschau KD, Carrington JC (2003) Negative Feedback Regulation of Dicer-Like1 in Arabidopsis by microRNA-Guided mRNA Degradation. Current Biology 13: 784-789.

6. Chen X (2004) A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development. Science 303: 2022-2025.

7. Aukerman MJ, Sakai H (2003) Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-Like Target Genes. Plant Cell 15: 2730-2741.

8. Gandikota M, Birkenbihl RP, Höhmann S, Cardon GH, Saedler H, et al. (2007) The miRNA156/157 recognition element in the 3′ UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. The Plant Journal 49: 683-693.

9. Mosher RA, Schwach F, Studholme D, Baulcombe DC (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. Proceedings of the National Academy of Sciences 105: 3145-3150.

10. Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, et al. (2004) Genetic and Functional Diversification of Small RNA Pathways in Plants. PLoS Biol 2: e104.

11. Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE (2007) Role of RNA polymerase IV in plant small RNA metabolism. Proceedings of the National Academy of Sciences 104: 4536-4541.

12. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, et al. (2008) Criteria for Annotation of Plant MicroRNAs. The Plant Cell Online 20: 3186-3190.

13. Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75: 843-854.

14. Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. Nat Rev Genet 10: 94-108.

15. Lares MR, Rossi JJ, Ouellet DL (2010) RNAi and small interfering RNAs in human disease therapeutic applications. Trends in biotechnology 28: 570-579.

16. Isshiki M, Kodama H (2009) Plant RNAi and Crop Improvement. In: Jain SM, Brar DS, editors. Molecular Techniques in Crop Improvement: Springer Netherlands. pp. 653-673.

17. Chen X (2009) Small RNAs and Their Roles in Plant Development. Annual Review of Cell and Developmental Biology 25: 21-44.

18. Lee Y, Kim M, Han J, Yeom K-H, Lee S, et al. (2004) MicroRNA genes are transcribed by RNA polymerase II. EMBO J 23: 4051-4060.

19. Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, et al. (2005) Expression of Arabidopsis MIRNA Genes. Plant Physiology 138: 2145-2154.

20. Malik S, Roeder RG (2010) The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet 11: 761-772.

21. Kim YJ, Zheng B, Yu Y, Won SY, Mo B, et al. (2011) The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. EMBO J 30: 814-822.

22. Park W, Li J, Song R, Messing J, Chen X (2002) CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in Arabidopsis thaliana. Current Biology 12: 1484-1495.

23. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes & Development 16: 1616-1626.

24. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes & Development 20: 3407-3425.

25. Kurihara Y, Watanabe Y (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proceedings of the National Academy of Sciences of the United States of America 101: 12753-12758.

26. Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, et al. (2009) Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. Genome Research 19: 57-69.

27. Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and Functional Diversification of MIRNA Genes. The Plant Cell Online 23: 431-442.

28. Han M-H, Goud S, Song L, Fedoroff N (2004) The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proceedings of the National Academy of Sciences of the United States of America 101: 1093-1098.

29. Vazquez F, Gasciolli V, CrÈtÈ P, Vaucheret H (2004) The Nuclear dsRNA Binding Protein HYL1 Is Required for MicroRNA Accumulation and Plant Development, but Not Posttranscriptional Transgene Silencing. Current biology : CB 14: 346-351.

30. Lobbes D, Rallapalli G, Schmidt DD, Martin C, Clarke J (2006) SERRATE: a new player on the plant microRNA scene. EMBO Rep 7: 1052-1058.

31. Yang L, Liu Z, Lu F, Dong A, Huang H (2006) SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. The Plant Journal 47: 841-850.

32. Yu B, Bi L, Zheng B, Ji L, Chevalier D, et al. (2008) The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proceedings of the National Academy of Sciences 105: 10073-10078.

33. Kurihara Y, Takashi Y, Watanabe Y (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. RNA 12: 206-212.

34. Dong Z, Han M-H, Fedoroff N (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proceedings of the National Academy of Sciences 105: 9970-9975.

35. Yu B, Yang Z, Li J, Minakhina S, Yang M, et al. (2005) Methylation as a Crucial Step in Plant microRNA Biogenesis. Science 307: 932-935.

36. Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS (2005) Nuclear processing and export of microRNAs in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 102: 3691-3696.

37. Vaucheret H (2008) Plant ARGONAUTES. Trends Plant Sci 13: 350 - 358.

38. Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. Nature Rev Mol Cell Biol 9: 22-32.

39. Baumberger N, Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proceedings of the National Academy of Sciences of the United States of America 102: 11928-11933.

40. Qi Y, Denli AM, Hannon GJ (2005) Biochemical Specialization within Arabidopsis RNA Silencing Pathways. Molecular Cell 19: 421-428.

41. Montgomery T, Howell M, Cuperus J, Li D, Hansen J, et al. (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell 133: 128 - 141.

42. Ji L, Liu X, Yan J, Wang W, Yumul RE, et al. (2011) <italic>ARGONAUTE10</italic> and <italic>ARGONAUTE1</italic> Regulate the Termination of Floral Stem Cells through Two MicroRNAs in <italic>Arabidopsis</italic>. PLoS Genet 7: e1001358.

43. Zhu H, Hu F, Wang R, Zhou X, Sze S-H, et al. (2011) Arabidopsis Argonaute10 Specifically Sequesters miR166/165 to Regulate Shoot Apical Meristem Development. Cell 145: 242-256.

44. Shen B, Goodman HM (2004) Uridine Addition After MicroRNA-Directed Cleavage. Science 306: 997.

45. Souret FF, Kastenmayer JP, Green PJ (2004) AtXRN4 Degrades mRNA in Arabidopsis and Its Substrates Include Selected miRNA Targets. Molecular cell 15: 173-183.

46. German MA, Pillay M, Jeong D-H, Hetawal A, Luo S, et al. (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotech 26: 941-946.

47. Rymarquis LA, Souret FF, Green PJ (2011) Evidence that XRN4, an Arabidopsis homolog of exoribonuclease XRN1, preferentially impacts transcripts with certain sequences or in particular functional categories. RNA 17: 501-511.

48. Song M-G, Kiledjian M (2007) 3′ Terminal oligo U-tract-mediated stimulation of decapping. RNA 13: 2356-2365.

49. Bartel DP (2009) MicroRNAs: Target Recognition and Regulatory Functions. Cell 136: 215-233.

50. Brodersen P, Voinnet O (2009) Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol 10: 141-148.

51. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, et al. (2008) Widespread Translational Inhibition by Plant miRNAs and siRNAs. Science 320: 1185-1190.

52. Yang L, Wu G, Poethig RS (2012) Mutations in the GW-repeat protein SUO reveal a developmental function for microRNA-mediated translational repression in Arabidopsis. Proceedings of the National Academy of Sciences 109: 315-320.

53. Li S, Liu L, Zhuang X, Yu Y, Liu X, et al. (2013) MicroRNAs Inhibit the Translation of Target mRNAs on the Endoplasmic Reticulum in Arabidopsis. Cell 153: 562-574.

54. Bazzini AA, Lee MT, Giraldez AJ (2012) Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. Science 336: 233-237.

55. Djuranovic S, Nahvi A, Green R (2012) miRNA-Mediated Gene Silencing by Translational Repression Followed by mRNA Deadenylation and Decay. Science 336: 237-240.

56. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research 39: D152-D157.

57. Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. Annual Review of Plant Biology 57: 19-53.

58. Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 13: 613-626.

59. Mallory AC, Vaucheret H (2006) Functions of microRNAs and related small RNAs in plants. Nat Genet.

60. Khraiwesh B, Zhu J-K, Zhu J (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1819: 137-148.

61. Sunkar R, Zhu J-K (2004) Novel and Stress-Regulated MicroRNAs and Other Small RNAs from Arabidopsis. The Plant Cell Online 16: 2001-2019.

62. Vaucheret H, Vazquez F, Crété P, Bartel DP (2004) The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes & Development 18: 1187-1197.

63. Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10: 126-139.

64. Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet 11: 597-610.

65. Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 9: 102-114.

66. Allen E, Xie Z, Gustafson AM, Sung G-H, Spatafora JW, et al. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat Genet 36: 1282-1290.

67. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, et al. (2007) High-Throughput Sequencing of <italic>Arabidopsis</italic> microRNAs: Evidence for Frequent Birth and Death of <italic>MIRNA</italic> Genes. PLoS ONE 2: e219.

68. Fenselau de Felippes F, Schneeberger K, Dezulian T, Huson DH, Weigel D (2008) Evolution of Arabidopsis thaliana microRNAs from random sequences. RNA 14: 2455-2459.

69. Axtell MJ (2008) Evolution of microRNAs and their targets: Are all microRNAs biologically relevant? Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1779: 725-734.

70. Maher C, Stein L, Ware D (2006) Evolution of Arabidopsis microRNA families through duplication events. Genome Research 16: 510-519.

71. Castel SE, Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet 14: 100-112.

72. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11: 204-220.

73. Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol 12: 483-492.

74. Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, et al. (2009) Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II. Molecular Cell 33: 192-203.

75. Huang L, Jones AME, Searle I, Patel K, Vogler H, et al. (2009) An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. Nat Struct Mol Biol 16: 91-93.

76. Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, et al. (2007) An SNF2 Protein Associated with Nuclear RNA Silencing and the Spread of a Silencing Signal between Cells in Arabidopsis. The Plant Cell Online 19: 1507-1521.

77. Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE (2011) SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4, and Chromatin Remodeling Factors, Associate with RNA Polymerase IV. PLoS Genet 7: e1002195.

78. Kasschau K, Fahlgren N, Chapman E, Sullivan C, Cumbie J, et al. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol 5: e57.

79. Qi Y, He X, Wang X-J, Kohany O, Jurka J, et al. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature 443: 1008-1012.

80. Ye R, Wang W, Iki T, Liu C, Wu Y, et al. (2012) Cytoplasmic Assembly and Selective Nuclear Import of Arabidopsis ARGONAUTE4/siRNA Complexes. Molecular cell 46: 859-870.

81. Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. Science 299: 716-719.

82. Eun C, Lorkovic ZJ, Naumann U, Long Q, Havecker ER, et al. (2011) AGO6 Functions in RNA-Mediated Transcriptional Gene Silencing in Shoot and Root Meristems in <italic>Arabidopsis thaliana</italic>. PLoS ONE 6: e25730.

83. Zheng X, Zhu J, Kapoor A, Zhu J-K (2007) Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. EMBO J 26: 1691-1701.

84. Olmedo-Monfil V, Duran-Figueroa N, Arteaga-Vazquez M, Demesa-Arevalo E, Autran D, et al. (2010) Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature 464: 628-632.

85. Wierzbicki A, Haag J, Pikaard C (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135: 635 - 648.

86. Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu J-K, et al. (2010) A Protein Complex Required for Polymerase V Transcripts and RNA- Directed DNA Methylation in Arabidopsis. Current biology : CB 20: 951-956.

87. Zheng B, Wang Z, Li S, Yu B, Liu J-Y, et al. (2009) Intergenic transcription by RNA Polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes & Development.

88. Azevedo J, Cooke R, Lagrange T (2011) Taking RISCs with Ago hookers. Current Opinion in Plant Biology 14: 594-600.

89. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nature Genet 41: 630-634.

90. El-Shami M, Pontier D, Lahmy S, Braun L, Picart C, et al. (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. Genes & Development 21: 2539-2544.

91. He X-J, Hsu Y-F, Zhu S, Wierzbicki AT, Pontes O, et al. (2009) An Effector of RNA-Directed DNA Methylation in Arabidopsis Is an ARGONAUTE 4- and RNA-Binding Protein. Cell 137: 498-508.

92. Ausin I, Mockler TC, Chory J, Jacobsen SE (2009) IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. Nat Struct Mol Biol 16: 1325-1327.

93. Ausin I, Greenberg MVC, Simanshu DK, Hale CJ, Vashisht AA, et al. (2012) INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in Arabidopsis. Proceedings of the National Academy of Sciences 109: 8374-8381.

94. Zhang C-J, Ning Y-Q, Zhang S-W, Chen Q, Shao C-R, et al. (2012) IDN2 and Its Paralogs Form a Complex Required for RNA–Directed DNA Methylation. PLoS Genet 8: e1002693.

95. Gao Z, Liu H-L, Daxinger L, Pontes O, He X, et al. (2010) An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. Nature 465: 106-109.

96. Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, et al. (2003) Role of the DRM and CMT3 Methyltransferases in RNA-Directed DNA Methylation. Current Biology 13: 2212-2217.

97. Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. Nature Rev Genet 6: 351-360.

98. Vongs A, Kakutani T, Martienssen RA, Richards EJ (1993) Arabidopsis thaliana DNA methylation mutants. Science 260: 1926-1928.

99. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, et al. (2001) Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. Science 292: 2077-2080.

100. Jeffrey AJ, Trevor LS, Eric JR (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nature Genetics 22: 94-97.

101. Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 416: 556-560.

102. Du J, Zhong X, Bernatavichute Yana V, Stroud H, Feng S, et al. (2012) Dual Binding of Chromomethylase Domains to H3K9me2-Containing Nucleosomes Directs DNA Methylation in Plants. Cell 151: 167-180.

103. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452: 215-219.

104. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell 133: 523-536.

105. Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8: 272-285.

106. Lisch D (2009) Epigenetic Regulation of Transposable Elements in Plants. Annual Review of Plant Biology 60: 43-66.

107. Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, et al. (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature 411: 212-214.

108. Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T (2003) Role of CG and Non-CG Methylation in Immobilization of Transposons in Arabidopsis. Current Biology 13: 421-426.

109. Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, et al. (2009) Bursts of retrotransposition reproduced in Arabidopsis. Nature 461: 423-426.

110. Soppe WJJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, et al. (2000) The Late Flowering Phenotype of fwa Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. Molecular Cell 6: 791-802.

111. Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, et al. (2004) One-Way Control of FWA Imprinting in Arabidopsis Endosperm by DNA Methylation. Science 303: 521-523.

112. Gutierrez-Marcos JF, Dickinson HG (2012) Epigenetic Reprogramming in Plant Reproductive Lineages. Plant and Cell Physiology 53: 817-823.

113. Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, et al. (2009) Genome-Wide Demethylation of Arabidopsis Endosperm. Science 324: 1451-1454.

114. Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, et al. (2012) Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. Science 337: 1360-1364.

115. Slotkin RK, Vaughn M, Borges F, Tanurd~i M, Becker JD, et al. (2009) Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell 136: 461-472.

116. Calarco JP, Borges F, Donoghue MTA, Van†Ex F, Jullien PE, et al. (2012) Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. Cell 151: 194-205.

117. Jullien PE, Mosquna A, Ingouff M, Sakata T, Ohad N, et al. (2008) Retinoblastoma and Its Binding Partner MSI1 Control Imprinting in <italic>Arabidopsis</italic>. PLoS Biol 6: e194.

118. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, et al. (2002) DEMETER, a DNA Glycosylase Domain Protein, Is Required for Endosperm Gene Imprinting and Seed Viability in Arabidopsis. Cell 110: 33-42.

119. Siomi MC, Sato K, Pezic D, Aravin AA (2011) PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol 12: 246-258.

120. Ishizu H, Siomi H, Siomi MC (2012) Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. Genes & Development 26: 2361-2373.

121. Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318: 761-764.

122. Allen E, Howell MD (2010) miRNAs in the biogenesis of trans-acting siRNAs in higher plants. Seminars in Cell & Developmental Biology 21: 798-804.

123. Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. Cell 121: 207-221.

124. Yoshikawa M, Peragine A, Park MY, Poethig RS (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes & Development 19: 2164-2175.

125. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes & Development 18: 2368-2379.

126. Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, et al. (2004) Endogenous trans-Acting siRNAs Regulate the Accumulation of Arabidopsis mRNAs. Molecular Cell 16: 69-79.

127. Gasciolli V, Mallory AC, Bartel DP, Vaucheret H (2005) Partially Redundant Functions of Arabidopsis DICER-like Enzymes and a Role for DCL4 in Producing trans-Acting siRNAs. Current biology : CB 15: 1494-1500.

128. Xie Z, Allen E, Wilken A, Carrington JC (2005) DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America 102: 12984-12989.

129. Cuperus J, Carbonell A, Fahlgren N, Garcia-Ruiz H, Burke R, et al. (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat Struct Mol Biol 17: 997 - U111.

130. Chen H-M, Chen L-T, Patel K, Li Y-H, Baulcombe DC, et al. (2010) 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proceedings of the National Academy of Sciences 107: 15269-15274.

131. Manavella PA, Koenig D, Weigel D (2012) Plant secondary siRNA production determined by microRNA-duplex structure. Proceedings of the National Academy of Sciences.

132. Zhang C, Ng DWK, Lu J, Chen ZJ (2012) Roles of target site location and sequence complementarity in trans-acting siRNA formation in Arabidopsis. The Plant Journal 69: 217-226.

133. Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, et al. (2004) Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis. The Plant Cell Online 16: 2089-2103.

134. Chen H-M, Li Y-H, Wu S-H (2007) Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. Proceedings of the National Academy of Sciences 104: 3318-3323.

135. Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, et al. (2007) Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway in Arabidopsis Reveals Dependency on miRNA- and tasiRNA-Directed Targeting. The Plant Cell Online 19: 926-942.

136. Fahlgren N, Montgomery TA, Howell MD, Allen E, Dvorak SK, et al. (2006) Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA Affects Developmental Timing and Patterning in Arabidopsis. Current biology : CB 16: 939-944.

137. Chapman EJ, Estelle M (2009) Mechanism of Auxin-Regulated Gene Expression in Plants. Annual Review of Genetics 43: 265-285.

138. Pulido A, Laufs P (2010) Co-ordination of developmental processes by small RNAs during leaf development. Journal of Experimental Botany 61: 1277-1291.

139. Pekker I, Alvarez JP, Eshed Y (2005) Auxin Response Factors Mediate Arabidopsis Organ Asymmetry via Modulation of KANADI Activity. The Plant Cell Online 17: 2899-2910.

140. Chitwood DH, Nogueira FTS, Howell MD, Montgomery TA, Carrington JC, et al. (2009) Pattern formation via small RNA mobility. Genes & Development 23: 549-554.

141. Hamilton AJ, Baulcombe DC (1999) A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. Science 286: 950-952.

142. Ding S-W, Voinnet O (2007) Antiviral Immunity Directed by Small RNAs. Cell 130: 413-426.

143. Burgyán J, Havelda Z (2011) Viral suppressors of RNA silencing. Trends in Plant Science 16: 265-272.

144. Ye K, Patel DJ (2005) RNA Silencing Suppressor p21 of Beet Yellows Virus Forms an RNA Binding Octameric Ring Structure. Structure 13: 1375-1384.

145. Lakatos L, Szittya G, Silhavy D, Burgyan J (2004) Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. EMBO J 23: 876-884.

146. Zhang X, Yuan Y-R, Pei Y, Lin S-S, Tuschl T, et al. (2006) Cucumber mosaic virus-encoded 2b suppressor inhibits Arabidopsis Argonaute1 cleavage activity to counter plant defense. Genes & Development 20: 3255-3268.

147. Fusaro AF, Correa RL, Nakasugi K, Jackson C, Kawchuk L, et al. (2012) The Enamovirus P0 protein is a silencing suppressor which inhibits local and systemic RNA silencing through AGO1 degradation. Virology 426: 178-187.

148. Jin H, Zhu J-K (2010) A viral suppressor protein inhibits host RNA silencing by hooking up with Argonautes. Genes & Development 24: 853-856.

149. Giner A, Lakatos L, García-Chapa M, López-Moya JJ, Burgyán J (2010) Viral Protein Inhibits RISC Activity by Argonaute Binding through Conserved WG/GW Motifs. PLoS Pathog 6: e1000996.

150. Azevedo J, Garcia D, Pontier D, Ohnesorge S, Yu A, et al. (2010) Argonaute quenching and global changes in Dicer homeostasis caused by a pathogen-encoded GW repeat protein. Genes & Development 24: 904-915.

151. Glick E, Zrachya A, Levy Y, Mett A, Gidoni D, et al. (2008) Interaction with host SGS3 is required for suppression of RNA silencing by tomato yellow leaf curl virus V2 protein. Proceedings of the National Academy of Sciences 105: 157-161.

152. Jamous RM, Boonrod K, Fuellgrabe MW, Ali-Shtayeh MS, Krczal G, et al. (2011) The helper component-proteinase of the Zucchini yellow mosaic virus inhibits the Hua Enhancer 1 methyltransferase activity in vitro. Journal of General Virology 92: 2222-2226.

153. Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. The Plant Cell Online 2: 279-289.

154. van der Krol AR, Mur LA, Beld M, Mol JN, Stuitje AR (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. The Plant Cell Online 2: 291-299.

155. Luo Z, Chen Z (2007) Improperly Terminated, Unpolyadenylated mRNA of Sense Transgenes Is Targeted by RDR6-Mediated RNA Silencing in Arabidopsis. The Plant Cell Online 19: 943-958.

156. Ji L, Chen X (2012) Regulation of small RNA stability: methylation and beyond. Cell Res 22: 624-636.

157. Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation Protects miRNAs and siRNAs from a 3'-End Uridylation Activity in Arabidopsis. Current Biology 15: 1501-1507.

158. Zhao Y, Yu Y, Zhai J, Ramachandran V, Dinh TT, et al. (2012) The Arabidopsis Nucleotidyl Transferase HESO1†Uridylates Unmethylated Small RNAs to Trigger Their Degradation. Current biology : CB 22: 689-694.

159. Ramachandran V, Chen X (2008) Degradation of microRNAs by a Family of Exoribonucleases in Arabidopsis. Science 321: 1490-1492.

160. Ren G, Chen X, Yu B (2012) Uridylation of miRNAs by HEN1 SUPPRESSOR1 in Arabidopsis. Current Biology 22: 695-700.

161. Yan J, Gu Y, Jia X, Kang W, Pan S, et al. (2012) Effective Small RNA Destruction by the Expression of a Short Tandem Target Mimic in Arabidopsis. Plant Cell 24: 415.

162. Chitwood DH, Timmermans MCP (2007) Target mimics modulate miRNAs. Nat Genet 39: 935-936.

163. Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, et al. (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet 39: 1033-1037.

164. Wu H-J, Wang Z-M, Wang M, Wang X-J (2013) Widespread Long Noncoding RNAs as Endogenous Target Mimics for MicroRNAs in Plants. Plant Physiology 161: 1875-1884.

CHAPTER 1

Development of a luciferase-based reporter of transcriptional gene silencing that enables bidirectional mutant screening in *Arabidopsis thaliana*

ABSTRACT

Cytosine methylation is an important chromatin modification that maintains genome integrity and regulates gene expression through transcriptional gene silencing. The major players in *de novo* methylation guided by siRNAs (known as RNA-directed DNA methylation, or RdDM), maintenance methylation, and active demethylation have been identified in *Arabidopsis*. However, active demethylation only occurs at a subset of RdDM loci, raising the question of how the homeostasis of DNA methylation is achieved at most RdDM loci. To identify factors that regulate the levels of cytosine methylation, we aimed to establish a transgenic reporter system that allows for forward genetic screens in *Arabidopsis*. We introduced a dual *35S* promoter (*d35S*) driven *luciferase* reporter into *Arabidopsis* and isolated a line, *LUCH*, with a moderate level of luciferase activity. *LUCH* produced transgene-specific 24 nt siRNAs and *d35S* contained methylated cytosine in CG, CHG and CHH contexts. Treatment with an inhibitor of cytosine methylation de-repressed luciferase activity. Mutations in several components of the RdDM pathway but not the maintenance methylation pathway resulted in reduced *d35S* methylation, especially CHH methylation, and de-repression of luciferase activity. A mutation in *MOM1* that is known to cooperate with RdDM to silence transposons

reduced *d35S* DNA methylation and de-repressed *LUCH* expression. A mutation in ROS1, a cytosine demethylation enzyme, increased *d35S* methylation and reduced *LUCH* expression. We developed a luciferase-based reporter system, *LUCH*, which reports both DNA methylation directed by small RNAs and active demethylation by ROS1 in *Arabidopsis*. The moderate basal level of *LUCH* expression allows for bi-directional genetic screens that dissect the mechanisms of DNA methylation as well as demethylation.

INTRODUCTION

Cytosine methylation is a major epigenetic mechanism that establishes transcriptional gene silencing (TGS) to maintain genome integrity and regulate gene expression in plants and mammals (reviewed in [1]). Well-known biological phenomena involving DNA methylation as an underlying mechanism include imprinting, paramutation and X chromosome inactivation. In plants, transposons and repetitive elements are hypermethylated, thereby keeping transposons silenced and immobilized and consequently protecting the genome from damage by these mobile elements. Also, when transposons or repeats are located in the regulatory regions of genes, DNA methylation at the transposons or repeats may influence the transcription of the nearby genes through TGS.

The enzymes that initiate, maintain, and erase DNA methylation in *Arabidopsis* have been identified and characterized (reviewed in [1]). *De novo* DNA methylation, also known as RNA-directed DNA methylation (RdDM), requires DOMAIN REARRANGED METHYLTRANSFERASE2 (DRM2), which is guided to specific genomic loci by 24 nucleotide (nt) small interfering RNAs (siRNAs). siRNAs are synthesized from repeats and transposons in a RNA polymerase IV (Pol IV)-, RNA DEPENDENT RNA POLYMERASE2 (RDR2)-, and DICERLIKE3 (DCL3)-dependent manner. Pol IV is thought to transcribe these loci into single-stranded RNAs, which are then rendered double-stranded by RDR2. DCL3 dices the double-stranded RNAs into 24 nt siRNAs, which are loaded into the ARGONAUTE4 (AGO4)-clade of AGO proteins

(reviewed in [2]). Base-pairing between the AGO4-loaded siRNAs and nascent transcripts produced by Pol V is thought to recruit AGO4/siRNAs and DRM2 to the RdDM targets, resulting in *de novo* methylation in a sequence-specific manner (reviewed in [2]). After the initial establishment of DNA methylation, hemimethylated cytosines in CG and CHG contexts resulting from DNA replication are fully methylated by METHYLTRANSFERASE1 (MET1) and CHROMOMETHYLTRANSFERASE3 (CMT3), respectively (reviewed in [1]). The positive feedback loop in which DNA methylation promotes siRNA biogenesis, which guides *de novo* DNA methylation, needs to be kept in check to prevent the expansion of heterochromatin and the sporadic silencing of genic regions. One such mechanism is DNA demethylation. Four DNA glycosylase/lyase enzymes remove methyl cytosine through a base excision repair mechanism (reviewed in [3]). DEMETER establishes imprinting during female gametogenesis and REPRESSOR OF SILENCING1 (ROS1), DEMETER-LIKE2 (DML2) and DML3 prevent aberrant hypermethylation in vegetative tissues.

Although the enzymes that deposit or erase DNA methylation are known, how these enzymes are regulated to achieve the proper homeostasis of DNA methylation is still nebulous. Although demethylation can keep DNA methylation in check, whole genome bisulfite sequencing in the *ros1 dml2 dml3* triple mutant revealed that only a few hundred loci are hypermethylated [4] and are thus targets of demethylation. Since thousands of loci harbor DNA methylation, generate siRNAs and are targets of RdDM, it remains to be determined how most RdDM loci achieve homeostasis of DNA

methylation. It is likely that other, as yet unknown, mechanisms prevent the hypermethylation of RdDM loci.

In addition to the RdDM pathway, *MORPHEUS' MOLECULE1* (*MOM1*) impacts TGS in a complex manner usually without affecting the levels of cytosine methylation at target loci [5]. It encodes a protein with similarities to chromatin remodeling ATPases and silences endogenous loci and transgenes by an unknown mechanism [5,6]. *MOM1* exhibits a complex relationship with RdDM depending on the target loci [7]. It functions either in the same pathway as RdDM or in a parallel pathway, or it could even antagonize the silencing by RdDM. Some loci are transcriptionally suppressed by *MOM1* independently of RdDM.

Forward genetic screens in *Arabidopsis* can help reveal mechanisms that regulate DNA methylation. In fact, most of the currently known genes involved in DNA methylation or demethylation were uncovered through genetic screens. However, most prior genetic screens were based on the isolation of mutations that release RdDM to result in de-repressed reporter gene expression, thus precluding the identification of negative regulators of DNA methylation. So far, the only known negative factors in DNA methylation, ROS1 and ROS3 (a protein required for ROS1-mediated demethylation), were isolated from genetic screens using the *RD29A::LUC* transgene system [8,9]. Therefore, *RD29A::LUC* happens to be a target of ROS1. As mentioned above, the relatively lower number of ROS1/DML2/DML3 target loci in the genome as compared to the number of RdDM loci suggests the presence of unknown negative factors for methylation acting independently of, or in combination with, active demethylation by

ROS1/DML2/DML3. Consequently, it is valuable to develop additional RdDM reporter transgenes inserted into different genomic locations to allow for the identification of these negative players.

Here, we report the establishment of a firefly *LUCIFERASE* (*LUC*)-based reporter transgene driven by a dual *35S* promoter that harbors DNA methylation in CG, CHG, and CHH contexts in *Arabidopsis*. We show that *LUC* expression is repressed mainly through CHH methylation in an RdDM-dependent manner. *MOM1* also plays a role in DNA methylation and TGS of the reporter. More importantly, the moderate level of basal *LUC* expression in wild-type plants allows for genetic screens that aim at the isolation of mutants with not only defective but also enhanced RdDM. In fact, a *ros1* allele with reduced transgene expression was isolated using this system. The reporter line will prove to be an effective tool in dissecting the mechanisms that regulate DNA methylation.

RESULTS AND DISCUSSION

Generation of the luciferase reporter line, *LUCH*

Initially, we aimed to establish a *LUC*-based transgene that reported both TGS by RdDM and post-transcriptional gene silencing by miRNAs to allow for forward genetic screens. A transgene was constructed such that *LUC* was C-terminally fused in frame to the partial *AP2* fragment containing the miR172 binding site [10] and the transgene was driven by a dual *35S* promoter, which will be referred to as *d35S*, from Cauliflower mosaic virus (*d35S::LUC-AP2*). In the same vector, *d35S*-driven *NEOMYCIN*

PHOSPHOTRANSFERASE II (*d35S::NPTII*) served as a selectable marker for plant transformation (Figure 1.1). This construct was introduced into the *RNA-dependent RNA polymerase6-11* (*rdr6-11*) [11-13] mutant background to prevent sense transgene post-transcriptional silencing (S-PTGS; [11-13]) and one line with moderate levels of LUC signal was isolated to enable bidirectional genetic screens based on higher or lower LUC signals. The *d35S::LUC-AP2* transgene in this line was named *LUCH* (*LUC* repressed by CHH methylation), as we found later that it was repressed by CHH methylation in *d35S*. *LUCH* was a one-copy insertion at a single genomic locus according to Southern blot analysis using the *LUC* sequence as a probe (Figure 1.2). TAIL-PCR followed by sequencing revealed that the transgene resided 20 nt before the stop codon of At3g07350, a gene of unknown function. This insertion did not cause any obvious morphological phenotypes.

LUCH does not report miRNA activity

Since *LUCH* contained a miR172 binding site, we first investigated whether it was able to report miRNA activity. If it were repressed by miR172, we would expect mutations in miRNA biosynthesis genes (reviewed in [14]), such as *DICERLIKE1* (*DCL1*), *HYPONASTIC LEAVES1* (*HYL1*), and *SERRATE* (*SE*) to de-repress *LUCH* expression. In the F2 population of *LUCH* crossed to *dcl1-7*, LUC luminescence was moderately increased in 12 out of 216 segregating seedlings (Figure 1.3A). Since *LUCH* and *DCL1* are not linked, the small proportion of seedlings with the moderately high LUC luminescence was not consistent with *dcl1-7* being able to de-repress *LUCH*

expression. Indeed, genotyping confirmed that only one of the 12 was homozygous for *dcl1-7*, and 3 of the 12 were homozygous for the wild-type *DCL1* allele. Therefore, the moderate increase was likely due to inherent variations in *LUCH* expression or other background mutations. *hyl1* and *se-1* mutations also failed to increase LUC luminescence (Figure 1.3B, C). These results demonstrate that *LUCH* was unable to report miRNA activities even though the *LUC* transcript contains a miRNA-binding site in the 3′ UTR.

LUCH is regulated by RdDM-mediated TGS

To evaluate whether *LUCH* was repressed by RdDM-mediated TGS, we first examined whether *LUCH* had the molecular characteristics associated with RdDM. When compared with other reporter systems (*NOSpro* and *α′pro* [15,16]), *d35S* is more than twice as long as those promoters but has a similar percentage of GC content. *d35S* has a relatively high non-CG composition (23 CG, 19 CHG and 138/128 CHH in forward/reverse strands), which was also observed in the α′pro system that was reported to be more sensitive to the regulation by RdDM than *NOSpro* [15,16]. McrBC-PCR was conducted using primers that specifically amplified the *d35S* in *LUCH* instead of that in *d35S::NPTII* to evaluate the DNA methylation status of the *LUCH* transgene. The results showed that *d35S* was methylated whereas the *LUC* coding region was not (Figure 1.4A). Bisulfite sequencing revealed the presence of DNA methylation in CG, CHG, and CHH contexts (Figure 1.4B). The levels of CHH methylation were 22%, which was particularly high as compared to other previously established reporter lines of RdDM. For example, the *clk-sk* line had 15% CHH methylation in the *SUPERMAN* 5′ region [17]; the

RD29A::LUC line had 1% and 6% CHH methylation in the *RD29A* promoter in wild type and *ros1*, respectively [8]. Treatment of *LUCH* seedlings with 5-aza-2′-deoxycytidine, an inhibitor of cytosine methylation increased LUC luminescence and *LUC* transcript levels, indicating that cytosine methylation transcriptionally silenced *LUCH* expression (Figure 1.5).

Next, since RdDM target loci produce siRNAs, we determined the accumulation of siRNAs from the *LUCH* and *d35S::NPTII* transgenes. Even though we did not artificially introduce any hairpin source of *d35S*-specific siRNAs, siRNAs were detected in the *LUCH* line by northern blotting using a *d35S*-specific probe (Figure 1.4C). High throughput sequencing was conducted to examine the small RNAs from the transgenes in more detail. siRNAs mapping to both DNA strands of the two transgenes were found; and 22 nt siRNAs were the most abundant small RNA species (Figure 1.6A, B). Even though *LUCH* was introduced into *rdr6-11* to prevent S-PTGS by blocking the biogenesis of secondary siRNAs, 21 nt and 22 nt siRNAs mapping to the transgene were present, which suggests that PTGS was still occurring. Perhaps the siRNAs were primary siRNAs resulting from sense and antisense transcription from the locus or secondary siRNAs from the activities of *RDR2*. 24 nt siRNAs, which are associated with RdDM, were also present. Among 18-27 nt small RNAs that mapped to *d35S* in *LUCH*, 24 nt siRNAs accounted for approximately 19% of the total (Figure 1.6A). The *d35S* promoters driving *LUC* and *NPTII* were 96% identical in sequences. We took advantage of the sequence differences to determine whether both regions generated siRNAs. Indeed, siRNAs specific to each *d35S* were found (Figure 1.6C; Table 1.2), indicating that each *d35S* gave

rise to siRNAs. The reverse strand 24 nt siRNAs were similar in quantity between the two transgenes (123 and 106 reads for *LUCH* and *d35S::NPTII*, respectively). Interestingly, forward strand 24 nt siRNAs were different in quantity between the two transgenes: 509 and 120 reads were from *d35S::NPTII* and *LUCH*, respectively. The abundance of *d35S::NPTII*-specific siRNAs was attributed to both higher diversity of siRNA species and higher levels of a subset of species. The basis for the differential siRNA levels is unknown but may be due to differences in read-through transcription at the two d35S. Taken together, *LUCH* exhibits the molecular characteristics associated with RdDM, such as CHH methylation and 24 nt siRNA production.

The regulation of *LUCH* by RdDM was further supported by the fact that mutations in known RdDM pathway components de-repressed *LUCH* expression. We mutagenized the *LUCH* line with either EMS or T-DNA and searched for mutants with higher LUC luminescence (Figure 1.7A). Genetic analyses demonstrated that each mutant with high LUC luminescence harbored a single, recessive mutation. Map-based cloning revealed that the mutations were in *HUA ENHANCER1*, *AGO4*, *DRM2* and *DEFECTIVE IN RNA-DIRECTED DNA METHYLATION1* (*DRD1*) (Figure 1.8), which are known genes in the RdDM pathway (reviewed in [1]). In addition, introducing *nrpe1-1*, a mutant of the largest subunit of Pol V (reviewed in [2]), into *LUCH* de-repressed LUC luminescence (Figure 1.7A). These mutants had higher levels of *LUC* transcripts as revealed by RT-PCR (Figure 1.7B), indicating that the de-repression of *LUCH* expression was at the transcriptional level. Since both *LUC* and *NPTII* are under the regulation of *d35S*, we analyzed the expression levels of *NPTII* by RT-PCR. The *NPTII* transcript

levels were also increased in these RdDM mutants (Figure 1.7B). We next analyzed the DNA methylation status of *d35S* in these mutants. Southern blot analysis with a *d35S*-specific probe showed that *d35S*-specific bands were downwardly shifted in *ago4-6*, *drd1-12* and *drm2-6* (Figure 1.9), indicating that DNA methylation at *d35S* was reduced in *ago4-6, drd1-12* and *drm2-6*. Bisulfite-sequencing with primers that allowed only amplification of *d35S* in *LUCH* showed that the levels of DNA methylation were decreased in all sequence contexts in *ago4-6*, with CHH methylation being the most drastically decreased (Figure 1.4B). These results show that *LUCH* is repressed by *de novo* DNA methylation at *d35S* and the repression requires RdDM components. To evaluate whether maintenance methylation at CG and CHG contexts by *MET1* and *CMT3*, respectively, contributes to the repression of *LUCH*, we crossed *met1-3* and *cmt3-7* mutations into *LUCH*. *met1-3* or *cmt3-7* did not affect *LUCH* expression (Figure 1.10), indicating that this reporter line was mainly repressed by *de novo* methylation through *DRM2*. These molecular and genetic results demonstrate that *LUCH* faithfully reports RdDM-mediated TGS.

LUCH is regulated by MOM1

Our genetic screen also resulted in the isolation of a new *mom1* allele (*mom1-5*) that displayed de-repressed LUC luminescence (Figure 1.11A; Figure 1.8). RT-PCR confirmed the increased levels of *LUC* and *NPTII* transcripts and the absence of *MOM1* transcripts in the mutant (Figure 1.11B). DNA methylation at *d35S* was moderately decreased in *mom1-5*, as revealed by McrBC-PCR and Southern blot analysis (Figure

1.7C; Figure 1.9). The reduction in DNA methylation in *mom1-5* was less severe than in RdDM mutants (Figure 1.9). Nonetheless, this shows that the DNA methylation and TGS of *LUCH* require *MOM1*.

LUCH is regulated by ROS1-mediated DNA demethylation

A major motivation to establish a *LUC*-based reporter was to enable the screening for mutants with enhanced silencing. The *LUCH* line, which exhibited a moderate basal level of LUC luminescence, was suitable for such a purpose. We performed T-DNA insertional mutagenesis of the *LUCH* line and isolated a recessive mutant allele with lower levels of LUC luminescence (Figure 1.12A). Map-based cloning identified this mutant as a new allele of *ROS1* (Figure 1.8), a gene required for DNA demethylation. This suggested that loss of demethylation resulted in the accumulation of cytosine methylation in *d35S* and reinforcement of TGS of *LUCH*. Indeed, there was an increase in DNA methylation of *d35S* in *LUCH* in *ros1-5* according to McrBC-PCR (Figure 1.7C). Levels of *LUC* and *NPTII* transcripts were decreased as determined by RT-PCR (Figure 1.12B). In addition, treatment of *LUCH ros1-5* seedlings with 5-aza-2′-deoxycytidine increased the expression of *LUCH* to wild-type levels (Figure 1.5), which further supported the notion that increased DNA methylation in *ros1-5* led to enhanced TGS of *LUCH*. Therefore, even though *LUCH* is transcriptionally repressed by RdDM, the basal expression of *LUCH* is relatively high such that the transgene can be used to screen for mutants with enhanced silencing.

CONCLUSIONS

We developed a transgenic *LUC* reporter system that reported both TGS by RdDM and *MOM1*, and *ROS1*-mediated demethylation. Moderate expression of the reporter enables genetic screens in two directions to isolate mutants with decreased as well as increased DNA methylation. Considering that existing TGS reporter systems, such as the *NOSpro*, *α′pro*, and *clk-sk* lines, are mainly suitable for the isolation of positive players in RdDM, *LUCH* is a useful genetic resource for the identification of negative players in RdDM, for which nothing is known. Moreover, *LUCH* will potentially contribute to the better understanding of *MOM1*-mediated TGS or the mechanisms of active demethylation. For the latter, although *RD29::LUC* reports *ROS1*-mediated DNA demethylation, as a second reporter of *ROS1*-mediated demethylation residing at a different genomic location, *LUCH* will enrich our resources to tackle the mechanisms of demethylation.

MATERIALS AND METHODS

Plant material

Arabidopsis mutants used in this study were *rdr6-11*[18], *dcl1-7*[19], *se-1*[20], *hyl1*[21], *met1-3*[22], *cmt3-7*[23] and *drd3-1*[24] and newly isolated *drm2-6*, *ago4-6*, *drd1-12*, *hen1-9*, *ros1-5* and *mom1-5*. For map-based cloning of newly isolated mutants, *LUCH rdr6-11* in the Columbia-0 (Col-0) accession was introgressed into Landsberg *erecta* (L*er*) by backcrossing to L*er* 5 times and one line with a similar level of LUC activity as *LUCH* in Col-0 was isolated. The isolated mutants from *LUCH rdr6-11* in Col were each crossed to *LUCH rdr6-11* in L*er*, and in the F2 population, seedlings with high

(for *drm2-6*, *ago4-6*, *drd1-12*, *hen1-9*, and *mom1-5)* or low (*ros1-5*) luciferase activities were identified and served as the mapping population. Polymorphisms between Col-0 and L*er* were utilized to map and clone the genes.

Growth conditions and luciferase live imaging

Arabidopsis thaliana seeds were surface-sterilized, planted on MS-agar plates containing 1% sucrose, and stratified at 4°C for 3 days. Seedlings were grown at 23°C under continuous lights for ten days. All experiments were performed with 10-day old seedlings unless otherwise specified. For luciferase live imaging, 1 mM luciferin (a substrate of luciferase; Promega) in 0.01% Triton X-100 was sprayed onto the seedlings, which were then transferred to a Stanford Photonics Onyx Luminescence Dark Box. Luciferase images were taken with a Roper Pixis 1024B camera controlled by the WinView32 software at a 2 min exposure time. Identical exposure conditions were used to capture all images in this study. The images were displayed and analyzed with WinView32 such that image contrast was adjusted to effectively distinguish the difference in intensities between different lines within a plate as previously described [25].

Construction of transgene, Southern analysis and TAIL-PCR

LUC coding region was amplified using Rlucp1 and Rlucp2 primers and pRL-SV40 (Promega) as the template. *d35S::LUC* was constructed by replacing *GFP* in pAVA321[26] with the *LUC* coding region using *Nco*I and *BamH*I restriction sites. The

d35S::LUC cassette was cloned into the pPZP211 binary vector[27] at the *Sal*I and *BamH*I restriction sites. An *AP2* fragment including the miR172 binding site was amplified from Col-0 genomic DNA with primers AP2p26 and AP2p28 and inserted downstream of *d35S::LUC* in pPZP211 using *BamH*I and *EcoR*I to generate *d35S::LUC-AP2*, which will be referred to as *LUCH*. The construct was introduced into *rdr6-11* plants by *Agrobacterium tumefaciens*-mediated transformation. Southern blot analysis was performed according to the standard protocol[28] to evaluate the copy number of *LUCH* using the full-length *LUC* coding region as the probe. The probe was amplified with primers lucp6 and lucp7, and radiolabeled with the RPN1633 Rediprime II Random Prime Labeling System (GE Healthcare). TAIL-PCR was performed as described [29]. Primers used are listed in Table 1.1.

Analysis of DNA cytosine methylation

For the McrBC-PCR assay, two reactions were set up for each genomic DNA sample: McrBC-treated and untreated reactions. 300 ng genomic DNA was digested with 3 units of McrBC (New England Biolabs) for 25 min at 37°C in a 20 µl reaction. Using 1 µl (15 ng) of restricted genomic DNA as the template, genomic regions corresponding to *d35S* or full length *LUC* in the *LUCH* transgene were amplified using 35Sf and LUC 0.13k R primers or lucp6 and lucp7 primers, respectively. *ACT1* was amplified with Actin1-F and Actin1-R primers and used as a loading control. PCR products were analyzed on a 2 % agarose gel stained with ethidium bromide. For Southern blot analysis, 15 µg of genomic DNA was digested with *Alu*I (NEB) and hybridization was performed

following standard methods[28]. The *d35S* promoter was PCR-amplified with 35Sf and 35Sr primers and radiolabeled using the RPN1633 Rediprime II random prime labeling system (GE Healthcare). For bisulfite sequencing, 1μg of genomic DNA was subjected to bisulfite conversion using the EpiTect Bisulfite Kit per manufacturer's instructions (Qiagen). Converted DNA was subjected to PCR reactions with primers YZ 35S Bis F and YZ LUC Bis R and the PCR products were cloned into the pGEM-T Easy vector (Promega). At least 26 colonies were sequenced for each sample. Unique clones were obtained and analyzed for DNA methylation with Kithmeth (http://katahdin.mssm.edu /kismeth/revpage.pl)[30]. For 5-aza-2'-deoxycytidine (Sigma) treatment, seeds were germinated and grown on MS-agar medium containing 7 μg/ml of the chemical for 2 weeks and luciferase images were taken. Primers used are listed in Table 1.1.

Analysis of small RNA accumulation

RNA isolation and hybridization to detect small RNAs were performed as described previously[31]. To detect siRNAs from the *d35S* promoter, a DNA fragment was amplified from the *d35S* promoter using 35Sf and 35Sr primers and cloned into the pGEM-T Easy vector (Promega). The plasmid was linearized by *Spe*I (NEB) and used as a template for *in vitro* transcription with T7 RNA polymerase (Promega) in the presence of [α-^{32}P] UTP. The labeled *in vitro* transcripts were used as the probe in northern blotting. Radioactive signals were detected with a Phosphorimager. For small RNA deep sequencing, a small RNA library was constructed using the TruSeq Small RNA Sample Prep Kit (Illumina) according to the manufacturer's instructions with some modifications.

Instead of total RNA, 15 to 40 nt long RNAs were used as the starting material. The small RNA library was sequenced by Illumina Hiseq2000 at the genomics core facility at UCR. After the raw reads were filtered by the Illumina quality control pipeline and the adaptor sequences were trimmed, 14,363,865 reads between 18 nt and 28 nt were matched to the *Arabidopsis* genome (TAIRv10) as well as the transgenes with SOAP2[32]. 8,710,699 and 22,245 reads were mapped to the *Arabidopsis* genome and the transgenes, respectively, with no mismatches.

RT-PCR

cDNA was synthesized from 5 μg of DNaseI (Roche)-treated total RNA using Reverse Transcriptase (Fermentas) and oligo-dT (Fermentas). Using cDNA and gene-specific primers, PCR was performed and RT-PCR products were analyzed on a 2% agarose gel stained with ethidium bromide. The sequences of primers are listed in Table 1.1.

FIGURES

Figure 1.1 Structure of *LUCH* and its neighboring transgene.

RB and LB, right border and left border of the T-DNA, respectively. The arrows indicate the directions of the coding regions. The *d35S* fragments (marked #1 to #3) specific for the *d35S* promoter upstream of *LUC* are amplified in McrBC-PCR and bisulfite sequencing.

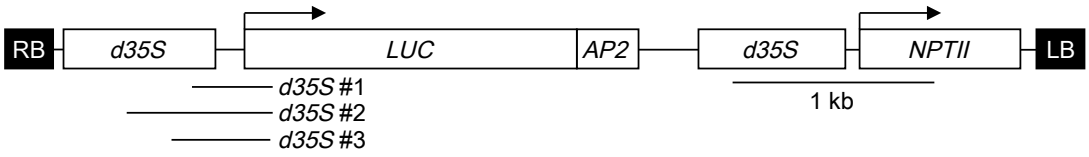
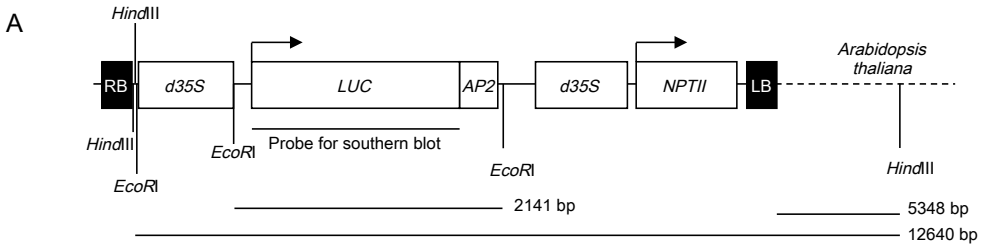


Figure 1.2 Southern blot analysis determines the *LUCH* transgene copy number.

(A) Map of *LUCH* and its neighboring transgene. The positions of the *Eco*RI and *Hind*III restriction sites, the expected sizes of restriction fragments and the position of the *LUC* probe are shown.

(B) Southern blot analysis of *LUCH*. Genomic DNA from Col-0 (wild type) or the *LUCH* line was digested with *Eco*RI or *Hind*III and hybridized with a radiolabeled full-length *LUC* probe. The radiolabeled DNA molecular marker is shown on the right. The sizes and numbers of bands are consistent with a single copy of *LUCH* at a single genomic location.



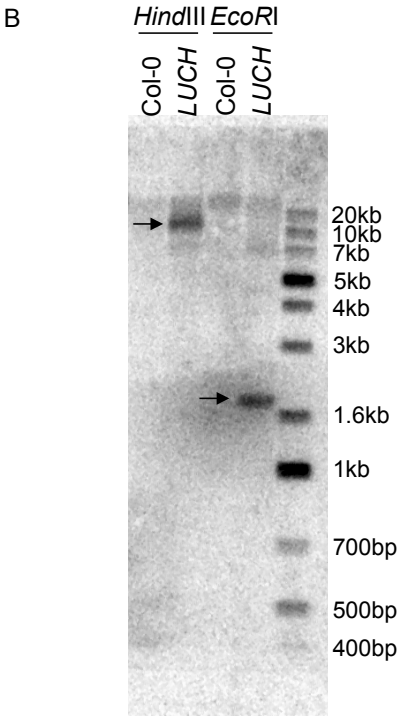


Figure 1.3 *LUCH* is not regulated by the miRNA pathway.

(A) LUC image of Col-0, *LUCH*, *LUCH ago4-6* (a positive control showing de-repression of LUC luminescence) and seedlings from the F2 population of *dcl1-7* crossed to *LUCH*. In the F2 population, LUC luminescence was moderately increased in 12 out of 216 segregating seedlings (only six are shown in circles here).
(B) LUC image of Col-0, *LUCH*, *LUCH ago4-6* and *LUCH hyl1*. The *hyl1* mutation did not result in de-repression of LUC luminescence.
(C) LUC image of Col-0, *LUCH* and seedlings from the F3 population of *se-1* crossed to *LUCH*. The F2 plant was genotyped to be homozygous for *LUCH* and *rdr6-11* and heterozygous for *se-1*. Therefore, one quarter of the F3 progenies are theoretically homozygous for *se-1*. There was no apparent de-repression of *LUCH* by *se-1*.

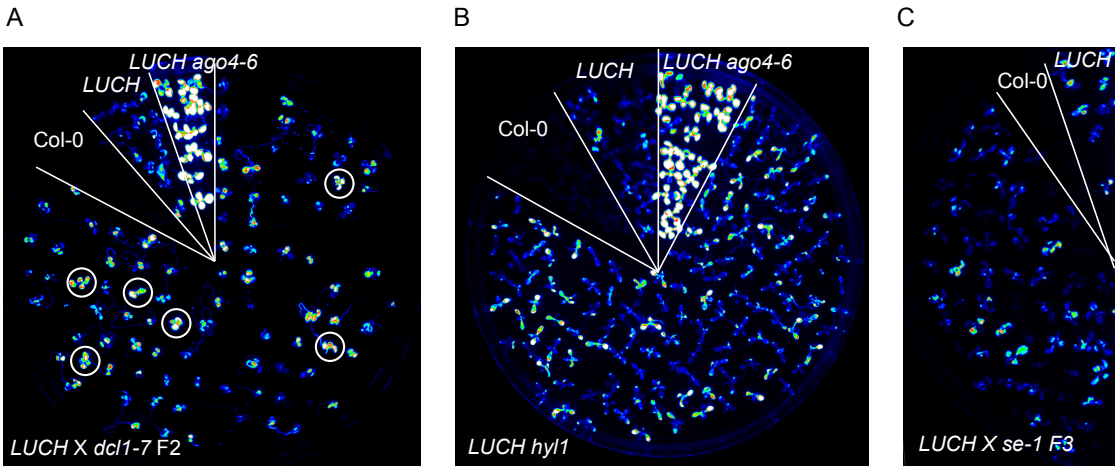


Figure 1.4 Molecular characteristics of *LUCH* associated with RdDM.

(A) Analysis of DNA methylation in *d35S* and the *LUC* coding region in *LUCH* by McrBC-PCR. The two *d35S* fragments are as diagrammed in Figure 1. − and + indicate McrBC-untreated and treated genomic DNA, respectively. "H_2O" is a negative control PCR without genomic DNA. McrBC digests methylated DNA to result in reduced PCR product amounts.

(B) Bisulfite sequencing analysis of cytosine methylation in *d35S* in *LUCH* in wild type and *ago4-6*. The top strand of *d35S* #3 in Figure 1 was analyzed.

(C) *d35S*-specific siRNA accumulation in the *LUCH* line as detected by northern blotting. The numbers indicate the amount of enriched small RNAs loaded into the gel. Col-0, wild type (with no transgene).

Figure 1.4

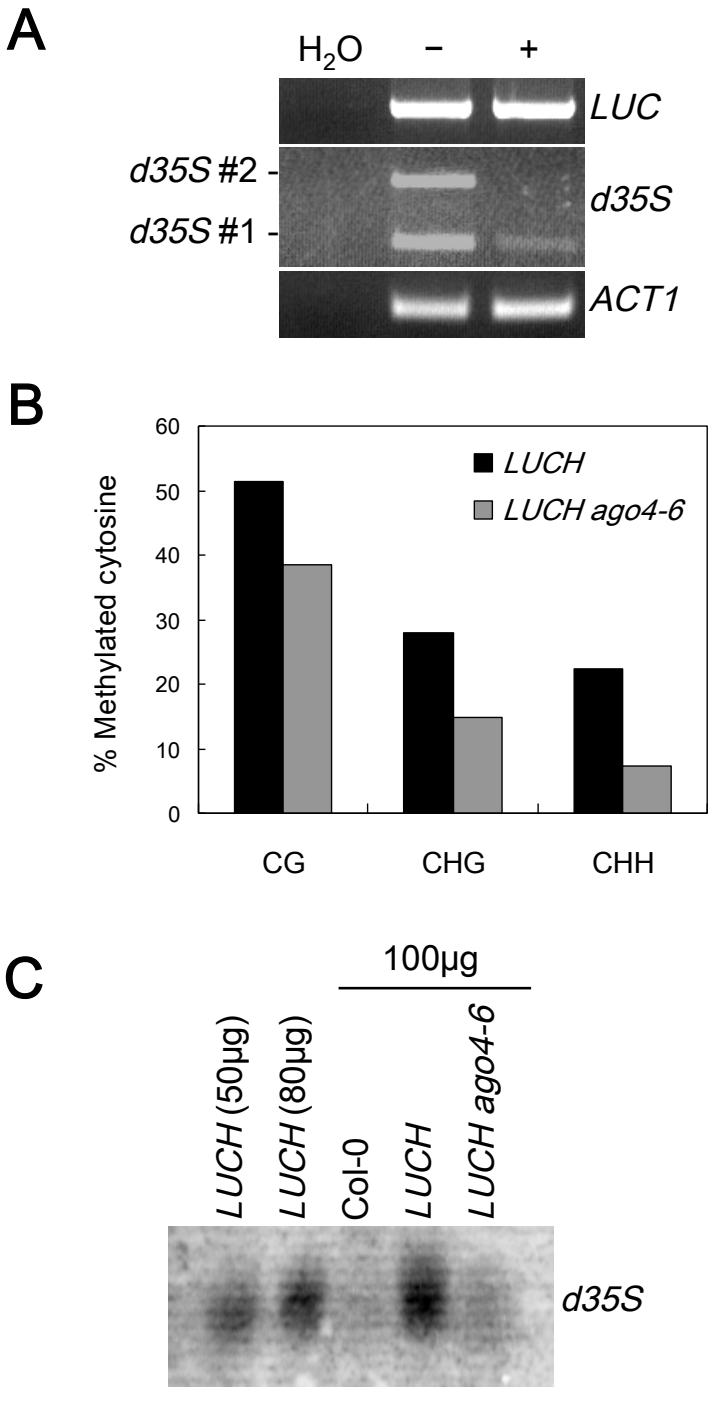

A
H2O
−
+
LUC
d35S #2
d35S #1
d35S
ACT1
B
% Methylated cytosine
60
50
40
30
20
10
0
LUCH
LUCH ago4-6
CG
CHG
CHH
C
100µg
LUCH (50µg)
LUCH (80µg)
Col-0
LUCH
LUCH ago4-6
d35S

Figure 1.5 De-repression of *LUCH* and *LUCH ros1-5* by the methylation inhibitor 5-aza-2′-deoxycytidine (5Aza-dC).

(A) Seedlings were grown on MS media for 10 days (mock) or on 7µg/ml 5Aza-dC-supplemented MS media for 2 weeks (5Aza-dC) followed by LUC luminescence imaging.
(B) RT-PCR analysis of *LUC* and *NPTII* expression in mock- or 5Aza-dC-treated *LUCH* and *LUCH ros1-5* seedlings. *UBIQUITIN5* (*UBQ5*) was used as a loading control. The RT (-) reactions were performed with *UBQ5* primers.

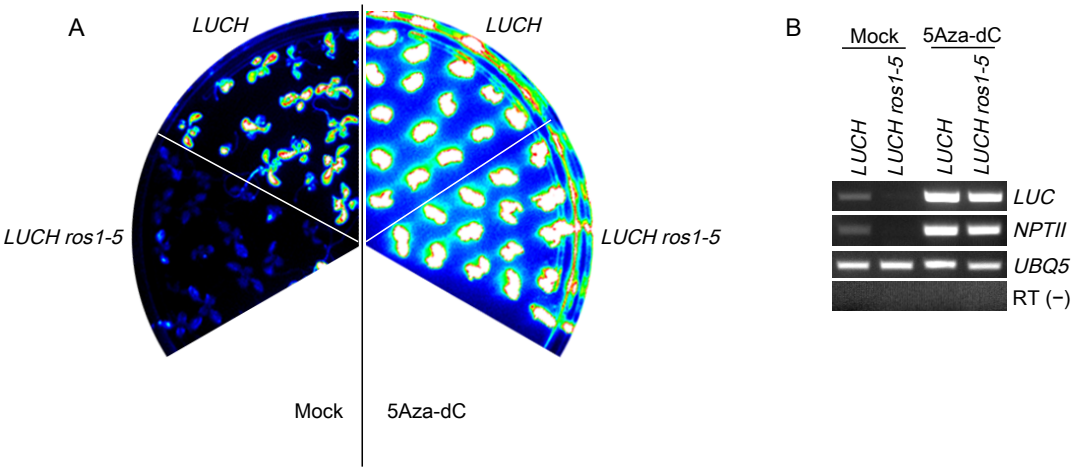


Figure 1.6 Transgene-specific small RNAs in the *LUCH* line as determined by deep sequencing.

(A) Size distribution of small RNAs mapping to the entire T-DNA containing *LUCH* and *d35S::NPTII* (total), *d35S* promoter in *LUCH* (*d35S*) or *LUC* coding sequence (*LUC*).
(B) Distribution and abundance of 24 nt small RNAs mapping to the *LUCH* and *d35S::NPTII* transgenes. Top and bottom figures indicate the distribution of 24 nt siRNAs from forward and reverse strands, respectively.
(C) Distribution and abundance of 24 nt small RNAs that are specific to each *d35S* promoter in the two transgenes. The 4% sequence variations between the *d35S* in the two transgenes allowed the identification of these transgene-specific *d35S* siRNAs. Small RNAs mapping to both strands were detected.

Figure 1.6 (A), (B)

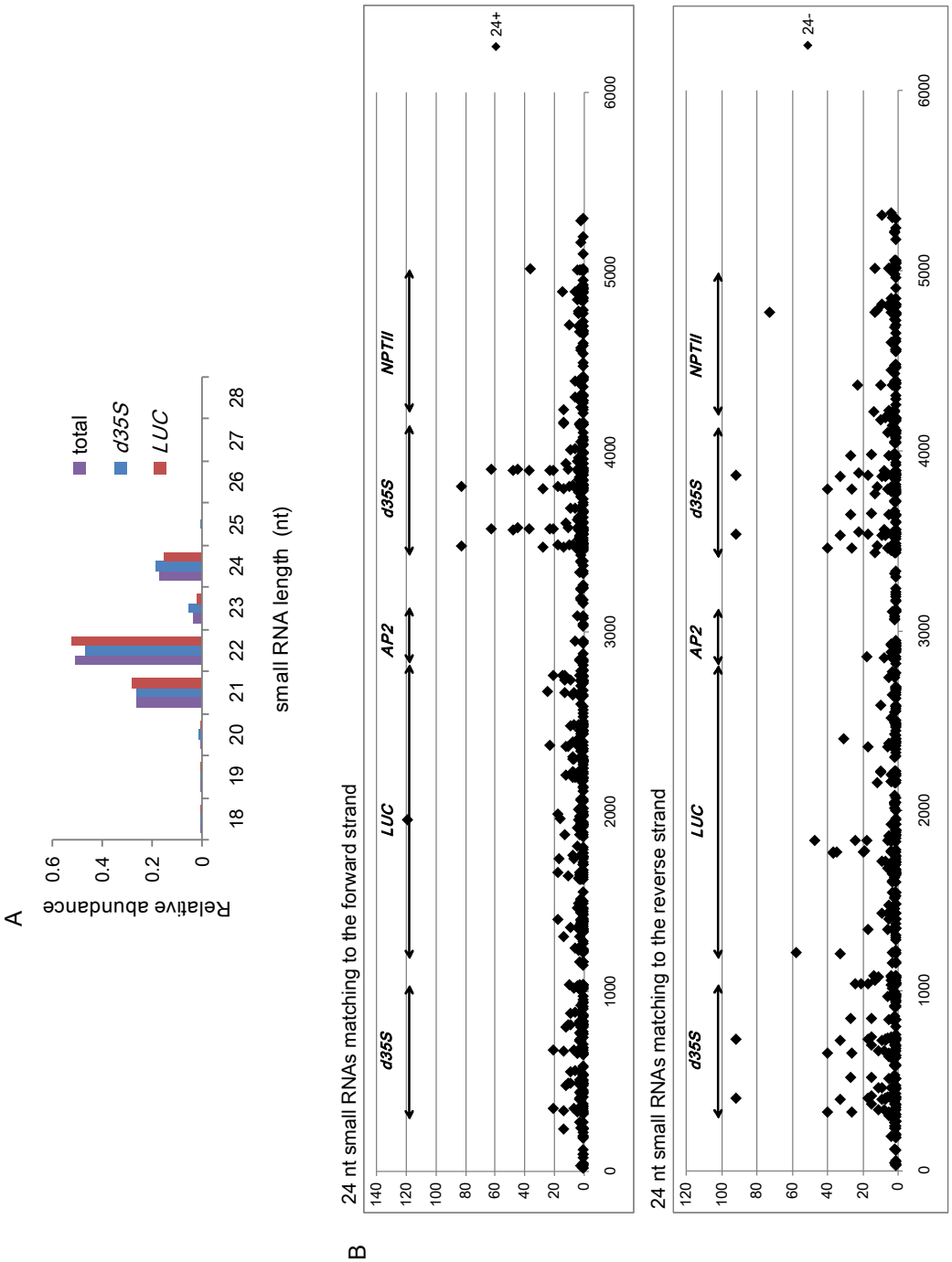

A
Relative abundance
0.6
0.4
0.2
0
18 19 20 21 22 23 24 25 26 27 28
small RNA length (nt)
total
d35S
LUC
B
24 nt small RNAs matching to the forward strand
d35S
LUC
AP2
d35S
NPTII
24+
24 nt small RNAs matching to the reverse strand
24-

Figure 1.6 (C)

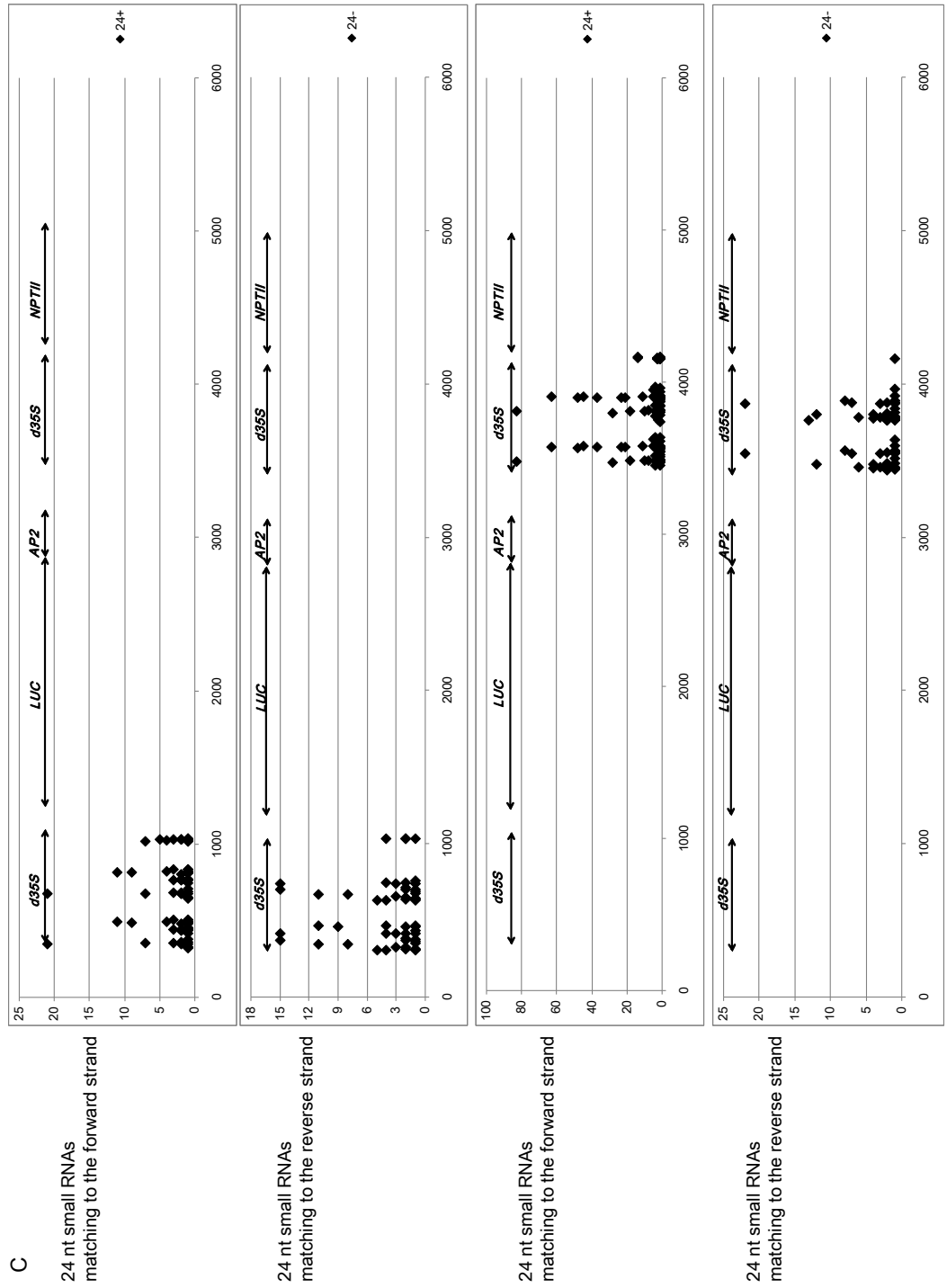

C
24 nt small RNAs matching to the forward strand
24 nt small RNAs matching to the reverse strand
24 nt small RNAs matching to the forward strand
24 nt small RNAs matching to the reverse strand
d35S
LUC
AP2
d35S
NPTII
24+
24-

Figure 1.7 The RdDM pathway is genetically required for the suppression of *LUCH* expression.

(A) De-repression of LUC luminescence in various RdDM mutants. Each spot represents an *Arabidopsis* seedling. The brighter the spots, the higher the LUC luminescence. Col-0, wild type (with no transgene).
(B) RT-PCR of *LUC* and *NPTII* in various RdDM mutants. *UBQ5* serves as a loading control. RT (-), *UBQ5* RT-PCR in which the reverse transcription was conducted in the absence of the reverse transcriptase.
(C) Analysis of cytosine methylation in *d35S* in *LUCH* in *ago4-6*, *ros1-5* and *mom1-5* mutants by McrBC-PCR. *ACT1* serves as an internal, unmethylated control.

Figure 1.7

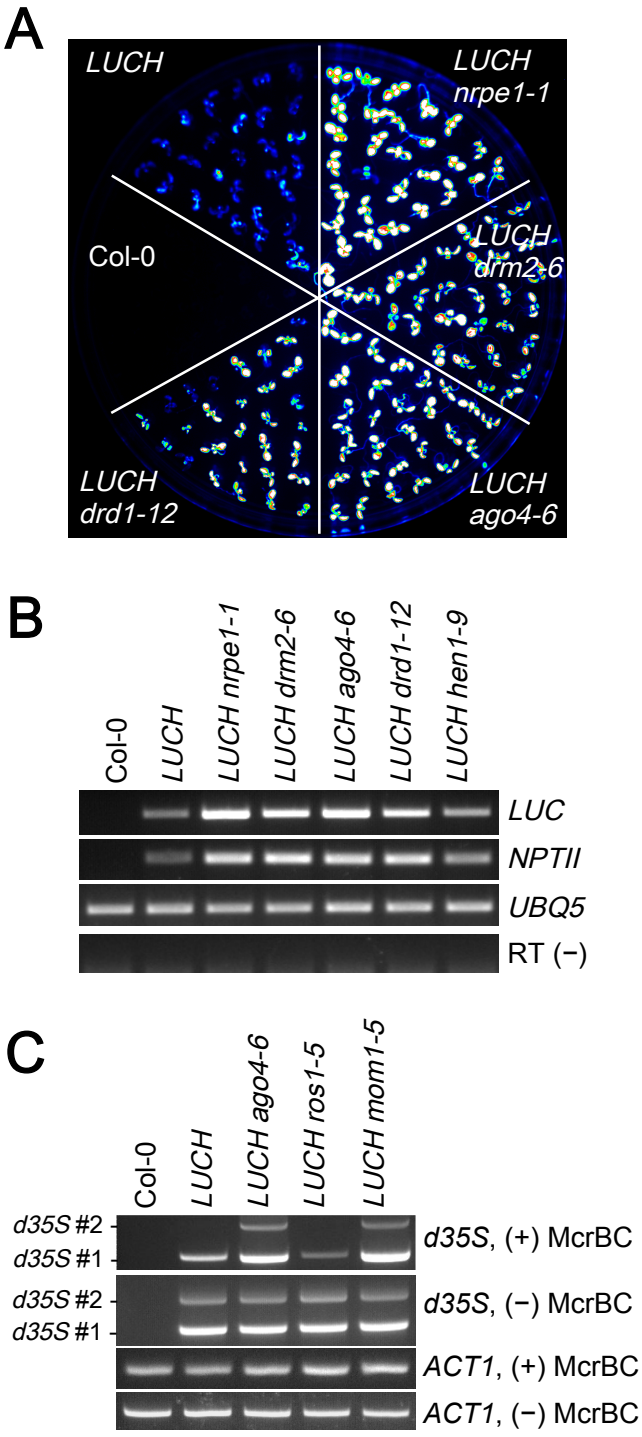

A
LUCH
LUCH nrpe1-1
Col-0
LUCH drm2-6
LUCH drd1-12
LUCH ago4-6
B
Col-0
LUCH
LUCH nrpe1-1
LUCH drm2-6
LUCH ago4-6
LUCH drd1-12
LUCH hen1-9
LUC
NPTII
UBQ5
RT (−)
C
Col-0
LUCH
LUCH ago4-6
LUCH ros1-5
LUCH mom1-5
d35S #2
d35S #1
d35S, (+) McrBC
d35S #2
d35S #1
d35S, (−) McrBC
ACT1, (+) McrBC
ACT1, (−) McrBC

Figure 1.8 Schematic diagrams of the gene structures and the mutations in the newly isolated mutant alleles in this study.

White and black rectangles indicate untranslated regions and coding exons, respectively. Lines represent introns. Arrows in *mom1-5* indicate the primers used for RT-PCR. The new *hen1* allele is not diagrammed because the exact nature of the mutation is not known. The allele was shown by a genetic complementation test with known *hen1* mutants to be a new *hen1* allele.

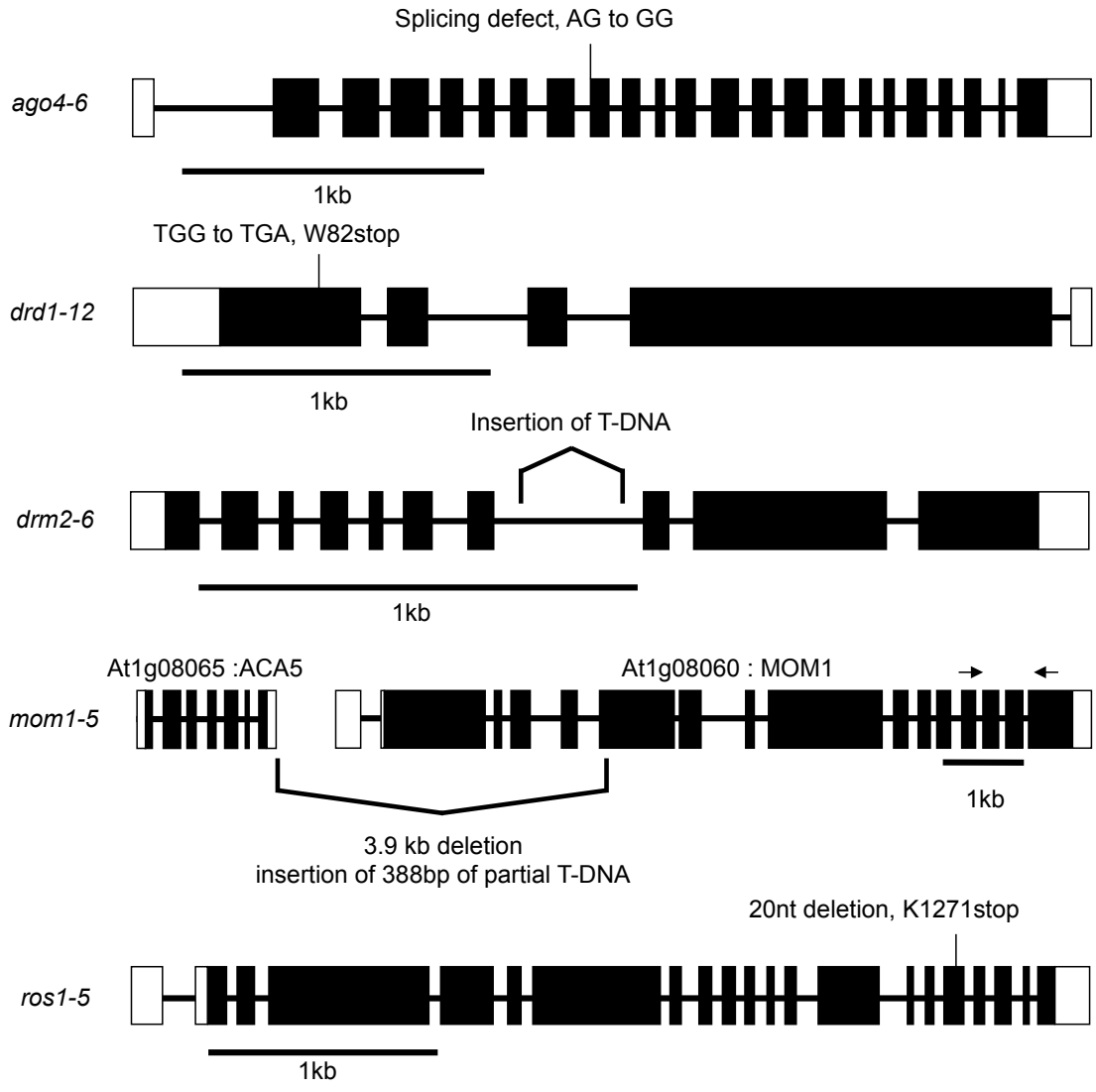


Figure 1.9 Southern blot analysis of cytosine methylation in *d35S*.

(A) Map of the transgenes. Bars represent the positions of the probe, which should hybridize to both transgene promoters.

(B) Genomic DNA was isolated from Col-0, *LUCH*, *LUCH ago4-6* and *LUCH drd1-12*, digested with cytosine methylation-sensitive *Alu*I and hybridized with the radiolabeled *d35S* probe. DNA bands are shifted downward in *ago4-6* and *drd1-12*, indicating that DNA methylation in *d35S* is decreased in *ago4-6* and *drd1-12*. Though the juxtaposed lanes are discontinuous, they are from a single gel. The phosphor-image was taken from a single membrane.

(C) Southern blot analysis of Col-0, *LUCH*, *LUCH ago4-6*, *LUCH mom1-5* and *LUCH drm2-6*. DNA bands are shifted downward to a lesser extent in *mom1-5* than in *ago4-6* or *drm2-6*.

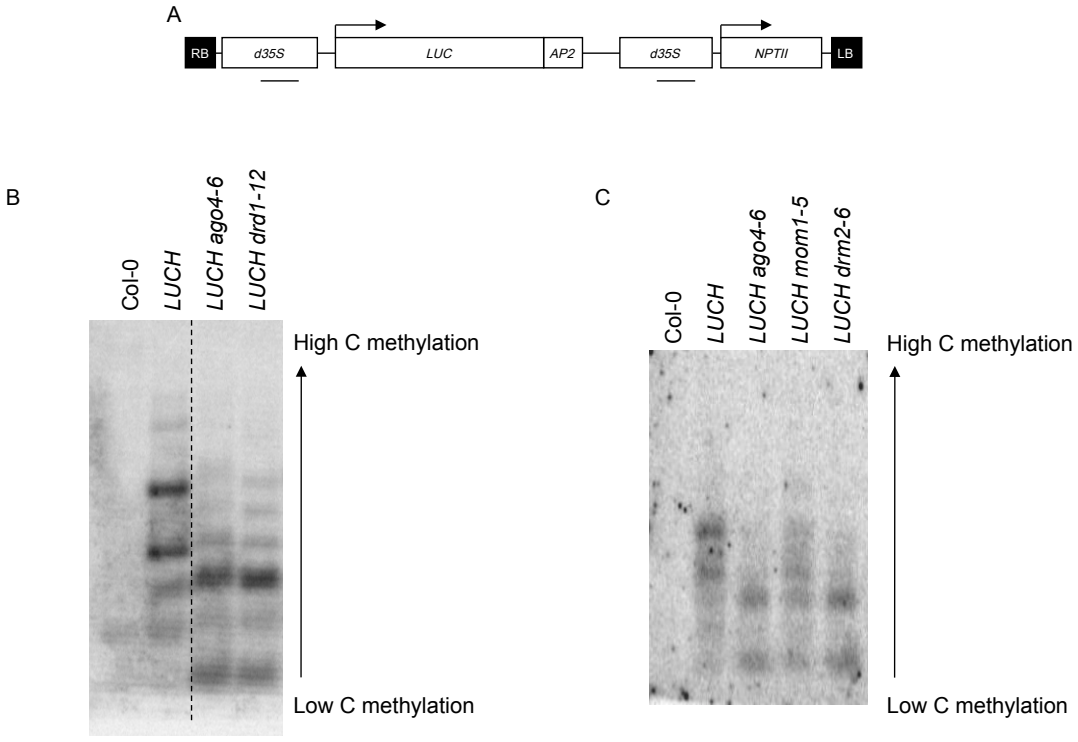


Figure 1.10 *LUCH* is not repressed by *MET1* or *CMT3*.

(A) LUC imaging of seedlings from an F3 population of *cmt3-7* crossed to *LUCH*. The F2 plant was genotyped to be homozygous for *LUCH* and *rdr6-11* and heterozygous for *cmt3-7*. Therefore, one quarter of the F3 progenies are theoretically homozygous for *cmt3-7*. If CMT3 represses *LUCH*, de-repression of *LUCH* is expected in one quarter of the seedlings. No such de-repression was observed.

(B) LUC imaging of an F2 population of *met1-3* crossed to *LUCH*. Seedlings with weakly de-repressed LUC signal were identified (circled), genotyped, and found not to be homozygous for *met1-3*. Note that *MET1* and *LUCH* are not linked, such that 3/16 of the seedlings are expected to be *LUCH* (or *LUCH*/+) *met1-3*.

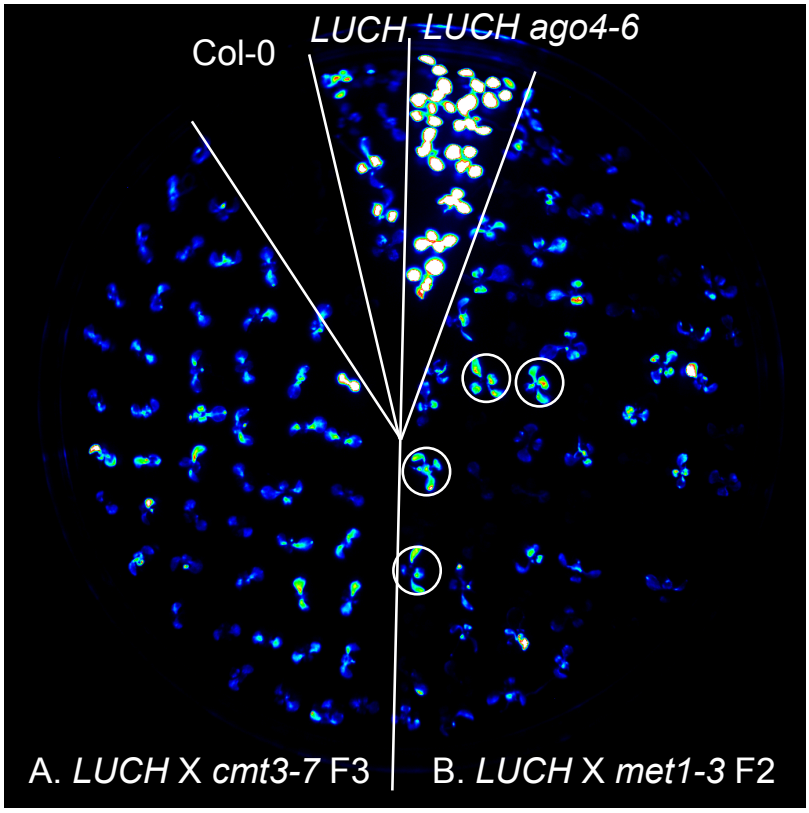


Figure 1.11 *LUCH* is regulated by MOM1.

(A) De-repression of LUC luminescence in *LUCH mom1-5*.

(B) RT-PCR of *LUC*, *NPTII* and *MOM1* in wild type (Col-0), *LUCH* and *LUCH mom1-5*.

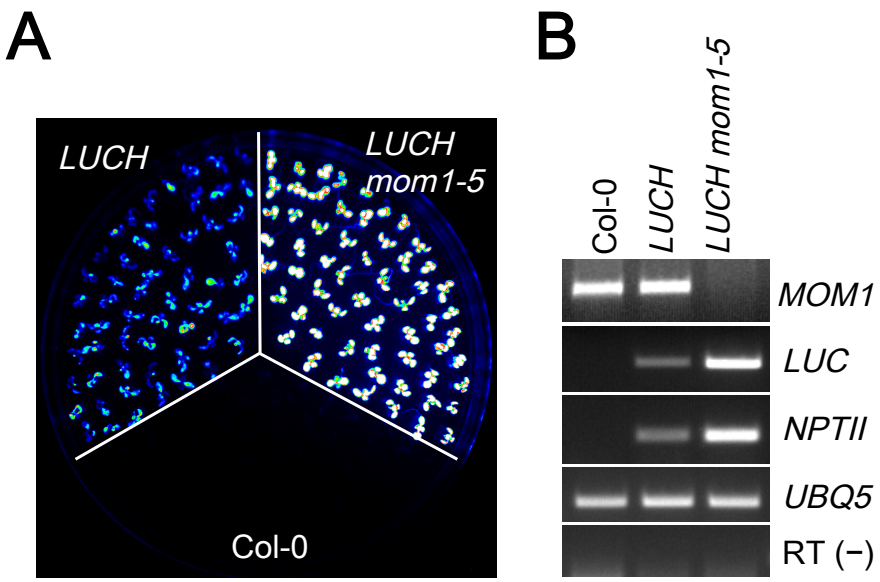
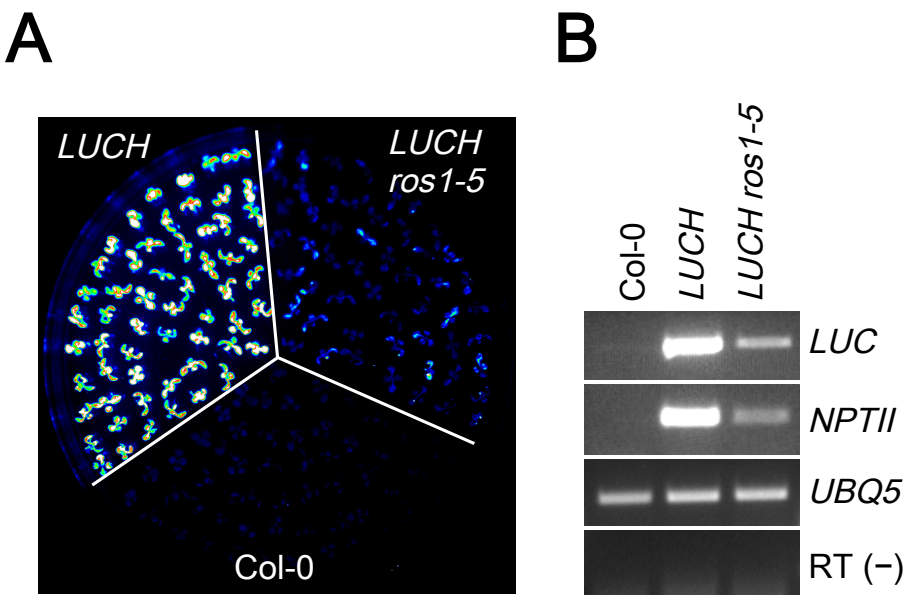


Figure 1.12 *LUCH* is targeted by ROS1-mediated DNA demethylation.

(A) Reduction of LUC luminescence in *LUCH ros1-5*.

(B) RT-PCR of *LUC* and *NPTII* in wild type (Col-0), *LUCH* and *LUCH ros1-5*.



TABLES

Table 1.1 DNA oligonucleotide used in this study.

Name	Sequence	Purpose
AP2p26	CCGGTTTGATGGTCGGGCCTCGAC	partial AP2 amplification
AP2p28	GTTTTTTTTAATTACCTTTAGAAAAAGGGA	partial AP2 amplification
Rlucp1	AGGGGATCCACCATGGCTTCGAAAGTTTATGATC CAGAAC	LUC amplification
Rlucp2	TCTAGGATCCTTGTTCATTTTTGAGAACTCGCTC	LUC amplification
lucp6	GCACCCGGGGAAGACGCCAAAAACATAAAAGAAA	McrBC-PCR, southern blot
lucp7	GGACCCGGGTGCGATCTTTCCGCCCTTCTTGGCCT	McrBC-PCR, southern blot
Actin1-F	CCAAGCAGCATGAAGATCAA	McrBC-PCR
Actin1-R	TGAACAATCGATGGACCTGA	McrBC-PCR
35Sf	CAAAGCAAGTGGATTGATGTGA	McrBC-PCR, southern blot
35Sr	TTTCCACGATGCTCCTCGT	Southern blot
LUC 0.13k R	TATGTGCATCTGTAAAAGCAA	McrBC-PCR
YZ 35S Bis F	AttAtTGTyGGtAGAGGtATtTTGAAyGATAGtt	Bisulfite sequencing
YZ LUC Bis R	CATCTaTAAAAaCAATTaTTCCAaaAACCAaa	Bisulfite sequencing
N_UBQ5	GGTGCTAAGAAGAGGAAGAAT	RT-PCR, loading control
C_UBQ5	CTCCTTCTTTCTGGTAAACGT	RT-PCR, loading control
LUCmF5	CTCCCCTCTCTAAGGAAGTCG	RT-PCR for LUC
LUCmR5	CCAGAATGTAGCCATCCATC	RT-PCR for LUC
Kan-RT-F	AGGTTCCATCTGCCAGGTATCA	RT-PCR for NPTII
Kan-RT-R	CCCGGTATCCAGATCCACAA	RT-PCR for NPTII
At1g08060-F10	CTCCTATGCCATTATCTTCG	RT-PCR for MOM1
At1g08060-R10	AACTGATGGAGTTGGAGCTA	RT-PCR for MOM1

Table 2.2 Sequences of 24 nt small RNAs mapping specifically to each *d35S* in the two transgenes.

Small RNAs were classified based on their transgene origin and strandedness. The sequences of the small RNAs, the number of reads and the positions of their 5′ nucleotides along the construct as shown in Figure 1.6 are shown.

Sequence	Reads	Position
LUCH-specific 24nt small RNAs matching to the forward strand of *d35S*		
GAAGACCAAAGGGCAATTGAGACT	21	349
GAAGACCAAAGGGCAATTGAGACT	21	676
AAAGGAAAGGCCATCGTTGAAGAT	11	490
AAAGGAAAGGCCATCGTTGAAGAT	11	817
ATAAAGGAAAGGCCATCGTTGAAG	9	488
ATAAAGGAAAGGCCATCGTTGAAG	9	815
AGACCAAAGGGCAATTGAGACTTT	7	351
AGACCAAAGGGCAATTGAGACTTT	7	678
ATTTCATTTGGAGAGGACCTCGAG	7	1020
AGGACCTCGAGAATTCTCAACACA	5	1033
AGGAAAGGCCATCGTTGAAGATGC	4	492
AGGAAAGGCCATCGTTGAAGATGC	4	819
TTCATTTGGAGAGGACCTCGAGAA	4	1022
AAAGGGCAATTGAGACTTTTCAAC	3	356
GAAGATAGTGGAAAAGGAAGGTGG	3	441
TGAAGATGCCTCTGCCGACAGTGG	3	507
AAAGGGCAATTGAGACTTTTCAAC	3	683
GAAGATAGTGGAAAAGGAAGGTGG	3	768
TGAAGATGCCTCTGCCGACAGTGG	3	834
AGAGGACCTCGAGAATTCTCAACA	3	1031
AAGACCAAAGGGCAATTGAGACTT	2	350
GACCAAAGGGCAATTGAGACTTTT	2	352
AGGGCAATTGAGACTTTTCAACAA	2	358
TATTGTGAAGATAGTGGAAAAGGA	2	435
GTGAAGATAGTGGAAAAGGAAGGT	2	439
ATAGTGGAAAAGGAAGGTGGCTCC	2	445
CATCATTGCGATAAAGGAAAGGCC	2	478
AAGACCAAAGGGCAATTGAGACTT	2	677
GACCAAAGGGCAATTGAGACTTTT	2	679
AGGGCAATTGAGACTTTTCAACAA	2	685
TATTGTGAAGATAGTGGAAAAGGA	2	762
GTGAAGATAGTGGAAAAGGAAGGT	2	766
ATAGTGGAAAAGGAAGGTGGCTCC	2	772
CATCATTGCGATAAAGGAAAGGCC	2	805
GGAGAGGACCTCGAGAATTCTCAA	2	1029
GAGGACCTCGAGAATTCTCAACAC	2	1032
ACTCCAAAAATATCAAAGATACAG	1	320
CAAAAATATCAAAGATACAGTCTC	1	324
AAATATCAAAGATACAGTCTCAGA	1	327
AGAAGACCAAAGGGCAATTGAGAC	1	348
AATTGAGACTTTTCAACAAAGGGT	1	363
ACAAAGGGTAATATCCGGAAACCT	1	378

Sequence	Reads	Position
CCCAGCTATCTGTCACTTTATTGT	1	417
ATTGTGAAGATAGTGGAAAAGGAA	1	436
TGAAGATAGTGGAAAAGGAAGGTG	1	440
AGTGGAAAAGGAAGGTGGCTCCTA	1	447
AGGAAGGTGGCTCCTACAAATGCC	1	455
ATTGCGATAAAGGAAAGGCCATCG	1	482
TAAAGGAAAGGCCATCGTTGAAGA	1	489
AAGGAAAGGCCATCGTTGAAGATG	1	491
GGAAAGGCCATCGTTGAAGATGCC	1	493
TTGAAGATGCCTCTGCCGACAGTG	1	506
GAAGATGCCTCTGCCGACAGTGGT	1	508
ACTCCAAAAATATCAAAGATACAG	1	647
CAAAAATATCAAAGATACAGTCTC	1	651
AAATATCAAAGATACAGTCTCAGA	1	654
AGAAGACCAAAGGGCAATTGAGAC	1	675
AATTGAGACTTTTCAACAAAGGGT	1	690
ACAAAGGGTAATATCCGGAAACCT	1	705
CCCAGCTATCTGTCACTTTATTGT	1	744
ATTGTGAAGATAGTGGAAAAGGAA	1	763
TGAAGATAGTGGAAAAGGAAGGTG	1	767
AGTGGAAAAGGAAGGTGGCTCCTA	1	774
ATTGCGATAAAGGAAAGGCCATCG	1	809
TAAAGGAAAGGCCATCGTTGAAGA	1	816
AAGGAAAGGCCATCGTTGAAGATG	1	818
GGAAAGGCCATCGTTGAAGATGCC	1	820
TTGAAGATGCCTCTGCCGACAGTG	1	833
GAAGATGCCTCTGCCGACAGTGGT	1	835
AGTTCATTTCATTTGGAGAGGACC	1	1015
TTTCATTTGGAGAGGACCTCGAGA	1	1021
TCATTTGGAGAGGACCTCGAGAAT	1	1023
GAGAGGACCTCGAGAATTCTCAAC	1	1030
ACCTCGAGAATTCTCAACACAACA	1	1036
LUCH-specific 24nt small RNAs matching to the reverse strand of *d35S*		
TCCGGATATTACCCTTTGTTGAAA	15	373
AATAAAGTGACAGATAGCTGGGCA	15	415
TCCGGATATTACCCTTTGTTGAAA	15	700
AATAAAGTGACAGATAGCTGGGCA	15	742
ATTGCCCTTTGGTCTTCTGAGACT	11	342
GCAATGATGGCATTTGTAGGAGCC	11	463
ATTGCCCTTTGGTCTTCTGAGACT	11	669
ATGATGGCATTTGTAGGAGCCACC	9	460
AATTGCCCTTTGGTCTTCTGAGAC	8	343
AATTGCCCTTTGGTCTTCTGAGAC	8	670
ATTTTTGGAGTAGACAAGTGTGTC	5	307
ATTTTTGGAGTAGACAAGTGTGTC	5	634
TTGGAGTAGACAAGTGTGTCGTGC	4	303
ACAATAAAGTGACAGATAGCTGGG	4	417
CGCAATGATGGCATTTGTAGGAGC	4	464
TTGGAGTAGACAAGTGTGTCGTGC	4	630
ACAATAAAGTGACAGATAGCTGGG	4	744
TGTGTTGAGAATTCTCGAGGTCCT	4	1033
TCTGAGACTGTATCTTTGATATTT	3	327

Sequence	Reads	Position
TAAAGTGACAGATAGCTGGGCAAT	3	413
TCTGAGACTGTATCTTTGATATTT	3	654
TAAAGTGACAGATAGCTGGGCAAT	3	740
GATATTTTTGGAGTAGACAAGTGT	2	310
CTGAGACTGTATCTTTGATATTTT	2	326
CCGGATATTACCCTTTGTTGAAAA	2	372
TCCGAGGAGGTTTCCGGATATTAC	2	385
CAATAAAGTGACAGATAGCTGGGC	2	416
TGATGGCATTTGTAGGAGCCACCT	2	459
GATATTTTTGGAGTAGACAAGTGT	2	637
CTGAGACTGTATCTTTGATATTTT	2	653
CCGGATATTACCCTTTGTTGAAAA	2	699
TCCGAGGAGGTTTCCGGATATTAC	2	712
CAATAAAGTGACAGATAGCTGGGC	2	743
GTGTTGAGAATTCTCGAGGTCCTC	2	1032
TTGTGTTGAGAATTCTCGAGGTCC	2	1034
TTTGGAGTAGACAAGTGTGTCGTG	1	304
TTTGATATTTTTGGAGTAGACAAG	1	313
TCTTTGATATTTTTGGAGTAGACA	1	315
AAAGTCTCAATTGCCCTTTGGTCT	1	351
ACCCTTTGTTGAAAAGTCTCAATT	1	363
TTCCGGATATTACCCTTTGTTGAA	1	374
AAAGTGACAGATAGCTGGGCAATG	1	412
TTTTCCACTATCTTCACAATAAAG	1	432
ATGGCATTTGTAGGAGCCACCTTC	1	457
TCGCAATGATGGCATTTGTAGGAG	1	465
TTTGGAGTAGACAAGTGTGTCGTG	1	631
TTTGATATTTTTGGAGTAGACAAG	1	640
TCTTTGATATTTTTGGAGTAGACA	1	642
AAAGTCTCAATTGCCCTTTGGTCT	1	678
ACCCTTTGTTGAAAAGTCTCAATT	1	690
TTCCGGATATTACCCTTTGTTGAA	1	701
AAAGTGACAGATAGCTGGGCAATG	1	739
TTTTCCACTATCTTCACAATAAAG	1	759
TGTTGTGTTGAGAATTCTCGAGGT	1	1036
d35S::NPTII-specific 24nt small RNAs matching to the forward strand of *d35S*		
GAAGACCAAAGGGCTATTGAGACT	83	3480
GAAGACCAAAGGGCTATTGAGACT	83	3807
ACAGTAGAAAAGGAAGGTGGCACC	63	3576
ACAGTAGAAAAGGAAGGTGGCACC	63	3903
AAAAGGACAGTAGAAAAGGAAGGT	48	3570
AAAAGGACAGTAGAAAAGGAAGGT	48	3897
AGTAGAAAAGGAAGGTGGCACCTA	45	3578
AGTAGAAAAGGAAGGTGGCACCTA	45	3905
AAGGACAGTAGAAAAGGAAGGTGG	37	3572
AAGGACAGTAGAAAAGGAAGGTGG	37	3899
ACAGTCTCAGAAGACCAAAGGGCT	28	3471
ACAGTCTCAGAAGACCAAAGGGCT	28	3798
AAAGGACAGTAGAAAAGGAAGGTG	23	3571
AAAGGACAGTAGAAAAGGAAGGTG	23	3898
AGGACAGTAGAAAAGGAAGGTGGC	21	3573
AGGACAGTAGAAAAGGAAGGTGGC	21	3900

Sequence	Reads	Position
AAGACCAAAGGGCTATTGAGACTT	18	3481
AAGACCAAAGGGCTATTGAGACTT	18	3808
AGAGGACACGCTGAAATCACCAGT	14	4162
GAGGACACGCTGAAATCACCAGTC	14	4163
CAGTAGAAAAGGAAGGTGGCACCT	11	3577
CAGTAGAAAAGGAAGGTGGCACCT	11	3904
AGACCAAAGGGCTATTGAGACTTT	10	3482
AGACCAAAGGGCTATTGAGACTTT	10	3809
AAAGGGCTATTGAGACTTTTCAAC	8	3487
AAAGGGCTATTGAGACTTTTCAAC	8	3814
GTAGAAAAGGAAGGTGGCACCTAC	5	3579
AGGAAAGGCTATCGTTCAAGATGC	5	3623
GTAGAAAAGGAAGGTGGCACCTAC	5	3906
AGGAAAGGCTATCGTTCAAGATGC	5	3950
CTCCAAGAATATCAAAGATACAGT	4	3452
GACCAAAGGGCTATTGAGACTTTT	4	3483
AATATCGGGAAACCTCCTCGGATT	4	3518
ATCAAAAGGACAGTAGAAAAGGAA	4	3567
AAAGGAAAGGCTATCGTTCAAGAT	4	3621
TCAAGATGCCTCTGCCGACAGTGG	4	3638
CTCCAAGAATATCAAAGATACAGT	4	3779
GACCAAAGGGCTATTGAGACTTTT	4	3810
AATATCGGGAAACCTCCTCGGATT	4	3845
ATCAAAAGGACAGTAGAAAAGGAA	4	3894
AAAGGAAAGGCTATCGTTCAAGAT	4	3948
TCAAGATGCCTCTGCCGACAGTGG	4	3965
CAGAAGACCAAAGGGCTATTGAGA	3	3478
AGAAGACCAAAGGGCTATTGAGAC	3	3479
AAGGAAAGGCTATCGTTCAAGATG	3	3622
CAGAAGACCAAAGGGCTATTGAGA	3	3805
AGAAGACCAAAGGGCTATTGAGAC	3	3806
AAGGAAAGGCTATCGTTCAAGATG	3	3949
TTCATTTGGAGAGGACACGCTGAA	3	4153
TCATTTGGAGAGGACACGCTGAAA	3	4154
GAGAGGACACGCTGAAATCACCAG	3	4161
ATCATTGCGATAAAGGAAAGGCTA	2	3610
CATTGCGATAAAGGAAAGGCTATC	2	3612
ATGGTGGAGCACGACACTCTCGTC	2	3753
ATCATTGCGATAAAGGAAAGGCTA	2	3937
CATTGCGATAAAGGAAAGGCTATC	2	3939
CATTTCATTTGGAGAGGACACGCT	2	4150
CAAGAATATCAAAGATACAGTCTC	1	3455
GTCTCAGAAGACCAAAGGGCTATT	1	3474
CTCAGAAGACCAAAGGGCTATTGA	1	3476
CAAAGGGCTATTGAGACTTTTCAA	1	3486
AAGGGCTATTGAGACTTTTCAACA	1	3488
AGGGCTATTGAGACTTTTCAACAA	1	3489
AAAGGGTAATATCGGGAAACCTCC	1	3511
TATCGGGAAACCTCCTCGGATTCC	1	3520
CATTGCCCAGCTATCTGTCACTTC	1	3543
ATTGCCCAGCTATCTGTCACTTCA	1	3544
TGCCCAGCTATCTGTCACTTCATC	1	3546

Sequence	Reads	Position
TCACTTCATCAAAAGGACAGTAGA	1	3560
CATCAAAAGGACAGTAGAAAAGGA	1	3566
TCAAAAGGACAGTAGAAAAGGAAG	1	3568
CAAAAGGACAGTAGAAAAGGAAGG	1	3569
GACAGTAGAAAAGGAAGGTGGCAC	1	3575
AGAAAAGGAAGGTGGCACCTACAA	1	3581
GAAAAGGAAGGTGGCACCTACAAA	1	3582
CATCATTGCGATAAAGGAAAGGCT	1	3609
TTCAAGATGCCTCTGCCGACAGTG	1	3637
GATGTGATAACATGGTGGAGCACG	1	3742
CAAGAATATCAAAGATACAGTCTC	1	3782
GTCTCAGAAGACCAAAGGGCTATT	1	3801
CTCAGAAGACCAAAGGGCTATTGA	1	3803
CAAAGGGCTATTGAGACTTTTCAA	1	3813
AAGGGCTATTGAGACTTTTCAACA	1	3815
AGGGCTATTGAGACTTTTCAACAA	1	3816
AAAGGGTAATATCGGGAAACCTCC	1	3838
TATCGGGAAACCTCCTCGGATTCC	1	3847
CATTGCCCAGCTATCTGTCACTTC	1	3870
ATTGCCCAGCTATCTGTCACTTCA	1	3871
TGCCCAGCTATCTGTCACTTCATC	1	3873
TCACTTCATCAAAAGGACAGTAGA	1	3887
CATCAAAAGGACAGTAGAAAAGGA	1	3893
TCAAAAGGACAGTAGAAAAGGAAG	1	3895
CAAAAGGACAGTAGAAAAGGAAGG	1	3896
GACAGTAGAAAAGGAAGGTGGCAC	1	3902
AGAAAAGGAAGGTGGCACCTACAA	1	3908
GAAAAGGAAGGTGGCACCTACAAA	1	3909
CATCATTGCGATAAAGGAAAGGCT	1	3936
TTCAAGATGCCTCTGCCGACAGTG	1	3964
ATTTCATTTGGAGAGGACACGCTG	1	4151
TTTCATTTGGAGAGGACACGCTGA	1	4152
ATTTGGAGAGGACACGCTGAAATC	1	4156
TTTGGAGAGGACACGCTGAAATCA	1	4157
AGGACACGCTGAAATCACCAGTCT	1	4164
d35S::NPTII-specific 24nt small RNAs matching to the reverse strand of *d35S*		
GATGAAGTGACAGATAGCTGGGCA	22	3546
GATGAAGTGACAGATAGCTGGGCA	22	3873
TCTTGGAGTAGACGAGAGTGTCGT	13	3763
ATAGCCCTTTGGTCTTCTGAGACT	12	3473
ATAGCCCTTTGGTCTTCTGAGACT	12	3800
TTTTCTACTGTCCTTTTGATGAAG	8	3563
TTTTCTACTGTCCTTTTGATGAAG	8	3890
TGATGAAGTGACAGATAGCTGGGC	7	3547
TGATGAAGTGACAGATAGCTGGGC	7	3874
CTGAGACTGTATCTTTGATATTCT	6	3457
CTGAGACTGTATCTTTGATATTCT	6	3784
ATCTTTGATATTCTTGGAGTAGAC	4	3447
AATAGCCCTTTGGTCTTCTGAGAC	4	3474
ATCTTTGATATTCTTGGAGTAGAC	4	3774
AATAGCCCTTTGGTCTTCTGAGAC	4	3801
AGACTGTATCTTTGATATTCTTGG	3	3454

Sequence	Reads	Position
GAAGTGACAGATAGCTGGGCAATG	3	3543
AGACTGTATCTTTGATATTCTTGG	3	3781
GAAGTGACAGATAGCTGGGCAATG	3	3870
ATTCTTGGAGTAGACGAGAGTGTC	2	3438
AGCCCTTTGGTCTTCTGAGACTGT	2	3471
AAAGTCTCAATAGCCCTTTGGTCT	2	3482
CTTTTGATGAAGTGACAGATAGCT	2	3551
TTGGAGTAGACGAGAGTGTCGTGC	2	3761
ATTCTTGGAGTAGACGAGAGTGTC	2	3765
AGCCCTTTGGTCTTCTGAGACTGT	2	3798
AAAGTCTCAATAGCCCTTTGGTCT	2	3809
CTTTTGATGAAGTGACAGATAGCT	2	3878
TTGATATTCTTGGAGTAGACGAGA	1	3443
TCTTTGATATTCTTGGAGTAGACG	1	3446
ACTGTATCTTTGATATTCTTGGAG	1	3452
TCTGAGACTGTATCTTTGATATTC	1	3458
AAAAGTCTCAATAGCCCTTTGGTC	1	3483
AGGAGGTTTCCCGATATTACCCTT	1	3512
TGAAGTGACAGATAGCTGGGCAAT	1	3544
TTTGATGAAGTGACAGATAGCTGG	1	3549
TTCTACTGTCCTTTTGATGAAGTG	1	3561
TTTCTACTGTCCTTTTGATGAAGT	1	3562
ATCGCAATGATGGCATTTGTAGGT	1	3597
ACTGTCGGCAGAGGCATCTTGAAC	1	3636
CTTGGAGTAGACGAGAGTGTCGTG	1	3762
TTGATATTCTTGGAGTAGACGAGA	1	3770
TCTTTGATATTCTTGGAGTAGACG	1	3773
ACTGTATCTTTGATATTCTTGGAG	1	3779
TCTGAGACTGTATCTTTGATATTC	1	3785
AAAAGTCTCAATAGCCCTTTGGTC	1	3810
AGGAGGTTTCCCGATATTACCCTT	1	3839
TGAAGTGACAGATAGCTGGGCAAT	1	3871
TTTGATGAAGTGACAGATAGCTGG	1	3876
TTCTACTGTCCTTTTGATGAAGTG	1	3888
TTTCTACTGTCCTTTTGATGAAGT	1	3889
ATCGCAATGATGGCATTTGTAGGT	1	3924
ACTGTCGGCAGAGGCATCTTGAAC	1	3963
AGACTGGTGATTTCAGCGTGTCCT	1	4164

REFERENCES

1. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews Genetics 11: 204-220.

2. Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nature reviews Molecular cell biology 12: 483-492.

3. Furner IJ, Matzke M (2011) Methylation and demethylation of the Arabidopsis genome. Current opinion in plant biology 14: 137-141.

4. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523-536.

5. Amedeo P, Habu Y, Afsar K, Scheid OM, Paszkowski J (2000) Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. Nature 405: 203-206.

6. Vaillant I, Schubert I, Tourmente S, Mathieu O (2006) MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. EMBO Rep 7: 1273-1278.

7. Yokthongwattana C, Bucher E, Caikovski M, Vaillant I, Nicolet J, et al. (2010) MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. EMBO J 29: 340-351.

8. Zheng X, Pontes O, Zhu J, Miki D, Zhang F, et al. (2008) ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. Nature 455: 1259-1262.

9. Gong Z, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, David L, et al. (2002) ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell 111: 803-814.

10. Chen X (2004) A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 303: 2022-2025.

11. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes & development 18: 2368-2379.

12. Mourrain P, Beclin C, Elmayan T, Feuerbach F, Godon C, et al. (2000) Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. Cell 101: 533-542.

13. Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC (2000) An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101: 543-553.

14. Chen X (2009) Small RNAs and their roles in plant development. Annual review of cell and developmental biology 25: 21-44.

15. Matzke M, Aufsatz W, Kanno T, Daxinger L, Papp I, et al. (2004) Genetic analysis of RNA-mediated transcriptional gene silencing. Biochimica et biophysica acta 1677: 129-141.

16. Aufsatz W, Mette MF, Matzke AJ, Matzke M (2004) The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. Plant molecular biology 54: 793-804.

17. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, et al. (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science 292: 2077-2080.

18. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes & Development 18: 2368-2379.

19. Robinson-Beers K, Pruitt RE, Gasser CS (1992) Ovule Development in Wild-Type Arabidopsis and Two Female-Sterile Mutants. The Plant Cell Online 4: 1237-1249.

20. Prigge MJ, Wagner DR (2001) The Arabidopsis SERRATE Gene Encodes a Zinc-Finger Protein Required for Normal Shoot Development. The Plant Cell Online 13: 1263-1280.

21. Lu C, Fedoroff N (2000) A Mutation in the Arabidopsis HYL1 Gene Encoding a dsRNA Binding Protein Affects Responses to Abscisic Acid, Auxin, and Cytokinin. The Plant Cell Online 12: 2351-2366.

22. Saze H, Scheid OM, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nature Genetics 34: 65-69.

23. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, et al. (2001) Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. Science 292: 2077-2080.

24. Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, et al. (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet 37: 761-765.

25. Chinnusamy V, Stevenson B, Lee BH, Zhu JK (2002) Screening for gene regulation mutants by bioluminescence imaging. Science's STKE : signal transduction knowledge environment 2002: pl10.

26. von Arnim AG, Deng XW, Stacey MG (1998) Cloning vectors for the expression of green fluorescent protein fusion proteins in transgenic plants. Gene 221: 35-43.

27. Hajdukiewicz P, Svab Z, Maliga P (1994) The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation. Plant Molecular Biology 25: 989-994.

28. Sambrook J, Fritsch, E.F., Maniatis, T., (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

29. Liu Y-G, Chen Y (2007) High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences BioTechniques 43: 649-656.

30. Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen R, et al. (2008) Kismeth: Analyzer of plant methylation states through bisulfite sequencing. BMC Bioinformatics 9: 371.

31. Park W, Li J, Song R, Messing J, Chen X (2002) CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in Arabidopsis thaliana. Current Biology 12: 1484-1495.

32. Li R, Yu C, Li Y, Lam T-W, Yiu S-M, et al. (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25: 1966-1967.

CHAPTER 2

A general transcription factor, TATA-binding protein-associated factor6, mediates RNA-directed DNA methylation and transcriptional gene silencing

ABSTRACT

Cytosine methylation is a critical mechanism for transcriptional gene silencing (TGS) of transposons and the precise regulation of gene expression throughout the plant life cycle. The initial determination of methylated targets in *Arabidopsis thaliana* is accomplished through RNA-directed DNA methylation (RdDM) guided by small and long non-coding RNAs. These RdDM-specific non-coding RNAs are generated by the plant-specific RNA polymerases Pol IV and Pol V. In a luciferase-based genetic screen employing the RdDM and TGS reporter *LUCH*, a T-DNA insertion disrupting the gene encoding the general transcription factor TATA-binding protein-associated factor6 (TAF6) was found to suppress TGS of *LUCH*. In addition to the RdDM reporter, several endogenous RdDM-targeted loci were found to require *TAF6* for proper DNA methylation and TGS. Genome-wide methylation analysis revealed that several hundred regions were hypomethylated at CHH sequence contexts in the *taf6* mutant and that the methylation at these regions was also largely dependent on Pol IV and/or Pol V. The decreases in CHH methylation could not be attributed to compromised Pol II or Pol IV function, as the expression of genes involved in gene silencing and the biogenesis of heterochromatic siRNAs were not dramatically altered in the *taf6* mutant. However, Pol

V-dependent transcripts were slightly decreased at several loci in the *taf6* mutant, which was also observed in plants in which *TAF6* expression was suppressed by *TAF6*-specific artificial miRNA. In addition, TAF6 was found to partially co-localize with the largest subunit of Pol V in nuclear foci. These findings show that TAF6 functions with Pol V at a subset of RdDM targets to establish cytosine methylation and induce TGS.

INTRODUCTION

Cytosine DNA methylation is a major mechanism underlying the transcriptional gene silencing (TGS) of repeats and transposons. In *Arabidopsis thaliana*, cytosine DNA methylation is established by RNA-directed DNA methylation (RdDM), in which 24-nucleotide (nt) small interfering RNAs (siRNAs) guide DNA methyltransferase activity in a sequence-specific manner [1]. RdDM is mediated by diverse modules including RNA polymerases, RNA interference (RNAi) factors, methyltransferases and chromatin remodelers. Of particular importance are the plant-specific DNA-dependent RNA polymerases, Pol IV and Pol V, which determine the targets of RdDM [2]. Pol IV is thought to generate long single-stranded RNAs (ssRNAs) from the RdDM targets [3,4]. These ssRNAs are subsequently processed into siRNAs by the RNAi machinery: RNA DEPENDENT RNA POLYMERASE2 (RDR2) generates double-stranded RNAs (dsRNAs) from the ssRNAs, and DICER-LIKE3 (DCL3) cleaves the dsRNAs into 24-nt siRNA duplexes [5,6]. The siRNA duplexes are stabilized by HUA ENHANCER1 (HEN1)-mediated 2'-*O*-methylation at the 3'ends, and one strand from the duplex is loaded into an ARGONAUTE4 (AGO4)-clade protein [7,8]. Independent of Pol IV activity, Pol V generates transcripts that traverse the methylated loci [9]. These Pol V-dependent transcripts hybridize with AGO4-bound siRNAs and function as scaffolds that recruit downstream players including DOMAIN REARRANGED METHYL TRANSFERASE2 (DRM2) for DNA methylation [10].

Pol IV and Pol V are evolutionarily and functionally related to Pol II, the enzyme responsible for mRNA transcription. Pol II, Pol IV and Pol V each contain 12 subunits,

and the largest subunits of the three polymerases are encoded by distinct genes [4]. Phylogenetic analysis of the largest subunits revealed that Pol IV and Pol V evolved from the duplication of Pol II [4]. Regarding the other subunits of the RNA polymerases, some subunits are specific to a single RNA polymerase, while others are common to two or all three polymerases. At some RdDM loci, Pol II, rather than Pol V, is required for the biogenesis of long non-coding RNAs to guide cytosine methylation, and a weak mutation in the gene encoding the second largest subunit of Pol II, NRPB2, compromises the recruitment of Pol IV and Pol V to some RdDM targets [11]. Additionally, a mutation in a subunit of Mediator, which integrates regulatory signals and mediates transcription by Pol II, was found to affect Pol II and Pol V occupancy as well as the biogenesis of scaffold transcripts at some RdDM targets [12].

Pol IV and Pol V are regulated by their assistant proteins. Pol IV requires the chromatin remodeler and helicase CLASSY1 and the homeodomain protein SAWADEE HOMEODOMAIN HOMOLOG1 (SHH1) to generate siRNAs [13,14]. SHH1 recognizes specific histone modifications and facilitates the recruitment of Pol IV to a subset of RdDM targets [15]. Pol V requires the DDR complex, which contains the chromatin remodeler DEFECTIVE IN RNA-DIRECTED DNA METHYLATION1 (DRD1), the hinge domain protein DEFECTIVE IN MERISTEM SILENCING3 (DMS3) and the single-stranded DNA-binding domain protein REQUIRED FOR DNA METHYLATION1 (RDM1), for its recruitment to specific targets and the biogenesis of scaffold transcripts [16]. Although ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequencing) for the largest subunit of Pol V uncovered a consensus

DNA sequence that was enriched at half of the Pol V-occupied loci, the motif was not an authentic cis-acting element for the recruitment of Pol V to its targets [17]. In this regard, the molecular mechanisms underlying Pol IV and Pol V activity require further investigation, such as the cis-acting elements and trans-acting factors that direct Pol IV and Pol V to the targets.

Transcription initiation by Pol II is elaborately regulated by cis-acting DNA elements, DNA-binding proteins and other accessory factors [18]. The core promoter of a given gene, i.e., the minimal DNA element required for transcription initiation that contains the transcription start site and other DNA elements such as the TATA box (reviewed in [19,20]), is recognized by general transcription factors, which recruit Pol II to trigger transcription. For instance, the recognition of the TATA box by TATA-binding protein (TBP) in transcription factor IID (TFIID) is a representative example of the initiation of Pol II-mediated transcription. In addition to TBP, TFIID contains 13 other TBP-associated factors (TAFs), some of which detect other core promoter elements for transcription initiation: the downstream promoter element (DPE) is recognized by the TAF6/TAF9 dimer, the initiator by the TAF1/TAF2 dimer and the downstream core element by TAF1 in animals. The interaction between the TATA box in the core promoter and TBP recruits the Pol II complex, with Pol II entry at the promoter representing the rate-limiting step for transcription initiation. In humans, only one-third of all genes contain a TATA box in the promoter region, and TATA-less promoters depend on other core promoter elements for the recognition of the transcription start site. For example, recognition of the DPE by the TAF6/TAF9 dimer facilitates Pol II

recruitment to the promoter. The rate of such TAF-dependent transcription is determined by TAF recognition of the promoter region.

The present findings show that *TAF6* is required for cytosine methylation and TGS. TAF6 was found to promote TGS of several endogenous RdDM targets and TGS of the luciferase-based RdDM reporter *LUCH*. Genome-wide methylation analysis revealed hundreds of TAF6-dependendent regions for CHH methylation, and these TAF6-dependent targets were found to require the action of Pol IV and Pol V for CHH methylation. Although *TAF6* was not required for the biogenesis of heterochromatic siRNAs, it was found to mediate the transcription of Pol V-dependent transcripts at some loci. TAF6 therefore promotes CHH methylation at a subset of RdDM targets by promoting the action of Pol V. This finding expands the current view on the functions of general transcription factors in epigenetic regulation and TGS.

RESULTS

Isolation of a weak mutant of *TAF6* with compromised TGS

To identify genes that facilitate RdDM and TGS in *Arabidopsis thaliana*, a forward genetic screen was performed using a line named *LUCH*, which contains a luciferase (*LUC*) transgene whose expression is repressed by CHH methylation [21]. In *LUCH*, the *LUC* transgene is driven by a dual cauliflower mosaic virus (CaMV) 35S promoter (*d35S*) and is in the *RNA-dependent RNA polymerase6-11* (*rdr6-11*) background to prevent post-transcriptional sense transgene silencing. The expression of *LUC* is influenced both by RdDM and by REPRESSOR OF SILENCING 1 (ROS1)-

mediated demethylation. Mutations disrupting RdDM decrease the level of methylation at *d35S* in the CHH sequence context and increase *LUC* expression and activity. Conversely, compromised ROS1 activity increases the level of cytosine methylation at *d35S* and decreases *LUC* expression. Employing LUC as a visible marker for the activity of RdDM and demethylation, a transfer DNA (T-DNA) mutagenesis screen was conducted, and a mutant with de-repressed LUC luminescence was isolated (Figure 2.1A). Real-time RT-PCR confirmed that *LUC* transcript levels were slightly increased in the mutant (Figure 2.1B). The mutant exhibited morphological defects including reduced plant size, delayed flowering and reduced fertility (Figure 2.1E, Figure 2.2). Some seedlings also exhibited more severe defects such as abnormal leaf shape (Figure 2.1E). According to genetic analysis, a single recessive mutation was responsible for both the de-repressed LUC activity and the observed morphological phenotype. Map-based cloning revealed that multiple copies of the T-DNA were inserted in the 5' untranslated region (5' UTR) of the *TAF6* gene (At1g04950, Figure 2.1C). RT-PCR confirmed that *TAF6* transcript levels were slightly reduced in the mutant (Figure 2.1D). The introduction of a genomic fragment containing the full-length coding region of *TAF6* and 1.6 kb of the upstream promoter region rescued the de-repressed LUC activity according to luciferase live imaging and RT-PCR of *LUC* (Figure 2.1A, B). The *TAF6* transgene also rescued the morphological defects of the mutant (data not shown). Therefore, the mutant is in the TAF6 gene and will be hereafter referred to as *taf6-1*. It was previously reported that no *taf6* homozygotes could be recovered in the progeny of heterozygous *taf6* mutants [22]. This is in part due to compromised growth of the *taf6* pollen tube, which reduces the

transmission of the mutant allele [22]. Because *taf6-1* still expressed *TAF6* RNA, albeit at reduced levels, and was able to complete the life cycle, this mutant represents a weak loss-of-function allele of *TAF6.*

The next question addressed was whether the de-repressed expression of *LUC* in *taf6-1* resulted from reduced methylation at *d35S*, as observed for mutants with disrupted RdDM. The cytosine methylation at *d35S* was examined by Southern blot analysis, McrBC-PCR and bisulfite sequencing, as previously described [21]. In Southern blot analysis, digestion of genomic DNA by methylation-sensitive *Alu*I and hybridization with radiolabeled *d35S* probe revealed that the *d35S*-specific DNA pattern was similar for *LUCH* and *LUCH taf6-1* in contrast to the downward shift in *LUCH ago4-6*, indicating that cytosine methylation at *Alu*I sites in *d35S* was not altered in *taf6-1* (Figure 2.3A). This finding was further confirmed using the McrBC-PCR method. Because McrBC specifically digests methylated DNA, differential amplification from McrBC-treated DNA can be used to measure methylation density at a given locus. Whereas unmethylated *ACTIN1* was similarly amplified using McrBC-treated and untreated DNA samples, reduced amplification was observed for methylated *d35S* in McrBC-treated DNA relative to untreated DNA. Similar amplification levels for *d35S* were observed for *LUCH* and *LUCH taf6-1* after McrBC-treatment (Figure 2.3B), indicating that cytosine methylation at *d35S* was not grossly affected in *taf6-1*. In order to examine *d35S* methylation at single nucleotide resolution, we performed bisulfite sequencing. Whereas increased CG methylation was observed, mild hypomethylation at *d35S* was observed for the CHG and CHH contexts in *LUCH taf6-1*. The overall methylation density was

unaffected in *LUCH taf6-1* (Figure 2.3C), which could explain why McrBC-PCR did not detect changes in DNA methylation. As *LUCH* is highly sensitive to CHH methylation, but not to CG methylation [21], the slight reduction in CHH methylation in *taf6-1* could be the reason for the mild *LUC* de-repression. However, the reduction in CHH methylation was so small that one could not be certain that it contributed to *LUC* de-repression.

TAF6 affects the expression and cytosine methylation of endogenous RdDM targets

In light of the compromised TGS of the *LUCH* RdDM reporter in *taf6-1*, the next question addressed was whether *TAF6* is required for TGS and DNA methylation at endogenous RdDM loci. Because disrupted TGS results in increased transcript accumulation, the state of TGS at RdDM loci can be assessed using RT-PCR. The *nrpe1-1* mutation disrupts the largest subunit of Pol V; as expected, transcript accumulation from seven known siRNA-generating loci was observed in *LUCH nrpe1-1* using either regular or real-time RT-PCR. *IGN5*, *AtLINE1-4*, *soloLTR*, *AtSN1* and *AtGP1* were also released from TGS in *LUCH taf6-1*, as indicated by increased transcript accumulation (Figures 2.4B to 2.4D). These findings indicate that *TAF6* is required for TGS at several endogenous RdDM loci. However, the impact of the *taf6-1* mutation on the TGS of *LUCH* and endogenous RdDM targets was less dramatic than that of *nrpe1-1*, except at *IGN5*, where a greater release of TGS was observed in *LUCH taf6-1*. In contrast, the accumulation of transcripts from *AtMU1* and *AtCopia4* was reduced in *LUCH taf6-1* (Figure 2.4A). One possible explanation for this finding is that TAF6 promotes Pol II-

mediated transcription through promoter recognition and recruitment of Pol II even at transcriptionally silenced regions.

We next performed McrBC-qPCR to determine whether TAF6 impacts DNA methylation at these RdDM targets. No changes in DNA methylation were observed for *IGN5*, *AtLINE1-4* or *soloLTR*. A reduction in cytosine methylation was observed for *AtSN1*, *AtGP1* and *AtMU1* in *taf6-1* (Figure 2.4E). In summary, although only several loci were analyzed, *TAF6* was found to affect the TGS and/or cytosine methylation of endogenous siRNA-generating regions.

Genome-wide cytosine methylation profile of *LUCH taf6-1*

In addition to the cytosine methylation analysis at several known RdDM targets, genome-wide analysis was also performed using whole-genome bisulfite sequencing (BS-seq) to further assess the role of *TAF6* in cytosine methylation. Libraries were constructed using 10-day-old Col-0, *sde4-3* (a Pol IV mutant), *nrpe1-11* (a Pol V mutant), *LUCH* and *LUCH taf6-1* seedlings, with two biological replicates each. *sde4-3* and *nrpe1-11* were compared to Col-0, and *LUCH taf6-1* was compared to *LUCH*. The bisulfite conversion efficiency (Table 1) and read coverage (Table 2) of each library were sufficiently high for subsequent analysis. Moreover, the two biological replicates yielded r^2 values greater than or equal to 99.6% for all of the genotypes analyzed, indicating that the experiments were reproducible (Figure 2.5A).

To investigate the function of *TAF6* in DNA methylation, differentially methylated regions (DMRs) were identified and subjected to further analysis. Because the

three cytosine contexts are methylated through different mechanisms, DMRs were identified for all three sequence contexts (i.e., CG, CHG and CHH DMRs). Furthermore, the identified DMRs were classified into two subgroups based on the direction of change in the methylation status: increased and reduced DMRs corresponding to an increase or decrease, respectively, in methylation relative to wild-type plants. After the DMRs were identified, TAF6-specific DMRs were selected by removing the hypervariability (HV) regions, which exhibit a tendency for spontaneous changes in methylation status [23,24], as well as *LUCH*-specific DMRs, which were identified by comparing the Col-0 and *LUCH* libraries (Figure 2.5B). Several hundreds of DMRs were subsequently identified in *LUCH taf6-1*: 268 CG, 3 CHG and 267 CHH reduced DMRs and 218 CG, 12 CHG and 9 CHH increased DMRs (Figure 2.6A). Among these six DMR classes, those with several hundred DMRs (increased CG, reduced CG and reduced CHH DMRs) were further analyzed to investigate the role of *TAF6* in cytosine methylation. The DMRs in these three classes were first partitioned according to their genomic location: transposons, intergenic regions or gene bodies. For the CG DMRs in *LUCH taf6-1*, 168 of the 218 reduced DMRs (77%) and 186 of the 268 increased DMRs (69%) were located in gene bodies (Figure 2.5C). In contrast, CHH reduced DMRs were significantly associated with transposons and repeats. Of the 267 CHH reduced DMRs in *LUCH taf6-1*, only 13 (5%) were located in gene bodies, while 216 (81%) and 38 (14%) were associated with repeat/transposon and intergenic regions, respectively (Figure 2.5C, Figure 2.6B). For the following reasons, CHH reduced DMRs were prioritized for subsequent analysis. First, whereas CG DMRs were almost evenly divided into reduced and increased DMRs, the

CHH DMRs identified in *LUCH taf6-1* were predominantly reduced DMRs. Second, CHH methylation is primarily established and maintained through the RdDM pathway, in contrast to the multiple regulatory mechanisms for CG methylation. Accordingly, the emphasis on CHH reduced DMRs facilitated downstream analysis and experiments for analyzing the role of *TAF6* in cytosine methylation. Lastly, *taf6-1* was initially isolated from a genetic screen aimed at identifying factors involved in RdDM, and TGS of known RdDM targets was compromised in the mutant.

Because CHH methylation is controlled by the RdDM pathway, the next question addressed was whether CHH reduced DMRs in *LUCH taf6-1* required the canonical RdDM pathway for cytosine methylation. Of the known players in RdDM, Pol IV and Pol V were used to address this question. The dependency of CHH reduced DMRs in *LUCH taf6-1* on RdDM was investigated by determining whether CHH reduced DMRs in *LUCH taf6-1* were represented among the CHH reduced DMRs identified in *sde4-3* and *nrpe1-11*. The impact of disrupted *TAF6* function on CHH methylation was less dramatic than that of compromised RdDM, as revealed by the large differences in the number of DMRs (Figure 2.6C). Nevertheless, *TAF6*-targeted DMRs significantly overlapped with Pol IV and Pol V-targeted DMRs (Figure 2.6C). Whereas 13% of DMRs were specific to TAF6, 87% of the DMRs identified in *LUCH taf6-1* were dependent on Pol IV, Pol V or both for cytosine methylation (Figure 2.6D). This indicates that RdDM is responsible for CHH cytosine methylation at the majority of TAF6-dependent loci.

Several TAF6 CHH DMRs were further analyzed to validate the changes in cytosine methylation in *taf6-1* and to examine the status of TGS at these loci. DDT2

(decreased CHH DMRs in *taf6-1*) is located in the intergenic region between At3g14205 and At3g14210, and DDT8 and DDT10 are located in transposon regions (Figure 2.6E). All three DDTs are dependent on Pol IV and Pol V for DNA methylation according to BS-seq. McrBC-qPCR revealed decreased amplification of these regions in *LUCH* compared to *LUCH nrpe1-1* and *LUCH taf6-1* (Figure 2.7A). This suggested that TAF6 and Pol V were required to methylate these loci. In addition, the level of transcripts from these loci was slightly increased in the mutants, except for DDT8 in *LUCH nrpe1-1* (Figure 2.7B, 2.7C). Thus, TAF6 is responsible for cytosine methylation at these regions and the transcriptional silencing of transposons.

TAF6 is not required for the accumulation of 24-nt siRNAs

To investigate how TAF6 mediates cytosine methylation at CHH sequence contexts, several experiments were conducted with respect to the function of the RNA polymerases Pol II, Pol IV and Pol V. To address the possibility that TAF6 promotes the transcription of genes required for RdDM as a general Pol II transcription factor, mRNA-seq (high-throughput sequencing of polyA-tailed RNA) was performed using 10-day-old *LUCH* and *LUCH taf6-1* seedlings to obtain a genome-wide profile of gene expression. In *LUCH taf6-1*, the transcript levels of 381 and 700 genes were up-regulated and down-regulated, respectively, by two-fold. Genes previously implicated in gene silencing [25] and other related genes were examined to determine whether they were differentially expressed in *LUCH taf6-1* (Table 2.5). *AGO3* transcript levels increased by 3.8-fold in *LUCH taf6-1* compared to *LUCH*. Among the 10 AGO proteins in *Arabidopsis*, AGO2

and AGO3 belong to the same clade and occur in tandem in the genome (reviewed in [26]). According to the mRNA-seq results, the *AGO2* transcript levels in *LUCH* were more than 10 times greater than the transcript levels of *AGO3*. In *LUCH taf6-1*, the levels of *AGO3* transcripts were one-quarter of those of *AGO2*. Thus, the increase in *AGO3* transcript levels did not greatly increase the total pool of *AGO2* and *AGO3* transcripts. Furthermore, AGO2 specifically binds small RNAs 21 nt in length with a 5' terminal adenosine, including some miRNA and tasiRNA species [27]. Although AGO3-bound small RNAs have not been characterized, it is unlikely that AGO3 is responsible for the altered methylation landscape in *LUCH taf6-1*.

ROS1 and *DML3* were among the genes down-regulated in *LUCH taf6-1*. Because both genes are responsible for cytosine demethylation, their reduced expression would correspond to an increase in cytosine methylation. In *LUCH taf6-1*, increased cytosine methylation occurred predominantly at the CG sequence context, and only 6.3% of the increased CG DMRs were targeted by ROS1-mediated demethylation. This indicates that the reduced transcript levels of *ROS1* and *DML3* were not responsible for the increased CG DMRs in *LUCH taf6-1*. Because most of the CHH DMRs in *LUCH taf6-1* exhibited hypomethylation, reduced expression of these demethylation genes could not explain a decrease in CHH methylation. Interestingly, down-regulation of *ROS1* expression appears to be a feedback mechanism to compensate for hypomethylation in RdDM mutants [28]. Therefore, the reduction in *ROS1* and *DML3* expression in *taf6-1* could be an indirect effect of the reduced CHH methylation.

Another possible mechanism is that TAF6 functions in Pol IV- and Pol V-mediated RdDM. As previously described, Pol IV and Pol V evolved from Pol II following gene duplication and neofunctionalization. That the three RNA polymerases have some common subunits raised the possibility that the Pol II regulatory framework may have been adopted to control certain functions of Pol IV and Pol V. The functions of Pol IV and Pol V in *LUCH taf6-1* were therefore addressed. To investigate Pol IV function, small RNA-seq was performed to examine the biogenesis of heterochromatic siRNAs in *LUCH taf6-1*. In *sde4-3* compared to wild-type, the level of 24-nt small RNAs dramatically decreased, while the accumulation of 21-nt small RNAs corresponding to miRNAs increased (Figure 2.8A). In contrast, *taf6-1* did not alter the composition of 21-nt and 24-nt small RNAs. Moreover, small RNA abundance globally decreased in *sde4-3* and *nrpe1-11* relative to Col-0, while *taf6-1* did not affect the abundance of small RNAs, compared to *LUCH taf6-1* (Figure 2.8B). Specifically, there were 8483 differential small RNA regions (DSRs) in *sde4-3* and 3524 DSRs in *nrpe1-1* with reduced accumulation of small RNAs. In contrast, *LUCH taf6-1* reduced and increased small RNA levels only in 20 and 17 regions, respectively, compared to *LUCH* (Table 6). Taken together, these findings indicate that TAF6 does not work through Pol IV to promote cytosine methylation.

TAF6 facilitates the production of Pol V-dependent transcripts

To address whether TAF6 worked together with Pol V to generate scaffold transcripts, the biogenesis of Pol V-dependent transcripts was tested by RT-PCR. In

LUCH nrpe1-1, the decrease in transcript levels was observed as expected (Figure 2.8C), except for soloLTR and siR02, loci where scaffold transcripts are generated by Pol II [11]. In *LUCH taf6-1*, the transcripts from *soloLTR*, *IGN5*, *IGN6*, *IGN23* and *IGN26* were not affected, while those from *IGN15*, *IGN24*, *IGN25* and *siR02* were reduced (Figure 2.8C). Because the effect of *taf6-1* on the biogenesis of scaffold transcripts was mild compared to that of *nrpe1-1*, the experiment was repeated with lines in which *TAF6* was suppressed by artificial miRNA targeted to *TAF6* (*amiR:TAF6*) to confirm that *TAF6* promotes the production of these transcripts. When the biogenesis of *amiR:TAF6* was induced by dexamethasone (DEX) treatment, the plants exhibited developmental defects such as smaller plant size, retarded growth and pale green color (Figure 2.9A). Real-time RT-PCR showed that the level of *TAF6* transcripts was reduced to approximately 50% of wild-type levels (Figure 2.9C). The accumulation of Pol V-dependent transcripts at *IGN25* and *IGN26* and Pol II-dependent transcripts *soloLTR* and *siR02* was suppressed in the *amiR:TAF6* lines (Figure 2.8D). These findings indicate that *TAF6* is required for Pol V-mediated transcription.

Subnuclear co-localization of NRPE1 and TAF6

To address whether TAF6 is present at the same nuclear location with NRPE1 and AGO4, immunolocalization was performed. A *LUCH taf6-1* line rescued by *pTAF6:TAF6-GFP* was used to examine the localization of TAF6. TAF6 appeared to be distributed throughout the nucleoplasm and form multiple, small foci (Figure 2.10). While no co-localization between AGO4 and TAF6 was observed, NRPE1 and TAF6

exhibited partial co-localization at some nuclear foci (Figure 2.10). The co-localization between TAF6 and Pol V further supports a functional association between the two.

DISCUSSION

Although many factors required for RdDM are known, how RdDM is targeted to transposable elements in the genome is still unclear. The present study begins to address this question by showing that a general transcription factor TAF6 promotes RdDM, the process by which methylated targets are initially determined. Several lines of evidence support the requirement of *TAF6* for RdDM, A mutation in *TAF6* suppressed TGS of the *LUCH* RdDM reporter as well as TGS of previously reported endogenous RdDM loci. Furthermore, several endogenous RdDM targets exhibited mild hypomethylation in *LUCH taf6-1*. BS-seq uncovered several hundred regions that required *TAF6* for cytosine methylation in the CHH context, and most of these regions were located in transposons, repeats and intergenic regions and required the action of Pol IV and Pol V. These findings reveal a role of TAF6 in RdDM-mediated TGS at transposons and repeats. mRNA-seq analysis ruled out an indirect effect of TAF6 as a general transcription factor. While Pol IV-mediated biogenesis of small RNAs was not affected, Pol V-dependent transcripts were down-regulated in *LUCH taf6-1*.

Although the molecular mechanism of TAF6 activity remains to be examined, one possibility is that TAF6 promotes the recruitment of Pol V to target regions as a transcription factor. In a previous report, mutations in Mediator subunits were found to compromise the occupancy of Pol V at chromatin [12]. The present findings provide

further evidence that the Pol II regulatory machinery may have been adopted by Pol II-derived Pol V. In turn, TAF6 may promote the association of Pol V to its targets. Another possible mechanism is that TAF6 contributes to Pol II-mediated guidance of Pol V. As previously described, the recruitment of or transcription by Pol II at scaffold transcript-generating loci is necessary for the recruitment of Pol V and AGO4 to chromatin [11]. In this way, TAF6 would indirectly influence the action of Pol V by directing Pol II to the RdDM targets.

As a general transcription factor, TAF6 is known to dimerize with TAF9 and recognize the DPE in core promoter elements to initiate transcription in *Drosophila*. The DPE is conserved from *Drosophila* to humans and compensates for the lack of TATA-box in the promoters of *Drosophila* genes in terms of transcriptional initiation [29]. Although genome-wide analysis of core promoters in *Arabidopsis* has not uncovered a significant enrichment of DPE sequences found in *Drosophila*, the interaction between TAF6 and TAF9 also occurs in *Arabidopsis* [30,31]. Transcriptional regulation by TAF6/TAF9-mediated promoter recognition requires further investigation. Regarding the regulation of transposons and repeats, the DPE consensus sequence is required for the transcription of several retrotransposons and LINE-like elements in *Drosophila* [31,32]. One possibility is that DNA elements in transposons can be recognized by a general transcription factor to initiate transcription for RdDM. Along with the recognition of histone marks for polymerase recruitment, as reported for SHH1 [15], the detection of DNA elements by unknown transcription factors may promote the recruitment of RdDM-specific polymerases to the targets and promote transcription initiation.

MATERIALS AND METHODS

Plant materials

All plants used in the present study were *Arabidopsis thaliana* plants in the Columbia-0 (Col-0) background. The *LUC*-based reporter *LUCH* [21] is in the *rdr6-11* mutant background [33]. The mutants used in this study were *LUCH nrpe1-1*, *LUCH ago4-6* [21], *sde4-3* [34] and *nrpe1-11* (originally reported as *nrpe1-1*, *nrpd1b-1* and SALK_029919) [35]. *nrpe1-1* (originally reported as *drd3-1*) [36] and the *NRPE1-Flag* transgenic line [37] were previously described.

Plant growth conditions and DEX treatment

Seeds were fumigated with chlorine gas for surface sterilization, planted on half-strength Murashige and Skoog (MS) medium containing 8% agar and 1% sucrose, and then stratified at 4°C for three days. Plants were grown in a growth chamber at 23°C with continuous light for 10 days, and all of the experiments were conducted using 10-day-old seedlings.

The stock solution of dexamethasone (DEX, Sigma, Cat# D1756) was prepared by dissolving DEX powder with dimethyl sulfoxide (DMSO, Amresco, Cat# 0231) to a final concentration of 100 mM. To induce the expression of the *TAF6*-targeting amiRNA, *amiR:TAF6* seeds were planted on MS medium containing 10 μM DEX and grown for 10 days. The mock control was performed in parallel, with the same amount of DMSO without DEX added to the MS medium.

T-DNA mutagenesis, genetic screening and luciferase live imaging

For use as an insertional mutagen, pEarleyGate303 plasmid [38] was modified to remove the Gateway cassette. The modified pEarleyGate303 plasmid was introduced into *Agrobacterium tumefaciens* strain GV3101, which was used to transform *LUCH* using the floral dip method [39] to generate the mutagenized populations. Transgenic plants were selected by spraying 0.1% BASTA herbicide, and seeds were harvested from individual surviving plants. Ten-day-old T2 seedlings and luciferase live imaging were used for the genetic screening procedure, as previously described [21]. *LUCH* (Col-0) was backcrossed to Landsberg *erecta* (L*er*) five times to obtain $LUCH^{Ler}$ for map-based cloning.

Mapping of the *taf6-1* mutation

The *LUCH taf6-1* (Col-0) mutant was isolated from the *LUCH* genetic screen and crossed to $LUCH^{Ler}$ to generate the mapping population. In the F2 population, plants with de-repressed LUC activity were used for mapping. Rough mapping using 27 plants revealed linkage to a marker in the F7G19 BAC clone, which is located on the top arm of chromosome 1. New SSLP and CAPS markers were designed based on identified polymorphisms between the Col-0 and L*er* accessions (http://arabidopsis.org/browse/Cereo/index.jsp). Fine mapping with a population of 415 plants narrowed the mapping region to a 146-kb window spanning three BAC clones, T1G11, F13M7 and T7A14. Among the 61 genes in the region, the 5' region encompassing the start codon of *TAF6*

(At1g04950) could not be amplified by PCR. The T-DNA insertion was subsequently confirmed by PCR using primers specific to the T-DNA and *TAF6*.

Plasmid construction

To construct *pTAF6:TAF6-GFP*, the *TAF6* genomic region including 1.6 kb of the upstream promoter region and the entire coding region without the stop codon was amplified using primers TAF6-proF1 and TAF6-fullR1 (Table 8), Phusion High-Fidelity DNA polymerase (New England Biolabs, Cat# M0530) and Col-0 genomic DNA as template. The PCR product was cloned into TSK108, an entry vector from Detlef Weigel's lab, using *Kpn*I and *Sac*I restriction sites at the 5' and 3' regions, respectively. Subsequently, the *TAF6* amplicon was introduced into the pGWB204 [40] destination vector to generate *pTAF6:TAF6-GFP*. *LUCH taf6-1* plants were transformed with the plasmid using the floral dip method [39].

To construct the *amiR:TAF6* plasmid for the expression of amiRNAs complementary to *TAF6*, two amiRNA sequences targeted to *TAF6* were designed using the Web MicroRNA Designer program (http://wmd.weigelworld.org/cgi-bin/mirnatools.pl). To generate the amiRNA precursors, two primary PCRs were conducted using six primers (A, B, I, II, III and IV) and pRS300 [41] as the template. Next, overlapping PCR was performed using two primers (A and B) and the two PCR products as the templates. The amiRNA precursor was directly cloned into pTA7002 using the *Xho*I and *Spe*I sites. After the sequence was confirmed, two plasmids were transformed to Col-0 using the floral dip method [39].

RNA extraction and RT-PCR

Total RNA was extracted from seedlings using TRI reagent (Molecular Research Center, Cat# TR118) then treated with DNase I (Roche, Cat# 04716728001) per the manufacturers' instructions. Reverse transcription was conducted using RevertAid Reverse Transcriptase (Thermo Scientific, Cat# EP0441) per the supplier's guidelines. Pol V-dependent transcripts were reverse-transcribed using locus-specific primers and the SuperScript III kit (Invitrogen, Cat# 18080-051) according to the manufacturer's instructions. cDNA amplification was performed using Crimson *Taq* DNA polymerase (New England Biolabs, Cat# M0324). Real-time RT-PCR was performed on a Bio-Rad C1000 thermal cycler and a CFX96 detection module using iQ SYBR Green Supermix (Bio-Rad, Cat# 170-0082) with three technical replicates. The primers used are listed in Table 2.8.

mRNA-seq library construction, data processing and identification of DEGs (differentially expressed genes)

DNase I-treated total RNA (2 μg) was used for library construction along with the TruSeq RNA Sample Preparation Kit v2 (Illumina, Cat# RS-122-2001) according to the manufacturer's guidelines. Sequencing of the mRNA-seq libraries was performed using an Illumina Hiseq2000 at the Genomics Core Facility at the University of California, Riverside.

The reads that passed the Illumina quality control were subjected to further analysis. All reads were mapped to the TAIR10 *Arabidopsis* genome using TopHat

v2.0.4 with default settings [42], and reads with multiple copies were treated as one read for the mapping procedure. Reads mapping to multiple regions were discarded. To calculate the transcript levels for each gene, reads mapped to each gene were counted and normalized to the RPKM (reads per kilobase per million) value using a Perl script. To identify DEGs in *LUCH taf6-1*, the relative expression levels were calculated by dividing the RPKM of *LUCH taf6-1* by that of *LUCH*. Genes that showed more than 2-fold changes were considered as DEGs in *LUCH taf6-1*. To reduce the false discovery of DEGs resulting from the background noise, genes with the extremely low expression levels were discarded from DEGs by using two cutoff values (the raw reads numbers and the RPKM values). First, genes with less than one RPKM values in both *LUCH* and *LUCH taf6-1* were discarded from the initial DEGs. Second, genes with less than 10 raw reads numbers in both lines were discarded. The remaining DEGs were considered as DEGs in *LUCH taf6-1*.

Construction of small RNA libraries, data processing and identification of DSRs (differential small RNA-generating regions)

Total RNA was extracted from 10-day-old seedlings using TRI reagent (Molecular Research Center, Cat# TR118), and 50 μg of total RNA was electrophoresed in a 15% denaturing polyacrylamide gel. The gel pieces containing RNA fragments 15 to 40 nt in length were incubated in 500 μl of 0.4 N NaCl at 4°C overnight to elute the RNA. The RNA was purified by ethanol precipitation and subjected to library

construction using the TruSeq Small RNA Sample Preparation Kit (Illumina, Cat# RS-200-0012) per the manufacturer's instructions.

The raw reads were processed as previously described [43] and aligned to the *Arabidopsis* genome (TAIR 10 release) using SOAP2 [44]. DSRs were identified by comparing the abundance of small RNAs within 500 bp windows between the wild-type and mutant libraries. Briefly, the *Arabidopsis* chromosomes were divided into continuous 500 bp segments, and small RNAs were assigned to specific windows based on the location of the 5' nucleotide. The number of total reads in each window was normalized as RPM (reads per million), which was then used to compare the wild-type and mutant libraries. For comparison of *sde4-3* and *nrpe-1-11* to Col-0, windows with less than 10 TPM (transcripts per million) in Col-0 library were not considered for DSR identification. For comparison of *LUCH taf6-1* to *LUCH*, windows with less than 10 TPM (transcripts per million) in both libraries were not considered for DSR identification. P-values were calculated [45] and adjusted using the false discovery rate (FDR) approach [46] as previously described. A fold change > 4 and an adjusted p-value (FDR) < 0.05 were required for DSRs between mutants and their respective controls.

Target-specific analysis of cytosine methylation

Southern blot analysis, McrBC-PCR and bisulfite sequencing (BS-seq) were performed as previously described [21] to examine the extent of cytosine methylation at individual loci. Genomic DNA was extracted from 10-day-old seedlings using the CTAB method [47]. For Southern blot analysis, 15 μg of genomic DNA was treated with *Alu*I

(New England Biolabs, Cat# R0137) and hybridized using the standard Southern method [48]. A *d35S*-specific radiolabeled probe was prepared by PCR amplification using the primers 35Sf and 35Sr and *LUCH* genomic DNA, and by the Rediprime II random prime labeling system (GE Healthcare, RPN1633).

For McrBC-PCR, 500 ng of genomic DNA was restricted with 3 units of McrBC (New England Biolabs, Cat# M0272) for 25 minutes at 37°C. Using the McrBC-treated DNA, PCR was performed with *ACTIN1* as an internal loading control. To quantify the relative level of undigested DNA after McrBC treatment, qPCR was conducted using iQ SYBR Green Supermix (Bio-Rad, Cat# 170-0082) with three technical replicates.

For BS-seq, 1 μg of genomic DNA was subjected to bisulfite conversion using the EpiTect Bisulfite Kit (Qiagen, Cat# 59104) per the manufacturer's instructions. Using the converted DNA and primers YZ 35S BisF and YZ LUC BisR, the *d35S* locus was amplified. The PCR products were cloned into the pGEM-T Easy vector (Promega, Cat# A1360), and sequences from 27 unique clones were analyzed to quantify cytosine methylation for each genotype using Kithmeth (http://katahdin.mssm.edu/kismeth/revpage.pl) [49].

Library construction for BS-seq and BS-seq data analysis

Genomic DNA was isolated using the DNeasy Plant Mini Kit (Qiagen, Cat# 69104). One microgram of genomic DNA was sheared into fragments 150 to 300 bp in length using a Diagenode Bioruptor for sonication. The DNA samples were subsequently treated using the PureLink PCR Purification Kit (Invitrogen, Cat# K3100-01) and the

End-It DNA End-Repair Kit (Epicentre, Cat# ER0720). The end-repaired DNA was purified using the Agencourt AMPure XP-PCR Purification system (Beckman Coulter, Cat# A63880) and adenylated at the 3' end using the polymerase activity of Klenow fragment (3'→5' exo-) (New England Biolabs, Cat# M0212). After purification using the Agencourt AMPure XP-PCR Purification system, the DNA fragments were ligated to the methylated adapters included in the TruSeq DNA Sample Preparation Kit (Illumina, Cat# FC-121-2001) using T4 DNA Ligase (New England Biolabs, Cat# M0202). After the ligated DNA was purified using AMPure XP beads, less than 400 ng DNA was subjected to bisulfite conversion using the MethylCode Bisulfite Conversion Kit (Invitrogen, Cat# MECOV-50) according to the manufacturer's instructions then PCR amplified using PfuTurbo Cx Hotstart DNA Polymerase (Agilent, Cat# 600414). The PCR products were purified using AMPure XP beads and constituted the final BS-seq libraries, which were sequenced using HiSeq2000 with the 101 cycle paired-end sequencing mode (Illumina) at the Genomics Core Facility at the University of California, Riverside. Image analysis and base calling were performed using the standard Illumina pipeline, version RTA 1.13.48.

After the raw data were processed using the Illumina quality control pipeline and multi-copy reads were reduced to a single copy, the reads were mapped to the *Arabidopsis* genome (TAIR10) using the BS-Seeker C-to-T genome conversion technique [50]. DMRs were identified as previously described [25]. To reduce the level of experimental noise, two biological replicates of each genotype were performed, and only DMRs identified in both replicates were considered for subsequent analysis. Hypervariability (HV) regions that tend to exhibit altered methylation status were

excluded from the list of DMRs [23,24] along with DMRs identified when comparing Col-0 and *LUCH*.

Immunolocalization

Protein localization was examined as previously described using mesophyll leaf nuclei [37]. DNA was stained using 1μg/ml DAPI (Invitrogen, Cat# D1306).

FIGURES

Figure 2.1 Isolation of the *taf6* mutant.

(A) LUC luminescence in *LUCH*, *LUCH taf6-1* and *LUCH taf6-1 TAF6* (*LUCH taf6-1* transformed with a *TAF6* genomic fragment). Each spot is an individual *Arabidopsis* seedling. Col-0 is the wild-type control without the *LUC* transgene.

(B) Real-time RT-PCR analysis of *LUC* transcript levels in *LUCH*, *LUCH taf6-1* and *LUCH taf6-1 TAF6*. The transcript levels of *LUC* were normalized to those of *UBQUITIN 5* (*UBQ5*), and *LUCH taf6-1* and *LUCH taf6-1 TAF6* were compared to *LUCH*. The standard deviation for *LUCH taf6-1* was calculated from three biological replicates. The standard deviation for *LUCH taf6-1 TAF6* was calculated from three technical replicates.

(C) Schematic diagram of the *TAF6* gene and the location of the T-DNA insertion in *taf6-1*. White and black rectangles represent untranslated regions and protein-coding exons, respectively. Thin lines indicate introns. Multiple T-DNAs were inserted in the 5' UTR of the *TAF6* gene. The two arrows represent the position of primers used for RT-PCR for the detection of *TAF6* transcript levels.

(D) RT-PCR analysis of *TAF6* transcript levels in *LUCH* and *LUCH taf6-1*. *UBQ5* was used as an internal loading control. For the -RT reaction, the reverse transcription reaction was performed without reverse transcriptase and *UBQ5* primers were used for PCR.

(E) Morphology of 10-day-old *LUCH* and *LUCH taf6-1* seedlings. Scale bars = 2 mm. Variations in phenotypes were observed in *LUCH taf6-1*, with the two seedlings on the right representing more severe phenotypes.

Figure 2.1

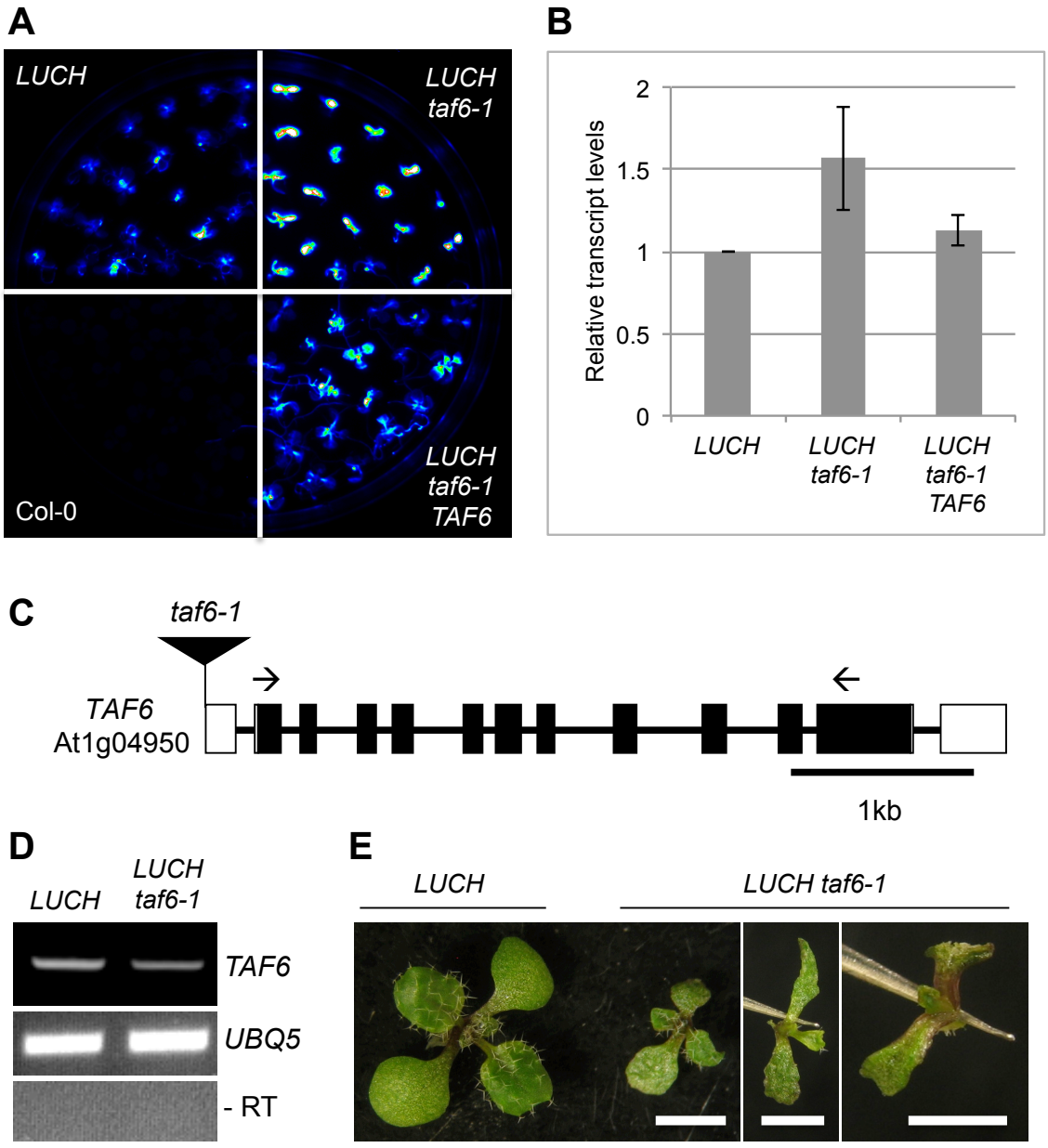

A
LUCH
LUCH taf6-1
Col-0
LUCH taf6-1 TAF6
B
Relative transcript levels
2
1.5
1
0.5
0
LUCH
LUCH taf6-1
LUCH taf6-1 TAF6
C
taf6-1
TAF6
At1g04950
1kb
D
LUCH
LUCH taf6-1
TAF6
UBQ5
- RT
E
LUCH
LUCH taf6-1

Figure 2.2 Comparison of *LUCH* and *LUCH taf6-1* phenotypes.

(A) Three-week-old plants.

(B) *LUCH* and *LUCH taf6-1* siliques. Whereas *LUCH* siliques were long and straight, the silique phenotype of *LUCH taf6-1* was highly variable, with normal, short and curved siliques observed.

(C) *LUCH* and *LUCH taf6-1* plants at the reproductive growth stage.

(D) Magnified image of the *LUCH taf6-1* plant shown in (C).

(E) Dissected siliques showing the fertility of *LUCH* and *LUCH taf6-1*. Aborted seeds were observed in some *LUCH taf6-1* siliques, indicating reduced fertility.

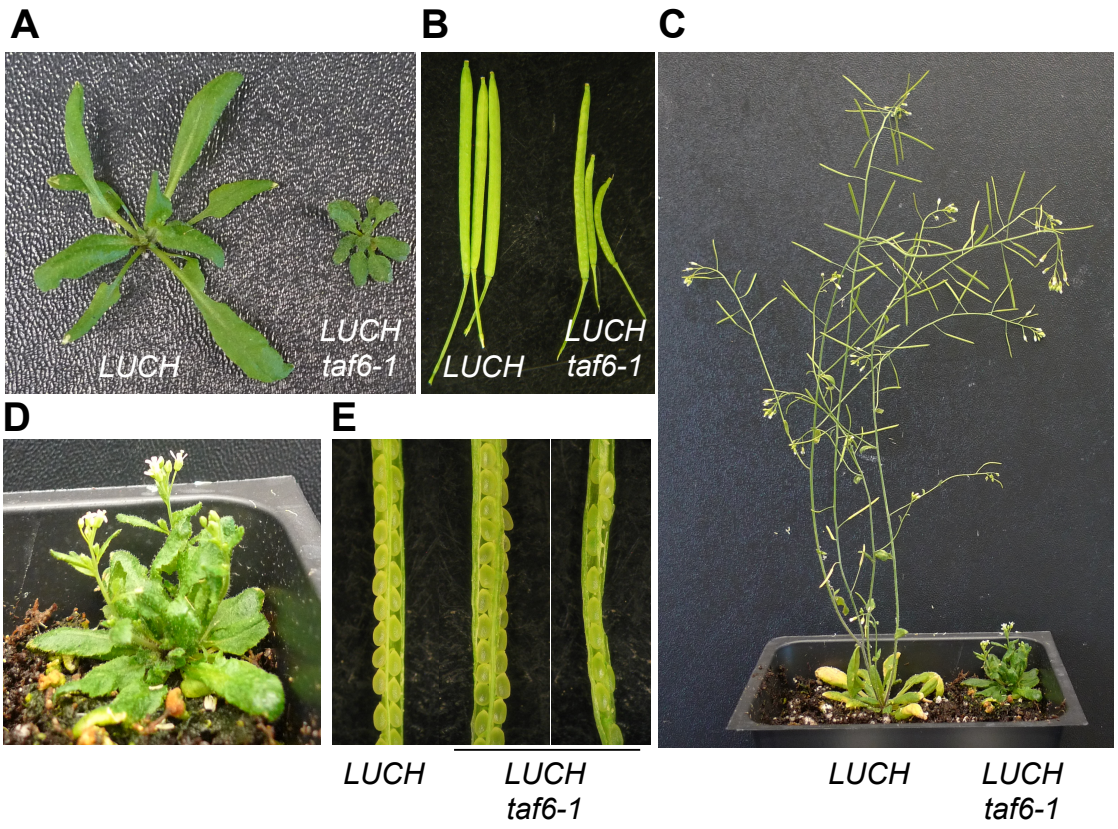


Figure 2.3 The T-DNA insertion disrupting *TAF6* in *taf6-1* had little effects on the level of cytosine methylation at *d35S*.

(A) Southern blot analysis of DNA methylation. *LUCH*, *LUCH ago4-6* and *LUCH taf6-1* genomic DNA was digested by *Alu*I and hybridized with a radiolabeled *d35S* probe. Because *Alu*I activity is inhibited by methylation at its target DNA, loss of cytosine methylation in *LUCH ago4-6* results in a downward shift of the DNA fragments compared to *LUCH*. *LUCH taf6-1* exhibited a similar pattern to *LUCH*, indicating that DNA methylation at *Alu*I sites in *d35S* was not altered in *taf6-1*.

(B) McrBC-PCR analysis of *LUCH* and *LUCH taf6-1*. Genomic DNA was treated with McrBC then PCR amplified using *d35S*-specific primers. McrBC cuts methylated DNA, resulting in reduced amplification from the methylated target compared to DNA not treated with McrBC (Mock). *ACTIN1*, which lacks cytosine methylation, was used as a loading control.

(C) Quantification of cytosine methylation at *d35S* in *LUCH* and *LUCH taf6-1* by bisulfite sequencing. The percentage of methylated cytosine was calculated in the three sequence contexts (CG, CHG and CHH) using 27 unique clones for each genotype.

Figure 2.3

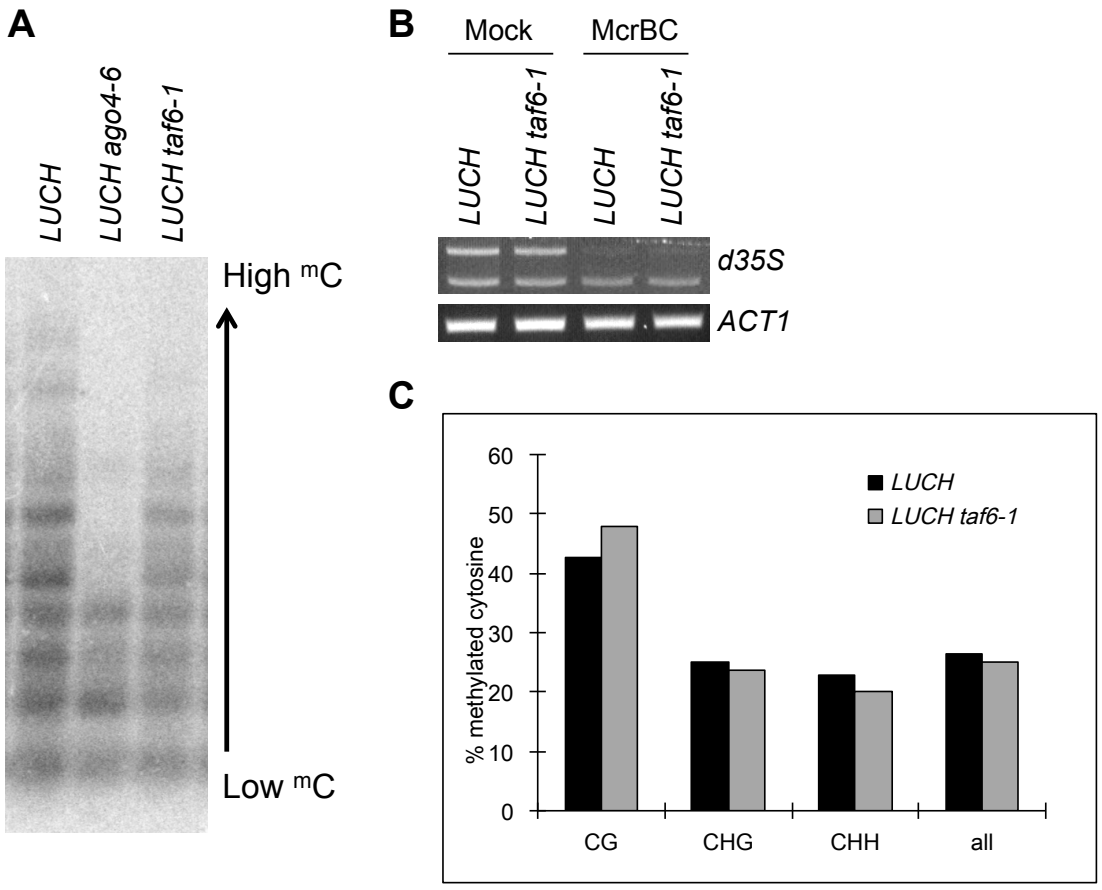

A
LUCH
LUCH ago4-6
LUCH taf6-1
High mC
Low mC
B
Mock
McrBC
LUCH
LUCH taf6-1
LUCH
LUCH taf6-1
d35S
ACT1
C
% methylated cytosine
60
50
40
30
20
10
0
LUCH
LUCH taf6-1
CG
CHG
CHH
all

Figure 2.4 Effect of *taf6-1* on TGS and cytosine methylation at endogenous RdDM targets.

(A) to (C) Real-time RT-PCR to quantify transcript levels from endogenous RdDM targets. The transcript levels were normalized to those of *UBQ5*, and *LUCH nrpe1-1* and *LUCH taf6-1* were compared to *LUCH*. Standard deviations were calculated from three technical replicates. Similar results were observed in three biological replicates.

(A) Real-time RT-PCR of *AtMU1* and *AtCopia4*. Reduced transcript levels were observed in *LUCH taf6-1* compared to *LUCH*.

(B), (C) Real-time RT-PCR of *IGN5*, *AtLINE1-4*, *soloLTR* and *AtSN1*. These four siRNA-generating loci were de-repressed in *LUCH taf6-1* compared to *LUCH*. In (C), the transcript levels in *LUCH nrpe1-1* and *LUCH taf6-1* are depicted in two separate graphs with different scales on the y-axes.

(D) RT-PCR of *AtGP1*. For the -RT reaction, the reverse transcription reaction was performed without reverse transcriptase and *UBQ5* primers were used for PCR. The result was reproduced three times.

(E) McrBC-qPCR analysis to examine cytosine methylation. Genomic DNA was treated with McrBC and subjected to qPCR. Because McrBC specifically digests methylated DNA, hypomethylated DNA results in a higher value of relative DNA levels. *ACTIN1*, which lacks cytosine methylation, was used to normalize the level of amplified product. *nrpe1-1* and *LUCH taf6-1* were compared to Col-0 and *LUCH*, respectively. Standard deviations were calculated from three technical replicates.

Figure 2.4

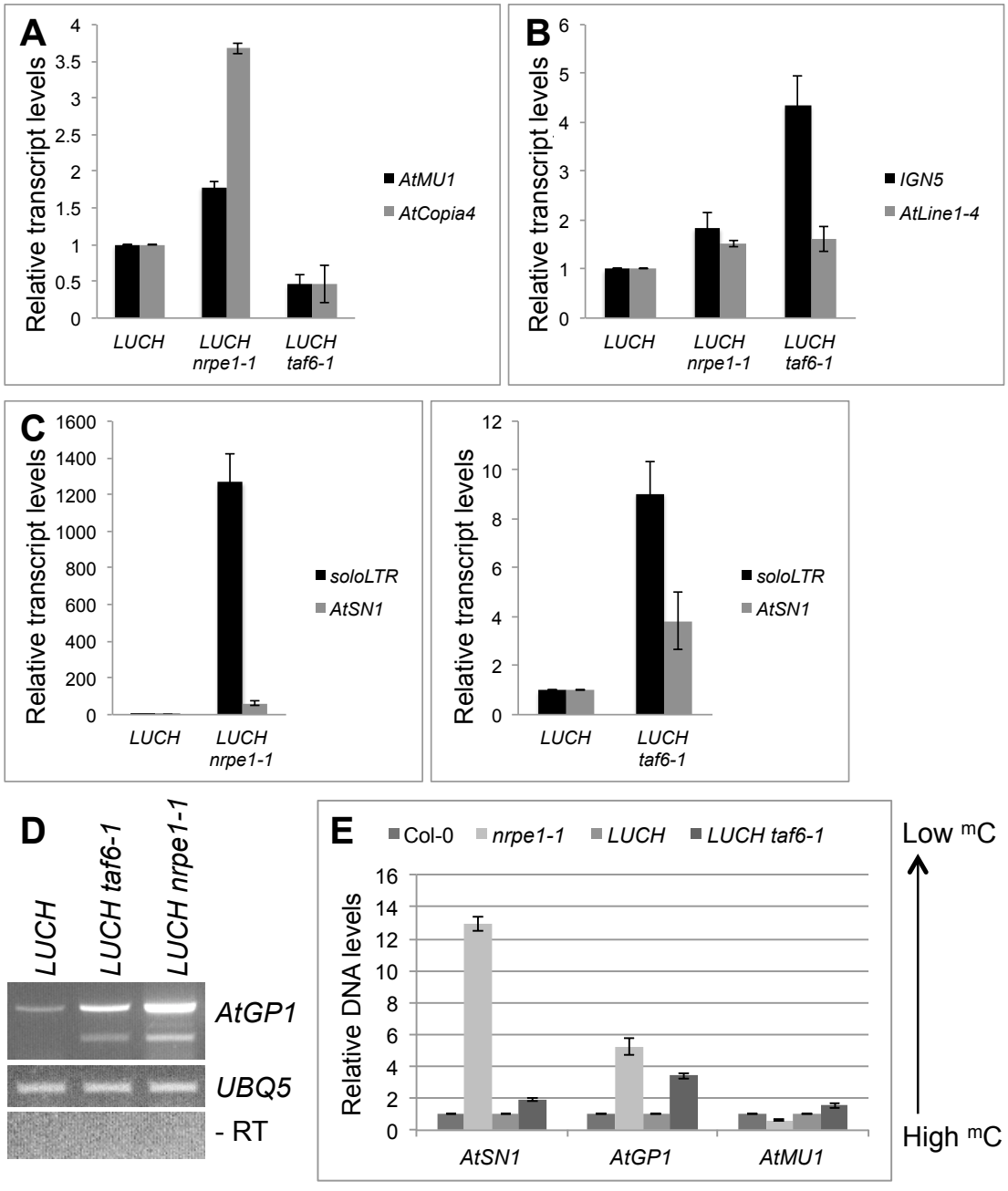

A
Relative transcript levels
4
3.5
3
2.5
2
1.5
1
0.5
0
AtMU1
AtCopia4
LUCH
LUCH nrpe1-1
LUCH taf6-1
B
Relative transcript levels
6
5
4
3
2
1
0
IGN5
AtLine1-4
LUCH
LUCH nrpe1-1
LUCH taf6-1
C
Relative transcript levels
1600
1400
1200
1000
800
600
400
200
0
soloLTR
AtSN1
LUCH
LUCH nrpe1-1
Relative transcript levels
12
10
8
6
4
2
0
soloLTR
AtSN1
LUCH
LUCH taf6-1
D
LUCH
LUCH taf6-1
LUCH nrpe1-1
AtGP1
UBQ5
- RT
E
Col-0
nrpe1-1
LUCH
LUCH taf6-1
Relative DNA levels
16
14
12
10
8
6
4
2
0
AtSN1
AtGP1
AtMU1
Low mC
High mC

Figure 2.5 Summary of whole-genome bisulfite sequencing.

(A) Correlation coefficients between two biological replicates (A and B) for each genotype. The experiments were highly reproducible, as indicated by the r^2 values (greater than 99%).
(B) Number of DMRs after the removal of HVR and *LUCH*-specific DMRs. TAF6-specific DMRs were identified for all three sequence contexts then classified as increased or reduced DMRs, indicating an increase or decrease, respectively, in the methylation density in *LUCH taf6-1* relative to *LUCH*.
(C) Classification of *taf6*-incurred DMRs based on annotation data.

Figure 2.5

A

Genotypes	Correlation coefficient
Col-0 (A, B)	0.9966
sde4-3 (A, B)	0.9978
nrpe1-11 (A, B)	0.9971
LUCH (A, B)	0.9970
LUCH taf6-1 (A, B)	0.9961

B

Initial DMRs			
	CG	CHG	CHH
Increased DMRs	235	13	11
Reduced DMRs	290	3	408
HVR is subtracted			
	CG	CHG	CHH
Increased DMRs	219	12	10
Reduced DMRs	269	3	400
LUCH-specific region is subtracted (FINAL DMRs)			
	CG	CHG	CHH
Increased DMRs	218	12	9
Reduced DMRs	268	3	267

C

		CG		CHH
		Increased	Reduced	Reduced
Repeat	Dispersed	10	13	159
	Inverted	3	2	11
	Tandem	1	1	8
TE		23	27	38
Subtotal (Repeat/TE)		27	43	216
Gene		168	186	13
Intergenic		13	39	38
Total		218	268	267

Figure 2.6 Genome-wide analysis of cytosine methylation in *taf6-1*.

(A) The number of *taf6*-incurred DMRs (differentially methylated regions) in the CG, CHG and CHH contexts. Blue and red bars represent DMRs with increased and reduced cytosine methylation, respectively, in *LUCH taf6-1* compared to *LUCH*.
(B) A pie chart showing the breakdown of CHH DMRs with reduced cytosine methylation in *LUCH taf6-1* by genomic location. A total of 267 DMRs were classified, and the genomic locations were transposon/repeat regions (blue), intergenic regions (red) and gene bodies (green).
(C) A Venn diagram showing the overlap of DMRs in *LUCH taf6-1*, *sde4-3* and *nrpe1-11*. Only CHH DMRs with reduced cytosine methylation in mutants relative to the respective wild-type were analyzed. The DMRs of *LUCH taf6-1*, *sde4-3* and *nrpe1-11* are displayed in blue, green and red, respectively.
(D) A pie chart showing the dependency on Pol IV and Pol V of CHH reduced DMRs in *LUCH taf6-1*. A total of 267 DMRs were classified based on the overlap with CHH reduced DMRs identified in *sde4-3* or *nrpe1-11*. Regions impacted by TAF6, TAF6/Pol IV, TAF6/Pol V, and TAF6/Pol IV/Pol V are shown in the blue, red, green and purple sections, respectively.
(E) Three representative loci with reduced CHH methylation in *LUCH taf6-1*. From top to bottom, Col-0, *nrpe1-11*, *LUCH* and *LUCH taf6-1* are plotted along the chromosome. Each peak represents the methylation density at each cytosine, with the scale of the y-axis adjusted to 50% to better visualize the differences between the mutants and their respective controls.

Figure 2.6

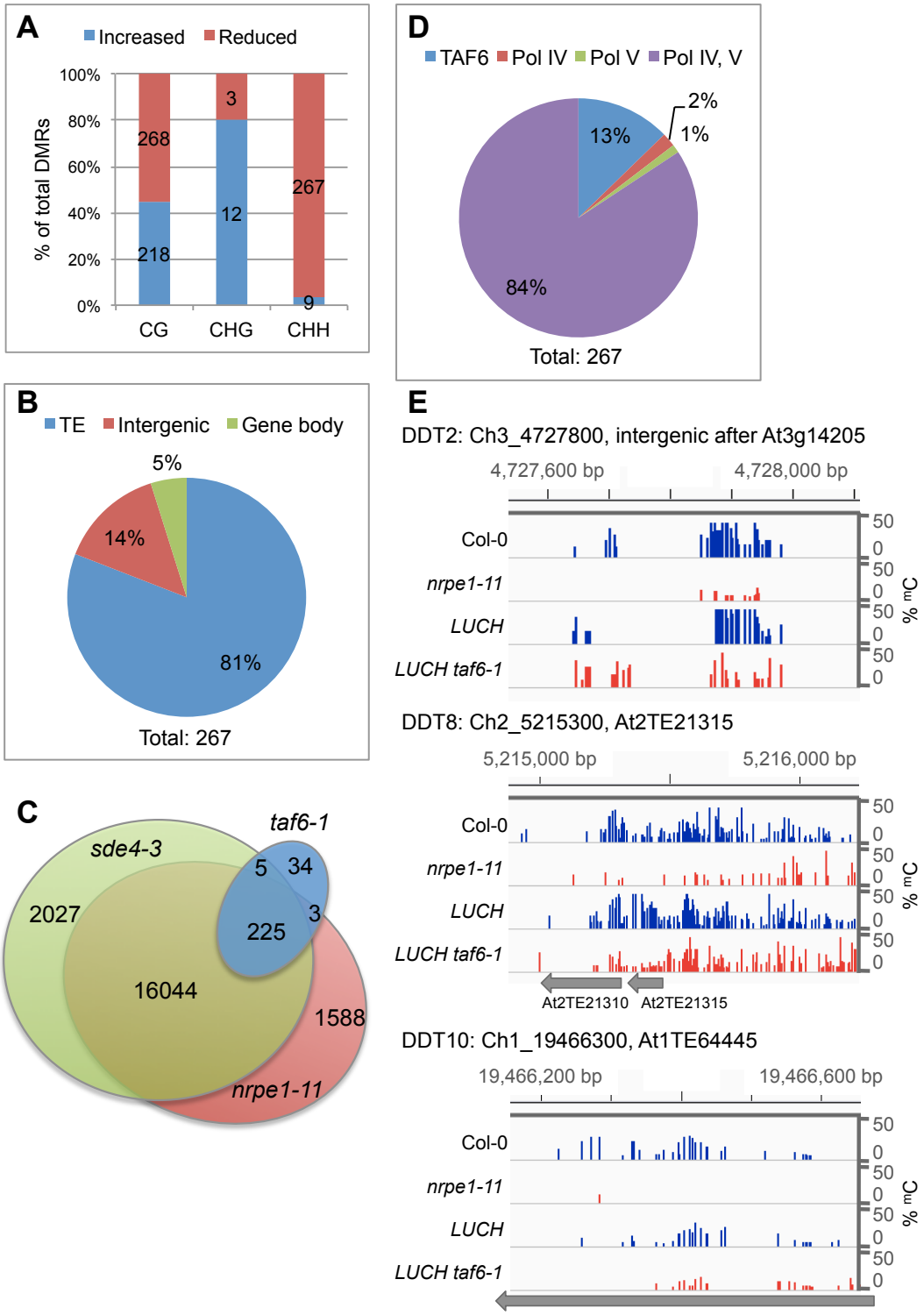

A
Increased
Reduced
% of total DMRs
100%
80%
60%
40%
20%
0%
268
218
3
12
267
9
CG
CHG
CHH
B
TE
Intergenic
Gene body
5%
14%
81%
Total: 267
C
taf6-1
sde4-3
nrpe1-11
5
34
3
225
2027
16044
1588
D
TAF6
Pol IV
Pol V
Pol IV, V
2%
1%
13%
84%
Total: 267
E
DDT2: Ch3_4727800, intergenic after At3g14205
4,727,600 bp
4,728,000 bp
Col-0
nrpe1-11
LUCH
LUCH taf6-1
% mC
DDT8: Ch2_5215300, At2TE21315
5,215,000 bp
5,216,000 bp
At2TE21310
At2TE21315
DDT10: Ch1_19466300, At1TE64445
19,466,200 bp
19,466,600 bp

Figure 2.7 Validation of *taf6*-incurred DMRs and the expression of the loci.

(A) McrBC-qPCR analysis to validate the BS-seq results in *LUCH taf6-1*. DDT2, DDT8 and DDT10 were identified as CHH DMRs with reduced cytosine methylation in *LUCH taf6-1*. Genomic DNA was treated with McrBC and subjected to qPCR. *ACTIN1*, which lacks DNA methylation, was used to normalize the level of amplified products. *nrpe1-1* and *LUCH taf6-1* were compared to Col-0 and *LUCH*, respectively. Standard deviations were calculated from three technical replicates.
(B) RT-PCR analysis for DDT2. *UBQ5* was used as a loading control. For the -RT reaction, the reverse transcription reaction was performed without reverse transcriptase and *UBQ5* primers were used for PCR.
(C) Real-time RT-PCR for DDT8 and DDT10. The transcript levels were normalized to those of *UBQ5*, and *LUCH nrpe1-1* and *LUCH taf6-1* were compared to *LUCH*. Standard deviations were calculated from three technical replicates.

Figure 2.7

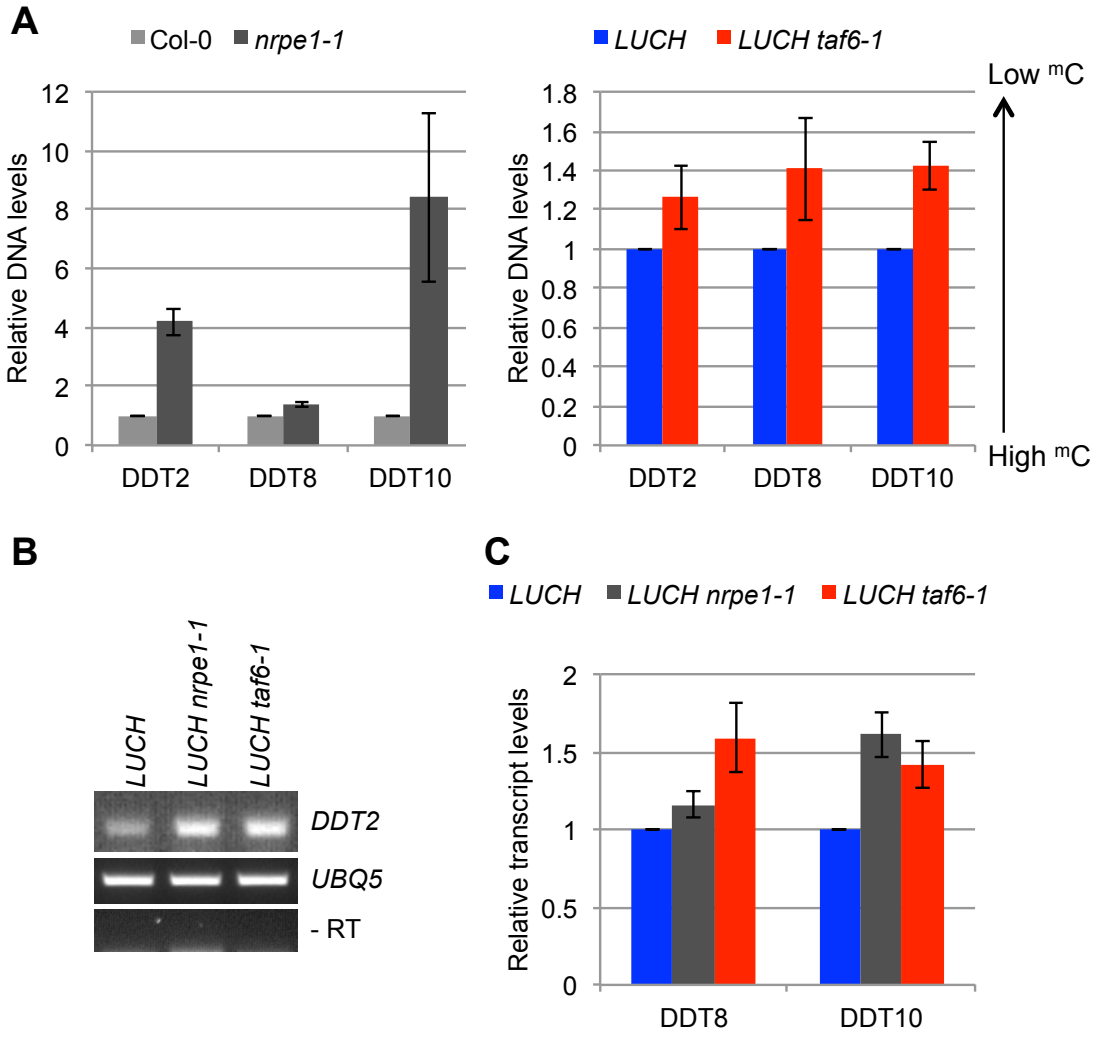

A
Col-0
nrpe1-1
Relative DNA levels
12
10
8
6
4
2
0
DDT2
DDT8
DDT10
LUCH
LUCH taf6-1
Low mC
High mC
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
B
LUCH
LUCH nrpe1-1
LUCH taf6-1
DDT2
UBQ5
- RT
C
LUCH
LUCH nrpe1-1
LUCH taf6-1
Relative transcript levels
2
1.5
1
0.5
0
DDT8
DDT10

Figure 2.8 Pol IV and Pol V function in *LUCH taf6-1*.

(A), (B) Analysis of small RNA accumulation by small RNA-seq. Libraries were constructed using 10-day-old seedlings and subjected to high-throughput sequencing.

(A) The size distribution of total small RNA reads in *LUCH*, *LUCH taf6-1*, Col-0 and *sde4-3*. The number of small RNAs 21 to 24 nt in length was quantified, and the relative abundance for each length is shown for the four genotypes.

(B) Box plots showing small RNA abundance in Col-0, *sde4-3*, *nrpe1-11*, *LUCH* and *LUCH taf6-1*. Small RNAs were mapped to 500 bp static windows in the genome. The read abundances in RPM (reads per million) were calculated for each 500 bp window and are shown on the y-axis. In Col-0, *sde4-3* and *nrpe1-11*, only windows with a read abundance greater than 10 RPM in Col-0 were used to generate the box plots. In *LUCH* and *LUCH taf6-1*, only windows with a read abundance greater than 10 RPM in either libraries were used to generate the box plots. *sde4-3* and *nrpe1-11* were compared to Col-0 whereas *LUCH taf6-1* was compared to *LUCH*. * indicates significant reduction by Mann-Whitney *U* test ($P < 10^{-10}$).

(C) RT-PCR for Pol V- or Pol II-dependent noncoding transcripts in *LUCH*, *LUCH taf6-1* and *LUCH nrpe1-1*. The transcripts at *soloLTR* and *siR02* are Pol II-dependent [11] whereas those at the other loci are Pol V-dependent.

(D) RT-PCR for Pol V- or Pol II-dependent noncoding transcripts in *amiR:TAF6*. Two independent lines (#226 and #233) were tested. Dexamethasone (DEX) treatment was used to induce the expression of *amiR:TAF6*, which ultimately suppressed the expression of *TAF6* (Figure 2.9C). DMSO is the mock control.

(C), (D) Long non-coding RNAs from previously reported loci were examined. For the -RT reaction, the reverse transcription reaction was performed without reverse transcriptase and *UBQ5* primers were used for PCR. The experiment was repeated twice, and similar results were observed for each replicate.

Figure 2.8

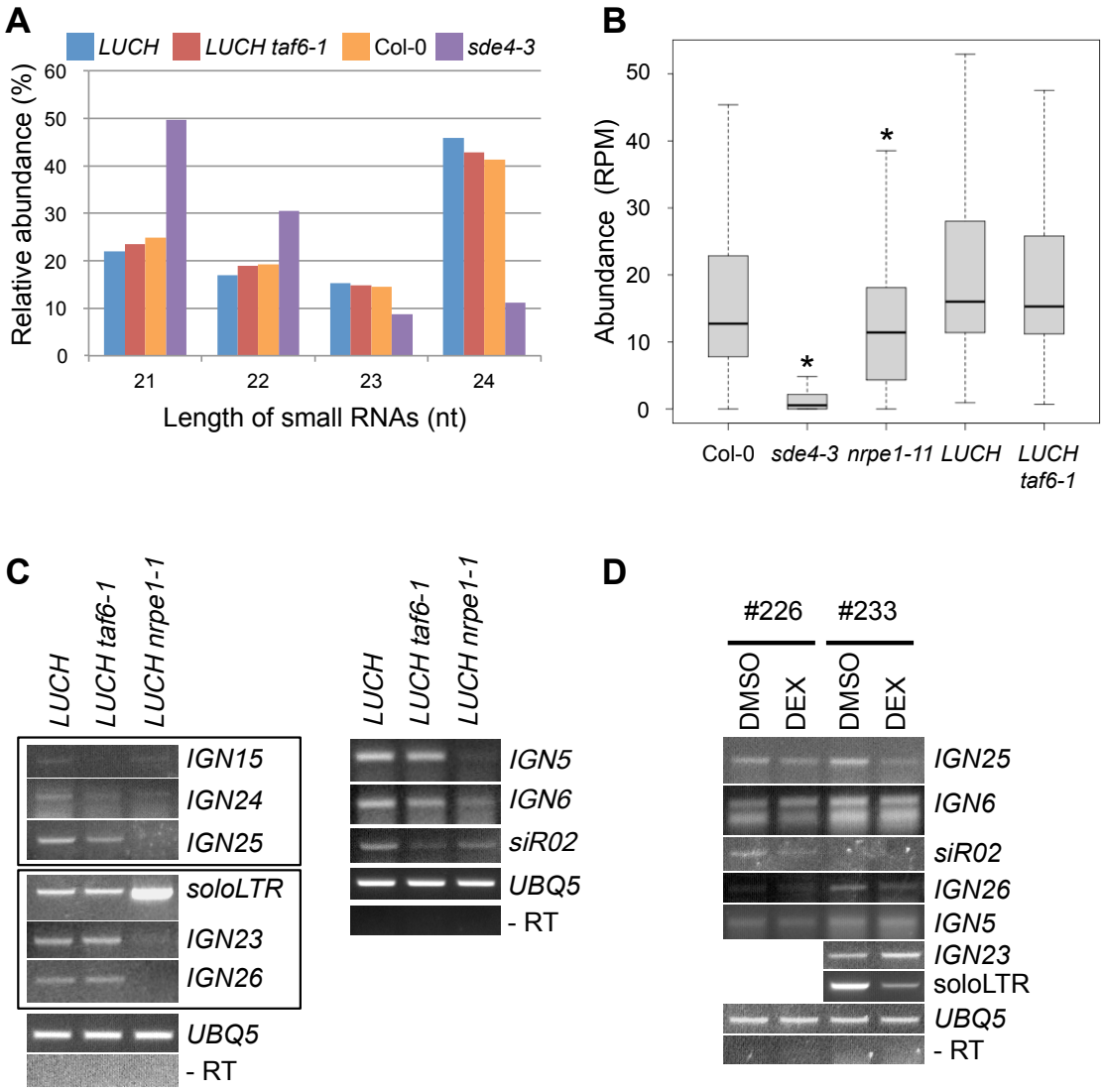

A
LUCH
LUCH taf6-1
Col-0
sde4-3
Relative abundance (%)
60
50
40
30
20
10
0
21
22
23
24
Length of small RNAs (nt)
B
Abundance (RPM)
50
40
30
20
10
0
*
*
Col-0
sde4-3
nrpe1-11
LUCH
LUCH taf6-1
C
LUCH
LUCH taf6-1
LUCH nrpe1-1
IGN15
IGN24
IGN25
soloLTR
IGN23
IGN26
UBQ5
- RT
LUCH
LUCH taf6-1
LUCH nrpe1-1
IGN5
IGN6
siR02
UBQ5
- RT
D
#226
#233
DMSO
DEX
DMSO
DEX
IGN25
IGN6
siR02
IGN26
IGN5
IGN23
soloLTR
UBQ5
- RT

Figure 2.9 Characterization of *amiR:TAF6*.

(A) *amiR:TAF6* plants in the segregating T2 population that were treated with DEX to induce the *TAF6*-specific artificial miRNA. Seeds were planted on MS media containing 10 μM DEX and grown for 10 days. The population segregated for certain morphological features including retarded growth and chlorosis. Plants grown on DMSO-treated MS media did not segregate for these phenotypes.
(B) Schematic diagram of the *TAF6* gene indicating the regions that were amplified for real-time RT-PCR. TAF6N and TAF6C were amplified to quantify the transcript levels of *TAF6*.
(C) Real-time RT-PCR for *TAF6*. Two independent *amiR:TAF6* lines were treated with DEX and DMSO. After RNA isolation and reverse transcription, real-time RT-PCR was performed. *UBQ5* was used to normalize the transcript levels of *TAF6*, and the DEX-treated samples were compared to the DMSO-treated samples. Induction of *amiR:TAF6* suppressed the expression of *TAF6* by approximately 50% or more.

Figure 2.9

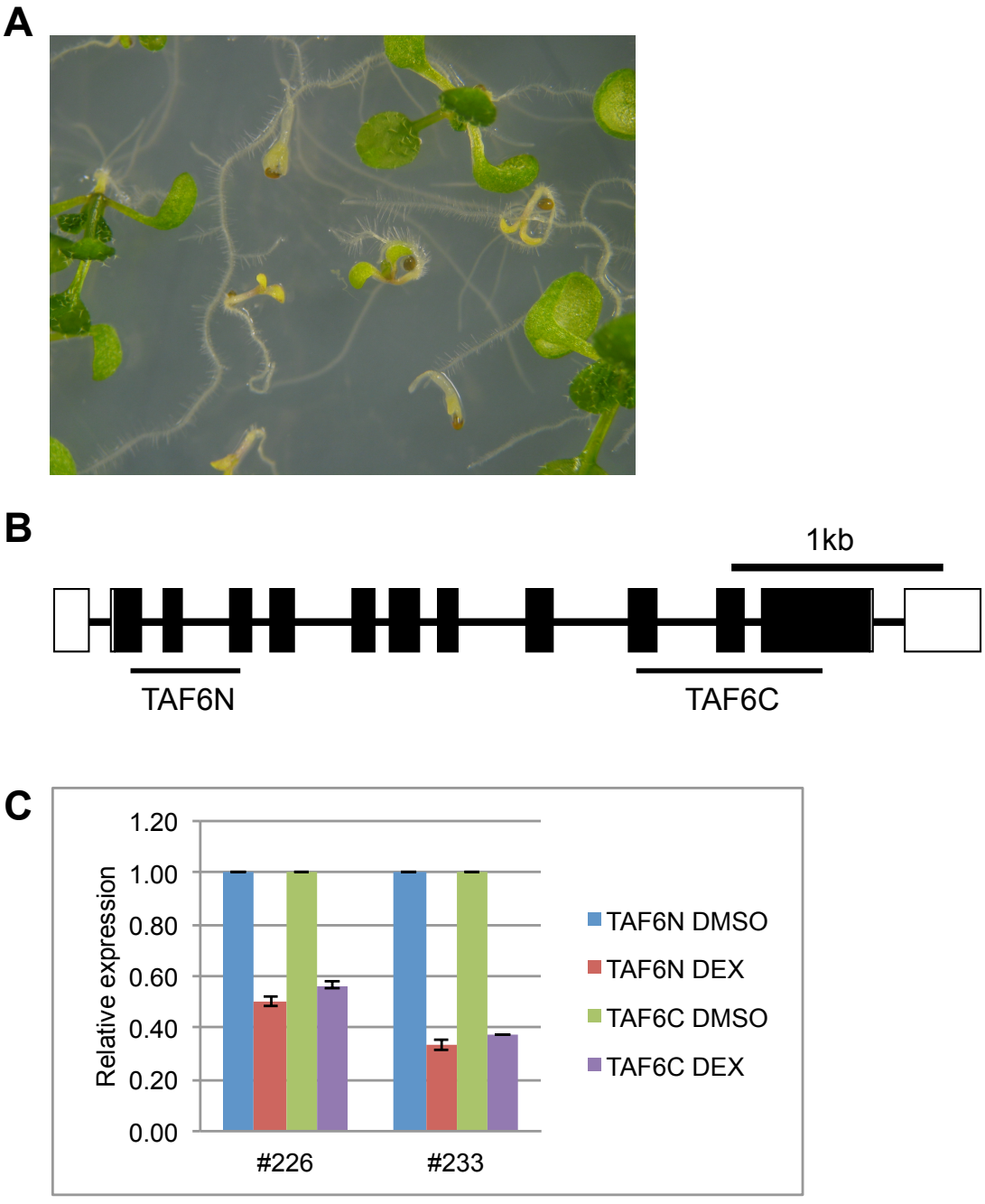
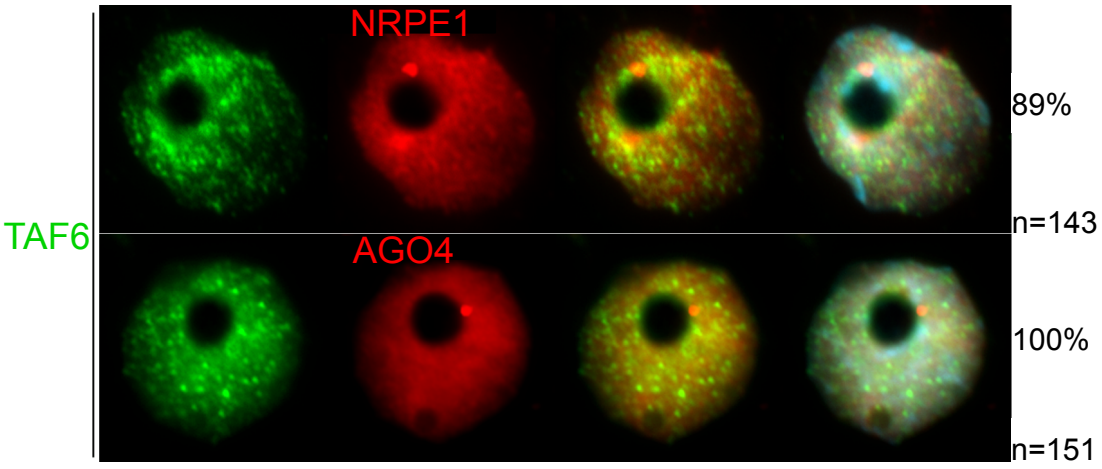

A
B
1kb
TAF6N
TAF6C
C
1.20
1.00
0.80
0.60
0.40
0.20
0.00
Relative expression
#226
#233
TAF6N DMSO
TAF6N DEX
TAF6C DMSO
TAF6C DEX

Figure 2.10 Subnuclear co-localization of TAF6 and NRPE1.

The localization of TAF6 (green), NRPE1 (red), or AGO4 (red) was examined by immunostaining. The distribution of NRPE1 and AGO4 is shown in the top and the bottom panel, respectively. In the merged image of TAF6 and NRPE1, some foci in yellow represent TAF6/NRPE1 co-localization. The nuclear DNA was stained with DAPI (blue) and shown in the merged images on the far right. "n" indicates the number of analyzed nuclei and "%" indicates the percentage of nuclei displaying the representative localization pattern.



TABLES

Table 2.1 Summary of bisulfite conversion efficiency for each genotype.

	CG	CHG	CHH	Total C
Col-0 (A)	98.2%	98.1%	97.8%	97.9%
Col-0 (B)	97.8%	97.7%	97.6%	97.6%
sde4-3 (A)	98.0%	98.0%	97.8%	97.9%
sde4-3 (B)	97.9%	97.8%	97.7%	97.7%
nrpe1-11 (A)	98.1%	98.1%	97.9%	98.0%
nrpe1-11 (B)	98.0%	98.0%	97.9%	97.9%
LUCH (A)	98.0%	98.0%	97.9%	97.9%
LUCH (B)	98.2%	98.2%	98.0%	98.0%
LUCH taf6-1 (A)	98.1%	98.0%	97.9%	98.0%
LUCH taf6-1 (B)	97.9%	97.9%	97.7%	97.8%

(A) and (B) indicate two independent biological replicates. All samples from the same biological replicates were processed at the same time and in the same manner.

Table 2.2 Read coverage of whole-genome bisulfite sequencing libraries.

CHH	# of sequenced ^{m}C	# of total sequenced C	31198380 [1]
			Coverage
Col-0 (A)	6974394	228191510	7.314
Col-0 (B)	8081587	276670598	8.868
sde4-3 (A)	3996254	221784087	7.109
sde4-3 (B)	3928539	238723704	7.652
nrpe1-11 (A)	5488089	293648870	9.412
nrpe1-11 (B)	3147078	175298018	5.619
LUCH (A)	8358149	276203474	8.853
LUCH (B)	7342876	281810404	9.033
LUCH taf6-1 (A)	8777346	283297569	9.081
LUCH taf6-1 (B)	6217968	247296968	7.927
CG	**# of sequenced ^{m}C**	**# of total sequenced C**	**5567714 [2]**
			Coverage
Col-0 (A)	11967876	40092653	7.201
Col-0 (B)	14292137	48943906	8.791
sde4-3 (A)	11233072	40655236	7.302
sde4-3 (B)	11410979	42322601	7.601
nrpe1-11 (A)	14992909	51859716	9.314
nrpe1-11 (B)	8591198	30902224	5.550
LUCH (A)	15786653	54455962	9.781
LUCH (B)	14068351	50695907	9.105
LUCH taf6-1 (A)	14969609	50784304	9.121
LUCH taf6-1 (B)	12027369	44818694	8.050
CHG	**# of sequenced ^{m}C**	**# of total sequenced C**	**6093657 [3]**
			Coverage
Col-0 (A)	4368090	43083879	7.070
Col-0 (B)	5122123	52159761	8.560
sde4-3 (A)	3536815	43885788	7.202
sde4-3 (B)	3458560	45355541	7.443
nrpe1-11 (A)	4982617	55895446	9.173
nrpe1-11 (B)	2750976	33011274	5.417
LUCH (A)	5982514	59537845	9.770
LUCH (B)	5188545	54916490	9.012
LUCH taf6-1 (A)	6224760	54652704	8.969
LUCH taf6-1 (B)	4561988	48552507	7.968
Total	**# of sequenced ^{m}C**	**# of total sequenced C**	**42859751 [4]**
			Coverage
Col-0 (A)	23310360	311368042	7.265
Col-0 (B)	27495847	377774265	8.814
sde4-3 (A)	18766141	306325111	7.147
sde4-3 (B)	18798078	326401846	7.616
nrpe1-11 (A)	25463615	401404032	9.366
nrpe1-11 (B)	14489252	239211516	5.581
LUCH (A)	30127316	390197281	9.104
LUCH (B)	26599772	387422801	9.039
LUCH taf6-1 (A)	29971715	388734577	9.070
LUCH taf6-1 (B)	22807325	340668169	7.948

(A) and (B) indicate two independent biological replicates. All samples from the same biological replicates were processed at the same time and in the same manner. Coverage = # of total sequenced C / # of total CXX sites in genome. [1] # of total CHH sites in genome. [2] # of total CG sites in genome. [3] # of total CHG sites in genome. [4] # of total C sites in genome

Table 2.3 Genes down-regulated in *LUCH taf6-1*.

AGI code	Relative levels (*taf6*/wt)	Description
AT1G01190	0.2962	cytochrome P450, family 78, subfamily A, polypeptide 8
AT1G01340	0.3900	cyclic nucleotide gated channel 10
AT1G01560	0.2316	MAP kinase 11
AT1G02230	0.3706	NAC domain containing protein 4
AT1G02400	0.2767	gibberellin 2-oxidase 6
AT1G04090	0.3824	Plant protein of unknown function (DUF946)
AT1G04570	0.3123	Major facilitator superfamily protein
AT1G05100	0.1310	mitogen-activated protein kinase kinase kinase 18
AT1G05575	0.4805	unknown protein
AT1G06100	0.4722	Fatty acid desaturase family protein
AT1G06160	0.3938	octadecanoid-responsive Arabidopsis AP2/ERF 59
AT1G06830	0.4356	Glutaredoxin family protein
AT1G07400	0.2817	HSP20-like chaperones superfamily protein
AT1G07620	0.4193	GTP-binding protein Obg/CgtA
AT1G08050	0.4180	Zinc finger (C3HC4-type RING finger) family protein
AT1G09950	0.3706	RESPONSE TO ABA AND SALT 1
AT1G10550	0.4700	xyloglucan:xyloglucosyl transferase 33
AT1G10585	0.2059	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT1G11080	0.3978	serine carboxypeptidase-like 31
AT1G11340	0.2918	S-locus lectin protein kinase family protein
AT1G11740	0.4805	ankyrin repeat family protein
AT1G12030	0.3295	Protein of unknown function (DUF506)
AT1G12160	0.3326	Flavin-binding monooxygenase family protein
AT1G12480	0.4152	C4-dicarboxylate transporter/malic acid transport protein
AT1G12570	0.3466	Glucose-methanol-choline (GMC) oxidoreductase family protein
AT1G12630	0.3881	Integrase-type DNA-binding superfamily protein
AT1G13140	0.4259	cytochrome P450, family 86, subfamily C, polypeptide 3
AT1G13420	0.3710	sulfotransferase 4B
AT1G13430	0.2507	sulfotransferase 4C
AT1G13470	0.2753	Protein of unknown function (DUF1262)
AT1G13608	0.4805	Defensin-like (DEFL) family protein
AT1G14240	0.2635	GDA1/CD39 nucleoside phosphatase family protein
AT1G14520	0.4219	myo-inositol oxygenase 1
AT1G14880	0.0312	PLANT CADMIUM RESISTANCE 1
AT1G14960	0.2162	Polyketide cyclase/dehydrase and lipid transport superfamily protein
AT1G15520	0.3666	pleiotropic drug resistance 12
AT1G15580	0.3603	indole-3-acetic acid inducible 5
AT1G16850	0.3961	unknown protein
AT1G16960	0.4900	Ubiquitin domain-containing protein
AT1G17380	0.1970	jasmonate-zim-domain protein 5
AT1G17420	0.2465	lipoxygenase 3
AT1G17600	0.4228	Disease resistance protein (TIR-NBS-LRR class) family
AT1G17610	0.2014	Disease resistance protein (TIR-NBS class)
AT1G17750	0.4260	PEP1 receptor 2
AT1G18710	0.3650	myb domain protein 47
AT1G18750	0.4418	AGAMOUS-like 65
AT1G19020	0.4748	unknown protein
AT1G19180	0.4491	jasmonate-zim-domain protein 1
AT1G19510	0.0940	RAD-like 5

AGI code	Relative levels (*taf6*/wt)	Description
AT1G19620	0.3284	unknown protein
AT1G19630	0.2102	cytochrome P450, family 722, subfamily A, polypeptide 1
AT1G19640	0.1836	jasmonic acid carboxyl methyltransferase
AT1G20515	0.3432	other RNA
AT1G20520	0.2507	Arabidopsis protein of unknown function (DUF241)
AT1G21120	0.2710	O-methyltransferase family protein
AT1G21240	0.1716	wall associated kinase 3
AT1G21550	0.3793	Calcium-binding EF-hand family protein
AT1G21910	0.4922	Integrase-type DNA-binding superfamily protein
AT1G22570	0.3566	Major facilitator superfamily protein
AT1G22590	0.4074	AGAMOUS-like 87
AT1G23060	0.4805	TPX2 (targeting protein for Xklp2) protein family
AT1G23110	0.4377	unknown protein
AT1G23200	0.3391	Plant invertase/pectin methylesterase inhibitor superfamily
AT1G23850	0.4633	unknown protein
AT1G24140	0.2831	Matrixin family protein
AT1G24145	0.3420	unknown protein
AT1G24147	0.2005	unknown protein
AT1G24400	0.4065	lysine histidine transporter 2
AT1G25530	0.4857	Transmembrane amino acid transporter family protein
AT1G26380	0.2469	FAD-binding Berberine family protein
AT1G26600	0.3855	CLAVATA3/ESR-RELATED 9
AT1G27730	0.3638	salt tolerance zinc finger
AT1G28480	0.1121	Thioredoxin superfamily protein
AT1G29600	0.1253	Zinc finger C-x8-C-x5-C-x3-H type family protein
AT1G29720	0.3161	Leucine-rich repeat transmembrane protein kinase
AT1G30135	0.0994	jasmonate-zim-domain protein 8
AT1G30280	0.4609	Chaperone DnaJ-domain superfamily protein
AT1G30840	0.2079	purine permease 4
AT1G31690	0.3910	Copper amine oxidase family protein
AT1G31750	0.3164	proline-rich family protein
AT1G32780	0.4096	GroES-like zinc-binding dehydrogenase family protein
AT1G33760	0.3164	Integrase-type DNA-binding superfamily protein
AT1G33930	0.3971	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT1G34060	0.4776	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
AT1G34580	0.4435	Major facilitator superfamily protein
AT1G35210	0.2845	unknown protein
AT1G35230	0.2429	arabinogalactan protein 5
AT1G35250	0.3363	Thioesterase superfamily protein
AT1G35710	0.4628	Protein kinase family protein with leucine-rich repeat domain
AT1G43160	0.0115	related to AP2 6
AT1G44350	0.4759	IAA-leucine resistant (ILR)-like gene 6
AT1G44970	0.4404	Peroxidase superfamily protein
AT1G47480	0.4633	alpha/beta-Hydrolases superfamily protein
AT1G47510	0.2607	inositol polyphosphate 5-phosphatase 11
AT1G48770	0.4710	Protein of unknown function (DUF1639)
AT1G48870	0.4596	Transducin/WD40 repeat-like superfamily protein
AT1G49450	0.2329	Transducin/WD40 repeat-like superfamily protein
AT1G51090	0.2376	Heavy metal transport/detoxification superfamily protein
AT1G51270	0.3486	structural molecules;transmembrane receptors;structural molecules
AT1G51670	0.1848	unknown protein

AGI code	Relative levels (*taf6*/wt)	Description
AT1G51780	0.0000	IAA-leucine resistant (ILR)-like gene 5
AT1G52060	0.2839	Mannose-binding lectin superfamily protein
AT1G52070	0.4188	Mannose-binding lectin superfamily protein
AT1G52290	0.4979	Protein kinase superfamily protein
AT1G52770	0.2053	Phototropic-responsive NPH3 family protein
AT1G52830	0.1264	indole-3-acetic acid 6
AT1G52890	0.1532	NAC domain containing protein 19
AT1G52920	0.4435	G protein coupled receptor
AT1G53470	0.3500	mechanosensitive channel of small conductance-like 4
AT1G53540	0.1747	HSP20-like chaperones superfamily protein
AT1G54660	0.2967	pseudogene
AT1G56060	0.1441	unknown protein
AT1G56150	0.4084	SAUR-like auxin-responsive protein family
AT1G56240	0.2883	phloem protein 2-B13
AT1G56242	0.2982	Potential natural antisense gene, locus overlaps with AT1G56240
AT1G56600	0.2147	galactinol synthase 2
AT1G56660	0.3969	unknown protein
AT1G56670	0.4922	GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G57630	0.0379	Toll-Interleukin-Resistance (TIR) domain family protein
AT1G58225	0.2677	unknown protein
AT1G58420	0.2402	Uncharacterised conserved protein UCP031279
AT1G59860	0.4447	HSP20-like chaperones superfamily protein
AT1G60190	0.4345	ARM repeat superfamily protein
AT1G61065	0.4539	Protein of unknown function (DUF1218)
AT1G61120	0.0470	terpene synthase 04
AT1G62400	0.3947	Protein kinase superfamily protein
AT1G62835	0.3793	pseudogene
AT1G65450	0.4118	HXXXD-type acyl-transferase family protein
AT1G65790	0.0933	receptor kinase 1
AT1G66090	0.2571	Disease resistance protein (TIR-NBS class)
AT1G66400	0.4759	calmodulin like 23
AT1G66460	0.2346	Protein kinase superfamily protein
AT1G66540	0.3335	Cytochrome P450 superfamily protein
AT1G66600	0.1860	ABA overly sensitive mutant 3
AT1G66650	0.4942	Protein with RING/U-box and TRAF-like domains
AT1G67070	0.3669	Mannose-6-phosphate isomerase, type I
AT1G67260	0.1713	TCP family transcription factor
AT1G67865	0.3580	unknown protein
AT1G67920	0.3179	unknown protein
AT1G69140	0.1641	pseudogene
AT1G69490	0.4239	NAC-like, activated by AP3/PI
AT1G69720	0.2480	heme oxygenase 3
AT1G70985	0.4211	hydroxyproline-rich glycoprotein family protein
AT1G70990	0.1965	proline-rich family protein
AT1G71000	0.0721	Chaperone DnaJ-domain superfamily protein
AT1G71015	0.4004	unknown protein
AT1G71520	0.1030	Integrase-type DNA-binding superfamily protein
AT1G72520	0.2613	PLAT/LH2 domain-containing lipoxygenase family protein
AT1G72910	0.2754	Toll-Interleukin-Resistance (TIR) domain-containing protein
AT1G73325	0.2682	Kunitz family trypsin and protease inhibitor protein
AT1G74150	0.4095	Galactose oxidase/kelch repeat superfamily protein

AGI code	Relative levels (*taf6*/wt)	Description
AT1G74545	0.4239	other RNA
AT1G74810	0.3535	HCO3- transporter family
AT1G74930	0.3028	Integrase-type DNA-binding superfamily protein
AT1G75030	0.3952	thaumatin-like protein 3
AT1G75250	0.1281	RAD-like 6
AT1G75960	0.4830	AMP-dependent synthetase and ligase family protein
AT1G76070	0.4416	unknown protein
AT1G76190	0.4387	SAUR-like auxin-responsive protein family
AT1G76530	0.0609	Auxin efflux carrier family protein
AT1G76600	0.4422	unknown protein
AT1G76640	0.2176	Calcium-binding EF-hand family protein
AT1G76650	0.4253	calmodulin-like 38
AT1G77145	0.2621	Protein of unknown function (DUF506)
AT1G78410	0.2402	VQ motif-containing protein
AT1G78450	0.3439	SOUL heme-binding family protein
AT1G78530	0.3789	Protein kinase superfamily protein
AT1G78815	0.1696	Protein of unknown function (DUF640)
AT1G79770	0.2517	Protein of unknown function (DUF1677)
AT1G79910	0.3286	Regulator of Vps4 activity in the MVB pathway protein
AT1G80120	0.2818	Protein of unknown function (DUF567)
AT1G80840	0.2362	WRKY DNA-binding protein 40
AT2G01200	0.2109	indole-3-acetic acid inducible 32
AT2G01520	0.0837	MLP-like protein 328
AT2G03530	0.4858	ureide permease 2
AT2G04450	0.4279	nudix hydrolase homolog 6
AT2G04495	0.3003	unknown protein
AT2G04800	0.1696	unknown protein
AT2G05160	0.1507	CCCH-type zinc fingerfamily protein with RNA-binding domain
AT2G05510	0.2883	Glycine-rich protein family
AT2G05518	0.3976	unknown protein
AT2G05914	0.0739	other RNA
AT2G05915	0.1264	unknown protein
AT2G07719	0.4239	Putative membrane lipoprotein
AT2G07774	0.2162	unknown protein
AT2G07777	0.4239	ATP synthase 9 mitochondrial
AT2G11810	0.2607	monogalactosyldiacylglycerol synthase type C
AT2G13550	0.4878	unknown protein
AT2G13570	0.4239	nuclear factor Y, subunit B7
AT2G14560	0.4584	Protein of unknown function (DUF567)
AT2G14580	0.2091	basic pathogenesis-related protein 1
AT2G14610	0.3620	pathogenesis-related gene 1
AT2G15040	0.2258	pseudogene
AT2G15042	0.2846	Leucine-rich repeat (LRR) family protein
AT2G15390	0.3712	fucosyltransferase 4
AT2G15830	0.3012	unknown protein
AT2G16005	0.0212	MD-2-related lipid recognition domain-containing protein
AT2G17470	0.2059	Aluminium activated malate transporter family protein
AT2G18210	0.1395	unknown protein
AT2G18480	0.1030	Major facilitator superfamily protein
AT2G18680	0.3760	unknown protein
AT2G18690	0.2625	unknown protein

AGI code	Relative levels (*taf6*/wt)	Description
AT2G20520	0.3089	FASCICLIN-like arabinogalactan 6
AT2G20880	0.3168	Integrase-type DNA-binding superfamily protein
AT2G21650	0.2176	Homeodomain-like superfamily protein
AT2G21880	0.4387	RAB GTPase homolog 7A
AT2G22810	0.2782	1-aminocyclopropane-1-carboxylate synthase 4
AT2G22821	0.4612	other RNA
AT2G22860	0.1471	phytosulfokine 2 precursor
AT2G22880	0.3535	VQ motif-containing protein
AT2G23560	0.4368	methyl esterase 7
AT2G24850	0.3472	tyrosine aminotransferase 3
AT2G25130	0.4004	ARM repeat superfamily protein
AT2G25150	0.4172	HXXXD-type acyl-transferase family protein
AT2G25160	0.2967	cytochrome P450, family 82, subfamily F, polypeptide 1
AT2G25460	0.3153	unknown protein
AT2G25735	0.3923	unknown protein
AT2G25780	0.2790	Protein of unknown function (DUF1677)
AT2G26150	0.1229	heat shock transcription factor A2
AT2G26560	0.3382	phospholipase A 2A
AT2G27690	0.2108	cytochrome P450, family 94, subfamily C, polypeptide 1
AT2G28085	0.0848	SAUR-like auxin-responsive protein family
AT2G28160	0.2162	FER-like regulator of iron uptake
AT2G28305	0.3716	Putative lysine decarboxylase family protein
AT2G29170	0.1880	NAD(P)-binding Rossmann-fold superfamily protein
AT2G29300	0.4624	NAD(P)-binding Rossmann-fold superfamily protein
AT2G29460	0.1533	glutathione S-transferase tau 4
AT2G29720	0.4849	FAD/NAD(P)-binding oxidoreductase family protein
AT2G29940	0.3805	pleiotropic drug resistance 3
AT2G30210	0.3745	laccase 3
AT2G30340	0.3326	LOB domain-containing protein 13
AT2G30400	0.4368	ovate family protein 2
AT2G30420	0.4667	Homeodomain-like superfamily protein
AT2G30432	0.2325	Homeodomain-like superfamily protein
AT2G30670	0.0000	NAD(P)-binding Rossmann-fold superfamily protein
AT2G30770	0.2570	cytochrome P450, family 71, subfamily A, polypeptide 13
AT2G30820	0.3805	unknown protein
AT2G30830	0.0449	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT2G31865	0.4826	poly(ADP-ribose) glycohydrolase 2
AT2G32140	0.4065	transmembrane receptors
AT2G32160	0.1566	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT2G32200	0.4204	unknown protein
AT2G32290	0.1378	beta-amylase 6
AT2G32510	0.3699	mitogen-activated protein kinase kinase kinase 17
AT2G32680	0.1993	receptor like protein 23
AT2G33205	0.4387	Serinc-domain containing serine and sphingolipid biosynthesis protein
AT2G33790	0.3760	arabinogalactan protein 30
AT2G34500	0.2607	cytochrome P450, family 710, subfamily A, polypeptide 1
AT2G34600	0.0920	jasmonate-zim-domain protein 7
AT2G34655	0.4931	unknown protein
AT2G34925	0.3540	CLAVATA3/ESR-RELATED 42
AT2G35290	0.3603	unknown protein

AGI code	Relative levels (*taf6*/wt)	Description
AT2G35345	0.4118	unknown protein
AT2G35585	0.4633	unknown protein
AT2G35980	0.3687	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
AT2G36490	0.4641	demeter-like 1
AT2G36590	0.2549	proline transporter 3
AT2G36750	0.2143	UDP-glucosyl transferase 73C1
AT2G37030	0.2059	SAUR-like auxin-responsive protein family
AT2G37950	0.3009	RING/FYVE/PHD zinc finger superfamily protein
AT2G38240	0.0985	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT2G38300	0.2752	myb-like HTH transcriptional regulator family protein
AT2G39250	0.4882	Integrase-type DNA-binding superfamily protein
AT2G39850	0.3714	Subtilisin-like serine endopeptidase family protein
AT2G40130	0.1752	Double Clp-N motif-containing P-loop nucleoside triphosphate hydrolases superfamily protein
AT2G40260	0.3800	Homeodomain-like superfamily protein
AT2G40750	0.2271	WRKY DNA-binding protein 54
AT2G44578	0.2217	RING/U-box superfamily protein
AT2G44810	0.0000	alpha/beta-Hydrolases superfamily protein
AT2G44840	0.4733	ethylene-responsive element binding factor 13
AT2G45660	0.4570	AGAMOUS-like 20
AT2G45900	0.4578	Phosphatidylinositol N-acetyglucosaminlytransferase subunit P-related
AT2G46940	0.4612	unknown protein
AT2G47050	0.2485	Plant invertase/pectin methylesterase inhibitor superfamily protein
AT2G47140	0.3931	NAD(P)-binding Rossmann-fold superfamily protein
AT2G47880	0.1614	Glutaredoxin family protein
AT3G01516	0.3391	unknown protein
AT3G01550	0.4331	phosphoenolpyruvate (pep)/phosphate translocator 2
AT3G02670	0.3203	Glycine-rich protein family
AT3G02800	0.4043	Tyrosine phosphatase family protein
AT3G02840	0.3363	ARM repeat superfamily protein
AT3G03480	0.4185	acetyl CoA:(Z)-3-hexen-1-ol acetyltransferase
AT3G04010	0.3503	O-Glycosyl hydrolases family 17 protein
AT3G04210	0.4522	Disease resistance protein (TIR-NBS class)
AT3G05685	0.4541	Cystatin/monellin superfamily protein
AT3G06145	0.3428	unknown protein
AT3G09010	0.4036	Protein kinase superfamily protein
AT3G09922	0.2942	induced by phosphate starvation1
AT3G09940	0.2653	monodehydroascorbate reductase
AT3G10150	0.2346	purple acid phosphatase 16
AT3G10185	0.0759	Gibberellin-regulated family protein
AT3G11110	0.4939	RING/U-box superfamily protein
AT3G11340	0.3378	UDP-Glycosyltransferase superfamily protein
AT3G11480	0.0390	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT3G11640	0.1310	unknown protein
AT3G12040	0.3080	DNA-3-methyladenine glycosylase (MAG)
AT3G13540	0.1897	myb domain protein 5
AT3G13950	0.1758	unknown protein
AT3G15270	0.2346	squamosa promoter binding protein-like 5
AT3G15358	0.3824	unknown protein

AGI code	Relative levels (*taf6*/wt)	Description
AT3G15500	0.2402	NAC domain containing protein 3
AT3G15540	0.3432	indole-3-acetic acid inducible 19
AT3G16070	0.2276	unknown protein
AT3G16120	0.4185	Dynein light chain type 1 family protein
AT3G17130	0.4424	Plant invertase/pectin methylesterase inhibitor superfamily protein
AT3G18070	0.3449	beta glucosidase 43
AT3G18145	0.3636	pseudogene
AT3G18950	0.3788	Transducin/WD40 repeat-like superfamily protein
AT3G19550	0.2682	unknown protein
AT3G20087	0.3603	pseudogene
AT3G20200	0.4894	Protein kinase protein with adenine nucleotide alpha hydrolases-like domain
AT3G20340	0.3203	unknown protein
AT3G21460	0.0257	Glutaredoxin family protein
AT3G21520	0.3326	DUF679 domain membrane protein 1
AT3G21890	0.4570	B-box type zinc finger family protein
AT3G22060	0.3287	Receptor-like protein kinase-related family protein
AT3G22231	0.0726	pathogen and circadian controlled 1
AT3G22235	0.2883	unknown protein
AT3G22275	0.0370	unknown protein
AT3G22570	0.3844	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT3G23010	0.3793	receptor like protein 36
AT3G23120	0.1948	receptor like protein 38
AT3G23230	0.0901	Integrase-type DNA-binding superfamily protein
AT3G23480	0.2773	Cyclopropane-fatty-acyl-phospholipid synthase
AT3G23550	0.3467	MATE efflux family protein
AT3G23840	0.4066	HXXXD-type acyl-transferase family protein
AT3G24715	0.4202	Protein kinase superfamily protein with octicosapeptide/Phox/Bem1p domain
AT3G24900	0.2193	receptor like protein 39
AT3G25010	0.2510	receptor like protein 41
AT3G25240	0.4324	Protein of unknown function (DUF506)
AT3G25620	0.3212	ABC-2 type transporter family protein
AT3G25780	0.3536	allene oxide cyclase 3
AT3G25882	0.4596	NIM1-interacting 2
AT3G25990	0.3540	Homeodomain-like superfamily protein
AT3G26200	0.2490	cytochrome P450, family 71, subfamily B, polypeptide 22
AT3G26490	0.4490	Phototropic-responsive NPH3 family protein
AT3G26500	0.4525	plant intracellular ras group-related LRR 2
AT3G26830	0.3234	Cytochrome P450 superfamily protein
AT3G27809	0.0450	unknown protein
AT3G27900	0.3276	Protein of unknown function (DUF1184)
AT3G28340	0.4704	galacturonosyltransferase-like 10
AT3G28510	0.4663	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT3G29000	0.4684	Calcium-binding EF-hand family protein
AT3G30460	0.3255	RING/U-box superfamily protein
AT3G42800	0.3159	unknown protein
AT3G43110	0.3603	unknown protein
AT3G43250	0.0000	Family of unknown function (DUF572)
AT3G44860	0.1553	farnesoic acid carboxyl-O-methyltransferase

AGI code	Relative levels (*taf6*/wt)	Description
AT3G45060	0.4004	high affinity nitrate transporter 2.6
AT3G45070	0.3313	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT3G45290	0.4752	Seven transmembrane MLO family protein
AT3G45430	0.2837	Concanavalin A-like lectin protein kinase family protein
AT3G45638	0.2574	other RNA
AT3G45860	0.2217	cysteine-rich RLK (RECEPTOR-like protein kinase) 4
AT3G45960	0.1647	expansin-like A3
AT3G46230	0.3171	heat shock protein 17.4
AT3G46370	0.1635	Leucine-rich repeat protein kinase family protein
AT3G46658	0.2522	other RNA
AT3G46660	0.2488	UDP-glucosyl transferase 76E12
AT3G46880	0.1696	unknown protein
AT3G48020	0.2135	unknown protein
AT3G48080	0.3346	alpha/beta-Hydrolases superfamily protein
AT3G48280	0.3940	cytochrome P450, family 71, subfamily A, polypeptide 25
AT3G48450	0.3479	RPM1-interacting protein 4 (RIN4) family protein
AT3G48520	0.0312	cytochrome P450, family 94, subfamily B, polypeptide 3
AT3G48640	0.0554	unknown protein
AT3G48650	0.2826	pseudogene
AT3G48660	0.4079	Protein of unknown function (DUF 3339)
AT3G49620	0.4318	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT3G49930	0.3793	C2H2 and C2HC zinc fingers superfamily protein
AT3G50280	0.1622	HXXXD-type acyl-transferase family protein
AT3G50470	0.3109	homolog of RPW8 3
AT3G50800	0.1822	unknown protein
AT3G50970	0.2419	dehydrin family protein
AT3G51750	0.1181	unknown protein
AT3G52450	0.3993	plant U-box 22
AT3G53600	0.0000	C2H2-type zinc finger family protein
AT3G53720	0.3844	cation/H+ exchanger 20
AT3G55646	0.3687	unknown protein
AT3G55710	0.2033	UDP-Glycosyltransferase superfamily protein
AT3G55940	0.2517	Phosphoinositide-specific phospholipase C family protein
AT3G55970	0.2732	jasmonate-regulated gene 21
AT3G56400	0.3556	WRKY DNA-binding protein 70
AT3G56710	0.3898	sigma factor binding protein 1
AT3G56790	0.3326	RNA splicing factor-related
AT3G57010	0.4712	Calcium-dependent phosphotriesterase superfamily protein
AT3G57260	0.3671	beta-1,3-glucanase 2
AT3G57460	0.2708	catalytics;metal ion binding
AT3G58070	0.4316	C2H2 and C2HC zinc fingers superfamily protein
AT3G59010	0.3762	pectin methylesterase 61
AT3G59250	0.4285	F-box/RNI-like superfamily protein
AT3G60420	0.4947	Phosphoglycerate mutase family protein
AT3G60550	0.2522	cyclin p3;2
AT3G61190	0.4435	BON association protein 1
AT3G61280	0.3211	Arabidopsis thaliana protein of unknown function (DUF821)
AT3G61920	0.1957	unknown protein
AT3G61970	0.4504	AP2/B3-like transcriptional factor family protein
AT3G62740	0.3133	beta glucosidase 7
AT3G62780	0.2574	Calcium-dependent lipid-binding (CaLB domain) family protein

AGI code	Relative levels (*taf6*/wt)	Description
AT3G62950	0.2760	Thioredoxin superfamily protein
AT3G62960	0.1716	Thioredoxin superfamily protein
AT3G63380	0.3983	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein
AT4G00970	0.4303	cysteine-rich RLK (RECEPTOR-like protein kinase) 41
AT4G01080	0.2797	TRICHOME BIREFRINGENCE-LIKE 26
AT4G02360	0.3284	Protein of unknown function, DUF538
AT4G03330	0.1948	syntaxin of plants 123
AT4G03450	0.0594	Ankyrin repeat family protein
AT4G03965	0.4036	RING/U-box superfamily protein
AT4G05020	0.4995	NAD(P)H dehydrogenase B2
AT4G05100	0.1426	myb domain protein 74
AT4G08040	0.3964	1-aminocyclopropane-1-carboxylate synthase 11
AT4G08555	0.3857	unknown protein
AT4G10390	0.3810	Protein kinase superfamily protein
AT4G10500	0.2346	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT4G11000	0.1758	Ankyrin repeat family protein
AT4G11521	0.4014	Receptor-like protein kinase-related family protein
AT4G11890	0.4287	Protein kinase superfamily protein
AT4G12005	0.4435	unknown protein
AT4G12510	0.1201	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G12520	0.4118	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G13280	0.2059	terpenoid synthase 12
AT4G13290	0.3654	cytochrome P450, family 71, subfamily A, polypeptide 19
AT4G13395	0.2204	ROTUNDIFOLIA like 12
AT4G13410	0.3803	Nucleotide-diphospho-sugar transferases superfamily protein
AT4G13560	0.4054	Late embryogenesis abundant protein (LEA) family protein
AT4G14060	0.1802	Polyketide cyclase/dehydrase and lipid transport superfamily protein
AT4G14400	0.2307	ankyrin repeat family protein
AT4G14650	0.4709	unknown protein
AT4G15440	0.4341	hydroperoxide lyase 1
AT4G15660	0.0676	Thioredoxin superfamily protein
AT4G15670	0.2023	Thioredoxin superfamily protein
AT4G15680	0.3120	Thioredoxin superfamily protein
AT4G15690	0.4358	Thioredoxin superfamily protein
AT4G15700	0.3943	Thioredoxin superfamily protein
AT4G15990	0.3459	unknown protein
AT4G16590	0.1157	cellulose synthase-like A01
AT4G17030	0.3894	expansin-like B1
AT4G17860	0.3904	Protein of Unknown Function (DUF239)
AT4G17970	0.4842	aluminum-activated, malate transporter 12
AT4G18050	0.4589	P-glycoprotein 9
AT4G19430	0.1373	unknown protein
AT4G19645	0.4227	TRAM, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein
AT4G19690	0.1880	iron-regulated transporter 1
AT4G20230	0.2988	Terpenoid cyclases/Protein prenyltransferases superfamily protein
AT4G20970	0.3255	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT4G21390	0.4303	S-locus lectin protein kinase family protein
AT4G21440	0.1395	MYB-like 102

AGI code	Relative levels (*taf6*/wt)	Description
AT4G21650	0.4189	Subtilase family protein
AT4G21830	0.2812	methionine sulfoxide reductase B7
AT4G21920	0.3133	unknown protein
AT4G22305	0.3950	alpha/beta-Hydrolases superfamily protein
AT4G22545	0.1441	pseudogene
AT4G22610	0.0901	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G22620	0.1310	SAUR-like auxin-responsive protein family
AT4G22666	0.3844	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G23140	0.1500	cysteine-rich RLK (RECEPTOR-like protein kinase) 6
AT4G23170	0.4805	receptor-like protein kinase-related family protein
AT4G23215	0.3363	pseudogene
AT4G23220	0.3317	cysteine-rich RLK (RECEPTOR-like protein kinase) 14
AT4G23230	0.3952	cysteine-rich RLK (RECEPTOR-like protein kinase) 15
AT4G24480	0.3504	Protein kinase superfamily protein
AT4G24540	0.4217	AGAMOUS-like 24
AT4G25940	0.4763	ENTH/ANTH/VHS superfamily protein
AT4G26320	0.4324	arabinogalactan protein 13
AT4G26470	0.4263	Calcium-binding EF-hand family protein
AT4G27410	0.2850	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
AT4G27654	0.0686	unknown protein
AT4G27657	0.4155	unknown protein
AT4G28085	0.3909	unknown protein
AT4G28140	0.1341	Integrase-type DNA-binding superfamily protein
AT4G28350	0.4250	Concanavalin A-like lectin protein kinase family protein
AT4G29270	0.3979	HAD superfamily, subfamily IIIB acid phosphatase
AT4G29690	0.4650	Alkaline-phosphatase-like family protein
AT4G29710	0.4036	Alkaline-phosphatase-like family protein
AT4G29740	0.3740	cytokinin oxidase 4
AT4G29930	0.1299	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT4G31870	0.3720	glutathione peroxidase 7
AT4G32800	0.4101	Integrase-type DNA-binding superfamily protein
AT4G32870	0.4144	Polyketide cyclase/dehydrase and lipid transport superfamily protein
AT4G32950	0.4065	Protein phosphatase 2C family protein
AT4G33905	0.1529	Peroxisomal membrane 22 kDa (Mpv17/PMP22) family protein
AT4G33980	0.4851	unknown protein
AT4G34060	0.2734	demeter-like protein 3
AT4G34410	0.2566	redox responsive transcription factor 1
AT4G34930	0.2932	PLC-like phosphodiesterases superfamily protein
AT4G35030	0.3844	Protein kinase superfamily protein
AT4G35160	0.3391	O-methyltransferase family protein
AT4G35290	0.4177	glutamate receptor 2
AT4G35783	0.4970	ROTUNDIFOLIA like 6
AT4G35810	0.3363	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT4G36570	0.2242	RAD-like 3
AT4G36950	0.0707	mitogen-activated protein kinase kinase kinase 21
AT4G37370	0.4805	cytochrome P450, family 81, subfamily D, polypeptide 8
AT4G37990	0.0465	elicitor-activated gene 3-2
AT4G38560	0.2786	Arabidopsis phospholipase-like protein (PEARLI 4) family

AGI code	Relative levels (*taf6*/wt)	Description
AT4G39250	0.0406	RAD-like 1
AT4G39670	0.4564	Glycolipid transfer protein (GLTP) family protein
AT4G40065	0.4347	other RNA
AT5G01015	0.4761	unknown protein
AT5G01100	0.3558	O-fucosyltransferase family protein
AT5G01250	0.3720	alpha 1,4-glycosyltransferase family protein
AT5G02180	0.4951	Transmembrane amino acid transporter family protein
AT5G02540	0.3569	NAD(P)-binding Rossmann-fold superfamily protein
AT5G03060	0.3931	unknown protein
AT5G03130	0.0703	unknown protein
AT5G03210	0.1030	unknown protein
AT5G03890	0.4185	unknown protein
AT5G04150	0.3009	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT5G04340	0.2432	zinc finger of Arabidopsis thaliana 6
AT5G05220	0.1696	unknown protein
AT5G05410	0.3292	DRE-binding protein 2A
AT5G06570	0.2883	alpha/beta-Hydrolases superfamily protein
AT5G06790	0.4018	unknown protein
AT5G08070	0.4504	TCP domain protein 17
AT5G08240	0.4848	unknown protein
AT5G09470	0.2109	dicarboxylate carrier 3
AT5G10130	0.0759	Pollen Ole e 1 allergen and extensin family protein
AT5G10250	0.3276	Phototropic-responsive NPH3 family protein
AT5G10410	0.4095	ENTH/ANTH/VHS superfamily protein
AT5G10600	0.3791	cytochrome P450, family 81, subfamily K, polypeptide 2
AT5G10605	0.4158	methyltransferases
AT5G10760	0.3704	Eukaryotic aspartyl protease family protein
AT5G10946	0.4805	unknown protein
AT5G11190	0.4518	Integrase-type DNA-binding superfamily protein
AT5G11920	0.3829	6-&1-fructan exohydrolase
AT5G12020	0.1671	17.6 kDa class II heat shock protein
AT5G13200	0.4929	GRAM domain family protein
AT5G13220	0.2393	jasmonate-zim-domain protein 10
AT5G13320	0.2091	Auxin-responsive GH3 family protein
AT5G14070	0.4970	Thioredoxin superfamily protein
AT5G15240	0.4420	Transmembrane amino acid transporter family protein
AT5G15500	0.3373	Ankyrin repeat family protein
AT5G16170	0.2059	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
AT5G16570	0.3603	glutamine synthetase 1;4
AT5G17040	0.2812	UDP-Glycosyltransferase superfamily protein
AT5G17350	0.4060	unknown protein
AT5G17390	0.2883	Adenine nucleotide alpha hydrolases-like superfamily protein
AT5G17490	0.4876	RGA-like protein 3
AT5G18060	0.3775	SAUR-like auxin-responsive protein family
AT5G18300	0.3687	NAC domain containing protein 88
AT5G18404	0.0534	unknown protein
AT5G18430	0.4515	GDSL-like Lipase/Acylhydrolase superfamily protein
AT5G18470	0.4655	Curculin-like (mannose-binding) lectin family protein
AT5G18540	0.4007	unknown protein
AT5G19580	0.2456	glyoxal oxidase-related protein

AGI code	Relative levels (*taf6*/wt)	Description
AT5G20410	0.4524	monogalactosyldiacylglycerol synthase 2
AT5G20670	0.4172	Protein of unknown function (DUF1677)
AT5G20790	0.4549	unknown protein
AT5G22250	0.4740	Polynucleotidyl transferase, ribonuclease H-like superfamily protein
AT5G22520	0.3203	unknown protein
AT5G22530	0.1201	unknown protein
AT5G23360	0.4048	GRAM domain-containing protein / ABA-responsive protein-related
AT5G23950	0.4805	Calcium-dependent lipid-binding (CaLB domain) family protein
AT5G24105	0.3766	arabinogalactan protein 41
AT5G24110	0.2655	WRKY DNA-binding protein 30
AT5G24150	0.4978	FAD/NAD(P)-binding oxidoreductase family protein
AT5G24270	0.4739	Calcium-binding EF-hand family protein
AT5G24530	0.4177	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT5G24920	0.2206	glutamine dumper 5
AT5G25250	0.2320	SPFH/Band 7/PHB domain-containing membrane-associated protein family
AT5G25830	0.3900	GATA transcription factor 12
AT5G25880	0.3851	NADP-malic enzyme 3
AT5G25970	0.2677	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
AT5G26660	0.3737	myb domain protein 86
AT5G28440	0.2276	unknown protein
AT5G34795	0.3603	pseudogene
AT5G35525	0.2883	PLAC8 family protein
AT5G37500	0.2952	gated outwardly-rectifying K+ channel
AT5G37990	0.2059	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT5G38320	0.0874	unknown protein
AT5G38710	0.3126	Methylenetetrahydrofolate reductase family protein
AT5G38930	0.2760	RmlC-like cupins superfamily protein
AT5G38940	0.4420	RmlC-like cupins superfamily protein
AT5G40330	0.4834	myb domain protein 23
AT5G40780	0.4569	lysine histidine transporter 1
AT5G42250	0.4262	Zinc-binding alcohol dehydrogenase family protein
AT5G42380	0.3506	calmodulin like 37
AT5G43620	0.3884	Pre-mRNA cleavage complex II
AT5G44050	0.3061	MATE efflux family protein
AT5G44565	0.3603	unknown protein
AT5G44610	0.3391	microtubule-associated protein 18
AT5G44973	0.3844	a defensin-like (DEFL) family protein
AT5G46500	0.3687	disease resistance protein (TIR-NBS-LRR class) family
AT5G46890	0.3133	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G46900	0.1281	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G47160	0.4271	YDG/SRA domain-containing protein
AT5G47220	0.4527	ethylene responsive element binding factor 2
AT5G47240	0.3480	nudix hydrolase homolog 8
AT5G47450	0.1910	tonoplast intrinsic protein 2;3
AT5G47990	0.4057	cytochrome P450, family 705, subfamily A, polypeptide 5
AT5G48430	0.2967	Eukaryotic aspartyl protease family protein

AGI code	Relative levels (*taf6*/wt)	Description
AT5G48940	0.3363	Leucine-rich repeat transmembrane protein kinase family protein
AT5G48950	0.2803	Thioesterase superfamily protein
AT5G51790	0.0613	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT5G52070	0.3984	Agenet domain-containing protein
AT5G52320	0.4319	cytochrome P450, family 96, subfamily A, polypeptide 4
AT5G52640	0.3332	heat shock protein 90.1
AT5G52740	0.2402	Copper transport protein family
AT5G52760	0.4663	Copper transport protein family
AT5G53200	0.4185	Homeodomain-like superfamily protein
AT5G53250	0.2217	arabinogalactan protein 22
AT5G53710	0.2217	unknown protein
AT5G53760	0.4572	Seven transmembrane MLO family protein
AT5G54165	0.2682	unknown protein
AT5G54585	0.3179	unknown protein
AT5G54610	0.1061	ankyrin
AT5G55250	0.4324	IAA carboxylmethyltransferase 1
AT5G55570	0.2147	unknown protein
AT5G56370	0.4564	F-box/RNI-like/FBD-like domains-containing protein
AT5G56840	0.1741	myb-like transcription factor family protein
AT5G56980	0.4870	unknown protein
AT5G56990	0.3432	Cystatin/monellin superfamily protein
AT5G57123	0.4225	unknown protein
AT5G57500	0.4989	Galactosyltransferase family protein
AT5G59050	0.4324	unknown protein
AT5G59220	0.3055	highly ABA-induced PP2C gene 1
AT5G59330	0.1910	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G59580	0.2691	UDP-glucosyl transferase 76E1
AT5G59670	0.1206	Leucine-rich repeat protein kinase family protein
AT5G60240	0.3603	unknown protein
AT5G60530	0.3931	late embryogenesis abundant protein-related / LEA protein-related
AT5G60780	0.1948	nitrate transporter 2.3
AT5G60900	0.2238	receptor-like protein kinase 1
AT5G60910	0.2258	AGAMOUS-like 8
AT5G61160	0.3777	anthocyanin 5-aromatic acyltransferase 1
AT5G61560	0.4118	U-box domain-containing protein kinase family protein
AT5G61600	0.4279	ethylene response factor 104
AT5G62040	0.2883	PEBP (phosphatidylethanolamine-binding protein) family protein
AT5G62340	0.2120	Plant invertase/pectin methylesterase inhibitor superfamily protein
AT5G62360	0.4935	Plant invertase/pectin methylesterase inhibitor superfamily protein
AT5G62850	0.3964	Nodulin MtN3 family protein
AT5G63450	0.1224	cytochrome P450, family 94, subfamily B, polypeptide 1
AT5G63660	0.1109	Scorpion toxin-like knottin superfamily protein
AT5G64110	0.2178	Peroxidase superfamily protein
AT5G64190	0.2669	unknown protein
AT5G64510	0.2426	unknown protein
AT5G64750	0.0746	Integrase-type DNA-binding superfamily protein
AT5G65140	0.4211	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
AT5G65300	0.2621	unknown protein
AT5G65600	0.2883	Concanavalin A-like lectin protein kinase family protein
AT5G65800	0.3603	ACC synthase 5

AGI code	Relative levels (*taf6*/wt)	Description
AT5G67080	0.1395	mitogen-activated protein kinase kinase kinase 19
AT5G67210	0.3820	Protein of unknown function (DUF579)
AT5G67310	0.1758	cytochrome P450, family 81, subfamily G, polypeptide 1
AT1TE10605	0.4324	ATIS112A
AT1TE10615	0.4805	ATLINE1_6
AT1TE20970	0.3326	ATLINE1_4
AT1TE25675	0.2359	HELITRONY1E
AT1TE36545	0.4504	ATHILA4D_LTR
AT1TE53070	0.3326	ATCOPIA87
AT1TE59820	0.4910	ATREP16
AT1TE70175	0.3851	ATENSPM1A
AT1TE80815	0.3881	ATREP10D
AT1TE81055	0.4118	LIMPET1
AT2G06045	0.0801	transposable element gene
AT2TE22815	0.4222	ATSINE2A
AT2TE25295	0.3303	ATHILA2
AT2TE25440	0.3373	ATREP3
AT2TE26610	0.3442	ATGP2N
AT2TE43295	0.4522	ATMU10
AT2TE54495	0.4435	ARNOLD3
AT2TE55520	0.4118	TAG2
AT2TE57460	0.3203	BRODYAGA1A
AT2TE57465	0.4368	ATREP11
AT2TE67330	0.1517	VANDAL3
AT2TE71850	0.1310	ATDNAI27T9A
AT3G15310	0.2995	transposable element gene
AT3G21050	0.4942	transposable element gene
AT3TE18945	0.4970	HELITRONY1C
AT3TE21700	0.2847	ATIS112A
AT3TE25130	0.4805	HELITRONY3
AT3TE30980	0.4942	Unassigned
AT3TE80275	0.2306	ATMU10
AT3TE91085	0.4805	ATDNAI27T9A
AT4G01490	0.1310	transposable element gene
AT4G08100	0.4680	transposable element gene
AT4G09480	0.3326	transposable element gene
AT4G28900	0.3057	transposable element gene
AT4TE03295	0.1201	ATLINE1_6
AT4TE09335	0.4435	RathE1_cons
AT4TE10030	0.4805	ATREP10
AT4TE10035	0.4118	ATHATN1
AT4TE11410	0.2217	ATLINE1_1
AT4TE21110	0.4680	Unassigned
AT4TE25325	0.3844	ATCOPIA90
AT4TE33010	0.4762	ATDNA1T9A
AT4TE39495	0.2982	ATREP13
AT4TE67490	0.3042	ATCOPIA46
AT4TE69880	0.4989	HELITRONY1B
AT5G34790	0.4225	transposable element gene
AT5G34800	0.3403	transposable element gene
AT5TE29575	0.4271	HELITRONY3

AGI code	Relative levels (*taf6*/wt)	Description
AT5TE39790	0.3931	ATLINE1A
AT5TE46155	0.3603	VANDAL20
AT5TE46165	0.3641	VANDAL20
AT5TE55640	0.1044	ATLINE1_3A
AT5TE57400	0.3720	ATENSPM5
AT5TE58950	0.4569	HELITRONY1B
AT5TE64925	0.3236	ATENSPM1A
AT5TE80010	0.4239	ARNOLDY2

Table 2.4 Genes up-regulated in *LUCH taf6-1*.

ID	Relative levels (*taf6*/wt)	Description
AT1G01670	2.2712	RING/U-box superfamily protein
AT1G02340	2.1921	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT1G02770	5.2413	Protein of unknown function (DUF626)
AT1G03790	18.7376	Zinc finger C-x8-C-x5-C-x3-H type family protein
AT1G04660	2.9428	glycine-rich protein
AT1G05490	3.1448	chromatin remodeling 31
AT1G05700	4.5042	Leucine-rich repeat transmembrane protein kinase protein
AT1G06350	2.0965	Fatty acid desaturase family protein
AT1G07985	2.4896	Expressed protein
AT1G08430	2.6889	aluminum-activated malate transporter 1
AT1G11070	4.8846	Hydroxyproline-rich glycoprotein family protein
AT1G12010	2.1589	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
AT1G12064	3.6034	unknown protein
AT1G12805	3.8251	nucleotide binding
AT1G13300	2.5893	myb-like transcription factor family protein
AT1G13448	2.0965	other RNA
AT1G14642	2.0820	unknown protein
AT1G15150	9.2658	MATE efflux family protein
AT1G15330	2.4896	Cystathionine beta-synthase (CBS) protein
AT1G16400	2.2956	cytochrome P450, family 79, subfamily F, polypeptide 2
AT1G17020	2.0948	senescence-related gene 1
AT1G17030	2.2135	unknown protein
AT1G17960	11.0035	Threonyl-tRNA synthetase
AT1G18830	30.1243	Transducin/WD40 repeat-like superfamily protein
AT1G19610	4.6123	Arabidopsis defensin-like protein
AT1G20350	2.0591	translocase inner membrane subunit 17-1
AT1G20400	2.9454	Protein of unknown function (DUF1204)
AT1G22980	3.9122	unknown protein
AT1G23600	25.9444	Domain of unknown function DUF220
AT1G24260	2.9520	K-box region and MADS-box transcription factor family protein
AT1G24270	2.9586	unknown protein
AT1G26250	4.3241	Proline-rich extensin-like family protein
AT1G29090	ND	Cysteine proteinases superfamily protein
AT1G29195	2.3062	unknown protein
AT1G30160	2.7131	Protein of unknown function (DUF295)
AT1G30220	2.4791	inositol transporter 2
AT1G30760	9.2395	FAD-binding Berberine family protein
AT1G31240	4.4682	Bromodomain transcription factor
AT1G31290	3.8628	ARGONAUTE 3
AT1G32560	4.5643	Late embryogenesis abundant protein, group 1 protein
AT1G36180	3.1791	acetyl-CoA carboxylase 2
AT1G47540	25.9444	Scorpion toxin-like knottin superfamily protein
AT1G48130	53.3301	1-cysteine peroxiredoxin 1
AT1G48660	3.4407	Auxin-responsive GH3 family protein
AT1G49470	2.0395	Family of unknown function (DUF716)
AT1G49570	7.7473	Peroxidase superfamily protein
AT1G49920	3.5610	MuDR family transposase
AT1G51850	2.2811	Leucine-rich repeat protein kinase family protein

ID	Relative levels (*taf6*/wt)	Description
AT1G52100	3.9520	Mannose-binding lectin superfamily protein
AT1G53080	10.8102	Legume lectin family protein
AT1G53830	2.7940	pectin methylesterase 2
AT1G54020	2.9775	GDSL-like Lipase/Acylhydrolase superfamily protein
AT1G54400	2.3422	HSP20-like chaperones superfamily protein
AT1G54575	2.2840	unknown protein
AT1G59730	4.8045	thioredoxin H-type 7
AT1G60060	2.1026	Serine/threonine-protein kinase WNK (With No Lysine)-related
AT1G60110	5.9145	Mannose-binding lectin superfamily protein
AT1G60130	4.8646	Mannose-binding lectin superfamily protein
AT1G60360	2.2175	RING/U-box superfamily protein
AT1G61275	3.6034	U12; snRNA
AT1G61280	2.1066	Phosphatidylinositol N-acetylglucosaminyltransferase, GPI19/PIG-P subunit
AT1G62290	2.4344	Saposin-like aspartyl protease family protein
AT1G62710	2.2148	beta vacuolar processing enzyme
AT1G64170	2.8923	cation/H+ exchanger 16
AT1G65500	2.1552	unknown protein
AT1G67100	5.7654	LOB domain-containing protein 40
AT1G67760	2.6158	TCP-1/cpn60 chaperonin family protein
AT1G68250	18.0169	unknown protein
AT1G68480	2.8827	C2H2 and C2HC zinc fingers superfamily protein
AT1G68880	2.2421	basic leucine-zipper 8
AT1G69310	2.6033	WRKY DNA-binding protein 57
AT1G69325	2.1934	Remorin family protein
AT1G69880	5.2980	thioredoxin H-type 8
AT1G70440	7.5271	similar to RCD one 3
AT1G71770	9.0325	poly(A)-binding protein 5
AT1G71890	3.2230	Major facilitator superfamily protein
AT1G72260	14.4135	thionin 2.1
AT1G73040	3.3427	Mannose-binding lectin superfamily protein
AT1G73120	6.4861	unknown protein
AT1G73190	2.8827	Aquaporin-like superfamily protein
AT1G74000	2.1751	strictosidine synthase 3
AT1G74010	3.0233	Calcium-dependent phosphotriesterase superfamily protein
AT1G74870	3.6034	RING/U-box superfamily protein
AT1G75430	3.6835	BEL1-like homeodomain 11
AT1G75830	5.4051	low-molecular-weight cysteine-rich 67
AT1G77960	9.8836	unknown protein
AT1G78000	2.1368	sulfate transporter 1;2
AT1G78206	2.2650	MIR775a; miRNA
AT1G79100	2.4709	arginine/serine-rich protein-related
AT1G80130	4.2325	Tetratricopeptide repeat (TPR)-like superfamily protein
AT2G01008	5.9317	unknown protein
AT2G01023	4.6844	unknown protein
AT2G04050	2.6206	MATE efflux family protein
AT2G16367	2.7226	pseudogene
AT2G17690	2.3517	F-box family protein with a domain of unknown function (DUF295)
AT2G18190	15.8549	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT2G18193	4.7987	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT2G18600	2.1810	Ubiquitin-conjugating enzyme family protein

ID	Relative levels (*taf6*/wt)	Description
AT2G21260	3.4833	NAD(P)-linked oxidoreductase superfamily protein
AT2G25000	2.1013	WRKY DNA-binding protein 60
AT2G27535	ND	ribosomal protein L10A family protein
AT2G27550	3.6152	centroradialis
AT2G28490	17.7767	RmlC-like cupins superfamily protein
AT2G31980	3.9993	PHYTOCYSTATIN 2
AT2G32487	2.6884	unknown protein
AT2G32490	3.3151	pseudogene
AT2G32660	3.3262	receptor like protein 22
AT2G32830	2.5624	phosphate transporter 1;5
AT2G34870	ND	hydroxyproline-rich glycoprotein family protein
AT2G35300	11.5308	Late embryogenesis abundant protein, group 1 protein
AT2G35570	12.4917	pseudogene
AT2G35950	3.0707	embryo sac development arrest 12
AT2G36110	9.1286	Polynucleotidyl transferase, ribonuclease H-like superfamily protein
AT2G36210	10.8102	SAUR-like auxin-responsive protein family
AT2G36610	2.6425	homeobox protein 22
AT2G37740	5.2850	zinc-finger protein 10
AT2G37870	3.2030	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT2G39030	2.4136	Acyl-CoA N-acyltransferases (NAT) superfamily protein
AT2G39320	3.1229	Cysteine proteinases superfamily protein
AT2G39330	3.4321	jacalin-related lectin 23
AT2G40170	20.1790	Stress induced protein
AT2G40200	4.8784	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT2G40970	2.1560	Homeodomain-like superfamily protein
AT2G41260	26.3047	glycine-rich protein / late embryogenesis abundant protein (M17)
AT2G41480	4.2946	Peroxidase superfamily protein
AT2G44195	4.3241	CBF1-interacting co-repressor CIR, N-terminal;Pre-mRNA splicing factor
AT2G44450	2.0909	beta glucosidase 15
AT2G46740	3.2219	D-arabinono-1,4-lactone oxidase family protein
AT2G46790	2.8186	pseudo-response regulator 9
AT2G46840	2.8827	DOMAIN OF UNKNOWN FUNCTION 724 4
AT2G46970	2.6534	phytochrome interacting factor 3-like 1
AT3G01600	5.8105	NAC domain containing protein 44
AT3G02240	6.2459	Encodes a root meristem growth factor (RGF
AT3G02480	3.2945	Late embryogenesis abundant protein (LEA) family protein
AT3G05150	2.4851	Major facilitator superfamily protein
AT3G06895	2.0179	unknown protein
AT3G06900	2.0179	U4.2; snRNA
AT3G08040	5.3482	MATE efflux family protein
AT3G08810	11.5308	Galactose oxidase/kelch repeat superfamily protein
AT3G08860	7.5500	PYRIMIDINE 4
AT3G08885	3.6835	pseudo-response regulator 9
AT3G09270	2.0920	glutathione S-transferase TAU 8
AT3G09450	2.8026	Fusaric acid resistance protein
AT3G09680	2.8827	Ribosomal protein S12/S23 family protein
AT3G11260	2.1277	WUSCHEL related homeobox 5
AT3G12320	2.3297	unknown protein
AT3G13090	2.9482	multidrug resistance-associated protein 8
AT3G13130	6.7263	unknown protein

ID	Relative levels (*taf6*/wt)	Description
AT3G15650	2.2264	alpha/beta-Hydrolases superfamily protein
AT3G15670	10.9131	Late embryogenesis abundant protein (LEA) family protein
AT3G15720	5.3389	Pectin lyase-like superfamily protein
AT3G16360	2.3565	HPT phosphotransmitter 4
AT3G16430	2.0277	jacalin-related lectin 31
AT3G17520	7.4688	Late embryogenesis abundant protein (LEA) family protein
AT3G18610	7.8427	nucleolin like 2
AT3G19350	2.2650	maternally expressed pab C-terminal
AT3G20210	2.3679	delta vacuolar processing enzyme
AT3G20370	2.4288	TRAF-like family protein
AT3G20470	2.2962	glycine-rich protein 5
AT3G21080	3.5004	ABC transporter-related
AT3G21090	2.5730	ABC-2 type transporter family protein
AT3G21351	5.6853	unknown protein
AT3G21370	8.8884	beta glucosidase 19
AT3G21720	82.7507	isocitrate lyase
AT3G22740	2.2858	homocysteine S-methyltransferase 3
AT3G27025	5.0023	unknown protein
AT3G27400	2.1472	Pectin lyase-like superfamily protein
AT3G27620	2.7226	alternative oxidase 1C
AT3G28100	2.0300	nodulin MtN21 /EamA-like transporter family protein
AT3G28570	5.5595	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT3G28580	4.5424	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT3G30730	3.7475	unknown protein
AT3G32920	5.2506	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT3G41761	16.9359	other RNA
AT3G41762	9.1286	unknown protein
AT3G43670	2.2894	Copper amine oxidase family protein
AT3G44300	3.8870	nitrilase 2
AT3G44790	3.7475	TRAF-like family protein
AT3G48700	5.9027	carboxyesterase 13
AT3G49580	3.7730	response to low sulfur 1
AT3G50450	4.6844	homolog of RPW8 1
AT3G50770	2.5750	calmodulin-like 41
AT3G51410	2.0965	Arabidopsis protein of unknown function (DUF241)
AT3G51860	2.5609	cation exchanger 3
AT3G53040	9.7291	late embryogenesis abundant protein, putative / LEA protein, putative
AT3G53980	2.0013	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT3G54940	50.4474	Papain family cysteine protease
AT3G55240	3.3774	Plant protein 1589 of unknown function
AT3G55290	2.2758	NAD(P)-binding Rossmann-fold superfamily protein
AT3G56080	2.1360	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT3G59190	2.8827	F-box/RNI-like superfamily protein
AT3G59340	6.1258	Eukaryotic protein of unknown function (DUF914)
AT3G59930	2.5944	Encodes a defensin-like (DEFL) family protein.
AT4G01335	2.4503	unknown protein
AT4G01430	2.7989	nodulin MtN21 /EamA-like transporter family protein
AT4G01920	2.2444	Cysteine/Histidine-rich C1 domain family protein
AT4G01930	2.2521	Cysteine/Histidine-rich C1 domain family protein

ID	Relative levels (*taf6*/wt)	Description
AT4G02330	2.0575	Plant invertase/pectin methylesterase inhibitor superfamily
AT4G02670	3.4432	indeterminate(ID)-domain 12
AT4G02700	3.0867	sulfate transporter 3;2
AT4G03060	2.3870	AOP2 (ALKENYL HYDROXALKYL PRODUCING 2)
AT4G03292	5.0928	Polynucleotidyl transferase, ribonuclease H-like superfamily protein
AT4G04810	4.9366	methionine sulfoxide reductase B4
AT4G04830	2.3106	methionine sulfoxide reductase B5
AT4G05370	8.6481	BCS1 AAA-type ATPase
AT4G05380	13.9331	P-loop containing nucleoside triphosphate hydrolases superfamily protein
AT4G06534	2.0591	unknown protein
AT4G08770	2.1740	Peroxidase superfamily protein
AT4G10265	2.0965	Wound-responsive family protein
AT4G10380	2.1876	NOD26-like intrinsic protein 5;1
AT4G11500	3.2945	pseudogene
AT4G12030	2.3872	bile acid transporter 5
AT4G12490	3.1322	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G12500	4.8211	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G12735	2.3283	unknown protein
AT4G12870	3.3965	Gamma interferon responsive lysosomal thiol (GILT) reductase family protein
AT4G12960	12.9722	Gamma interferon responsive lysosomal thiol (GILT) reductase family protein
AT4G14690	3.1681	Chlorophyll A-B binding family protein
AT4G15248	2.0965	B-box type zinc finger family protein
AT4G16240	6.0357	unknown protein
AT4G18490	2.5144	unknown protein
AT4G20690	3.1371	unknown protein
AT4G21020	ND	Late embryogenesis abundant protein (LEA) family protein
AT4G21926	2.6692	unknown protein
AT4G21930	3.2321	Protein of unknown function, DUF584
AT4G22020	3.5233	pseudogene
AT4G22390	2.4896	F-box associated ubiquitination effector family protein
AT4G22470	4.2776	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein
AT4G22960	4.0358	Protein of unknown function (DUF544)
AT4G24110	2.1909	unknown protein
AT4G24420	7.9274	RNA-binding (RRM/RBD/RNP motifs) family protein
AT4G24450	2.1478	phosphoglucan, water dikinase
AT4G25140	20.1790	oleosin 1
AT4G25480	2.6768	dehydration response element B1A
AT4G25580	8.2363	CAP160 protein
AT4G26260	2.3422	myo-inositol oxygenase 4
AT4G27140	73.5091	seed storage albumin 1
AT4G27150	85.0399	seed storage albumin 2
AT4G27160	ND	seed storage albumin 3
AT4G27170	ND	seed storage albumin 4
AT4G28040	2.6347	nodulin MtN21 /EamA-like transporter family protein
AT4G28520	36.0339	cruciferin 3
AT4G28790	2.3024	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT4G29770	3.4467	Target of trans acting-siR480/255.

ID	Relative levels (*taf6*/wt)	Description
AT4G31520	6.4861	SDA1 family protein
AT4G31640	3.1710	transcriptional factor B3 family protein
AT4G33070	4.0790	Thiamine pyrophosphate dependent pyruvate decarboxylase family protein
AT4G33550	6.6302	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT4G33560	2.3109	Wound-responsive family protein
AT4G34320	3.3066	Protein of unknown function (DUF677)
AT4G36060	2.2421	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT4G36600	12.9722	Late embryogenesis abundant (LEA) protein
AT4G36700	6.2158	RmlC-like cupins superfamily protein
AT4G36880	2.4983	cysteine proteinase1
AT4G38340	2.0668	Plant regulator RWP-RK family protein
AT4G38780	2.3638	Pre-mRNA-processing-splicing factor
AT4G39320	2.0076	microtubule-associated protein-related
AT4G39480	5.0680	cytochrome P450, family 96, subfamily A, polypeptide 9
AT4G39500	25.9444	cytochrome P450, family 96, subfamily A, polypeptide 11
AT5G01870	2.5005	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G02580	2.3586	Plant protein 1589 of unknown function
AT5G03010	8.6481	Galactose oxidase/kelch repeat superfamily protein
AT5G03860	8.5373	malate synthase
AT5G05340	2.0458	Peroxidase superfamily protein
AT5G05400	2.4823	LRR and NB-ARC domains-containing disease resistance protein
AT5G06250	2.3422	AP2/B3-like transcriptional factor family protein
AT5G06720	2.0044	peroxidase 2
AT5G06730	2.3740	Peroxidase superfamily protein
AT5G07700	2.4423	myb domain protein 76
AT5G08030	2.7386	PLC-like phosphodiesterases superfamily protein
AT5G09970	2.8172	cytochrome P450, family 78, subfamily A, polypeptide 7
AT5G10040	3.6034	unknown protein
AT5G10140	2.2013	K-box region and MADS-box transcription factor family protein
AT5G10340	6.1258	F-box family protein
AT5G10580	2.6425	Protein of unknown function, DUF599
AT5G11100	2.0719	Calcium-dependent lipid-binding (CaLB domain) family protein
AT5G11320	2.1260	Flavin-binding monooxygenase family protein
AT5G11410	5.4051	Protein kinase superfamily protein
AT5G13170	5.1248	senescence-associated gene 29
AT5G13330	4.0171	related to AP2 6l
AT5G14180	3.1312	Myzus persicae-induced lipase 1
AT5G14470	3.9637	GHMP kinase family protein
AT5G14490	ND	NAC domain containing protein 85
AT5G15120	2.3695	Protein of unknown function (DUF1637)
AT5G18270	2.3851	Arabidopsis NAC domain containing protein 87
AT5G21280	2.3367	hydroxyproline-rich glycoprotein family protein
AT5G23020	2.5342	2-isopropylmalate synthase 2
AT5G23810	3.3331	amino acid permease 7
AT5G24240	7.9930	Phosphatidylinositol 3- and 4-kinase ;Ubiquitin family protein
AT5G24280	2.1021	gamma-irradiation and mitomycin c induced 1
AT5G25230	2.0914	Ribosomal protein S5/Elongation factor G/III/V family protein
AT5G26280	2.5638	TRAF-like family protein
AT5G35940	5.3536	Mannose-binding lectin superfamily protein

ID	Relative levels (*taf6*/wt)	Description
AT5G39330	3.0629	Protein of unknown function (DUF1163)
AT5G39860	3.2671	basic helix-loop-helix (bHLH) DNA-binding family protein
AT5G39890	3.0228	Protein of unknown function (DUF1637)
AT5G40420	40.3579	oleosin 2
AT5G42840	2.3857	Cysteine/Histidine-rich C1 domain family protein
AT5G44120	19.7859	RmlC-like cupins superfamily protein
AT5G44980	8.3861	F-box/RNI-like/FBD-like domains-containing protein
AT5G45990	2.5131	crooked neck protein, putative / cell cycle protein, putative
AT5G46260	2.0291	disease resistance protein (TIR-NBS-LRR class) family
AT5G46590	2.4297	NAC domain containing protein 96
AT5G47170	3.1710	unknown protein
AT5G49200	10.5699	WD-40 repeat family protein / zfwd4 protein (ZFWD4)
AT5G49690	2.8827	UDP-Glycosyltransferase superfamily protein
AT5G49700	2.2265	Predicted AT-hook DNA-binding family protein
AT5G50760	2.0349	SAUR-like auxin-responsive protein family
AT5G51174	3.1229	SNOR30; snoRNA
AT5G51920	2.5694	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
AT5G52300	3.1422	CAP160 protein
AT5G53230	21.2600	Protein of unknown function (DUF295)
AT5G53240	ND	Protein of unknown function (DUF295)
AT5G53730	2.3496	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
AT5G53740	2.9586	unknown protein
AT5G53820	4.6844	Late embryogenesis abundant protein (LEA) family protein
AT5G54075	3.1710	U3D; snoRNA
AT5G54190	11.5863	protochlorophyllide oxidoreductase A
AT5G54740	ND	seed storage albumin 5
AT5G55270	10.3297	Protein of unknown function (DUF295)
AT5G55410	13.6929	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G59320	2.8273	lipid transfer protein 3
AT5G59390	2.4983	XH/XS domain-containing protein
AT5G59520	2.1044	ZRT/IRT-like protein 2
AT5G59590	4.6844	UDP-glucosyl transferase 76E2
AT5G60250	6.4861	zinc finger (C3HC4-type RING finger) family protein
AT5G61290	2.3346	Flavin-binding monooxygenase family protein
AT5G61890	3.5479	Integrase-type DNA-binding superfamily protein
AT5G64060	4.4936	NAC domain containing protein 103
AT5G64870	3.3278	SPFH/Band 7/PHB domain-containing membrane-associated protein family
AT5G66350	2.1690	Lateral root primordium (LRP) protein-related
AT5G66780	6.7263	unknown protein
AT5G66985	3.1830	unknown protein
AT5G67060	2.3532	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
AT1G20390	2.5358	transposable element gene
AT1TE22850	2.4956	ATLANTYS2
AT1TE22855	2.9728	ATLANTYS1
AT1TE80655	3.1229	ATREP1
AT1TE88535	4.6844	HELITRONY3
AT2G04135	4.4682	transposable element gene
AT2G04460	3.5313	transposable element gene

ID	Relative levels (*taf6*/wt)	Description
AT2G05280	3.0080	transposable element gene
AT2G05290	3.1448	transposable element gene
AT2G15810	2.6521	transposable element gene
AT2TE07145	3.5849	ATCOPIA95
AT2TE08840	2.9482	TA11
AT2TE08845	2.7226	TA11
AT2TE22875	14.4135	ATLANTYS3
AT2TE28020	2.4152	ATMU1
AT2TE41840	2.8827	ATLINE1A
AT2TE52200	2.4023	ATREP4
AT3G32925	5.9256	transposable element gene
AT3TE24450	2.9115	ATGP1
AT3TE29990	2.5173	ATGP1
AT3TE32785	2.2242	ATCOPIA5
AT3TE43450	2.4023	TAG2
AT3TE44530	3.7836	ATENSPM5
AT3TE54905	5.1889	ATLINEIII
AT3TE91870	4.6223	ATCOPIA13
AT4G04410	2.6141	transposable element gene
AT4TE02990	2.0965	ATREP11
AT4TE04425	10.0895	ATGP1
AT4TE04430	5.2850	ATGP1
AT4TE04665	2.3062	ATCOPIA20
AT4TE06880	3.4762	ATREP3
AT4TE10320	2.6768	ATCOPIA93
AT4TE22170	2.1620	ATHILA6A
AT4TE31690	2.1620	ATCOPIA56
AT4TE32815	5.0447	ATLINEIII
AT4TE35205	2.3062	ATREP4
AT4TE50935	3.9237	TAT1_ATH
AT4TE70265	2.1620	ATMU10
AT4TE88430	3.4593	ATREP10D
AT5G27845	32.6707	transposable element gene
AT5G28145	2.4023	transposable element gene
AT5TE02230	2.4323	VANDAL17
AT5TE36085	2.0179	ATLANTYS1
AT5TE36475	ND	ATCOPIA95
AT5TE36920	2.4152	ATCOPIA12
AT5TE48725	2.1453	ATIS112A
AT5TE78560	2.7517	VANDAL6

ND: The raw reads numbers in *LUCH* were zero but the raw reads and/or RPKM values in *LUCH taf6-1* pass the quality control for DEGs identification Although relative levels were not determined, these genes were considered as DEGs in *LUCH taf6-1*.

Table 2.5 Transcript levels of known genes in various gene silencing pathways.

AGI code	Name(s)	RPKM in *LUCH*	RPKM in *LUCH taf6-1*	Relative levels (*LUCH taf6-1* /*LUCH*)[1]
AT1G48410	AGO1	20.878	26.556	1.272
AT1G31280	AGO2	10.821	11.709	1.082
AT1G31290	AGO3[2]	0.742	2.868	3.863
AT2G27040	AGO4	12.683	14.213	1.121
AT2G27880	AGO5	0.099	0.143	1.441
AT2G32940	AGO6	1.145	1.289	1.126
AT1G69440	AGO7	4.613	6.247	1.354
AT5G21030	AGO8	0.082	0.098	1.201
AT5G21150	AGO9[3]	0.255	0.524	2.051
AT5G43810	AGO10	14.402	16.677	1.158
AT5G09790	ATXR5	7.546	6.501	0.861
AT5G24330	ATXR6	2.333	4.618	1.979
AT3G18730	BRU1	2.232	2.499	1.119
AT3G42670	CLSY1	8.529	8.738	1.024
AT1G80740	CMT1[3]	0.103	0.037	0.360
AT4G19020	CMT2	9.245	9.794	1.059
AT1G69770	CMT3	9.665	12.656	1.309
AT1G01040	DCL1	11.505	12.587	1.094
AT3G03300	DCL2	8.392	9.508	1.133
AT3G43920	DCL3	4.209	3.687	0.876
AT5G20320	DCL4	5.348	5.284	0.988
AT5G66750	DDM1	6.943	8.670	1.249
AT3G49250	DMS3	7.465	6.386	0.856
AT2G30280	DMS4/RDM4	7.437	7.272	0.978
AT5G25480	DNMT2	4.626	4.505	0.974
AT2G16390	DRD1	7.969	7.917	0.993
AT5G15380	DRM1	0.052	0.099	1.922
AT5G14620	DRM2	8.621	9.457	1.097
AT3G17310	DRM3	10.848	9.564	0.882
AT5G64630	FAS2	5.286	6.250	1.183
AT4G16280	FCA	6.906	7.704	1.115
AT2G43410	FPA	12.118	12.281	1.013
AT2G19520	FVE	17.029	19.219	1.129
AT5G63110	HDA6	13.470	14.799	1.099
AT4G20910	HEN1	5.500	6.274	1.141
AT3G07610	IBM1	7.037	6.695	0.951
AT3G48670	IDN2	15.882	18.145	1.143
AT1G15910	IDNL1	11.038	13.068	1.184
AT4G00380	IDNL2	4.438	6.540	1.474
AT5G04290	KTF1	11.081	12.361	1.116
AT5G49160	MET1	12.906	18.672	1.447
AT4G14140	MET2[3]	0.475	0.193	0.407
AT1G08060	MOM1	6.211	5.255	0.846

AGI code	Name(s)	RPKM in *LUCH*	RPKM in *LUCH taf6-1*	Relative levels (*LUCH taf6-1* /*LUCH*)[1]
AT4G36290	MORC1	7.597	6.290	0.828
AT1G19100	MORC6	4.492	3.954	0.880
AT2G16780	MSI2	15.261	18.765	1.230
AT4G35800	NRPB1	23.802	29.511	1.240
AT4G21710	NRPB2	14.816	18.443	1.245
AT2G15430	NRPB3/D3/E3a	15.039	17.676	1.175
AT5G09920	NRPB4	18.876	22.199	1.176
AT5G57980	NRPB5-like	1.662	1.797	1.081
AT3G22320	NRPB5/D5	25.813	31.227	1.210
AT2G41340	NRPB5b/E5b	9.643	9.694	1.005
AT5G51940	NRPB6a/D6a/E6a	18.792	21.143	1.125
AT2G04630	NRPB6b/E6b	18.214	21.396	1.175
AT5G59180	NRPB7	9.104	9.710	1.066
AT4G14520	NRPB7-like	8.178	7.208	0.881
AT1G54250	NRPB8a/E8a	8.079	11.166	1.382
AT3G59600	NRPB8b/D8b/E8b	14.766	17.016	1.152
AT3G16980	NRPB9a/D9a/E9a	1.631	1.856	1.138
AT4G16265	NRPB9b/D9b/E9b	5.121	4.589	0.896
AT1G61700	NRPB10-like[3]	0.045	0.000	0.000
AT1G11475	NRPB10/D10/E10	10.216	11.928	1.168
AT3G52090	NRPB11/D11/E11	19.405	23.848	1.229
AT1G53690	NRPB12-like[3]	0.932	0.122	0.131
AT5G41010	NRPB12/D12/E12	11.251	12.496	1.111
AT1G63020	NRPD1	2.897	2.842	0.981
AT3G23780	NRPD2/E2	8.607	8.355	0.971
AT2G15400	NRPD3B/E3B	1.484	2.452	1.652
AT4G15950	NRPD4/E4	11.275	10.477	0.929
AT3G22900	NRPD7	5.105	4.079	0.799
AT4G14660	NRPD7b/E7b	10.481	11.059	1.055
AT2G40030	NRPE1	5.023	4.160	0.828
AT3G57080	NRPE5	11.684	13.162	1.127
AT3G54490	NRPE5C	0.454	0.561	1.235
AT2G36490	ROS1[2]	17.557	8.148	0.464
AT5G04560	DME	13.861	13.525	0.976
AT3G10010	DML2	8.231	7.491	0.910
AT4G34060	DML3[2]	1.278	0.349	0.273
AT5G58130	ROS3	18.761	19.678	1.049
AT3G14890	ZDP	7.094	8.282	1.168
AT2G39740	HESO1	10.004	9.920	0.992
AT4G20400	JMJ14	13.505	14.923	1.105
AT5G59380	MBD6	9.582	9.048	0.944
AT3G22680	RDM1	17.980	16.883	0.939
AT1G14790	RDR1	7.294	7.488	1.027
AT4G11130	RDR2	6.054	6.726	1.111
AT3G49500	RDR6	7.028	6.155	0.876
AT3G48430	REF6	12.864	12.087	0.940

AGI code	Name(s)	RPKM in *LUCH*	RPKM in *LUCH taf6-1*	Relative levels (*LUCH taf6-1* /*LUCH*)[1]
AT2G24490	RPA2	6.298	7.587	1.205
AT1G05460	SDE3	9.534	10.057	1.055
AT3G15390	SDE5	4.228	4.402	1.041
AT4G15180	SDG2	10.464	10.268	0.981
AT1G77300	SDG8	10.205	9.503	0.931
AT3G50100	SDN1	6.717	4.072	0.606
AT5G05540	SDN2	13.029	12.943	0.993
AT5G67240	SDN3	16.064	14.794	0.921
AT5G23570	SGS3	9.371	10.506	1.121
AT1G15215	SHH1	7.068	7.760	1.098
AT5G04940	SUVH1	14.157	15.220	1.075
AT2G33290	SUVH2	9.886	8.244	0.834
AT1G73100	SUVH3	14.806	13.987	0.945
AT5G13960	SUVH4/KYP	5.835	4.941	0.847
AT2G35160	SUVH5	3.275	2.941	0.898
AT2G22740	SUVH6	10.043	8.050	0.802
AT2G24740	SUVH8	0.000	0.038	ND[4]
AT4G13460	SUVH9	19.853	18.410	0.927
AT2G05900	SUVH10	0.000	0.000	ND[4]
AT1G04050	SUVR1	2.778	1.909	0.687
AT5G43990	SUVR2	5.520	4.406	0.798
AT3G03750	SUVR3	6.163	5.748	0.933
AT3G04380	SUVR4	3.104	3.306	1.065
AT2G23740	SUVR5	11.870	11.239	0.947
AT3G49600	SUP32/UBP26	9.355	8.116	0.868
AT1G57820	VIM1	12.120	16.534	1.364
AT1G66050	VIM2	0.513	0.766	1.493
AT5G39550	VIM3	5.281	7.636	1.446

[1]Relative levels were calculated by dividing RPKM in *LUCH taf6-1* by RPKM in *LUCH*.
[2]Differentially expressed genes (DEGs) in *LUCH taf6-1* (2-fold change).
[3]Although expression levels in *LUCH taf6-1* were changed by more than 2-fold relative to those in *LUCH*, these genes were not considered as DEGs in *LUCH taf6-1*. Genes with less than one RPKM or ten raw reads number in both libraries were not considered for DEG identification.
[4]Relative levels were not determined because the raw reads numbers were zero in either or both plants.

Table 2.6 Only 37 differential small RNA regions (DSRs) were found in *LUCH taf6-1*.

Small RNA-seq was performed to profile small RNA populations in Col-0, *sde4-3* (a Pol IV mutant), *nrpe1-11* (a Pol V mutant), *LUCH* and *LUCH taf6-1*. Small RNAs were mapped to the *Arabidopsis* genome. The genome was divided into 500bp static windows and small RNAs were assigned to specific windows based on the location of the 5' nucleotide. The number of total reads in each window was counted and normalized as RPM (reads per million) in each genotype. To identify DSRs, *sde4-3* and *nrpe1-11* were compared to Col-0, and *LUCH taf6-1* was compare to *LUCH*. For comparison of *sde4-3* and *nrpe-1-11* to Col-0, windows with less than 10 TPM (transcripts per million) in Col-0 library were not considered for DSR identification. For comparison of *LUCH taf6-1* to *LUCH*, windows with less than 10 TPM in both libraries were not considered for DSR identification. Total 10729 and 13960 windows were identified in Col-0/*sde4-3* and Col-0/*nrpe1-11* comparison, respectively. Total 13577 windows were identified in the pair of *LUCH* and *LUCH taf6-1*. A fold change > 4 and an adjusted p-value (FDR) < 0.05 were required for DSRs between mutants and their respective controls. "Reduced" and "Increase" refer to DSRs with reduced and increased read count of small RNAs, respectively, in mutants.

	***sde4-3*/Col-0**	***nrpe1-11*/Col-0**	***LUCH taf6-1 /LUCH***
Reduced	8483	3524	20
Increased	50	385	17

Table 2.7 Hypomethylated regions in *LUCH taf6-1* in CHH contexts.

Location	*LUCH taf6-1*		*sde4-3*		*nrpe1-11*	
Chromosome#_nt	(A)	(B)	(A)	(B)	(A)	(B)
Chr1_372500	-0.1535	-0.1466	-0.2488	-0.2146	-0.2153	-0.2477
Chr1_1522900	-0.1627	-0.2052	-0.3423	-0.3504	-0.3455	-0.3409
Chr1_1882100	-0.2474	-0.2262	-0.5408	-0.6067	-0.6094	-0.5028
Chr1_3933200	-0.1208	-0.1292	-0.4521	-0.2099	-0.5833	-0.6023
Chr1_4094300	-0.1648	-0.1744	-0.6130	-0.5569	-0.5840	-0.5919
Chr1_4112700	-0.2365	-0.1516	-0.3856	-0.3936	-0.4005	-0.3906
Chr1_5198700	-0.1552	-0.1278	-0.4405	-0.4507	-0.4457	-0.4374
Chr1_5359100	-0.1496	-0.1015	-0.2408	-0.2934	-0.2885	-0.2461
Chr1_5531300	-0.1918	-0.2460	-0.5791	-0.5286	-0.5224	-0.5730
Chr1_5911700	-0.2011	-0.3002	-0.6231	-0.7239	-0.7173	-0.6160
Chr1_7282100	-0.2214	-0.2781	-0.5122	-0.5492	-0.5554	-0.5696
Chr1_7712200	-0.2742	-0.1112	-0.1900	-0.2274	-0.2271	-0.1661
Chr1_8238600	-0.2377	-0.3317	-0.2945	-0.3192	-0.3293	-0.3068
Chr1_8267400	-0.1764	-0.3256	-0.6542	-0.6924	-0.6919	-0.6413
Chr1_8539600	-0.2137	-0.2982	-0.4438	-0.2954	-0.3718	-0.4519
Chr1_8733400	-0.1839	-0.2478	-0.2706	-0.2377	-0.2025	-0.2690
Chr1_9001100	-0.1034	-0.1567	-0.1833	-0.0616	-0.0890	-0.1893
Chr1_9138200	-0.1486	-0.2463	-0.5234	-0.4315	-0.4328	-0.5525
Chr1_9587600	-0.1824	-0.1403	-0.3327	-0.2811	-0.2901	-0.3382
Chr1_9746400	-0.2329	-0.2045	-0.2833	-0.2698	-0.2770	-0.2813
Chr1_9746500	-0.2755	-0.2039	-0.3281	-0.3237	-0.3236	-0.3368
Chr1_10490700	-0.1259	-0.1635	-0.2528	-0.2738	-0.2871	-0.2406
Chr1_10696700	-0.2328	-0.2002	-0.1102	-0.2142	-0.2253	-0.1060
Chr1_10802500	-0.1577	-0.1771	-0.1672	-0.1432	-0.1362	-0.1790
Chr1_10814300	-0.2827	-0.2694	-0.5574	-0.6227	-0.6334	-0.5378
Chr1_11174600	-0.1470	-0.1798	-0.1837	-0.1543	-0.1506	-0.2080
Chr1_11661300	-0.1802	-0.2272	-0.3079	-0.1479	-0.1356	-0.2926
Chr1_12084300	-0.1365	-0.1656	-0.1929	-0.1293	-0.1475	-0.1958
Chr1_12097500	-0.1827	-0.1888	-0.2398	-0.2223	-0.2458	-0.1805
Chr1_12133200	-0.2124	-0.1687	-0.2257	-0.3014	-0.3037	-0.2292
Chr1_12330000	-0.1390	-0.1575	-0.1948	-0.2171	-0.2388	-0.1855
Chr1_12401800	-0.1527	-0.1163	-0.1043	-0.0519	-0.0361	-0.0674
Chr1_12561500	-0.2071	-0.2034	0.0003	0.0107	0.0038	-0.0005
Chr1_12594700	-0.1152	-0.1254	-0.2130	-0.2438	-0.2456	-0.2014
Chr1_12646000	-0.1640	-0.1456	-0.0708	-0.1536	-0.1559	-0.0590
Chr1_12662800	-0.2948	-0.2564	-0.2935	-0.5073	-0.5075	-0.3305
Chr1_13148400	-0.1386	-0.1554	-0.2027	-0.1668	-0.1606	-0.2084
Chr1_14164500	-0.1268	-0.1633	-0.4216	-0.4899	-0.4962	-0.4121
Chr1_15935600	-0.1878	-0.1346	-0.3923	-0.3235	-0.3460	-0.3923
Chr1_16449200	-0.1244	-0.1272	-0.1650	-0.0884	-0.1186	-0.1723
Chr1_16599500	-0.1377	-0.1111	-0.1791	-0.0847	-0.0802	-0.1696
Chr1_16712500	-0.1256	-0.1131	-0.2505	-0.2464	-0.2444	-0.2630
Chr1_16779500	-0.1405	-0.1048	-0.1959	-0.1685	-0.1611	-0.1959
Chr1_17335600	-0.1844	-0.2138	-0.5307	-0.4951	-0.5081	-0.5115
Chr1_17507100	-0.1320	-0.1136	-0.2274	-0.1826	-0.1876	-0.2290
Chr1_17766200	-0.1852	-0.1395	-0.2873	-0.3348	-0.3382	-0.2839
Chr1_17863100	-0.1671	-0.2005	-0.1976	-0.3184	-0.3197	-0.1943
Chr1_18320600	-0.1778	-0.1554	-0.6654	-0.7118	-0.7244	-0.6592

Location	*LUCH taf6-1*		*sde4-3*		*nrpe1-11*	
Chromosome#_nt	**(A)**	**(B)**	**(A)**	**(B)**	**(A)**	**(B)**
Chr1_19780800	-0.1511	-0.1267	-0.3686	-0.4506	-0.4552	-0.3748
Chr1_19944400	-0.1671	-0.1735	-0.4867	-0.5232	-0.5232	-0.5028
Chr1_20024300	-0.1947	-0.2153	-0.3609	-0.4625	-0.4594	-0.3742
Chr1_20363000	-0.2891	-0.1946	-0.6135	-0.6095	-0.6095	-0.6022
Chr1_20412300	-0.1550	-0.1792	-0.2751	-0.3158	-0.3095	-0.2784
Chr1_20525300	-0.1446	-0.1097	-0.1702	-0.2169	-0.2308	-0.1714
Chr1_20625000	-0.2751	-0.1865	-0.3111	-0.3580	-0.3785	-0.3049
Chr1_20982800	-0.1553	-0.1405	-0.0821	-0.1671	-0.1352	-0.0957
Chr1_21055100	-0.1541	-0.1682	-0.2500	-0.2548	-0.2480	-0.2397
Chr1_21105600	-0.2276	-0.2229	-0.3167	-0.4550	-0.4896	-0.3824
Chr1_21341500	-0.3830	-0.2605	-0.3408	-0.3117	-0.3479	-0.3754
Chr1_21342400	-0.1787	-0.1982	-0.6123	-0.4508	-0.4725	-0.6288
Chr1_21682400	-0.1411	-0.2379	-0.2799	-0.2807	-0.2739	-0.2426
Chr1_21741800	-0.1949	-0.2489	-0.4672	-0.4630	-0.4575	-0.4756
Chr1_22224800	-0.2520	-0.1872	-0.5616	-0.5489	-0.5412	-0.5510
Chr1_22526000	-0.1522	-0.1051	0.0133	0.0032	-0.0015	0.0017
Chr1_22694100	-0.1667	-0.1732	-0.5797	-0.5785	-0.5625	-0.5896
Chr1_22872200	-0.1835	-0.1884	-0.3145	-0.2551	-0.2552	-0.3009
Chr1_23197600	-0.1409	-0.1377	-0.3485	-0.3698	-0.3630	-0.3571
Chr1_23907200	-0.2777	-0.2087	-0.5153	-0.5186	-0.5655	-0.5092
Chr1_24488200	-0.2192	-0.1741	-0.2230	-0.3016	-0.3297	-0.2727
Chr1_24565400	-0.1014	-0.1372	-0.0631	-0.0505	-0.0510	-0.0577
Chr1_24605500	-0.1532	-0.2569	-0.3107	-0.3207	-0.4599	-0.5192
Chr1_24917700	-0.1756	-0.2717	-0.2799	-0.2828	-0.2711	-0.2704
Chr1_25401300	-0.2398	-0.1879	-0.4160	-0.4815	-0.5055	-0.4462
Chr1_26374900	-0.1197	-0.1384	-0.1924	-0.1613	-0.1646	-0.2048
Chr1_26722400	-0.1252	-0.2646	-0.3143	-0.4819	-0.5069	-0.4414
Chr1_27710600	-0.2807	-0.2117	-0.6059	-0.6741	-0.6752	-0.6014
Chr2_72700	-0.1663	-0.1708	-0.4280	-0.4300	-0.4338	-0.4270
Chr2_179800	-0.1970	-0.1710	-0.2999	-0.4170	-0.4060	-0.2993
Chr2_1078900	-0.2264	-0.1823	-0.4269	-0.4340	-0.4812	-0.4634
Chr2_1383700	-0.1888	-0.1922	-0.2789	-0.2866	-0.3403	-0.3231
Chr2_1459900	-0.1026	-0.1289	-0.1653	-0.1784	-0.1732	-0.1489
Chr2_1543100	-0.2087	-0.1649	-0.1475	-0.1857	-0.1944	-0.1520
Chr2_1982900	-0.1940	-0.1489	-0.2630	-0.2665	-0.2649	-0.2499
Chr2_2035100	-0.1376	-0.2217	-0.3061	-0.2886	-0.3036	-0.3084
Chr2_2240700	-0.1640	-0.1498	-0.1241	-0.1846	-0.1772	-0.1189
Chr2_2696400	-0.1702	-0.1951	-0.1349	-0.2576	-0.2513	-0.1537
Chr2_4301600	-0.1643	-0.2153	-0.4291	-0.4173	-0.4124	-0.4338
Chr2_4340600	-0.1626	-0.1949	-0.2262	-0.1429	-0.1670	-0.2333
Chr2_5215300	-0.1953	-0.2517	-0.1815	-0.2542	-0.2581	-0.1657
Chr2_5874200	-0.1336	-0.1629	-0.1408	-0.2060	-0.2058	-0.1470
Chr2_5923300	-0.3263	-0.2994	-0.2594	-0.3704	-0.3886	-0.2539
Chr2_6468700	-0.1500	-0.1970	-0.3129	-0.2621	-0.2630	-0.3121
Chr2_6523100	-0.1292	-0.1381	-0.1840	-0.2630	-0.2600	-0.1875
Chr2_6546200	-0.1176	-0.1909	-0.1899	-0.2045	-0.2045	-0.1453
Chr2_6677600	-0.2017	-0.1629	0.1036	-0.1309	-0.1538	-0.0292
Chr2_6935900	-0.2705	-0.1690	-0.0346	-0.0905	0.0242	0.0959
Chr2_7295200	-0.1688	-0.1152	-0.2109	-0.2234	-0.2279	-0.1997
Chr2_7538600	-0.1078	-0.1245	-0.1889	-0.1646	-0.1622	-0.1922
Chr2_7584000	-0.1530	-0.1340	-0.4217	-0.4257	-0.4162	-0.4245

Location	*LUCH taf6-1*		*sde4-3*		*nrpe1-11*	
Chromosome#_nt	**(A)**	**(B)**	**(A)**	**(B)**	**(A)**	**(B)**
Chr2_7875300	-0.2608	-0.1724	-0.3252	-0.3498	-0.3564	-0.3275
Chr2_7889700	-0.1518	-0.1644	-0.5648	-0.5973	-0.6166	-0.5974
Chr2_8198400	-0.2287	-0.1658	-0.4215	-0.4657	-0.4642	-0.4169
Chr2_8200100	-0.1316	-0.1695	-0.3176	-0.2986	-0.3018	-0.3262
Chr2_8371300	-0.1444	-0.1085	-0.0303	-0.0376	-0.0268	-0.0243
Chr2_8907700	-0.1199	-0.1396	-0.2143	-0.1752	-0.1812	-0.2323
Chr2_9432300	-0.2488	-0.1917	-0.0959	-0.0987	-0.1014	-0.1024
Chr2_9592000	-0.2144	-0.1318	-0.1915	-0.1600	-0.1171	-0.1886
Chr2_10682200	-0.2116	-0.1351	-0.2346	-0.1735	-0.1806	-0.2456
Chr2_10996600	-0.2288	-0.1524	-0.2307	-0.2218	-0.2208	-0.2261
Chr2_11056400	-0.2311	-0.1668	-0.2664	-0.3283	-0.3305	-0.2646
Chr2_11161000	-0.2594	-0.2376	-0.1864	-0.3809	-0.3216	-0.1164
Chr2_12320500	-0.1664	-0.1564	-0.2112	-0.2378	-0.2370	-0.2029
Chr2_12556100	-0.1649	-0.1471	-0.1408	-0.1690	-0.1671	-0.1408
Chr2_12725500	-0.1579	-0.1998	-0.5097	-0.5426	-0.5374	-0.5419
Chr2_13211500	-0.1974	-0.1887	-0.3082	-0.3181	-0.3138	-0.3075
Chr2_13634000	-0.2218	-0.1652	-0.3718	-0.3227	-0.3215	-0.3658
Chr2_14596600	-0.1562	-0.1850	-0.1992	-0.1564	-0.1467	-0.2110
Chr2_14682600	-0.3595	-0.2060	-0.2323	-0.3647	-0.3558	-0.2210
Chr2_16798600	-0.2224	-0.1631	-0.5485	-0.5022	-0.5238	-0.5908
Chr2_17418000	-0.2040	-0.2350	-0.5735	-0.4625	-0.5776	-0.6073
Chr2_19143600	-0.1174	-0.1320	-0.1296	-0.1808	-0.1807	-0.1239
Chr3_43000	-0.1116	-0.1613	-0.7172	-0.7338	-0.7406	-0.7275
Chr3_1057900	-0.2148	-0.2034	-0.4346	-0.4249	-0.4380	-0.4379
Chr3_4727900	-0.3247	-0.2790	-0.2756	-0.3057	-0.3057	-0.2329
Chr3_4753000	-0.1635	-0.1802	-0.3483	-0.4211	-0.4193	-0.3623
Chr3_4977800	-0.3309	-0.2207	-0.3347	-0.2753	-0.2850	-0.3383
Chr3_8752300	-0.1566	-0.1718	-0.3930	-0.3948	-0.4094	-0.4291
Chr3_8985700	-0.1431	-0.1937	-0.3684	-0.4428	-0.5040	-0.3711
Chr3_8985800	-0.3173	-0.3056	-0.5884	-0.5915	-0.6912	-0.6402
Chr3_9300400	-0.1269	-0.1118	-0.1471	-0.1467	-0.1439	-0.1416
Chr3_9459400	-0.2753	-0.1760	-0.3541	-0.5032	-0.4946	-0.3515
Chr3_9682100	-0.1895	-0.3661	-0.3900	-0.5257	-0.5223	-0.3900
Chr3_9769400	-0.1644	-0.2167	-0.4159	-0.5607	-0.5434	-0.4083
Chr3_10509900	-0.3082	-0.2264	-0.3416	-0.3422	-0.3455	-0.3247
Chr3_10542200	-0.3243	-0.1565	0.0015	-0.0021	-0.0027	0.0010
Chr3_10553800	-0.2122	-0.2061	-0.0247	0.0050	-0.0008	-0.0300
Chr3_10753300	-0.1613	-0.1643	-0.2962	-0.3739	-0.3836	-0.3069
Chr3_11089500	-0.1145	-0.1457	-0.1495	-0.1285	-0.1432	-0.1481
Chr3_11321600	-0.1616	-0.2342	-0.2525	-0.2147	-0.2060	-0.2924
Chr3_11374800	-0.2218	-0.2276	-0.1271	-0.3107	-0.3167	-0.1620
Chr3_11626100	-0.1617	-0.1412	-0.1979	-0.1524	-0.1615	-0.1973
Chr3_11657900	-0.2028	-0.1394	-0.4632	-0.4687	-0.5080	-0.4610
Chr3_11968400	-0.3245	-0.1849	-0.3088	-0.2881	-0.2954	-0.2643
Chr3_12126800	-0.1965	-0.1678	-0.3150	-0.3882	-0.4196	-0.3052
Chr3_13531100	-0.1139	-0.1272	-0.0243	-0.0685	-0.0768	-0.0417
Chr3_14677600	-0.2168	-0.1632	-0.2129	-0.2393	-0.2090	-0.2309
Chr3_15086900	-0.1857	-0.1900	-0.1337	-0.1908	-0.2217	-0.1960
Chr3_15797900	-0.1602	-0.1799	-0.4171	-0.4487	-0.4488	-0.4284
Chr3_15874100	-0.1270	-0.1419	-0.1874	-0.1712	-0.1595	-0.1853
Chr3_16079300	-0.2292	-0.1916	-0.2589	-0.1164	-0.1224	-0.2829

Location	*LUCH taf6-1*		*sde4-3*		*nrpe1-11*	
Chromosome#_nt	**(A)**	**(B)**	**(A)**	**(B)**	**(A)**	**(B)**
Chr3_16158600	-0.1339	-0.1146	0.0230	0.0103	-0.0637	-0.0576
Chr3_16459400	-0.1536	-0.1645	-0.2509	-0.1763	-0.2051	-0.2857
Chr3_17098700	-0.2010	-0.1376	-0.2451	-0.2609	-0.2596	-0.2352
Chr3_17516700	-0.3242	-0.2560	-0.5788	-0.5705	-0.5705	-0.5677
Chr3_17730500	-0.1950	-0.1919	-0.4403	-0.4663	-0.4776	-0.4526
Chr3_19243600	-0.1310	-0.1320	-0.2701	-0.2953	-0.6600	-0.5879
Chr3_20144200	-0.1566	-0.2129	-0.6381	-0.6148	-0.5974	-0.6318
Chr3_20260400	-0.2140	-0.2448	-0.5708	-0.5430	-0.5604	-0.5763
Chr3_21694700	-0.1879	-0.1541	-0.2557	-0.2936	-0.2984	-0.2567
Chr3_22119900	-0.1934	-0.2790	-0.1914	-0.4067	-0.4035	-0.2549
Chr3_23115700	-0.1646	-0.2183	-0.5995	-0.6647	-0.6863	-0.6197
Chr4_555500	-0.2300	-0.1654	-0.1910	-0.1758	-0.1758	-0.2000
Chr4_743400	-0.2856	-0.2372	-0.2348	-0.1878	-0.1792	-0.2313
Chr4_869900	-0.2010	-0.1609	-0.4824	-0.5390	-0.5408	-0.4933
Chr4_929600	-0.1579	-0.1512	-0.5703	-0.5281	-0.5209	-0.5730
Chr4_1292100	-0.1379	-0.1017	-0.0606	-0.1377	-0.1353	-0.0575
Chr4_1366000	-0.2430	-0.1656	-0.6280	-0.6734	-0.6758	-0.6221
Chr4_1882300	-0.1891	-0.1920	-0.3461	-0.3017	-0.3021	-0.3541
Chr4_1887100	-0.1529	-0.1436	-0.0881	-0.1198	-0.1330	-0.0981
Chr4_2169100	-0.3251	-0.2100	-0.4106	-0.4396	-0.4396	-0.4892
Chr4_2321700	-0.1284	-0.1751	-0.0949	-0.1301	-0.1364	-0.1000
Chr4_2393800	-0.1544	-0.1188	-0.1033	-0.1486	-0.1460	-0.1160
Chr4_2674300	-0.1790	-0.1798	-0.1247	-0.0174	-0.0326	-0.1267
Chr4_3360200	-0.1277	-0.1234	-0.1996	-0.2202	-0.2259	-0.1895
Chr4_4063000	-0.1293	-0.1367	-0.1136	-0.1128	-0.1123	-0.1074
Chr4_4339900	-0.1711	-0.1189	-0.2364	-0.1153	-0.1256	-0.2375
Chr4_4548100	-0.1593	-0.1635	-0.2895	-0.1886	-0.2263	-0.2959
Chr4_5194100	-0.1881	-0.2006	-0.2881	-0.3345	-0.3128	-0.3186
Chr4_5194200	-0.1453	-0.1419	-0.2305	-0.2570	-0.2773	-0.2476
Chr4_5456500	-0.1911	-0.2006	-0.4411	-0.3278	-0.5380	-0.5433
Chr4_5642500	-0.1837	-0.2332	-0.2877	-0.2354	-0.4225	-0.3618
Chr4_5670000	-0.1701	-0.1518	-0.2038	-0.1678	-0.1664	-0.2067
Chr4_6103600	-0.2471	-0.1631	-0.4216	-0.3983	-0.3914	-0.4372
Chr4_6140700	-0.1846	-0.1356	-0.2068	-0.3124	-0.2992	-0.1998
Chr4_6238700	-0.1505	-0.1147	-0.1619	-0.1081	-0.1337	-0.1562
Chr4_6333200	-0.1548	-0.1289	-0.1151	-0.1601	-0.1398	-0.1037
Chr4_6890500	-0.2257	-0.1537	-0.4026	-0.3872	-0.3920	-0.4127
Chr4_7139600	-0.1034	-0.2030	-0.1004	-0.1877	-0.1853	-0.1039
Chr4_7908600	-0.1503	-0.1348	-0.1574	-0.0629	-0.0855	-0.1772
Chr4_8141900	-0.1712	-0.2271	-0.2560	-0.2779	-0.2795	-0.2649
Chr4_8421500	-0.1601	-0.1783	-0.4259	-0.4054	-0.4321	-0.4193
Chr4_8881800	-0.1945	-0.1227	-0.2843	-0.2631	-0.2721	-0.2834
Chr4_8947600	-0.1733	-0.2048	-0.1811	-0.1762	-0.1750	-0.1855
Chr4_8960400	-0.1214	-0.1480	-0.3507	-0.3159	-0.3220	-0.3560
Chr4_8969000	-0.2583	-0.3007	0.0018	-0.0013	-0.0137	0.0077
Chr4_9079400	-0.1838	-0.2515	-0.3092	-0.2565	-0.2410	-0.3503
Chr4_9189100	-0.1646	-0.1234	-0.1869	-0.1956	-0.1911	-0.2093
Chr4_9981900	-0.3236	-0.2309	-0.0391	0.0000	0.0101	-0.0270
Chr4_10959500	-0.1677	-0.3558	-0.5263	-0.5333	-0.5333	-0.5172
Chr4_11414200	-0.1121	-0.1156	-0.1305	-0.1228	-0.1272	-0.1269
Chr4_11416000	-0.1849	-0.2569	-0.3943	-0.4096	-0.4052	-0.3960

Location	*LUCH taf6-1*		*sde4-3*		*nrpe1-11*	
Chromosome#_nt	**(A)**	**(B)**	**(A)**	**(B)**	**(A)**	**(B)**
Chr4_11699200	-0.1316	-0.1484	-0.0490	0.0019	-0.0266	-0.0541
Chr4_11699300	-0.1674	-0.1519	-0.0409	-0.0687	-0.0784	-0.0422
Chr4_11757300	-0.1195	-0.1643	-0.0999	-0.1799	-0.1844	-0.1053
Chr4_11932300	-0.1325	-0.1218	-0.0974	-0.0809	-0.0809	-0.1011
Chr4_12365500	-0.1922	-0.2053	-0.5632	-0.5395	-0.5852	-0.5741
Chr4_14045600	-0.1223	-0.1589	-0.6482	-0.6378	-0.6169	-0.6502
Chr4_14683800	-0.1408	-0.1340	-0.1563	-0.2543	-0.2528	-0.1633
Chr4_14911700	-0.1943	-0.1587	-0.4243	-0.4041	-0.4075	-0.4215
Chr4_16827500	-0.1306	-0.1618	-0.1147	-0.1085	-0.1092	-0.1087
Chr4_18182400	-0.1121	-0.1555	-0.1939	-0.1138	-0.1112	-0.1288
Chr5_113900	-0.2992	-0.2674	-0.1089	-0.1497	-0.1497	-0.1089
Chr5_676400	-0.2301	-0.3179	-0.0109	0.0000	0.0000	-0.0151
Chr5_766000	-0.1152	-0.1667	-0.7464	-0.7519	-0.7591	-0.7405
Chr5_1108700	-0.2562	-0.3105	-0.3296	-0.3172	-0.3196	-0.3483
Chr5_2348100	-0.1283	-0.1357	-0.3613	-0.2918	-0.3018	-0.4016
Chr5_2348300	-0.2778	-0.2026	-0.3146	-0.2935	-0.3179	-0.2498
Chr5_2560000	-0.1856	-0.1546	-0.2475	-0.2629	-0.2544	-0.2589
Chr5_2560100	-0.1631	-0.1443	-0.1286	-0.2288	-0.2348	-0.1412
Chr5_3715000	-0.1311	-0.1046	-0.1722	-0.2246	-0.2207	-0.1815
Chr5_3903100	-0.1212	-0.2842	-0.7745	-0.7869	-0.7830	-0.7778
Chr5_5155900	-0.3258	-0.2434	-0.3281	-0.4567	-0.4604	-0.3650
Chr5_5969900	-0.2031	-0.1854	-0.2629	-0.3749	-0.3769	-0.2689
Chr5_6361400	-0.1336	-0.1049	-0.0971	-0.1559	-0.1540	-0.0971
Chr5_7097700	-0.2294	-0.1825	-0.3875	-0.3550	-0.3542	-0.3874
Chr5_7416000	-0.1830	-0.1637	-0.4533	-0.4812	-0.4740	-0.4534
Chr5_7478500	-0.1936	-0.2357	-0.4167	-0.2963	-0.2850	-0.4066
Chr5_8179300	-0.1361	-0.1488	-0.2406	-0.2780	-0.2877	-0.2347
Chr5_8374700	-0.1151	-0.1293	-0.1039	-0.1078	-0.1310	-0.1150
Chr5_8399100	-0.2450	-0.2028	-0.2849	-0.2613	-0.2550	-0.3014
Chr5_8566000	-0.2943	-0.3254	-0.4153	-0.4325	-0.5327	-0.4515
Chr5_8611800	-0.1415	-0.1495	-0.3052	-0.3445	-0.3470	-0.3115
Chr5_8893300	-0.1732	-0.1963	-0.5697	-0.6132	-0.6116	-0.5887
Chr5_9309500	-0.3242	-0.3333	-0.2194	-0.3546	-0.3971	-0.3026
Chr5_10206000	-0.1421	-0.1748	-0.1997	-0.2952	-0.2929	-0.2138
Chr5_10214500	-0.1768	-0.2032	-0.4247	-0.5282	-0.6200	-0.5714
Chr5_10613200	-0.3075	-0.2451	-0.4207	-0.4536	-0.4562	-0.4182
Chr5_10653800	-0.1541	-0.1566	-0.2452	-0.1991	-0.2133	-0.2554
Chr5_10903300	-0.2335	-0.1507	-0.1916	-0.3820	-0.3508	-0.2379
Chr5_13617300	-0.1525	-0.2055	-0.1687	-0.1962	-0.1998	-0.1807
Chr5_13890800	-0.1025	-0.1063	-0.1711	-0.1020	-0.0901	-0.1747
Chr5_13921500	-0.1531	-0.1745	-0.1549	-0.1686	-0.1715	-0.1544
Chr5_13991400	-0.1628	-0.1279	-0.1048	-0.1096	-0.1540	-0.1180
Chr5_14079400	-0.2591	-0.3191	-0.3846	-0.3116	-0.3158	-0.3801
Chr5_14107900	-0.1160	-0.1160	-0.1313	-0.1217	-0.1194	-0.1310
Chr5_14560600	-0.1394	-0.1240	-0.0572	-0.0513	-0.0336	-0.0292
Chr5_14754300	-0.1316	-0.1294	-0.1501	-0.2473	-0.2639	-0.1556
Chr5_14847000	-0.1518	-0.1862	-0.2886	-0.3077	-0.3111	-0.2876
Chr5_15387300	-0.1245	-0.1055	-0.1431	-0.2103	-0.2137	-0.1473
Chr5_15841200	-0.2474	-0.1548	-0.3208	-0.2428	-0.2509	-0.3264
Chr5_16013400	-0.2588	-0.1020	-0.0038	-0.0093	0.0070	0.0077
Chr5_16357800	-0.2778	-0.4246	-0.3010	-0.3235	-0.3123	-0.3141

Location	*LUCH taf6-1*		*sde4-3*		*nrpe1-11*	
Chromosome#_nt	**(A)**	**(B)**	**(A)**	**(B)**	**(A)**	**(B)**
Chr5_16480800	-0.1610	-0.1632	-0.2274	-0.3034	-0.3004	-0.2314
Chr5_17327200	-0.1421	-0.1351	-0.1223	-0.0725	-0.0805	-0.1223
Chr5_17710700	-0.1538	-0.1348	-0.2332	-0.2714	-0.2665	-0.2470
Chr5_17833600	-0.1393	-0.1974	-0.4893	-0.4759	-0.4755	-0.4898
Chr5_19672700	-0.1879	-0.1873	-0.5812	-0.5060	-0.5034	-0.5812
Chr5_19952500	-0.2856	-0.1975	-0.7040	-0.5891	-0.6223	-0.7004
Chr5_20089800	-0.2916	-0.2245	-0.5518	-0.5403	-0.6136	-0.5730
Chr5_21028100	-0.2668	-0.2339	-0.3841	-0.4299	-0.4316	-0.3695
Chr5_21697200	-0.1348	-0.1802	-0.3439	-0.3059	-0.2955	-0.3553
Chr5_21776400	-0.2510	-0.3541	-0.4325	-0.4470	-0.4937	-0.4288
Chr5_22367000	-0.1794	-0.1172	-0.1357	-0.1458	-0.1445	-0.1448
Chr5_22408800	-0.1444	-0.2605	-0.5431	-0.5427	-0.5412	-0.5495
Chr5_22429900	-0.1288	-0.1266	-0.0916	-0.0556	-0.0606	-0.0808
Chr5_24118000	-0.1669	-0.1602	-0.3355	-0.4185	-0.4172	-0.3546
Chr5_24383100	-0.1796	-0.1671	-0.1725	-0.1204	-0.1205	-0.1666

The location of DMRs is displayed in the manner of “chromosome number_the position of first nucleotide”. The values indicate the methylation difference between mutants and respective wild-type plants. The larger absolute values, the less methylation levels in mutants relative to those in wild-type plants. The (A) and (B) indicate two independent biological replicates. All samples from the same biological replicates were processed at the same time and in the same manner.

Table 2.8 Oligonucleotide sequences used in this study.

Name	Oligonucleotide sequences	Purpose
LUCmF5	CTCCCCTCTCTAAGGAAGTCG	RT-PCR for LUC
LUCmR5	CCAGAATGTAGCCATCCATC	RT-PCR for LUC
N_UBQ5	GGTGCTAAGAAGAGGAAGAAT	RT-PCR, loading control
C_UBQ5	CTCCTTCTTTCTGGTAAACGT	RT-PCR, loading control
35Sf	CAAAGCAAGTGGATTGATGTGA	McrBC-PCR, southern blot
35Sr	TTTCCACGATGCTCCTCGT	Southern blot
LUC 0.13k R	TATGTGCATCTGTAAAAGCAA	McrBC-PCR
YZ 35S Bis F	ATTATTGTYGGTAGAGGTATTTTGAAYGATAGTT	Bisulfite sequencing
YZ LUC Bis R	CATCTATAAAAACAATTATTCCAAAAACCAAA	Bisulfite sequencing
Actin1-F	CCAAGCAGCATGAAGATCAA	McrBC-PCR, loading control
Actin1-R	TGAACAATCGATGGACCTGA	McrBC-PCR, loading control
TAF6-RT-F1	TGCTTGCTCCTGATGTTGAG	RT-PCR for TAF6
TAF6-RT-R1	AAATCGCGATGTCCAATAGC	RT-PCR for TAF6
TAF6-RT-F2	CACTGCTGGACCCAAAGAAG	RT-PCR for TAF6
TAF6-RT-R2	ATCTTTTCTGCGGTGATTGG	RT-PCR for TAF6
TAF6-proF1	ATGGTACCTCTCTTTGGCCCTGAACTGT	Cloning of *pTAF6:TAF6-GFP*
TAF6-fullR1	TAGAGCTCGAGGAATACTGACATCTCTG	Cloning of *pTAF6:TAF6-GFP*
AtMu1F1	CCGAGAACTGGTTGTGGTTT	RT-PCR, McrBC-qPCR
AtMu1R1	GCTCTTGCTTTGGTGATGGT	RT-PCR, McrBC-qPCR
ATCOPIA4F	CTCACTCAAGCTTCGGTTCC	RT-PCR
ATCOPIA4R	TGTTGGTGAAGGACCGTACA	RT-PCR
IGN5 A_F	CGCAGCGGAATTGACATCCTATC	RT-PCR
IGN5 A_R	TCGGAAAGAGACTCTCCGCTAGAAA	RT-PCR
AtLINE1-4F	CCGATGGTGACCAAGAGTTT	RT-PCR
AtLINE1-4R	TCAATGTCGGAGACCTCCTC	RT-PCR

Name	Oligonucleotide sequences	Purpose
soloLTR A221	ATCAATTATTATGTCATGTTAAAACCGATTG	RT-PCR, McrBC-PCR
soloLTR A222	TGTTTCGAGTTTTATTCTCTCTAGTCTTCATT	RT-PCR, McrBC-PCR
AtSN1 F1	ACCAACGTGCTGTTGGCCCAGTGGTAAATC	RT-PCR, McrBC-qPCR
AtSN1 R1	AAAATAAGTGGTGGTTGTACAAGC	RT-PCR, McrBC-qPCR
ATGP1F	TGGTTTTTCCTGTCCAGTTTG	RT-PCR, McrBC-qPCR
ATGP1R	AACAATCCTAACCGGGTTCC	RT-PCR, McrBC-qPCR
DDT2F	GAACCGATTCCAAAACAAACA	RT-PCR, McrBC-qPCR
DDT2R	AAGGTCACAGAAAATGAATACCG	RT-PCR, McrBC-qPCR
DDT8F	GAGGAGAGGCTTGTGAATCG	RT-PCR, McrBC-qPCR
DDT8R	GAGAGAATCAGGGACATCCAA	RT-PCR, McrBC-qPCR
DDT10F	ACTGTGAGTTGTATATGTTTTGTTGA	RT-PCR, McrBC-qPCR
DDT10R	AAAATATTATAACGGGTGAAAAATAGA	RT-PCR, McrBC-qPCR
IGN5-RT-R	CTGAGGTATTCCATAGCCCCTGATCC	RT of IGN5 (B region)
IGN5 realtime F1	ATGAAGAAAGCCCAAACCAT	PCR of IGN5 (B region)
IGN5 realtime R1	GCCGAATAACAGCAAGTCCT	PCR of IGN5 (B region)
IGN6-RT-R	TTTGTAATTCTCAGTTCGGGTATCTGCTTG	RT of IGN6 (B region)
IGN6 realtime F1	GCAAACATAGCAACCGAGAA	PCR of IGN6 (B region)
IGN6 realtime R1	GTTAGTGACGGCGAAAAAGC	PCR of IGN6 (B region)
IGN15-F	CCATAGCATAGAAACTTGGCGATATATGAA	RT-PCR (B region)
IGN15-R	CGGAAAAGGTAAGGTGGTTGGAAAA	RT-PCR (B region)
IGN23-F	ACTGAAAATTGTAAACAAAGAAACGGCACTACA	RT-PCR (B region)
IGN23-R	GATCGGTCCATAAACTTGTTGGGTTT	RT-PCR (B region)
IGN24-F	CGCATACGATGGTCGGAGAGTT	RT-PCR (B region)
IGN24-R	GCTTATCATTATCCAAACTTGATCCTATCCTAAA	RT-PCR (B region)
IGN25-F	CTTCTTATCGTGTTACATTGAGAACTCTTTCC	RT-PCR (B region)

Name	Oligonucleotide sequences	Purpose
IGN25-R	ATTCGTGTGGGCTTGGCCTCTT	RT-PCR (B region)
IGN26-F	CTCTTTCAGTGCGACAGCCTCAT	RT-PCR (B region)
IGN26-R	CGGCCAGGAAACCCTAACTTCC	RT-PCR (B region)
siR02F4	TGAAAACCACTCTACGTA	RT-PCR (B region)
siR02F4	CGATGGTGAAGAACATAT	RT-PCR (B region)
soloLTR B A217	CATATAACCGAAGCCGAAGGATGTGAAA	RT-PCR (B region)
soloLTR B A218	CAGAAACCTAAGGAACCATTACACGCTAAACC	RT-PCR (B region)
Taf6-1-I	GATTCTAAAACGAAATGGCGCTCTCTCTCTTTT GTATTCC	Cloning of *amiR:TAF6*
Taf6-1-II	GAGAGCGCCATTTCGTTTTAGAATCAAAGAGA ATCAATGA	Cloning of *amiR:TAF6*
Taf6-1-III	GAGAACGCCATTTCGATTTAGATTCACAGGTC GTGATATG	Cloning of *amiR:TAF6*
Taf6-1-IV	GAATCTAAATCGAAATGGCGTTCTCTACATAT ATATTCCT	Cloning of *amiR:TAF6*
Taf6-2-I	GATAATCTACACTCAGGTTCCGCTCTCTCTTTT GTATTCC	Cloning of *amiR:TAF6*
Taf6-2-II	GAGCGGAACCTGAGTGTAGATTATCAAAGAGA ATCAATGA	Cloning of *amiR:TAF6*
Taf6-2-III	GAGCAGAACCTGAGTCTAGATTTTCACAGGTC GTGATATG	Cloning of *amiR:TAF6*
Taf6-2-IV	GAAAATCTAGACTCAGGTTCTGCTCTACATATA TATTCCT	Cloning of *amiR:TAF6*
Taf6-3-I	GATAATATATACATTGCAACCGTTCTCTCTTTT GTATTCC	Cloning of *amiR:TAF6*
Taf6-3-II	GAACGGTTGCAATGTATATATTATCAAAGAGA ATCAATGA	Cloning of *amiR:TAF6*
Taf6-3-III	GAACAGTTGCAATGTTTATATTTTCACAGGTCG TGATATG	Cloning of *amiR:TAF6*
Taf6-3-IV	GAAAATATAAACATTGCAACTGTTCTACATATA TATTCCT	Cloning of *amiR:TAF6*
amiR-A	CTGCAAGGCGATTAAGTTGGGTAAC	Cloning of *amiR:TAF6*
amiR-B	GCGGATAACAATTTCACACAGGAAACAG	Cloning of *amiR:TAF6*

REFERENCES

1. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11: 204-220.

2. Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol 12: 483-492.

3. Huang L, Jones AME, Searle I, Patel K, Vogler H, et al. (2009) An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. Nat Struct Mol Biol 16: 91-93.

4. Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, et al. (2009) Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II. Molecular Cell 33: 192-203.

5. Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, et al. (2004) Genetic and Functional Diversification of Small RNA Pathways in Plants. PLoS Biol 2: e104.

6. Qi Y, Denli AM, Hannon GJ (2005) Biochemical Specialization within Arabidopsis RNA Silencing Pathways. Molecular Cell 19: 421-428.

7. Yu B, Yang Z, Li J, Minakhina S, Yang M, et al. (2005) Methylation as a Crucial Step in Plant microRNA Biogenesis. Science 307: 932-935.

8. Qi Y, He X, Wang X-J, Kohany O, Jurka J, et al. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature 443: 1008-1012.

9. Wierzbicki A, Haag J, Pikaard C (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135: 635 - 648.

10. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nature Genet 41: 630-634.

11. Zheng B, Wang Z, Li S, Yu B, Liu J-Y, et al. (2009) Intergenic transcription by RNA Polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes & Development.

12. Kim YJ, Zheng B, Yu Y, Won SY, Mo B, et al. (2011) The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. EMBO J 30: 814-822.

13. Smith LM, Pontes O, Searle I, Yelina N, Yousafzai FK, et al. (2007) An SNF2 Protein Associated with Nuclear RNA Silencing and the Spread of a Silencing Signal between Cells in Arabidopsis. The Plant Cell Online 19: 1507-1521.

14. Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE (2011) SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4, and Chromatin Remodeling Factors, Associate with RNA Polymerase IV. PLoS Genet 7: e1002195.

15. Law JA, Du J, Hale CJ, Feng S, Krajewski K, et al. (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498: 385-389.

16. Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu J-K, et al. (2010) A Protein Complex Required for Polymerase V Transcripts and RNA- Directed DNA Methylation in Arabidopsis. Current biology : CB 20: 951-956.

17. Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, et al. (2012) Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes & Development 26: 1825-1836.

18. Wittkopp PJ, Kalay G (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet 13: 59-69.

19. Goodrich JA, Tjian R (2010) Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet 11: 549-558.

20. Thomas MC, Chiang C-M (2006) The General Transcription Machinery and General Cofactors. Critical Reviews in Biochemistry and Molecular Biology 41: 105-178.

21. Won SY, Li S, Zheng B, Zhao Y, Li D, et al. (2012) Development of a luciferase-based reporter of transcriptional gene silencing that enables bidirectional mutant screening in Arabidopsis thaliana. Silence 3: 6.

22. Lago C, Clerici E, Dreni L, Horlow C, Caporali E, et al. (2005) The Arabidopsis TFIID factor AtTAF6 controls pollen tube growth. Developmental Biology 285: 91-100.

23. Becker C, Hagmann J, Muller J, Koenig D, Stegle O, et al. (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245-249.

24. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, et al. (2011) Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. Science 334: 369-373.

25. Stroud H, Greenberg Maxim VC, Feng S, Bernatavichute Yana V, Jacobsen Steven E (2013) Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell 152: 352-364.

26. Vaucheret H (2008) Plant ARGONAUTES. Trends Plant Sci 13: 350 - 358.

27. Mi S, Cai T, Hu Y, Chen Y, Hodges E, et al. (2008) Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5' Terminal Nucleotide. Cell 133: 116-127.

28. Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJM, et al. (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. EMBO J 25: 2828-2836.

29. Burke TW, Kadonaga JT (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes & Development 10: 711-724.

30. Molina C, Grotewold E (2005) Genome wide analysis of Arabidopsis core promoters. BMC Genomics 6: 25.

31. Lawit S, O'Grady K, Gurley W, Czarnecka-Verner E (2007) Yeast two-hybrid map of Arabidopsis TFIID. Plant Molecular Biology 64: 73-87.

32. Kadonaga JT (2002) The DPE, a core promoter element for transcription by RNA polymerase II. Exp Mol Med 34: 259-264.

33. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes & Development 18: 2368-2379.

34. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. Science 308: 118-120.

35. Pontier D (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev 19: 2030-2040.

36. Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, et al. (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet 37: 761-765.

37. Pontes O, Li CF, Nunes PC, Haag J, Ream T, et al. (2006) The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. Cell 126: 79-92.

38. Earley KW, Haag JR, Pontes O, Opper K, Juehne T, et al. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. The Plant Journal 45: 616-629.

39. Clough SJ, Bent AF (1998) Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. The Plant Journal 16: 735-743.

40. Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, et al. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. Journal of Bioscience and Bioengineering 104: 34-41.

41. Schwab R, Ossowski S, Riester M, Warthmann N, Weigel D (2006) Highly Specific Gene Silencing by Artificial MicroRNAs in Arabidopsis. The Plant Cell Online 18: 1121-1133.

42. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14: R36.

43. Lertpanyasampatha M, Gao L, Kongsawadworakul P, Viboonjun U, Chrestin H, et al. (2012) Genome-wide analysis of microRNAs in rubber tree (Hevea brasiliensis L.) using high-throughput sequencing. Planta 236: 437-445.

44. Li R, Yu C, Li Y, Lam T-W, Yiu S-M, et al. (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25: 1966-1967.

45. Audic S, Claverie J-M (1997) The Significance of Digital Gene Expression Profiles. Genome Research 7: 986-995.

46. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) 57: 289-300.

47. Rogers S, Bendich A (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Molecular Biology 5: 69-76.

48. Sambrook J, Fritsch, E.F., Maniatis, T., (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

49. Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen R, et al. (2008) Kismeth: Analyzer of plant methylation states through bisulfite sequencing. BMC Bioinformatics 9: 371.

50. Chen P-Y, Cokus S, Pellegrini M (2010) BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics 11: 203.

CHAPTER 3

An atypical HEAT SHOCK PROTEIN 20 homolog suppresses cytosine methylation and transcriptional gene silencing in *Arabidopsis*

ABSTRACT

DNA methylation is crucial for the transcriptional silencing of transposons. While positive feedback mechanisms help maintain DNA methylation, mechanisms that counteract hypermethylation are equally important but are less well understood. From genetic screens using two luciferase (LUC)-based cytosine methylation reporters, the gene *LOW IN LUCIFERASE EXPRESSION* (*LIL*), which encodes a HEAT SHOCK PROTEIN 20 (HSP20) homolog, was identified as a negative regulator of DNA methylation and transcriptional gene silencing (TGS). *LIL* was found to suppress the transcriptional silencing of the two *LUC* reporter transgenes as well as endogenously methylated targets. Moreover, *LIL* was found to decrease the methylation levels of one of the *LUC* reporters and endogenous targets. Additionally, a mutation in *LIL* led to the mislocalization of AGO4 within the nucleus. LIL was found to localize to the nucleus and physically interact with three methyl-CpG-binding domain (MBD) proteins, MBD5, MBD6 and MBD7. Taken together, these findings demonstrate that *LIL* encodes a molecular chaperone of MBD proteins and prevents hypermethylation and TGS.

.

INTRODUCTION

Epigenetic modifications represent a crucial regulatory mechanism that affects the level of transcription by controlling the state of chromatin [1,2]. As repressive epigenetic marks, cytosine methylation and histone H3 lysine 9 dimethylation (H3K9me2) are found in transposons and repeats and mediate transcriptional gene silencing (TGS) to maintain genomic integrity in plants [3]. Various epigenetic marks are precisely maintained by regulated deposition and removal, and by the crosstalk with other epigenetic marks.

In *Arabidopsis thaliana*, cytosine residues may be subjected to *de novo* methylation, maintenance methylation and demethylation (reviewed in [4]). The methylated targets are initially determined by RNA-directed DNA methylation (RdDM) with the guidance of small and long non-coding RNAs. The available evidence indicates that Pol IV generates long non-coding RNAs from these methylated targets, which are subsequently processed into 24-nucleotide (nt) small interfering RNAs (siRNAs). These siRNAs are then incorporated into ARGONAUTE4 (AGO4). In a manner independent of Pol IV activity, Pol V generates long non-coding RNAs from RdDM targets, and these nascent transcripts interact with the AGO4-loaded siRNAs. The interaction between these two types of non-coding RNAs and other subsidiary proteins recruits methyltransferase enzymes to the methylated targets, resulting in methylation specifically at the target regions. In plants, methylated cytosines occur in three sequence contexts, CG, CHG and CHH (where H indicates A, T or C), with specific regulatory signals and catalytic enzymes required for methylation maintenance at each of the three contexts. While CG and CHG require distinct maintenance mechanisms, CHH maintenance depends on the

continuous action of RdDM. Acting in reverse to these DNA methylation mechanisms, demethylation actively removes methylation marks to ensure proper levels of cytosine methylation. REPRESSOR OF SILENCING 1 (ROS1) and three homologs (DEMETER, DML2, and DML3) are responsible for this demethylation activity in *Arabidopsis*.

The regulation of DNA methylation is coordinated by epigenetic marks in histone proteins. The genome-wide distribution of H3K9me2 revealed significant overlap with repeats, siRNAs and cytosine methylation, which promotes chromatin condensation and TGS [5,6]. In the early phase of RdDM, SAWADEE HOMEODOMAIN HOMOLOG1 (SHH1) detects the methylation status of histones H3K4 and H3K9 and recruits Pol IV to a subset of methylated targets for the biogenesis of siRNAs and cytosine methylation [7]. For methylation maintenance, CHROMOMETHYLASE3 (CMT3) binds to nucleosomes harboring H3K9me2 and preferentially methylates cytosine in the CHG context [8]. In turn, KRYPTONITE (KYP) binds to methylated cytosine and mediates H3K9 methylation, establishing a positive feedback loop for heterochromatin [9]. For demethylation, INCREASED DNA METHYLATION1 (IDM1) recognizes both CG methylation and histone marks; the subsequent acetylation of H3 helps recruit ROS1 to the chromatin for the removal of methylated cytosine [10].

Methylated cytosine is recognized by downstream factors that induce subsequent events, including methyl-CpG-binding domain (MBD) proteins, which detect methylated cytosine. Among the 13 MBD proteins in *Arabidopsis*, MBD5, MBD6 and MBD7 have been found to bind methylated cytosine and to localize to hypermethylated chromocenters [11-14]. Additionally, MBD6 was co-purified with a histone deacetylation

activity and implicated in rRNA gene silencing in the context of nucleolar dominance [11,15]. In mammals, MBD proteins have been associated with histone deacetylases and methyltransferases involved in the establishment of repressive histone marks [16-19]. Together, these previous findings indicate that chromatin is transcriptionally silenced through the concerted action of cytosine methylation, histone modifications and MBD protein function.

In the present study, a gene whose protein product contains a HEAT SHOCK PROTEIN 20 (HSP20)-like chaperone domain, hereafter referred to as *LOW IN LUCIFERASE EXPRESSION* (*LIL*), was identified as a negative regulator of TGS and cytosine methylation. *LIL* was required to prevent TGS of two luciferase (LUC)-based TGS reporters and endogenous RdDM loci. A mutation in *LIL* increased the level of cytosine methylation at one of the LUC reporters, and at hundreds of endogenous loci as determined by whole genome bisulfite sequencing (BS-seq). While *LIL* was not found to affect Pol IV or Pol V activity, mutations in Pol V and AGO4 suppress the molecular defects of *lil* to a large extent, suggesting that LIL acts downstream of RdDM. Consistent with this notion, a mutation in *LIL* led to the mislocalization of AGO4 within the nucleus. LIL was found to physically interact with MBD5, MBD6 and MBD7. Furthermore, some regions that were hypermethylated in *lil* were also affected in *mbd* mutants in terms of cytosine methylation and TGS. These findings indicate that *LIL* negatively regulates cytosine methylation and TGS by mediating the function of MBD proteins.

RESULTS

Genetic screens using luciferase-based TGS reporters

To identify genes that prevent cytosine methylation and TGS in *Arabidopsis*, forward genetic screens were conducted using two LUC-based TGS reporters. As previously described, *LUCH* contains a transgene in which the *LUC*-coding sequence is fused to a miR172 binding site and is under the control of a dual cauliflower mosaic virus 35S promoter (*d35S*) [20]. Genetic and molecular analysis revealed that *LUC* expression in *LUCH* is impacted by RdDM and ROS1. While a mutation disrupting RdDM decreased the level of cytosine methylation at *d35S* and subsequently increased the expression of *LUC*, a mutation in *ROS1* induced the opposite effects.

The second LUC-based reporter used in the study, *YJ*, has not been previously reported. *YJ* also contains a *d35S*-driven *LUC* transgene but differs from *LUCH* in that the *LUC*-coding region in *YJ* is fused to a miR173 binding site rather than a miR172 binding site (Figure 3.1A). When a weak mutant of *ARGONAUTE1* (*AGO1*, the gene encoding the effector protein of miRNAs) was crossed into *YJ* (*YJ ago1-45*), LUC activity was not altered, suggesting that *LUC* in *YJ* is not under the regulation of the miRNA pathway (data not shown). *YJ* also differs from *LUCH* in that the basal level of *LUC* expression in *YJ* is much higher than that in *LUCH* (Figure 3.2A). *YJ* is a single-locus insertion into the gene At1g02740; this insertion did not cause any morphological defects. *YJ* is also in the *rdr6-11* background to prevent posttranscriptional silencing of the transgene.

We first evaluated genetically whether the expression of LUC in *YJ* is under the control of RdDM or demethylation. As mutations in IDM1 cause reduced LUC expression in *YJ* [10], we suspected that LUC in *YJ* is a target of ROS1-mediated dememthylation. Indeed, when *YJ* was crossed into the *ros1 dml2 dml3* background, *LUC* expression was extremely low (Figure 3.1B, C). The requirement for dememthylation genes in *LUC* expression suggests that *LUC* in *YJ* is subjected to methylation. However, real-time RT-PCR analysis revealed that mutations in RdDM genes *DRD1* and *NRPE1* did not de-repress LUC activity in *YJ*, while the *ago4-6* mutation slightly increased *LUC* transcript levels (Figure 3.1B, C). Based on these findings, we conclude that LUC in *YJ* is efficiently demethylated by ROS1 and its homologs and this ensures the high levels of LUC expression. The lack of a strong effect of RdDM mutations in the de-repression of this reporter could be due to the strong demethylation activity at this locus. In fact, bisulfite sequencing studies showed that *d35S* from LUC had little DNA methylation in *YJ* (see below).

LIL suppresses transcriptional silencing of the _LUC_ transgenes

From two independent genetic screens, namely, an ethyl methanesulfonate mutagenesis screen of *YJ* and a transfer-DNA (T-DNA) mutagenesis screen of *LUCH*, two mutants with reduced luciferase activity of the respective reporter lines were isolated (Figure 3.2A). The same gene was disrupted in the two mutants, and the two alleles isolated from the *YJ* and *LUCH* screens are hereafter referred to as *lil-1* and *lil-2 (LIL, LOW IN LUCIFERASE EXPRESSION*), respectively. Real-time RT-PCR confirmed that

LUC transcript levels were reduced in the mutants relative to the respective controls (Figure 3.2B). Map-based cloning and candidate gene sequencing revealed that *lil-1* was a G-to-A mutation disrupting the single splice acceptor site of At1g20870 and that *lil-2* was a C-to-T mutation that introduced a premature stop codon at the 532nd nt of the same gene (Figure 3.2D). Although *lil-2* was isolated from a T-DNA mutagenesis screen, the mutation was not caused by a T-DNA insertion. The gene is predicted to encode a protein containing an HSP20-like chaperone domain at the C-terminal region. Furthermore, there are three proteins homologous to LIL in *Arabidopsis* (Figure 3.3).

To determine whether the reduced *LUC* expression in the mutants was associated with increased DNA methylation, McrBC-PCR was performed to examine the state of methylation at *d35S* and *LUC*. Because McrBC specifically digests methylated DNA, reduced PCR amplification is expected of hypermethylated targets after McrBC treatment. The reduced amplification of *d35S* in the two mutants relative to their respective controls indicated that the mutations in *LIL* increased the methylation levels at *d35S* (Figure 3.2C). For the *LUC* coding region, an increase in the level of methylation was only observed in *LUCH lil-2*, as indicated by the lack of PCR amplification when McrBC-treated DNA from *LUCH lil-2* was used (Figure 3.2C). Whole-genome BS-seq confirmed the McrBC-PCR results. After the raw data were processed, the sequencing reads were mapped to both the original sequence and the C-to-T converted sequence of the *LUC* transgene for the determination of the methylation levels. Because the methylation density was extremely low throughout the entire transgene in both *YJ* and *YJ lil-1*, the analysis of *LUCH* and *LUCH lil-2* was prioritized. In *LUCH lil-2*, the

methylation density throughout the transgene was increased for all three sequence contexts compared to *LUCH* (Figure 3.4D). At *d35S*, the level of methylation was slightly increased for the CG and CHG contexts in *LUCH lil-2* (Figure 3.4E). However, this was not sufficient to alter the overall methylation state as cytosine in the CHH context is predominant in *d35S*. Inconsistencies between the McrBC-PCR and BS-seq analyses can be attributed to differences between the two methods and in the regions analyzed. For McrBC-PCR, methylation density was examined in the region encompassing a portion of *d35S* and 140 bp of the 5' coding region of *LUC* (Figure 3.1A). In contrast, methylation density was calculated for the entire *d35S* region in the BS-seq analysis. It is also important to note that the calculation of methylation density using the values from BS-seq is complicated by the presence of two copies of *d35S*. When the methylation density analysis was restricted to the region used for McrBC-PCR, the two methods yielded similar results. Based on these findings, the mutation in *LIL* increased the methylation levels of both *d35S* and/or *LUC*.

To evaluate whether increased cytosine methylation was responsible for the suppression of *LUC* expression in the *lil* mutants, the mutants and their respective controls were treated with the cytosine methylation inhibitor 5-aza-2'-deoxycytidine. After the chemical treatment, *LUC* expression in *YJ lil-1* and *LUCH lil-2* was similar to that in the respective controls, as shown by LUC live imaging of luciferase activity and real-time RT-PCR to examine *LUC* transcript levels (Figure 3.2A, B). Therefore, the role of *LIL* in the promotion of *LUC* expression was totally alleviated when the reporter genes

lacked DNA methylation. These findings indicate that *LIL* suppresses TGS of the two *LUC* reporters by reversing or preventing cytosine methylation.

LIL affects the expression of endogenous RdDM and ROS1 loci

To determine whether *LIL* also prevented TGS of endogenous RdDM targets, real-time RT-PCR was conducted to quantify the transcript levels from previously known RdDM loci. Mutations disrupting Pol V (*nrpe1-11* and *nrpe1-1*) resulted in the failure of TGS at the RdDM targets analyzed, as revealed by the de-repression of their transcript levels (Figure 3.5A). Although the impact of the two *lil* mutations on TGS was different, the four endogenous siRNA-generating regions analyzed were further repressed transcriptionally in *YJ lil-1* and *LUCH lil-2* relative to *YJ* and *LUCH*, respectively (Figure 3.5A). This finding indicated that *LIL* is required to prevent TGS at the examined RdDM loci.

Because the role of ROS1 and its homologs in counteracting DNA methylation and TGS is well established, the next question addressed was whether LIL was involved in preventing TGS at ROS1 targets. Hypermethylation in mutants disrupting *ROS1* is known to result in the suppression of expression *ROS1* targets. *At2g32160* and *At3g47300* are two known targets of ROS1-mediated demethylation [21]. The expression of these genes was greatly reduced in the *ros1 dml2 dml3* triple mutant (also referred to as *rdd*) (Figure 3.5B). At *At2g32160*, the *lil* mutations slightly enhanced TGS while they had no effects on *At3g47300* (Figure 3.5B). Although the analysis was limited to a few RdDM and ROS1 target loci, the requirement of *LIL* in promoting the expression of

RdDM target loci but not all ROS1 loci suggests that *LIL* acts as a negative regulator of TGS rather than a positive factor of ROS1-mediated demethylation.

Genetic interactions of *LIL* with factors involved in RdDM and demethylation

To further interrogate the relationship between LIL and RdDM or ROS1-mediated demethylation and investigate how *LIL* impacts TGS, genetic interaction analysis was performed by introducing *lil-1* into loss-of-function mutants of RdDM pathway genes and *ROS1*. We first combined *YJ lil-1* with *ros1-3 dml2-1 dml3-1* (*rdd*) to generate the quadruple mutant (referred to as *YJ lil rdd*). If LIL facilitates demethylation by ROS1 and its homologs, we would expect the quadruple mutant to behave identically to the *rdd* triple mutant. The transcript levels of the *YJ LUC* reporter and endogenous RdDM and ROS1 target loci were measured by qRT-PCR. The levels of *LUC* transcripts in the *rdd* triple mutant were extremely low and was not further reduced by the *lil-1* mutation (Figure 3.6A). However, transcript levels from endogenous RdDM or *ROS1* target loci in *YJ rdd* was further reduced by the *lil-1* mutation in *YJ lil rdd* (Figure 3.6B, C and E), which indicates that *LIL* and the *ROS1* homologs act in separate genetic pathways to prevent TGS. The lack of a difference in *LUC* expression between *YJ rdd* and *YJ lil rdd* could be due to the extremely low levels of expression in the *rdd* background. In fact, when *YJ lil* was compared to *YJ lil rdd*, an additive effect could be seen between *lil* and *rdd* mutations. Findings from these genetic analyses are consistent with LIL acting in parallel to ROS1-mediated demethylation.

We next combined *lil-1* with mutations in RdDM pathway genes *AGO4*, *DRD1*, and *NRPE1*. *YJ lil-1 ago4-6*, *YJ lil drd1-12*, and *YJ lil nrpe1-11* were generated and the transcript levels of *LUC* as well as endogenous RdDM target loci were analyzed. Although the extents of the effects varied among the RdDM pathway mutations *ago4*, *drd1* and *nrpe1*, these mutations caused de-repression of *LUC*, *soloLTR*, and *AtSN1* in the *YJ lil-1* background (Figure 3.6A, B, D). For example, *LUC* expression was much higher in *YJ lil ago4*, *YJ lil drd1* and *YJ lil nrpe1* as compared to *YJ lil*, suggesting that *LUC* was repressed by RdDM in *YJ lil-1*. The key question to be addressed here was whether *lil-1* had an effect at these loci when RdDM was compromised. For this, we compared *YJ ago4*, *YJ drd1*, and *YJ nrpe1* to *YJ lil ago4*, *YJ lil drd1*, and *YJ lil nrpe1*, respectively. In these RdDM pathway mutant backgrounds, in general, *lil-1* still caused a small reduction in the expression of *LUC* and endogenous RdDM targets (Figure 3.6A, B, C, D). This suggested that LIL acted, in part, in parallel to RdDM. However, it was evident from *LUC* and *AtSN1* (Figure 3.6A, D) that the effects of *lil-1* were much weaker when RdDM was compromised. For example, the ratio of *LUC* or *AtSN1* expression was much lower in *YJ lil* compared to *YJ* than in *YJ lil ago4* compared to *YJ ago4*. This suggested that the need for *LIL* to promote the expression of these loci was largely alleviated in the absence of RdDM. This is consistent with the notion that *LIL* acts as a negative factor in RdDM or acts downstream of RdDM to promote the expression of RdDM target genes.

Genome-wide impact of *LIL* on DNA methylation

To globally examine the role of *LIL* in cytosine methylation, BS-seq was performed for Col-0, *sde4-3*, *nrpe1-11*, *rdd*, *YJ*, *YJ lil-1*, *LUCH* and *LUCH lil-2*. Differentially methylated regions (DMRs) were identified by comparing the methylation density between the mutants and their corresponding controls. Because different cytosine sequence contexts are regulated by different mechanism, DMRs were identified for all three contexts (CG, CHG and CHH DMRs). Furthermore, DMRs were divided into two groups, increased DMRs corresponding to hypermethylated regions and reduced DMRs corresponding to hypomethylated regions in mutants relative to wild-type. The number of *LIL*-dependent DMRs in the three cytosine methylation contexts was as follows: 106 CG, 15 CHG and 134 CHH hypermethylated DMRs in *lil* and 38 CG, 4 CHG and 10 CHH hypomethylated DMRs in *lil* (Figure 3.7A). The higher frequency of hypermethylated DMRs rather than hypomethylated ones in *lil* indicates that LIL mainly acts as a negative regulator in DNA methylation, although the impact of LIL is mild, as revealed by the small number of DMRs. Among the six DMR classes, CG and CHH hypermethylated DMRs in *lil* were subjected to the further analysis to investigate the role of LIL in DNA (de)methylation, because *lil* mutations impacted specifically on the two DMRs (see below).

Some of the hypermethylated *LIL*-specific DMRs identified by BS-seq were verified by McrBC-qPCR analysis. *IDL* (increased DMRs in *lil*, or DMRs with increased DNA methylation in *lil*) *1*, *2*, *10* and *13* were located in transposons, and *IDL4*, *5*, *6* and *12* were located in protein-coding genes. The regions covered by *IDL3* and *9*

encompassed both transposon and protein-coding gene regions (Figure 3.8A). All of these loci were confirmed to be hypermethylated in *YJ lil-1* and *LUCH lil-2* relative to the respective controls (Figure 3.8B-D), indicating that *LIL* is required to prevent DNA methylation at the examined regions.

Relationship of LIL with RdDM and demethylation

The hypermethylated CG and CHH DMRs in *lil* were classified based on their genomic locations. For 106 CG DMRs, 21 and 6 DMRs were located in transposons/repeats and intergenic regions, respectively, while 79 DMRs (74%) were mapped to the gene bodies (Table 3.3, Figure 3.7F). However, all of CHH DMRs except one in gene body were located in transposons/repeats (116 loci, 86%) or intergenic regions (17 loci, 13%) (Table 3.3, Figure 3.7C). This indicates that LIL represses hypermethylation both in gene bodies at CG context and in transposons at CHH context.

Because siRNAs initially determines the methylated targets in RdDM, the next question addressed was whether LIL-targeted DMRs depend on siRNAs for methylation. For this, the distances between each LIL-dependent DMR and its closest small RNA-generating regions were calculated. Small RNA-generating regions were defined as the genomic regions in which small RNAs are reduced or depleted in *sde4-3* (a Pol IV mutant) relative to Col-0. After the distances from the small RNA regions were calculated for every DMRs and for randomly chosen genomic regions excluding DMRs in *lil*, the distances were compared between the two. It was found that CHH DMRs were significantly closely associated with the small RNA regions, compared to the randomly

chosen loci, as revealed by shorter distance of LIL-DMRs from small RNA regions (Figure 3.7B). More than half of CHH DMRs (76 loci, 56.7%) were observed to be located within 1 kb from the small RNA regions. This indicates LIL-targeted regions require small RNAs for methylation. However, CG DMRs did not exhibit the close association with the small RNA regions (Figure 3.7B).

Because LIL-targeted CHH DMRs are closely located from the small RNA regions and CHH cytosine is methylated by RdDM, the next question was whether LIL-targeted CHH DMRs are regulated by RdDM mechanism. Of the known RdDM factors, Pol IV and Pol V were used to examine the dependency of LIL-specific DMRs on RdDM for DNA methylation. When examined for the overlap with CHH hypomethylated DMRs in *sde4-3* (a Pol IV mutant) or *nrpe1-11* (a Pol V mutant), 72% CHH hypermethylated DMRs in *lil* were found to occur in Pol IV or Pol V DMRs (Figure 3.7D). This indicates that LIL represses CHH cytosine methylation in these specific RdDM-dependent loci, suggesting the negative action of LIL in RdDM. Consistent with the no association with small RNA regions, however, hypermethylated CG DMRs in *lil* were not represented in hypomethylated CG DMRs in *sde4-3* or *nrpe1-11* (Figure 3.7G).

Because ROS1 and its homologs also act as a negative regulator of DNA methylation in their targets, the next question addressed was whether LIL-dependent regions are demethylated by ROS1 and its homologs. The dependency of LIL-targeted DMRs on ROS1 families was examined by determining how LIL-target DMRs are over-represented in DMRs hypermethylated in *rdd* (*ros1 dml2 dml3*). Of the hypermethylated CG and CHH DMRs in *lil*, only 15 CHH (11%) and 25 CG (24%) DMRs were observed

to commonly increase CG and CHH methylation, respectively, in *rdd* (Figure 3.7E, H). This demonstrates that LIL represses cytosine methylation mainly in the genomic regions that are independent from ROS1-mediated demethylation, although LIL shares some targets with ROS1 to prevent hypermethylation.

LIL does not impact siRNA biogenesis or the production of Pol V-dependent transcripts

Global gene expression profiles were obtained for the *lil* mutants to determine whether *LIL* affects the expression of genes in gene silencing pathways. mRNA-seq (deep-sequencing of polyA-tailed RNA) was performed using 10-day-old *YJ*, *YJ lil-1*, *LUCH* and *LUCH lil-2* seedlings. Considering only the changes that were detected in both *YJ lil-1* and *LUCH lil-2*, the *lil* mutations induced no significant change in the transcripts profile (fold change > 2, adjusted p-value < 0.05) (Table 3.4). The differentially expressed genes in the two mutants did not include gene silencing pathway genes [22] or other related genes. Thus, the changes in cytosine methylation and TGS observed in the *lil* mutants cannot be attributed to altered transcript levels of genes known or implicated in gene silencing.

Pol IV and Pol V activity were examined in *lil-1* to determine if either of these mechanisms is affected by *LIL* in terms of its activity in reversing or preventing TGS of RdDM loci. Pol IV function was assessed genome-wide by high-throughput sequencing of small RNA populations. Small RNA libraries were constructed for Col-0, *sde4-3*, *nrpe1-11*, *YJ* and *YJ lil-1*, and the accumulation of siRNAs was analyzed. The small

RNA size distribution revealed that in *sde4-3*, the accumulation of 24-nt small RNAs decreased, while that of 21-nt small RNAs mostly mapped to miRNAs greatly increased (Figure 3.9A). In contrast to *sde4-3*, *YJ lil-1* showed a similar pattern to *YJ* with no shift in the length distribution of small RNAs (Figure 3.9A). A Pol IV mutation dramatically reduced the abundance of small RNAs observed in *sde4-3* in each small RNA-generating window. *YJ lil-1* also statistically significantly changed the small RNA abundance but the impact was milder than Pol IV (*sde4-3*) (Figure 3.9B). Only 15 and 45 regions showed reduction and increase of small RNA accumulation in *YJ lil-1*, while *8483* and 3524 regions were identified as differential small RNA regions (DSRs) in *sde4-3* and *nrpe1-11*, respectively, (Table 3.5). Taken together, these findings suggest that *LIL* does not affect the accumulation of siRNAs.

Whether *LIL* regulates the biogenesis of Pol V-dependent long non-coding transcripts was also addressed. RT-PCR was performed to measure the accumulation of transcripts from nine previously reported loci [23-25]. As expected, a mutation disrupting the largest Pol V subunit (*nrpe1-1*) resulted in a reduction or complete loss of transcripts. In *YJ lil-1* relative to *YJ*, the accumulation of Pol V-dependent transcripts was not affected at any of the tested loci (Figure 3.9C). Thus, *LIL* was not found to contribute to the production of Pol V-dependent transcripts.

Localization of AGO4 and LIL

At RdDM target loci, the nascent Pol V-dependent transcripts are recognized by AGO4-bound siRNAs, resulting in the recruitment of AGO4 and downstream factors to

the chromatin. In light of these downstream events of RdDM, the impact of *LIL* on the subnuclear localization of AGO4 was investigated. Previously reported immunolocalization analysis revealed that AGO4 exhibits a strong nucleolar dot signal and a diffuse nucleoplasmic signal in wild-type plants [26]. In *YJ lil-1*, AGO4 was found to be mis-localized within the nucleus. Multiple prominent AGO4 dots were present, with some apparently being close to the surface of the nucleolus (Figure 3.10A, middle panel of *lil-1*). In some nuclei, the AGO4 nucleolar dot was not present, and instead, one large AGO4 aggregate (Figure 3.10A, left panel of *lil-1*) or scattered AGO4 signal (Figure 3.10A, right panel of *lil-1*) was observed in the nucleoplasm in *YJ lil-1*. Although the nature of these AGO4 bodies needs to be further characterized and the occupancy of AGO4 at chromatin remains to be examined, this finding indicates that *LIL* contributes to the proper localization of AGO4 in the nucleus.

The subcellular localization of LIL was examined by transiently expressing *GFP*-fused *LIL* under the control of the *35S* promoter (*35S:LIL-GFP*) in tobacco leaves. Microscopic analysis revealed GFP signal in the nucleus (Figure 3.10B). Moreover, several discrete regions of GFP signal were observed, including one enlarged patch of strong GFP signal. It is possible that the expression of *LIL* under the *35S* promoter resulted in the formation of protein aggregates. Although neither the endogenous expression of LIL in *Arabidopsis* nor the nature of the protein body containing LIL has been characterized, the present findings indicate that LIL functions in the nucleus.

LIL is associated with MBD5, MBD6 and MBD7

To better understand the molecular functions of LIL, we performed a yeast two-hybrid screen to identify LIL-interacting proteins. We used full-length LIL as bait and an *Arabidopsis* cDNA library as prey. From this analysis, LIL was found to interact with three MBD proteins: MBD5, MBD6 and MBD7 (Figure 3.11B). When the full-length proteins and partial fragments were subjected to the yeast two-hybrid assay to map the interaction domain, interaction was only observed for full-length LIL. Although the final MBD domain of MBD7 weakly exhibited non-specific reporter activation, as observed in a few colonies for the BD and MBD7d3 pair, a greater survival of yeast cells was observed for the combination of MBD7d3 and LIL. This indicated that the final MBD7 MBD domain was capable of interacting with LIL. The interaction of MBD5 and MBD6 with LIL was further confirmed using a pull-down assay. GST-MBD5 and GST-MBD6 were expressed in *E. coli* and captured using glutathione-agarose beads. His-LIL was also expressed in *E. coli*, and the lysate was incubated with GST-MBD5-bound or GST-MBD6-bound beads and subjected to a pull-down assay. Western blot analysis revealed that GST-MBD5 and GST-MBD6 pulled down His-LIL (Figure 3.11C), while GST alone did not, indicating that LIL directly interacts with MBD5 and MBD6.

To examine the functional relationship between *LIL* and the three *MBD* genes in the context of DNA methylation and TGS, we obtained T-DNA insertion lines of *MBD5*, *MBD6* and *MBD7* from the Arabidopsis Biological Resource Center and GABI-Kat [27,28]. The locations of the T-DNA insertions in *mbd5* (SALK_111193), *mbd6* (SALK_043927) and *mbd7* (GABI_067_A09) are approximately 30 bp upstream of the

transcription start site of *MBD5*, in the first exon of *MBD6* and in the second exon of *MBD7*, respectively (Figure 3.12A). Transcripts were not detected from the *MBD6* and *MBD7* loci by RT-PCR in *mbd6* and *mbd7*, respectively (Figure 3.12B), indicating that the T-DNA insertions in these lines led to a loss of function. However, *MBD5* transcript levels were increased in *mbd5*, probably due to transcription initiation from the T-DNA and the accumulation of aberrant transcripts (Figure 3.12B). Although the function of these transcripts in *mbd5* was not addressed, *mbd5* and *mbd6* were subjected to further analysis.

To test whether MBD proteins are required for the regulation of *LIL*-dependent methylation at *LIL* targets, McrBC-qPCR was performed in *mbd5* and *mbd6*. The impact of *mbd5* and *mbd6* on methylation at 13 *LIL*-specific DMRs were generally less effective than that of *lil* (Figure 3.13). McrBC-qPCR revealed that decreased amplification (with a fold change greater than or equal to 1.5) was observed for four DMRs (*IDL1*, *3*, *6* and *12*) in both *mbd5* and *mbd6* (Figure 3.14A), indicating increased DNA methylation in the mutants. *IDL1* was located in transposon region and *IDL3* was encompassed by both transposon and protein-coding region, while *IDL6* and *IDL12* were located in protein-coding gene regions. For this reason, *IDL1* and *IDL3* were further examined to address whether hypermethylation of the two transposons enhanced TGS. RT-PCR using the *LUCH lil-2*, *YJ lil-1*, *mbd5* and *mbd6* lines revealed reduced transcript levels from *IDL1* in *LUCH lil-2*, *mbd5* and *mbd6* and reduced transcript levels from *IDL3* in *YJ lil-1* and *mbd6* (Figure 3.14B). These findings indicate that both LIL and MBD proteins prevent hypermethylation and TGS at these two transposons.

DISCUSSION

The self-reinforcing loop associated with repressive epigenetic marks also requires negative regulators to achieve a balanced chromatin state. In the present study, *LIL* was identified as a negative regulator of cytosine methylation and TGS. Mutations in *LIL* enhanced the TGS of *LUC* in two LUC-based reporter lines and the TGS of several endogenous RdDM targets. BS-seq uncovered hundreds of genomic regions that require the action of LIL to repress hypermethylation in both CG and CHH contexts. Particularly, most of CHH targets by LIL are closely associated with small RNA regions including transposons and methylated by RdDM pathway.

How does *LIL* function to suppress DNA methylation and TGS? Genetic interaction analysis of *LIL* with demethylation enzymes (*ROS1*, *DML2* and *DML3*) showed that *LIL* suppresses TGS in the separate genetic pathway of active demethylation. Although examined only in several loci, however, *LIL* was found to act as a negative factor in RdDM or act downstream of RdDM to prevent TGS of RdDM targets. Consistent with these genetic analyses, BS-seq revealed that LIL-dependent CHH DMRs were over-represented in Pol IV or Pol V-dependent DMRs but not in ROS1-dependent DMRs. In RdDM, *LIL* does not impact on the action of Pol IV and Pol V, while *LIL* impacts the proper localization of AGO4 in the nucleus.

As the RdDM effector protein, AGO4 were previously reported to colocalize with two nuclear bodies: Cajal bodies and AB-bodies [29]. Cajal bodies were originally identified as the nuclear body where protein-RNA complex assembly takes place, and they were hypothesized to be the site where Pol IV-dependent transcripts are processed

into siRNAs [26,30]. Independent from the Cajal bodies, AGO4 colocalizes with NRPD1b (NRPE1) to form AB-bodies (AGO4/NRPD1b bodies), where NRPD2 and DRM2 also colocalize [29]. Whether the AGO4 bodies in *YJ lil-1* correspond to Cajal bodies, AB-bodies, unidentified bodies or protein aggregates remains unclear. However, the unique AGO4 localization observed in *YJ lil-1* has not been reported in any RdDM mutants [29]. The only similar reported increase in the size of AGO4 bodies was the result of *COILIN-mRFP* overexpression and Cajal body enlargement [29]. The mRNA-seq analysis in the present study indicated no change in *COILIN* transcript levels in the *lil* mutants, indicating that the abnormal nuclear distribution of AGO4 in *YJ lil-1* was not attributable to changes in *COILIN* expression.

The following evidence supports the notion that *LIL* negatively impacts RdDM. First, the RdDM reporter LUC in *LUCH* and endogenous RdDM targets undergo hypermethylation and enhanced TGS in *lil* mutants. Second, the enhanced TGS of LUC in *LUCH* or *YJ* in *lil* mutants was completely suppressed by the DNA methylation inhibitor 5'-aza-deoxycytidine. Third, hundreds of endogenous loci undergo hypermethylation in *lil* mutants. Finally, AGO4 protein is mis-localized in *lil* mutants, and loss of function in AGO4 largely suppresses the effects of *lil*-1 in terms of the expression of LUC and endogenous RdDM loci. These findings suggest that *LIL* inhibits AGO4-dependent DNA methylation at RdDM targets to antagonize the repressive effects of RdDM.

Interestingly, three MBD proteins (MBD5, 6 and 7) were identified as the LIL-interacting proteins. Particularly, MBD6 was previously reported to predominantly

localize in perinucleolar chromocenter and promote rRNA gene silencing for nucleolar dominance in the manner of recognition of methylated target DNA and alteration of histone marks [15]. Because LIL contains a conserved α-crystallin HSP20 domain at C-terminal region, LIL probably assists the proper folding of MBDs and prevent the formation of protein aggregates as a molecular chaperone. If LIL really helps the action of MBD6, and MBD6 promotes gene silencing such as at the rRNA gene cluster, LIL would be a positive regulator for TGS. However, LIL was identified as a negative regulator for TGS and DNA methylation and even MBD6 was found to prevent hypermethylation and TGS albeit only at several regions. Does MBD6 differentially regulate gene silencing based on the genomic locations? A scenario consistent with existing data is that MBD6 promotes gene silencing at the rRNA gene cluster, while it suppresses TGS at other genomic regions.

If MBD6 prevents DNA methylation and TGS with the assistance of LIL, why does *mbd6* not affect cytosine methylation to the same extent as *lil*? It is possible that MBD5, MBD6, and MBD7 have overlapping functions and that cytosine methylation and TGS would be affected at more regions in double or triple mutants. Additionally, disrupting *LIL* function as a chaperone may affect other MBD proteins and consequently result in a more dramatic effect on DNA methylation than either the *mbd5* or *mbd6* single mutant.

How do MBDs prevent hypermethylation and TGS? Genome-wide analysis in mouse revealed that the occupancy of MBD proteins is mainly determined by methylation density [31]. In this manner, the degree of MBD recruitment may be a metric

of methylation density for downstream events in *Arabidopsis*. It is possible that MBD proteins protect methylated regions from further epigenetic modification by blocking the action from other enzymes. One interesting previous report is that *MBD6* knockdown slightly increased the rRNA-specific siRNA levels in *Arabidopsis suecica*, although it released the silencing of the rRNA genes [15]. One hypothesis is that MBD6 restricts the action of Pol IV or other RNA polymerases to generate siRNA precursors. If MBD6 does not occupy at these methylated loci, Pol IV would be recruited to the targets and generate siRNA precursors. Although the accumulation of siRNAs did not promote the silencing of the rRNA gene cluster in the MBD6 knockdown lines, other genomic loci could be affected by the positive feedback between siRNAs and DNA methylation.

MBDs could also influence ROS1-mediated demethylation. ROS1 is likely involved in limiting the expansion of DNA methylation at transposons into surrounding regions, as in *ros1* mutants, DNA methylation at transposons tends to occupy larger regions than in wild type [10,21]. Perhaps, the generation of non-coding RNAs as part of the RdDM process is prone to extend beyond the transposon region to result in methylation outside the region, and ROS1 is predicted to access only the outside or edge of the transposon regions for demethylation where the occupancy of MBD proteins would be lower. In a manner consistent with the hypermethylation observed in *mbd5*, *mbd6* and *lil*, LIL may help channel MBD proteins to the highly methylated regions and thereby enhance demethylation at less methylated regions.

The action of LIL on MBD would become more complicated with the action by the chromatin remodeler DDM1. DDM1 physically interacts with MBD6 and is required

for the localization of MBD6 to perinucleolar chromocenters [14]. Furthermore, the disruption of MET1 function in CG methylation maintenance is less disruptive than a mutation in *DDM1* in terms of the localization of MBD6; thus, DDM1 directly assists the recruitment of MBD6 to targets in addition to the indirect effect by cytosine methylation [14]. DDM1 probably transports MBD6 to the perinucleolar chromocenter and may also help alter the chromatin structure to expose methylated cytosine to MBD6. Along this same line, it is possible that LIL would compete or collaborate with DDM1 to properly localize MBDs to the genomic targets. This hypothesis could be tested by analyzing the nuclear localization of MBD proteins or the global occupancy of MBD proteins at chromatin in *lil*.

MATERIALS AND METHODS

Plant materials

Arabidopsis thaliana Columbia-0 (Col-0) ecotype plants were used in the present study. The LUC-based reporters *LUCH* [20] and *YJ* are in the *rdr6-11* mutant background [32]. *ago4-6* and *drd1-12* in *LUCH* [20] were introduced into *YJ* and *YJ lil-1*. *ago1-45* [33], *sde4-3* [34], *nrpe1-1* (also known as *drd3-1* [35]), *nrpe1-11* [36] and *ros1-3 dml2-1 dml3-1* [37] were previously described. *mbd5* (SALK_111193) and *mbd6* (SALK_043927) were obtained from the Arabidopsis Biological Resource Center [27], and *mbd7* (GABI_067_A09) was obtained from GABI-Kat [28]. *Nicotiana benthamiana* was used for the transient expression assay.

Plant growth conditions, 5-aza-2'-deoxycytidine treatment and luciferase live imaging

Seeds were surface-sterilized and planted on half-strength Murashige and Skoog (MS) media supplemented with 8% agar and 1% sucrose then stratified at 4°C for three days. Plants were grown in a growth incubator at 23°C under continuous light. Unless otherwise specified, 10-day-old seedlings were used for all of the experiments. For 5-aza-2'-deoxycytodine (5-aza-dC) (Sigma, Cat# A3656) treatment, plants were grown in MS media containing 8% agar, 1% sucrose and 7 μg/ml 5-aza-dC for 2 weeks. Luciferase live imaging was performed as previously described [20].

***LUCH* and *YJ* genetic screens**

Mutant populations in the *LUCH* and *YJ* backgrounds were generated via ethyl methanesulfonate (EMS) or T-DNA mutagenesis. For EMS mutagenesis, 1 ml of seeds (around 10,000 seeds) was washed with 0.1% Tween20 for 15 min, treated with 0.2% EMS for 12 hr and washed three times with 10 ml water for 1 hr with gentle agitation. For T-DNA mutagenesis, pEarleygate303 was modified to remove the Gateway cassette then transformed into the two *LUC* lines. TGS enhancers with lower LUC activity, based on LUC live imaging, were isolated in the M2 generation of the EMS population and the T2 generation of the T-DNA population. The isolated mutants were backcrossed to the respective parental lines two times before further analysis.

Mapping of the *lil-1* and *lil-2* mutations

The *YJ lil-1* and *LUCH lil-2* mutants were isolated from the *YJ* EMS screen and the *LUCH* T-DNA screen, respectively, and crossed to the corresponding *LUC* lines in the L*er* background to generate the F2 mapping populations. For *YJ lil-1*, 32 F2 plants with reduced LUC activity were used for rough mapping, and the mutation was linked to the center of the upper arm of chromosome 1. SSLP and dCAPS markers were designed using identified polymorphisms between the Col and L*er* accessions (http://arabidopsis.org/browse/Cereon/index.jsp). Fine mapping narrowed the region to a 160-kb window spanning the F5M15, F2D10 and F9H16 BAC clones. Candidate gene sequencing uncovered a G-to-A mutation in the splice acceptor site of At1g20870, and the mutation was subsequently referred to as *lil-1*. For *LUCH lil-2*, 27 F2 plants with reduced LUC activity were used for rough mapping, and linkage to the same mapping region of *lil-1* was observed. Sequencing of At1g20870 revealed a C-to-T mutation that introduced a premature stop codon in the first exon.

Plasmid construction

To generate *35S:LIL-GFP*, the *HSP20* coding region without the stop codon was amplified by PCR using the primer pair 35SHSP20-*F* and 35SHSP20-*R* and Col cDNA as the template. The PCR product was cloned into pENTR1A (Invitrogen, Cat# 11813011) in *Kpn*I and *Not*I sites and the resulting plasmid was recombined into pEarleyGate103 [38] using a Gateway LR Clonase kit (Invitrogenm Cat# 11791-019).

For yeast two-hybrid experiment, full-length or truncated *LIL* cDNAs were cloned into pGBKT7, and full-length or truncated *MBD5*, *MBD6* and *MBD7* cDNAs were cloned into pGADT7. For pull-down assay, GST-MBD5 and GST-MBD6 were kindly provided by Dr. Assaf Zemach.

RNA extraction and RT-PCR

RNA was extracted using TRI reagent (Molecular Research Center, Cat# TR118) and treated with DNaseI (Roche, Cat# 04716728001). cDNA was synthesized using RevertAid Reverse Transcriptase (Thermo Scientific, Cat# EP0441) and oligo-dT primer (Thermo Scientific, Cat# SO131). Quantitative RT-PCR was performed with three technical replicates on a Bio-Rad C1000 thermal cycler equipped with a CFX detection module using iQ SYBR Green Supermix (Bio-Rad, Cat# 170-0082). To detect Pol V-dependent transcripts, reverse transcription was performed using locus-specific primers and the SuperScript III kit (Invitrogen, Cat# 18080-051). PCR for this analysis was performed using Crimson *Taq* DNA polymerase (New England Biolabs, Cat# M0324). All experiments were conducted per the manufacturers' instructions. The primers used in the study are listed in Table 3.6.

Library construction for mRNA-seq, data processing and identification of differentially expressed genes (DEGs)

Libraries were generated using 2 μg of DNaseI-treated RNA and the TruSeq RNA Sample Preparation Kit v2 (Illumina, Cat# RS-122-2001) according to the manufacturer's

instructions. All high-throughput sequencing in this study was performed using an Illumina HiSeq 2000 system at the Genomics Core Facility at the University of California, Riverside. Image analysis and base calling were performed using the standard Illumina pipeline, version RTA 1.13.48. Only reads that passed the Illumina quality control steps were included in subsequent analyses, and reads with multiple copies were considered as a single read for the mapping procedure. The reads were mapped to the TAIR10 *Arabidopsis* genome using TopHat v2.0.4 with default settings [39]. Reads that mapped to multiple regions were discarded. After the number of reads mapped to each gene was counted, the RPKM (reads per kilobase per million) value of each gene was calculated using a Perl script. DEGs were identified, as previously described [40]. A fold change > 2 and FDR < 0.05 were used for the identification of DEGs and DEGs that were found in both *YJ lil-1* and *LUCH lil-2* were considered as LIL-specific DEGs.

Construction of small RNA libraries, data processing and identification of differential small RNA regions (DSRs)

Total RNA was size-fractionated by electrophoresis, and RNAs 15 to 40 nt in length were purified and subjected to the library construction procedure using the TruSeq Small RNA Sample Preparation Kit (Illumina, Cat# RS-200-0012) according to the manufacturer's instructions. Perl scripts were used to process the sequence reads, as previously described [41]. After the adapter sequences were trimmed from the reads, reads corresponding to rRNA, tRNA, snRNAs and snoRNAs were discarded. The subset of raw small RNA reads, including all remaining reads 20 to 24 nt in length, was aligned

to the TAIR10 *Arabidopsis* genome using SOAP2 [42]. The entire *Arabidopsis* genome was divided into 500 bp static and continuous windows, and reads were assigned to the window in which the 5' end was located. The number of reads was counted in each window and normalized as RPM (reads per million), which was used to indicate the abundance of small RNAs in any given window. For Col-0, *sde4-3* and *nrpe1-11* analysis, windows with fewer than 10 TPMs (transcripts per million) in Col-0 were discarded. For *YJ* and *YJ lil-1* analysis, windows with fewer than 10 TPMs (transcripts per million) in both plants were discarded. P-values were calculated using the Audic-Claverie method [43] and adjusted to the false discovery rate (FDR) [44]. A fold change > 4 and FDR < 0.05 were used for the identification of DSRs when comparing the RPM values of a given mutant and the respective control.

McrBC-PCR

Genomic DNA was extracted using the CTAB method [45]. Three units of McrBC (New England Biolabs, Cat# M0272) were used to treat 500 ng of DNA for 25 minutes at 37°C. A mock experiment was performed in parallel using DNA that had not been treated with McrBC. Either regular PCR or Quantitative PCR using iQ SYBR Green Supermix (Bio-Rad, Cat# 170-0082) was used to determine the methylation levels. For both regular PCR and qPCR, *ACTIN1*, which lacks cytosine methylation, was used as the internal loading control.

BS-seq library construction, data processing and identification of differentially methylation regions (DMRs)

Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Cat# 69104). One microgram of genomic DNA was sonicated into fragments 150 to 300 bp in length using a Diagenode Bioruptor, followed by purification with the PureLink PCR Purification Kit (Invitrogen, Cat# K3100-01). DNA ends were repaired using the End-It DNA End-Repair Kit (Epicentre, Cat# ER0720), and the DNA fragments were purified using the Agencourt AMPure XP-PCR Purification system (Beckman Coulter, Cat# A63880). The purified DNAs were adenylated at the 3' end using the polymerase activity of Klenow Fragment (3'→5' exo-) (New England Biolabs, Cat# M0212), followed by purification using the Agencourt AMPure XP-PCR Purification system. The methylated adapters in the TruSeq DNA Sample Preparation Kit (Illumina, Cat# FC-121-2001) were ligated to the DNA fragments using T4 DNA Ligase (New England Biolabs, Cat# M0202). After purification using AMPure XP beads, less than 400 ng DNA was subjected to bisulfite conversion using the MethylCode Bisulfite Conversion Kit (Invitrogen, Cat# MECOV-50). PfuTurbo Cx Hotstart DNA polymerase (Agilent, Cat# 600414) and the following PCR conditions were used for amplification: 95°C for 2 minutes; 9 cycles of 98°C for 15 seconds, 60°C for 30 seconds and 72°C for 4 minutes; and 72°C for 10 minutes. After purification using AMPure XP beads, the library was sequenced using HiSeq 2000 with the 101-cycle paired end sequencing mode (Illumina). Raw reads that failed to pass the Illumina quality control steps were discarded, and multi-copy reads were treated as a single copy for the mapping procedure. The filtered reads

were mapped to the TAIR10 *Arabidopsis* genome and a C-to-T converted genome using the BS-Seeker technique with default conditions [46]. Only perfectly and uniquely matched reads were included in subsequent analyses.

DMRs were identified as previously described [22] with some modifications. The *Arabidopsis* genome was divided into continuous and static 100-bp windows, and the methylation level in each window was calculated by dividing the number of methylated cytosine by the total number of cytosine. DMRs were determined by comparing methylation levels in the CG, CHG and CHH sequence contexts using cutoff values of 0.4, 0.2 and 0.1, respectively, between a mutant and its corresponding control, and an adjusted p-value (FDR) < 0.01 (Fisher's exact test). The windows with at least four cytosines that were sequenced by at least 4 reads in wild-type were selected for DMR identification. If closely located within 200 bp, multiple DMRs were merged into a single DMR. Hypervariability (HV) regions, which exhibit spontaneous fluctuations in methylation density, were discarded from the merged DMRs [47,48]. DMRs that were commonly found in *YJ lil-1* and *LUCH lil-2* were considered as LIL-specific DMRs. These final DMRs in *lil* were used to examine the overlap with DMRs in *sde4-3*, *nrpe1-11* and *rdd*.

Yeast two-hybrid screen and pull-down assays

The full-length coding sequence of *LIL* was cloned into the bait vector pGBKT7 at the *Nde*I and *Bam*HI sites then fused in-frame with the sequence encoding the GAL4 DNA-binding domain (BD). The *Arabidopsis* cDNA library cloned into the prey vector

pGADT7-RecAB was constructed by Clontech. All experiments using the yeast two-hybrid system were carried out according to the manufacturer's instructions (Clontech, Matchmaker GAL4 Two-Hybrid System 3 & Libraries User Manual, Cat# PT3247-1). The bait plasmid pGBKT7-LIL and the prey library DNA were co-transformed into the yeast strain AH109. The resulting progeny were first selected on SD/-Leu/-Trp/-His/-Ade plates then re-selected for β-galactosidase activity to eliminate false interactions. Positive clones harboring target cDNA were isolated, and the cDNA was sequenced to ensure that the sequences were accurate and had been fused in-frame with the sequence encoding the GAL4 AD domain. To identify the interaction domains, truncated *LIL* cDNAs were cloned into pGBKT7, and full-length or truncated *MBD5*, *MBD6* and *MBD7* cDNAs were cloned into pGADT7. Pairs of the constructed plasmids were co-transformed into yeast, and positive strains were used for the spot assay.

For pull-down assay, GST-MBD5, GST-MBD6 or GST alone were expressed in *E.coli* BL21 and purified using glutathione particles according to the manufacture's protocol (GE healthcare, Cat# 17-0756-01). Cellular extracts were prepared from *E.coli* BL21 containing the 6x His-HSP20 and an aliquot of the cell lysate was then incubated with the bound GST-MBD5, GST-MBD6 or GST respectively for 60 min at room temperature. The precipitates were washed six times with cell lysis buffer and resolved by SDS/PAGE. Anti-His antibody was used to detect His-HSP20. Anti-GST was used to detect GST-MBD5, GST-MBD6 and GST.

Immunolocalization

Subnuclear localization was examined as previously described [26].

Transient expression in *N. benthamiana*

Agrobacterium containing *35S:LIL-GFP* was cultured overnight at 28°C in 10 ml LB media with antibiotics. *Agrobacterium* cells were pelleted by centrifugation at 1000g for 10 minutes at room temperature and washed with infiltration media (10 mM $MgCl_2$) twice. Cells were resuspended in infiltration media at a final O.D. of 0.5-0.8. The resuspended *Agrobacterium* was infiltrated into *N. benthamiana* leaves. After 2 to 3 days, the expression of GFP-fused LIL was detected by a Leica SP5 confocal microscope. The excitation/emission wavelengths were 488nm/500-550nm for GFP.

FIGURES

Figure 3.1 Characterization of the *YJ* luciferase-based reporter.

(A) Structure of the T-DNA harboring *LUC* in *YJ*. RB and LB indicate the right and left borders of the T-DNA, respectively. The *dual 35S* promoter (*d35S*) drives the transcription of *LUC* and *NPTII*. A miR173 binding site is fused to the 3' end of *LUC*. The arrows indicate the direction of transcription. The lines underneath *d35S* indicate the regions examined by McrBC-PCR for the locus-specific methylation study.

(B) LUC live image of *YJ* and *YJ hsp20-1* combined with known mutations disrupting RdDM and demethylation. The left- and right-hand sides of the panel indicate *YJ* and *YJ hsp20-1* containing the indicated mutations in RdDM or demethylation, respectively. *rdd* represents the *ros1-3 dml2-1 dml3-1* triple mutant.

(C) Real-time RT-PCR to measure *LUC* transcript levels in *YJ*, *YJ ago4-6*, *YJ drd1-12*, *YJ nrpe1-11* and *YJ rdd*. The transcript levels were normalized using *UBIQUITIN5*, and each mutant was compared to *YJ*. Standard deviations were calculated from three technical replicates and are indicated by the error bars.

Figure 3.1

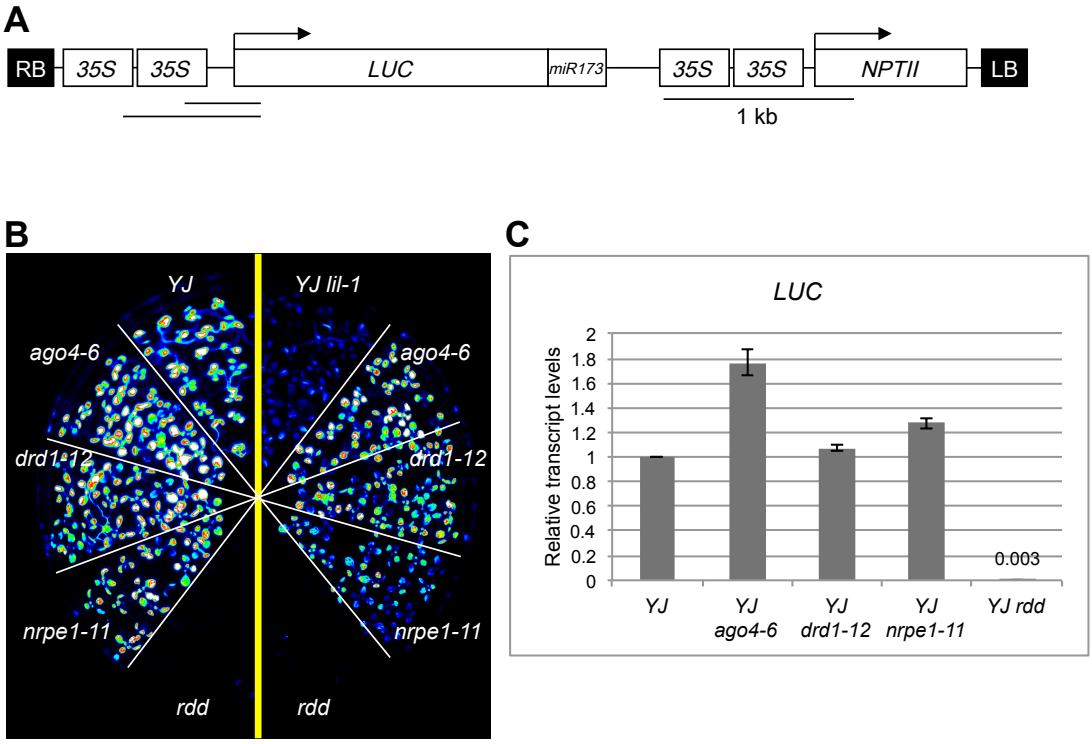

A
RB
35S
35S
LUC
miR173
35S
35S
NPTII
LB
1 kb
B
YJ
YJ lil-1
ago4-6
ago4-6
drd1-12
drd1-12
nrpe1-11
nrpe1-11
rdd
rdd
C
LUC
Relative transcript levels
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.003
YJ
YJ ago4-6
YJ drd1-12
YJ nrpe1-11
YJ rdd

Figure 3.2 Isolation of *lil* mutants as enhancers of TGS.

(A) Luciferase (LUC) luminescence in mock and 5-aza-2'-deoxycytidine (5-aza-dC)-treated *YJ*, *YJ lil-1*, *LUCH* and *LUCH lil-2*. Seeds were planted and grown on MS media with or without 5-aza-dC for 2 weeks, and a CCD camera was used for LUC live imaging. Each individual spot corresponds to an *Arabidopsis* seedling.

(B) Real-time RT-PCR to quantify the transcript levels of *LUC* in mock (top) and 5-aza-dC-treated (bottom) seedlings in (A). The transcript levels were normalized to *UBIQUITIN5* and compared to the corresponding control line. Error bars indicate the standard deviation from three technical replicates.

(C) Cytosine methylation analysis by McrBC-PCR. Genomic DNA was treated with McrBC, and PCR was performed to determine the methylation status of *d35S* and *LUC*. Because McrBC specifically digests methylated DNA, hypermethylation results in reduced PCR amplification. *LUC* is driven by two copies of *35S* in *YJ* and *LUCH* (Figure 3.1A), and as a result, two PCR bands were amplified for *d35S*. For *LUC*, the entire coding region was amplified. Mock experiments without McrBC treatment were performed in parallel. *ACTIN1*, which lacks methylation, was used as the internal loading control.

(D) The structure of the *LIL* gene and the positions of the mutations. The rectangles and the thin line represent the exons and intron of *LIL*, respectively. The sequence corresponding to the HSP20-like chaperone domain is indicated. In *lil-1*, a G-to-A mutation disrupted the splice acceptor site, and in *lil-2*, a C-to-T mutation introduced a premature stop codon in the first exon.

Figure 3.2

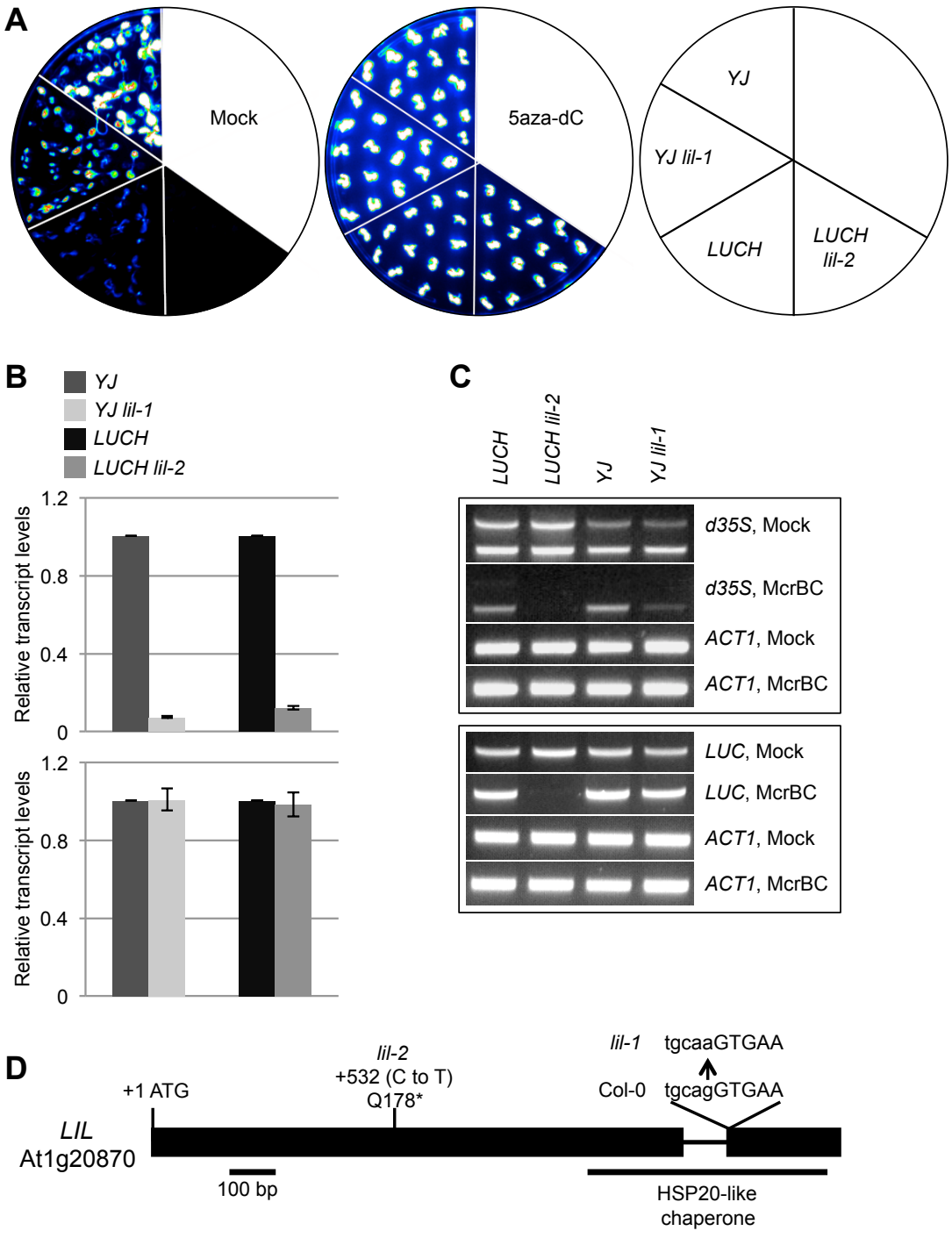

A
Mock
5aza-dC
YJ
YJ lil-1
LUCH
LUCH lil-2
B
YJ
YJ lil-1
LUCH
LUCH lil-2
Relative transcript levels
1.2
0.8
0.4
0
C
LUCH
LUCH lil-2
YJ
YJ lil-1
d35S, Mock
d35S, McrBC
ACT1, Mock
ACT1, McrBC
LUC, Mock
LUC, McrBC
ACT1, Mock
ACT1, McrBC
D
LIL
At1g20870
+1 ATG
lil-2
+532 (C to T)
Q178*
lil-1 tgcaaGTGAA
Col-0 tgcagGTGAA
100 bp
HSP20-like
chaperone

Figure 3.3 Sequence alignment of LIL and its homologous proteins.

The LIL homologs were identified using the *Arabidopsis thaliana* WU-BLAST2 Search function, the TAIR10 Proteins dataset and the full-length amino-acid sequence of LIL as a query (http://www.arabidopsis.org). The protein sequences were aligned using ClustalW2 (http://embnet.vital-it.ch/software/ClustalW.html) and displayed using Boxshade (http://embnet.vital-it.ch/software/BOX_form.html). Black and grey boxes indicate identical and similar residues, respectively. The region encoding the HSP20-like chaperone domain is indicated by the red line. The consensus sequences are indicated underneath the protein alignment.

Figure 3.3

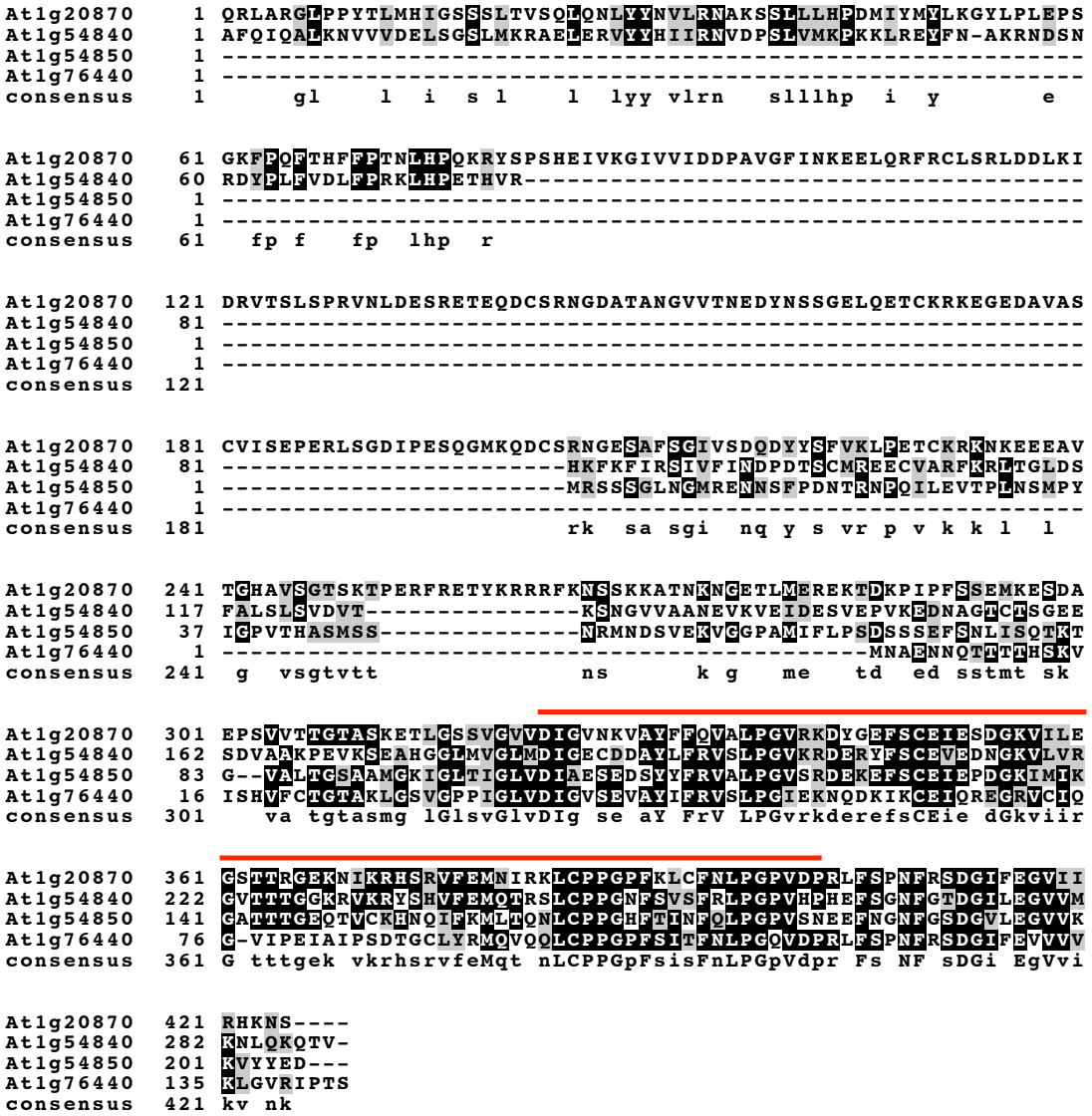

At1g20870 1 QRLARGLPPYTLMHIGSSSLTVSQLQNLYYNVLRNAKSSLLLHPDMIYMYLKGYLPLEPS
At1g54840 1 AFQIQALKNVVVDELSGSLMKRAELERVYYHIIRNVDPSLVMKPKKLREYFN-AKRNDSN
At1g54850 1 ----------
At1g76440 1 ----------
consensus 1 gl l i s l l lyy vlrn slllhp i y e
At1g20870 61 GKFPQFTHFFPTNLHPQKRYSPSHEIVKGIVVIDDPAVGFINKEELQRFRCLSRLDDLKI
At1g54840 60 RDYPLFVDLFPRKLHPETHVR----------
At1g54850 1 ----------
At1g76440 1 ----------
consensus 61 fp f fp lhp r
At1g20870 121 DRVTSLSPRVNLDESRETEQDCSRNGDATANGVVTNEDYNSSGELQETCKRKEGEDAVAS
At1g54840 81 ----------
At1g54850 1 ----------
At1g76440 1 ----------
consensus 121
At1g20870 181 CVISEPERLSGDIPESQGMKQDCSRNGESAFSGIVSDQDYYSFVKLPETCKRKNKEEEAV
At1g54840 81 ----------HKFKFIRSIVFINDPDTSCMREECVARFKRLTGLDS
At1g54850 1 ----------MRSSSGLNGMRENNSFPDNTRNPQILEVTPLNSMPY
At1g76440 1 ----------
consensus 181 rk sa sgi nq y s vr p v k k l l
At1g20870 241 TGHAVSGTSKTPERFRETYKRRRFKNSSKKATNKNGETLMEREKTDKPIPFSSEMKESDA
At1g54840 117 FALSLSVDVT----------KSNGVVAANEVKVEIDESVEPVKEDNAGTCTSGEE
At1g54850 37 IGPVTHASMSS----------NRMNDSVEKVGGPAMIFLPSDSSSEFSNLISQTKT
At1g76440 1 ----------MNAENNQTTTTHSKV
consensus 241 g vsgtvtt ns k g me td ed sstmt sk
At1g20870 301 EPSVVTTGTASKETLGSSVGVVDIGVNKVAYFFQVALPGVRKDYGEFSCEIESDGKVILE
At1g54840 162 SDVAAKPEVKSEAHGGLMVGLMDIGECDDAYLFRVSLPGVKRDERYFSCEVEDNGKVLVR
At1g54850 83 G--VALTGSAAMGKIGLTIGLVDIAESEDSYYFRVALPGVSRDEKEFSCEIEPDGKIMLK
At1g76440 16 ISHVFCTGTAKLGSVGPPIGLVDIGVSEVAYIFRVSLPGIEKNQDKIKCEIQREGRVCIQ
consensus 301 va tgtasmg lGlsvGlvDIg se aY FrV LPGvrkderefsCEie dGkviir
At1g20870 361 GSTTRGEKNIKRHSRVFEMNIRKLCPPGPFKLCFNLPGPVDPRLFSPNFRSDGIFEGVII
At1g54840 222 GVTTTGGKRVKRYSHVFEMQTRSLCPPGNFSVSFRLPGPVHPHEFSGNFGTDGILEGVVM
At1g54850 141 GATTTGEQTVCKHNQIFKMLTQNLCPPGHFTINFQLPGPVSNEEFNGNFGSDGVLEGVVK
At1g76440 76 G-VIPEIAIPSDTGCLYRMQVQQLCPPGPFSITFNLPGQVDPRLFSPNFRSDGIFEVVVV
consensus 361 G tttgek vkrhsrvfeMqt nLCPPGpFsisFnLPGpVdpr Fs NF sDGi EgVvi
At1g20870 421 RHKNS----
At1g54840 282 KNLQKQTV-
At1g54850 201 KVYYED---
At1g76440 135 KLGVRIPTS
consensus 421 kv nk

Figure 3.4 Cytosine methylation density at the *LUCH*-containing transgene.

(A) to (C) Methylation level at each cytosine in the transgene. The sequencing reads from BS-seq were mapped to the transgene sequence as well as the C-to-T converted sequence, and the methylation levels at each cytosine were calculated. The methylation levels are plotted along the transgene for the CG (A), CHG (B) and CHH (C) sequence contexts. The regions corresponding to *d35S*, *LUC*, partial *AP2* and *NPTII* are indicated.
(D) to (F) Methylation density in the entire T-DNA (D), *d35S* for *LUC* (E) and the *LUC*-coding region (F). The number of methylated cytosines was divided by the total number of cytosines in all sequence contexts, in the CG context, in the CHG context and in the CHH context. *LUCH* and *LUCH lil-2* are shown in black and red, respectively.

Figure 3.4

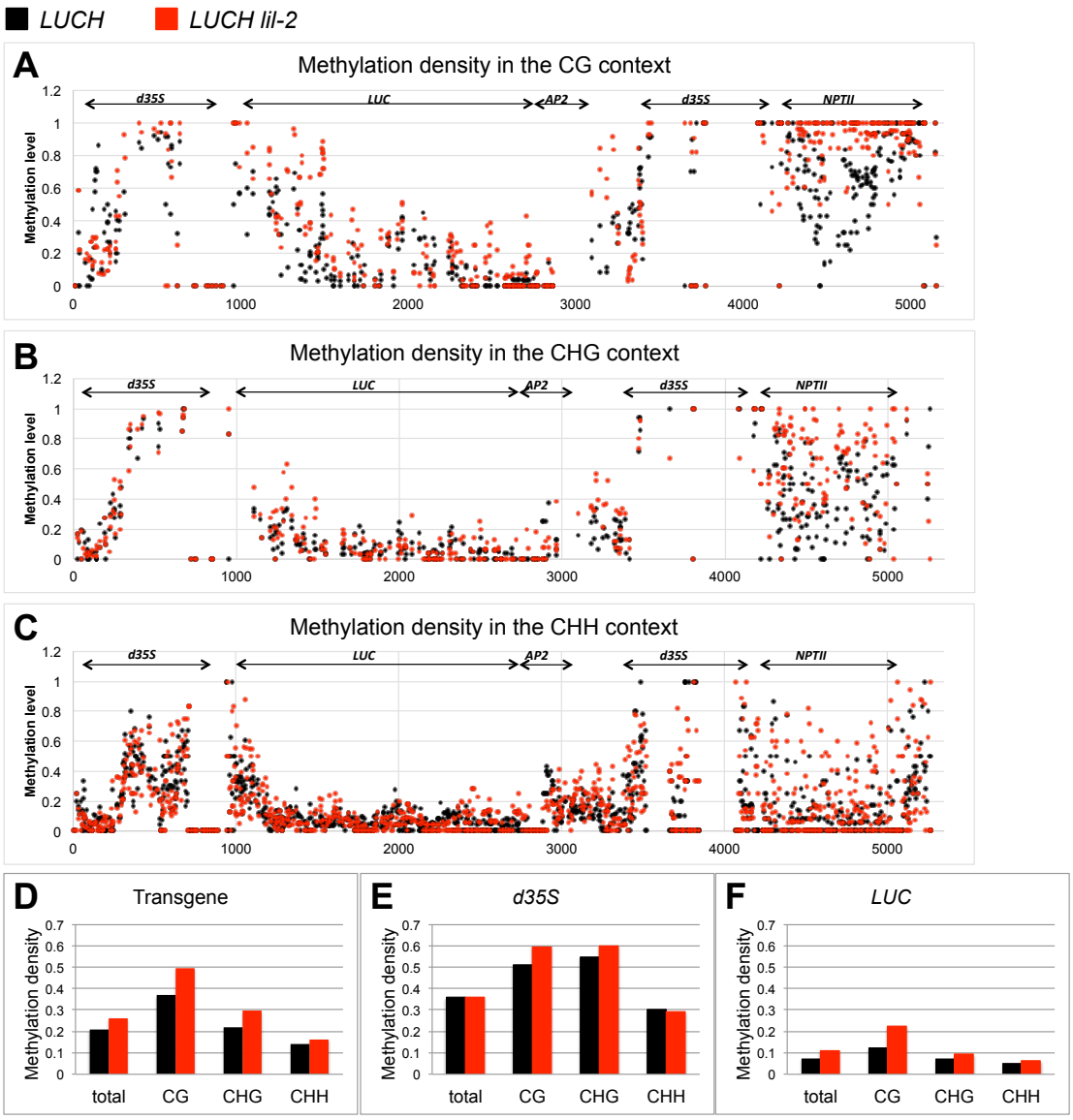

LUCH
LUCH lil-2
A
Methylation density in the CG context
B
Methylation density in the CHG context
C
Methylation density in the CHH context
d35S
LUC
AP2
d35S
NPTII
Methylation level
D
Transgene
E
d35S
F
LUC
Methylation density
total
CG
CHG
CHH

Figure 3.5 Effect of *LIL* on the expression of endogenous RdDM and ROS1 target loci.

(A) Real-time RT-PCR analysis of endogenous RdDM target loci. The effect of the mutations in *LIL* on the transcript levels of *AtMU1*, *AtSN1*, *AtGP1* and *soloLTR* was analyzed (top). Mutations disrupting Pol V activity (*nrpe1-11* and *nrpe1-1*) were used for the control experiment (bottom).
(B) Real-time RT-PCR analysis of endogenous ROS1 targets. The effect of the mutations in *LIL* on the transcript levels of *At2g32160* and *At3g47300* was analyzed (left). The relative transcript levels in *YJ ros1-3 dm2-1 dml3-1* (*YJ rdd*) are also shown (right).
(A), (B) The transcript levels were normalized to *UBIQUITIN5* and compared to the respective control line. Error bars indicate the standard deviation from three technical replicates.

Figure 3.5

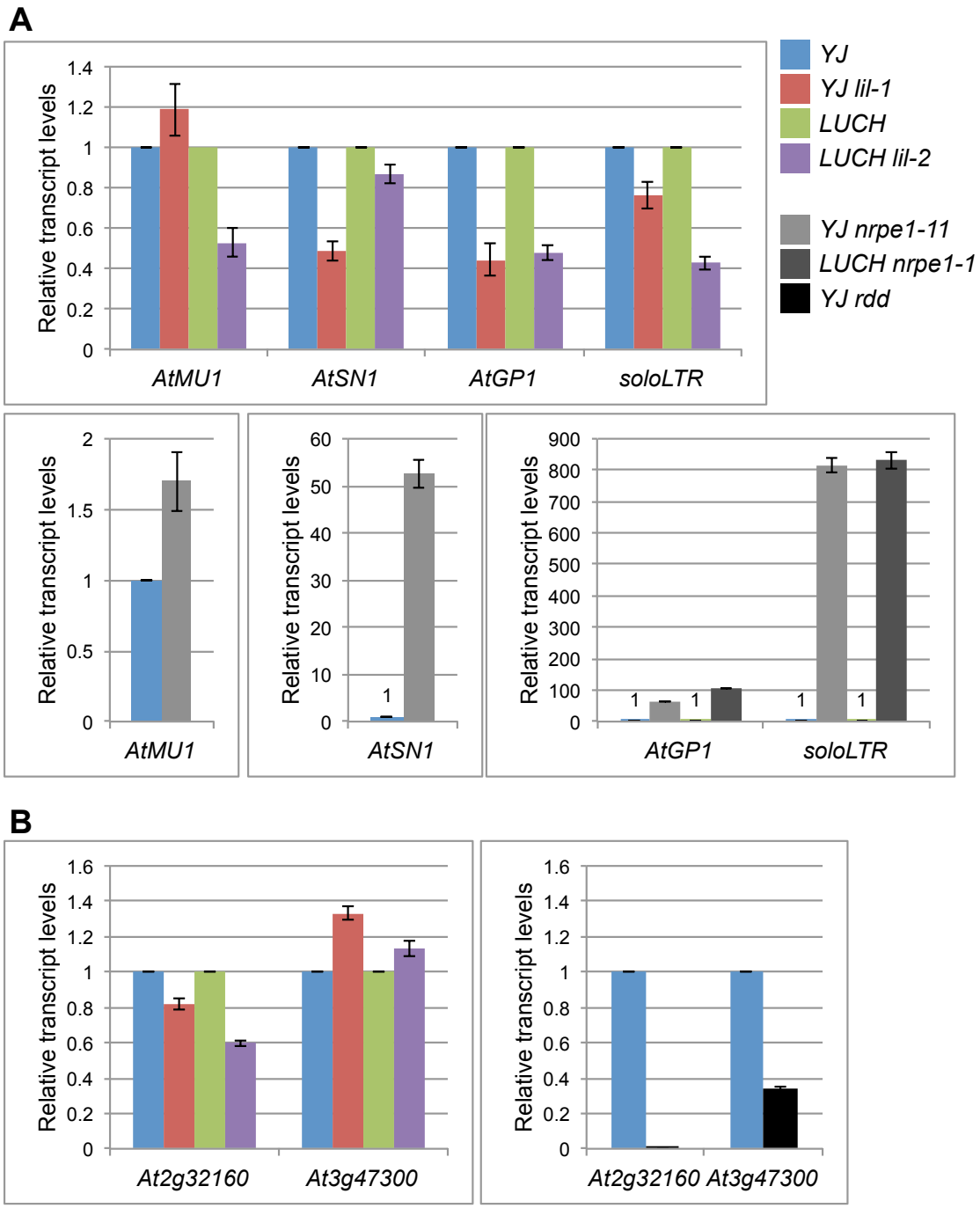

A
Relative transcript levels
1.4
1.2
1
0.8
0.6
0.4
0.2
0
AtMU1
AtSN1
AtGP1
soloLTR
YJ
YJ lil-1
LUCH
LUCH lil-2
YJ nrpe1-11
LUCH nrpe1-1
YJ rdd
Relative transcript levels
2
1.5
1
0.5
0
AtMU1
Relative transcript levels
60
50
40
30
20
10
0
1
AtSN1
Relative transcript levels
900
800
700
600
500
400
300
200
100
0
1
1
1
1
AtGP1
soloLTR
B
Relative transcript levels
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
At2g32160
At3g47300
Relative transcript levels
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
At2g32160
At3g47300

Figure 3.6 Genetic interaction analysis of *lil-1*.

(A) Real-time RT-PCR analysis of *LUC*.

(B) to (D) Real-time RT-PCR analysis of RdDM targets. *soloLTR* (B), *AtMU1* (C) and *AtSN1* (D) were examined.

(E) Real-time RT-PCR analysis of ROS1 targets. *At2g32160* and *At3g47300* were examined.

(A) to (E) *rdd* indicates *ros1-3 dml2-1 dml3-1*. For all of the tested loci and genotypes, the transcript levels were normalized to *UBIQUITIN5* and compared to *YJ*. Error bars indicate the standard deviation from three technical replicates. Due to the large differences in the values for the different genotypes, the scale of the y-axis was adjusted to improve visualization.

Figure 3.6

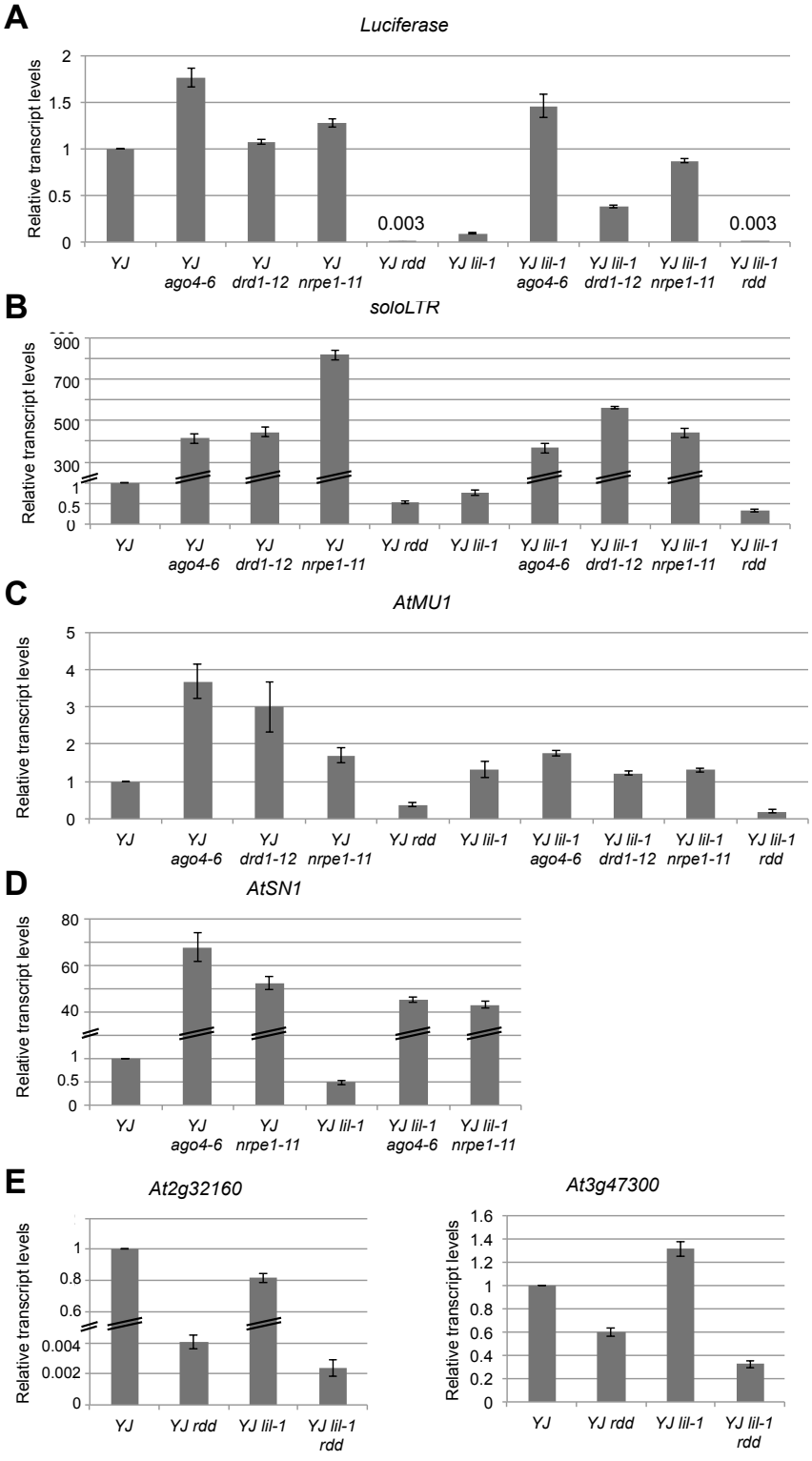

A
Luciferase
Relative transcript levels
2
1.5
1
0.5
0
0.003
0.003
YJ
YJ ago4-6
YJ drd1-12
YJ nrpe1-11
YJ rdd
YJ lil-1
YJ lil-1 ago4-6
YJ lil-1 drd1-12
YJ lil-1 nrpe1-11
YJ lil-1 rdd
B
soloLTR
Relative transcript levels
900
700
500
300
1
0.5
0
C
AtMU1
Relative transcript levels
5
4
3
2
1
0
D
AtSN1
Relative transcript levels
80
60
40
1
0.5
0
E
At2g32160
Relative transcript levels
1
0.8
0.6
0.004
0.002
0
At3g47300
Relative transcript levels
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0

Figure 3.7 Genome-wide analysis of cytosine methylation in mutants of LIL.

(A) The number of CG, CHG and CHH DMRs (differentially methylated regions) in *lil*. Blue and red bars indicate DMRs with increased and reduced cytosine methylation, respectively, in *lil* relative to wild-type.

(B) Box plots showing the association between hypermethylated CHH DMRs in *lil* and small RNA regions. The distances between hypermethylated DMRs in *lil* and small RNA regions were calculated. Small RNA regions correspond to the regions in which small RNA levels are reduced in *sde4-3* (a Pol IV mutants), compared to Col-0. CHH-DMRs and CG-DMRs in *lil* were plotted together with the randomly chosen genomic regions (CTL, control) for comparision. The same number of loci (106 for CG and 134 for CHH) were selected as a control and these control regions do not include LIL-specific DMRs. * indicates the significant difference by Mann-Whitney U test ($P < 10^{-10}$).

(C), (F) A pie chart showing genomic locations of hypermethylated DMRs in *lil* at CHH (C) and CG (F) cytosine contexts. Transposons (TE), intergenic regions and gene body are plotted in blue, red and green, respectively.

(D), (G) A pie chart showing the dependency on Pol IV or Pol V of hypermethylated DMRs in *lil*. The overlap between hypomethylated DMRs in *sde4-3*/*nrpe1-1* and hypermethylated DMRs in *lil* were examined in CHH (D) and CG (G) sequence contexts.

(E), (H) A pie chart showing the dependency on ROS1 of hypermethylated DMRs in *lil*. The overlap between hypermethylated DMRs in *rdd (ros1 dml2 dml3)* and hypermethylated DMRs in *lil* were examined in CHH (E) and CG (H) sequence contexts.

Figure 3.7

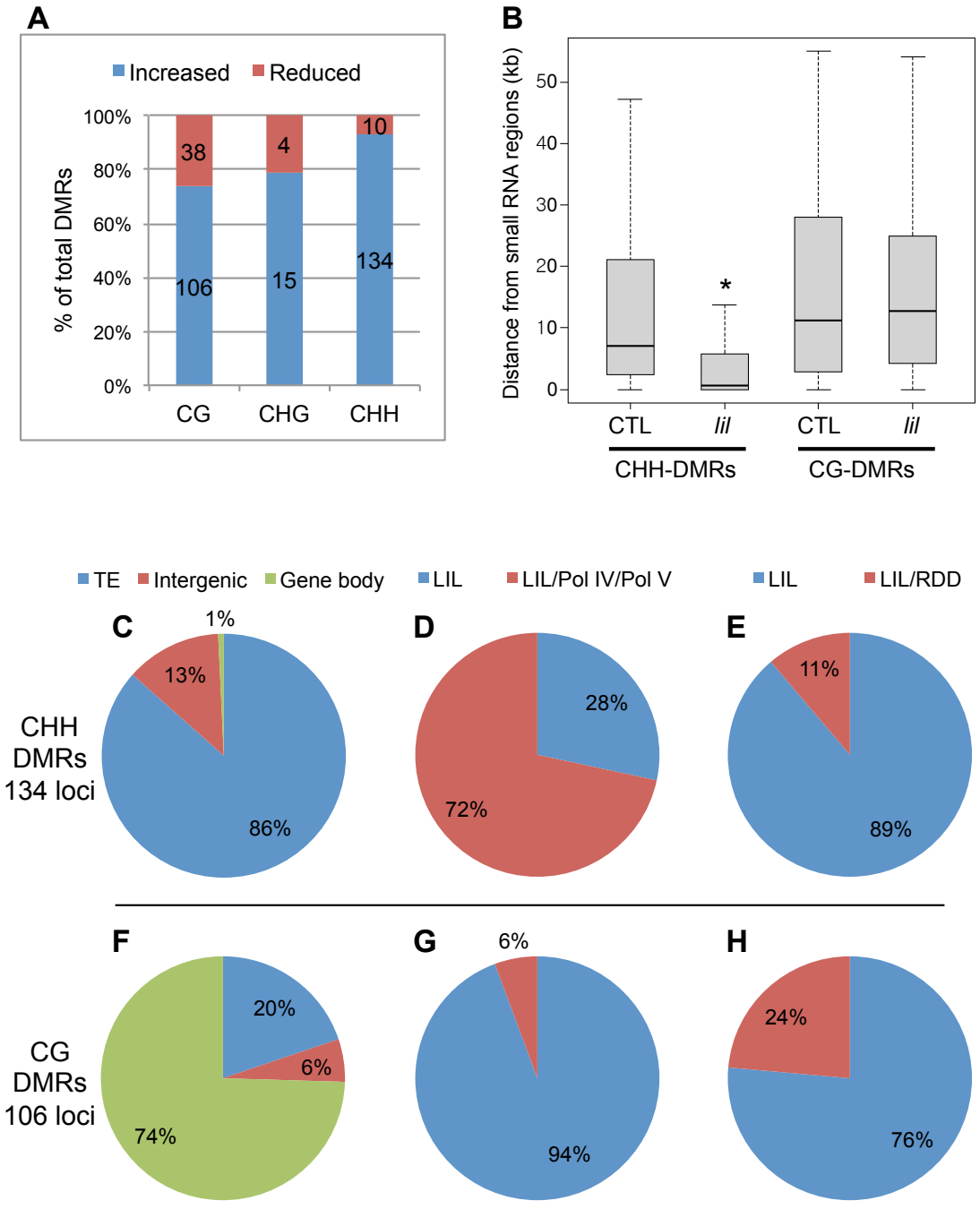

A
Increased
Reduced
% of total DMRs
100%
80%
60%
40%
20%
0%
38
106
4
15
10
134
CG
CHG
CHH
B
Distance from small RNA regions (kb)
50
40
30
20
10
0
*
CTL
lil
CTL
lil
CHH-DMRs
CG-DMRs
TE
Intergenic
Gene body
LIL
LIL/Pol IV/Pol V
LIL
LIL/RDD
C
1%
13%
86%
CHH
DMRs
134 loci
D
28%
72%
E
11%
89%
F
20%
6%
74%
CG
DMRs
106 loci
G
6%
94%
H
24%
76%

Figure 3.8 Validation of the hypermethylated regions identified in the *lil* mutants.

(A) Four representative regions with increased cytosine methylation. From top to bottom, the methylation levels in *YJ*, *YJ lil-1*, *LUCH* and *LUCH lil-2* are displayed, with the height of each peak indicating methylation density at each cytosine. *IDH1*, *2* and *4* are DMRs in the CG context. *IDH13* is a DMR in the CHH context, with the y-axis scale adjusted to 50% for better visualization. Grey arrows beneath the plots indicate transposons in *IDH1*, *2* and *13*. In the *IDH4* panel, the blue bars and thin lines represent exons and introns, respectively.

(B) to (D) McrBC-qPCR analysis to confirm the BS-seq results. DMRs with increased methylation mapped to transposons (B), regions encompassing both transposons and protein-coding genes (C) or protein-coding genes (D). Cytosine methylation in *IDH2* was increased at all cytosine contexts. *IDH9* and *IDH13* were DMRs in the CHH and CHG/CHH contexts, respectively. All other DMRs shown are CG DMRs. Genomic DNA was treated with McrBC and subjected to qPCR. *ACTIN1*, which lacks cytosine methylation, was used to normalize the amplified product. *YJ lil-1* and *LUCH lil-2* were compared to *YJ* and *LUCH*, respectively. Standard deviations were calculated from three technical replicates and are indicated by the error bars.

Figure 3.8

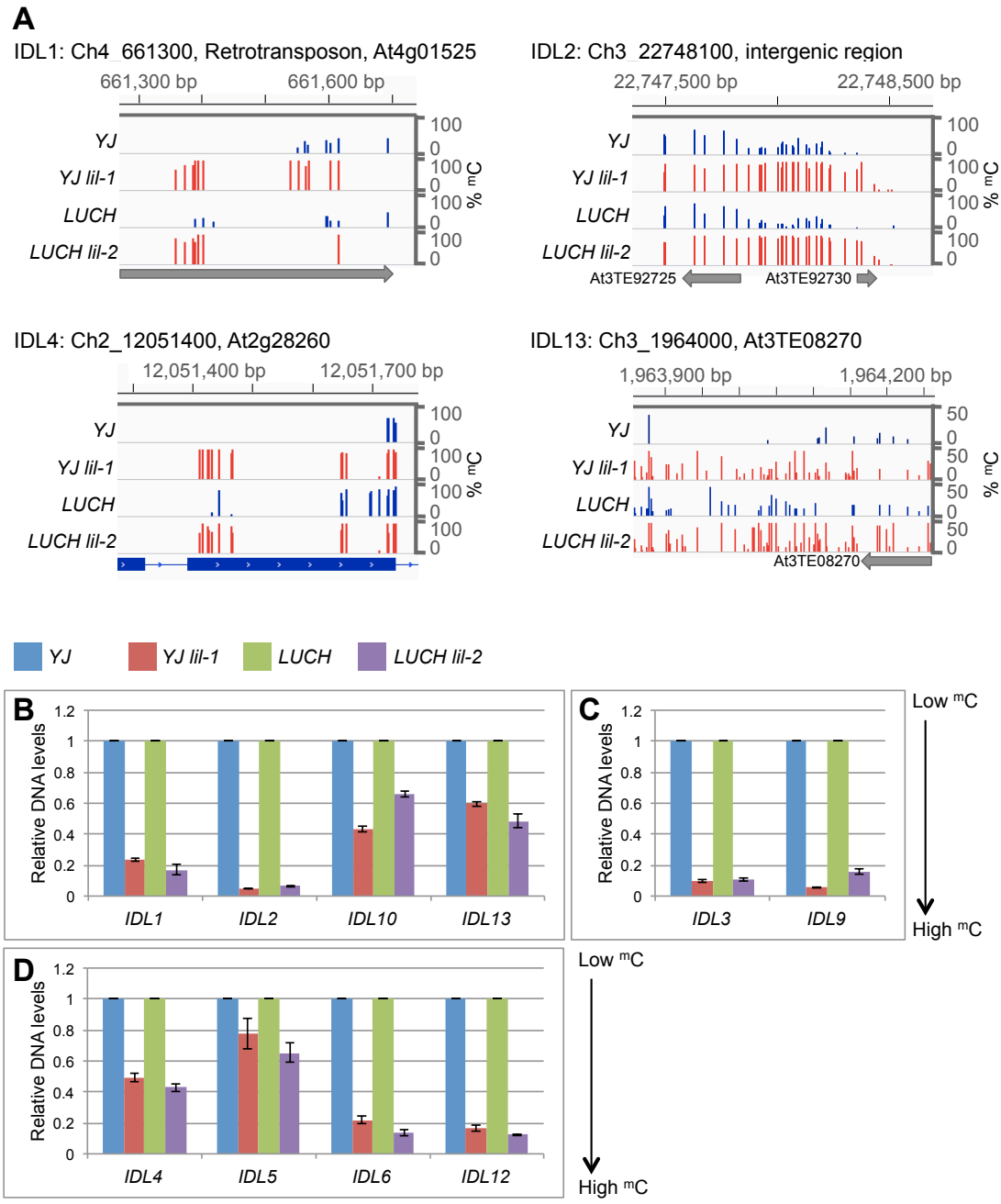

A
IDL1: Ch4_661300, Retrotransposon, At4g01525
661,300 bp
661,600 bp
YJ
YJ lil-1
LUCH
LUCH lil-2
% mC
IDL2: Ch3_22748100, intergenic region
22,747,500 bp
22,748,500 bp
At3TE92725
At3TE92730
IDL4: Ch2_12051400, At2g28260
12,051,400 bp
12,051,700 bp
IDL13: Ch3_1964000, At3TE08270
1,963,900 bp
1,964,200 bp
At3TE08270
YJ
YJ lil-1
LUCH
LUCH lil-2
B
C
D
Relative DNA levels
IDL1
IDL2
IDL10
IDL13
IDL3
IDL9
IDL4
IDL5
IDL6
IDL12
Low mC
High mC

Figure 3.9 *LIL* does not affect Pol IV or Pol V activity.

(A) Size distribution of total small RNA reads in Col-0, *sde4-3*, *YJ* and *YJ lil-1*. The genome-wide accumulation of small RNAs was analyzed by small RNA-seq. The number of small RNAs 20 to 24 nt in length was determined, and the relative abundance of each size class was calculated in each genotype.

(B) Box plots showing small RNA abundance in Col-0, *sde4-3*, *nrpe1-11*, *YJ* and *YJ lil-1*. The *Arabidopsis* genome was divided into 500-bp continuous windows, and the small RNA reads were mapped to the windows. For each window, read abundance was calculated as RPM (reads per million), which is plotted on the y-axis. For Col-0, *sde4-3* and *nrpe1-11*, only windows with RPM values greater than 10 in Col-0 were included in the analysis and plotted in the box plots. For YJ and YJ *lil-1*, only windows with RPM values greater than 10 in either plant were included in the analysis and plotted in the box plots. * indicates significant reduction by Mann-Whitney *U* test ($P < 10^{-10}$).

(C) RT-PCR analysis of Pol V-dependent transcripts in Col-0, *nrpe1-1*, *YJ* and *YJ lil-1*. In contrast to *nrpe1-1*, the *hsp20-1* mutation did not impact the accumulation of Pol V-dependent transcripts. *UBIQUITIN5* (*UBQ5*) served as the loading control, and the -RT reaction was performed using *UBQ5* primers.

Figure 3.9

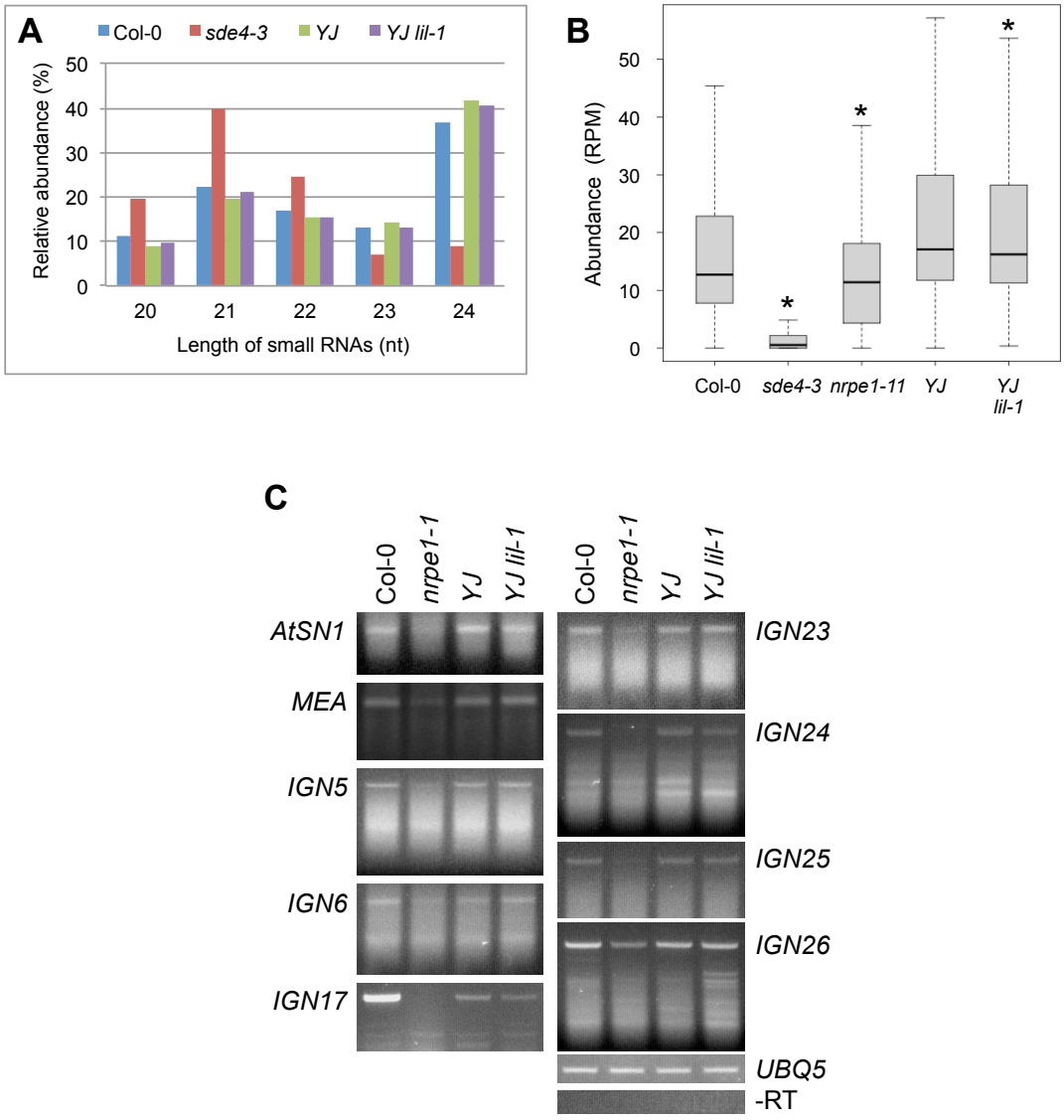

A
Col-0
sde4-3
YJ
YJ lil-1
Relative abundance (%)
50
40
30
20
10
0
20
21
22
23
24
Length of small RNAs (nt)
B
Abundance (RPM)
50
40
30
20
10
0
*
*
*
Col-0
sde4-3
nrpe1-11
YJ
YJ lil-1
C
Col-0
nrpe1-1
YJ
YJ lil-1
Col-0
nrpe1-1
YJ
YJ lil-1
AtSN1
MEA
IGN5
IGN6
IGN17
IGN23
IGN24
IGN25
IGN26
UBQ5
-RT

Figure 3.10 Localization of AGO4 and LIL.

(A) Subnuclear localization of AGO4. Immunostaining analysis of the nuclear distribution of AGO4 using anti-AGO4 antibody (red) in wild-type and *YJ lil-1*. DNA was stained by DAPI (blue).

(B) Subcellular localization of LIL. *LIL* was fused to *GFP* and transcribed by the *35S* promoter. The resulting plasmid was agro-infiltrated into *Nicotiana benthamiana* leaves for transient expression assays, and GFP signal was examined using confocal microscopy. Red and green signals correspond to chloroplast autofluorescence and GFP expression, respectively. Bright field and merged images are shown on the bottom left and bottom right, respectively, in all three panels.

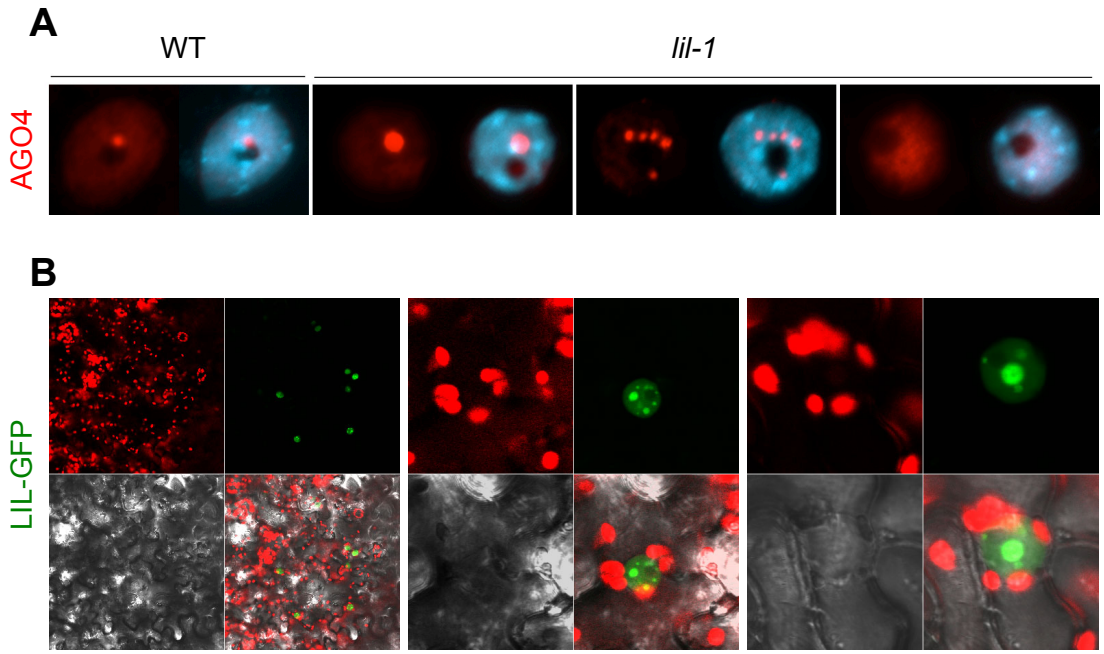


Figure 3.11 LIL interacts with MBD proteins.

(A) Constructs used for yeast two-hybrid assays in (B). Full-length and partial *LIL* fragments were fused to the sequence encoding the GAL4 DNA binding domain (BD). The black box indicates the region encoding the HSP20-like chaperone domain. Full-length and partial MBD fragments were fused to the sequence encoding the GAL4 activation domain (AD). Grey boxes indicate methyl-CpG-binding domains. aa indicates amino acid.

(B) Yeast two-hybrid interactions. *LIL-BD* and *MBD-AD* plasmids were co-transformed into the yeast strain AH109. Yeast growth on plates containing SD/-Leu/-Trp confirms the transformation of both plasmids into yeast. Yeast growth on plates containing SD/-Leu/-Trp/-Ade/-His indicates protein interaction.

(C) Pull-down assay of His-LIL by GST-MBD5 and GST-MBD6. GST-MBD5 and GST-MBD6 were expressed in *E. coli* and captured by glutathione-agarose beads. The lysate from *E. coli* expressing His-LIL was incubated with the GST-MBD-bound beads. After pull-down, GST-MBD5/6 and His-LIL were detected by Western blot analysis using anti-GST and anti-His antibodies, respectively. The panel on the left shows the expression and immobilization of GST-MBD5/6 by the glutathione-agarose beads. In the panel on the right, the first lane shows the expression of His-LIL, and the last three lanes show the specific interaction of His-LIL with GST-MBD5 and GST-MBD6.

Figure 3.11

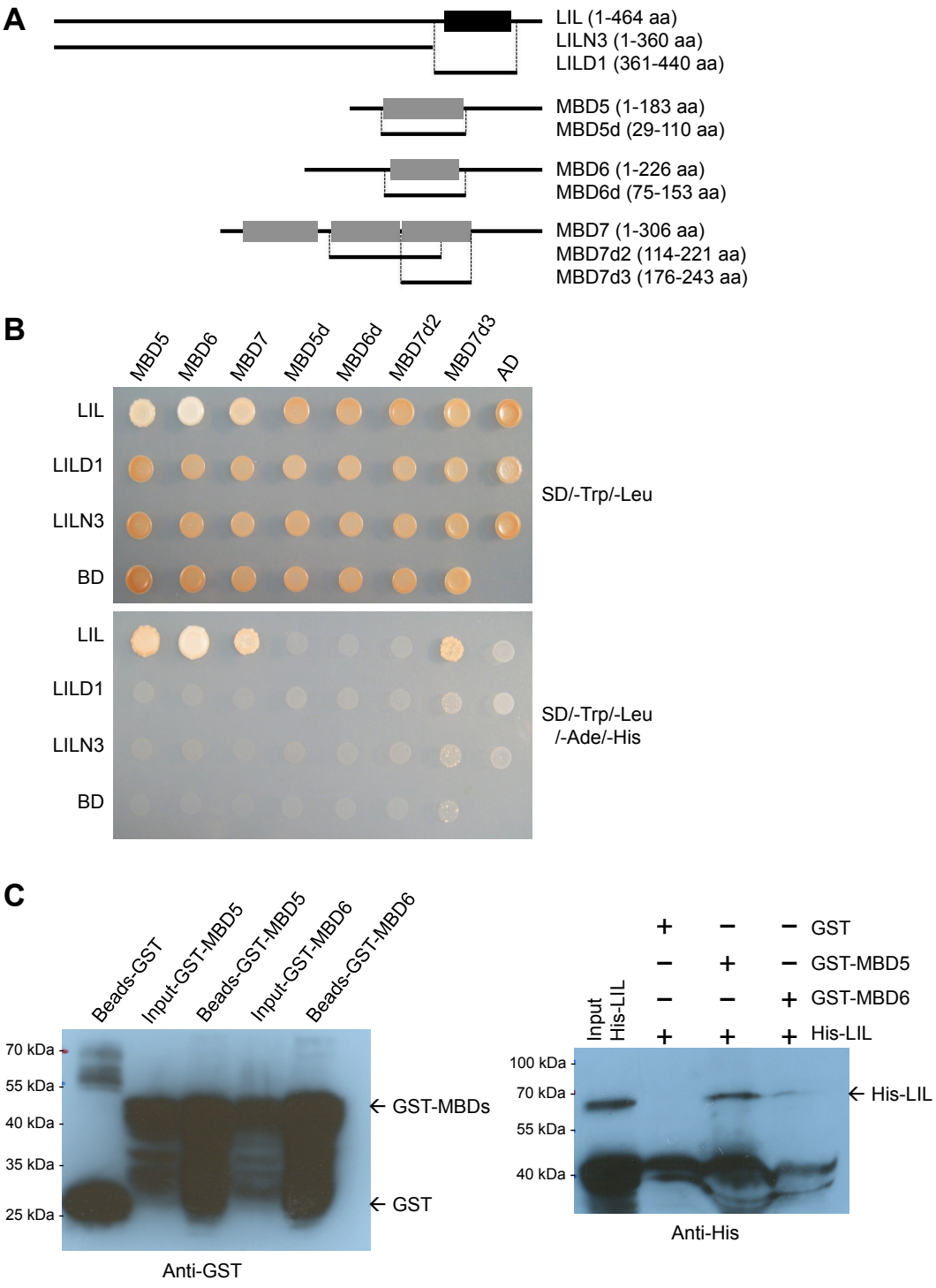

A
LIL (1-464 aa)
LILN3 (1-360 aa)
LILD1 (361-440 aa)
MBD5 (1-183 aa)
MBD5d (29-110 aa)
MBD6 (1-226 aa)
MBD6d (75-153 aa)
MBD7 (1-306 aa)
MBD7d2 (114-221 aa)
MBD7d3 (176-243 aa)
B
MBD5
MBD6
MBD7
MBD5d
MBD6d
MBD7d2
MBD7d3
AD
LIL
LILD1
LILN3
BD
SD/-Trp/-Leu
SD/-Trp/-Leu
/-Ade/-His
C
Beads-GST
Input-GST-MBD5
Beads-GST-MBD5
Input-GST-MBD6
Beads-GST-MBD6
70 kDa
55 kDa
40 kDa
35 kDa
25 kDa
GST-MBDs
GST
Anti-GST
Input
His-LIL
GST
GST-MBD5
GST-MBD6
His-LIL
100 kDa
70 kDa
55 kDa
40 kDa
His-LIL
Anti-His

Figure 3.12 Isolation and characteristics of the *mbd* mutants.

(A) Schematic diagram showing the site of the T-DNA insertion in *mbd5*, *mbd6* and *mbd7*. White and black rectangles indicate untranslated regions and exons, respectively, and the thin lines indicate introns. The methyl-CpG-binding domains (MBDs) are indicated beneath each diagram. The locations of the primers for RT-PCR are indicated with arrowheads.

(B) RT-PCR of *MBD5*, *MBD6* and *MBD7* genes in the respective mutants. *UBIQUITIN5* (*UBQ5*) was used as the loading control, and the -RT reaction was performed using *UBQ5* primers. For *MBD5*, the *MBD5*-specific reverse primer (R1) was also paired with a T-DNA-specific primer (Vector) to detect aberrant transcripts that were initiated from the T-DNA.

Figure 3.12

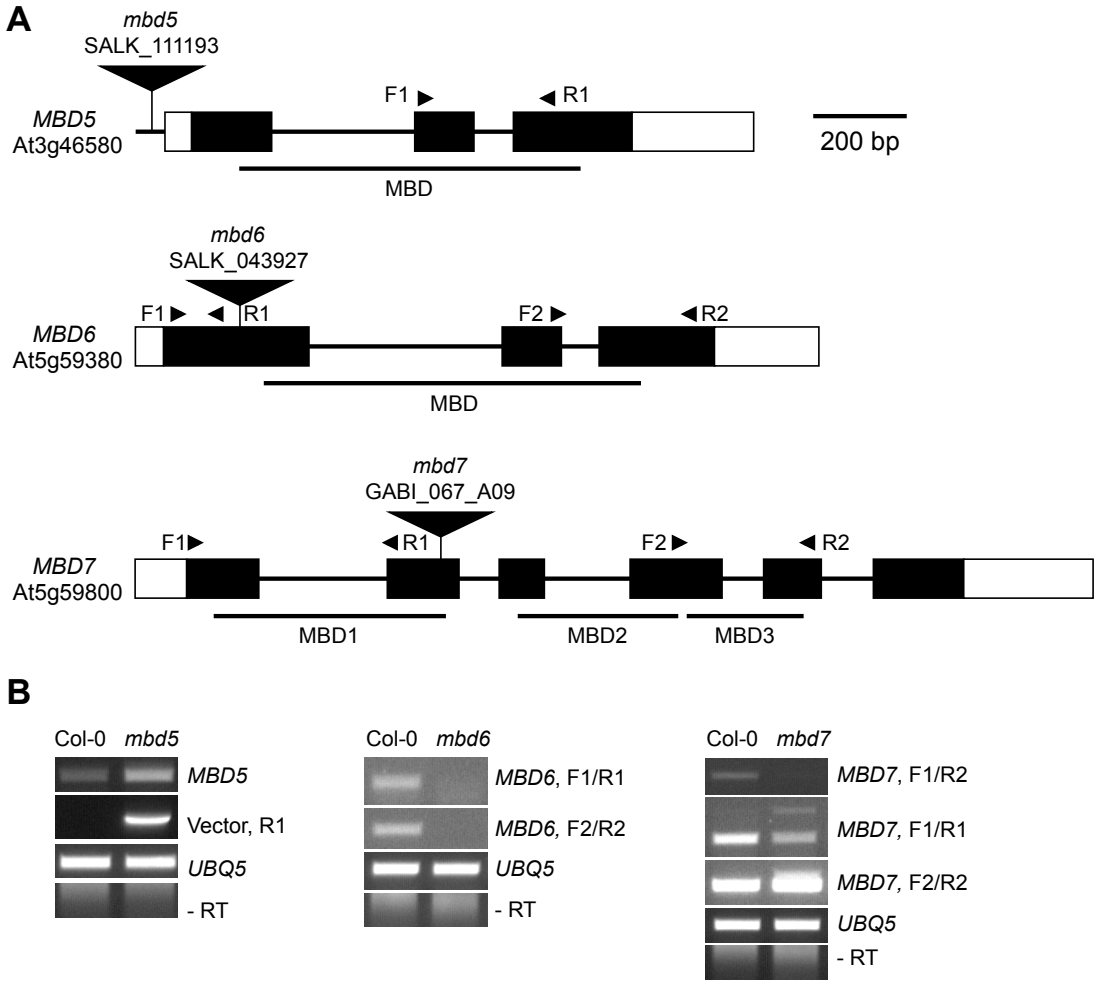

A
mbd5
SALK_111193
MBD5
At3g46580
F1
R1
200 bp
MBD
mbd6
SALK_043927
MBD6
At5g59380
F1
R1
F2
R2
MBD
mbd7
GABI_067_A09
MBD7
At5g59800
F1
R1
F2
R2
MBD1
MBD2
MBD3
B
Col-0 mbd5
MBD5
Vector, R1
UBQ5
- RT
Col-0 mbd6
MBD6, F1/R1
MBD6, F2/R2
UBQ5
- RT
Col-0 mbd7
MBD7, F1/R2
MBD7, F1/R1
MBD7, F2/R2
UBQ5
- RT

Figure 3.13 McrBC-qPCR analysis of methylation density in the *mbd* mutants at regions showing increased DNA methylation in *lil* mutants.

(A) McrBC-qPCR analysis of *YJ*, *YJ lil-1*, *LUCH* and *LUCH lil-2*.

(B) McrBC-qPCR analysis of Col-0, *nrpe1-1*, *mbd5* and *mbd6*.

(A), (B) DMRs with increased methylation in the *lil* mutants were classified into transposons, protein-coding genes or regions encompassed by both. Genomic DNA was treated with McrBC and subjected to qPCR. *ACTIN1*, which lacks cytosine methylation, was used to normalize the amplified product. *Mutants* were compared to their respective control plants or the wild-type plant. Standard deviations were calculated from three technical replicates and are indicated by the error bars.

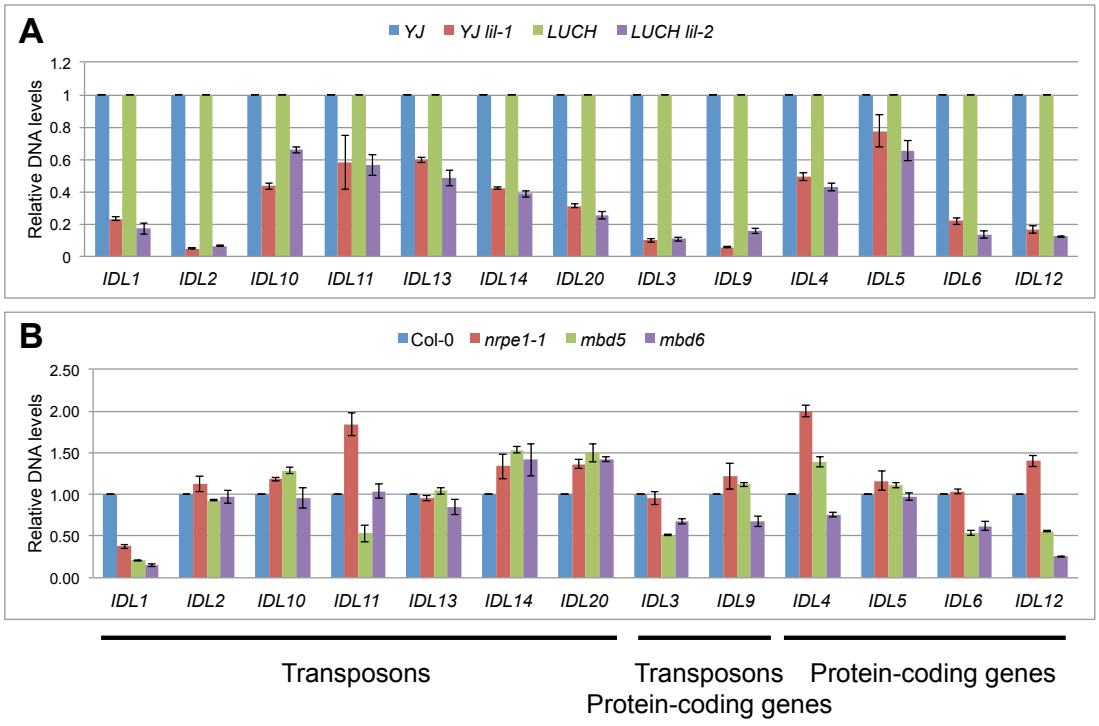
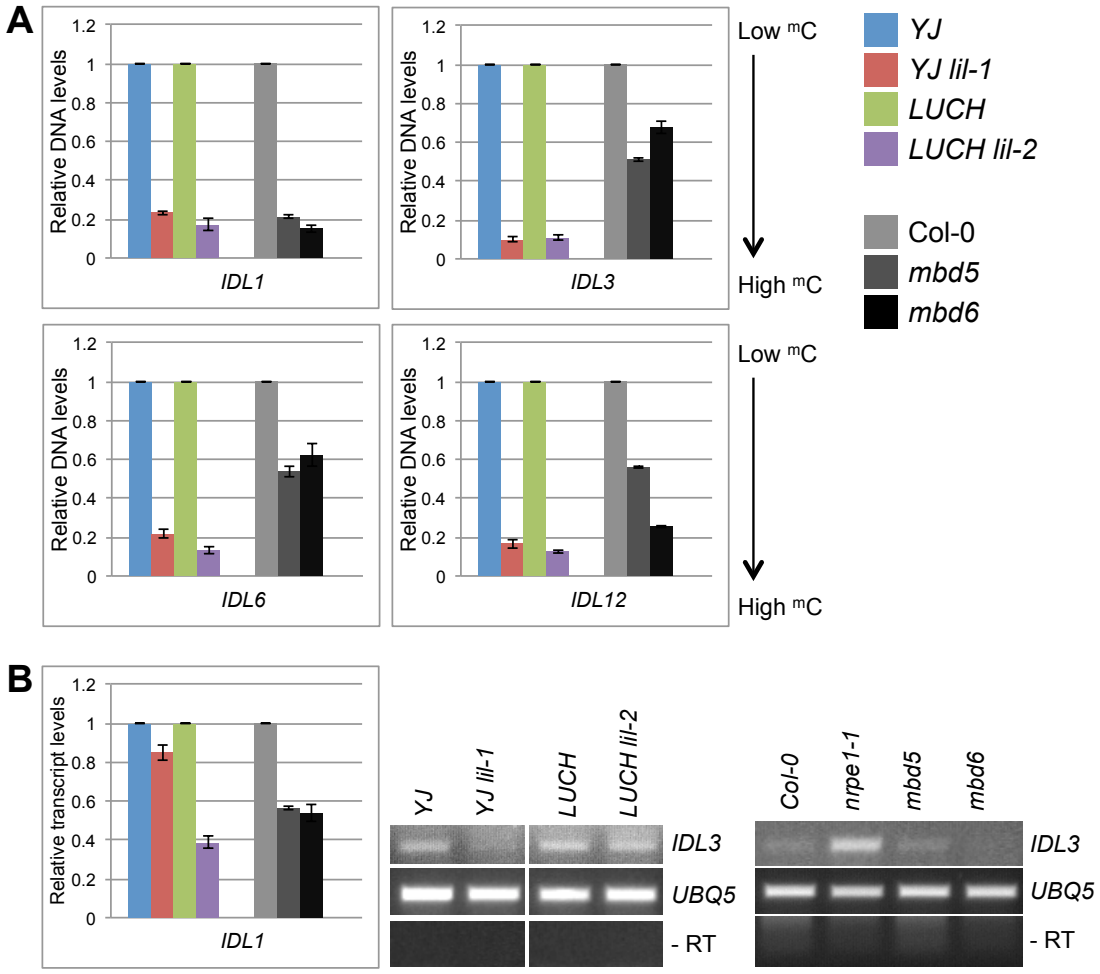


Figure 3.14 Impact of *MBD5* and *MBD6* on cytosine methylation and TGS at regions impacted by LIL.

(A) Analysis of cytosine methylation by McrBC-qPCR. The methylation status of regions displaying increased DNA methylation in *lil* mutants was examined in Col-0, *mbd5* and *mbd6*, and in the *lil* mutants and their corresponding controls. The results for the *lil* mutants are shown on the left-hand side in each graph. Genomic DNA was treated with McrBC and subjected to qPCR. *ACTIN1*, which lacks cytosine methylation, was used to normalize the amplified product. Mutants were compared to wild-type or their respective controls. Standard deviations were calculated from three technical replicates and are indicated by the error bars.

(B) RT-PCR to examine the expression of DMRs. The transcripts from *IDL1* were quantified by real-time RT-PCR, normalized using *UBIQUITIN 5* (*UBQ5*) and compared to wild-type or the respective control. Regular RT-PCR was conducted to examine the transcript levels of *IDL3*. *UBQ5* was used as the loading control, and the -RT reaction was performed using *UBQ5* primers.

Figure 3.14


A
Relative DNA levels
IDL1
IDL3
IDL6
IDL12
Low mC
High mC
YJ
YJ lil-1
LUCH
LUCH lil-2
Col-0
mbd5
mbd6
B
Relative transcript levels
IDL1
YJ
YJ lil-1
LUCH
LUCH lil-2
Col-0
nrpe1-1
mbd5
mbd6
IDL3
UBQ5
- RT

TABLES

Table 3.1 Summary of bisulfite conversion efficiency for each genotype.

	CG	CHG	CHH	Total C
Col-0 (A)	98.2%	98.1%	97.8%	97.9%
Col-0 (B)	97.8%	97.7%	97.6%	97.6%
sde4-3 (A)	98.0%	98.0%	97.8%	97.9%
sde4-3 (B)	97.9%	97.8%	97.7%	97.7%
nrpe1-11 (A)	98.1%	98.1%	97.9%	98.0%
nrpe1-11 (A)	98.0%	98.0%	97.9%	97.9%
YJ	98.1%	98.0%	97.9%	97.9%
YJ lil-1	98.0%	97.9%	97.7%	97.8%
LUCH	98.0%	98.0%	97.9%	97.9%
LUCH lil-2	98.1%	98.1%	97.9%	98.0%
Col-0 (C)	98.6%	98.5%	98.4%	98.4%
ros1 dml2 dml3 (C)	98.2%	98.2%	98.1%	98.1%

(A), (B) and (C) indicate different biological replicates. All samples from the same biological replicates were processed at the same time and in the same manner.

Table 3.2 Read coverage of whole-genome bisulfite sequencing libraries.

CHH	# of sequenced mC	# of total sequenced C	31198380 [1]
			Coverage
Col-0 (A)	6974394	228191510	7.314
Col-0 (B)	8081587	276670598	8.868
sde4-3 (A)	3996254	221784087	7.109
sde4-3 (B)	3928539	238723704	7.652
nrpe1-11 (A)	5488089	293648870	9.412
nrpe1-11 (B)	3147078	175298018	5.619
YJ	7591767	253807442	8.135
YJ lil-1	7214717	247728128	7.94
LUCH	7342876	281810404	9.033
LUCH lil-2	9007295	251352047	8.057
Col-0 (C)	10578187	238998500	7.661
rdd (C)	11671566	256604056	8.225
CG	# of sequenced mC	# of total sequenced C	5567714 [2]
			Coverage
Col-0 (A)	11967876	40092653	7.201
Col-0 (B)	14292137	48943906	8.791
sde4-3 (A)	11233072	40655236	7.302
sde4-3 (B)	11410979	42322601	7.601
nrpe1-11 (A)	14992909	51859716	9.314
nrpe1-11 (B)	8591198	30902224	5.55
YJ	13539764	46510382	8.354
YJ lil-1	14228416	49050371	8.81
LUCH	15786653	54455962	9.781
LUCH lil-2	16424963	51444033	9.24
Col-0 (C)	14918204	41610035	7.473
rdd (C)	18343806	45090383	8.099
CHG	# of sequenced mC	# of total sequenced C	6093657 [3]
			Coverage
Col-0 (A)	4368090	43083879	7.07
Col-0 (B)	5122123	52159761	8.56
sde4-3 (A)	3536815	43885788	7.202
sde4-3 (B)	3458560	45355541	7.443
nrpe1-11 (A)	4982617	55895446	9.173
nrpe1-11 (B)	2750976	33011274	5.417
YJ	5125708	50315801	8.257
YJ lil-1	5092482	53543374	8.787
LUCH	5982514	59537845	9.77
LUCH lil-2	6506327	55975060	9.186
Col-0 (C)	5940881	42542339	6.981
rdd (C)	7659881	46000253	7.549
Total	# of sequenced mC	# of total sequenced C	42859751 [4]
			Coverage
Col-0 (A)	23310360	311368042	7.265
Col-0 (B)	27495847	377774265	8.814

sde4-3 (A)	18766141	306325111	7.147
sde4-3 (B)	18798078	326401846	7.616
nrpe1-11 (A)	25463615	401404032	9.366
nrpe1-11 (B)	14489252	239211516	5.581
YJ	26257239	350633625	8.181
YJ lil-1	26535615	350321873	8.174
LUCH	30127316	390197281	9.104
LUCH lil-2	31938585	358771140	8.371
Col-0 (C)	31437272	323150874	7.54
rdd (C)	37675253	347694692	8.112

(A) and (B) indicate two independent biological replicates. All samples from the same biological replicates were processed at the same time and in the same manner.
Coverage = # of total sequenced C / # of total CXX sites in genome.
[1] # of total CHH sites in genome.
[2] # of total CG sites in genome.
[3] # of total CHG sites in genome.
[4] # of total C sites in genome

Table 3.3 Genomic locations of DMRs in *lil*.

		Increased			Reduced		
		CG	**CHG**	**CHH**	**CG**	**CHG**	**CHH**
Repeat	Dispersed	10	6	96	1	3	6
	Inverted	3	2	9	1	0	1
	Tandem	2	0	2	0	0	0
TE		6	1	9	4	0	0
Subtotal (Repeat/TE)		21	9	116	6	3	7
Gene		79	6	1	31	1	0
Intergenic		6	0	17	1	0	3
Total		106	15	134	38	4	10

Table 3.4 Only several genes were differentially expressed in the *lil* mutants.

	Down-regulated in *lil*	**Up-regulated in *lil***
YJ lil-1 / *YJ*	102	94
LUCH lil-2 / *LUCH*	22	8
Overlapped genes	0	3

Table 3.5 Only 60 differential small RNA regions (DSRs) were found in *YJ lil-1*.

Small RNA-seq was performed to profile small RNA populations in Col-0, *sde4-3* (a Pol IV mutant), *nrpe1-11* (a Pol V mutant), *YJ* and *YJ lil-1*. Total 10729 and 13960 windows were identified in Col-0/*sde4-3* and Col-0/*nrpe1-11* comparison, respectively. Total 12415 windows were identified in the pair of *YJ* and *YJ lil-1*. A fold change > 4 and an adjusted p-value (FDR) < 0.05 were required for DSRs between mutants and their respective controls. "Reduced" and "Increased" refer to DSRs with reduced and increased read count of small RNAs, respectively, in mutants.

	***sde4-3*/Col-0**	***nrpe1-11*/Col-0**	***YJ lil-1*/*YJ***
Reduced	8483	3524	15
Increased	50	385	45

Table 3.6 Oligonucleotide sequences used in this study.

Name	Oligonucleotide sequence	Purpose
LUCmF5	CTCCCCTCTCTAAGGAAGTCG	RT-PCR for LUC
LUCmR5	CCAGAATGTAGCCATCCATC	
N_UBQ5	GGTGCTAAGAAGAGGAAGAAT	RT-PCR, loading control
C_UBQ5	CTCCTTCTTTCTGGTAAACGT	
35Sf	CAAAGCAAGTGGATTGATGTGA	McrBC-PCR of d35S
LUC 0.13k R	TATGTGCATCTGTAAAAGCAA	
lucp6	GCACCCGGGGAAGACGCCAAAAACATAAAAGAAA	McrBC-PCR of LUC
lucp7	GGACCCGGGTGCGATCTTTCCGCCCTTCTTGGCCT	
Actin1-F	CCAAGCAGCATGAAGATCAA	McrBC-PCR, loading control
Actin1-R	TGAACAATCGATGGACCTGA	
AtMu1F1	CCGAGAACTGGTTGTGGTTT	RT-PCR of AtMU1
AtMu1R1	GCTCTTGCTTTGGTGATGGT	
soloLTR A221	ATCAATTATTATGTCATGTTAAAACCGATTG	RT-PCR of soloLTR
soloLTR A222	TGTTTCGAGTTTTATTCTCTCTAGTCTTCATT	
AtSN1 F1	ACCAACGTGCTGTTGGCCCAGTGGTAAATC	RT-PCR of AtSN1
AtSN1 R1	AAAATAAGTGGTGGTTGTACAAGC	
ATGP1F	TGGTTTTTCCTGTCCAGTTTG	RT-PCR of AtGP1
ATGP1R	AACAATCCTAACCGGGTTCC	
AtSN1_B F	TGAGAGATTTACCACTGGGCCAACA	RT-PCR of AtSN1 (Pol V-dependent)
AtSN1_B R	TGAGGAGCTCAACACATAAATGGCAATA	
IGN5-RT-R	CTGAGGTATTCCATAGCCCCTGATCC	RT of IGN5 (Pol V-dependent)
IGN5 realtime F1	ATGAAGAAAGCCCAAACCAT	PCR of IGN5 (Pol V-dependent)
IGN5 realtime R1	GCCGAATAACAGCAAGTCCT	
IGN6-RT-R	TTTGTAATTCTCAGTTCGGGTATCTGCTTG	RT of IGN6 (Pol V-dependent)
IGN6 realtime F1	GCAAACATAGCAACCGAGAA	PCR of IGN6 (Pol V-dependent)
IGN6 realtime R1	GTTAGTGACGGCGAAAAAGC	
IGN17-F	AACCCTAGCCTTTCATTAAAACCCTCTC	RT-PCR of IGN17 (Pol V-dependent)
IGN17-R	CATAGATAGGAAACTCAATCTCTTCGCATTT	
IGN23-F	ACTGAAAATTGTAAACAAAGAAACGGCACTACA	RT-PCR of IGN23 (Pol V-dependent)
IGN23-R	GATCGGTCCATAAACTTGTTGGGTTT	
IGN24-F	CGCATACGATGGTCGGAGAGTT	RT-PCR of IGN24 (Pol V-dependent)
IGN24-R	GCTTATCATTATCCAAACTTGATCCTATCCTAAA	
IGN25-F	CTTCTTATCGTGTTACATTGAGAACTCTTTCC	RT-PCR of IGN25 (Pol V-dependent)
IGN25-R	ATTCGTGTGGGCTTGGCCTCTT	
IGN26-F	CTCTTTCAGTGCGACAGCCTCAT	RT-PCR of

Name	Oligonucleotide sequence	Purpose
IGN26-R	CGGCCAGGAAACCCTAACTTCC	IGN26 (Pol V-dependent)
MEA-realtime-F1	cgcgaacgactattgctaaa	RT-PCR of MEA (Pol V-dependent)
MEA-realtime-R1	acgattccacaaatccaaca	
At2G32160-F	AGCTGTCTCCTGCGCACAA	RT-PCR of At2g32160
At2G32160-R	GCTCGCGCATGTGTTTACTC	
At3G47300-F	CCTGCTCGTACAAGGGAACTG	RT-PCR of At3g47300
At3G47300-R	CAATAACACCCACCTGAGCAACT	
IDL1F	TTCGGTAAATCGGAGATTCG	McrBC-qPCR, qRT-PCR
IDL1R	ACCGGAAAAGAGGATCTTGC	
IDL2F	cggtgctcagctcataagaa	McrBC-qPCR
IDL2R	cctcttctccgtgctctctc	
IDL3F	tgaccatgcccactctctact	McrBC-qPCR, qRT-PCR
IDL3R	GGACTCTAGCGACCCGATAA	
IDL4F	ACTGTCTATGGTGGGGCTTG	McrBC-qPCR
IDL4R	CGGCTTGTCTTAGCTCTGGT	
IDL5F	CATGCCATGTCAAAGTCCCTA	McrBC-qPCR
IDL5R	TGGTTCAACTTCTGCTCATCA	
IDL6F	TTGTAGTGGGTGGTTCACGA	McrBC-qPCR, qRT-PCR
IDL6R	agGAACATTTGACCATGCAA	
IDL9F	GGCAAGGGACGTATAGCAAC	McrBC-qPCR
IDL9R	ATTCCTTGATGCGATGATCC	
IDL10F	GTTTGGTTGCTGGAGTTTCC	McrBC-qPCR
IDL10R	ACCCAACTCACAACACAGCA	
IDL11F	cgagactttattggagcttgaga	McrBC-qPCR
IDL11R	ccaaatttaacccaaattaacaaaa	
IDL12F	TCATTGGTCGTCATCCACAT	McrBC-qPCR, qRT-PCR
IDL12R	gatgttcgtacatgtttgattcg	
IDL13F	tgggcctactaatagggcatt	McrBC-qPCR
IDL13R	ggctatatggctactcccaaa	
IDL14F	cactgctctgataccatgtga	McrBC-qPCR
IDL14R	gcgatgtctcgtgaagatga	
IDL20F	CGAATTGAGCCAGAGTCAGA	McrBC-qPCR
IDL20R	CTCCATCTTGAGGGGGTGTA	
35SHSP20-F	cggggtaccATGAGTCTGTATAGCGACGG	Cloning *35S:LIL-GFP*
35SHSP20-R	atttgcggccgcaaAGAGTTTTTGTGTCGGATG	
mbd5-193 LP	ACGGACTTCTTTTTAGGCGTC	genotyping *mbd5*
mbd5-193 RP	TTTGTGTCACTATGTTGAAACATAATTC	
mbd6-927 LP	AAACAATTTCCACTCCCAATG	genotyping *mbd6*
mbd6-927 RP	TGCTTAAGCAGAACCAACCAC	
mbd7-09 LP	CAAGTATCTCTCAACGGCTGC	genotyping *mbd7*
mbd7-09 RP	TCTTCCTCTCCTTCGGCTAAC	genotyping *mbd7*
LBa1	tggttcacgtagtgggccatcg	genotyping *mbd*, RT-PCR of MBD5

Name	Oligonucleotide sequence	Purpose
mbd5-RT-F1	GATTACGGGTCGTAAGTTCC	RT-PCR of MBD5
mbd5-RT-R1	GCACATTCAGAAAATCGAAG	
mbd6-RT-F1	CCTCCCGATCCACTTCTC	RT-PCR of MBD6
mbd6-RT-R1	AACCAGAAATTCCGATTCC	
mbd6-RT-F2	TTTCAATCCAGACCATTTTG	
mbd6-RT-R2	TCGGCTTGTGATAAAGGTAA	
mbd7-RT-F1	CAGACGAGATCCTCTTCCTC	RT-PCR of MBD7
mbd7-RT-R1	GTCCCTGGCTCGATAAAGTA	
mbd7-RT-F2	CAGCAGCTTAGGGTTTTACA	
mbd7-RT-R2	CACCATAGAAACCGAGTCAA	
HSP20T7fl F	GGAATTCCATATGAGTCTGTATAGCGACG	Y2H, pGBKT7-LIL
HSP20T7fl R	CGGGATCCTTAAGAGTTTTTGTGTCGGATG	
HSP20T7N3F	CACCCATATGATGAGTCTGTATAGCGACGG	Y2H, pGBKT7-LIL$^{1-360}$
HSP20T7N3R	GGATCCTTAAACAACACCAACAGATGATCCT	
HSP20T7D1F	CACCCATATGGACATTGGTGTCAACAAGGTTG	Y2H, pGBKT7-LIL$^{361-440}$
HSP20T7D1R	GGATCCTTACGGATCAACTGGTCCCG	
MBD5ADfl F	CACCCATATGATGTCGAACGGCACGGAT	Y2H, pGADT7-MBD5
MBD5ADfl R	GGATCCCTAGAACATCGTTTTTCCAGCGT	
EcoRI-MBD5d	AGTgaattcCCAGGGGATGATAATTGG	Y2H, pGADT7-MBD5$^{29-110}$
MBD5d-BamHI	CGCggatccCTGCCGCCTAGCGCTTC	
MBD6ADfl F	CACCCATATGATGTCAGATTCTGTGGCCG	Y2H, pGADT7-MBD6
MBD6ADfl R	GGATCCTCAAGCCGACACTTTACTAGGGT	
MBD6d-BamHI	CGCggatccTCTATTGCTTCCTTGGC	Y2H, pGADT7-MBD6$^{75-153}$
EcoRI-MBD6d1	AGTgaattcTTCCGATTGCCTAGAGG	
MBD7ADfl F	CACCCATATGATGCAGACGAGATCCTCTTCC	Y2H, pGADT7-MBD7
MBD7ADfl R	GAATTC TTAAGAGCGGTCTTCGATCA	
EcoRI-MBD7d1	AGTgaattcTTCCGATTGCCTAGAGG	Y2H, pGADT7-MBD7$^{114-221}$
MBD7d1-BamHI	CGCggatccTAAGTATCTCTCAACGG	
EcoRI-MBD7d2	AGTgaattcCATTCCAAAGATTTCAGG	Y2H, pGADT7-MBD7$^{176-243}$
MBD7d2-BamHI	CGCggatccAAGAGGAAGCCGCTCC	

REFERENCES

1. Bird A (2007) Perceptions of epigenetics. Nature 447: 396-398.

2. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. Genes & Development 23: 781-783.

3. Lippman Z, Martienssen R (2004) The role of RNA interference in heterochromatic silencing. Nature 431: 364-370.

4. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11: 204-220.

5. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS ONE 3: e3156.

6. Zhou J, Wang X, He K, Charron J-B, Elling A, et al. (2010) Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in Arabidopsis reveals correlation between multiple histone marks and gene expression. Plant Molecular Biology 72: 585-595.

7. Law JA, Du J, Hale CJ, Feng S, Krajewski K, et al. (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498: 385-389.

8. Du J, Zhong X, Bernatavichute Yana V, Stroud H, Feng S, et al. (2012) Dual Binding of Chromomethylase Domains to H3K9me2-Containing Nucleosomes Directs DNA Methylation in Plants. Cell 151: 167-180.

9. Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, et al. (2007) The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. Current Biology 17: 379-384.

10. Qian W, Miki D, Zhang H, Liu Y, Zhang X, et al. (2012) A Histone Acetyltransferase Regulates Active DNA Demethylation in Arabidopsis. Science 336: 1445-1448.

11. Zemach A, Grafi G (2003) Characterization of Arabidopsis thaliana methyl-CpG-binding domain (MBD) proteins. The Plant Journal 34: 565-572.

12. Scebba F, Bernacchia G, De Bastiani M, Evangelista M, Cantoni R, et al. (2003) Arabidopsis MBD proteins show different binding specificities and nuclear localization. Plant Molecular Biology 53: 755-771.

13. Ito M, Koike A, Koizumi N, Sano H (2003) Methylated DNA-Binding Proteins from Arabidopsis. Plant Physiology 133: 1747-1754.

14. Zemach A, Li Y, Wayburn B, Ben-Meir H, Kiss V, et al. (2005) DDM1 Binds Arabidopsis Methyl-CpG Binding Domain Proteins and Affects Their Subnuclear Localization. The Plant Cell Online 17: 1549-1558.

15. Preuss SB, Costa-Nunes P, Tucker S, Pontes O, Lawrence RJ, et al. (2008) Multimegabase Silencing in Nucleolar Dominance Involves siRNA-Directed DNA Methylation and Specific Methylcytosine-Binding Proteins. Molecular cell 32: 673-684.

16. Jones PL, Jan Veenstra GC, Wade PA, Vermaak D, Kass SU, et al. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet 19: 187-191.

17. Nan X, Ng H-H, Johnson CA, Laherty CD, Turner BM, et al. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393: 386-389.

18. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, et al. (2003) The Methyl-CpG-binding Protein MeCP2 Links DNA Methylation to Histone Methylation. Journal of Biological Chemistry 278: 4035-4040.

19. Zhang Y, Ng H-H, Erdjument-Bromage H, Tempst P, Bird A, et al. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes & Development 13: 1924-1935.

20. Won SY, Li S, Zheng B, Zhao Y, Li D, et al. (2012) Development of a luciferase-based reporter of transcriptional gene silencing that enables bidirectional mutant screening in Arabidopsis thaliana. Silence 3: 6.

21. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell 133: 523-536.

22. Stroud H, Greenberg Maxim VC, Feng S, Bernatavichute Yana V, Jacobsen Steven E (2013) Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell 152: 352-364.

23. Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, et al. (2012) Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes & Development 26: 1825-1836.

24. Wierzbicki A, Haag J, Pikaard C (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135: 635 - 648.

25. Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu J-K, et al. (2010) A Protein Complex Required for Polymerase V Transcripts and RNA- Directed DNA Methylation in Arabidopsis. Current biology : CB 20: 951-956.

26. Pontes O, Li CF, Nunes PC, Haag J, Ream T, et al. (2006) The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. Cell 126: 79-92.

27. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, et al. (2003) Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. Science 301: 653-657.

28. Kleinboelting N, Huep G, Kloetgen A, Viehoever P, Weisshaar B (2012) GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. Nucleic Acids Research 40: D1211-D1215.

29. Li CF, Henderson IR, Song L, Fedoroff N, Lagrange T, et al. (2008) Dynamic Regulation of ARGONAUTE4 within Multiple Nuclear Bodies inArabidopsis thaliana. PLoS Genet 4: e27.

30. Pontes O, Vitins A, Ream TS, Hong E, Pikaard CS, et al. (2013) Intersection of Small RNA Pathways in Arabidopsis thaliana Sub-Nuclear Domains. PLoS ONE 8: e65652.

31. Baubec T, Ivánek R, Lienert F, Schübeler D (2013) Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. Cell 153: 480-492.

32. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes & Development 18: 2368-2379.

33. Smith MR, Willmann MR, Wu G, Berardini TZ, Möller B, et al. (2009) Cyclophilin 40 is required for microRNA activity in Arabidopsis. Proceedings of the National Academy of Sciences 106: 5424-5429.

34. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. Science 308: 118-120.

35. Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, et al. (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat Genet 37: 761-765.

36. Pontier D (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev 19: 2030-2040.

37. Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, et al. (2007) DNA demethylation in the Arabidopsis genome. Proceedings of the National Academy of Sciences 104: 6752-6757.

38. Earley KW, Haag JR, Pontes O, Opper K, Juehne T, et al. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. The Plant Journal 45: 616-629.

39. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14: R36.

40. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.

41. Lertpanyasampatha M, Gao L, Kongsawadworakul P, Viboonjun U, Chrestin H, et al. (2012) Genome-wide analysis of microRNAs in rubber tree (Hevea brasiliensis L.) using high-throughput sequencing. Planta 236: 437-445.

42. Li R, Yu C, Li Y, Lam T-W, Yiu S-M, et al. (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25: 1966-1967.

43. Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Research 7: 986-995.

44. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57: 289-300.

45. Rogers S, Bendich A (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Molecular Biology 5: 69-76.

46. Chen P-Y, Cokus S, Pellegrini M (2010) BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics 11: 203.

47. Becker C, Hagmann J, Muller J, Koenig D, Stegle O, et al. (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245-249.

48. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, et al. (2011) Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. Science 334: 369-373.

CONCLUSIONS AND PERSPECTIVES

My thesis research aimed to understand the molecular mechanisms underlying DNA methylation and transcriptional gene silencing (TGS) in the model plant *Arabidopsis thaliana*. In Chapter 1, thorough genetic characterizations of a luciferase (LUC)-containing transgenic plant, *LUCH*, established it as a reporter for DNA methylation and TGS and formed the basis for subsequent research reported in Chapters 2 and 3. From a forward genetic screen with *LUCH*, TATA-binding proteins-associated factor6 (TAF6) and a heat shock protein20 (HSP20) were identified as a positive and a negative factor, respectively, in DNA methylation and TGS. The studies of the two genes (in Chapters 2 and 3) contributed to a better understanding of how plants achieve a precise control of epigenetic modifications and also raised a number of open questions.

Characterization of an RdDM and TGS reporter, *LUCH*

LUCH was found to be a *LUC*-based TGS reporter that is regulated by RNA-directed DNA methylation (RdDM) and active demethylation through *ROS1*. In *LUCH*, the promoter of *LUC* is methylated and generates siRNAs. In addition, mutations in RdDM and *ROS1* alter the expression of *LUC* by affecting the levels of DNA methylation in the promoter of *LUC*. These molecular and genetic studies demonstrate that *LUCH* is a suitable system for genetic screens that aim to isolate mutations in genes involved in DNA (de)methylation and TGS. The moderate basal level of *LUC* expression makes it

suitable for the identification of both positive and negative regulators of DNA methylation and TGS.

Although transgene-specific siRNAs were not artificially introduced, for unknown reasons, they were spontaneously produced from the promoter of *LUC* and they likely subject *LUCH* to the regulation by RdDM. What initiated the biogenesis of transgene-specific siRNAs and DNA methylation in *LUCH*? Better understanding of how plants recognize this newly introduced transgene and regulate the expression of the transgene would help our research and applications with transgenic plants.

Identification of a positive factor in DNA methylation and TGS

TAF6 was identified as a positive factor that promotes DNA methylation and TGS. From a genetic screen with *LUCH*, a mutation in *TAF6* was found to release the TGS of *LUCH*. Moreover, *TAF6* appeared to promote cytosine methylation and TGS of several endogenous RdDM targets. Genome-wide methylation profiling revealed that hundreds of loci require *TAF6* for methylation at CHH contexts and most of them also depend on RdDM-specific RNA polymerases Pol IV and Pol V for DNA methylation. At several previously reported RdDM targets, *TAF6* is required for the production of Pol V-dependent long non-coding RNAs that recruit siRNAs to RdDM targets. Moreover, TAF6 was observed to partially colocalize with NRPE1, the largest subunit of Pol V, in nuclear foci. These findings indicate that the TAF6, which is well known for its function as a Pol II transcription factor, has been adapted to aid the Pol II homolog Pol V to participate in DNA methylation and TGS.

With regards to the function of TAF6, there are many questions to be answered. First of all, does *Arabidopsis* contain a specific cis-acting element that is recognized by TAF6 as *Drosophila* does? If so, what DNA sequence does TAF6 recognize to initiate Pol II-mediated transcription and which genes are regulated by TAF6 in *Arabidopsis*? For RdDM, how does TAF6 promote the action of Pol V? Does it recognize any specific DNA sequence in the RdDM targets?

Identification of a negative factor in DNA methylation and TGS

LOW IN LUCIFERASE EXPRESSION (*LIL*), which encodes a HEAT SHOCK PROTEIN 20 (HSP20) homolog, was identified as a negative factor in DNA methylation and TGS. Two independent genetic screens with *LUCH* and another *LUC*-based TGS reporter resulted in the isolation of two *lil* alleles, in which *LUC* was further silenced. In addition to the transgenic reporters, several endogenously methylated loci require the action of *LIL* to suppress their TGS. *LIL* was found to prevent hypermethylation at *LUCH* and endogenous targets and precisely localize the RdDM effector protein ARGONAUTE4 in the nucleus. In addition, LIL was observed to be localized in the nucleus. Moreover, LIL physically interacted with three methyl CpG-binding domain (MBD) proteins. MBD proteins are known to bind methylated DNA and recruit factors that deposit repressive histone marks at the underlying chromatin. Among the diverse classes of heat shock proteins, the HSP20 class is reported to prevent the aggregation of their interacting proteins. These findings suggest that LIL influences epigenetic modifications as a molecular chaperone of MBDs.

To gain a better understanding of LIL-mediated epigenetic regulation, LIL's molecular functions still need to be further investigated. First of all, does LIL directly associate with chromatin to prevent cytosine methylation and TGS? If so, does DNA methylation affect the binding efficiency of LIL? With regards to the action of MBD proteins, does LIL regulate the pool of MBD proteins as a protein chaperone or directly affect the function of MBD proteins? Does LIL enhance or suppress the ability of MBD proteins to bind methylated DNA? Lastly, does LIL impact other epigenetic marks such as histone modifications by affecting the functions of MBD proteins?