UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mechanisms of Small RNA Degradation and Characterization of THO Complex Mutants in *Arabidopsis*

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

Doctor of Philosophy

in

Plant Biology

by

Yuanyuan Zhao

August 2013

Dissertation Committee:
Dr. Xuemei Chen, Chairperson
Dr. Julia Bailey-Serres
Dr. Shou-Wei Ding

The Dissertation of Yuanyuan Zhao is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

I would like to express my appreciation to everyone who provided me help and support to make this work possible.

First and foremost, I am deeply grateful to my advisor and dissertation Chair Dr. Xuemei Chen. I am indebted to her for her continuous guidance and constant encouragement throughout my PhD study. Her patience, encouragement and thoughtful advices helped me through many challenging times. From her, I learned how to think critically and work efficiently as a scientist. I deeply appreciate the mentoring and support she provided that would benefit my future career.

I would also like to thank the other members of my dissertation committee: Dr. Shouwei Ding and Dr. Julia Bailey-Serres for their valuable comments to my research and dissertation. In addition, I would like to thank my qualify examination committee members, Dr. Patricia Springer, Dr. Renyi Liu, Dr. Harley Smith, Dr. Hailing Jin, Dr. Howard S Judelson and my guidance committee member Dr. Jiankang Zhu for their time and insightful suggestions to keep me on track.

Furthermore, I would like to express my gratitude to all professors who taught me during my graduate study at UC Riverside and my undergraduate study at China Agricultural University. I am also indebted to Dr. Xiujie Wang at the Institute of Genetics and Developmental Biology in Beijing for giving me the first opportunity to do research in college.

I also wish to thank every member in Dr. Chen's lab for offering help in the lab. I would particularly like to thank Dr. Shengben Li for teaching me a lot of experiments at the beginning of my PhD research and for the daily discussions in research. I would like to thank the senior students Drs. YunJu Kim, Lijuan Ji and Theresa Dinh for their valuable advice and help. Furthermore, I would like to thank a fellow student SoYoun Won who has accompanied me

throughout my five years of study and research. I would also like to thank two undergraduate students Tammy Luu and Renee Zamore who provided technical assistance to my research.

I sincerely appreciate the contributions from all collaborators and the kind help from all those who provided materials to my work.

The first Chapter of my dissertation was reprinted from two published papers with some rearrangements and minor alterations. I gratefully acknowledge permission to use the items from the following publishers: for permission to use a paper published in Current Biology (Curr Biol. 2012 Apr 24;22(8):689-94), Cell press; for permission to use a paper published in RNA biology (RNA Biol. 2012 Oct;9(10):1218-23), Landes Bioscience.

Many friends including my lab mates have helped me in my life in the past five years. I would like to thank Yu Yu, Lin Liu and all the other friends for their accompany. I deeply appreciate their care and support and I greatly value their friendship.

Last but not least, I want to thank my family for their unconditional love and support. I would especially like to thank my parents for motivating me when I had tough times and for being a constant source of love, concern and support all these years.

Author Contributions

The following is a document stating the author contributions for each part of my thesis.

For Chapter One, a former postdoc Dr. Vanitha Ramachandran performed the crosses between nine nucleotidyl transferase mutants and the *hen1-8* mutant. I performed the cross between the *heso1-1* mutant and the *hen1-8* mutant. I performed the genotyping for all mutant crosses and isolated *heso1-1* as a *hen1* suppressor. I constructed the complementation vector for *HESO1* and performed plant transformation and screened transgenic plants. I performed the real-time RT PCR for miRNA targets. A fellow graduate student Yu Yu performed small RNA northern blots and beta-elimination essay. The small RNA libraries were constructed by Yu Yu and all the sequencing data was analyzed by a graduate student Jixian Zhai from Dr. Blake Meyers' lab. I performed AGO1 immunoprecipitation and small RNA northern blots. I also constructed vectors for *in vitro* enzymatic activity assays, purified the recombinant proteins and performed the enzymatic activity assays.

For Chapter Two, the EMS mutagenesis of *LUCH* and *YJ11-3F* were performed by two postdocs, Dr. Binglian Zheng and Dr. YunJu Kim, respectively. I performed the T-DNA insertional mutagenesis of *YJ11-3F*. The *tex1* mutant line *28-6L* was isolated by Dr. Shengben Li. The *hpr1* mutant line *P204R* was isolated by Dr. YunJu Kim. The *tho5a* mutant line *TL525L* was isolated by an undergraduate student Tammy Luu. I identified the mutation site in *28-6L*. A fellow graduate student Shaofang Li identified the mutation site in P204R. Tammy Luu and I together identified the mutation site in *TL525L*. I constructed complementation vectors for *HPR1* and *THO5A*, transformed mutant lines and screened transgenic plants. I performed the genetic interaction analysis between THO mutants and RdDM mutants. I also performed RT-PCR, Real-time PCR and small RNA northern blots. The small RNA libraries were constructed by a visiting

student Dongming Li. Dr. Bob Schmitz from Dr. Joseph Ecker's lab constructed the bisulfite sequencing libraries for the first biological replicates for preliminary analysis. I constructed the mRNA sequencing libraries and all the other bisulfite sequencing libraries including the ones mentioned in Chapter Two and Appendix B. Most high throughput sequencing data was analyzed by a postdoc Dr. Lei Gao under the direction of Dr. Xuemei Chen and me.

For Appendix A, I isolated the mutant line and performed rough mapping and constructed DNA sequencing library of the mutant. Shaofang Li analyzed the DNA sequencing data and identified the mutation site. I performed the real-time PCR and the 5-aza-2' deoxycytidine treatment. Dr. Lei Gao analyzed the bisulfite sequencing data.

For Appendix C, the antibody cell lines were cultured by a former graduate student Dr. Lijuan Ji before ascites production. I purified the antibodies from ascites and performed all the experiments. Dr. Lei Gao processed the sequencing data.

ABSTRACT OF THE DISSERTATION

Mechanisms of Small RNA Degradation and Characterization of THO Complex Mutants in *Arabidopsis*

by

Yuanyuan Zhao

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

The expression of genes in eukaryotes is regulated at multiple layers. Transcriptional regulation and post-transcriptional regulation affect the mRNA levels of genes prior to protein translation. In my thesis work, I have worked on two research projects to study gene silencing at transcriptional and post-transcriptional levels in *Arabidopsis.* First, through a reverse genetic approach, I identified a nucleotidyl transferase HESO1 that functions in the degradation of microRNAs in the *Arabidopsis hen1* mutant. As important regulators of gene expression, the accumulation of miRNAs affects the levels of their target mRNAs. I found that a mutation in *HESO1* partially rescued the *hen1* developmental defects, which result from decreased miRNA levels. The tailing of unmethylated miRNAs in *hen1* by HESO1 promoted their degradation by an unknown nuclease, which differed from the one known to cause miRNA 3'-end truncation in *hen1*. I also confirmed the enzymatic activity of the HESO1 protein *in vitro* and found that 2'-*O*-methylation at the 3'-end of small RNAs completely blocked the activity of HESO1. These

findings unveiled the protein implicated in small RNA tailing in *hen1* and contributed to the knowledge of miRNAs degradation in *Arabidopsis*. Second, from forward genetic screens using two luciferase reporter lines that are regulated by DNA methylation, I isolated mutants with decreased luminescence and cloned three genes encoding subunits of the THO complex — *TEX1*, *HPR1* and *THO5A*. I found that DNA methylation was not affected in the mutants at either the *LUCIFERASE* transgene locus or most endogenous loci known to harbor DNA methylation. From mRNA sequencing, I identified genes that were up-regulated or down-regulated in *tex1* or *hpr1*. I found that genes associated with dispersed repeats and inverted repeats were enriched among the down-regulated genes in *tex1* and *hpr1*. Additionally, genes with H3K27me3 were enriched in both the up-regulated and the down-regulated genes in *tex1* and *hpr1*. These findings provided insights into the potential targets of THO complex in *Arabidopsis*.

Table of Contents

List of Figures

List of Tables

Chapter 1
Mechanisms of small RNA degradation

Abstract

Small RNAs are 20-30nt non-coding RNA molecules that regulate many biological processes in eukaryotes. They transcriptionally or post-transcriptionally repress the expression of their targets in a sequence specific manner. Although the biogenesis pathway of small RNAs is well characterized in plants, the mechanism of small RNA degradation is largely unclear. The small RNA methyltransferase HEN1 methylates small RNA at the 3' terminal nucleotide. In *Arabidopsis hen1* mutant, small RNAs with 3' end truncation and 3' end untemplated tailing have been observed, and these species are hypothesized to be degradation intermediates. Here we identified a nucleotidyl transferase gene *HESO1* that functions in small RNA degradation in the *Arabidopsis hen1* mutant. HESO1 recombinant protein shows nucleotidyl transferase activity on unmethylated small RNAs *in vitro* and its activity is completely blocked by small RNA methylation. Compared with *hen1,* the *hen1 heso1* double mutant shows elevated small RNA accumulation and lower levels of small RNA tailing, suggesting that *HESO1* functions in adding tails to unmethylated small RNAs to promote their degradation *in vivo*. In the *hen1 heso1* double mutant, although 3' tailing is reduced, 3' truncation is not obviously affected, suggesting that 3' truncation does not depend on 3' tailing. Given that *HESO1* promotes the degradation of unmethylated small RNA without affecting small RNA 3' truncation, we hypothesize that a nuclease that is different from the one that causes 3' truncation of small RNAs functions in degrading the tailed small RNAs.

Introduction

In eukaryotes, the regulation of gene expression is controlled at multiple levels through various regulators. Among these regulators, small RNAs are a type of negative regulators that recognize their targets through sequence complementarity and silence their targets at transcriptional or post-transcriptional levels. There are three major types of small RNAs: microRNA (miRNA), small interfering RNA (siRNA) and piwi-interacting RNA (piRNA), and they are defined by their distinct biogenesis pathways. MiRNAs are generated from a single-stranded RNA precursor with a stem-loop structure by the RNase III endonuclease Dicer. SiRNAs are also generated by Dicer, but from a long double-stranded RNA precursor. Although the biogenesis of piRNAs are not clearly understood, it is proposed to be generated from a long single-stranded RNA precursor in a Dicer-independent manner (reviewed in [1]).

MiRNAs and siRNAs exist in plants and they play important roles during the life of a plant. MiRNAs silence their targets at the post-transcriptional level. By down-regulating their target genes, miRNAs are involved in various biological processes. Since many miRNA pathway mutants exhibit pleiotropic developmental defects, it is not surprising that one important impact of miRNAs in plants is the regulation of developmental processes. In *Arabidopsis*, many key transcription factors that function during development are miRNA targets [2-7]. In addition, miRNAs are also involved in other processes like abiotic stress response [8-10] and plant immune response [11]. Plant siRNAs consist 21-nt trans-acting siRNAs (tasiRNA) and 24-nt heterochromatic siRNAs. The 21-nt tasiRNAs originate from TAS transcripts and act on their targets *in trans* in a similar way as miRNAs; while 24-nt siRNAs are generated mainly from transposons and repeats, and they silence their target DNA at the transcriptional level to maintain genome stability [12].

Because many protein-coding genes involved in various biological processes are targeted by miRNAs, it is important to maintain the miRNAs at proper levels to ensure that their target mRNAs levels are well controlled. The regulation of miRNA abundance should involve both biogenesis and degradation.

Biogenesis of miRNAs

MiRNAs are the final products of non-coding RNA genes. Following transcription, the primary transcript (pri-miRNA) is processed into the hairpin-structured precursor (pre-miRNA), which is further processed into the mature miRNA, by RNase III-type nucleases [13]. In animals, the maturation of the miRNA from the pri-miRNA involves the nuclear enzyme Drosha [14] and the cytoplasmic enzyme Dicer [15, 16]. In *Arabidopsis*, the formation of the pre-miRNA from the pri-miRNA and the processing of the pre-miRNA to the mature miRNA are both performed by the Dicer-like protein, DCL1 [17-19]. The processing of the pre-miRNA by Dicer or DCL1 yields the miRNA/miRNA* duplex, which has a 2-nt overhang at the 3' end of each strand, as well as a 5' phosphate (P) and a 3' OH on each strand [15, 18]. The accumulation of miRNAs in *Arabidopsis* requires a protein named HEN1 [19]. HEN1 is a methyltransferase that acts on the miRNA/miRNA* duplex to deposit a methyl group onto the 2' OH of the 3' terminal ribose on each strand [20]. *In vitro* methylation assays [20] and structures of HEN1 in complex with a miRNA/miRNA* duplex [21] indicate that the two termini of the miRNA/miRNA* duplex are separately and independently methylated. Therefore, HEN1 does not distinguish the miRNA strand from miRNA* or contribute to strand selection. After methylation, the miRNA/miRNA* duplex is loaded into the RNA-induced silencing complex (RISC) that contains the ARGONAUTE1 (AGO1) protein as the core component, where the miRNA, after shedding of the miRNA* strand, guides the cleavage or translation repression of its target mRNAs through base-pairing with specific targets [22-25].

2'-*O*-methylation increases miRNA stability

The abundance of a certain miRNA in cells is governed by its rates of biogenesis and degradation. Studies on one of the biogenesis steps in plants, miRNA 2'-*O*-methylation, provided the first glimpse into processes that lead to miRNA degradation. Methylation on the ribose of the last nucleotide by HEN1 is a universal step in the biogenesis of miRNAs in plants [20, 26]. In *Arabidopsis*, almost all miRNAs are fully 2'-*O*-methylated at their 3' ends *in vivo*. In order to understand the function of miRNA methylation, northern blots were conducted to examine various miRNAs in wild type and *hen1* mutants. It was found that miRNAs are reduced in abundance and heterogeneous in size when they are unmethylated [27]. Primer extension experiments showed that the size heterogeneity resides in the 3' ends of miRNAs [27]. Cloning and sequencing (by the traditional Sanger method) of miR173 and miR167 in wild type and *hen1* mutants revealed that miRNAs undergo 3' truncation and 3' uridylation, the addition of a short U-rich tail [27]. This led to the hypothesis that the 2'-*O*-methyl moiety in plant miRNAs promotes their stability by protecting them from a 3'-to-5' exonucleolytic activity and a uridylation activity [27]. In addition to miRNAs, siRNAs in plants and *Drosophila* and piRNAs in animals also undergo 2'-*O*-methylation by HEN1 or HEN1 orthologs [27-35]. In both plants and animals, the lack of small RNA methylation is often associated with small RNA instability as inferred from reduced accumulation, 3' truncation, and 3' uridylation [27, 28, 35, 36].

siRNAs negatively influence HEN1-mediated miRNA methylation

siRNAs are similar to miRNAs in structure, biogenesis and function. But unlike miRNAs, which come from hairpin pre-miRNAs, siRNAs originate from long double-stranded RNA precursors [37, 38]. Both miRNAs and siRNAs in plants carry a 2'-*O*-methyl group on the 3' -terminal nucleotide, a modification introduced by the methyltransferase HEN1 [20, 27]. siRNAs

compete with miRNAs for methylation in *Arabidopsis* when HEN1 function is compromised [39]. This conclusion was drawn from studies of mutations in DNA-dependent RNA polymerase IV or RNA-dependent RNA polymerase 2, both of which are essential for the biogenesis of endogenous 24-nt siRNAs [40-42]. These mutations can rescue the defects in miRNA methylation of *hen1-2*, a weak *hen1* allele, revealing the negative influence of siRNAs on HEN1-mediated miRNA methylation [39]. These mutations do not exhibit large effects on miRNA levels in wild type, suggesting that HEN1 activity is not limiting. However, HEN1 activity may be rendered limiting under conditions when large quantities of siRNAs are produced.

SDNs degrade mature miRNAs

It is now known that multiple enzymes, such as 3' and 5' exonucleases, can engage in miRNA degradation. For example, in *Caenorhabditis elegans*, XRN-2 (a 5'-to-3' exonuclease) is involved in the degradation of mature miRNAs [43]. SDNs, a family of 3'-to-5' exonucleases encoded by the *SMALL RNA DEGRADING NUCLEASE* (*SDN*) genes, degrade mature miRNAs in *Arabidopsis* [44]. SDN1 has specificity for short, single-stranded RNAs and does not digest small RNA duplexes, pre-miRNAs, or longer RNAs *in vitro* [44]. Knock-down of multiple SDN genes in *Arabidopsis* results in elevated miRNA levels and pleiotropic developmental abnormalities [44]. This suggests that maintaining proper miRNA levels through miRNA turnover is crucial to plant development. Are SDNs the enzymes responsible for the production of the 3' truncated species in *hen1* mutants? This has not yet been determined. If they were responsible for the production of these 3' truncated species, knocking down *SDN*s in a *hen1* mutant background would reduce the levels of 3' truncated species, which can be determined through small RNA high-throughput sequencing. It is known, however, that the *SDN*s impact the steady-state levels of miRNAs. Knocking down multiple members of the *SDN* gene family using an artificial miRNA results in elevated miRNA accumulation [44]. Since miRNA species are

nearly all methylated in wild type [20, 27], the effect of *SDN* knockdown on miRNA accumulation in the wild type background suggests that SDNs can degrade methylated miRNAs. Biochemical assays show that SDN1 can degrade 2'-*O*-methylated miRNAs *in vitro*, although at a lower efficiency than unmethylated miRNAs [44].

Small RNA tailing at 3'-terminal nucleotide in eukaryotes

Although the protein that uridylates unmethylated miRNAs is not known in *Arabidopsis*, several nucleotidyl transferases have been reported to be implicated in small RNA uridylation in other organisms. In *C. elegans*, one nucleotidyl transferase PUP-2 uridylates let-7 pre-miRNA and promotes its turnover [45]; another nucleotidyl transferase CDE-1 uridylates siRNAs that are bound by the Argonaute protein CSR-1 [46], in *cde-1* mutant, siRNAs loses untemplated 3' terminal residues and are over accumulated [46]. In addition, In *Chlamydomonas reinhardtii*, a nucleotidyl transferase MUT68 uridylates miRNAs and siRNAs and stimulate the degradation of these small RNAs by the RRP6 exosome [47]. These proteins contain the nucleotidyl transferase domain and exhibit homology to the yeast non-canonical poly(A) polymerase TRF4, which polyadenylate aberrant RNAs to stimulate their degradation [48]. The function of these nucleotidyl transferases from other eukaryotes suggests the presence of a similar protein that uridylates small RNAs and stimulates their degradation in *Arabidopsis.*

In order to gain insights into the small RNA degradation process, we tried to identify the gene that functions in uridylating miRNAs in *hen1* mutant and study its impact on small RNA degradation. Here we report the identification of a nucleotidyl transferase gene *HESO1* in *Arabidopsis* that tails miRNAs and siRNAs in *hen1* mutant background. The *hen1 heso1* double mutant show increased miRNA abundance and decreased miRNA tailing level compared with *hen1* mutant. Decreased tailing level in *hen1 heso1* does not affect the accumulation of truncated

miRNA species. In addition, the HESO1 recombinant protein exhibit nucleotidyl transferase activity *in vitro*. We proposed that HESO1 is a nucleotidyl transferase that tails unmethylated miRNAs and siRNAs in *hen1* mutant. Tailing of small RNAs stimulates their degradation through a nuclease that is different from the one that causes small RNA 3' truncation.

Results

Loss of function in *HESO1* suppresses the morphological defects of a partial loss-of-function *hen1* mutant

We hypothesized that a terminal nucleotidyl transferase uridylates miRNAs in *Arabidopsis hen1* mutants and that uridylation serves as a signal to trigger miRNA degradation. Hence, we predicted that loss of function in this nucleotidyl transferase gene would lead to increased miRNA accumulation in *hen1* mutants and may rescue the morphological phenotypes associated with reduced miRNA accumulation in *hen1*. To identify the nucleotidyl transferase, we first searched for *Arabidopsis* proteins with sequence similarities to known yeast TRF4-like nucleotidyl transferases with uridylation activity such as the human TUT4 [49], *C. elegans* CDE1 [46] and *Chlamydomonas* MUT-68 [47]. Ten putative *Arabidopsis* nucleotidyl transferases of unknown biological function were identified. IDs of the ten putative *Arabidopsis* nucleotidyl transferases and the evolutionary distances between the nucleotidyl transferase (NT) domains from the ten *Arabidopsis* proteins are shown in Figure 1.1 together with the NT domain from poly(A) polymerases and terminal uridylyl transferases in other eukaryotes.

Next, we obtained T-DNA mutants in each of the ten putative nucleotidyl transferase genes (Table 1.1) and crossed each mutant with *hen1-8*, a partial loss-of-function *hen1* mutant in the Col accession [39], the same accession in which the nucleotidyl transferase mutants were generated. We found that only the mutation in *At2g39740* (Figures 1.2A and Figure 1.2B)

partially rescued the morphological phenotypes of *hen1-8* (Figure 1.3A and Figure 1.3B). Therefore, we named this gene *HEN1 SUPPRESSOR1 (HESO1)* and this mutant allele is referred as *heso1-1*. The *hen1-8 heso1-1* double mutant was larger and had better fertility than the *hen1-8* single mutant (Figure 1.3A and Figure 1.3B). To ensure that the *heso1-1* mutation was solely responsible for the phenotypic rescue of *hen1-8*, we introduced a *HESO1* genomic fragment fused in-frame to *GFP* (*pHESO1:HESO1-GFP*) into the *hen1-8 heso1-1* double mutant. Many independent transgenic lines showed phenotypes that resembled *hen1-8* mutants (Figure 1.3B), indicating that it was the *heso1-1* mutation that suppressed the defects of *hen1-8*. Some transgenic lines showed phenotypes even more severe than *hen1-8* mutants (data not shown). It is possible that these lines had higher levels of HESO1 expression, which enhanced *hen1-8* defects.

Loss of function in *HESO1* results in an increase in miRNA abundance

Suppression of the morphological defects of *hen1-8* by *heso1-1* suggested that miRNA activities were partially rescued by *heso1-1*. One activity of plant miRNAs is the cleavage of their target transcripts to lead to their degradation. We examined the levels of four miRNA target transcripts in wild type (WT), *hen1-8*, and *hen1-8 heso1-1* by real-time RT-PCR. As expected, in *hen1-8*, these transcripts accumulated to higher levels as compared to WT (Figure 1.4A), due to the reduced levels of miRNAs [39]. Indeed, the increase in the transcript levels in *hen1-8* was partially or fully rescued by *heso1-1* (Figure 1.4A).

Next, we tested whether the *heso1-1* mutation resulted in higher abundance of miRNAs. Northern blotting was performed to detect miR166, miR167, and miR173 in WT, *heso1-1*, *hen1-8*, and *hen1-8 heso1-1*. Like other *hen1* alleles [27], the *hen1-8* mutation resulted in both a reduction in miRNA levels and the accumulation of heterogeneous species—bands corresponding to 3' truncated as well as 3' tailed species were present (Figure 1.4B). The *heso1-1* mutation did not

affect miRNA accumulation in the WT background, but it resulted in an obvious increase in the abundance of the heterogeneous miRNA species in the *hen1-8* background (Figure 1.4B). The increase in miRNA levels in the double mutant was rescued by the *pHESO1::HESO1-GFP* transgene (Figure 1.4C). For miR167 and miR173, it was also apparent that the increase in abundance was restricted to normal-sized species, 3' truncated species, or species with short tails (Figure 1.4B). This suggested that 3' tailing was reduced in the *hen1-8 heso1-1* double mutant relative to *hen1-8*.

Previous studies show that mutations in RNA polymerase IV (Pol IV) subunit genes, *NRPD1* and *NRPD2*, rescue the morphological, as well as miRNA accumulation, defects of a weak *hen1* allele, *hen1-2*, which harbors the same molecular lesion as *hen1-8* but is in the Landsberg accession [39]. Pol IV is a key factor in the biogenesis of a predominant class of endogenous small interfering RNAs (siRNAs), the 24 nucleotide (nt) siRNAs [41, 42, 50]. As natural substrates of HEN1 [27], the 24 nt siRNAs compete with miRNAs for methylation by the mutant HEN1-2 protein such that elimination of these siRNAs in *nrpd1* or *nrpd2* mutants allows miRNAs to gain methylation and thus increase in abundance in the *hen1-2* background. With β-elimination followed by northern blotting [20], miR165/166 and miR173 were found to be unmethylated in both *hen1-8* and *hen1-8 heso1-1* (Figure 1.5). Therefore, unlike *nrpd1* or *nrpd2* mutations, the *heso1-1* mutation led to an increase in miRNA accumulation without affecting their methylation status.

Loss of function in *HESO1* results in reduced 3' uridylation without affecting 3' truncation

Although northern blotting suggested that the *heso1-1* mutation caused a reduction in 3' tailing, the concomitant increase in miRNA accumulation made it difficult to determine the degree of reduction in 3' tailing. To obtain a global and quantitative view of 3' modifications of

miRNAs, we constructed small RNA libraries from WT, *heso1-1*, *hen1-8*, and *hen1-8 heso1-1* and subjected them to deep sequencing. Two biological replicates were performed for each genotype. An algorithm was developed to classify reads corresponding to a known miRNA into four categories: full-length, 3' truncated only, 3' tailed only, and 3' truncated and tailed (see materials and methods).

We first focused on the status of 3' tailing of miRNAs. For each biological replicate, 44 miRNAs that were represented by at least 50 transcripts per million in all four libraries were included in the analysis. 3' tailing was quantified for each miRNA as a percentage of 3' tailed only reads + 3' truncated and tailed reads divided by total reads. In this analysis, the lengths of the tails were not taken into account—as long as the read had at least one nontemplated nucleotide, it was considered a tailed species. For each of the miRNAs, representatives of which are shown in Figure 1.6A (replicate 1) and Figure 1.6B (replicate 2), the percentage of tailed species was reduced in *hen1-8 heso1-1* relative to *hen1-8*, indicating that *HESO1* was partially responsible for 3' tailing in *hen1-8*. The *heso1-1* mutation did not affect the tailing status of most miRNAs in the WT background, with the exception of miR319a, which showed higher levels of tailing relative to other miRNAs in WT (Figure 1.6A and Figure 1.6B).

We next examined the nature of the nontemplated nucleotides in the tails and the lengths of the tails. For each miRNA, all tailed reads (both full-length species with tails and 3' truncated species with tails) were categorized by tail length, and the frequencies of the four nucleotides in the tails were calculated. Although the tailing patterns differed to some extent among the miRNAs, two observations were obvious (Figure 1.7). First, uridine was the predominant nucleotide of the tails in *hen1-8* as previously reported for other *hen1* alleles [27]. Second, there was a shift toward shorter tails in *hen1-8 heso1-1* as compared to *hen1-8*. For miR166a and many other miRNAs, tails of 1 to 7 nt were found in *hen1-8*, but 1 nt tails were the most predominant in

hen1-8 heso1-1 (Figure 1.7). These results demonstrated that *HESO1* was responsible for 3' uridylation of miRNAs in *hen1*.

We next examined the status of 3' truncation in relationship to 3' tailing by *HESO1* for all 44 miRNAs abundantly represented in the libraries. We found similar results for nearly all miRNAs; ten representative miRNAs from the two biological replicates are shown in Figure 1.8 (Figure 1.8-1 to Figure 1.8-5). In WT and the *heso1-1* single mutant, in which miRNA methylation was unaffected, most miRNA reads belonged to the full-length category. In *hen1-8*, the lack of methylation led to an increase in truncated only, tailed only, and truncated and tailed species. In *hen1-8 heso1-1*, 3' tailing was greatly reduced, such that species with 0 or 1 nt tails were most abundant. Intriguingly, reads representing 3' truncation were largely unaffected by the *heso1-1* mutation, which was inconsistent with the hypothesis that 3' uridylation triggers 3' truncation.

Endogenous siRNAs are also 3' uridylated in *hen1* mutants [27] . To determine whether HESO1 was responsible for the uridylation of 24 nt siRNAs, we focused on three abundantly represented 24 nt siRNAs (Table 1.2). All three siRNAs were truncated and tailed in *hen1-8*. In *hen1-8 heso1-1*, an obvious reduction in 3' tailing was observed, but 3' truncated species were still present (Figure 1.8-6 and Figure 1.8-7). Therefore, siRNAs and miRNAs were similarly affected by loss of function in *HESO1*.

miRNAs with 3' truncation and tailing in *hen1* are associated with AGO1

Based on the observations that *heso1* mutation increases miRNA accumulation and miRNA activity in *hen1* mutant without affecting miRNA 3' methylation and 3' truncation, it is possible that unmethylated miRNAs in *hen1* mutant are bound by AGO proteins and are functional in silencing their target mRNAs. To test this possibility, we immunoprecipitated

AGO1 from WT, *hen1* and *hen1 heso1*. Small RNA was isolated from immunoprecipitates and miRNA 165/166 and miRNA 167 were tested by northern blots. Similar to the patterns in total RNA samples, the AGO1 associated miRNAs in *hen1* and *hen1 heso1* also exhibit size heterogeneity (Figure 1.9), suggesting that most miRNAs with 3' truncation and 3' tailing are associated with AGO1 protein. Together with the result that *heso1* mutation increases miRNA activity in *hen1* mutant, it is likely that unmethylated miRNAs without or with 3' truncation and 3' tailing could function in recognizing and down-regulating their target mRNAs.

HESO1 exhibits terminal nucleotidyl transferase activity

The above genetic evidence suggested that *HESO1* is responsible for 3' uridylation of unmethylated miRNAs and siRNAs in *hen1* mutants. We sought biochemical evidence that HESO1 uridylated small RNAs with a recombinant His-HESO1 protein produced in *E. coli*. His-HESO1 was incubated with an RNA oligonucleotide corresponding to unmethylated miR173 that was radiolabeled at its 5′ end. HESO1 was able to add a U tail to miR173 that lengthened with time and reached up to approximately 60 nt after 40 min of incubation (Figure 1.10A, lanes 1–6). To rule out that the terminal nucleotidyl transferase activities were due to a contaminating protein from *E. coli*, we mutated two aspartate residues in His-HESO1 that are part of the metal binding pocket in the nucleotidyl transferase domain to alanine (Figure 1.10B). The mutant His-HESO1m protein was similarly purified and assayed for activity. The His-HESO1m protein was unable to add a U tail to unmethylated miR173 (Figure 1.10A, lanes 13 and 14).

Next, we examined the enzymatic properties of HESO1. HESO1 was able to use ATP, CTP, or GTP, but the tail lengths with these nucleotides were much shorter than U tails, indicating that HESO1 preferred uridine (Figure 1.10A, lanes 7–9). When dATP was used in the reaction, only one nucleotide was added (Figure 1.10A, lane 10), suggesting that the 3' OH of the

substrate RNA was the site of phosphodiester bond formation. More importantly, when 2′-*O*-methylated miR173 was incubated with HESO1 in the presence of UTP, no tailing was observed (Figure 1.10A, lanes 15–18). This indicated that HESO1 activity was completely inhibited by 2′-*O*-methylation.

Discussion

Genetic and biochemical evidence in this study supports HESO1 as the nucleotidyl transferase that uridylates miRNAs and siRNAs when they lack 2′-*O*-methylation. The large increase in miRNA abundance in *hen1-8 heso1-1* versus *hen1-8* demonstrates that uridylation leads to degradation. The fact that the *heso1-1* mutation caused a large reduction in 3’ uridylation without strong effects on 3’ truncation suggests that 3’ truncation observed in *hen1* mutants is largely independent of 3’ uridylation. Two implications from these observations are as follows: (1) uridylation triggers the degradation of small RNAs by a nuclease distinct from the one that causes 3’ truncation, and (2) if the nuclease that degrades uridylated small RNAs is a 3’-to-5′ exonuclease, then it should be highly processive such that few or no degradation intermediates (i.e., 3’ truncated species) would be present *in vivo*. In *C. reinhardtii*, it has been shown that siRNA uridylation by the nucleotidyl transferase MUT68 stimulates its degradation by the RNA exosome subunit RRP6 [47]. Based on the conservation of cooperation between TRF4 like nucleotidyl transferase and the exosome in RNA degradation in eukaryotes [47, 48], one possibility is that the exosome also mediates the degradation of uridylated small RNAs in plants. A previous study on the *Arabidopsis* exosome target RNAs revealed that the degradation of pri-miRNA processing intermediates depends on the exosome, however, the levels of mature miRNAs are not affected in the *RRP4* or *RRP41* knockdown plants [51], suggesting that the exosome does not have a strong impact on miRNA accumulation in *Arabidopsis* wild type plants.

Recently, a RNase II-type exonuclease Dis3L2 that preferentially degrades uridylated RNAs has been identified in *Schizosaccharomyces pombe* [52]. In human cells, Dis3l2 mediates the degradation of uridylated pre-let-7 miRNA [53]. Intriguingly, Dis3L2 does not interact with exosome components and co-localizes with processing bodies (P-bodies) in yeast [52]. P-bodies are cytoplasmic foci where mRNA turnover occurs and miRNA-mediated mRNA translation repression may also occur at P-bodies [23] [54]. The Dis3l2 homolog in *Arabidopsis* SOV [52] is also located at distinct cytoplasmic foci that resemble P-bodies [55]. Further studies on the targets of exosome subunit genes and SOV in *Arabidopsis* wild type as well as *hen1* mutants may help to understand more about small RNA degradation downstream of uridylation.

The presence of truncated and tailed miRNAs in AGO1 protein in *hen1* mutants suggests the possibility that 3' truncation and 3' tailing of unmethylated miRNAs occur while they are bound by AGO1. The PAZ domain of AGO1 binds to the 3' end of miRNAs. A structural study of a *Thermus thermophilus* AGO protein complexed with 21-nt guide DNA and a fully complementary 15-nt target RNA suggests that during target recognition, the 3' end of the guide strand is released from the PAZ domain [56]. In D*rosophila*, miRNAs are unmethylated and have low levels of sequence complementarity with their target mRNAs at the 3' end. Introduction of an artificial RNA target with extensive sequence complementarity to a certain miRNA leads to the 3' truncation and tailing of the miRNA [36]. It is likely that low sequence complementarity between miRNA and its target RNA prevents the release of 3' end from the PAZ domain so that the 3' end of miRNA is not accessible by nucleotidyl transferases and 3' to 5' exonucleases in animals. In plants, miRNAs are highly complementary with their target RNAs, during the recognition between miRNA and its target, the extensive sequence complementarity may lead to the release of the 3' end. Without the protection from the methyl group as in the case of *hen1* mutants, the 3' end of miRNAs may be targeted by 3' modifiers like SDN1 and HESO1 while they are still

bound by AGO1. Although direct interaction between AGO1 and HESO1 has not been reported, a recent work indicated that in *hen1* mutant background, mutations in *AGO1* suppressed the 3' truncation and tailing of miRNAs [57], suggesting that miRNA 3' modifications are AGO1-dependent. Based on the evidence mentioned above, it is conceivable that the small RNA 3' modifiers act on AGO1-bound miRNAs in *hen1*. The miRNAs with 3' short truncation and tailing may still be bound by AGO1 and are relatively stable and even functional. However, it is possible that after the small RNA is tailed by HESO1 to a certain length that exceeds the binding capacity of AGO1, it may be released from the RISC complex and degraded by a nuclease with high processivity. However, the possibility that HESO1 functions on free small RNAs still cannot be excluded. A recent study on small RNA release showed that introduction of a target RNA with a high degree of sequence complementarity to a miRNA promotes the specific release of the miRNA from AGO2 protein [58]. Unlike animal miRNAs, the sequence complementarity between plant miRNAs and their target mRNAs are very high. If a highly complementary target mRNA also induces the release of the miRNA from AGO1 in *Arabidopsis*, the released miRNA may be subsequently modified at the 3' end by SDNs and HESO1 in a non-processive manner, so that some modified miRNA species could be rebound by AGO1 and act on their targets again.

The lack of an effect on miRNA accumulation by *heso1-1* in WT suggests that *HESO1* may not be part of the normal small RNA degradation machinery. However, given the presence of other nucleotidyl transferase genes with potentially overlapping functions, a role of *HESO1* in miRNA degradation in WT plants cannot be excluded. HESO1 activity is completely inhibited by 2′-*O*-methylation, it may only act on unmethylated miRNAs as a surveillance mechanism in WT plants. Alternatively, it is possible that HESO1 acts cooperatively with SDN1, which truncates small RNAs [44], or another nucleotidyl transferase that is not inhibited by 2′-*O*-methylation, to degrade miRNAs. The activities of these enzymes would generate 3' end unmethylated miRNAs

on which HESO1 can uridylate. In addition, in wild type plans, the *HESO1* may function in degrading some small RNAs that are unmethylated. For example, a virus silencing suppressor protein P1/helper component-proteinase (P1/Hc-Pro) could inhibit the activity of HEN1 during viral infection [59, 60]. The expression Hc-Pro in tobacco impairs the methylation of viral small RNAs without significantly affecting endogenous plant small RNAs, however, the Hc-Pro protein also enhances the accumulation of viral small RNAs [59]. It has been hypothesized that unmethylated viral small RNAs may induce the production of secondary small RNAs, which may allow the infected cell to be immune from repeated infections [59]. This may decrease the damage to the host plant and increase the viability of the virus [59]. If the unmethylated primary small RNAs from virus could be targeted by HESO1 and further degraded by host nucleases, this could prevent the production of secondary siRNAs from the virus. Further studies on the *heso1* mutant during viral infection are needed to determine whether *HESO1* could have a function in antagonizing the function of certain viral silencing suppressors like Hc-Pro during viral infection.

In addition to small RNA degradation in *hen1* mutants, the HESO1 protein may also function in tailing and degradation of other RNAs. In fact, the *C. reinhardtii* TRF4-like nucleotidyl transferase MUT68 not only uridylates siRNAs, it also adenylates the 5' cleavage products of siRNA-targeted RNAs [61], suggesting that this type of nucleotidyl transferase could be multifunctional. Since most miRNAs are methylated and cannot be directly tailed by HESO1 in wild type *Arabidopsis* plants, the HESO1 protein may mainly act on other types of RNAs rather than small RNAs. Since miRNAs function in target mRNA cleavage in plants, the function of HESO1 in *hen1* mutants and the involvement of AGO1 in miRNA 3' modifications [57] indicate that HESO1 may also have access to other RNAs at the AGO1 catalytic site. Thus, it is possible that HESO1 functions in degrading miRNA target RNAs in wild type plants and because of the presence of HESO1 near the AGO1 protein, miRNAs need to be methylated in order to

prevent them from being tailed by HESO1. In addition, it remains to be determined whether HESO1 also mediates the degradation of RNAs unrelated to post-transcriptional gene silencing. Determination of HESO1's subcellular localization and identification of its interacting proteins would provide more insights into the mechanisms of HESO1 mediated RNA degradation. Furthermore, a systematic analysis of RNA tailing in wild type and *heso1* would help to identify the targets of HESO1 protein under normal conditions.

Materials and Methods

Phylogenetic analysis

For ten putative *Arabidopsis* nucleotidyl transferases and selected homologs from other eukaryotes, sequences corresponding to the NT_PAP_TUTase (cd05402) domain44 were aligned using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) with default parameters [62]. The phylogenetic tree in Figure 1.1 was generated using TreeView [63].

Arabidopsis strains and growth conditions

Plants were grown under long day (16 h light/ 8 h darkness) conditions at 22°C. All *Arabidopsis* strains are in the Columbia background. Seeds of *heso1-1* were obtained from the Gabi-Kat (GK-367H02-017041) collection. To obtain the *hen1-8 heso1-1* double mutant, *hen1-8* was crossed with *heso1-1* and the F2 population was genotyped for *hen1-8* and *heso1-1* mutations. The *hen1-8* mutation was genotyped by PCR using primers T13K14-P19 and T13K14-P20 (Table 1.2) followed by restriction digestion with HpaI (only the mutant sequence can be digested). *heso1-1* was genotyped by PCR with the primer pair GK367_GT-F/GK367_GT-R (Table 1.2), which amplifies the wild type genomic fragment (absent in *heso1-1* homozygous plants), and PCR amplification with the primer pair GK367_GT-R/Gabi-kat_T-DNA (Table 1.2), which

amplifies part of the T-DNA insertion (present in the *heso1-1* homozygous and heterozygous plants).

Plasmid construction

To construct the *pHESO1:HESO1-GFP* plasmid, the *HESO1* genomic region from a 2kb promoter to the end of the coding region (without the stop codon) was amplified by PCR with primers at2g39740_SacII and at2g39740_EcoRI_R_Nstop (Table 1.2), and cloned into pJL-blue, a pBluescript-based Gateway entry vector[64], at *Sac*II and *Eco*RI sites. Then the DNA fragment was moved into pMDC107[65] by the LR Gateway reaction. This plasmid was used to transform *hen1-8 heso1-1* plants by *Agrobacterium* (GV3101) mediated floral dip transformation. To construct the *6xHis-HESO1_WT* plasmid, the *HESO1* coding sequence was amplified using the cDNA clone S67294 (obtained from ABRC) as a template with primers HESO1-EcoRI_F-28b and HESO1-EcoRI_R-28b (Table 1.2). The amplified fragment was cloned into *pET-28b* at the *Eco*RI site. The orientation of the insertion was confirmed by sequencing.

To construct the *6xHis-HESO1_M* plasmid, the *HESO1* coding sequence with the D66A and D68A mutations was first generated by PCR using primers that incorporated the mutations and clone S67294 as a template. The left fragment (primers: HESO1-EcoRI_F-28b and HESO1_DADA_R) and right fragment (primers: HESO1_DADA_F and HESO1-EcoRI_R-28b) (Table 1.2) were annealed and used as the template to amplify *HESO1 CDS_M* using HESO1-EcoRI_F-28b and HESO1-EcoRI-28b. This fragment was cloned into *pET28b* at the *Eco*RI site. The orientation of insertion was confirmed by sequencing.

Reverse transcription and PCR

Reverse transcription and real-time PCR were performed as previously described [66]. *UBIQUITIN5* was used as an internal control for the real-time RT-PCR to examine the expression

of miRNA target genes. *ACTIN8* was used as an internal control for the *HESO1* RT-PCR. Primers used are listed in Table 1.2.

Northern blotting

RNA isolation and northern blotting to detect small RNAs were performed as described [19]. Total RNAs were extracted from inflorescences using TRIzol reagent (Invitrogen). 5′-end-labeled (32P) antisense DNA oligonucleotides were used to detect miRNAs from total RNAs. Sodium periodate treatment and β-elimination were done as described [20].

Small RNA library construction and sequencing

Small RNAs of a desired size range (15–40 nt) were fractionated from total RNAs by 15% polyacrylamide gel electrophoresis [67]. These small RNAs were sequentially ligated with the 3' and 5' adapters using the Small RNA Sample Preparation Kit (Illumina) according to the manufacturer's instructions. A reverse transcription reaction followed by a low cycle PCR was performed to obtain sufficient amounts of products for deep sequencing. The wild type, *heso1-1*, *hen1-8*, and *hen1-8 heso1-1* libraries were barcoded and sequenced in one channel on an Illumina Hiseq2000. Two biological replicates were performed. The small RNA sequence data are available at GEO under the accession number GSE35479.

Analysis of miRNA and siRNA tailing and truncation

Small RNA reads were filtered for quality and separated into different genotypes according to the barcodes using Illumina's built-in pipeline. Reads that matched to tRNAs and rRNAs were removed. In biological replicate one, the total reads that passed the quality and tRNA/rRNA filters for wild type, *heso1-1*, *hen1-8*, and *hen1-8 heso1-1* were 2525504, 1695482, 2518194, and 2840703, respectively. In biological replicate two, the total reads that passed the filters for the same four genotypes were 7638946, 13470428, 23249325, and 15341388. The

small RNA reads were then mapped to the *Arabidopsis* genome using Bowtie [68]. To analyze miRNA 3' derivatives, any small RNA read that could not be perfectly mapped back to the genome was chopped one nucleotide at a time from the 3' end until the remaining sequence was perfectly mapped to the genome. Thus, for any non-genome-matched small RNA read, the sequence was split into two parts: the longest 5' genome-matched component (5GMC) part, and a 3' "tail" part. For any read that could be mapped to the genome without chopping, the 5GMC would be the same as the read and there would be no tail. With all reads processed into the format of 5GMC plus tail (the tail could be null), the 5GMC of each read was compared to all annotated miRNAs in miRBase to ascertain those originating from known miRNAs. miRNAs with a 5GMC frequency of 50 TPM (transcript per million) and above were included in the tailing analyses.

To determine the extent of "tailing" (addition of non-templated 3' nucleotides) and "truncation" (shortening from the 3' end), *Arabidopsis* miRNAs annotated in miRBase v17 were examined. Small RNA reads with the 5GMC portion perfectly aligning to each one of the annotated miRNAs were identified, and the proportion of tailing and truncation was calculated based on read frequencies for the different categories of forms (full-length, tailed only, truncated only, and tailed and truncated) for each miRNA. For siRNAs, we chose a small set of 24 nt siRNAs (see Table 1.2 for their sequences) that are abundantly represented for the analysis. Small RNA reads matching to these siRNA sequences were separated into 5GMCs and 3' tails as for miRNAs.

AGO1 immunoprecipitation and RNA extraction

AGO1 was immunoprecipitated as previously described [69] from *hen1* and *hen1 heso1* inflorescences and wild type seedlings with mouse anti-AGO1 antibodies generated by Dr. Xiaofeng Cao's group at Institute of Genetics and Developmental Biology, Chinese Academy of

Sciences. RNAs were isolated from AGO1 immunoprecipitates as well as input samples (total RNA) using Tri-reagent (Molecular Research Center, Inc.).

Protein purification and enzymatic assays

The *6xHis-HESO1_WT* and *6xHis-HESO1_M* plasmids were transformed into the *E. coli* strain BL21 Star™(DE3) (Invitrogen) for protein expression. The transgenic *E. coli* strains were cultured at 37°C until the OD reached 0.5. IPTG was added to a final concentration of 0.2 mM and the culture was incubated at 16°C overnight. The recombinant proteins were purified using Ni-NTA agarose (Qiagen) under native conditions following the manufacturer's instructions with some modifications. During the wash step, the column was washed in succession with the following buffers: 10 volume of wash buffer, 10 volume of wash buffer with 50% glycerol and 1% NP-40, 5 volume of wash buffer, 10 volume of wash buffer with 3M Urea and 1% NP-40, 5 volume of wash buffer, 8 volume wash buffer with 100 mM imidazole, and 2 volume of wash buffer with 120 mM imidazole. Then the protein was eluted with elution buffer.

miR173 RNA oligonucleotides (either with a 2'-OH or with a 2'-*O*-methyl group at the 3' terminal nucleotide) were labeled in a 50 µl reaction mixture containing 100 µM miR173, 50 U T4 Polynucleotide Kinase (New England Biolabs), 1X T4 PNK buffer and 5 µl ATP [γ-32P] (3000Ci/mmol 10mCi/ml from PerkinElmer). After incubating at 37°C for 1 hour, the mixture was purified with Illustra MicroSpin™ G-25 Columns (GE Healthcare) according to the manufacturer's instructions.

The HESO1 enzymatic activity assays were conducted in 20 µl reaction mixtures containing 1 µg recombinant wild type or mutated 6XHis-HESO1, 1 µl labeled miR173, 20 mM Tris-HCl (pH 8.0), 50 mM KCl, 0.7 mM MnCl2, 2.5 mM MgCl2, 0.5 mM DTT, 200 µg/mL BSA, and 0.5 mM of different nucleotide triphosphates. Reactions were incubated at room temperature

for 0-40 min, and stopped by adding the formamide small RNA loading dye to the reaction mix. Then, the mix was denatured at 75°C, incubated on ice for 5 min, then the RNAs were separated on a 15% acrylamide urea gel and visualized by autoradiography.

References

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

2. Mallory, A.C., Reinhart, B.J., Jones-Rhoades, M.W., Tang, G., Zamore, P.D., Barton, M.K., and Bartel, D.P. (2004). MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. The EMBO journal 23, 3356-3364.

3. Laufs, P., Peaucelle, A., Morin, H., and Traas, J. (2004). MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems. Development (Cambridge, England) 131, 4311-4322.

4. Guo, H.S., Xie, Q., Fei, J.F., and Chua, N.H. (2005). MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. The Plant cell 17, 1376-1386.

5. Millar, A.A., and Gubler, F. (2005). The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. The Plant cell 17, 705-721.

6. Wang, J.W., Wang, L.J., Mao, Y.B., Cai, W.J., Xue, H.W., and Chen, X.Y. (2005). Control of root cap formation by MicroRNA-targeted auxin response factors in Arabidopsis. The Plant cell 17, 2204-2216.

7. Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science (New York, N.Y.) 303, 2022-2025.

8. Lu, S., Sun, Y.H., Shi, R., Clark, C., Li, L., and Chiang, V.L. (2005). Novel and mechanical stress-responsive MicroRNAs in Populus trichocarpa that are absent from Arabidopsis. The Plant cell 17, 2186-2203.

9. Sunkar, R., and Zhu, J.K. (2004). Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. The Plant cell 16, 2001-2019.

10. Yamasaki, H., Abdel-Ghany, S.E., Cohu, C.M., Kobayashi, Y., Shikanai, T., and Pilon, M. (2007). Regulation of copper homeostasis by micro-RNA in Arabidopsis. The Journal of biological chemistry 282, 16369-16378.

11. Li, Y., Zhang, Q., Zhang, J., Wu, L., Qi, Y., and Zhou, J.M. (2010). Identification of microRNAs involved in pathogen-associated molecular pattern-triggered plant innate immunity. Plant physiology 152, 2222-2231.

12. Chen, X. (2010). Small RNAs - secrets and surprises of the genome. The Plant journal : for cell and molecular biology 61, 941-958.

13. Kim, V.N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. Nature reviews. Molecular cell biology 6, 376-385.

14. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415-419.

15. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science (New York, N.Y.) 293, 834-838.

16. Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes & development 15, 2654-2659.

17. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes & development 16, 1616-1626.

18. Kurihara, Y., and 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.

19. Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Current biology : CB 12, 1484-1495.

20. Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science (New York, N.Y.) 307, 932-935.

21. Huang, Y., Ji, L., Huang, Q., Vassylyev, D.G., Chen, X., and Ma, J.B. (2009). Structural insights into mechanisms of the small RNA methyltransferase HEN1. Nature 461, 823-827.

22. Baumberger, N., and Baulcombe, D.C. (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.

23. Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L., and Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. Science (New York, N.Y.) 320, 1185-1190.

24. Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. The Plant cell 14, 1605-1619.

25. Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science (New York, N.Y.) 297, 2053-2056.

26. Abe, M., Yoshikawa, T., Nosaka, M., Sakakibara, H., Sato, Y., Nagato, Y., and Itoh, J. (2010). WAVY LEAF1, an ortholog of Arabidopsis HEN1, regulates shoot development

by maintaining MicroRNA and trans-acting small interfering RNA accumulation in rice. Plant physiology 154, 1335-1346.

27. Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Current biology : CB 15, 1501-1507.

28. Kamminga, L.M., Luteijn, M.J., den Broeder, M.J., Redl, S., Kaaij, L.J., Roovers, E.F., Ladurner, P., Berezikov, E., and Ketting, R.F. (2010). Hen1 is required for oocyte development and piRNA stability in zebrafish. The EMBO journal 29, 3688-3700.

29. Billi, A.C., Alessi, A.F., Khivansara, V., Han, T., Freeberg, M., Mitani, S., and Kim, J.K. (2012). The Caenorhabditis elegans HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. PLoS genetics 8, e1002617.

30. Montgomery, T.A., Rim, Y.S., Zhang, C., Dowen, R.H., Phillips, C.M., Fischer, S.E., and Ruvkun, G. (2012). PIWI associated siRNAs and piRNAs specifically require the Caenorhabditis elegans HEN1 ortholog henn-1. PLoS genetics 8, e1002616.

31. Kurth, H.M., and Mochizuki, K. (2009). 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. RNA (New York, N.Y.) 15, 675-685.

32. Kirino, Y., and Mourelatos, Z. (2007). The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. RNA (New York, N.Y.) 13, 1397-1401.

33. Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes & development 21, 1603-1608.

34. Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Current biology : CB 17, 1265-1272.

35. Kamminga, L.M., van Wolfswinkel, J.C., Luteijn, M.J., Kaaij, L.J., Bagijn, M.P., Sapetschnig, A., Miska, E.A., Berezikov, E., and Ketting, R.F. (2012). Differential impact of the HEN1 homolog HENN-1 on 21U and 26G RNAs in the germline of Caenorhabditis elegans. PLoS genetics 8, e1002702.

36. Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science (New York, N.Y.) 328, 1534-1539.

37. Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nature reviews. Molecular cell biology 10, 126-139.

38. Chen, X. (2005). MicroRNA biogenesis and function in plants. FEBS letters 579, 5923-5931.

39. Yu, B., Bi, L., Zhai, J., Agarwal, M., Li, S., Wu, Q., Ding, S.W., Meyers, B.C., Vaucheret, H., and Chen, X. (2010). siRNAs compete with miRNAs for methylation by HEN1 in Arabidopsis. Nucleic acids research 38, 5844-5850.

40. Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS biology 2, E104.

41. Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell 120, 613-622.

42. Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science (New York, N.Y.) 308, 118-120.

43. Chatterjee, S., and Grosshans, H. (2009). Active turnover modulates mature microRNA activity in Caenorhabditis elegans. Nature 461, 546-549.

44. Ramachandran, V., and Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. Science (New York, N.Y.) 321, 1490-1492.

45. Lehrbach, N.J., Armisen, J., Lightfoot, H.L., Murfitt, K.J., Bugaut, A., Balasubramanian, S., and Miska, E.A. (2009). LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans. Nature structural & molecular biology 16, 1016-1020.

46. van Wolfswinkel, J.C., Claycomb, J.M., Batista, P.J., Mello, C.C., Berezikov, E., and Ketting, R.F. (2009). CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. Cell 139, 135-148.

47. Ibrahim, F., Rymarquis, L.A., Kim, E.J., Becker, J., Balassa, E., Green, P.J., and Cerutti, H. (2010). Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. Proceedings of the National Academy of Sciences of the United States of America 107, 3906-3911.

48. Belostotsky, D. (2009). Exosome complex and pervasive transcription in eukaryotic genomes. Current opinion in cell biology 21, 352-358.

49. Heo, I., Joo, C., Kim, Y.K., Ha, M., Yoon, M.J., Cho, J., Yeom, K.H., Han, J., and Kim, V.N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. Cell 138, 696-708.

50. Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes & development 19, 2030-2040.

51. Chekanova, J.A., Gregory, B.D., Reverdatto, S.V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., et al. (2007). Genome-wide high-resolution

mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. Cell 131, 1340-1353.

52. Malecki, M., Viegas, S.C., Carneiro, T., Golik, P., Dressaire, C., Ferreira, M.G., and Arraiano, C.M. (2013). The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. The EMBO journal 32, 1842-1854.

53. Chang, H.M., Triboulet, R., Thornton, J.E., and Gregory, R.I. (2013). A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature 497, 244-248.

54. Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nature cell biology 7, 719-723.

55. Zhang, W., Murphy, C., and Sieburth, L.E. (2010). Conserved RNaseII domain protein functions in cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant phenotypes. Proceedings of the National Academy of Sciences of the United States of America 107, 15981-15985.

56. Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature 461, 754-761.

57. Zhai, J., Zhao, Y., Simon, S.A., Huang, S., Petsch, K., Arikit, S., Pillay, M., Ji, L., Xie, M., Cao, X., et al. (2013). Plant MicroRNAs Display Differential 3' Truncation and Tailing Modifications That Are ARGONAUTE1 Dependent and Conserved Across Species. The Plant cell.

58. De, N., Young, L., Lau, P.W., Meisner, N.C., Morrissey, D.V., and MacRae, I.J. (2013). Highly complementary target RNAs promote release of guide RNAs from human Argonaute2. Molecular cell 50, 344-355.

59. Ebhardt, H.A., Thi, E.P., Wang, M.B., and Unrau, P.J. (2005). Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression. Proceedings of the National Academy of Sciences of the United States of America 102, 13398-13403.

60. Jamous, R.M., Boonrod, K., Fuellgrabe, M.W., Ali-Shtayeh, M.S., Krczal, G., and Wassenegger, M. (2011). The helper component-proteinase of the Zucchini yellow mosaic virus inhibits the Hua Enhancer 1 methyltransferase activity in vitro. The Journal of general virology 92, 2222-2226.

61. Ibrahim, F., Rohr, J., Jeong, W.J., Hesson, J., and Cerutti, H. (2006). Untemplated oligoadenylation promotes degradation of RISC-cleaved transcripts. Science (New York, N.Y.) 314, 1893.

62. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England) 23, 2947-2948.

63. Page, R.D. (1996). TreeView: an application to display phylogenetic trees on personal computers. Computer applications in the biosciences : CABIOS 12, 357-358.

64. Yant, L., Mathieu, J., Dinh, T.T., Ott, F., Lanz, C., Wollmann, H., Chen, X., and Schmid, M. (2010). Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. The Plant cell 22, 2156-2170.

65. Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant physiology 133, 462-469.

66. Kim, Y.J., Zheng, B., Yu, Y., Won, S.Y., Mo, B., and Chen, X. (2011). The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. The EMBO journal 30, 814-822.

67. Lu, C., Meyers, B.C., and Green, P.J. (2007). Construction of small RNA cDNA libraries for deep sequencing. Methods (San Diego, Calif.) 43, 110-117.

68. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10, R25.

69. Ji, L., Liu, X., Yan, J., Wang, W., Yumul, R.E., Kim, Y.J., Dinh, T.T., Liu, J., Cui, X., Zheng, B., et al. (2011). ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in Arabidopsis. PLoS genetics 7, e1001358.

Figures and Tables

Figure 1.1 Phylogenetic analysis of the nucleotidyl transferase (NT) domain of poly(A) polymerases and terminal uridylyl transferases

Ten *Arabidopsis* proteins and homologous proteins from other eukaryotes were used in the analysis. The *Arabidopsis* proteins all have the "AT" prefix, and *HESO1* is *At2g39740*. Evolutionary distance is indicated by the scale bar.

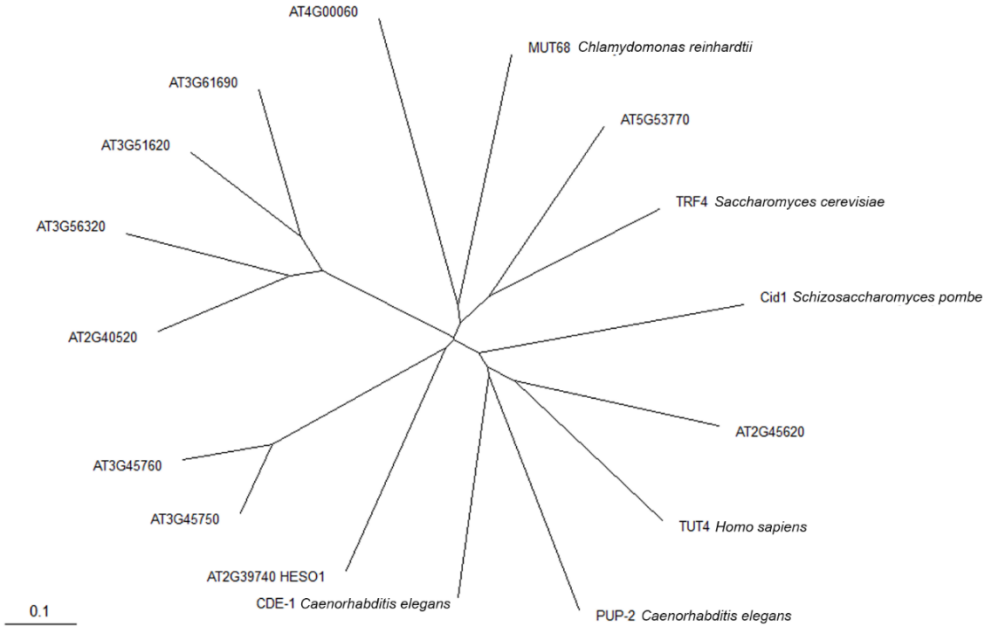
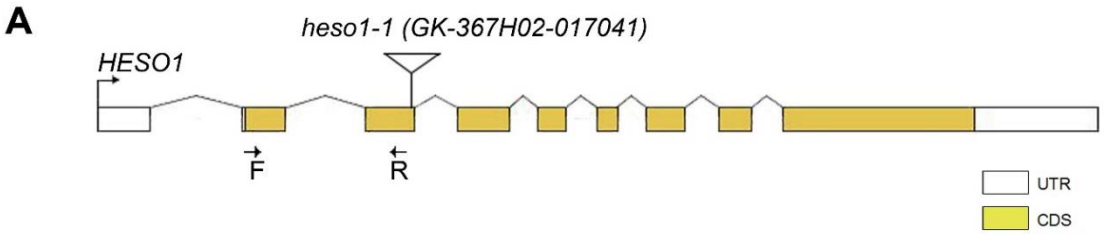

AT4G00060
MUT68 Chlamydomonas reinhardtii
AT3G61690
AT5G53770
AT3G51620
TRF4 Saccharomyces cerevisiae
AT3G56320
Cid1 Schizosaccharomyces pombe
AT2G40520
AT2G45620
AT3G45760
AT3G45750
TUT4 Homo sapiens
AT2G39740 HESO1
CDE-1 Caenorhabditis elegans
PUP-2 Caenorhabditis elegans
0.1

Figure 1.2 *HESO1* gene structure and mutation of *heso1-1* allele

(A) The *HESO1* gene structure and the T-DNA insertion site in the *heso1-1* mutant. Rectangles represent exons and lines represent introns. The T-DNA insertion is indicated by a triangle. The catalog number for the T-DNA mutant at the Nottingham *Arabidopsis* Stock Center is in the parentheses. The arrow on the far left indicates the transcription start site. The arrows in the coding region mark the positions of primers used for RT-PCR in (B).

(B) RT-PCR to examine the *HESO1* transcript in wild type (Col) and *heso1-1*. *ACTIN8* was used as a loading control. Although the primers in *HESO1* were 5' to the T-DNA insertion site, no transcripts were detected in the *heso1-1* mutant, suggesting that *heso1-1* is likely a null allele.


A
heso1-1 (GK-367H02-017041)
HESO1
F
R
UTR
CDS

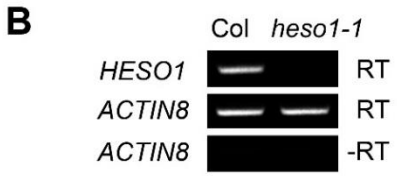

B
Col
heso1-1
HESO1
RT
ACTIN8
RT
ACTIN8
-RT

Figure 1.3 The *heso1-1* mutation partially rescues the developmental defects of *hen1-8* plants

(A) Twenty-day-old plants of WT (Col), *hen1-8*, and *hen1-8 heso1-1* genotypes, as indicated. The *heso1-1* mutation partially rescues the morphological phenotypes of *hen1* at vegetative stage.

(B) Siliques from wild type (Col), *hen1-8*, and *hen1-8 heso1-1* plants as well as three independent *hen1-8 heso1-1* T1 transgenic lines harboring the *HESO1* genomic DNA. The *heso1-1* mutation largely restored the fertility of *hen1-8* plants, and the enhanced fertility in the *hen1-8 heso1-1* double mutant was abrogated by the *HESO1* transgene.

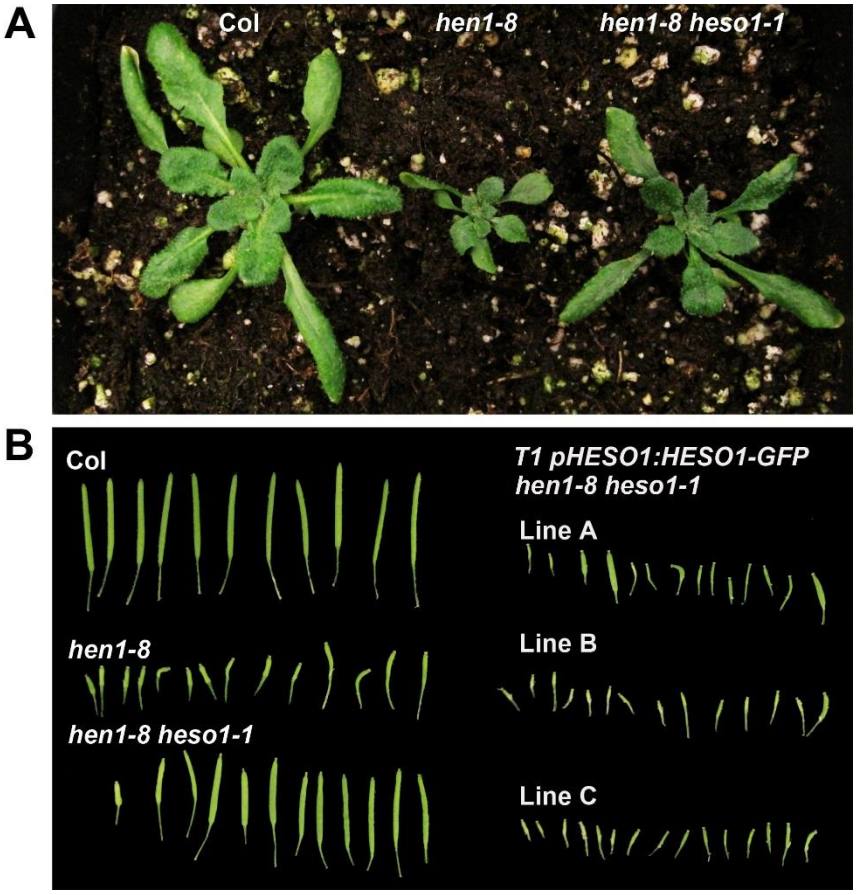

A
Col
hen1-8
hen1-8 heso1-1
B
Col
hen1-8
hen1-8 heso1-1
T1 pHESO1:HESO1-GFP
hen1-8 heso1-1
Line A
Line B
Line C

Figure 1.4 The *heso1-1* mutation partially rescues the molecular defects of *hen1-8* plants

(A) Relative transcript levels from four miRNA target genes in the three genotypes as determined by real-time RT-PCR. The derepression of the genes in *hen1-8*, due to reduced miRNA accumulation, was fully or partially rescued by *heso1-1*. *PHB, CUC2, SCL6*, and *SPL10* are targets of miR165/166, miR164, miR171, and miR156/157, respectively. The error bars represent SD calculated from three technical replicates.

(B) The accumulation of miR165/166, miR167 and miR173 in the four genotypes as determined by northern blotting. Note that 50 μg of total RNA was used for *hen1-8* and *hen1-8 heso1-1*, whereas 5 μg of total RNA was used for Col and *heso1-1*. The abundance of three miRNAs was increased in *hen1-8 heso1-1* compared with *hen1-8*.

(C) The accumulation of miR165/166 and miR173 in wild type (Col), *hen1-8*, *hen1-8 heso1-1* and two independent T1 lines (T1 line A and line B) of *hen1-8 heso1-1* harboring the *pHESO1:HESO1-GFP* transgene. The abundance of the two miRNAs was decreased in the two T1 lines compared to *hen1-8 heso1-1*, indicating that *pHESO1:HESO1-GFP* rescued the molecular phenotypes of *heso1-1*.

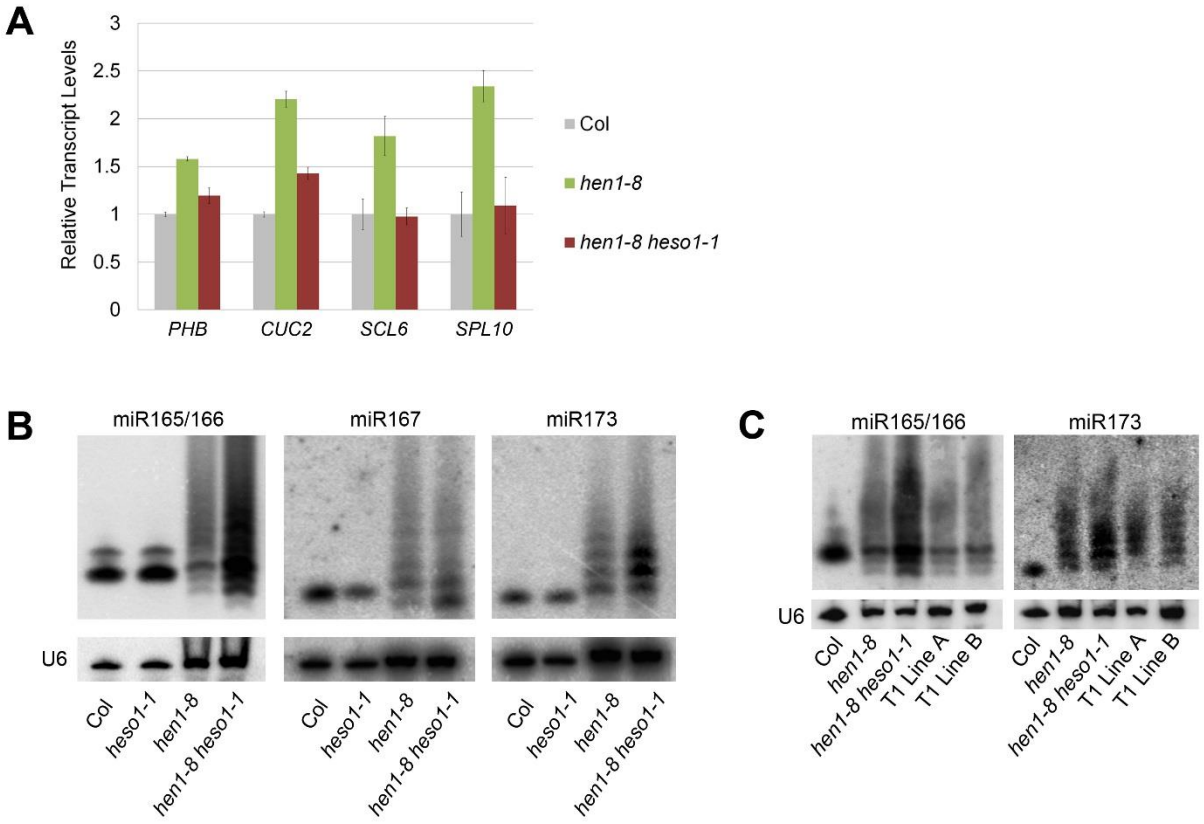

A
Relative Transcript Levels
3
2.5
2
1.5
1
0.5
0
PHB
CUC2
SCL6
SPL10
Col
hen1-8
hen1-8 heso1-1
B
miR165/166
miR167
miR173
U6
Col
heso1-1
hen1-8
hen1-8 heso1-1
C
miR165/166
miR173
U6
Col
hen1-8
hen1-8 heso1-1
T1 Line A
T1 Line B

Figure 1.5 The *heso1-1* mutation does not rescue the miRNA methylation defects of *hen1-8*

Total RNAs were subjected (+) or not (-) to the β-elimination reactions followed by northern blotting. In *hen1-8*, the lack of 2'-*O*-methylation of the miRNAs rendered the miRNAs susceptible to the reactions such that a mobility shift was observed between the treated and untreated samples. In wild type (Col) and *heso1-1*, 2'-*O*-methylation protected the miRNAs from the chemical reactions and no mobility shifts were observed. In *hen1-8 heso1-1*, mobility shifts were observed, indicating that the miRNAs lacked 2'-*O*-methylation. Note that 50 μg of total RNA was used for *hen1-8* and *hen1-8 heso1-1* whereas 5 μg was used for wild type (Col) and *heso1-1*. The general reduction in miRNA levels after β-elimination was due to loss incurred during the procedure.

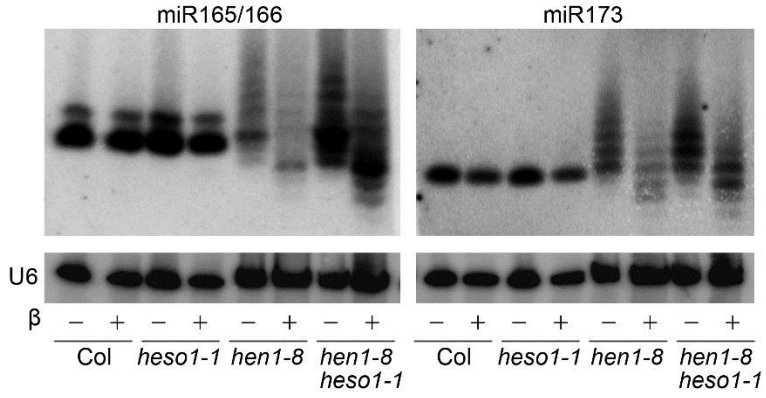

miR165/166
miR173
U6
β
– + – + – + – +
– + – + – + – +
Col
heso1-1
hen1-8
hen1-8 heso1-1
Col
heso1-1
hen1-8
hen1-8 heso1-1

Figure 1.6 The *heso1-1* mutation reduces 3' uridylation of miRNAs in *hen1-8* as revealed by small RNA high-throughput sequencing

Small RNA reads corresponding to known miRNAs were categorized into four classes: full-length (class 1), tailed only (full-length reads plus tails) (class 2), truncated only (class 3), and truncated and tailed (class 4) The proportion of tailed and untailed reads for ten miRNAs in *hen1-8* and *hen1-8 heso1-1*. The proportions of tailed and untailed species were calculated as %(sum of classes 2 and 4 read numbers divided by total read number) and %(sum of classes 1 and 3 read numbers divided by total read number), respectively. Data from the biological one and biological replicate two are shown in figure (A) and (B) respectively.

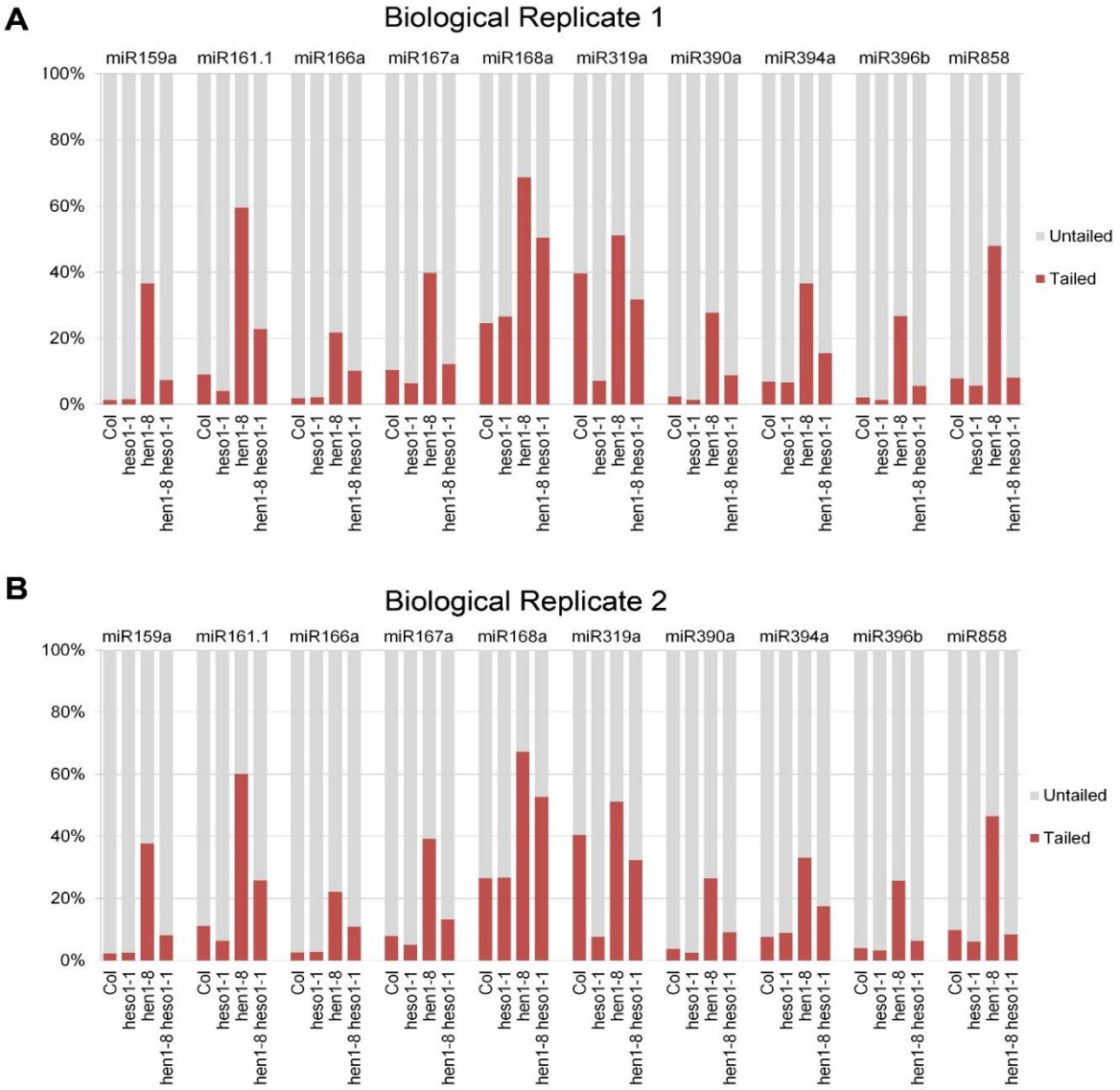

A
Biological Replicate 1
miR159a
miR161.1
miR166a
miR167a
miR168a
miR319a
miR390a
miR394a
miR396b
miR858
100%
80%
60%
40%
20%
0%
Col
heso1-1
hen1-8
hen1-8 heso1-1
Untailed
Tailed
B
Biological Replicate 2
miR159a
miR161.1
miR166a
miR167a
miR168a
miR319a
miR390a
miR394a
miR396b
miR858
100%
80%
60%
40%
20%
0%
Col
heso1-1
hen1-8
hen1-8 heso1-1
Untailed
Tailed

Figure 1.7 Tail length distribution and nucleotide frequencies in the tails of four miRNAs

For each miRNA species, all tailed miRNA reads (classes 2 and 4) were analyzed for frequencies of tail length. The proportions of microRNA species with tails of 1–7 nt are shown. Nucleotide frequencies (represented by the color codes) in the tails were calculated as % (number of a nucleotide in the tail divided by tail length).

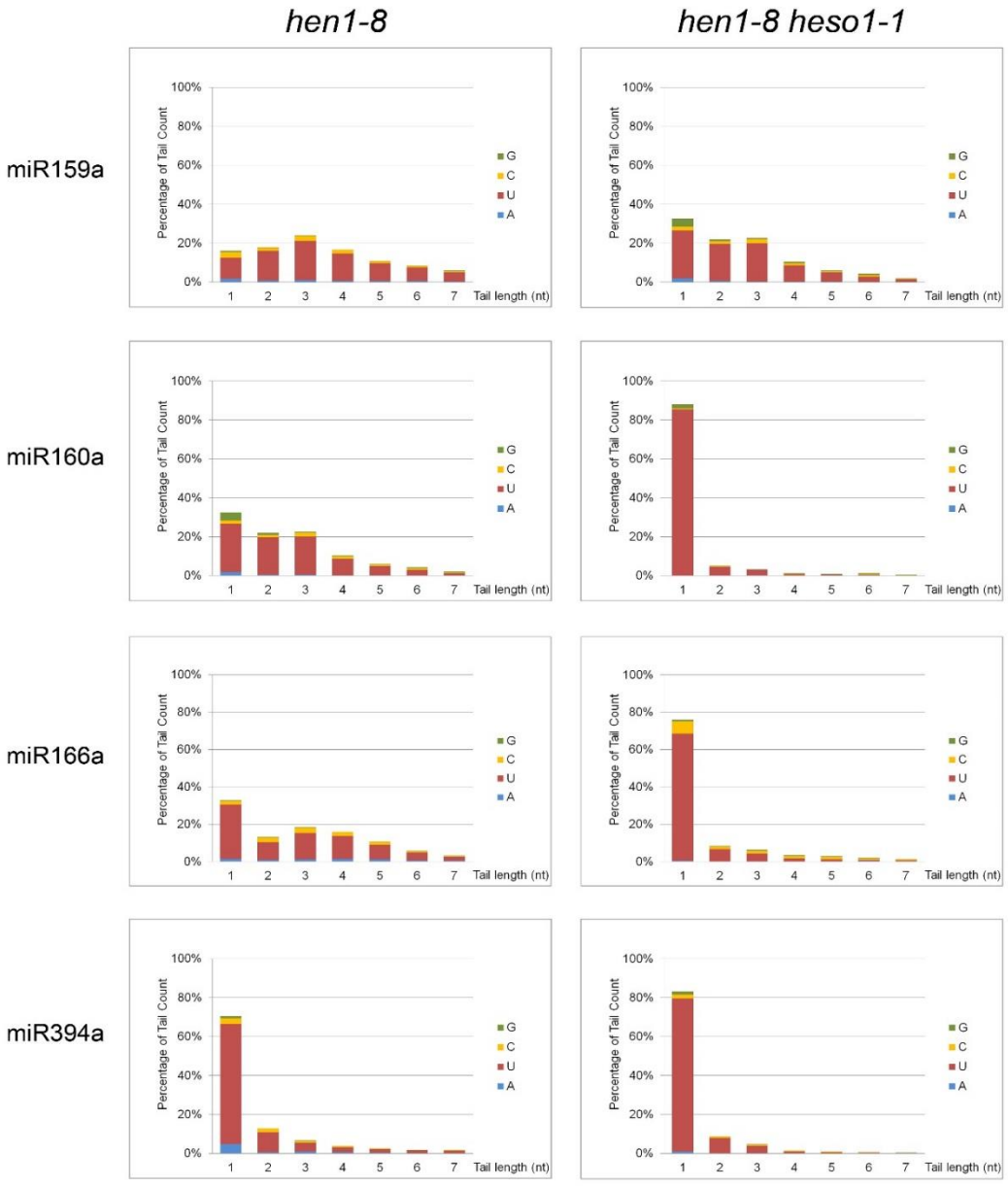

hen1-8
hen1-8 heso1-1
miR159a
miR160a
miR166a
miR394a
Percentage of Tail Count
100%
80%
60%
40%
20%
0%
1 2 3 4 5 6 7
Tail length (nt)
G
C
U
A

Figure 1.8 The *heso1-1* mutation reduces 3' tailing of three miRNAs and an siRNA without much effects on 3' truncation in *hen1-8*

The status of 3' truncation and/or 3' tailing for each small RNA is represented by a two-dimensional matrix in which the x axis represents the extents of 3' truncation and the y axis represents the extents of 3' tailing. The sizes of the circles indicate the relative abundance of the small RNA species. In *hen1-8*, both 3' truncation and 3' tailing occurred at a much higher frequency than in WT. In *hen1-8 heso1-1*, 3' tailing was drastically reduced, but 3' truncation was largely unaffected. Data from both biological replicates were used for the analysis here.

Figure 1.8-1

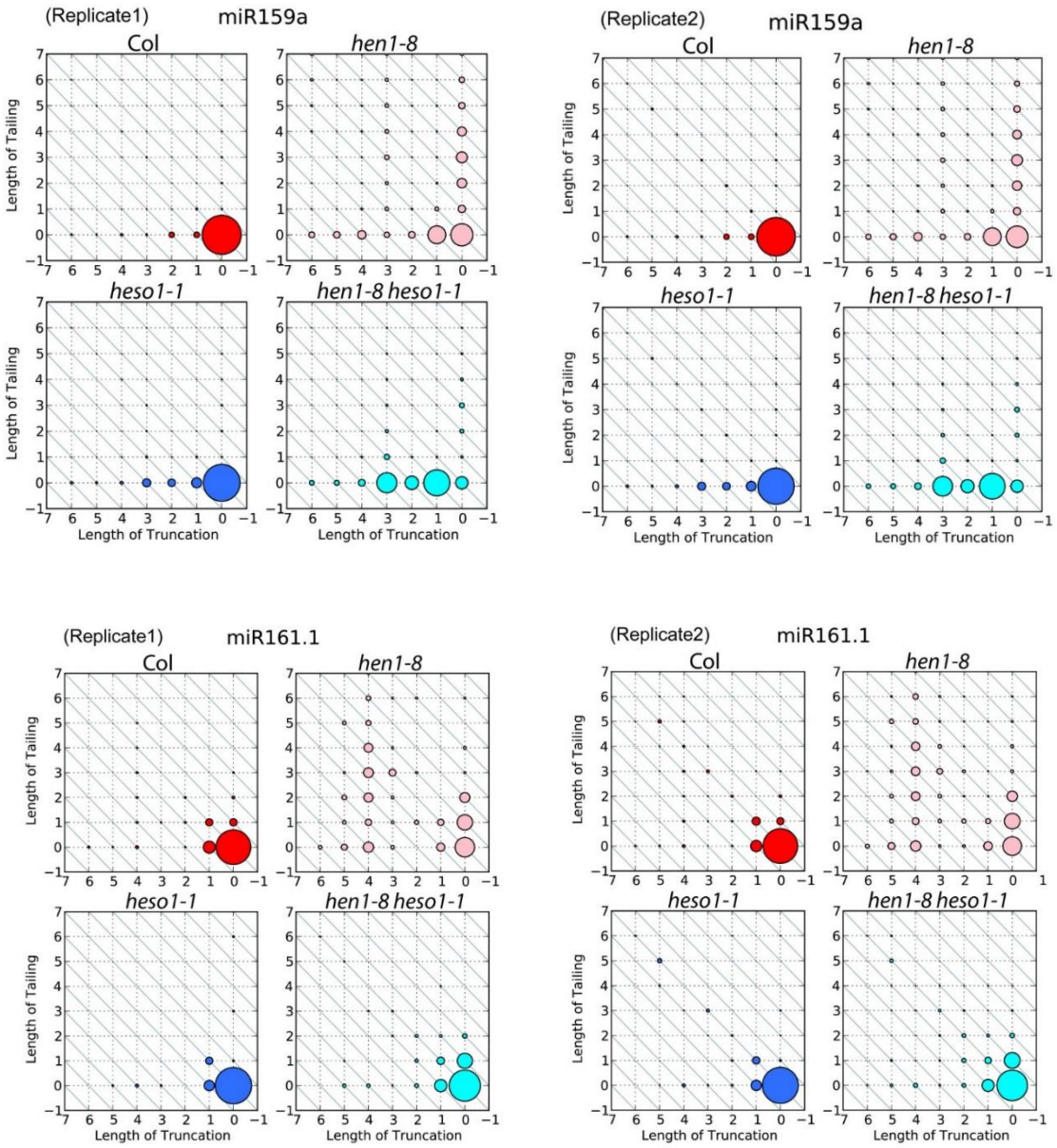

(Replicate1)
miR159a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR159a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate1)
miR161.1
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR161.1
Col
hen1-8
heso1-1
hen1-8 heso1-1
Length of Tailing
Length of Truncation

Figure 1.8-2

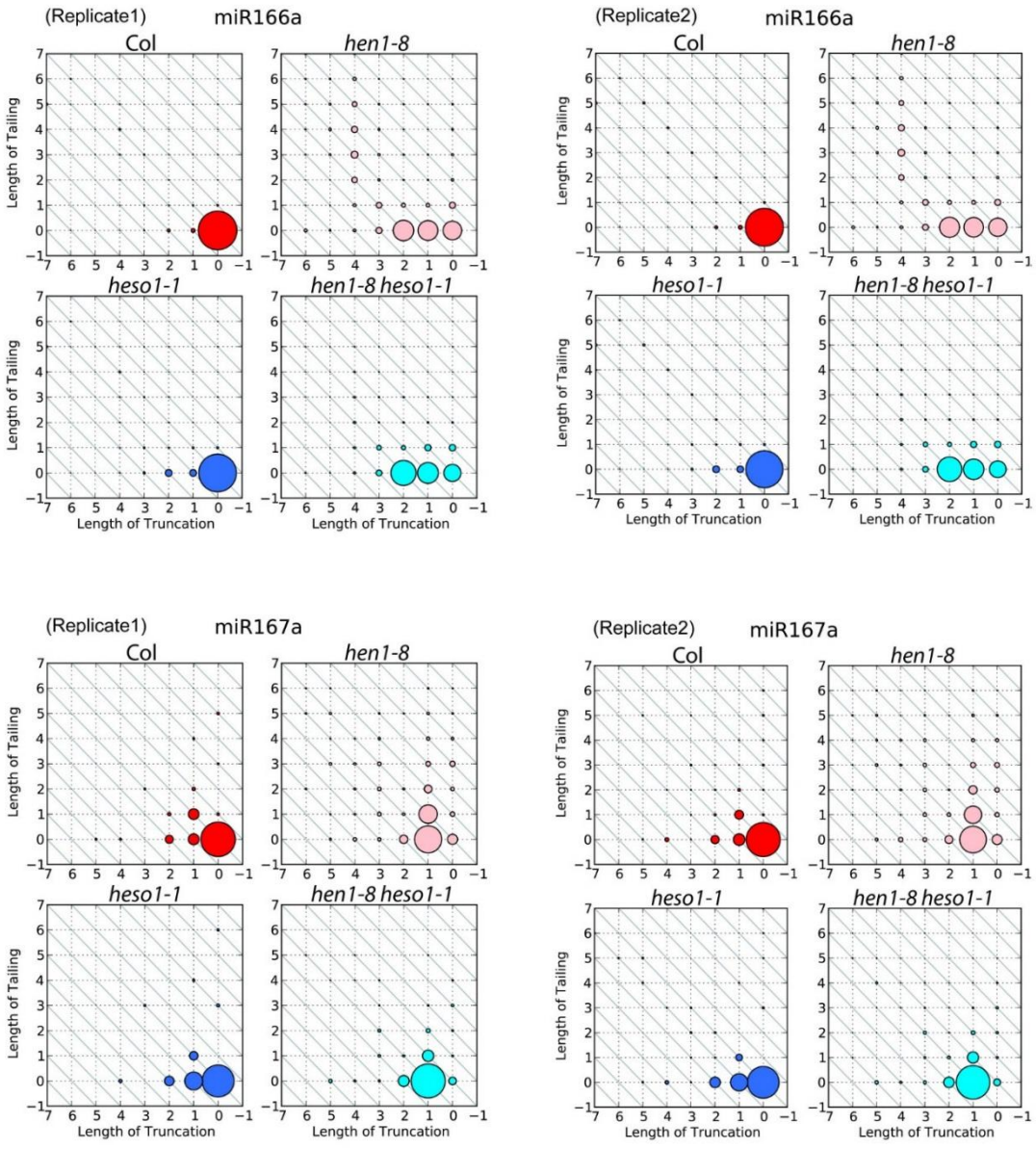

(Replicate1)
miR166a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR166a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate1)
miR167a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR167a
Col
hen1-8
heso1-1
hen1-8 heso1-1
Length of Tailing
Length of Truncation

Figure 1.8-3

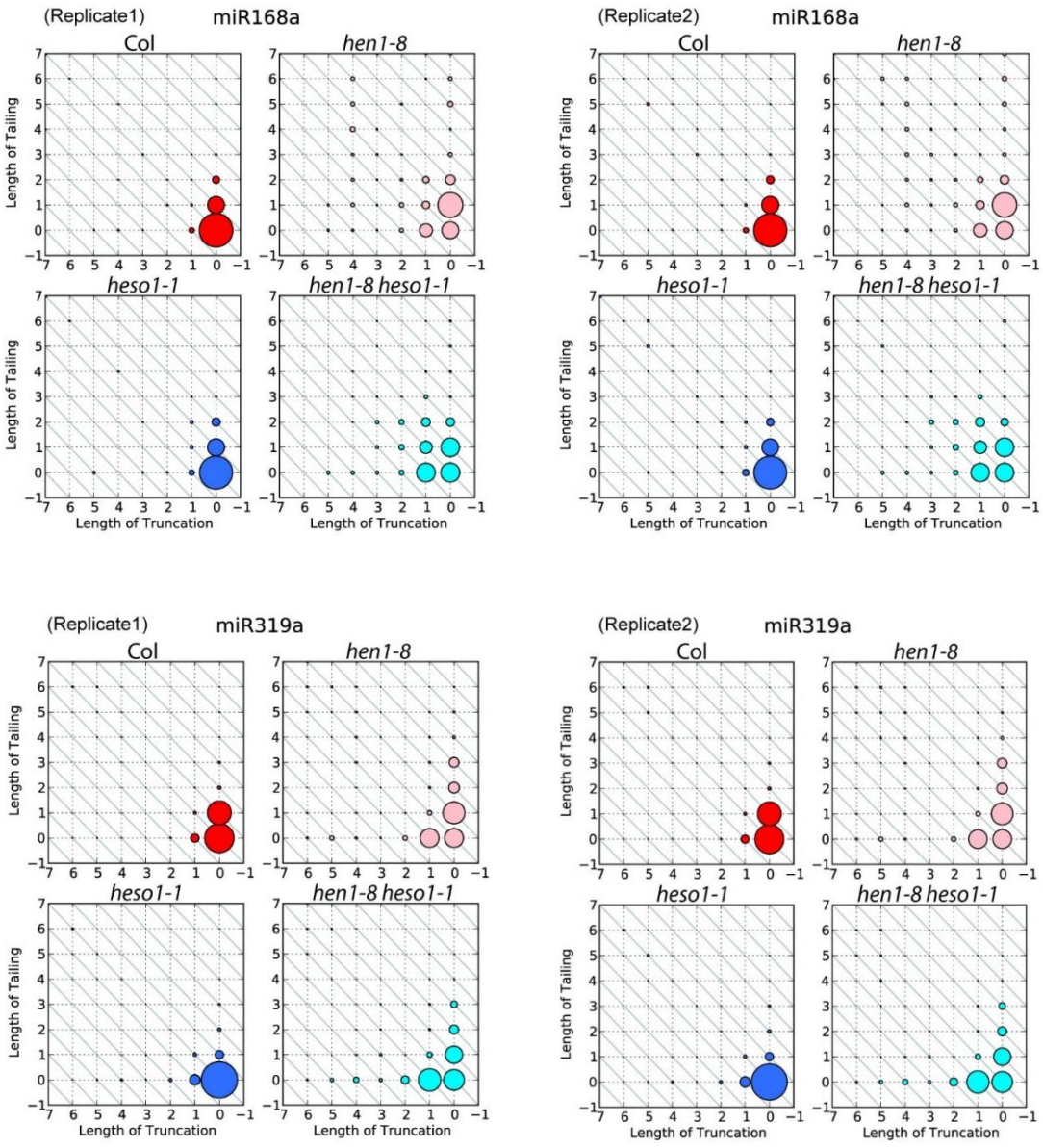

(Replicate1)
miR168a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR168a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate1)
miR319a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR319a
Col
hen1-8
heso1-1
hen1-8 heso1-1
Length of Tailing
Length of Truncation

Figure 1.8-4

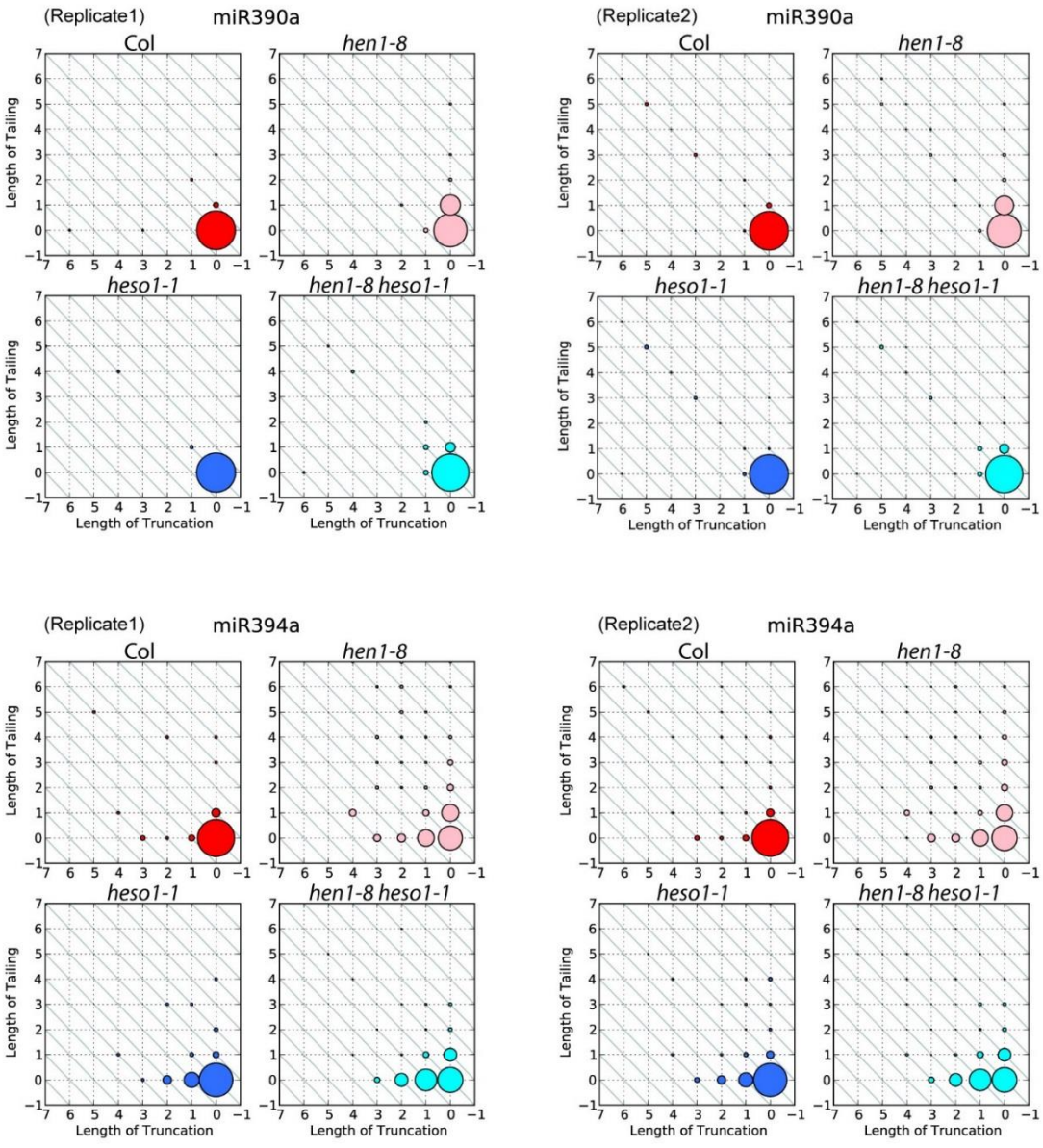

(Replicate1)
miR390a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR390a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate1)
miR394a
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR394a
Col
hen1-8
heso1-1
hen1-8 heso1-1
Length of Tailing
Length of Truncation

Figure 1.8-5

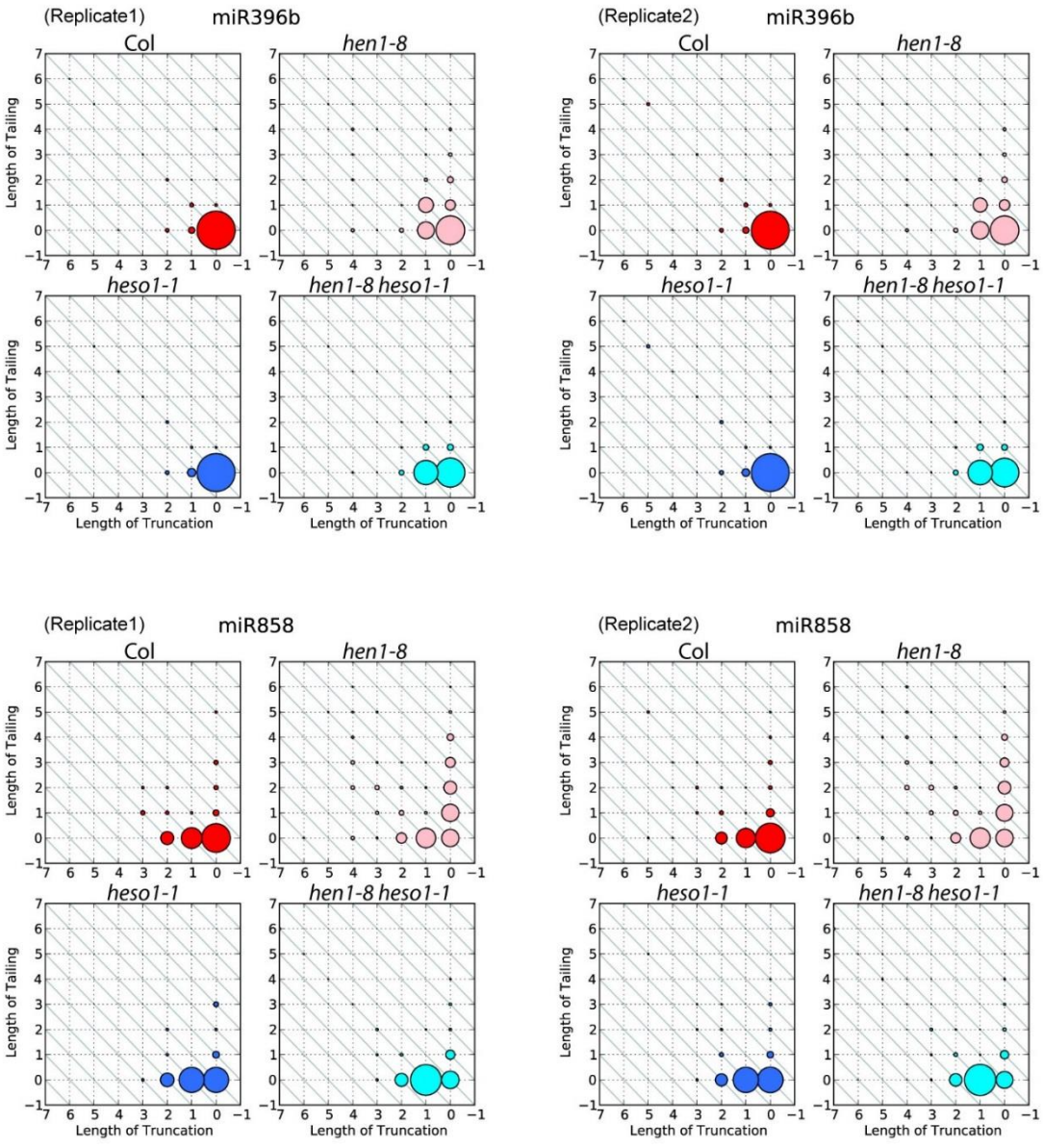

(Replicate1)
miR396b
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR396b
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate1)
miR858
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
miR858
Col
hen1-8
heso1-1
hen1-8 heso1-1
Length of Tailing
Length of Truncation

Figure 1.8-6

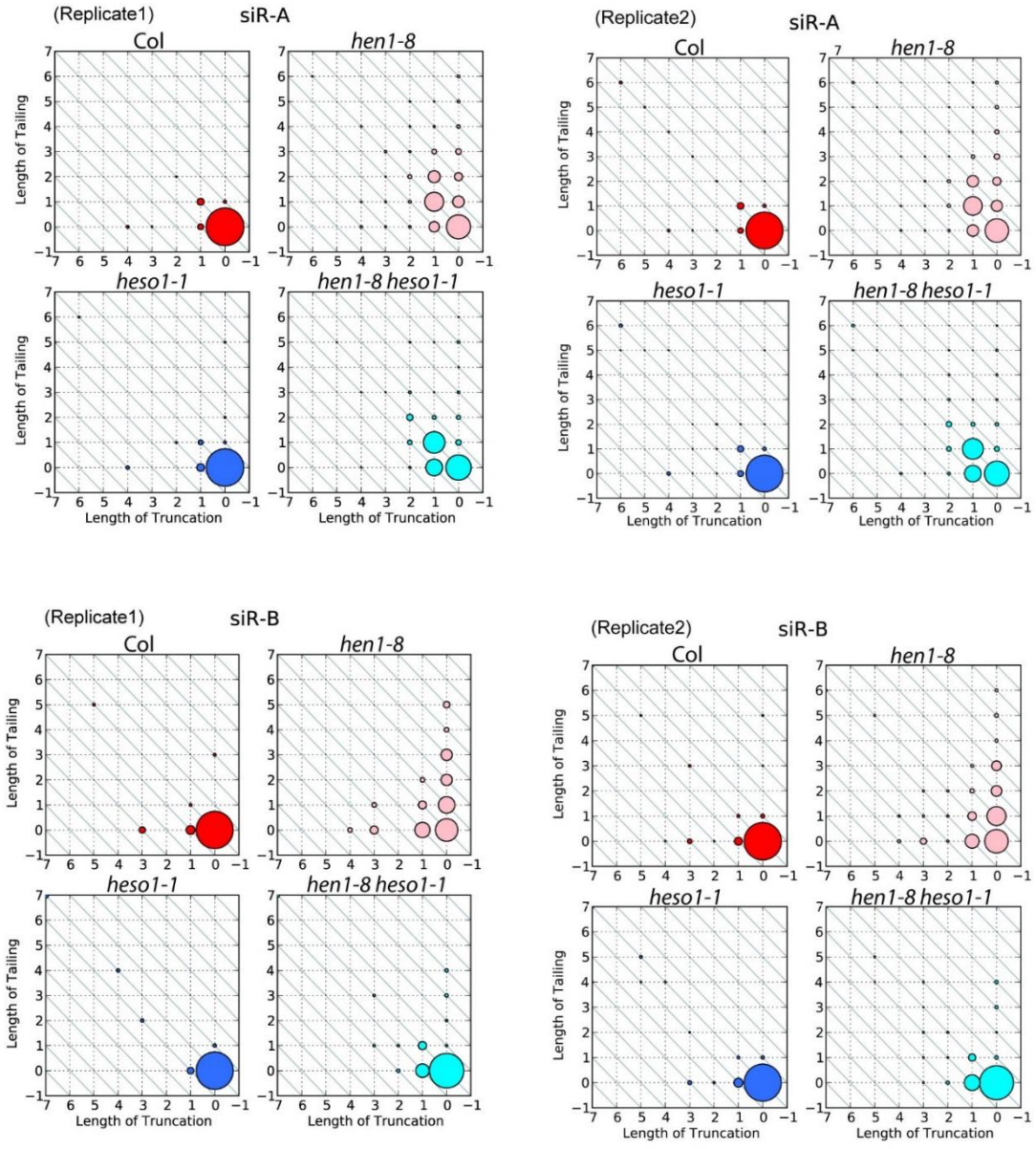

(Replicate1)
siR-A
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
siR-A
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate1)
siR-B
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
siR-B
Col
hen1-8
heso1-1
hen1-8 heso1-1
Length of Tailing
Length of Truncation

Figure 1.8-7

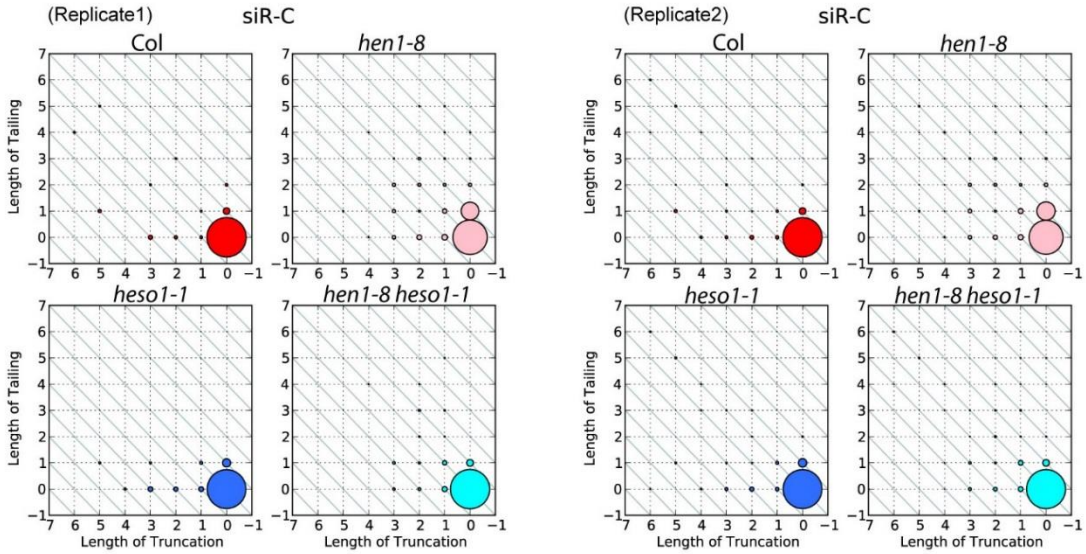

(Replicate1)
siR-C
Col
hen1-8
heso1-1
hen1-8 heso1-1
(Replicate2)
siR-C
Col
hen1-8
heso1-1
hen1-8 heso1-1
Length of Tailing
Length of Truncation

Figure 1.9 Northern blots of miR167 and miR165/166 from total RNA or AGO1 immunoprecipitates from *hen1–8* and *hen1–8 heso1–1* inflorescences

Note that the Col (wild type) samples only serve as a size reference; RNA from seedlings was used and was not in the equivalent quantity as the *hen1–8* or *hen1–8 heso1–1* samples. The bands representing miRNAs of the wild type size are indicated by asterisks. The arrows above and below the asterisks mark bands that represent 3' uridylated and 3' truncated species, respectively.

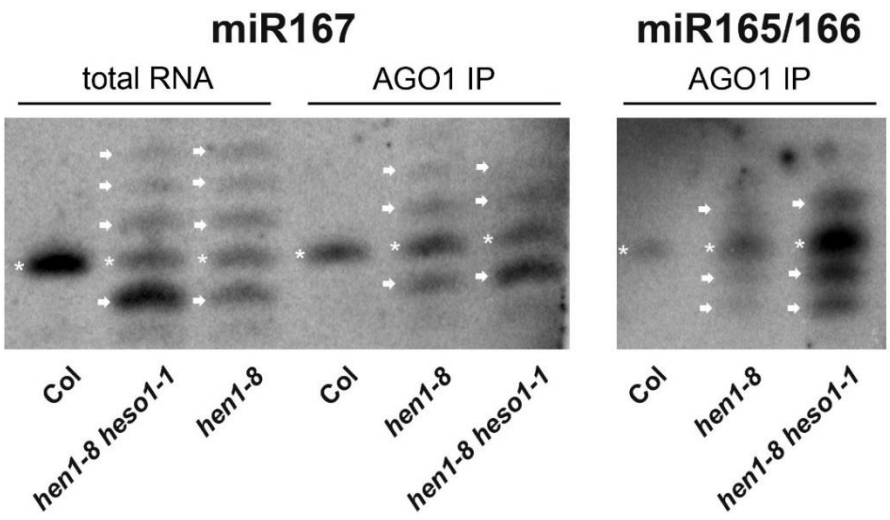
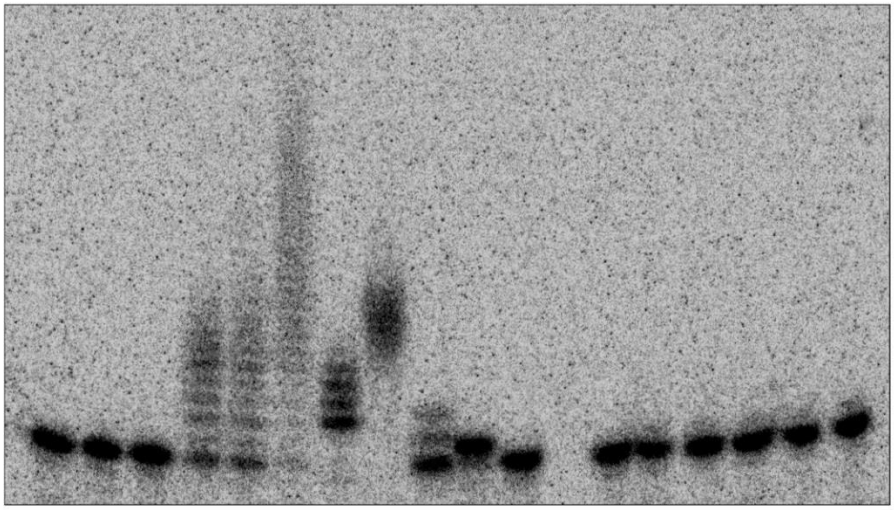

miR167
miR165/166
total RNA
AGO1 IP
AGO1 IP
Col
hen1-8 heso1-1
hen1-8
Col
hen1-8
hen1-8 heso1-1
Col
hen1-8
hen1-8 heso1-1

Figure 1.10 HESO1 exhibited terminal nucleotidyl transferase activity

(A) The nucleotidyl transferase assay was conducted with 5′ radiolabeled miR173 without or with 2′-*O*-methylation. Recombinant WT (W) or mutant His-HESO1, in which two catalytic residues were mutated (M; Figure 1.10B), was incubated with the miRNA in the presence of various nucleotide triphosphates. Nucleotidyl transferase activity is represented by the presence of higher molecular weight bands relative to the input miRNA. The "-" signs indicate the absence of the nucleotide triphosphate, protein, or RNA oligonucleotide.

(B)The nucleotidyl transferase domain of HESO1 is shown as the green box. Part of the NT domain sequence is shown below. Three conserved aspartic acids (D66, D68 and D130 in red) form a metal binding triad in the wild type (W) protein. In the mutated version (M) used for the *in vitro* enzymatic assay, the first two aspartic acid residues were mutated to alanine (marked by *).

A



Lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Time (min)	0	40	0	5	20	40	40	40	40	40	40	40	0	40	0	40	0	40
Nucleotide	U	U	U	U	U	U	A	C	G	dA	-	U	U	U	U	U	U	U
Protein	-	-	W	W	W	W	W	W	W	W	W	W	M	M	-	-	W	W
Oligo	miR173											-	miR173		miR173-2'-*O*-Me			

B


```
W   61  FTRWGDLDISVDLFSGSSILFTGKKQKQTLLGHLLRALRASGLWYKLQFVIHARVPILKVVSGHQRISCDISIDNLDGLL 140
M   61  FTRWGALAISVDLFSGSSILFTGKKQKQTLLGHLLRALRASGLWYKLQFVIHARVPILKVVSGHQRISCDISIDNLDGLL 140
            * *
```

Table 1.1 T-DNA lines used for each nucleodidyl transferase genes in *Arabidopsis*

Gene ID	Mutant line	Source
At2g40520	CS810224	ABRC
At2g39740	GK-367H02	NASC
At2g45620	SALK_119177, SALK_087647	ABRC
At3g45750	SALK_057397, SALK_105213	ABRC
At3g45760	SALK_018808	ABRC
At3g56320	SALK_111188, SALK_122314	ABRC
At3g61690	SALK_074440	ABRC
At4g00060	Sup4	Unpulished T-DNA line from from Dr. Jiankang Zhu's Lab
At5g53770	GK-297E11	NASC
At3g51620	SALK_085337	ABRC

Table 1.2 Oligo sequences in this study

Experiment	Name	Sequence (5'-3')
Genotyping	GK367_GT-F	CTGGTTCTGTGATTGTTAGGTG
Genotyping	GK367_GT-R	GAGACCAACAGCTCCGAGA
Genotyping	Gabi-kat_T-DNA	ATATTGACCATCATACTCATTGC
Genotyping	T13K14-P19	TAGTCTCATCAAGTTATGTCT
Genotyping	T13K14-P20	GCAGAGAAGCGTGTTCAATC
Plasmid construction	at2g39740_SacII	tccccgcggTGAGGATTGAGGTAAGCCA
Plasmid construction	at2g39740_EcoRI_R_Nstop	cggaattcCTGCTCATGTCTCGGTCTC
Plasmid construction	HESO1-EcoRI_F-28b	cggaattcgATGAGTAGAAACCCTTTCC
Plasmid construction	HESO1-EcoRI_R-28b	cggaattcCTACTGCTCATGTCTCGGT
Plasmid construction	HESO1_DADA_F	TCACACGATGGGGAgccTTAgctATCTCTGTTGACTT
Plasmid construction	HESO1_DADA_R	AAGTCAACAGAGATagcTAAggcTCCCCATCGTGTGA
Northern blot	miRNA166_AS	GGGGAATGAAGCCTGGTCCGA
Northern blot	miRNA173_AS	GTGATTTCTCTCTGTAAGCGA
Northern blot	miRNA167_AS	TAGATCATGTTGGCAGTTTCA
Truncation and tailing analysis	siR-A	AACTTGGATACTGTGAATGATGCA
Truncation and tailing analysis	siR-B	ATTGAGGCTGAGTCGAGTAGAGCT
Truncation and tailing analysis	siR-C	AAGAATCAACGAACTGCAGACACC
Enzymatic activity assay	miR173 RNA oligo	UUCGCUUGCAGAGAGAAAUCAC
Real-time PCR Ubiquitin 5	UBIQUITIN5-F	GGTGCTAAGAAGAGGAAGAAT
	UBIQUITIN5-R	CTCCTTCTTTCTGGTAAACGT
Real-time PCR CUC2	CUC2F2	CTCACTCCCACCTCTCCT
	CUC2R2	GAAAAGGGTCAAAGTCAAAC
Real-time PCR SCL6	SCL6_F	GAGGTCATAGAGAGCGACAC
	SCL6_R	GGAATAAGGGTTTAGGGTTT
Real-time PCR SPL10	SPL10_F	TGAGACAAAGCCTACACAGATGGA
	SPL10_R	GATGATGCAACCCGACTTTTTTATG
Real-time PCR PHB	PHB_realtimeF1	ATGAAACATCAACTTCACACTG
	PHB_realtimeR2	CGTTTCCAGCAGGTATCACA
RT-PCR	NTP4-5'RT_F	GGAGCAAGACGATGAGTAGA
RT-PCR	NTP4-5'RT_R	GTAGTAAATGTCCGAGTAAAGT
RT-PCR	ACTIN 8 F	CACATGCTATCCTCCGTCTC
RT-PCR	ACTIN 8 R	CAATGCCTGGACCTGCTT

Chapter 2
Characterization of THO complex mutants in *Arabidopsis*

Abstract

DNA methylation and histone modifications are important epigenetic marks that regulate gene expression. To identify new players involved in the regulation of gene expression through DNA methylation in *Arabidopsis*, we performed genetic screens using two *LUCIFERASE* (*LUC*) reporter lines that are regulated by DNA methylation and isolated mutants with decreased luminescence. We cloned three genes, *TEX1*, *HPR1* and *THO5A*, which encode three different subunits of the conserved eukaryotic nuclear protein complex THO. Genetic analysis and whole genome DNA methylation profiling indicated that the THO complex does not have a significant function in regulating DNA methylation at the *LUC* locus or endogenous loci. Small RNA sequencing of the THO complex mutants confirmed that the mutations in these genes do not affect the overall accumulation of Pol IV-dependent 24-nt heterochromatic siRNAs, which are key players in DNA methylation. From mRNA sequencing, we identified genes that were differentially expressed in *tex1* and *hpr1* and found that repeat-associated genes were enriched among the down-regulated genes in two mutants of the THO complex. Additionally, genes marked by H3K27 trimethylation were enriched among both up-regulated and down-regulated genes in *tex1* and *hpr1*.

Introduction

In eukaryotes, the transcription of a DNA sequence is determined not only by the genetic information stored in the sequence but also by epigenetic modifications of the DNA and associated histone proteins. In plants, DNA cytosine methylation is recognized as the most important type of epigenetic DNA modification. Nevertheless, various types of histone modifications in plants and animals are also known, including methylation, phosphorylation, acetylation, ubiquitination and SUMOylation, which occur at specific amino acids of histone subunits in the nucleosome. The epigenetic status of a DNA sequence is dynamic. Some epigenetic modifications may be maintained during mitotic cell division or change in response to different developmental stages [1, 2] or external stimuli [3]. In reproductive cells, genome-wide reprogramming occurs at certain epigenetic marks to ensure that the correct epigenetic information is transmitted to the next generation [4]. Different epigenetic modifications may have different effects on the transcription of the underlying DNA sequence. For example, DNA methylation at gene promoters may inhibit the binding of transcription factors to the sequence [5]. In contrast, histone acetylation and phosphorylation may reduce the positive charge of a histone, thereby weakening the interaction between the histone and DNA and providing a favorable environment for transcription (reviewed in [6]).

The repression of DNA transcription by epigenetic modifications such as DNA methylation and histone modifications is referred to as transcriptional gene silencing (TGS). In plants, the TGS mechanisms of DNA methylation, histone H3K9 dimethylation (H3K9me2) and histone H3K27 trimethylation (H3K9me3) are well studied. These modifications silence transposons and some protein-coding genes to regulate developmental processes and responses to environmental changes and to help maintain genome integrity. This introduction provides a summary of the

mechanisms of DNA methylation and repressive histone modifications in *Arabidopsis* along with the biological functions of these modifications. Prior studies on THO complex genes from other eukaryotes will also be introduced to provide a framework for the cloning and characterization of orthologous genes in *Arabidopsis* performed in the present study.

DNA methylation

1. Maintenance of DNA methylation

In plants, DNA methylation is important for the transcriptional regulation of a small number of genes and is largely associated with the maintenance of genomic integrity through transposon silencing. Plant DNA cytosine methylation occurs at three sequence contexts—CG, CHG and CHH (where H indicates A, T or C) [7]. CG methylation is commonly found in gene bodies, transposable elements (TEs) and repeats, while CHG and CHH methylation are generally found in transposons and repeats, which are silenced under normal conditions [7, 8]. Immediately following DNA replication, newly synthesized DNA strands lack methyl-cytosine and must be methylated by DNA methyltransferases to ensure the maintenance of this epigenetic mark in both daughter cells. In *Arabidopsis*, the maintenance of DNA methylation at the three sequence contexts requires different mechanisms.

CG methylation is primarily maintained by *DNA METHYLTRANSFERASE 1* (*MET1*), which is a homolog of the methyltransferase gene *DNMT1* in mammals. After DNA replication, MET1 is recruited to hemi-methylated loci to methylate the cytosine on the newly synthesized strand (reviewed in [9]). In *met1* mutants, DNA methylation at the CG context is lost throughout the genome [10]. If a wild-type *MET1* gene is introduced back into the mutant, CG methylation can only be restored at transposons and repeats where RNA-directed DNA methylation (RdDM) takes place; in contrast, CG methylation at transcribed gene bodies cannot be restored [10]. In addition

to MET1, three VARIATION IN METHYLATION family proteins, VIM1, VIM2 and VIM3, are also required for the maintenance of CG methylation. These VIM proteins contain a PHD domain, a SET and RING finger-associated (SRA) domain and two RING domains [11, 12]. *In vitro* assays have shown that VIM1 preferentially binds hemi-methylated CG sites [13]. The methylation profile of the *vim1 vim2 vim3* triple mutant resembles that of the *met1* mutant, in which a global loss of CG methylation is observed [10].

The maintenance of cytosine methylation at the CHG context is mediated mainly by CHROMOMETHYLASE 3 (CMT3), which is guided by H3K9me2. H3K9me2 occurs at TEs and repeats in both the pericentromeric region and euchromatic chromosomal arms [14], and the profile of CMT3 localization in the genome strongly resembles the distribution pattern of H3K9me2 [15]. In *Arabidopsis*, the histone methyltransferase KRYPTONITE (KYP)/SUVH4, together with its homologs SU(VAR)3-9 HOMOLOG 5 (SUVH5) and SUVH6, mediates the methylation of H3K9. In the *kyp* single mutant and in the *kyp suvh5 suvh6* triple mutant, H3K9me2 levels are reduced [16], and a dramatic loss of CHG methylation can be observed [10]. A recent study showed that the chromo domain and BAH domain of CMT3 bind to the methylated tails of histone H3 and that the DNMT domain of CMT3 methylates DNA [15]. Furthermore, the SRA domain of KYP preferentially binds to methylated cytosines [17]. In DNA methyltransferase mutants, decreased H3K9 methylation is observed at some loci [17]. These multiple lines of evidence suggest that the DNA methylation and histone H3K9me2 pathways form a self-reinforcing loop at TEs and transposons. In addition to CMT3, DOMAINS REARRANGED METHYLASE 1 (DRM1) and DRM2 also contribute to the maintenance of CHG methylation at some loci through the RdDM pathway [10, 18].

Unlike CG and CHG methylation, CHH methylation is asymmetric. After DNA replication, methylation is only retained in one of the new DNA molecules; in the other DNA molecule, the

original template lacks a cytosine at the site, and methylation needs to be re-established to ensure that the same methylation pattern is maintained in both daughter cells. In *Arabidopsis*, CHH methylation often occurs alongside CHG methylation at TEs and repeats [7, 8]. The RdDM pathway plays an important role in maintaining CHH methylation in transposons and repeats that reside in the euchromatic arms and at the edges of long heterochromatic blocks [9, 19]. The RdDM pathway involves non-coding RNA molecules that guide DNA methyltransferases to the target loci, including 24-nt heterochromatic siRNAs generated from TEs and repeats. Thus far, several proteins are known to be required for the accumulation of 24-nt siRNAs at all RdDM target loci. One current model of the underlying mechanism proposes that a long RNA is transcribed from the silenced region by Pol IV, a specialized RNA polymerase that is only found in plants [20]. The SNF2 domain-containing protein CLASSY1 (CLSY1), which is a putative chromatin remodeling protein [21], and the IWR (Interacts with Pol II)-type transcription factor RNA-DIRECTED DNA METHYLATION 4 (RDM4) may be required for Pol IV transcription [22]. SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1) may function in Pol IV recruitment to a large number of RdDM loci. More specifically, SHH1 may bridge histone modification and Pol IV recruitment through the recognition of histone H3 tails with unmethylated K4 and methylated K9 by the SAWADEE domain [22, 23]. Following Pol IV transcription, the long RNA is made double -stranded by RNA DEPENDENT RNA POLYMERASE 2 (RDR2) [20, 24]. In an affinity purification assay of NRPD1, the largest subunit of Pol IV, SHH1, CLSY1, RDM4 and RDR2 were co-purified, evidencing the involvement of these proteins in the production of small RNA precursors [22]. The long double-stranded RNA generated by RDR2 is subsequently diced by DICER-LIKE 3 (DCL3) into 24-nt siRNA duplexes [25]. After HEN1 methylation at the 3’ ends [26], the siRNA duplex is exported into the cytoplasm, where it is loaded into AGO4 with the help of HSP90. Next, AGO4 cleaves the siRNA passenger strand and is subsequently

translocated into the nucleus and recruited to the target site of the siRNA [27]. The recruitment process involves a second non-coding RNA molecule known as the scaffold RNA. While the scaffold RNA is transcribed by the plant-specific RNA polymerase Pol V at most RdDM target loci, Pol II also contributes to the production of scaffold RNA at some loci [25] [28]. AGO4 may recognize the scaffold RNA through sequence complementarity between the AGO4-bound siRNA and the scaffold RNA. Through an unknown mechanism, this interaction recruits the DNA methyltransferase DRM2 or its homolog DRM1 to the locus, resulting in the deposition of methyl groups at unmethylated cytosines [9, 25]. The production of Pol V-dependent scaffold RNA requires the putative chromatin remodeling complex known as the DDR complex, which is made up of the following components: DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), a SWI2/SNF2-like chromatin remodeling protein, DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), a structural maintenance of chromosomes (SMC) hinge domain-containing protein; and RNA-DIRECTED DNA METHYLATION 1 (RDM1) [29, 30]. KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1) is recruited to the RdDM target loci in an NRPE1-dependent manner and may promote the recruitment of chromatin-modifying enzymes [31, 32]. INVOLVED IN DE NOVO 2 (IDN2) is a double-stranded RNA-binding protein. Through its interaction with the scaffold RNA and the SWI/SNF complex, it may regulate transcriptional gene silencing by altering nucleosome positioning at target loci [33].

In the body of long TEs, which mainly occur in pericentromeric regions, the SWI2/SNF2-like chromatin-remodeling protein DECREASED DNA METHYLATION 1 (DDM1) allows methyltransferases to access the DNA [19]. At these loci as well as TEs in euchromatic regions, MET1 and CMT3 are responsible for maintaining DNA methylation at the CG and CHG sequence contexts, respectively [10, 19], as they are at TEs residing at euchromatic regions.

However, because the chromatin remodeling protein DRD1, which is required for RdDM, is not able to function effectively in the body of long TEs, cytosine methylation in the CHH context at these loci is mainly maintained by chromomethylase 2 (CMT2), a DNA methyltransferase homologous to CMT3 [19].

Although plants have evolved specialized pathways for the maintenance of DNA methylation in different sequence contexts, these different pathways cannot be completely classified based on sequence contexts. A recent study in which the methylomes of different DNA methylation pathway mutants were profiled provided evidence of crosstalk between the different methylation pathways. Non-CG methylation (i.e., CHG and CHH) at some loci was found to be dependent on CG methylation, as the loss of non-CG methylation was associated with the loss of CG methylation at certain sites in *met1*. In a *cmt3* mutant, CHG methylation at some loci remained unaffected, and the remaining methylation was *DRM1*- and/or *DRM2*-dependent. On the other hand, mutations in the CHG methylation maintenance pathway genes *CMT3*, *KYP* and *SUVH5/6* also resulted in the loss of CHH methylation at some loci, while KYP and SUVH5/6 appeared to affect CHH methylation through a pathway distinct from that of CMT3 [10].

2. DNA demethylation

The DNA methylation status throughout the genome is dynamic. At some developmental stages, DNA needs to be demethylated, and this is accomplished by either passive DNA demethylation during replication or active DNA demethylation by DNA glycosylases.

In somatic tissues, three DNA glycosylases, REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2) and DML3, actively remove DNA methylation at some loci [34]. REPRESSOR OF SILENCING 4 (ROS4), a histone acetyltransferase with an MBD domain and a PHD finger domain, acts on a subset of ROS1 targets. It is capable of binding histones without

H3K4 dimethylation or trimethylation marks through its PhD domain and also capable of binding methylated cytosines through its MBD domain. The subsequent acetylation by the histone acetyltransferase domain may provide a favorable environment for DNA glycosylases [35]. During DNA demethylation, ROS1 excises the methylated cytosine through its glycosylase/lyase activity [36, 37], the DNA phosphatase ZDP removes the 3' phosphate generated by ROS1 [38], and DNA repair system may fill in the gap with an unmethylated cytosine [9]. ATXRCC1, the *Arabidopsis* homolog of animal X-RAY REPAIR CROSS COMPLEMENTING 1 (XRCC1), may also have a role in stimulating the activity of ROS1 and ZDP [39]. How proteins involved in demethylation are guided to specific loci remains enigmatic, but the identification of the RNA-binding protein REPRESSOR OF SILENCING 3 (ROS3) as a DNA demethylation factor provided some clues. ROS3 colocalizes with ROS1 and binds small RNAs *in vivo* [40], which may mediate ROS3 guidance of ROS1 to target loci. However, how these small RNAs are produced and how target recognition is achieved is still unknown.

During male gametogenesis, the expression level of the DDM1 gene is down-regulated in the vegetative nuclei. This results in the loss of DNA methylation and activation of TEs. Twenty-one-nucleotide siRNAs generated from these TEs move into the sperm nuclei and reinforce the silencing of the sperm cell target loci [41]. The CG methylation maintenance methyltransferase gene *MET1* is down-regulated during female gametogenesis, which results in passive DNA demethylation during replication [9]. Additionally, the DNA glycosylase DEMETER (DME) actively removes DNA methylation in the central cell prior to double fertilization, causing a global reduction of CG methylation in the central cell and eventual endosperm [9]. It has been proposed that the siRNAs generated from activated TEs in the endosperm may move into the embryo to reinforce the silenced state of TEs [42, 43].

3. DNA methylation and gene expression

Although DNA methylation maintenance has been studied extensively in plants, little is known about how DNA methylation alters gene transcription. A study on DNA methylation in humans and transcription factors indicated that empirical transcription factor binding sites in the human genome are more likely to be located in CpG islands where cytosines are unmethylated than predicted transcription factor binding sites; this finding led to the hypothesis that DNA methylation may prevent transcription factor binding to DNA [5]. In the *Arabidopsis* genome, about one-third of all expressed genes have DNA methylation within the gene bodies. In contrast, the fraction of expressed genes with DNA methylation in the promoter region is only about 5%, and these genes tend to exhibit tissue-specific expression patterns [44]. Thus, DNA methylation may also influence transcription factor binding in plants. Another mode by which DNA methylation may regulate gene expression is through its effect on histone modifications. Methylated cytosine may be recognized and bound by methyl-DNA-binding protein domains such as the methyl-CpG-binding domain (MBD) and the SET and RING finger-associated (SRA) domain. In *Arabidopsis,* AtMBD5, AtMBD6 and AtMBD7 exhibit methyl-CpG binding capacity *in vitro*. Additionally, proteins with histone deacetylase activity were precipitated in a pull-down assay using MBD6 and *Arabidopsis* nuclear extracts, indicating that MBD6 may recruit histone deacetylases to methylated DNA [45]. The SRA domain is present in some histone methyltransferase proteins in *Arabidopsis*, such as KYP/SUVH4, SUVH5 and SUVH6, and the SRA domain of KYP is known to bind methylated cytosine in all sequence contexts *in vitro* [17, 46]. Thus, the recruitment of histone modifiers by DNA methylation may alter histone modification patterns and establish a local chromatin environment that represses transcription.

Histone modifications

The basic assembly unit of the chromatin is the nucleosome. The nucleosome core particle consists of about 147 bp of DNA wrapped around a histone octamer, and the DNA between two core nucleosome particles is referred to as linker DNA. Interactions with histone proteins affect the accessibility of the nucleosomal DNA to the transcription machinery. Two copies each of histones H2A, H2B, H3 and H4 comprise the histone octamer. Post-translational modifications may occur at histone H2A, H2B and H3 subunits and contribute to the epigenetic regulation of gene expression. Furthermore, transcriptional regulation by histone modifications may be accomplished on two levels: through changes in the local chromatin structure or through the recruitment of effector proteins (reviewed in [6]). In plants, H3K9 dimethylation (H3K9me2) and H3K27 trimethylation (H3K27me3) are two types of repressive modification that have been well characterized.

1. Histone H3K9me2

Whole-genome chromatin IP-chip (ChIP-chip) studies of *Arabidopsis* H3K9me2 have shown an association of this epigenetic mark with heterochromatic regions in *Arabidopsis* [14, 47]. As described above, H3K9me2 is highly correlated with CHG methylation and primarily targets transposons and repeats. While the level of H3K9me2 is relatively low at transposons and repeats on the euchromatic arms, heterochromatin usually contains long, uninterrupted regions with a high level of H3K9me2 [14].

In *Arabidopsis*, H3K9me2 is maintained by three histone methyltransferases, KYP/SUVH4, SUVH5 and SUVH6. In addition to the SET domain, which functions in catalyzing H3K9 methylation, these histone modifiers also contain the SRA domain, which facilitates the recruitment of the proteins to their target loci by binding to methylated cytosines [17, 46].

As previously mentioned, The SRA domains from H3K9 histone methyltransferases bind to methylated cytosines in all sequence contexts (CG, CHG and CHH) [17, 46]. Given that histone H3K9me2- and CMT3-mediated CHG methylation may form a feed-forward regulatory loop, it is essential for plants to prevent ectopic deposition of H3K9me2 marks at protein-coding genes with CG methylation. The Jumonji domain-containing histone demethylase INCREASE IN BONSAI METHYLATION 1 (IBM1) has been shown to prevent ectopic H3K9me2 and DNA methylation at protein-coding genes. A mutation in *IBM1* results in an increased level of DNA methylation in protein-coding genes, and the hypermethylation phenotype can be rescued by the *suvh4* mutation [48]. Whole-genome DNA methylation profiling of the *ibm1* mutant revealed that IBM1 primarily targets transcribed genes with CG methylation in the gene bodies [10, 49], suggesting that the recruitment of IBM1 to its targets is a transcription-coupled process. However, how IBM1 is precisely recruited to the target genes is still unknown.

2. Histone H3K27me3

Histone H3K27me3 is another repressive mark of gene transcription in *Arabidopsis*. Unlike H3K9me2, however, H3K27me3 is primarily found at genes that are facultatively silenced [50, 51]. In *Arabidopsis* seedling tissue, approximately 4400 genes with H3K27me3 have been identified; many of these genes encode transcription factors that regulate developmental processes and exhibit tissue-specific expression, indicating dynamic regulation of H3K27me3 over the course of plant development [50].

In both animals and plants, H3K27me3 deposition and subsequent gene silencing require Polycomb-group (PcG) proteins. Similar to what has been reported in animals, the histone methyltransferase activity of Polycomb repressive complex 2 (PRC2) maintains the H3K27me3 level at target genes in *Arabidopsis*. Among the PRC2 complex subunit genes in *Arabidopsis*,

CURLY LEAF (*CLF*), *SWINGER* (*SWN*) and *MEDEA* (*MEA*) encode homologs of the H3 methyltransferase Enhancer of Zeste (E(z)) in animals. *FERTILIZATION-INDEPENDENT ENDOSPERM* (*FIE*) encodes a homolog of the WD40 domain protein Extra Sex Combs (ESC) in animals. Three genes, *REDUCED VERNALIZATION RESPONSE 2* (*VRN2*), *FERTILIZATION INDEPENDENT SEED 2* (*FIS2*) and *EMBRYONIC FLOWER 2* (*EMF2*), encode homologs of the zinc finger protein subunit Suppressor of Zeste 12 (Su(z)12). Finally, the counterpart of the animal WD40 domain protein P55 is encoded by *MULTICOPY SUPRESSOR OF IRA1 (MSI1)* in *Arabidopsis*. These proteins form at least three types of PRC2 complexes, namely, the VRN, FIS and EMF complexes, that regulate H3K27me3 at different tissues or development stages (reviewed in [52]). The VRN complex consists of CLF/SWN, VRN2, FIE and MSI1 and functions in silencing the *FLOWERING LOCUS C* (*FLC*) gene after vernalization. The FIS complex contains MEA/SWN, FIS2, FIE and MSI1. It represses its target genes, including *PHERES1* (*PHE1*), *FUSCA3* (*FUS3*) and *MEA*, in the female gametophyte and endosperm. The EMF complex, which is likely to be composed of CLF/SWN, EMF2, FIE and MSI1, functions in maintaining vegetative growth during early development.

In addition to the PRC2 complex, the silencing of PcG targets also requires the PRC1 complex, which binds to H3K27me3 marks. Although homologs of animal PRC1 complex subunits have not been identified in *Arabidopsis*, a functional analog exists and consists of the following components: LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), EMBRYONIC FLOWER 1 (EMF1), REDUCED VERNALIZATION RESPONSE 1 (VRN1), AtRING1A/AtRING1B and AtBMI1A/AtBMI1B [53]. LHP1 contains a chromodomain and a chromo shadow domain. Although LHP1 protein binds H3K9me2, H3K9me3 and H3K27me2 *in vitro*, it only associates with loci marked with H3K27me3 *in vivo* [51]. This specificity may be conferred by the interaction of LHP1 with other proteins. In *Arabidopsis*, LHP1 is required for

the repression of PcG target genes without affecting H3K27me3 [51], but how the PRC1 complex mediates this repression is unclear. In animals, PRC1-mediated histone H2A ubiquitination is crucial for the repression of PcG target genes [54-56]. However, in Arabidopsis, the repression of PcG targets could be either dependent or independent of H2A ubiquitination [1].

Genes marked with H3K27me3 are not constitutively repressed. At certain developmental stages or in response to certain stimuli, H3K27me3 can be removed from some genes, resulting in gene activation. In *Arabidopsis*, the Jumonji domain-containing histone demethylase RELATIVE OF EARLY FLOWERING 6 (REF6) is one such enzyme that demethylates H3K27me3 and H3K27me2 to regulate gene expression [57]. In addition, *EARLY FLOWERING 6* (*ELF6*) and *JUMONJI DOMAIN-CONTAINING PROTEIN 13* (*JMJ13*), two close homologous genes of *REF6* in *Arabidopsis*, may have redundant functions [57]. It is not yet known how REF6 is recruited to certain loci to demethylate H3K27me2 and H3K27me3.

The THO complex in eukaryotes

As discussed in a later section, genes encoding *Arabidopsis* THO complex subunits were cloned from genetic screens using epigenetically regulated reporter lines in this study.

The THO complex is a nuclear complex that exists in eukaryotes. In yeast, HPR1, THO2, THP1 and MFT1 are recognized as the core components of the THO complex [58]. Additional proteins, including TEX1, SUB2 and ALY, associate with the THO complex to form the TREX (transcription-export) complex. The core human THO complex consists of HPR1/THO COMPLEX SUBUNIT 1 (THOC1), THOC2, THOC5, THOC6 and THOC7. Of these, THOC1 and THOC2 are homologous to yeast HPR1 and THO2, respectively. The animal THO complex also associates with UAP56 (a yeast SUB2 ortholog), REF1/ALY and TEX1 [59] in the TREX complex. Several lines of evidence indicate that the export factors UAP56 and ALY associate

with the THO complex in a transient manner in human cells; for example, they do not exist in the same fraction as other THO components when analyzed by gel filtration, and they only co-immunoprecipitate with other THO subunits under low salt conditions [60]. Although TEX1 is not generally considered a core component of the THO complex, it is tightly associated with the THO complex in yeast [61]. In humans and *Drosophila*, TEX1 exists in sub-stoichiometric amounts in the purified THO complex [59, 62]. Unlike *Thoc2*, *Tex1* knockdown in human cells does not affect the accumulation of other THO components. Nevertheless, TEX1 exists in the same fraction as other THO components in gel filtration analysis and co-immunoprecipitates with THOC2 even under high salt conditions [60]. Thus, TEX1 can be considered a peripheral subunit of the THO complex. In *Arabidopsis*, orthologs of each animal THO complex subunit gene exist in the genome, suggesting that the *THO5*, *THO6* and *THO7* genes evolved before the divergence of plants and animals. Through immunoprecipitation of HA-tagged TEX1 in *Arabidopsis* followed by mass spectrometry, peptides from all other THO complex core subunits (HPR1, THO2, THO5, THO6 and THO7) were identified, indicating that these subunit proteins also form a protein complex in plants [63].

1. Function of the THO complex in yeast

In yeast, the THO complex is co-transcriptionally recruited to almost all transcribed genes by RNA Pol II, but it is only required for the efficient expression of some genes. Genes that are affected in THO complex mutants are generally characterized by long gene length, high GC content, high expression levels [64] or internal repeats [65]. The THO complex is implicated in mRNA maturation and export through the recruitment of other factors such as SUB2 and ALY [66, 67]. THO complex mutants exhibit decreased accumulation of some transcripts and transcription-associated hyper-recombination [68]. Additionally, mutations in THO complex subunit genes and *SUB2* lead to defects in mRNA 3’ maturation, which in turn triggers mRNA

degradation mediated by RNA surveillance factors, such as TRF4 and RRP6 [69]. The THO complex is also recruited to snoRNA genes in yeast. However, unlike its effect on mRNAs, it negatively regulates the accumulation of snoRNAs by bridging snoRNA transcription and TRAMP complex-mediated RNA decay [70]. Another consequence of THO complex gene mutation is the formation of RNA-DNA hybrids (R-loops) during transcription [71]. The existence of R-loop during transcription diminishes the processivity of RNA Pol II. R-loop formation in THO complex mutants also accounts for the hyper-recombination phenotype in the *hpr1* mutant, as the frequency of recombination can be reduced by overexpressing RNase H, which degrades the RNA in R-loops [71].

2. Function of the THO complex in animals

In animals, components of the TREX complex, including core subunits of the THO complex, have been identified from the spliceosome [72], suggesting a role of the THO complex in mRNA splicing. The results of *in vitro* assays using human cell nuclear extracts indicate that the THO complex only associates with spliced mRNAs, as opposed to unspliced pre-mRNAs [62], promoting the hypothesis that THO complex recruitment to mRNA depends on mRNA splicing. In *Drosophila* S2 cells, depletion of THOC2 and HPR1 results in the accumulation of poly (A) mRNA in the nucleus [59]. Genome-wide microarray analyses of mRNA levels indicate that less than 20% of genes require the THO complex for proper expression [59], in contrast to UAP56 depletion, which affects the accumulation of most mRNAs in *Drosophila* S2 cells [73]. Notably, the accumulation and export of both HSP70, which lacks introns, and HSP83, which contains an intron, depends on the THO complex, indicating that the function of the THO complex may not be restricted to mRNAs with introns in *Drosophila in vivo* [59]. Similar to the findings reported in yeast, genome instability and DNA breaks are induced by the formation of R-loops in THO-depleted human cells [74]. It has also been reported that mouse THOC5/FMIP (Fms-interacting

protein) may be phosphorylated by Fms tyrosine kinase and protein kinase C (PKC). PKC-mediated phosphorylation at two serine residues located near the nuclear localization signal leads to the translocation of THOC5 from the nucleus to the cytosol [75]. However, it is not clear whether THO5 is a free protein or a component of the THO complex during this translocation.

3. Function of the THO complex in plants

So far, the understanding of the THO complex in plants is limited. A mutation in *HPR1* has been shown to affect the splicing pattern of some alternatively spliced genes, suggesting that the THO complex is involved in splicing in *Arabidopsis* [76]. Two recent studies indicate that the THO complex functions in the biogenesis of some trans-acting siRNAs (ta-siRNAs) as well as the biogenesis of siRNAs involved in other post-transcriptional gene silencing (PTGS) pathways, including sense transgene PTGS (S-PTGS) and inverted repeat-induced PTGS (IR-PTGS). Mutations in *HPR1*, *TEX1* and *THO6* result in the decreased accumulation of siRNAs involved in PTGS pathways [63, 77]. The accumulation of TAS1 and TAS2 transcripts, which serve as precursors for ta-siRNA generation, is increased in THO complex mutants. However, the accumulation of miR173, which cleaves TAS1 and TAS2 transcripts to trigger ta-siRNA formation, is not affected. Notably, the effect of a mutation in *TEX1* on siRNA biogenesis is much stronger than mutations in *HPR1* and *THO6* (both of which are single-copy genes in the *Arabidopsis* genome; furthermore, the mutant alleles discussed are both likely to be null alleles). Additionally, a mutation in the THOC2 gene is known to be embryonic lethal in homozygous plants [63, 76]. The structure of the THO complex in yeast indicates that THOC2 directly interacts with nucleic acids [61]. Considering all of these previous findings in aggregate, the THO complex is likely to have a broad effect on mRNAs in *Arabidopsis*, while some subunits may have regulatory functions at specific loci. It remains to be determined where the THO complex is recruited in the genome and which genes are affected by mutations disrupting the THO complex.

Despite the reported functions of the THO complex in plants, no genome-wide studies of THO complex mutants aimed at characterizing the potential target genes of the THO complex in plants have been reported thus far. In the present study, we used two *LUCIFERASE* reporter lines that are regulated by DNA methylation to screen for mutants with decreased luciferase activity. We isolated three mutations that are located in three genes encoding THO complex subunits. We performed genome-wide analyses of small RNA accumulation, DNA methylation and mRNA transcript levels using wild-type plants and the THO complex mutants isolated from the screens. We found that the mutations in THO complex subunit genes do not have obvious impacts on DNA methylation, despite having been isolated from screens employing reporter lines that are regulated by DNA methylation. From the mRNA sequencing data, we identified genes with altered transcript levels in the *tex1* and *hpr1* mutants, and we characterized some of the features of the down-regulated genes in the two mutants.

Results

Genetic screens using two *LUCIFERASE* reporter lines to detect changes in DNA methylation

In the last decade, significant progress has been made in the study of DNA methylation, and many factors involved in establishing and maintaining DNA methylation have been identified. The available evidence indicates that plants have elaborate pathways for faithfully maintaining DNA methylation at target loci over multiple generations while simultaneously preventing DNA methylation from spreading to neighboring regions. Nevertheless, it remains unclear how DNA methylation, and the RdDM pathway in particular, is restricted to certain loci rather than expanding into neighboring regions through sporadic events. To identify potential regulators that

negatively regulate DNA methylation, we carried out genetic screens using *LUCH* and *YJ11-3F*, two luciferase reporter lines that are regulated by DNA methylation.

A construct containing the firefly *LUCIFERASE* (*LUC*) gene fused with the miR172 binding site of *AP2* under the regulation of a dual *35S* promoter (Figure 2.1A) was used to generate the *LUCH* reporter line, and the gene *AT3G07350* was identified as the site of insertion. To prevent S-PTGS, *LUCH* was combined with a mutation in *RNA-DEPENDENT RNA POLYMERASE 6* (*RDR6*), *rdr6-11*. As previously reported [78], this reporter line is not regulated by the miRNA pathway despite the presence of a miRNA binding site. Instead, the promoter region of the reporter construct is methylated, and mutations in RdDM pathway genes, such as those in *AGO4, DRD1* and *DRM2*, were found to de-repress the silencing of *LUC*. These findings indicate that *LUCH* is regulated by the RdDM pathway. Whole-genome bisulfite sequencing of two biological replicates of *LUCH* revealed methylation at all three sequence contexts in both the promoter and the *LUC* coding region. Furthermore, CHH methylation at the transgene was reduced when either *AGO4* or *DRD1* was mutated (Figure 2.2A). A mutation in *ROS1* further enhanced the silencing of the reporter gene, indicating that the reporter is also targeted by ROS1 [78]. Thus, this reporter line represents an effective tool for the isolation of mutations that affect DNA methylation through either the RdDM or ROS1 pathway. A construct similar to that in *LUCH* was used to generate the second reporter line, *YJ11-3F*, except that the 3' UTR of *LUC* was replaced with a miR173 binding site instead of a miR172 binding site (Figure 2.1A). As with *LUCH*, *YJ11-3F* was also introduced into the *rdr6-11* mutant background. In the *YJ11-3F* line, the transgene is inserted into the 3' UTR of the gene *AT1G02740*, and both luminescence and *LUC* transcript levels were found to be higher in *YJ11-3F* than in *LUCH* (Figure 2.3B). The introduction of a weak allele of *AGO1*, *ago1-45*, did not have any obvious effect on *LUC* gene expression (Figure 2.1B). In addition, the introduction of *drd1-12* did not affect *LUC* expression

either (Figure 2.1B). Nonetheless, *ros1* and *ros4* mutants isolated from the genetic screen exhibited decreased luminescence compared to *YJ11-3F* (data not shown), indicating that this reporter line is regulated by the DNA demethylation pathway. Bisulfite sequencing data showed that the promoter region of the transgene in *YJ11-3F* was methylated in all three cytosine sequence contexts, whereas the *LUC* coding region was unmethylated (Figure 2.2B). The *YJ11-3F* reporter line was therefore employed for the isolation of negative regulators of DNA methylation. Although both *LUCH* and *YJ11-3F* are regulated by DNA methylation, the possibility that the two reporter lines are also regulated by other mechanisms cannot be excluded.

Using these two reporter lines, we performed both EMS and T-DNA insertional mutagenesis and screened for mutants with reduced luminescence. Three mutant lines, *28-6L* (from the EMS mutagenesis of *LUCH*), *P204R* (from the EMS mutagenesis of *YJ11-3F*) and *TL525L* (from the T-DNA insertional mutagenesis of *YJ11-3F*), with reduced luminescence phenotypes were isolated (Figure 2.3A). Semi-quantitative RT-PCR confirmed that the accumulation of *LUC* mRNA was reduced in the mutant lines compared to the respective reporter lines (i.e., *LUCH* or *YJ11-3F*) (Figure 2.3B).

Cloning of the mutated genes in *28-6L, P204R* and *TL525L*

For *28-6L* (hereafter referred to as *LUCH tex1-5*), the mutation was identified by map-based cloning and Sanger sequencing. A G-to-A mutation at the splice acceptor site in the sixth intron of the *TEX1* gene (*AT5G56130*) was identified (Figure 2.4A). The mutation resulted in intron retention and reduced *TEX1* transcript accumulation (Figure 2.4B). Like TEX1 proteins in other eukaryotes, *Arabidopsis* TEX1 contains a WD40 domain that is predicted to mediate protein interactions. A transgene encoding N-terminal HA-tagged TEX1 under the endogenous promoter (published in [63]) was transformed into the mutant line for complementation analysis.

In the line *P204R* (hereafter referred to as *YJ11-3F hpr1-5*), a G-to-A mutation in the 16th exon of the gene *HPR1* (*AT5G09860*) was identified (Figure 2.4A). This mutation introduced a premature stop codon and reduced *HPR1* transcript accumulation (Figure 2.4B). The single predicted domain in *Arabidopsis* HPR1 is the efThoc1 domain, which is only present in HPR1 homologs in eukaryotes. A transgene encoding N-terminal HA-tagged HPR1 driven by its own promoter (the 1.4 kb region upstream of the transcription start site of *HPR1*) was cloned into a binary vector and transformed into the mutant line. The transformed plants exhibited luminescence levels similar to that of the reporter line *YJ11-3F* (Figure 2.3A).

A third THO complex subunit gene was cloned from the line *TL525L* (hereafter referred to as *YJ11-3F tho5a-1*), in which the T-DNA was inserted into the fifth exon of the *THO5A* gene (*AT5G42920*) (Figure 2.4A). No transcripts were detected from the region downstream of the T-DNA insertion site (Figure 2.4B). The *Arabidopsis* genome encodes two *THO5* homologs, and we named the second gene (*AT1G45233*) *THO5B*. Both THO5A and THO5B contain FimP domains, which are only present in THO5 homologs in eukaryotes. However, due to the low sequence similarity between the plant and animal THO5 proteins, it is not clear whether *Arabidopsis* THO5 proteins also contain phosphorylation sites that mediate subcellular localization signaling. A C-terminal *GFP*-tagged *THO5A* transgene driven by its own promoter (the 2.3 kb region upstream of the transcription start site of *THO5A*) was found to complement the low luminescence phenotype of the mutant (Figure 2.3A).

Genes encoding putative THO complex subunits in *Arabidopsis* and their protein sequence similarity with human THO complex orthologs are summarized in Table 2.1.

DNA methylation levels at the *LUC* reporter and endogenous loci were unaffected by the THO complex mutations

In light of the regulation of the reporter lines by DNA methylation and the regulation of *LUCH* by the RdDM pathway in particular, one possibility was that the THO complex mutations affected the level of DNA methylation and, consequently, *LUC* reporter activity. To test the hypothesis that the mutations affected the RdDM pathway, the *ago4-6* and *drd1-12* mutations were introduced into the *LUCH tex1-5* and *LUCH hpr1-5* lines (the *hpr1-5* mutation was introduced into the *LUCH* background). As shown in Figure 2.5A, the luminescence levels of *LUCH tex1-5 ago4-6* and *LUCH tex1-5 drd1-12* were higher than *LUCH tex1-5* and lower than *LUCH ago4-6* and *LUCH drd1-12*, respectively. Similar luminescence patterns were observed for the *hpr1-5* mutant combinations (Figure 2.5A). The lack of epistatic effects suggests that *TEX1* and *HPR1* function through a mechanism distinct from the RdDM pathway in terms of the regulation of the *LUC* reporter gene.

To assess the DNA methylation status at the *LUC* gene as well as endogenous loci, we performed whole-genome bisulfite sequencing for two biological replicates of each of the THO complex mutants and related controls. We first analyzed the DNA methylation status at the dual 35S promoter and the *LUC* coding region in *LUCH tex1-5*, *YJ11-3F hpr1-5* and their respective control lines. No significant changes in DNA methylation at the transgene were observed in the THO complex mutants (Figure 2.5B), indicating that the DNA methylation status of the transgene was not affected by the mutations in *TEX1* and *HPR1*.

We also analyzed the DNA methylation status of endogenous loci in the *tex1-5*, *hpr1-5* and *tho5a-1* mutants to determine whether the THO complex is a general regulator of DNA methylation. We compared the DNA methylation levels of the mutant lines and their respective

controls and identified differentially methylated regions (DMRs) in the mutants. The number of common DMRs between the two biological replicates performed for each mutant is shown in Figure 2.5C and Table 2.2. As expected, the RdDM mutants included in the analysis (*nrpe1-11*, *ago4-6* and *drd1-12*) had large numbers of hypomethylated CHH and CHG DMRs. On the other hand, very few CHH and CHG DMRs were identified in the THO complex mutants (Figure 2.5C and Table 2.2), indicating that these mutations do not have obvious effects on DNA methylation throughout the genome. Based on these findings, it is unlikely that the THO complex directly participates in the DNA methylation pathways.

The abundance of most heterochromatic siRNAs was not affected by the THO complex mutations

The THO complex has been implicated in the biogenesis of some siRNAs involved in S-PTGS and IR-PTGS as well as some ta-siRNAs. Previous studies have also shown that a mutation in *TEX1* does not affect the abundance of Pol IV-dependent 24-nt heterochromatic siRNAs [21, 63]. To confirm this finding on a genome-wide scale, we sequenced the small RNAs isolated from each of the THO complex mutants and the control lines. Because 24-nt heterochromatic siRNAs are much more abundant than other types of small RNAs in the 18- to 28-nt size range, mutations in heterochromatic siRNA biogenesis genes significantly alter the small RNA length distribution. Small RNA sequencing revealed similar length distributions for the *tex1-5* and *hpr1-5* mutants and the wild-type control (Figure 2.6A), indicating that the THO complex is not a general player in the heterochromatic siRNA biogenesis pathway. Additionally, northern blots for two Pol IV-dependent heterochromatic siRNAs demonstrated that *TEX1* and *HPR1* were not required for their accumulation (Figure 2.6B).

Identification of endogenous loci affected by the mutations in *TEX1* and *HPR1*

Although the *Arabidopsis tho2* mutant is embryonic lethal, mutations in other single-copy THO complex subunit genes do not seem to affect the viability of *Arabidopsis* [63]. Nevertheless, unlike mutants that only affect ta-siRNA biogenesis and exhibit subtle developmental defects, such as *rdr6* and *suppressor of silencing 3* (*sgs3*), *hpr1-5* and *tex1-5* mutant plants were small and exhibited reduced fertility compared to wild-type plants (Figure 2.7). Thus, the mutations in *HPR1* and *TEX1* appear to affect a broader range of genes, as opposed to specifically affecting ta-siRNA-regulated genes. This finding also raised the possibility that the mechanism(s) by which the THO complex regulates endogenous genes may be related to the mechanism underlying the regulation of the *LUC* reporter.

To identify endogenous genes with altered expression in the *hpr1* and *tex1* mutants, we performed mRNA sequencing for Col, *hpr1-5*, *tex1-5* and *nrpe1-11* (as a control). As expected, the transcript level changes in *tex1-5* were well correlated with those in *hpr1-5*. In other words, most genes that were up/down-regulated in *hpr1-5* were also up/down-regulated in *tex1-5*, respectively, and *vice versa* (Figure 2.8A). This is consistent with the hypothesis that TEX1 and HPR1 work together in the THO complex to regulate gene expression.

From the mRNA sequencing analysis, a few hundred genes were identified as having a greater than 2-fold increase or decrease (FDR<0.05) in *tex1-5* or *hpr1-5* (Table 2.3). The up-regulated or down-regulated loci are listed in Table 2.5 and Table 2.6 respectively. The transcript level changes of some genes were further validated by RT-PCR (Figure 2.8B). However, these analyses do not resolve whether the affected genes were directly or indirectly affected by the mutations in *TEX1* or *HPR1*.

Genes that depend on the THO complex in yeast tend to be long genes or genes with a high GC content [64]. To test whether these features also characterize the THO complex-dependent genes identified in *Arabidopsis*, we analyzed the genes that were down-regulated in *tex1-5* and *hpr1-5* along with 500 randomly selected genes that did not exhibit a greater than 2-fold change in the two mutants (see materials and methods). As shown in Figure 2.9A and Figure 2.9B, the length distribution and GC content of down-regulated genes in *tex1-5* and *hpr1-5* did not obviously differ from the corresponding features of the control genes. In animals, the THO complex associates with the splicing complex, so we also analyzed the number of exons in genes down-regulated in *hpr1-5* and *tex1-5*. As shown in Figure 2.9C, genes with reduced transcript levels in *tex1-5* or *hpr1-5* did not seem to have more exons compared to control genes.

It has also been shown that genes with internal repeats require the THO complex for proper expression in yeast [65]. We therefore analyzed the distance between the down-regulated genes in the THO complex mutants to the repeats closest to these genes. As shown in Figure 2.10A top panel, there was a slight enrichment of genes located close to repeats among the genes down-regulated in *hpr1-5* and *tex1-5*. A similar enrichment was not observed for genes that were up-regulated in *hpr1-5* and *tex1-5* (Figure 2.10A bottom panel). Next, we classified the repeats as tandem repeats, dispersed repeats or inverted repeats and analyzed the percentage of down-regulated genes in *tex1-5* and *hpr1-5* that are located within 1 kb of each repeat type. For dispersed repeats and inverted repeats, we observed an enrichment of genes closely located to these repeat types (i.e., located within 1 kb of the repeats). For tandem repeats, no obvious enrichment was observed (Figure 2.10B). These findings suggest that the THO complex is required for the expression of some genes that are associated with or in close proximity to certain types of repeats. In the *LUC* reporter transgene, the dual *35S* promoter is consisted of two

tandemly located, long repeats. The *35S* promoter is similar to dispersed repeats because the marker gene *NPTII* is driven by a similar promoter.

In the genome, dispersed repeats are often found in multi-copy loci (genes and transposons) as well as the terminal repeats of LTR retrotransposons, while inverted repeats are characteristic of DNA transposons. As a result, some repeats are targeted for silencing by DNA methylation or H3K9me2. We therefore analyzed the distance between the down-regulated genes in *tex1-5* and *hpr1-5* to RdDM target loci represented by hypomethylated CHH DMRs identified in *nrpe1-11* and *sde4-3*. From this analysis, we observed a slight enrichment of genes associated with *nrpe1* and *sde4* hypomethylated CHH DMRs among the down-regulated genes in the *hpr1-5* and *tex1-5* mutants (Figure 2.11A). However, the percentage of genes located within 1 kb of *nrpe1* and *sde4* CHH DMRs among the down-regulated genes in *hpr1-5* and *tex1-5* did not show a statistically significant difference from the control percentage (data not shown). A similar observation was observed for the H3K9me2 mark (Figure 2.11B). We analyzed the composition of the down-regulated genes in both *hpr1-5* and *tex1-5* in terms of association with repeats, RdDM target loci and H3K9me2 (Figure 2.11C). Among the 134 down-regulated genes in both *hpr-5* and *tex1-5*, 59 genes associate with dispersed repeats or inverted repeats, of which 27 genes were not in close proximity to H3K9me2 or RdDM target loci. On the contrary, most H3K9me2- or RdDM target loci-associated genes were also repeat-associated. Therefore, the slight enrichment of genes in closely proximity to H3K9me2 or RdDM target loci was likely an indirect effect of the enrichment of repeat-associated genes among the down-regulated genes in *hpr1-5* and *tex1-5*.

Next, we analyzed the distance of differentially expressed genes in *tex1-5* and *hpr1-5* from regions containing H3K27me3, another well-characterized histone mark. Surprisingly, there was a marked enrichment of genes in close proximity to H3K27me3 marks for both down-regulated and up-regulated genes (Figure 2.12A and B) in *hpr1-5* and *tex1-5*. The percentage of genes with

H3K27me3 among the up-regulated and down-regulated genes in *hpr1-5* and *tex1-5* was calculated and a significant increase was observed compared to the control percentage (Figure 2.12C). Furthermore, the composition of the down-regulated genes in both *hpr1-5* and *tex1-5* in terms of repeat and H3K27me3 association was analyzed. We found that among all the down-regulated genes in *hpr1-5* and *tex1-5* with H3K27me3, about half of the genes are associated with repeats, and *vice versa* (Figure 2.12D).

We subsequently analyzed the expression patterns of dispersed or inverted repeat-associated loci in *tex1-5* and *hpr1-5* to see whether mutations in *TEX1* and *HPR1* alter the expression of genes associated with repeats in general. We plotted the transcript fold change in *tex1-5* (Figure 2.13A left panel) and *hpr1-5* (Figure 2.13A right panel) of all dispersed repeat-associated genes against average transcript levels in Col (RPKM). Among the dispersed repeat-associated genes that were not expressed at very low levels (the fold change of genes with very low expression levels in Col may not be accurate), up-regulated and down-regulated genes as well as genes whose transcript levels were not altered in the mutants were identified. Similar expression patterns were also observed in inverted repeat-associated loci in *tex1-5* and *hpr1-5* (Figure 2.13B). We also identified genes with H3K27me3 marks in 10-day old seedlings [50] (the same stage used for mRNA sequencing) then analyzed the expression patterns of these genes in *hpr1-5* and *tex1-5* relative to Col (Figure 2.13C). Similarly, among the H3K27me3-marked genes, we did not observe a particular pattern of expression change in either *tex1-5* or *hpr1-5*. Thus, the association with a dispersed repeat, an inverted repeat or H3K27me3 itself is not indicative of the requirement of TEX1 and HPR1 at a given locus.

Discussion

In this study, we used two luciferase reporter lines, *LUCH* and *YJ11-3F*, in screens for mutants with altered DNA methylation. We isolated three mutant lines with reduced luminescence compared to their respective controls and identified the disrupted genes. The three genes were found to encode three subunits (TEX1, HPR1 and THO5) of the *Arabidopsis* THO complex. The THO complex is a conserved nuclear protein complex that functions in transcription elongation and mRNA export in yeast and animals. In yeast, the THO complex also prevents transcription-associated genome instability by preventing R-loop formation during transcription. Although the two luciferase reporter lines are regulated by genes involved in DNA methylation pathways, this does not preclude the possibility that the reporter transgenes are also regulated by other pathways affecting gene expression. Based on the function of the THO complex in yeast and animals, the three THO complex subunits encoded by the genes isolated in the study are likely to function co-transcriptionally or post-transcriptionally in regulating *LUC* mRNA levels. We performed several analyses to rule out the possibility that the regulation of *LUC* by the THO complex was accomplished through DNA methylation pathways. Our genetic analysis of the THO complex mutants and RdDM mutants indicated that *tex1-5* and *hpr1-5* affect *LUC* expression through a pathway distinct from the RdDM pathway. Although *TEX1* and *HPR1* are known to be required for the accumulation of some endogenous siRNAs involved in PTGS, we confirmed that they are not general regulators affecting the accumulation of most 24-nt heterochromatic siRNAs, which function in DNA methylation. Most importantly, through whole-genome DNA methylation profiling, we found that DNA methylation was not significantly affected at the *LUC* reporter or endogenous loci in the THO complex mutants. Taken together, these findings indicate that the THO complex does not regulate DNA methylation.

To identify genes affected in *tex1-5* and *hpr1-5*, we performed mRNA sequencing analysis for the two mutants and wild-type. Many genes were up-regulated or down-regulated in *tex1-5* and *hpr1-5*, and the differentially expressed genes in the two mutants were well correlated. This finding was consistent with the hypothesis that TEX1 and HPR1 function together in the THO complex. Among the differentially expressed genes in the two mutants, some genes may be direct targets of the THO complex, while others may be indirectly affected. Although the THO complex is recruited to almost all Pol II-transcribed genes in yeast [64], whether the same is true in plants is unknown. Future studies using chromatin IP followed by high-throughput sequencing (ChIP-seq) or RNA IP-seq (RIP-seq) will provide insight into the direct targets of the THO complex in plants.

From studies in yeast and animals, it is clear that the THO complex is involved in multiple processes during mRNA biogenesis and export. The complex functions as a platform for the recruitment of proteins with specialized functions to nascent RNA. Thus, mutations in THO complex genes may lead to multiple defects in mRNA processing and export. Although the THO complex is a stable complex and all the core subunits work together during mRNA transcription and export, it is nevertheless possible that some THO complex subunits have specialized functions and are only required at a subset of THO complex-associated loci. Indeed, the phenotypic difference between *tho2*, which is embryonic lethal, and other mutants disrupting single-copy genes encoding THO complex subunits (e.g., *tex1*, *hpr1* and *tho6*) [63] indicates that some genes that depend on *THO2* do not stringently depend on other subunit genes of the THO complex. In addition, although most genes with transcript level changes were well correlated between *tex1* and *hpr1* in the mRNA sequencing analysis, some genes were differentially regulated in the two mutants (data not shown). Further analysis of genes that are differentially

affected in different THO complex subunit mutants may help improve our understanding of the specialized functions of the individual subunits.

Using the mRNA sequencing data, we analyzed genes that were down-regulated or up-regulated by more than 2-fold in *tex1* and *hpr1*. No significant differences were observed between the down-regulated genes in the mutants and unaffected control genes in terms of GC content, gene length and exon number. There was an enrichment of genes associated with or in close proximity to dispersed repeats, inverted repeats and H3K27me3 among the genes down-regulated in *tex1* and *hpr1*. However, not all of the down-regulated genes in *tex1* and *hpr1* exhibited these particular features, and, conversely, genes characterized by one of these features were not always down-regulated in *tex1* and *hpr1*. It is unlikely that the THO complex is directly recruited by repeats and H3K27me3. Instead, certain genes with these features are more likely to be regulated by TEX1 and HPR1, possibly through the interaction between the THO complex and some unknown factors existing at these loci.

Some repeats co-localize with RdDM target loci and histone H3K9me2. The *LUC* reporter locus is similar to these types of loci in this regard because the dual *35S* promoter resembles a dispersed repeat and is regulated by DNA methylation. However, genetic analysis indicated that TEX1 and HPR1 do not act through the RdDM pathway at the *LUC* locus. Thus, DNA methylation itself may not be the factor that determines the dependency of the *LUC* locus on THO. Furthermore, many *TEX1*- and *HPR1*-dependent repeat-associated genes were not closely associated with DNA methylation and H3K9me2.

H3K27me3 often occurs at genes that are facultatively silenced. Both the up-regulated and down-regulated genes in *tex1* and *hpr1* were enriched for genes with H3K27me3. On the other hand, biased expression patterns were not observed for genes with H3K27me3 marks in *tex1* and

hpr1. It is not clear whether the THO complex directly functions at target loci marked by H3K27me3. One possibility is that the THO complex indirectly regulate some H3K27me3 target loci by regulating a protein that functions at these loci. In this case, the downstream effector is likely to be a regulator of H3K27me3-mediated silencing rather than a component directly involved in H3K27 methylation or demethylation. Alternatively, the THO complex may be a direct regulator at some loci silenced by H3K27me3. Studies on the CUL4-DDB1 (CULLIN 4-DAMAGED DNA BINDING PROTEIN 1) E3 ligase complex shed light on the potential interaction between the THO complex and the H3K27me3 pathway. *Arabidopsis* THO6 has been found to interact with DDB1B *in vivo*, and TEX1 is also a potential target of DDB1 [79]. Furthermore, the CUL4-DDB1 complex interacts with MSI1, a subunit of the PRC2 complex, which methylates H3K27me3 [80]. It remains to be determined whether the CUL4-DDB1 complex at H3K27me3 sites regulates the occupancy or activity of the THO complex.

Preventing R-loop formation may also represent a major function of the THO complex at its target loci in plants. A recent study suggested that the stabilization of the R-loop generated from a non-coding RNA near *FLOWER LOCUS C* (*FLC*) inhibits the transcription of the non-coding RNA, which in turn affects the expression of *FLC* [81]. Non-coding RNAs often associate with PcG target genes. The THO complex may also affect the transcription of non-coding RNAs by preventing R-loop stabilization. In this context, genome-wide profiling of R-loops in wild-type and THO complex mutants may help improve the understanding the role of R-loop formation and gene regulation in plants.

Materials and Methods

Arabidopsis strains and growth conditions

Plants were grown on Murashige and Skoog (MS) medium under continuous light at 22°C. 10 days after germination, the seedlings were used for experiments or transferred to soil to grow under long-day (16 h light/ 8 h darkness) conditions at 22°C for harvesting seeds.

All *Arabidopsis* plants used for mutagenesis and genetic analysis are in Columbia accession. *LUCH* and *YJ11-3F* were crossed with wild type Ler for at least 5 times and plants that were homozygote for the *LUC* transgene and *rdr6-11* were selected in the progenies to use in map-based cloning.

Genotyping information for mutant lines can be found in table 2.4. Mutation in *tex1-5* was genotyped by PCR amplification using primer pair At5g56130seq_F2 and At5g56130seq_R1 followed by AlwNI digestion. Products were separated on 1% gel and band from *tex1-5* is larger than WT band. Mutation in *hpr1-5* was genotyped by PCR amplification using primer pair HPR1-F and HPR1-R followed by MboII digestion. Products were separated on 4% agarose gel and *hpr1-5* mutant DNA band was smaller than WT band. Mutation in *tho5a-1* were genotyped by PCR amplification using primer pairs AT5G42920seq_F2 / AT5G42920seq_R2 (only amplifies WT DNA) and AT5G42920seq_F2 / pEGLmR1 (only amplifies DNA with T-DNA insertion in *tho5a-1*). Mutation in *ago1-45* was genotyped by PCR amplification using primer pair ago1-45 dCAP SphI F and ago1-45 dCAP sphI R followed by SphI digestion. Digested products were separated on 4% agarose gel and wild-type DNA band was smaller. Mutation in *ago4-6* was genotyped by PCR amplification using primer pair 115-1H gtF and 115-1H gtR followed by AluI digestion. Digested products were separated on 4% gel and *ago4-6* mutant band was smaller. Mutation in *drd1-12* was genotyped by PCR amplification using primer pair 86-1H gtF1-TaqI

and 86-1H gtR1 followed by TaqαI digestion. Digested products were separated on 4% agarose gel and *drd1-12* mutant band was smaller. Mutation in *rdr6-11* was genotyped by PCR amplification using primer pair rdr6-11_gtF4 and rdr6-11_gtR4 followed by ClaI digestion. Digested products were separated in 4% gel and wild-type band was smaller. *LUC* insertion in *LUCH* was genotyped by PCR amplification using primer pairs TailcR / 9-7-2gtF (only amplifies DNA without transgene insertion) and TailcR / R1 (only amplifies DNA with *LUC* transgene insertion). *LUC* insertion in *YJ11-3F* was genotyped using primer pairs YJ11-3F_For / YJ11-3F_Rev (only amplifies DNA without transgene insertion) and YJ11-3F_Rev / R1 (only amplifies DNA with *LUC* transgene insertion).

T-DNA insertation mutagenesis and mutant screening

A pEarleyGate 303 vector [82] without *ccdB* gene was transformed into Agrobacterium strain GV3101. This strain was used to transform 5-week old *YJ11-3F* plants by floral dip transformation. F0 Seeds were harvested and selected with Basta and seeds from Basta-resistant F1 plants were harvested independently to get F2 lines for screening.

Mutagenized F2 lines from *YJ11-3F* were planted on MS plates together with *YJ11-3F*. 10 days after germination, F2 lines were screened for plants with lower luminescence than *YJ11-3F* by luminescence imaging as previously described [78].

Map-based cloning

Recessive mutant with lower luminescence phenotype were crossed with the reporter line with L*er* background. F2 plants with lower luminescence phenotype were identified by luminescence imaging and used for mapping population. Map-based cloning were performed by using SNP and InDel polymorphisms between Col and L*er*. Candidate genes were sequenced by Sanger sequencing using PCR product amplified from mutant DNA with sequencing primers.

Oligo sequences for PCR amplification and sequencing of *TEX1* (*AT5G56130*), *HPR1* (*AT5G09860*) and *THO5A* (*AT5G42920*) are listed in table 2.4.

Plasmid construction and complementation

pTEX1:HA-TEX1 was obtained from Balcombe lab [63] and transformed into *rdr6-11*. The transgenic plants were selected on MS plates with kanamycin. Transgenic plants were crossed with *LUCH tex1-5* and F2 plants from different transgenic parents will be genotyped for complementation study.

To construct *pHPR1:HA-HPR1* plasmid, the 1.4kb promoter region and 5'UTR of the *HPR1* gene was amplified using Phusion DNA polymerase with the primer pair HPR1_PRO-SacII_F and HPR1_PRO-SpeI_R from total genomic DNA. The amplified fragment (1.6kb) was cloned into pGEM®-T Easy Vector to get *pGEM-T-HPR1_PRO*. pJL-Blue and *pGEM-T-HPR1_PRO* vector were digested with SacII and SpeI. The HPR1_PRO fragment (1.6kb) was ligated with pJl-Blue fragment to generate *pJL-HPR1_PRO*. The coding region and 3'UTR of the *HPR1* gene was amplified with the primer pair HA_HPR1_CDS-SpeI_F (with HA tag sequence) and HPR1_CDS-XhoI_R from total genomic DNA. The amplified fragment (4.1kb) cloned into pGEM®-T Easy Vector to generate *pGEM-T-HA-HPR1*. The *pJL-HPR1_PRO* and the *pGEM-T-HA-HPR1* plasmids were digested with SpeI and XhoI. The HA-HPR1 fragment (4.1k) was ligated with the pJL-HPR1_PRO fragment (4k) to generate *pJL-pHPR1:HA-HPR1*. This clone was confirmed by Sanger sequencing. Then the plasmid was linearized and cloned into the destination vector pEG301 by LR reaction to generate the binary plasmid *pHPR1:HA-HPR1*. The plasmid was used for transforming *YJ11-3F hpr1-5* by agrobacterium-mediated transformation and T1 transgenic plants were selected by Basta.

To generate *pTHO5A:THO5A-GFP*, genomic sequence of *at5g42920* with ~2.3kb promoter (total length ~5.5kb) was amplified using Phusion DNA polymerase with the following primers thoc5a_SacII_F and thoc5a_SpeI_R_N_stop to get the fragment pTHO5:THO5 with a mutated stop codon for C-terminal fusion. The pTHO5A:THO5A fragment and pJL-Blue vector were digested using SacII and SpeI and the pTHO5A:THO5A fragment was ligated with the pJL-blue fragment to get the *pJL-pTHO5A:THO5A*. The *pJL-pTHO5A:THO5A* vector was linearized and cloned in the destination vector *pMDC107* by LR reaction to get the binary plasmid *pTHO5A:THO5A-GFP*. The plasmid was used to transform *YJ11-3F tho5a-1* by agrobacterium-mediated transformation and T1 transgenic plants were selected on MS medium with hygromycin B.

Reverse transcription and PCR

RNA extraction, Reverse transcription, semi-quantitative PCR and real-time PCR were performed as described [83]. RNA was extracted from the aerial tissues of 10-day old seedlings. *UBIQUTIN 5* was used as an internal control for PCR. Primers used are listed in Table 2.4.

Small RNA northen blotting

Small RNA northern blotting were performed as previously described [83]. RNA from the aerial tissues of 10-day old seedlings were used. 5'-end-labeled antisense DNA probes were used to detect small RNAs from total RNA. Sequences of probes are listed in table 2.4.

Bisulfite sequencing library construction and data analysis

To generate whole genome bisulfite sequencing library, 1~5μg of genomic DNA was sonicated to 100~300 bp using the Diagenode Bioruptor using the parameters: intensity=high, on=30s, off=30s, total time=60min. Fragmented DNA was purified using PureLink® PCR Purification Kit (Life Sciences). Purified DNA was end repaired using End-It kit (Epicentre) with dATP,dGTP and dTTP replaced with dNTPs for 45 min at room temperature The end-repaired

DNA was purified with Agencourt AMPure XP beads (Beckman Coulter) and adenylated with dATP and Klenow 3'-5' exo- (New England Biolabs) for 30 min at 37°C and then purified with a Agencourt AMPure XP beads. The purified DNA was ligated overnight at 16°C to genomic DNA adapters from Illumina Kit A (Illumina) with T4 DNA Ligase (New England Biolabs). Ligation products were purified with AMPure XP beads twice. Less than 400ng ligated product was used for bisulfite conversion using the MethylCode Kit (Invitrogen) following the manufacturer's guidelines except 12ug carrier RNA (Qiagen) was added into the conversion product before column purification. The final conversion product was amplified using Pfu Cx Turbo (Agilent) using the following PCR conditions (2 minutes at 95°C, 9 cycles of 15 seconds at 98°C, 30 seconds at 60°C, 4 minutes at 72°C and 10 minutes at 72°C). PCR product was purified by AMPure XP beads. BS-seq libraries were sequenced for 101 cycles (single end) using Illumina HiSeq 2000.

Raw reads from sequencing were processed to remove low-quality reads and sibling reads. Then the reads were mapped to original Tair 10 *Arabidopsis* genome and C-T converted genome with default settings. Only reads were perfectly mapped to unique locus of the genomes were retained.

Differentially Methylated Regions (DMRs) were identified between wild type and mutant as described [10] with modifications. Briefly, the genome was divided into 100bp windows. Only windows with at least 4 Cs that were sequenced for 4 times each were kept for DMR analysis. Methylation level was the total methylated "C"s against total "C"s. "C"s in CG, CHG and CHH sequence contexts were calculated separately. Regions with methylation differences above 0.4, 0.2 and 0.1 for CG, CHG and CHH respectively between wild-type and mutant (adjusted p-value(FDR)<0.01) were defined as DMRs. Two biological replicates were used in whole genome bisulfite sequencing and only DMRs that exist in both replicates were retained.

Small RNA sequencing library construction and data analysis

Small RNA libraries were generated as previously described [83] using the aerial tissues of 10-day old seedlings. Small RNA libraries were sequenced for 50 cycles (single end) using Illumina Hiseq2000.

Raw reads from small RNA sequencing were processed as previously described [84]. Small RNA reads after processing were mapped to Tair 10 *Arabidopsis* genome with SOAP2 [85]. 18-28nt small RNAs were selected for small RNA analysis.

mRNA sequencing and data anlysis

RNAs extracted from aerial tissues of 10-day old seedlings were used for library construction. mRNA sequencing libraries were prepared using Illumina TruSeq RNA Sample Preparation Kits (V2) following manufacturer's instruction.

Raw reads were filtered to remove low-quality reads and sibling reads. Then reads are mapped to Tair 10 *Arabidopsis* genome (up to 2 mismatches were allowed) and reads that map to multiple loci were removed. Sequencing reads whose 5' nucleotide falls into an annotated locus (gene or TE) were counted as those belonging to this locus. Expression difference between mutant and wild type from three biological replicates were analyzed by EdgeR [86]. Loci with 2-fold transcript increase in mutant (adjusted p-value (FDR) <0.05) were identified as up-regulated loci in mutant. Loci with 2-fold transcript decrease in mutant (FDR<0.05) and more than 5 RPKM (reads per kb per million) in Col (average number from 3 replicates) were identified as down-regulated loci in mutant. The control gene sets were selected from genes that were neither up-regulated nor down-regulated in *hpr1-5* and *tex1-5*. The control gene set for up-regulated genes contain 500 randomly selected genes. The control gene set for down-regulated genes

contain 500 randomly selected genes with transcript level above 5 RPKM in Col (average number from 3 replicates).

Distance between gene and a locus (repeat/DMR/H3K9me2 mark/H3K27me3 mark) was defined as 0 if the two annotated regions have overlap, otherwise the physical distance between the two loci was used regardless of the gene orientation. Repeats annotations used in this study were the same as in [10]. Locations of Pol IV and Pol V hypomethylated CHH DMRs were obtained from analysis of whole genome bisulfite sequencing data. Locations of H3K9me2 were obtained from [87]. Locations from H3K27me3 were obtained from data in [88].

References

1. Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013). VAL- and AtBMI1-Mediated H2Aub Initiate the Switch from Embryonic to Postgerminative Growth in Arabidopsis. Current biology : CB.

2. Berr, A., McCallum, E.J., Menard, R., Meyer, D., Fuchs, J., Dong, A., and Shen, W.H. (2010). Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development. The Plant cell 22, 3232-3248.

3. Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., and Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. Nature 427, 164-167.

4. Martienssen, R.A., Kloc, A., Slotkin, R.K., and Tanurdzic, M. (2008). Epigenetic inheritance and reprogramming in plants and fission yeast. Cold Spring Harbor symposia on quantitative biology 73, 265-271.

5. Choy, M.K., Movassagh, M., Goh, H.G., Bennett, M.R., Down, T.A., and Foo, R.S. (2010). Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. BMC genomics 11, 519.

6. Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell research 21, 381-395.

7. Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133, 523-536.

8. Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452, 215-219.

9. Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews. Genetics 11, 204-220.

10. Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152, 352-364.

11. Woo, H.R., Pontes, O., Pikaard, C.S., and Richards, E.J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes & development 21, 267-277.

12. Woo, H.R., Dittmer, T.A., and Richards, E.J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS genetics 4, e1000156.

13. Yao, Q., Song, C.X., He, C., Kumaran, D., and Dunn, J.J. (2012). Heterologous expression and purification of Arabidopsis thaliana VIM1 protein: in vitro evidence for its inability to recognize hydroxymethylcytosine, a rare base in Arabidopsis DNA. Protein expression and purification 83, 104-111.

14. Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PloS one 3, e3156.

15. Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A., et al. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell 151, 167-180.

16. Inagaki, S., Miura-Kamio, A., Nakamura, Y., Lu, F., Cui, X., Cao, X., Kimura, H., Saze, H., and Kakutani, T. (2010). Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. The EMBO journal 29, 3496-3506.

17. Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., and Jacobsen, S.E. (2007). The SRA methyl-cytosine-binding domain links DNA and histone methylation. Current biology : CB 17, 379-384.

18. Cao, X., and Jacobsen, S.E. (2002). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proceedings of the National Academy of Sciences of the United States of America 99 Suppl 4, 16491-16498.

19. Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153, 193-205.

20. Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science (New York, N.Y.) 308, 118-120.

21. Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C. (2007). An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. The Plant cell 19, 1507-1521.

22. Law, J.A., Vashisht, A.A., Wohlschlegel, J.A., and Jacobsen, S.E. (2011). SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS genetics 7, e1002195.

23. Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M., Strahl, B.D., Patel, D.J., and Jacobsen, S.E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498, 385-389.

24. Haag, J.R., Ream, T.S., Marasco, M., Nicora, C.D., Norbeck, A.D., Pasa-Tolic, L., and Pikaard, C.S. (2012). In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. Molecular cell 48, 811-818.

25. Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. Current opinion in cell biology 21, 367-376.

26. Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Current biology : CB 15, 1501-1507.

27. Ye, R., Wang, W., Iki, T., Liu, C., Wu, Y., Ishikawa, M., Zhou, X., and Qi, Y. (2012). Cytoplasmic assembly and selective nuclear import of Arabidopsis Argonaute4/siRNA complexes. Molecular cell 46, 859-870.

28. Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.Y., and Chen, X. (2009). Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes & development 23, 2850-2860.

29. Zhong, X., Hale, C.J., Law, J.A., Johnson, L.M., Feng, S., Tu, A., and Jacobsen, S.E. (2012). DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nature structural & molecular biology 19, 870-875.

30. Law, J.A., Ausin, I., Johnson, L.M., Vashisht, A.A., Zhu, J.K., Wohlschlegel, J.A., and Jacobsen, S.E. (2010). A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in Arabidopsis. Current biology : CB 20, 951-956.

31. Rowley, M.J., Avrutsky, M.I., Sifuentes, C.J., Pereira, L., and Wierzbicki, A.T. (2011). Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. PLoS genetics 7, e1002120.

32. He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H., and Zhu, J.K. (2009). An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. Cell 137, 498-508.

33. Zhu, Y., Rowley, M.J., Bohmdorfer, G., and Wierzbicki, A.T. (2013). A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. Molecular cell 49, 298-309.

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

35. Qian, W., Miki, D., Zhang, H., Liu, Y., Zhang, X., Tang, K., Kan, Y., La, H., Li, X., Li, S., et al. (2012). A histone acetyltransferase regulates active DNA demethylation in Arabidopsis. Science (New York, N.Y.) 336, 1445-1448.

36. Agius, F., Kapoor, A., and Zhu, J.K. (2006). Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. Proceedings of the National Academy of Sciences of the United States of America 103, 11796-11801.

37. Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L., and Zhu, J.K. (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell 111, 803-814.

38. Martinez-Macias, M.I., Qian, W., Miki, D., Pontes, O., Liu, Y., Tang, K., Liu, R., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., et al. (2012). A DNA 3' phosphatase functions in active DNA demethylation in Arabidopsis. Molecular cell 45, 357-370.

39. Martinez-Macias, M.I., Cordoba-Canero, D., Ariza, R.R., and Roldan-Arjona, T. (2013). The DNA repair protein XRCC1 functions in the plant DNA demethylation pathway by stimulating cytosine methylation (5-meC) excision, gap tailoring, and DNA ligation. The Journal of biological chemistry 288, 5496-5505.

40. Zheng, X., Pontes, O., Zhu, J., Miki, D., Zhang, F., Li, W.X., Iida, K., Kapoor, A., Pikaard, C.S., and Zhu, J.K. (2008). ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. Nature 455, 1259-1262.

41. Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136, 461-472.

42. Mosher, R.A., and Melnyk, C.W. (2010). siRNAs and DNA methylation: seedy epigenetics. Trends in plant science 15, 204-210.

43. Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., and Baulcombe, D.C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature 460, 283-286.

44. Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126, 1189-1201.

45. Zemach, A., and Grafi, G. (2003). Characterization of Arabidopsis thaliana methyl-CpG-binding domain (MBD) proteins. The Plant journal : for cell and molecular biology 34, 565-572.

46. Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D., Patel, D.J., and Jacobsen, S.E. (2011). A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes & development 25, 137-152.

47. Zhou, J., Wang, X., He, K., Charron, J.B., Elling, A.A., and Deng, X.W. (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.

48. Saze, H., Shiraishi, A., Miura, A., and Kakutani, T. (2008). Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. Science (New York, N.Y.) 319, 462-465.

49. Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H., and Kakutani, T. (2009). An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. The EMBO journal 28, 1078-1086.

50. Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS biology 5, e129.

51. Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS genetics 3, e86.

52. Hennig, L., and Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? Trends in genetics : TIG 25, 414-423.

53. Zheng, B., and Chen, X. (2011). Dynamics of histone H3 lysine 27 trimethylation in plant development. Current opinion in plant biology 14, 123-129.

54. Cao, R., Tsukada, Y., and Zhang, Y. (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Molecular cell 20, 845-854.

55. de Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Developmental cell 7, 663-676.

56. Endoh, M., Endo, T.A., Endoh, T., Isono, K., Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W.A., et al. (2012). Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS genetics 8, e1002774.

57. Lu, F., Cui, X., Zhang, S., Jenuwein, T., and Cao, X. (2011). Arabidopsis REF6 is a histone H3 lysine 27 demethylase. Nature genetics 43, 715-719.

58. Piruat, J.I., and Aguilera, A. (1998). A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. The EMBO journal 17, 4859-4872.

59. Rehwinkel, J., Herold, A., Gari, K., Kocher, T., Rode, M., Ciccarelli, F.L., Wilm, M., and Izaurralde, E. (2004). Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. Nature structural & molecular biology 11, 558-566.

60. Chi, B., Wang, Q., Wu, G., Tan, M., Wang, L., Shi, M., Chang, X., and Cheng, H. (2013). Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. Nucleic acids research 41, 1294-1306.

61. Pena, A., Gewartowski, K., Mroczek, S., Cuellar, J., Szykowska, A., Prokop, A., Czarnocki-Cieciura, M., Piwowarski, J., Tous, C., Aguilera, A., et al. (2012). Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. The EMBO journal 31, 1605-1616.

62. Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes & development 19, 1512-1517.

63. Yelina, N.E., Smith, L.M., Jones, A.M., Patel, K., Kelly, K.A., and Baulcombe, D.C. (2010). Putative Arabidopsis THO/TREX mRNA export complex is involved in transgene and endogenous siRNA biosynthesis. Proceedings of the National Academy of Sciences of the United States of America 107, 13948-13953.

64. Gomez-Gonzalez, B., Garcia-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M., and Aguilera, A. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. The EMBO journal 30, 3106-3119.

65. Voynov, V., Verstrepen, K.J., Jansen, A., Runner, V.M., Buratowski, S., and Fink, G.R. (2006). Genes with internal repeats require the THO complex for transcription. Proceedings of the National Academy of Sciences of the United States of America 103, 14423-14428.

66. Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R., et al. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature 417, 304-308.

67. Zenklusen, D., Vinciguerra, P., Wyss, J.C., and Stutz, F. (2002). Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. Molecular and cellular biology 22, 8241-8253.

68. Chavez, S., Beilharz, T., Rondon, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. The EMBO journal 19, 5824-5834.

69. Saguez, C., Schmid, M., Olesen, J.R., Ghazy, M.A., Qu, X., Poulsen, M.B., Nasser, T., Moore, C., and Jensen, T.H. (2008). Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. Molecular cell 31, 91-103.

70. Larochelle, M., Lemay, J.F., and Bachand, F. (2012). The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. Nucleic acids research 40, 10240-10253.

71. Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Molecular cell 12, 711-721.

72. Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature 419, 182-185.

73. Herold, A., Teixeira, L., and Izaurralde, E. (2003). Genome-wide analysis of nuclear mRNA export pathways in Drosophila. The EMBO journal 22, 2472-2483.

74. Dominguez-Sanchez, M.S., Barroso, S., Gomez-Gonzalez, B., Luna, R., and Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. PLoS genetics 7, e1002386.

75. Mancini, A., Koch, A., Whetton, A.D., and Tamura, T. (2004). The M-CSF receptor substrate and interacting protein FMIP is governed in its subcellular localization by protein kinase C-mediated phosphorylation, and thereby potentiates M-CSF-mediated differentiation. Oncogene 23, 6581-6589.

76. Furumizu, C., Tsukaya, H., and Komeda, Y. (2010). Characterization of EMU, the Arabidopsis homolog of the yeast THO complex member HPR1. RNA (New York, N.Y.) 16, 1809-1817.

77. Jauvion, V., Elmayan, T., and Vaucheret, H. (2010). The conserved RNA trafficking proteins HPR1 and TEX1 are involved in the production of endogenous and exogenous small interfering RNA in Arabidopsis. The Plant cell 22, 2697-2709.

78. Won, S.Y., Li, S., Zheng, B., Zhao, Y., Li, D., Zhao, X., Yi, H., Gao, L., Dinh, T.T., and Chen, X. (2012). Development of a luciferase-based reporter of transcriptional gene silencing that enables bidirectional mutant screening in Arabidopsis thaliana. Silence 3, 6.

79. Lee, J.H., Terzaghi, W., Gusmaroli, G., Charron, J.B., Yoon, H.J., Chen, H., He, Y.J., Xiong, Y., and Deng, X.W. (2008). Characterization of Arabidopsis and rice DWD proteins and their roles as substrate receptors for CUL4-RING E3 ubiquitin ligases. The Plant cell 20, 152-167.

80. Dumbliauskas, E., Lechner, E., Jaciubek, M., Berr, A., Pazhouhandeh, M., Alioua, M., Cognat, V., Brukhin, V., Koncz, C., Grossniklaus, U., et al. (2011). The Arabidopsis CUL4-DDB1 complex interacts with MSI1 and is required to maintain MEDEA parental imprinting. The EMBO journal 30, 731-743.

81. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science (New York, N.Y.) 340, 619-621.

82. Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. The Plant journal : for cell and molecular biology 45, 616-629.

83. Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T.T., Meyers, B.C., Mo, B., and Chen, X. (2012). The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. Current biology : CB 22, 689-694.

84. Lertpanyasampatha, M., Gao, L., Kongsawadworakul, P., Viboonjun, U., Chrestin, H., Liu, R., Chen, X., and Narangajavana, J. (2012). Genome-wide analysis of microRNAs in rubber tree (Hevea brasiliensis L.) using high-throughput sequencing. Planta 236, 437-445.

85. Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics (Oxford, England) 25, 1966-1967.

86. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England) 26, 139-140.

87. Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., and Jacobsen, S.E. (2012). Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. PLoS genetics 8, e1003062.

88. Lafos, M., Kroll, P., Hohenstatt, M.L., Thorpe, F.L., Clarenz, O., and Schubert, D. (2011). Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. PLoS genetics 7, e1002040.

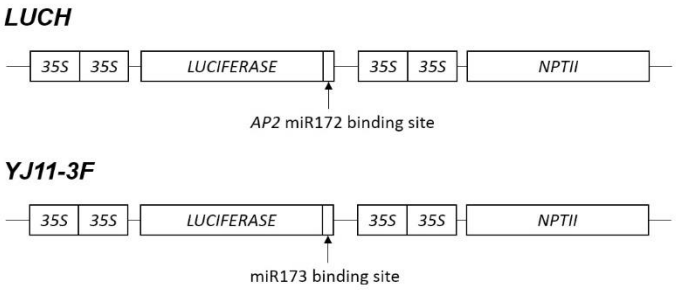
Figures and Tables

Figure 2.1 The *LUCH* and *YJ11-3F* luciferase reporter lines

(A) Transgenes in the reporter lines *LUCH* and *YJ11-3F*. In *LUCH*, the *LUCIFERASE* gene was fused to the *AP2* miR172 binding site at the 3' end and driven by a dual *35S* promoter. The *NPTII* marker gene was also driven by a dual *35S* promoter. In *YJ11-3F*, the transgene was similar to that in *LUCH* except that a miR173 binding site was used instead of a miR172 binding site.

(B) Genetic analysis of the reporter line *YJ11-3F*. Introduction of *ago1-45* and *drd1-12* into *YJ11-3F* did not obviously affect the luminescence of the reporter line.

A


LUCH
35S
35S
LUCIFERASE
35S
35S
NPTII
AP2 miR172 binding site
YJ11-3F
35S
35S
LUCIFERASE
35S
35S
NPTII
miR173 binding site

B

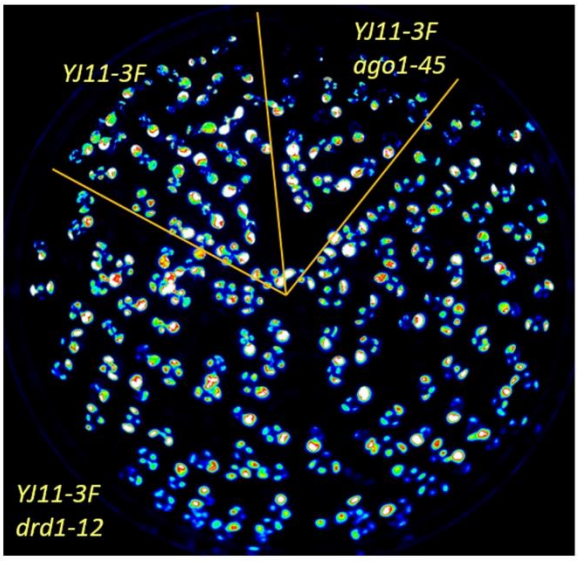

YJ11-3F
YJ11-3F
ago1-45
YJ11-3F
drd1-12

Figure 2.2 DNA methylation levels at the dual *35S* promoter and *LUC* coding region in *LUCH* and *YJ11-3F*

(A) DNA methylation level in *LUCH*, *LUCH ago4-6* and *LUCH drd1-12* at the dual *35S* promoter driving the *LUC* reporter gene and at the *LUC* coding region in all cytosine sequence contexts. DNA methylation was present at both the promoter and coding regions. Mutations in *AGO4* and *DRD1* reduced CHH methylation at the promoter region.

(B) DNA methylation level in *YJ11-3F* in all cytosine sequence contexts. DNA methylation was present at the promoter region but not at the coding region.

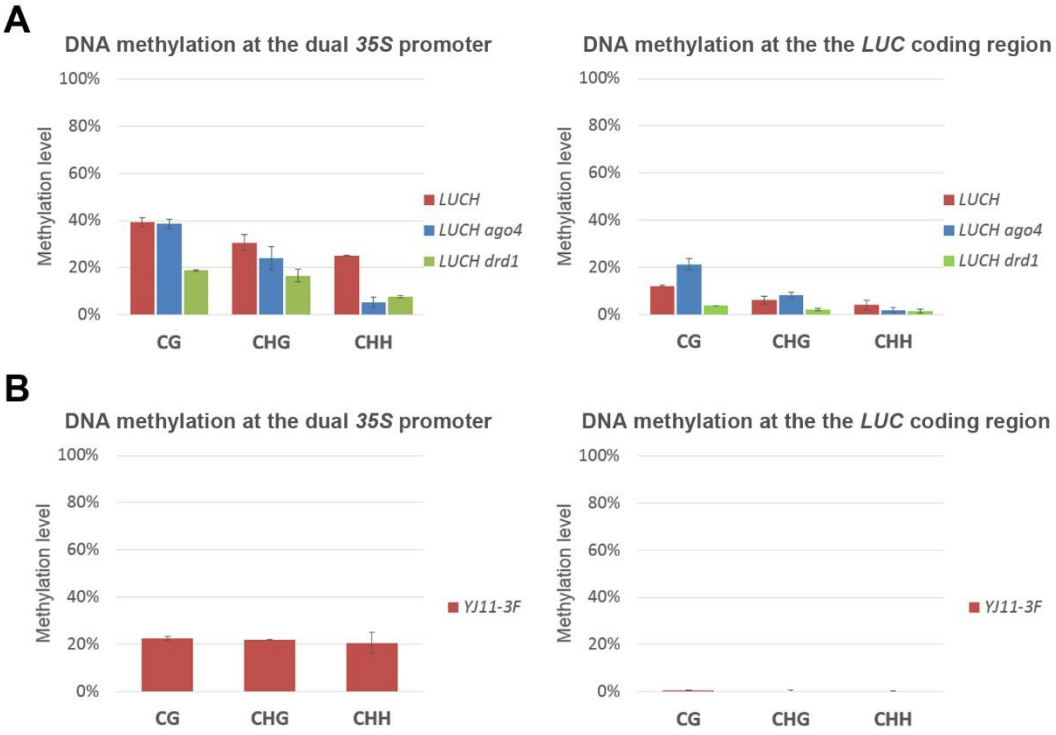

A
DNA methylation at the dual 35S promoter
Methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
LUCH
LUCH ago4
LUCH drd1
DNA methylation at the the LUC coding region
Methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
LUCH
LUCH ago4
LUCH drd1
B
DNA methylation at the dual 35S promoter
Methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
YJ11-3F
DNA methylation at the the LUC coding region
Methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
YJ11-3F

Figure 2.3 Luminescence and *LUC* transcript levels in the three mutant lines

(A) Luminescence of mutant lines *P204R*, *TL525L* and *28-6L* and the respective reporter lines. The three mutant lines had decreased luminescence compared to the respective reporter lines. The reduced luminescence phenotypes of *YJ11-3F hpr1-5* and *YJ11-3F tho5a-1* were rescued by p*HPR1:HA-HPR1* and p*THO5A:THO5A-GFP*, respectively.

(B) Semi-quantitative RT-PCR to examine *LUC* transcript levels in the mutant lines and the respective reporter lines. *YJ11-3F* had higher *LUC* transcript levels than *LUCH*. *YJ11-3F hpr1-5* and *YJ11-3F tho5a-1* had decreased *LUC* transcript levels compared to *YJ11-3F*. Similarly, *LUCH tex1-5* had decreased *LUC* transcript levels compared to *LUCH*. *UBQ5* was used as a control.

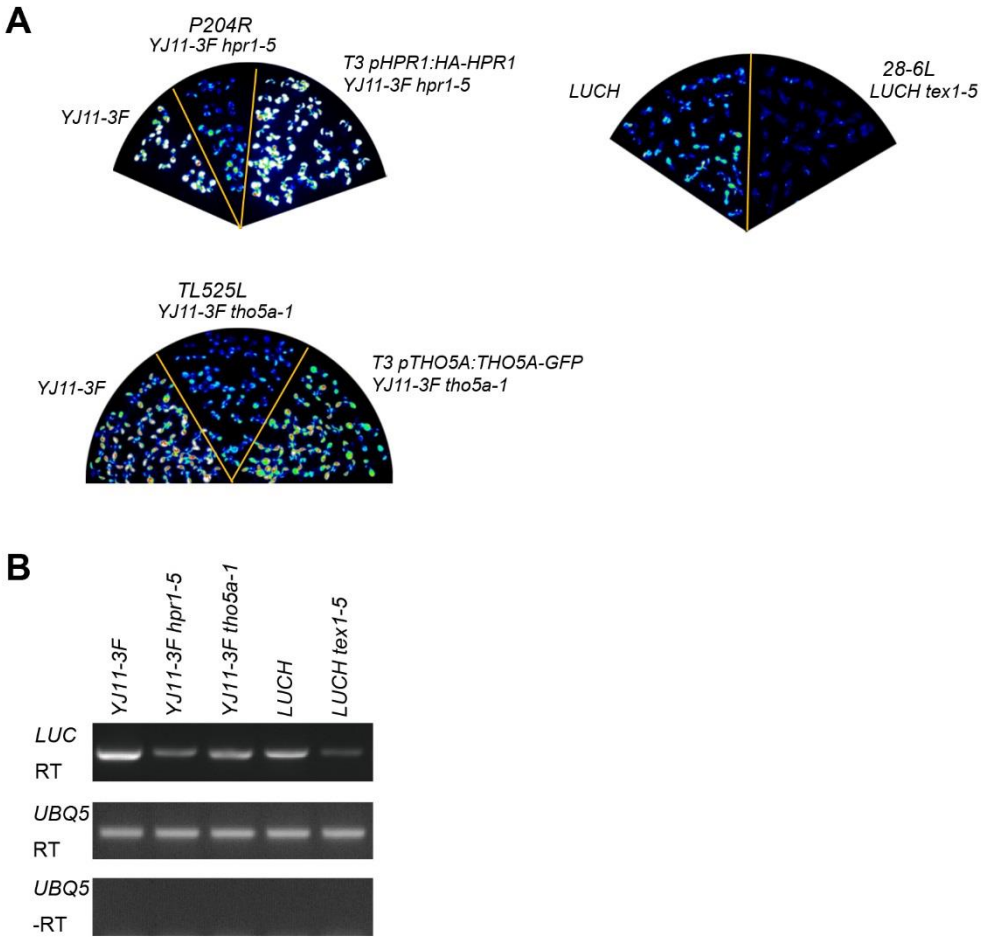

A
P204R
YJ11-3F hpr1-5
T3 pHPR1:HA-HPR1
YJ11-3F hpr1-5
YJ11-3F
LUCH
28-6L
LUCH tex1-5
TL525L
YJ11-3F tho5a-1
T3 pTHO5A:THO5A-GFP
YJ11-3F tho5a-1
YJ11-3F
B
YJ11-3F
YJ11-3F hpr1-5
YJ11-3F tho5a-1
LUCH
LUCH tex1-5
LUC
RT
UBQ5
RT
UBQ5
-RT

Figure 2.4 Gene structures of *TEX1, HPR1* and *THO5A* and the mutation sites

(A) Gene structures of *TEX1*, *HPR1* and *THO5A*. Boxes represent exons. Red and white regions indicate coding regions and UTRs, respectively. The sites of the mutations are indicated by arrows. Regions covered by RT-PCR in (B) are indicated by the lines below each gene diagram.

(B) Semi-quantitative RT-PCR to examine the transcript levels of the THO complex genes in the mutant lines. *tex1-5* led to a splicing defect and reduced the transcript accumulation of the *TEX1* gene. *hpr1-5* introduced a premature stop codon and reduced the transcript accumulation of the *HPR1* gene. The T-DNA insertion in *tho5a-1* disrupted transcription, as no transcript was detected downstream of the insertion site.

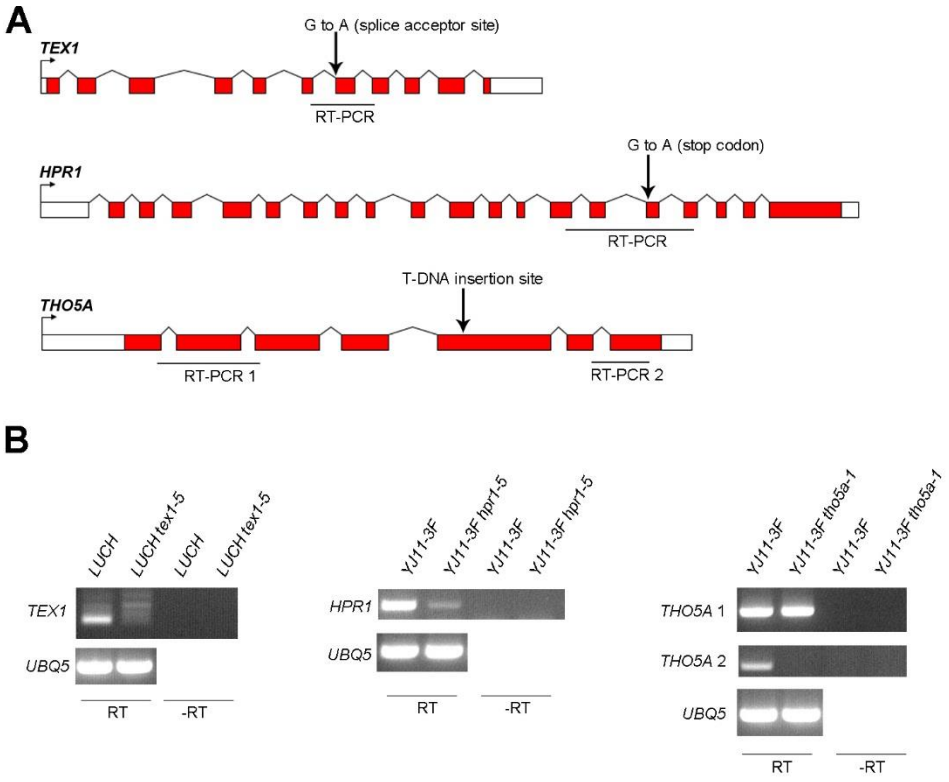

A
TEX1
G to A (splice acceptor site)
RT-PCR
HPR1
G to A (stop codon)
RT-PCR
THO5A
T-DNA insertion site
RT-PCR 1
RT-PCR 2
B
LUCH
LUCH tex1-5
LUCH
LUCH tex1-5
TEX1
UBQ5
RT
-RT
YJ11-3F
YJ11-3F hpr1-5
YJ11-3F
YJ11-3F hpr1-5
HPR1
UBQ5
RT
-RT
YJ11-3F
YJ11-3F tho5a-1
YJ11-3F
YJ11-3F tho5a-1
THO5A 1
THO5A 2
UBQ5
RT
-RT

Figure 2.5 DNA methylation at the *LUC* reporter locus and endogenous loci was not affected in the THO complex mutants

(A) Genetic interactions between the THO complex mutants and RdDM pathway mutants. Note that the luminescence levels of the *LUCH tex1-5 ago4-6* and *LUCH tex1-5 drd1-12* double mutants were higher than *LUCH tex1-5* and lower than *LUCH ago4-6* and *LUCH drd1-12*, respectively. These findings indicate that *TEX1* does not act in the RdDM pathway. Similar genetic interaction patterns were observed for the *hpr1-5* mutant.

(B) DNA methylation levels in *LUCH, LUCH tex1-5, YJ11-3F* and *YJ11-3F hpr1-5* at the dual *35S* promoter driving *LUC* and at the *LUC* coding region in all cytosine sequence contexts. The mutations in *TEX1* and *HPR1* did not affect the DNA methylation levels at the promoter or coding region of the *LUC* reporter transgene.

(C) Number of differentially methylated regions (DMRs) in the THO complex mutants and RdDM pathway mutants compared to the control lines (refer to Table 2.2). In contrast to the number of DMRs in the RdDM mutants, only a few CHG and CHH DMRs were identified in the THO complex mutants.

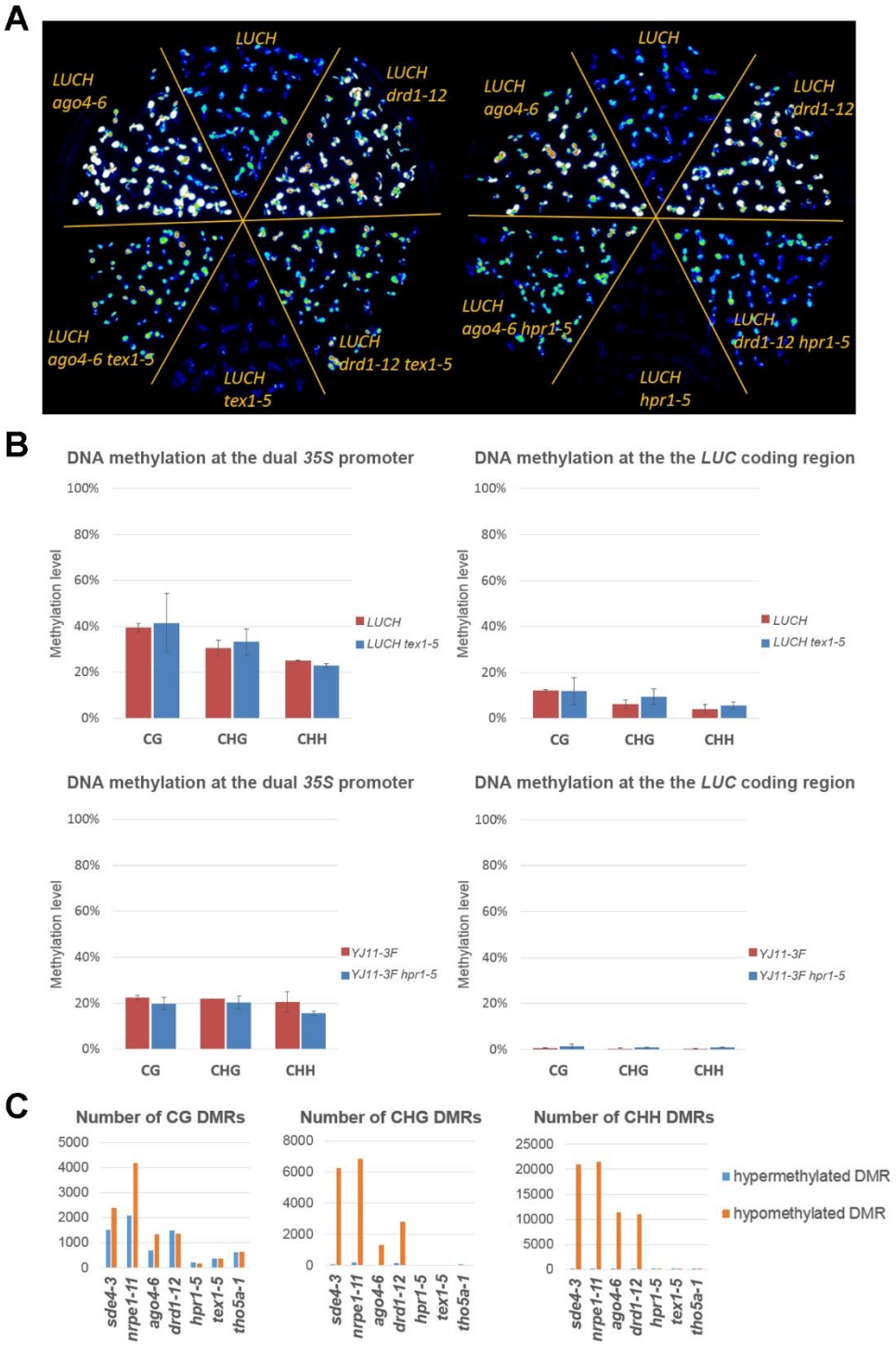

A
LUCH
LUCH ago4-6
LUCH drd1-12
LUCH ago4-6 tex1-5
LUCH drd1-12 tex1-5
LUCH tex1-5
LUCH
LUCH ago4-6
LUCH drd1-12
LUCH ago4-6 hpr1-5
LUCH drd1-12 hpr1-5
LUCH hpr1-5
B
DNA methylation at the dual 35S promoter
DNA methylation at the the LUC coding region
Methylation level
100%
80%
60%
40%
20%
0%
CG
CHG
CHH
LUCH
LUCH tex1-5
YJ11-3F
YJ11-3F hpr1-5
C
Number of CG DMRs
Number of CHG DMRs
Number of CHH DMRs
5000
4000
3000
2000
1000
0
8000
6000
25000
20000
15000
10000
sde4-3
nrpe1-11
ago4-6
drd1-12
hpr1-5
tex1-5
tho5a-1
hypermethylated DMR
hypomethylated DMR

Figure 2.6 The accumulation of 24-nt heterochromatic siRNAs was not affected in the THO complex mutants

(A) Length distribution of small RNAs in wild-type Col, *sde4-3*, *nrpe1-11*, *tex1-5* and *hpr1-5*. Unlike the *sde4-3* mutant, in which the accumulation of 24-nt siRNAs was reduced, the *tex1-5* and *hpr1-5* mutants had similar small RNA length distributions as wild-type Col.

(B) Small RNA accumulation at two Pol IV-dependent siRNA loci in *sde4-3*, Col, *hpr1-5*, *tex1-5* and *tho5a-1* detected by northern blot analysis. The accumulation of small RNAs was not obviously affected at these two loci in the THO complex mutants. U6 was used as a loading control.

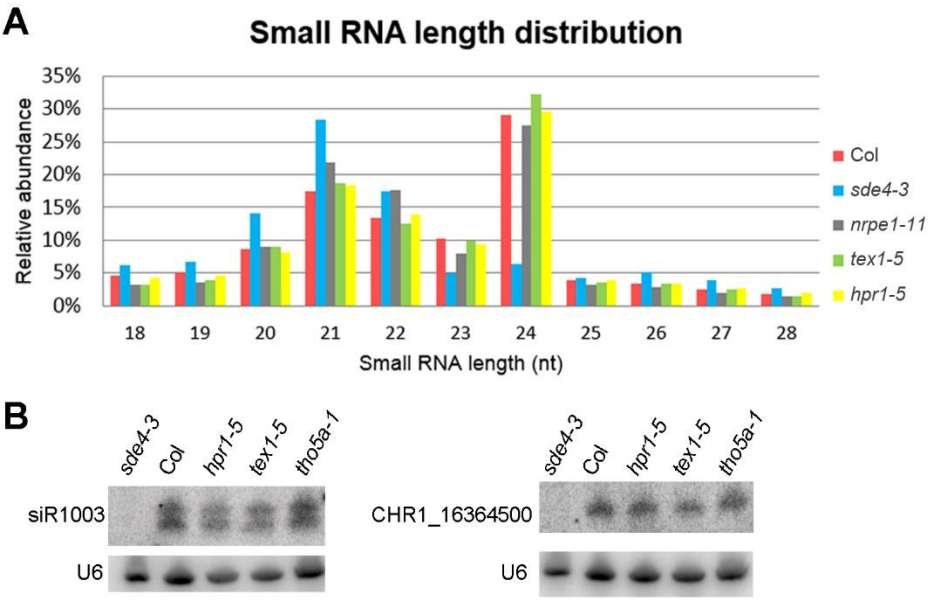

A
Small RNA length distribution
Relative abundance
35%
30%
25%
20%
15%
10%
5%
0%
18
19
20
21
22
23
24
25
26
27
28
Small RNA length (nt)
Col
sde4-3
nrpe1-11
tex1-5
hpr1-5
B
sde4-3
Col
hpr1-5
tex1-5
tho5a-1
siR1003
U6
sde4-3
Col
hpr1-5
tex1-5
tho5a-1
CHR1_16364500
U6

Figure 2.7 Morphological phenotypes of *hpr1-5* and *tex1-5*

The *hpr1-5* and *tex1-5* mutants exhibited developmental defects, namely, smaller plant size compared to wild-type Col plants.

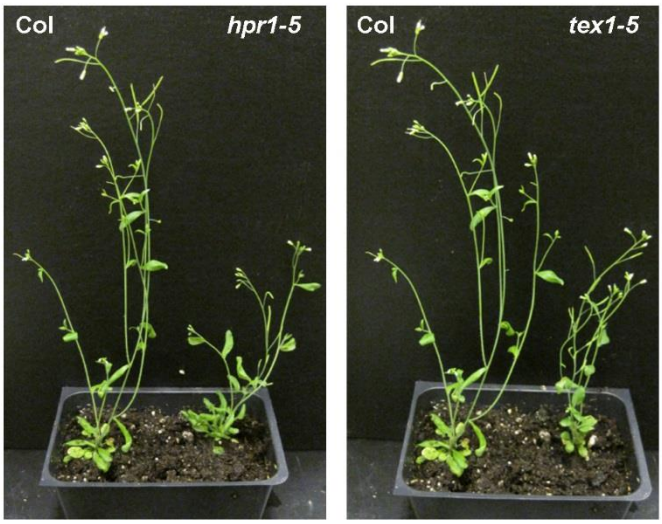
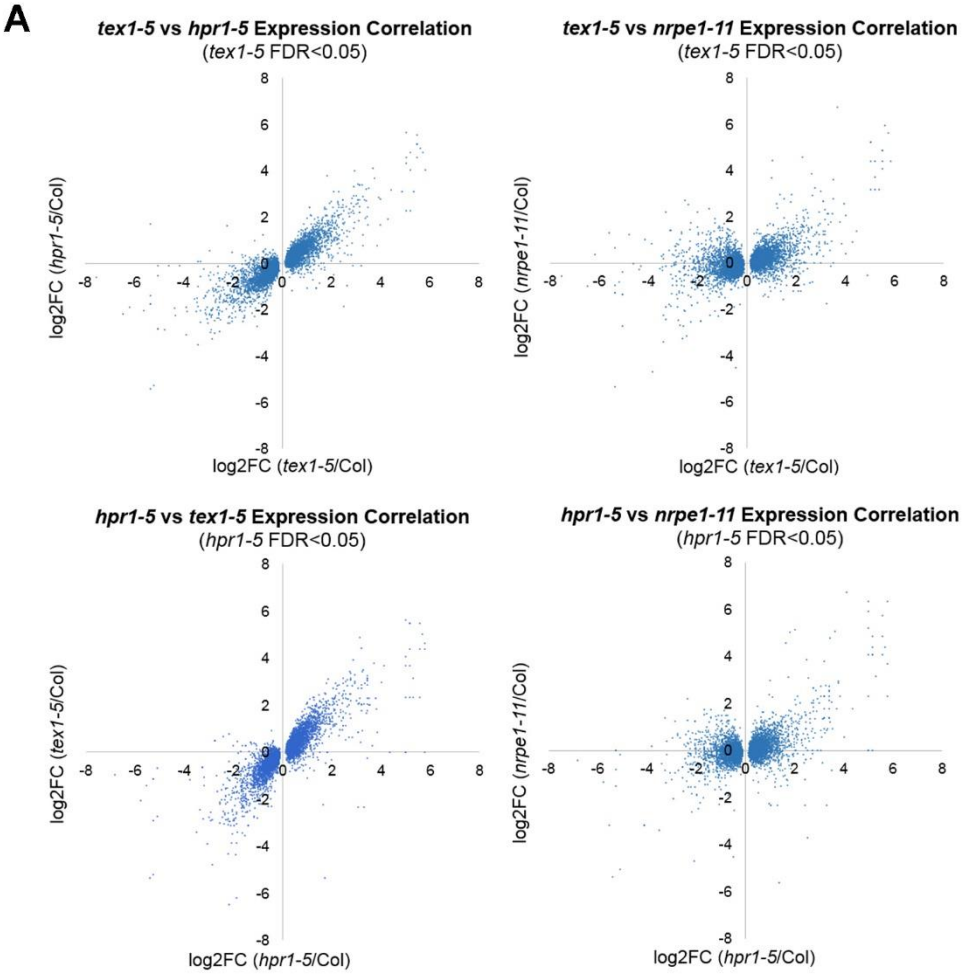

Col
hpr1-5
Col
tex1-5

Figure 2.8 Correlation of *hpr1-5* and *tex1-5* mRNA sequencing data

(A) Dot plots of transcript fold changes (represented by log2(fold change)) in *hpr1-5* and *tex1-5*. Only genes with an FDR value less than 0.05 in *tex1-5* vs. Col (top two panels) and in *hpr1-5* vs. Col (bottom two panels) were included in the analysis. The transcript level changes in *hpr1-5* and *tex1-5* were well correlated, as shown in the two panels on the left. Most of the up-regulated/down-regulated genes in one mutant were also up-regulated/down-regulated, respectively, in the other mutant. In comparison, no obvious correlation between *hpr1-5* and *nrpe1-11* or between *tex1-5* and *nrpe1-11* was observed (two panels on the right).

(B) Validation of mRNA sequencing data by real-time RT-PCR. The transcript levels of *AT1G07000, AT5G24030* and *AT1G72416* were decreased in *YJ11-3F hpr1-5* and *LUCH tex1-5* compared to the respective reporter lines. The transcript levels of the three genes analyzed were normalized to *UBQ5*.


A
tex1-5 vs hpr1-5 Expression Correlation
(tex1-5 FDR<0.05)
log2FC (hpr1-5/Col)
log2FC (tex1-5/Col)
tex1-5 vs nrpe1-11 Expression Correlation
(tex1-5 FDR<0.05)
log2FC (nrpe1-11/Col)
log2FC (tex1-5/Col)
hpr1-5 vs tex1-5 Expression Correlation
(hpr1-5 FDR<0.05)
log2FC (tex1-5/Col)
log2FC (hpr1-5/Col)
hpr1-5 vs nrpe1-11 Expression Correlation
(hpr1-5 FDR<0.05)
log2FC (nrpe1-11/Col)
log2FC (hpr1-5/Col)

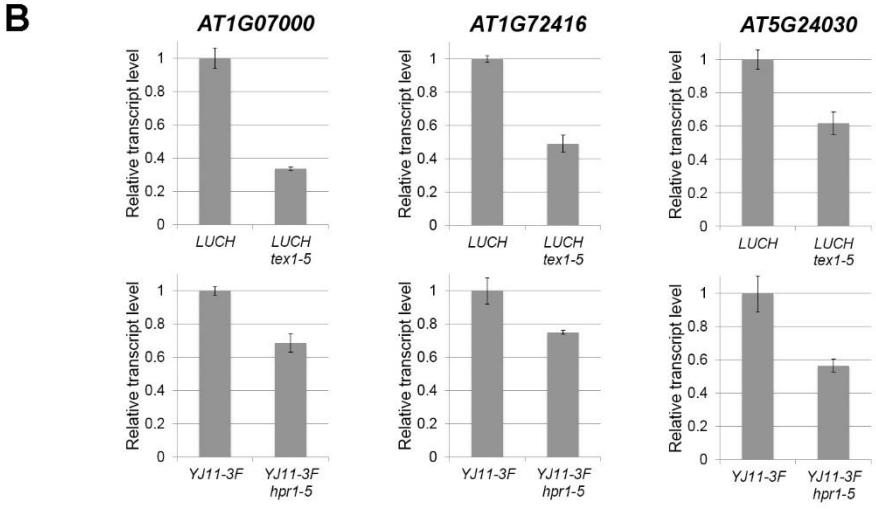

B
AT1G07000
AT1G72416
AT5G24030
Relative transcript level
LUCH
LUCH tex1-5
YJ11-3F
YJ11-3F hpr1-5

Figure 2.9 GC content, gene length and exon number of genes down-regulated in *hpr1-5* and *tex1-5*

(A) GC content of down-regulated genes in *hpr1-5* (yellow), down-regulated genes in *tex1-5* (green) and a set of control genes that were not affected by the THO complex mutations (red). The height of each bar indicates the proportion of genes that fall into a certain GC content range. No obvious differences were observed between the control genes and the down-regulated genes in *hpr1-5* and *tex1-5* in terms of GC content.

(B) The length distribution of down-regulated genes in *hpr1-5* (yellow), down-regulated genes in *tex1-5* (green) and a set of control genes that were not affected by the THO complex mutations (red). The height of each bar indicates the proportion of genes that fall into a certain gene length category. The down-regulated genes in *hpr1-5* and *tex1-5* did not tend to be long genes.

(C) The number of exons in down-regulated genes in *hpr1-5* (yellow), down-regulated genes in *tex1-5* (green) and a set of control genes that were not affected by the THO complex mutations (red). The Y-axis indicates the proportion of genes with a certain number of exons. The down-regulated genes in *hpr1-5* and *tex1-5* did not tend to have more exons.

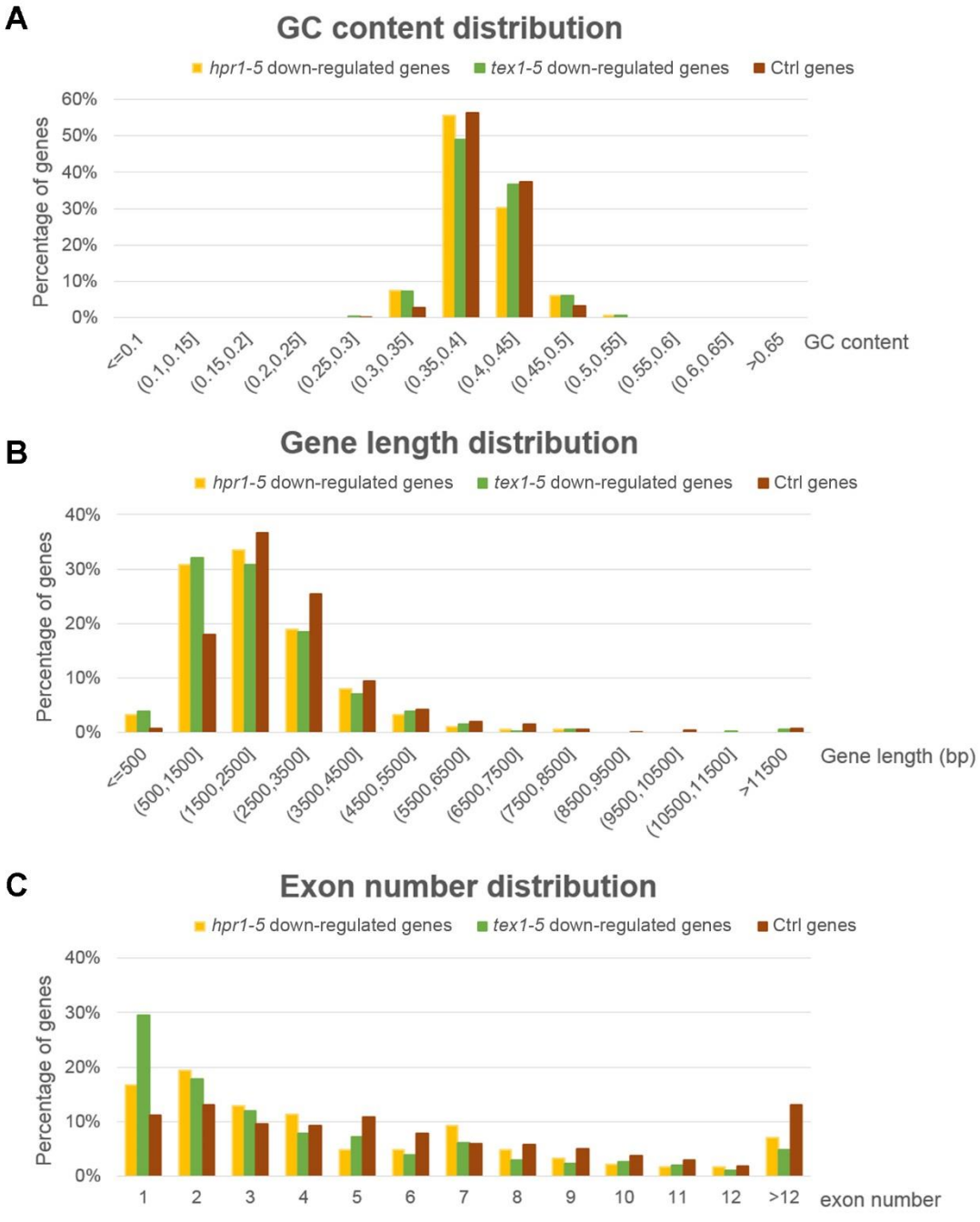
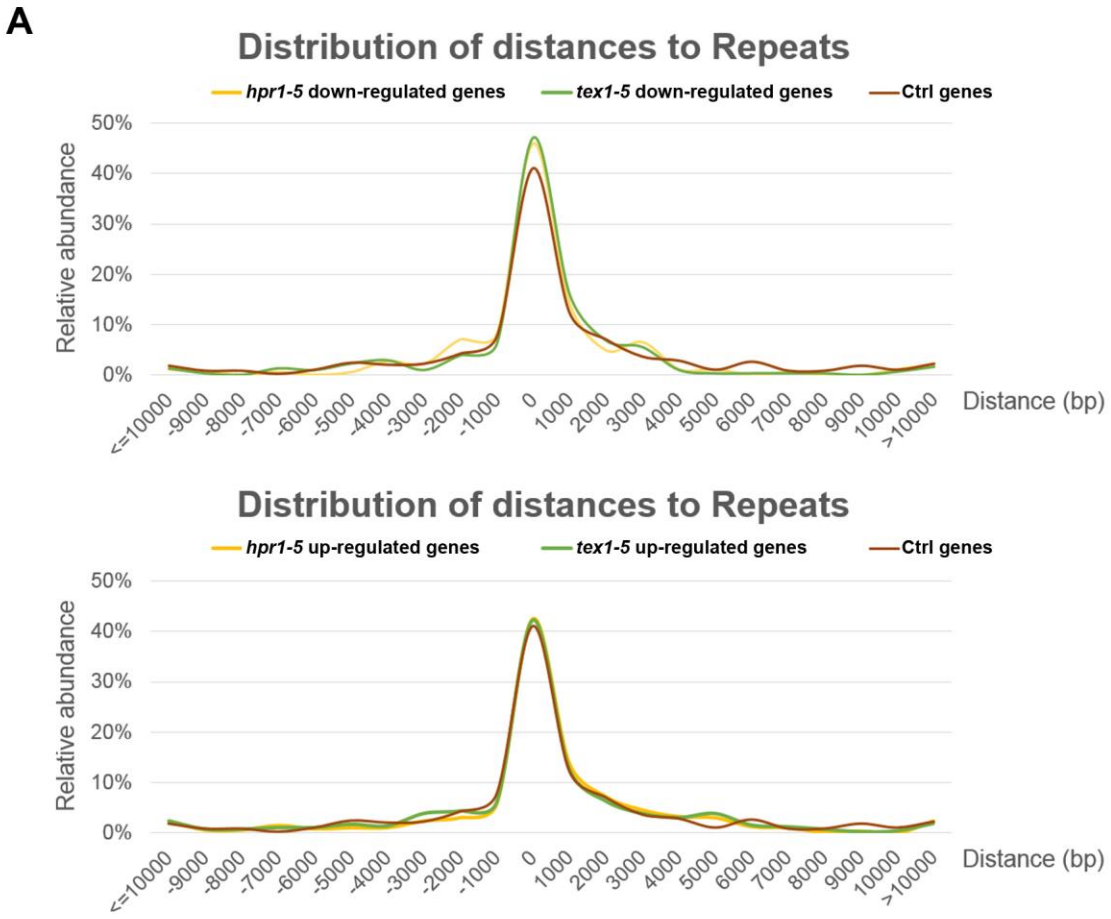

A
GC content distribution
hpr1-5 down-regulated genes
tex1-5 down-regulated genes
Ctrl genes
Percentage of genes
60%
50%
40%
30%
20%
10%
0%
<=0.1
(0.1,0.15]
(0.15,0.2]
(0.2,0.25]
(0.25,0.3]
(0.3,0.35]
(0.35,0.4]
(0.4,0.45]
(0.45,0.5]
(0.5,0.55]
(0.55,0.6]
(0.6,0.65]
>0.65
GC content
B
Gene length distribution
hpr1-5 down-regulated genes
tex1-5 down-regulated genes
Ctrl genes
Percentage of genes
40%
30%
20%
10%
0%
<=500
(500,1500]
(1500,2500]
(2500,3500]
(3500,4500]
(4500,5500]
(5500,6500]
(6500,7500]
(7500,8500]
(8500,9500]
(9500,10500]
(10500,11500]
>11500
Gene length (bp)
C
Exon number distribution
hpr1-5 down-regulated genes
tex1-5 down-regulated genes
Ctrl genes
Percentage of genes
40%
30%
20%
10%
0%
1
2
3
4
5
6
7
8
9
10
11
12
>12
exon number

Figure 2.10 Distances between down-regulated genes in *tex1-5* and *hpr1-5* and repeats

(A) Distribution of distances of control genes and down-regulated genes in *hpr1-5* and *tex1-5* to the closest repeat loci (top panel). For the down-regulated genes in *hpr1-5* and *tex1-5*, the proportion of genes in close proximity to repeat regions was greater compared to the set of control genes. A similar pattern was not observed for the up-regulated genes in *hpr1-5* and *tex1-5* compared to the control genes (bottom panel).

(B) Percentage of genes located within 1 kb of dispersed repeats, inverted repeats and tandem repeats among the down-regulated genes in *hpr1-5* and *tex1-5* compared to the control genes. The proportion of genes that overlaps with repeats was indicated in blue. Among the down-regulated genes in *hpr1-5* and *tex1-5*, the proportion of genes located within 1 kb of dispersed repeats and inverted repeats was greater compared to the set of control genes. Differences were not observed for tandem repeats.


A
Distribution of distances to Repeats
hpr1-5 down-regulated genes
tex1-5 down-regulated genes
Ctrl genes
Relative abundance
50%
40%
30%
20%
10%
0%
<=10000
-9000
-8000
-7000
-6000
-5000
-4000
-3000
-2000
-1000
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
>10000
Distance (bp)
Distribution of distances to Repeats
hpr1-5 up-regulated genes
tex1-5 up-regulated genes
Ctrl genes
Relative abundance
Distance (bp)

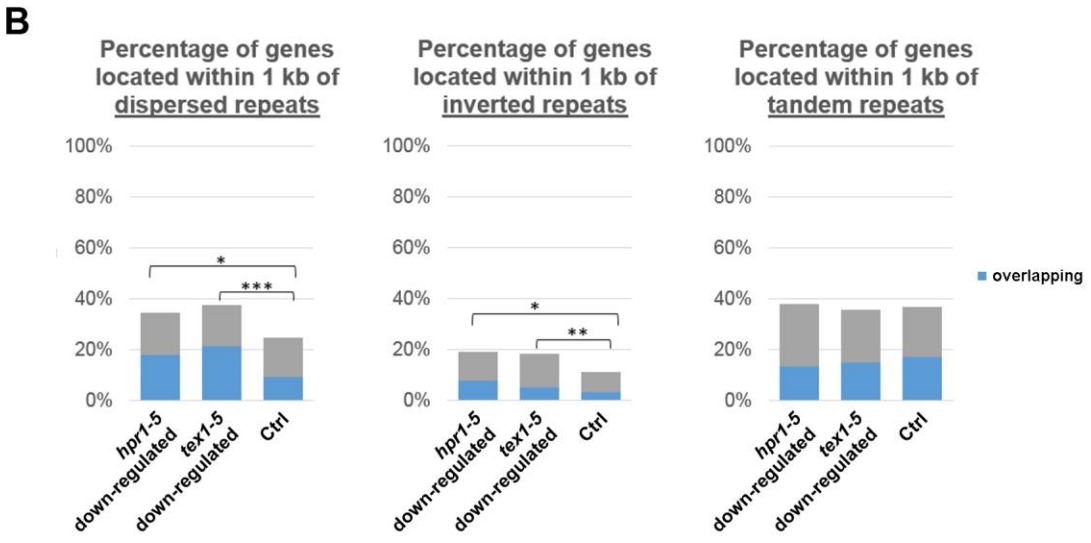

B
Percentage of genes located within 1 kb of dispersed repeats
Percentage of genes located within 1 kb of inverted repeats
Percentage of genes located within 1 kb of tandem repeats
100%
80%
60%
40%
20%
0%
*
***
**
hpr1-5 down-regulated
tex1-5 down-regulated
Ctrl
overlapping

Figure 2.11 Distances between down-regulated genes in *tex1-5* and *hpr1-5* and RdDM target loci and H3K9me2 loci

(A) Distribution of distances of control genes and down-regulated genes in *hpr1-5* and *tex1-5* to the closest RdDM target loci (identified as CHH hypomethylated DMRs in *sde4-3* and *nrpe1-11*). The proportion of genes in close proximity to RdDM target loci was slightly higher for the down-regulated genes of *hpr1-5* and *tex1-5* compared to the set of control genes.

(B) Distribution of distances of control genes and down-regulated genes in *hpr1-5* and *tex1-5* to the closest H3K9me2 loci. The proportion of genes in close proximity to H3K9me2 was slightly higher for the down-regulated genes of *hpr1-5* and *tex1-5* compared to the set of control genes.

(C) Venn diagram showing the overlapping between dispersed or inverted repeat-associated genes, RdDM target loci-associated genes and H3K9me2-associated genes among all the down-regulated genes in *hpr1-5* and *tex1-5*. Among the 134 genes that are down-regulated both in *hpr1-5* and *tex1-5* (fold change>2, FDR<0.05), 59 genes are located within 1 kb of dispersed repeats or inverted repeats (green circle), 25 genes are located within 1 kb of RdDM target loci (identified as CHH hypomethylated DMRs in *sde4-3* and *nrpe1-11*, yellow circle) and 27 genes are located within 1 kb of H3K9me2 loci (red circle). Most RdDM loci-associated genes and H3K9me2-associated genes are also in close proximity to dispersed or inverted repeats. However, around half (27) of the dispersed or inverted repeats associated genes are not in close proximity to RdDM loci or H3K9me2.

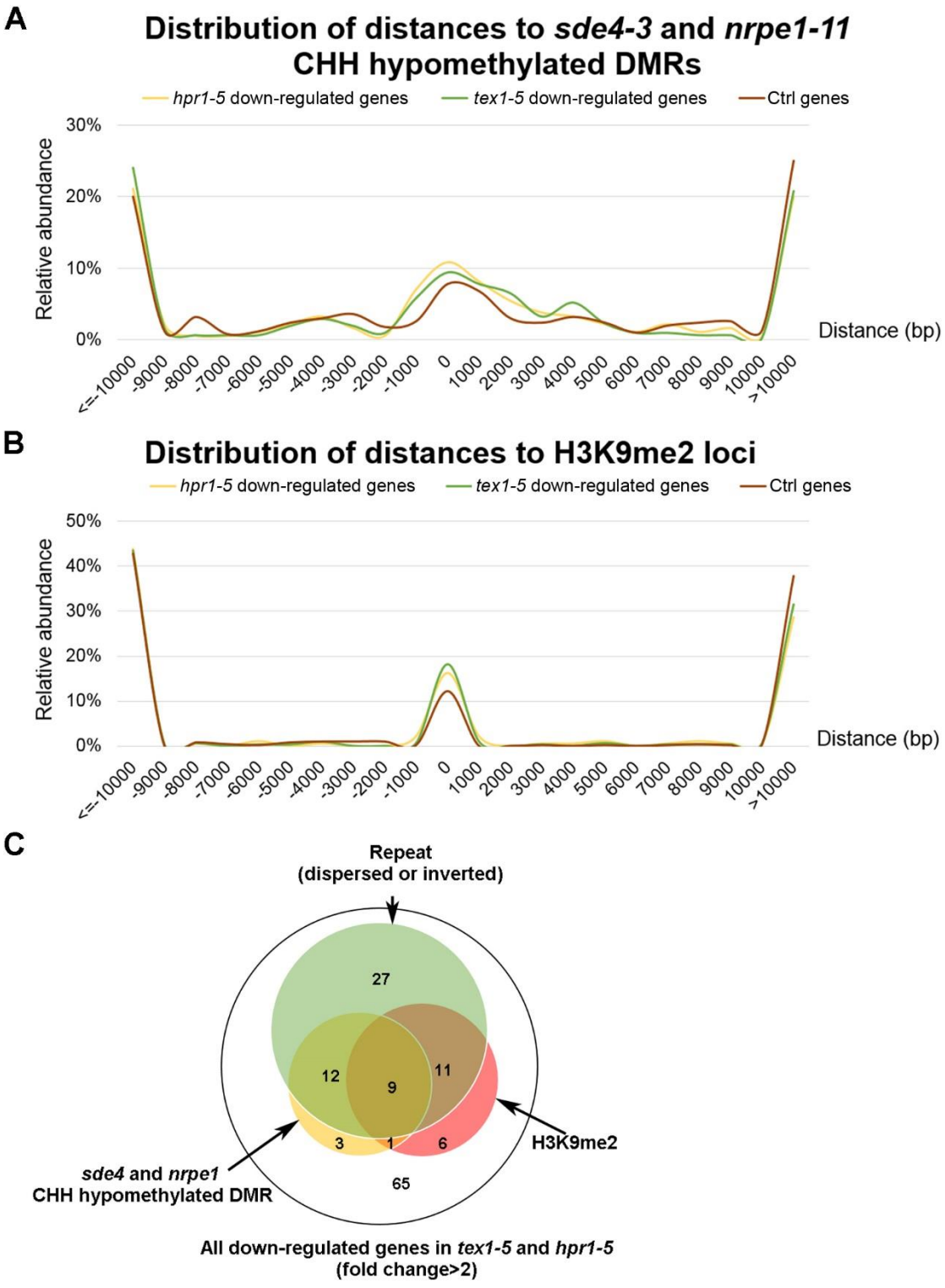

A
Distribution of distances to sde4-3 and nrpe1-11
CHH hypomethylated DMRs
hpr1-5 down-regulated genes
tex1-5 down-regulated genes
Ctrl genes
Relative abundance
30%
20%
10%
0%
Distance (bp)
<=-10000
-9000
-8000
-7000
-6000
-5000
-4000
-3000
-2000
-1000
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
>10000
B
Distribution of distances to H3K9me2 loci
hpr1-5 down-regulated genes
tex1-5 down-regulated genes
Ctrl genes
Relative abundance
50%
40%
30%
20%
10%
0%
Distance (bp)
C
Repeat
(dispersed or inverted)
27
12
9
11
3
1
6
65
H3K9me2
sde4 and nrpe1
CHH hypomethylated DMR
All down-regulated genes in tex1-5 and hpr1-5
(fold change>2)

Figure 2.12 The proportion of genes with H3K27me3 was increased among up-regulated and down-regulated genes in *hpr1-5* and *tex1-5*

(A) Distribution of distances of control genes and down-regulated genes in *hpr1-5* and *tex1-5* to the closest H3K27me3 loci. The proportion of genes in close proximity to H3K27me3 loci was higher among the down-regulated genes in *hpr1-5* and *tex1-5* compared to the set of control genes.

(B) Distribution of distances of control genes and up-regulated genes in *hpr1-5* and *tex1-5* to the closest H3K27me3 loci. The proportion of genes in close proximity to H3K27me3 loci was higher among the up-regulated genes in *hpr1-5* and *tex1-5* compared to the set of control genes.

(C) Percentage of genes overlapping with H3K27me3 loci among the control genes and among the up-regulated or down-regulated genes in *hpr1-5* and *tex1-5*. For both up-regulated and down-regulated genes in *hpr1-5* and *tex1-5*, a greater proportion of genes overlap with H3K27me3 marks compared to the set of control genes.

(D) Venn diagram showing the overlapping between dispersed or inverted repeat-associated genes and H3K27me3-associated genes among all the down-regulated genes in *hpr1-5* and *tex1-5*. Among the 134 genes that are down-regulated both in *hpr1-5* and *tex1-5* (fold change>2, FDR<0.05), 59 genes are located within 1 kb of dispersed repeats or inverted repeats (green circle) and 58 genes overlap with H3K27me3 marks (blue circle). 31 genes associate with both H3K27me3 and dispersed or inverted repeats.

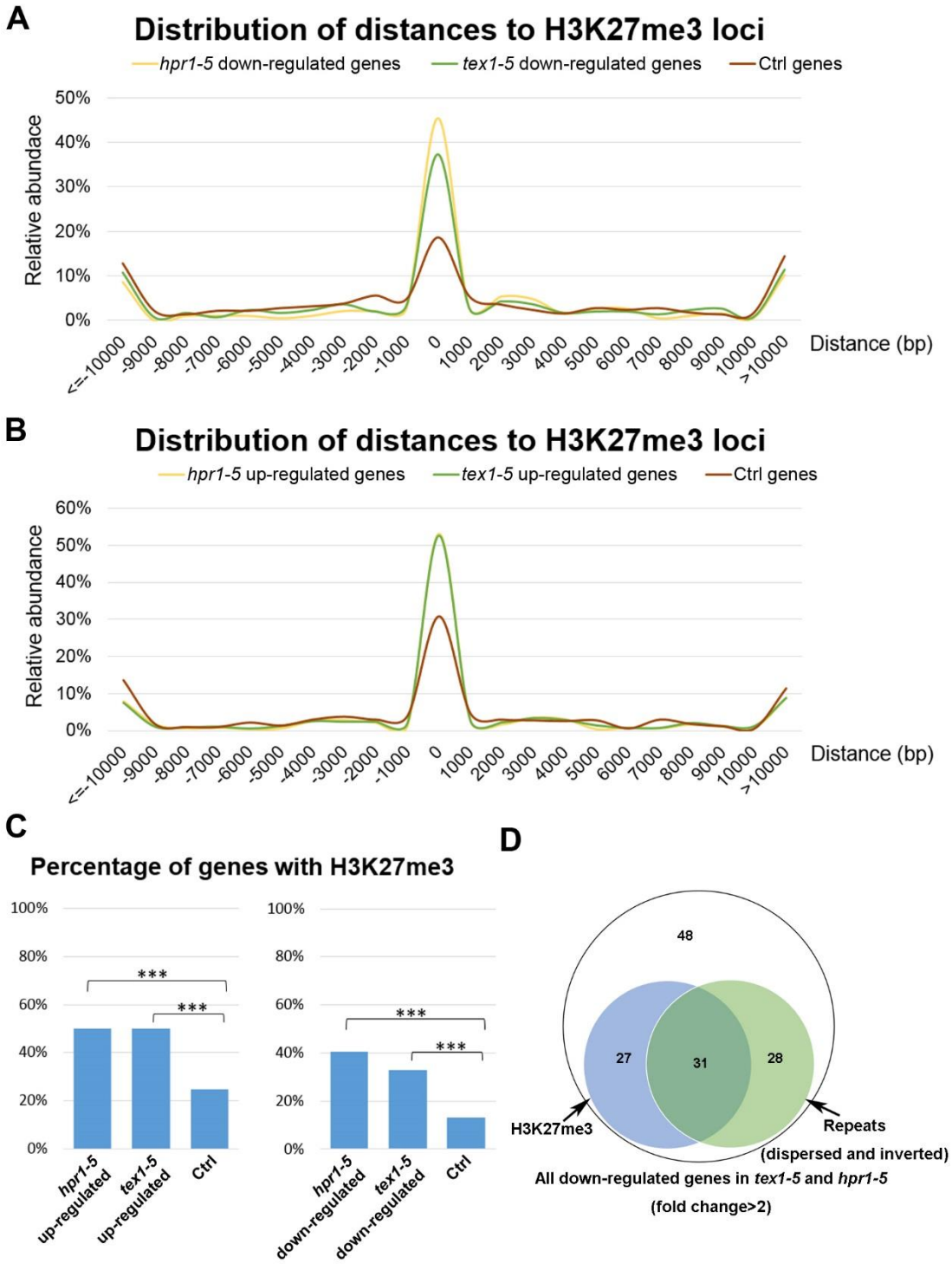

A
Distribution of distances to H3K27me3 loci
hpr1-5 down-regulated genes
tex1-5 down-regulated genes
Ctrl genes
Relative abundace
50%
40%
30%
20%
10%
0%
<=-10000
-9000
-8000
-7000
-6000
-5000
-4000
-3000
-2000
-1000
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
>10000
Distance (bp)
B
Distribution of distances to H3K27me3 loci
hpr1-5 up-regulated genes
tex1-5 up-regulated genes
Ctrl genes
Relative abundance
60%
50%
40%
30%
20%
10%
0%
Distance (bp)
C
Percentage of genes with H3K27me3
100%
80%
60%
40%
20%
0%
***
***
hpr1-5 up-regulated
tex1-5 up-regulated
Ctrl
***
***
hpr1-5 down-regulated
tex1-5 down-regulated
Ctrl
D
48
27
31
28
H3K27me3
Repeats
(dispersed and inverted)
All down-regulated genes in tex1-5 and hpr1-5
(fold change>2)

Figure 2.13 Expression patterns of all dispersed repeat-associated loci, inverted repeat-associated loci and H3K27me3-associated loci in *tex1-5* and *hpr1-5*

(A) Dot plots of transcript fold changes (represented by log2 (fold change)) in *tex1-5* (left panel) and *hpr1-5* (right panel) vs. average transcript level (RPKM: reads per 1 kb per million reads) in Col for all genes located within 1 kb of dispersed repeats. Expressed genes overlapping or in close proximity with dispersed repeats were not found to exhibit a particular pattern of expression change in either *tex1-5* or *hpr1-5* (i.e., these genes were up regulated, down-regulated or did not have altered transcript levels in *tex1-5* or *hpr1-5*).

(B) Transcript fold change in *tex1-5* (left panel) and *hpr1-5* (right panel) vs. average transcript level (RPKM: reads per 1 kb per million reads) in Col for all genes located within 1 kb of inverted repeats. Expressed genes overlapping with or in close proximity to inverted repeats were not found to exhibit a particular pattern of expression change in either *tex1-5* or *hpr1-5*.

(C) Transcript fold change in *tex1-5* (left panel) and *hpr1-5* (right panel) vs. average transcript level (RPKM: reads per 1 kb per million reads) in Col for all genes with H3K27me3. Genes with H3K27me3 were not found to exhibit a particular pattern of expression change in either *tex1-5* or *hpr1-5*.

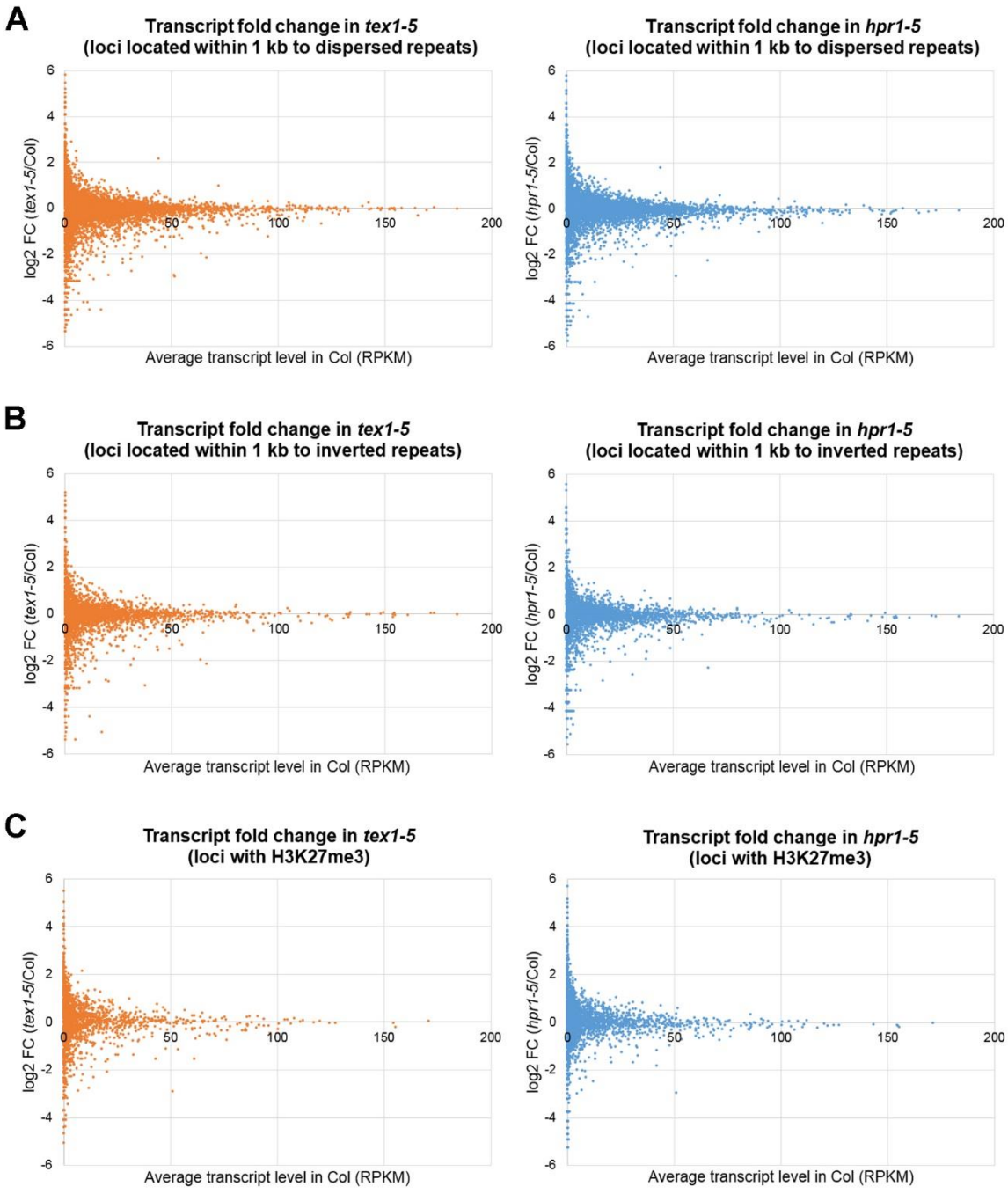

A
Transcript fold change in tex1-5
(loci located within 1 kb to dispersed repeats)
log2 FC (tex1-5/Col)
Average transcript level in Col (RPKM)
Transcript fold change in hpr1-5
(loci located within 1 kb to dispersed repeats)
log2 FC (hpr1-5/Col)
Average transcript level in Col (RPKM)
B
Transcript fold change in tex1-5
(loci located within 1 kb to inverted repeats)
log2 FC (tex1-5/Col)
Average transcript level in Col (RPKM)
Transcript fold change in hpr1-5
(loci located within 1 kb to inverted repeats)
log2 FC (hpr1-5/Col)
Average transcript level in Col (RPKM)
C
Transcript fold change in tex1-5
(loci with H3K27me3)
log2 FC (tex1-5/Col)
Average transcript level in Col (RPKM)
Transcript fold change in hpr1-5
(loci with H3K27me3)
log2 FC (hpr1-5/Col)
Average transcript level in Col (RPKM)

Table 2.1 Genes encoding THO complex subunits in *Arabidopsis*, their conserved domains and protein sequence identifty with human orthologs

	***Arabidopsis* gene ID**	**Conserved domain**	**Amino acid sequence identity with human ortholog in homologous region**
HPR1	*AT5G09860*	efThoc1 (pfam11957)	33%
THO2	*AT1G24706*	Tho2/Thoc2 (pfam11262, pfam11732)	36%
THO5	*AT5G42920 (THO5A)*	FimP (pfam09766)	28%
	AT5G45233 (THO5B)	FimP (pfam09766)	27%
THO6	*AT2G19430*	WD40 (cd00200)	31%
THO7	*AT5G16790*	Thoc7 (pfam05615)	36%
	AT3G02950	Thoc7 (pfam05615)	34%
TEX1	*AT5G56130*	WD40 (cd00200)	49%

Table 2.2 Number of differentially methylated regions (DMRs) in each genotype

	CG		**CHG**		**CHH**	
	hyper-methylated	hypo-methylated	hyper-methylated	hypo-methylated	hyper-methylated	hypo-methylated
sde4-3	1513	2389	81	6229	14	21031
nrpe1-11	2092	4181	179	6842	17	21515
ago4-6	695	1343	3	1310	10	11442
drd1-12	1496	1359	141	2794	5	11065
hpr1-5	210	175	14	6	41	55
tex1-5	359	362	4	3	36	55
tho5a-1	616	631	54	26	89	41

Table 2.3 Number of differentially expressed loci in *hpr1-5*, *tex1-5* and *nrpe1-11*

	Increased (>2-fold)			**Reduced (>2-fold) (RPKM_Col>5)**		
	hpr1-5	***tex1-5***	***nrpe1-11***	***hpr1-5***	***tex1-5***	***nrpe1-11***
Total	516	664	566	213	350	39
Gene	453	605	374	185	308	32
TE	63	59	192	28	42	7

Table 2.4 Oligonucleotides used in this study

Name of oligo	5' to 3' sequence	Purpose
N_UBQ5	GGTGCTAAGAAGAGGAAGAAT	RT PCR
C_UBQ5	CTCCTTCTTTCTGGTAAACGT	RT PCR
LUCmF5	CTCCCCTCTCTAAGGAAGTCG	RT PCR
LUCmR5	CCAGAATGTAGCCATCCATC	RT PCR
tex1_RT F	TTTCTTCTTGACTACTGGACTT	RT PCR
TEX1 RT R2	ATCCGCACTCCCGACAGC	RT PCR
HPR1_RT F	AGTTGAGCATATACTGGAGCG	RT PCR
HPR1_RT R	ATCTTCTGCTAATGGTTTCC	RT PCR
At5g42920 RT F	CCTCAACTTCGTCAATCTTCG	RT PCR
At5g42920 RT R	TCTGTTGCTCCAATCTCGCTC	RT PCR
AT5G42920 RT_F2	CGAGCAATGGAAGCAGA	RT PCR
AT5G42920 RT_R2	CACATCCTCTGCCCTTCCAT	RT PCR
At5g56130seq F1	CGGAATGACTAAACGCTGC	PCR and sequencing
At5g56130seq F2	GCCAGATTCCATAAGAGAGG	Genotyping, PCR and sequencing
At5g56130seq R1	GTGATACTGTAAGATTATTCCAAC	Genotyping, PCR and sequencing
At5g09860-F1	GTTCTTCCTCCACTCTTCTTC	PCR and sequencing
At5g09860-S1	CCTCATCTTCTAGGTTAGC	PCR and sequencing
At5g09860-R1	AACATAATGAGGCACTGGAG	PCR and sequencing
At5g09860-F2	ACCACTAAGCGAAGAAGAGG	PCR and sequencing
At5g09860-S2	AGCTGACCTTGTAATCTAAC	PCR and sequencing
At5g09860-R2	CTAGCAGATTTGCACTCATAC	PCR and sequencing
at5g42920seq F1	ACATAAAAATCATAAATAGTTTGAC	PCR and sequencing
at5g42920seq R1	GTCTGACCATCACCAAGGAT	PCR and sequencing
at5g42920seq F2	TTTGTAACTTGTTTCCAGATG	Genotyping, PCR and sequencing
at5g42920seq R2	CTTTTGCTTGTACCACCAGAG	Genotyping, PCR and sequencing
At5g42920seq F3	CGGTCAATTTAATCCCTC	PCR and sequencing
ago1-45 dCAP SphI F	TGAGCCATGGTCTCGGATGTTTCA	Genotyping
ago1-45 dCAP sphI R	GAGACTATGCCGAGTTCAGTCTCACGCATG	Genotyping
115-1H gtF	GTTGGTTCCACTTCAGAGGTACA	Genotyping
115-1H gtR	GATCGAAATGCCACAGGATC	Genotyping
86-1H gtF1-TaqI	TGGATTTGGTTAAGCGTTCTATTCG	Genotyping
86-1H gtR1	ATACATAGGCCTCTGCCTCTCA	Genotyping
rdr6-11_gtF4	CTTGCATGGCTTTGTGCTTA	Genotyping
rdr6-11_gtR4	CATACCGCGGAGATATGAGCAATC	Genotyping

TailcR	GATGATCTGAAGCGGATCTTG	Genotyping
9-7-2gtF	GGATCTTACTCTTGGATCATATCG	Genotyping
YJ11-3F_For	CTGCCCATTAAACATATAGTTCTG	Genotyping
YJ11-3F_Rev	GTGAAATTACCGGAGTTGCT	Genotyping
R1	CCTCTAGAGTCGACCTGCAGGCATG	Genotyping
HPR1-F	GTGATCTCTGGTCGCTAG	Genotyping
HPR1-R	ATCTGCCCATCTCCAAAG	Genotyping
pEG-LmR1	GATCAATAAAGCCACTTACTTTGC	Genotyping
chr1_16364500	tcatcctctcttccttgccatgattccacaatc	Small RNA northern blot
siR1003	ATGCCAAGTTTGGCCTCACCGT	Small RNA northern blot
U6	AGGGGCCCATGCTAATCCTTCCTC	Small RNA northern blot
HPR1_PRO-SacII_F	TCCccgcggTCGGAAATCATCCAAATCCAT	Complementation
HPR1_PRO-SpeI_R	GGactagtTATTTCTCAATGAACCAGCTTCCA	Complementation
HA_HPR1_CDS-SpeI_F	GactagtATGTATCCATATGACGTTCCAGATTACGCTATGGTAACTCTTTTGGGATTAT	Complementation
HPR1_CDS-XhoI_R	CCGctcgagATTACGCAAAAGAACTTAATTGTATGATG	Complementation
thoc5a_SacII_F	tccccgcggGACACAAGTACAGATCCAAAC	Complementation
thoc5a_SpeI_R_N_stop	ggactagtGCATGGATAACCCGATGCA	complementation

Table 2.5 Genes and that are more than 2 fold up-regulated in *hpr1-5* or *tex1-5* and their association with repeats, H3K9me2, RdDM target loci and H3K27me3.

ID	Up-regulated in *hpr1*	Up-regulated in *tex1*	Distance to dispersed repeat	Distance to inverted repeat	Distance to tandem repeat	Distance to H3K9me2	Distance to sde4 and nrpe1 CHH DMR	H3K27me3 target
AT1G01250		yes	27476	8602	3670	317721	-26892	yes
AT1G01480	yes		19569	35410	0	244651	59900	
AT1G01520	yes	yes	5830	21671	103	230912	46161	
AT1G01580		yes	-11210	769	3294	210010	25259	yes
AT1G01660	yes	yes	12934	-5902	-1077	180443	-758	
AT1G01670		yes	9554	-8688	-3863	177063	-3544	
AT1G01695	yes	yes	1047	963	-1499	168556	1605	
AT1G01810	yes		13437	-36583	4529	127192	-3022	
AT1G02310	yes	yes	-24312	-5099	13908	-22881	-24234	
AT1G02340	yes	yes	31528	-12684	6760	-30466	31556	yes
AT1G02470		yes	0	0	0	-76513	0	
AT1G02570	yes		-348	-28203	1713	-106083	-336	yes
AT1G02770		yes	0	-5227	-2527	-169621	5907	yes
AT1G03710		yes	-7957	-10544	0	-488174	19474	yes
AT1G04180	yes	yes	0	64304	-4722	324012	54212	yes
AT1G04445	yes		-12224	-35786	-907	222138	-12145	yes
AT1G04560	yes		3602	47744	-1643	183941	3641	yes
AT1G04570	yes	yes	886	45028	885	181225	925	
AT1G05260		yes	-6282	-2313	-6237	-83215	-6268	yes
AT1G05490		yes	-480	-3930	0	-172050	-703	
AT1G05660		yes	9343	-17299	939	179980	9342	yes
AT1G06310		yes	7816	7266	7400	-44138	-44492	yes
AT1G06980	yes		13288	12859	2892	-260378	89979	yes
AT1G07485	yes	yes	5166	-1810	2555	227850	33079	
AT1G07490	yes	yes	5266	-1815	2655	227950	33179	
AT1G07710	yes	yes	25112	-44303	4773	141985	-45776	
AT1G08035		yes	-7949	28734	14426	34822	-21103	
AT1G08860		yes	1049	136870	9796	-36364	-41454	yes
AT1G08990	yes	yes	-41385	92049	-1788	-82964	-88054	yes
AT1G09930	yes		17312	12749	-2958	-421701	20557	yes
AT1G10140		yes	6080	-78979	4724	455978	18852	
AT1G10300	yes	yes	13377	-134319	0	399300	-15843	yes
AT1G10585		yes	2608	23603	2743	284382	2556	yes
AT1G10690		yes	706	1110	-2883	228894	968	
AT1G11070		yes	35067	80634	0	85878	68452	yes
AT1G11700	yes		-12256	-12268	10145	-11855	-12238	
AT1G12013	yes		31764	2794	-8527	50852	2984	
AT1G12040	yes		17613	-7876	0	36701	-8425	yes
AT1G12150	yes	yes	-10470	-10701	-9343	-8865	-10431	
AT1G12290		yes	0	0	-4890	-63694	26507	
AT1G12560	yes	yes	-14651	3492	-7289	50252	15401	yes
AT1G12805	yes	yes	-10781	15278	-2284	-30949	-32098	yes
AT1G13150	yes	yes	-13030	21439	-1388	-65771	41404	
AT1G13480	yes	yes	-15051	-40305	27940	-206530	34130	

AT1G13590	yes		34307	33629	3270	-231845	9460	yes
AT1G13607		yes	21829	21151	-1166	-244674	-2279	
AT1G13608	yes	yes	21018	20340	476	-245578	-3183	
AT1G14640	yes		0	18727	-3390	71404	30045	
AT1G14750		yes	16321	-34316	442	14839	15780	
AT1G15380	yes	yes	-5357	64290	-970	-187028	-5346	yes
AT1G15870		yes	3002	-64409	-7149	219575	-19285	
AT1G16400	yes	yes	8843	46243	4464	70531	49827	yes
AT1G16440	yes		0	36044	-3829	60332	39628	
AT1G16830	yes		-2187	57005	4750	-57810	36081	
AT1G17180	yes	yes	0	-6669	0	37764	-8546	
AT1G17260	yes		0	2819	-4153	1952	2802	
AT1G18320	yes	yes	-1203	-34318	27	-84572	-30661	
AT1G18330		yes	-2809	-35924	-181	-86178	-32267	
AT1G18710	yes		1563	89	0	84109	1486	yes
AT1G18850	yes		8051	14500	5730	30799	7976	
AT1G19490	yes	yes	5169	5167	5485	-7043	5101	
AT1G19750	yes	yes	4101	-67145	10119	-82367	-67198	
AT1G19830		yes	-17205	53266	-11902	-107525	53228	
AT1G19968	yes		7072	-8082	-739	92001	-23762	
AT1G20070		yes	-17529	-4606	1275	63863	-51511	yes
AT1G20150	yes	yes	20692	33185	4323	32625	33039	yes
AT1G20400	yes	yes	-509	-13858	0	0	1962	yes
AT1G20515		yes	12005	46295	-2884	-14687	24523	
AT1G20520		yes	12365	46655	-3641	-15444	24883	
AT1G21210		yes	0	-488	14909	-47841	-29407	
AT1G21260		yes	-636	-18600	-228	-65953	-47519	
AT1G22340	yes	yes	-38138	-18323	14614	66300	43610	yes
AT1G22500	yes	yes	0	9733	-5842	7490	9000	yes
AT1G22640	yes	yes	9005	-18504	-554	-19141	8972	
AT1G23350		yes	-2025	-5271	-3476	-1236	-1996	yes
AT1G24260	yes	yes	-4084	0	0	47137	-2338	yes
AT1G24570	yes	yes	13528	3685	3687	12977	13511	
AT1G24577		yes	10085	242	244	9534	10068	yes
AT1G25175		yes	0	14344	0	0	-28987	
AT1G26360	yes	yes	-11919	8356	-1934	51591	14883	yes
AT1G26680		yes	0	12422	3987	-39769	12251	
AT1G26790	yes	yes	-13533	-21940	-14750	87881	29388	yes
AT1G26970		yes	0	-1559	-569	1360	-1583	yes
AT1G27110	yes	yes	559	-33670	0	-27136	-321	
AT1G27140		yes	0	-48708	-61	-42174	3484	
AT1G27740	yes		20388	-18989	1888	17484	20311	yes
AT1G28220	yes		1260	473	5646	-41831	1266	
AT1G29195	yes	yes	-1598	2484	3400	-45424	3304	yes
AT1G29460	yes		6584	6584	0	38155	6357	
AT1G29540	yes		-9618	-11948	730	18193	-11872	
AT1G29560		yes	11416	-16948	0	10604	11406	
AT1G29640	yes	yes	1217	-40280	0	0	518	
AT1G30110		yes	-1594	22049	-1723	6798	-1536	
AT1G30160	yes	yes	-1224	0	-451	-396	-7911	yes
AT1G30390	yes	yes	0	-15189	-23	-13243	84	
AT1G30515		yes	4344	-5946	7136	-99078	4285	
AT1G30750	yes		9290	9516	0	7616	9291	
AT1G30870	yes	yes	-7872	-7872	10029	50142	-7886	
AT1G30960	yes	yes	-4568	3909	4226	3132	-5113	yes
AT1G31040		yes	-2655	-2512	-2525	0	10809	yes
AT1G31760		yes	4611	35194	-1521	-47627	4717	

AT1G32870	yes		0	26335	747	-7924	-9680	
AT1G32950	yes		0	0	7551	35320	-10539	yes
AT1G33102		yes	-3583	6648	-9154	5583	6323	yes
AT1G34060		yes	3943	28100	311	1298	-4721	
AT1G34510	yes		-621	3631	736	7778	-434	yes
AT1G34670		yes	203	21583	-4028	0	-16429	yes
AT1G35910		yes	-521	-5370	1017	0	-3604	yes
AT1G36060		yes	5177	5138	0	0	-5197	yes
AT1G44830		yes	1118	-3242	0	0	1589	yes
AT1G46768		yes	0	3401	779	0	464	
AT1G47400		yes	-3587	610	926	-829	-1218	yes
AT1G47510		yes	230	10394	0	0	2904	yes
AT1G47600	yes	yes	0	-406	-660	0	-533	yes
AT1G47920		yes	-60	2356	2980	0	-796	yes
AT1G47950	yes	yes	0	-1046	0	0	3647	yes
AT1G48330		yes	1505	-18548	340	-12797	-538	
AT1G48930	yes		-5207	-6426	-375	0	-5244	
AT1G49370	yes		-1994	-4011	-4513	26831	-3651	
AT1G49920		yes	0	4851	0	-89808	9367	yes
AT1G50060	yes		0	0	0	82942	0	
AT1G50300		yes	193	-1198	15086	3574	154	
AT1G50400		yes	-5767	-344	0	-11637	-19456	
AT1G51470	yes	yes	0	-6525	-6774	-18988	-514	yes
AT1G51530		yes	4128	-28966	15257	-41429	-22955	
AT1G52050	yes		0	-24141	-4866	0	-1420	yes
AT1G52060	yes	yes	0	25293	3582	-94	-5540	yes
AT1G52070	yes	yes	0	19044	-592	-5943	1476	yes
AT1G52270	yes		735	-19846	2263	-77072	745	
AT1G52315		yes	635	2078	2222	-94582	-4589	
AT1G52650	yes		0	-10674	11739	-6171	-15943	yes
AT1G53035	yes	yes	210	5435	-7853	-8125	449	
AT1G53130	yes		376	24363	1781	-9475	3356	yes
AT1G53540	yes		-8935	23861	-8730	-30374	5791	yes
AT1G53541	yes		-9858	23475	-9653	-31297	5405	yes
AT1G53700	yes		5429	-6857	-2535	-10850	-14887	yes
AT1G53887		yes	-3393	-24355	-907	0	31234	
AT1G54120	yes	yes	458	18633	497	-75886	-121	yes
AT1G54260	yes	yes	-286	-5166	-178	50103	18683	
AT1G54330		yes	-886	-886	-696	28568	-868	yes
AT1G54540	yes		-47	-36906	25236	-27145	-73	yes
AT1G54890	yes		11097	11097	1758	4774	3530	yes
AT1G54970	yes		2349	2349	0	0	2312	yes
AT1G55175	yes	yes	-394	21150	0	-68945	-13666	
AT1G55990		yes	-27468	1080	0	32313	-27288	yes
AT1G56555		yes	-394	0	0	51443	17983	yes
AT1G56600	yes	yes	621	318	887	34064	604	yes
AT1G58170	yes	yes	-1690	-11533	1383	0	3564	
AT1G58400	yes		0	-65955	1037	0	-12666	yes
AT1G58889	yes		0	110210	-8968	0	21438	
AT1G61290	yes		536	-12590	5429	90500	-11810	
AT1G61590	yes	yes	5942	17076	-3065	-19418	5926	yes
AT1G61830		yes	0	4522	-468	6236	3040	
AT1G61860		yes	7591	8074	8249	-3250	-7158	
AT1G61930	yes	yes	-2045	-11627	-4423	-33494	-2402	
AT1G62690		yes	475	4505	2416	-77654	2211	yes
AT1G62720	yes	yes	0	-7152	2542	-94745	-7133	
AT1G62980	yes	yes	-9986	-9133	-6137	94530	-9977	yes

AT1G63250	yes		4905	22413	-7484	0	-7870	
AT1G64210		yes	0	-44632	-367	91937	23537	
AT1G64220	yes	yes	-873	-46687	423	91479	23079	
AT1G64405	yes		881	-43269	-1553	464	864	yes
AT1G64930	yes		-1189	12809	-19440	150939	3539	yes
AT1G65860		yes	0	5049	-5637	-56468	3655	yes
AT1G65870		yes	-3589	3601	7032	-61026	2207	yes
AT1G65890	yes		129	-612	-1208	-69859	-4158	yes
AT1G66650	yes	yes	0	1581	0	-17074	9679	yes
AT1G66830	yes		476	-9457	54	-7771	-12066	
AT1G67105		yes	-214	-23656	2225	101516	-180	
AT1G67260		yes	-6561	-3980	28	0	-3963	yes
AT1G67265		yes	4138	-1214	4138	-5765	3898	yes
AT1G67370		yes	0	3748	-9694	-69324	3663	
AT1G68180		yes	-21581	-18149	-2456	-384928	-17922	yes
AT1G68250	yes		14057	4982	0	-409580	1106	yes
AT1G68510		yes	-3171	-21169	-2502	-537089	-8383	yes
AT1G68795	yes		0	-16975	1757	524028	11848	yes
AT1G69280	yes		-6365	27550	2528	317490	-12861	
AT1G69320		yes	-23067	13226	2011	303166	21486	yes
AT1G69490		yes	18839	-45596	1460	242259	16579	
AT1G69710	yes	yes	2212	-21189	4081	139028	-24673	
AT1G69880	yes	yes	12130	-25206	2981	42687	12907	yes
AT1G70440	yes	yes	1053	32451	5791	-167740	980	yes
AT1G70460		yes	0	23687	0	-174634	2016	
AT1G71200	yes		3815	416	1156	-454761	42892	
AT1G71400	yes		0	-3717	0	-528623	13503	yes
AT1G71528		yes	0	20075	1002	-558350	-13263	
AT1G71530		yes	0	20075	1002	-558355	-13268	
AT1G71770		yes	0	0	0	-609313	0	yes
AT1G71890	yes	yes	-2039	-5430	-6767	-677227	-21840	yes
AT1G71910	yes		-8532	-11923	8774	-683720	-28333	
AT1G72200		yes	8263	15446	-486	-788583	15379	yes
AT1G72490	yes		11540	-101558	-5700	-907688	42958	yes
AT1G72660	yes	yes	28933	63834	-475	-972579	-19692	
AT1G72720	yes		8208	43109	-4992	-995543	-42656	
AT1G72960		yes	-1194	-11800	-15292	1000100	-5437	
AT1G73066		yes	14634	-47349	15159	1000100	-40986	
AT1G73120		yes	399	53226	924	1000100	-56387	yes
AT1G73410	yes	yes	-23144	-4165	-136	926847	83	yes
AT1G73860	yes		-60304	57499	0	753307	-60287	yes
AT1G74088		yes	-275	331	-489	667411	-217	
AT1G74460		yes	4667	12865	-10761	540363	-14962	
AT1G74500	yes	yes	-2644	3845	3906	531343	-25176	
AT1G74550	yes	yes	11761	-13286	-13284	512453	33689	
AT1G74660		yes	3368	7322	-8653	482037	3273	yes
AT1G75580		yes	-2970	19323	-7719	151986	19322	
AT1G76610		yes	-673	33683	-1531	-208480	-689	yes
AT1G77330		yes	13878	60886	-6842	315803	13863	
AT1G77380		yes	1025	48033	2463	302950	1010	
AT1G77570	yes		-2476	-10217	1078	236337	-17451	
AT1G78750	yes		0	16442	-1040	-125989	-623	yes
AT1G78930	yes	yes	411	7184	12976	-190964	-65598	
AT1G79470	yes	yes	4952	46498	1317	-407465	43134	
AT1G79770	yes	yes	-21751	4323	1068	-527228	-29662	yes
AT1G80130	yes		57443	58752	-1145	-653828	57354	yes
AT1G80240	yes	yes	27250	28559	-2387	-684224	27161	yes

AT1G80660		yes	0	60911	0	-828893	-60527	
AT2G01580		yes	260	-10500	7407	-14829	195	
AT2G01810	yes		147	-804	-1135	16857	121	yes
AT2G01818	yes		-1358	-369	-1507	11514	-358	
AT2G02061	yes		427	3761	565	202	341	
AT2G02850		yes	-130	-2802	-1763	0	-105	
AT2G03010	yes		0	5833	-3119	2904	-1348	yes
AT2G03505	yes		251	22661	0	0	228	yes
AT2G04050	yes	yes	1629	8795	948	0	-4287	yes
AT2G04070	yes		-127	-4527	-965	0	-5656	yes
AT2G04135	yes		0	44	0	0	0	
AT2G06002	yes		-464	-3853	652	0	521	
AT2G06200		yes	875	-11674	0	0	1177	
AT2G14820	yes	yes	4679	4675	-5049	0	-5065	yes
AT2G15030	yes		-2657	-12992	-1300	0	-1268	
AT2G15060	yes		0	-2259	1136	0	0	
AT2G15070	yes		0	704	0	0	619	
AT2G15880		yes	2129	-25114	0	0	2081	
AT2G16720	yes	yes	0	-494	-490	0	720	
AT2G16980		yes	0	-4958	3563	-15070	-650	yes
AT2G17010		yes	-15122	-21459	592	-31571	-17151	yes
AT2G18060		yes	-2713	-6590	1757	-56661	-3600	yes
AT2G18100		yes	3538	-6171	926	-76090	3558	
AT2G18193	yes	yes	0	472	0	96140	-206	
AT2G18196		yes	-1744	-599	1651	93091	-609	yes
AT2G18370		yes	-2777	-5994	-2792	33659	-2826	yes
AT2G18540	yes		686	-377	0	0	-383	yes
AT2G18800		yes	4638	4380	4383	58572	11335	
AT2G18969		yes	-2802	-8061	-1515	-18758	2226	
AT2G18970		yes	-2871	-8130	-1584	-18827	2045	
AT2G19240		yes	-2951	-17715	-14	33123	27138	
AT2G20180		yes	-6943	553	0	113539	1000	yes
AT2G20250		yes	13793	23738	-643	83862	-25651	
AT2G20520	yes		4594	25375	1081	-7576	12839	yes
AT2G20750		yes	14265	-824	762	-16208	-1420	
AT2G20815		yes	-1680	1188	-3687	-34344	-1456	
AT2G20880	yes	yes	6415	-21983	-168	11987	10737	
AT2G20940		yes	6229	-38042	5116	0	2899	
AT2G21045	yes	yes	787	-63862	619	-21358	780	yes
AT2G21100		yes	7848	51011	7201	4986	0	yes
AT2G21320	yes	yes	-465	-25779	0	0	-264	
AT2G21430	yes	yes	-530	-10759	-284	18631	-2086	
AT2G21640	yes	yes	6231	6216	22207	-47669	7375	
AT2G22482		yes	9426	16864	4092	-4223	316	
AT2G23630	yes		-3320	-25503	1522	-3977	-11867	yes
AT2G24300	yes	yes	2378	7607	1137	14322	-2817	
AT2G24430		yes	-3679	3434	0	-6682	10475	yes
AT2G24660	yes	yes	0	11889	0	-66491	257	yes
AT2G24800		yes	-4669	9958	-4754	19608	7230	yes
AT2G25680	yes	yes	-740	15663	151	-44440	-662	yes
AT2G25980	yes	yes	0	1007	-60	-188867	-689	yes
AT2G26480	yes	yes	0	358	-329	44247	316	
AT2G26695		yes	-2075	690	-2111	27556	1235	yes
AT2G27370		yes	-599	-28282	1904	-300355	-14051	yes
AT2G28400		yes	-3526	2255	-4232	306344	-3505	
AT2G28870		yes	-829	-12950	-1205	60481	-957	yes
AT2G29090		yes	-3289	-1165	0	-32498	-17352	yes

AT2G29480	yes		12966	-48888	-7198	-43950	-7199	
AT2G29620	yes		-16806	51253	-1123	74780	-39742	yes
AT2G30340		yes	-1086	4080	9523	-106946	-820	yes
AT2G30690	yes		-5789	-12330	-956	146975	-7830	yes
AT2G30770	yes	yes	0	-2416	-7561	113559	1289	yes
AT2G31830	yes	yes	2937	-19672	-8991	-97584	4359	yes
AT2G31910		yes	24110	41329	1740	-135948	5081	yes
AT2G31980	yes	yes	-10730	5446	-2945	-173997	24098	
AT2G32620	yes	yes	0	21881	-304	-405648	21876	yes
AT2G33260		yes	-5454	-734	10500	303555	7248	
AT2G33790	yes	yes	15902	-723	0	110536	-2762	yes
AT2G34390	yes	yes	6805	6800	2615	-87151	37785	
AT2G34830	yes		5114	-10316	-2559	-2510	-10057	yes
AT2G35075	yes		-10127	21728	1458	-95436	21677	yes
AT2G35550	yes	yes	3078	20936	-3044	-238035	-8382	yes
AT2G35742	yes		-22005	-7161	-7522	-336343	70132	
AT2G35850	yes	yes	11281	-36325	34515	-365507	40805	
AT2G36080	yes	yes	-6869	1556	-3186	-457013	-6860	yes
AT2G36100	yes	yes	-18201	-6105	1720	-468345	-18192	yes
AT2G36640	yes		12482	28362	0	647557	12488	yes
AT2G36690	yes	yes	-7813	4987	1184	624182	-7795	yes
AT2G36815	yes		4692	35081	-17954	564893	7524	
AT2G37130		yes	-738	-738	0	406092	-522	
AT2G37750		yes	-9365	11644	9599	174611	-9190	
AT2G38180		yes	7630	23726	0	7263	4494	
AT2G38250		yes	863	3132	3389	0	1200	yes
AT2G38255	yes	yes	0	0	0	0	0	
AT2G38340	yes	yes	33347	251	-6627	-29194	-36797	
AT2G39040	yes	yes	953	1121	1154	-261149	973	yes
AT2G39200		yes	-8094	13316	0	-317946	4187	yes
AT2G39230	yes		-745	-752	-1009	-343445	-748	
AT2G39430	yes	yes	2628	6375	1009	-425145	2569	
AT2G39675		yes	5971	17723	-6411	-499086	17723	
AT2G39681		yes	3831	15583	-8507	-501182	15583	
AT2G40370		yes	-6231	5329	1747	477404	15394	yes
AT2G40390		yes	6998	-429	-4831	468921	6911	yes
AT2G40610	yes	yes	0	-3402	13344	387453	2043	yes
AT2G41210		yes	-21127	-42514	-28513	161203	-25714	
AT2G41230		yes	10299	-64027	11195	142336	10226	yes
AT2G41231		yes	10058	-64393	10954	142095	9985	yes
AT2G41240		yes	6653	-67383	7549	138690	6580	
AT2G41250	yes	yes	3355	66225	4251	135392	3282	
AT2G41260	yes		-338	61145	0	130312	-311	yes
AT2G41970	yes	yes	7874	29147	0	-165818	7944	yes
AT2G42560	yes		31157	35815	-5951	-360077	-38639	yes
AT2G42980	yes	yes	-6001	27896	-7457	-520444	27712	
AT2G43050		yes	0	41	198	-547923	0	yes
AT2G44110	yes	yes	19435	19432	-14569	407471	20071	yes
AT2G44745		yes	16015	25254	-11819	206791	48991	yes
AT2G45920	yes	yes	-7777	266	-600	-231168	-7785	
AT2G46192	yes		-15364	20508	-656	-305391	-608	
AT2G46290		yes	-11646	-2451	1568	-337855	-11972	
AT2G46530	yes		-20255	52087	-1898	-436749	-32566	
AT2G46790	yes	yes	753	-12975	740	-564733	713	
AT2G46860	yes		-10554	-34014	0	-585772	-10489	yes
AT2G46970	yes	yes	14695	640	7694	-627574	24574	yes
AT2G47140		yes	-5592	-26959	-8076	-682824	2212	

AT2G47200		yes	-32581	-53948	2449	-709813	-3330	
AT2G47270		yes	1766	-87779	0	-743644	1749	
AT2G47540		yes	21343	77938	1076	-837944	21302	yes
AT2G48140		yes	-2405	0	-1351	1000100	-2510	yes
AT3G01190	yes	yes	5349	-27610	0	1000100	-23959	yes
AT3G01220	yes	yes	0	-34033	0	1000100	30062	yes
AT3G01260	yes		1187	31679	0	1000100	23595	yes
AT3G01345	yes		0	-7695	-17749	1000100	1651	yes
AT3G01600	yes	yes	4230	-107728	-9353	934832	-72972	
AT3G02000		yes	-38919	69951	3513	832915	89578	yes
AT3G02010	yes		-44609	62531	-263	825495	82158	
AT3G03370	yes		-724	0	-5221	366789	27652	
AT3G03830		yes	0	0	-2302	185449	5012	
AT3G03840		yes	0	0	-3390	184139	3702	
AT3G03850		yes	0	0	2504	182186	1749	
AT3G04485		yes	15836	-25426	-10	-25879	-26915	
AT3G04530		yes	-5222	-48858	830	-49311	-50347	yes
AT3G04700	yes		-2234	58082	3678	-104433	5276	
AT3G05920		yes	11571	46219	0	-160844	-56491	
AT3G05937		yes	4382	39030	-3685	-167956	-63603	
AT3G06390		yes	-10139	-14421	-343	18559	-10133	yes
AT3G06460		yes	-10261	-10828	2279	-9508	-30860	yes
AT3G06630		yes	-21924	-17606	16859	-97237	64600	
AT3G07070	yes	yes	0	19506	13350	-264807	11226	
AT3G07730	yes		-76559	161090	6802	336755	82194	
AT3G08810		yes	0	42493	19050	129668	-27032	yes
AT3G09450	yes	yes	4803	-23128	-10020	-81929	9809	yes
AT3G09680		yes	194	188	458	140937	229	yes
AT3G09925	yes	yes	-11759	-59428	-1975	58524	27216	yes
AT3G10110		yes	-273	2924	0	0	-35411	
AT3G10320		yes	7400	55872	-6808	-72022	7354	
AT3G10710	yes		12613	31803	3120	-228997	69263	
AT3G11050	yes	yes	6457	-33016	1055	-340306	12694	yes
AT3G11260		yes	9632	-4135	0	277470	9567	yes
AT3G11430	yes	yes	-8321	1824	3047	207951	-8366	yes
AT3G11550	yes	yes	46905	18993	18982	166706	46903	
AT3G12460	yes		9598	31434	0	-140217	9677	yes
AT3G12510		yes	-947	19035	-3722	-152617	-896	
AT3G12720		yes	-11380	-55780	0	-228832	42691	yes
AT3G12820		yes	11772	-86628	4144	200290	11686	
AT3G13065		yes	-4161	31195	-12369	84606	25002	
AT3G13090	yes	yes	0	14322	7779	67733	8129	
AT3G13100	yes	yes	0	8320	1777	61731	2127	
AT3G13210		yes	15157	15908	6437	28058	15154	
AT3G13680	yes	yes	0	18093	15524	100186	-8790	
AT3G13900		yes	-854	-89225	15049	-653	-852	yes
AT3G14140	yes	yes	-40995	37264	-15437	-103048	37196	
AT3G14510	yes		-532	56509	-4013	58519	-27679	yes
AT3G14630	yes	yes	0	5514	8693	7524	8091	
AT3G14640		yes	0	3136	6315	5146	5713	
AT3G14780		yes	-10095	-31118	-5664	-27374	13656	yes
AT3G15170		yes	-12773	26056	318	40094	25853	
AT3G15357	yes	yes	-21993	-21993	4697	-20883	-21917	
AT3G15670	yes		-23154	66209	0	-21848	17612	yes
AT3G15720		yes	22034	49558	-10338	-37252	961	
AT3G16440	yes	yes	0	7610	3402	246843	2950	yes
AT3G17220	yes	yes	-12592	-29890	-10428	-42441	-16029	

AT3G17225	yes	yes	-13638	29441	10046	-43487	-17075	
AT3G17280		yes	0	9759	0	-62566	24787	yes
AT3G17580		yes	-5303	37326	-2368	-174297	-32185	yes
AT3G18200	yes	yes	-15200	6340	2122	-393419	10809	
AT3G18400		yes	24181	24166	-8463	-477858	15407	yes
AT3G18510		yes	-10716	10144	9998	-514777	-10665	
AT3G18610	yes	yes	47637	23542	0	-563196	-59084	yes
AT3G18715	yes		14111	-7932	7746	-600180	44952	yes
AT3G18840	yes		-8953	-7774	-6267	-655311	-8899	
AT3G19040	yes		1016	1020	0	-726270	1018	yes
AT3G19270		yes	-4749	0	0	-832840	28780	yes
AT3G19430		yes	-4938	1581	0	-895297	-6585	yes
AT3G19440		yes	6513	-406	-2618	-899808	8847	
AT3G20130		yes	0	-32922	-22496	771081	12064	
AT3G20898	yes	yes	-4412	30720	-721	475283	-4384	
AT3G20900	yes		-4713	30684	-1022	475247	-4685	
AT3G20975		yes	0	22	8907	444585	168	yes
AT3G20978		yes	0	15	8900	444578	161	yes
AT3G21020	yes	yes	0	-12395	-3979	430391	-12368	yes
AT3G21040		yes	0	-16888	100	425553	-16861	yes
AT3G21150	yes		-3728	-56764	-4980	386378	-14337	
AT3G21180	yes		-5058	-69647	0	367717	-27220	
AT3G21330	yes	yes	2672	2199	-4412	290742	-4441	
AT3G21460	yes	yes	2287	-45836	9651	241959	-385	yes
AT3G21660		yes	7906	38217	7702	173263	7846	
AT3G21720	yes		2246	8860	-2434	143906	2889	yes
AT3G21890	yes	yes	-232	-31041	4334	90184	-5014	yes
AT3G22100	yes	yes	-1132	-7054	264	15448	-1312	
AT3G22540	yes	yes	-1977	-2192	248	9871	-1977	
AT3G22820		yes	7865	64257	-768	19120	7862	yes
AT3G22860		yes	259	45147	157	10	2552	yes
AT3G23260	yes		0	15285	-879	52440	-11135	
AT3G23470	yes		-7204	7618	0	-39455	7635	yes
AT3G24230	yes		0	7419	-1900	58125	-1717	yes
AT3G25590	yes		6054	19026	0	18873	-1511	
AT3G25855		yes	-695	14439	-2065	-5308	0	
AT3G25930	yes	yes	4730	-4810	-2538	-38732	-17646	
AT3G26235		yes	1689	-1311	0	5257	5227	yes
AT3G26290		yes	0	-3148	-3203	-3597	9130	yes
AT3G26445	yes		0	-15769	-2879	9004	1154	
AT3G26614	yes		-377	2821	0	0	329	
AT3G27025		yes	-2653	5662	-4500	34262	-2906	yes
AT3G27473	yes	yes	0	52382	-786	0	-243	yes
AT3G27550	yes		-914	13713	-5913	-3763	-3482	
AT3G27620		yes	9274	-5356	7242	-25027	1298	
AT3G27650	yes	yes	0	-14857	-507	22274	-247	yes
AT3G27710	yes	yes	-5915	-279	-229	-928	-5107	yes
AT3G28750		yes	-42	11175	4881	-12631	-92	
AT3G29035		yes	1095	0	-214	0	892	
AT3G29644		yes	-1268	-5543	-2472	0	1334	
AT3G30122		yes	1695	2283	-66	0	2442	
AT3G42180		yes	0	0	0	0	-3775	yes
AT3G44540		yes	0	3918	3928	-4882	0	yes
AT3G44790	yes	yes	0	-832	0	8700	1335	yes
AT3G44900		yes	4341	-5218	-3845	15169	4340	
AT3G44990	yes	yes	-1163	1490	-748	0	-1176	yes
AT3G45440		yes	0	-20512	9788	-4827	2816	

AT3G45660	yes	yes	363	9272	2204	-18562	6159	
AT3G46280		yes	6001	-9719	100	0	2576	yes
AT3G46680		yes	0	518	-187	-36460	5957	
AT3G46880	yes		1867	3899	5437	100558	3834	yes
AT3G47100	yes	yes	6846	9173	11562	20718	6894	
AT3G47110	yes	yes	2656	4983	7372	16528	2704	
AT3G47420	yes		-1215	696	5235	-11894	-1143	
AT3G47790		yes	-6083	-6083	127	38481	2089	
AT3G47870	yes		-648	-16945	16184	11384	-681	yes
AT3G47875	yes	yes	0	-20740	10301	5501	0	
AT3G48240		yes	6317	-24999	-5631	-113923	13113	
AT3G48250	yes		2775	-27747	-8379	113135	9571	
AT3G48340	yes		171	8429	1740	85856	92	yes
AT3G48550		yes	-7122	103	8731	0	-7300	yes
AT3G48580	yes	yes	6355	-6195	-3662	-7297	6312	
AT3G48970		yes	-5180	17843	12772	28327	-5189	yes
AT3G49055		yes	-368	-259	-1719	0	-4467	
AT3G49570	yes		3562	-15355	-6174	-105634	-14834	
AT3G49580	yes		802	-18091	-8910	-108370	-17570	
AT3G49845		yes	6849	-20695	0	151422	402	
AT3G49960	yes	yes	4654	3734	-6612	114033	-4585	yes
AT3G49970	yes		1275	355	-9744	110654	-7717	
AT3G50840	yes	yes	-8659	-8914	-7557	-10875	0	
AT3G52740		yes	-8559	-1267	-5960	0	-8475	yes
AT3G53140	yes	yes	-25108	-36224	-14	78256	42158	
AT3G53690		yes	-9018	-97846	11226	-108058	59235	
AT3G53940		yes	-11376	50386	-7040	-180289	-11220	
AT3G54770	yes	yes	1166	-39090	11350	38190	1102	yes
AT3G55230	yes	yes	10156	-701	2925	18741	9411	yes
AT3G55240		yes	8282	1048	1051	16867	7537	yes
AT3G55500		yes	0	-1128	1636	-74334	53518	
AT3G55580		yes	-7128	29413	4114	46418	13780	yes
AT3G55720	yes		4158	4153	9767	-7166	353	
AT3G56600	yes	yes	-10461	-11294	-4959	-257438	-10445	
AT3G57540		yes	-2232	3021	11488	29979	3039	yes
AT3G58270	yes	yes	0	31959	0	-233015	-23188	yes
AT3G58980	yes	yes	0	-863	-14269	271302	85594	
AT3G59200		yes	0	598	-3926	178964	-3673	
AT3G59370		yes	-7609	0	1225	125737	26329	yes
AT3G59670	yes		9817	-7227	1402	25595	15187	
AT3G59710		yes	762	763	0	11152	744	yes
AT3G59750	yes		0	-10320	6094	0	4479	
AT3G60140	yes		0	-108650	-4584	4526	-673	yes
AT3G60650		yes	-30341	24157	9926	-14843	-41652	yes
AT3G60670		yes	-34902	18143	3912	-19404	-46213	yes
AT3G61230	yes	yes	-2583	12549	-1730	38033	203	
AT3G61450	yes	yes	-3805	20741	-89	-26088	7504	
AT3G61900	yes	yes	-10665	64182	-3749	192589	5721	
AT3G62090		yes	-10609	0	0	128259	-10548	
AT3G62455	yes	yes	0	4277	-290	14388	4220	yes
AT3G62680		yes	7085	-47231	0	-51410	-65123	yes
AT3G63210	yes	yes	48975	22704	4915	-222303	-236016	
AT3G63320		yes	12579	-11872	1162	-258526	-272239	
AT4G00250	yes	yes	-1052	0	-1063	-57529	-1006	yes
AT4G00390	yes	yes	12443	4311	0	-116321	757	yes
AT4G01430		yes	1695	3646	-302	-2902	-1999	yes
AT4G01680		yes	8070	3901	3928	-46784	10721	

AT4G01990	yes		-1070	-1077	-3256	0	-1046	
AT4G02270	yes	yes	309	7251	-6426	0	362	yes
AT4G02960		yes	0	-12413	0	0	-9648	yes
AT4G03060	yes	yes	-324	0	-940	-36097	-281	yes
AT4G03320	yes	yes	-1095	0	0	0	-1068	
AT4G03480	yes	yes	0	-4572	-5130	14947	-4525	
AT4G03490	yes		-2173	-7724	-8282	11034	-7677	
AT4G03823		yes	58	3848	-2071	0	-2127	
AT4G04760	yes		0	2982	5910	-13226	-565	
AT4G08100	yes		0	-3782	665	0	0	
AT4G08400	yes		1119	-2056	0	0	1059	
AT4G09500	yes		0	-804	-709	9609	-1536	
AT4G10150	yes	yes	2346	491	492	1515	-1537	yes
AT4G10260		yes	1146	456	3662	0	1130	
AT4G10310		yes	0	-3390	0	-1608	0	yes
AT4G10580	yes		0	-2564	-2595	0	-983	yes
AT4G11210		yes	-10161	-782	-154	-38018	-792	yes
AT4G11230	yes	yes	298	-8582	5192	-45818	409	
AT4G11910		yes	-876	3826	-2055	73205	-6053	
AT4G12050		yes	0	-1252	0	9919	-2225	yes
AT4G12090	yes		-5086	9731	1015	-4985	976	
AT4G13235		yes	-2291	4674	5014	4884	-8067	yes
AT4G13260		yes	0	445	-488	-11982	-3019	yes
AT4G13390	yes	yes	-20956	1238	0	26745	-9019	
AT4G13420	yes	yes	8691	4946	0	9820	11749	yes
AT4G13620	yes	yes	10579	12039	-7260	-21131	-6039	yes
AT4G13880		yes	0	14084	0	101302	1686	
AT4G14380	yes		-2384	-22804	0	-18461	8528	
AT4G15248	yes	yes	4623	4307	3444	15803	-1149	
AT4G15340		yes	0	-1380	0	-24034	-3781	yes
AT4G15396		yes	-3726	-40255	-3985	-77081	788	yes
AT4G16000	yes	yes	4460	24631	2195	0	14319	
AT4G16670	yes	yes	-15876	-16153	-16157	-64605	-16020	yes
AT4G16680	yes	yes	-18050	-18327	-18331	-66779	-18194	yes
AT4G16780	yes		-8589	-1916	-2942	33092	-8615	
AT4G16910	yes		0	-16882	-1200	0	10840	yes
AT4G17215	yes	yes	-3562	31209	-258	72321	-18338	
AT4G17905	yes		4367	11336	124	3637	4415	
AT4G17920		yes	-5237	-653	4465	-4460	-5622	yes
AT4G18510	yes		-26647	-36574	11275	-35661	12279	yes
AT4G18550	yes	yes	21023	42962	-1177	-48713	-122	yes
AT4G18900	yes		-4703	-7430	-495	36192	-4466	yes
AT4G19680		yes	4767	662	2175	-139138	684	yes
AT4G19690		yes	568	-202	-342	-143202	-1727	yes
AT4G20390		yes	-6189	-60841	-22550	0	-6136	
AT4G20470	yes		2497	48232	0	0	-31433	
AT4G21020	yes		1121	1696	0	0	1123	yes
AT4G21215		yes	29738	-1535	-9266	43443	47191	
AT4G21323	yes	yes	0	23104	-12312	11520	15268	
AT4G22620		yes	6881	-8753	2079	-60323	5887	
AT4G22666	yes	yes	-1667	3464	2167	-69749	-2958	yes
AT4G22758	yes		3461	2841	3719	-111002	3364	
AT4G22810	yes		0	-20794	-5004	-136832	-4541	yes
AT4G23420		yes	-3584	-34918	-2798	-378207	32982	
AT4G23560	yes		8707	-16605	-194	-446024	-31333	yes
AT4G23700	yes		-5009	6900	1505	-495120	-6529	yes
AT4G24910		yes	10497	-41094	-3305	507030	10499	

AT4G25250		yes	13534	13533	4219	390558	13527	yes
AT4G25420		yes	13797	35279	-8615	333473	-22385	yes
AT4G25470		yes	-8969	11517	-729	309711	3780	
AT4G25480		yes	-11906	8616	1882	306810	879	
AT4G25707	yes	yes	979	23754	2274	228169	26038	
AT4G25790	yes		522	0	723	202584	453	yes
AT4G25820	yes	yes	0	-7883	-5963	194143	-5798	yes
AT4G26010	yes	yes	6031	30314	-430	123990	7859	
AT4G26220		yes	-1879	-2102	-13383	40608	31277	
AT4G26260		yes	4345	-1571	7369	25544	16213	yes
AT4G26800	yes	yes	0	2196	2407	131731	45346	
AT4G26930		yes	7501	-16848	605	94613	8228	yes
AT4G26950	yes	yes	2314	-22555	-3670	89426	3041	
AT4G27400		yes	-47536	4708	-3063	-71603	-42	
AT4G28380		yes	33720	35768	1883	236216	4669	
AT4G28530	yes	yes	-15397	8474	0	182371	27324	yes
AT4G28720	yes	yes	0	42004	-2067	82891	-28484	yes
AT4G28850		yes	4006	4005	21475	31190	3943	yes
AT4G28890	yes		1090	-5338	9697	19412	-5334	
AT4G28930		yes	-7242	883	0	5495	16548	yes
AT4G28940	yes	yes	3088	-1303	0	52	11105	yes
AT4G29180	yes	yes	164	8323	419	447	211	
AT4G29760		yes	6895	-55742	573	-173964	-4715	yes
AT4G29770		yes	4400	-57346	0	-175568	4557	yes
AT4G29800		yes	-8264	44428	-8378	-190634	-8285	
AT4G29880		yes	-24558	28622	10747	-206928	-24579	
AT4G29990	yes	yes	0	-28378	-5831	-265541	13670	
AT4G30180	yes	yes	2656	-289	-289	207717	-18037	yes
AT4G30230	yes		6329	-23009	-1482	169940	6375	yes
AT4G30450	yes		195	30132	521	90690	125	
AT4G31320	yes	yes	-20157	-12994	-5210	-203311	6138	yes
AT4G31380		yes	-55949	-18343	-2806	-239103	1382	yes
AT4G31470	yes	yes	48515	-50304	-9570	-271064	-29447	
AT4G31520	yes		29079	37701	0	270421	-45982	yes
AT4G31940		yes	4998	13239	10707	98000	-9789	yes
AT4G32280	yes	yes	7424	11104	-3601	-17855	7131	yes
AT4G32860	yes	yes	-36501	-32905	1537	196786	2741	yes
AT4G32990	yes	yes	174	203	442	131487	142	
AT4G33040		yes	-16518	-16040	-17151	112815	-16462	yes
AT4G34330	yes		-28500	-36812	-11393	-364697	2310	
AT4G34332	yes		-28503	-36815	-11396	-364700	2310	
AT4G34419	yes	yes	9828	12812	-6325	-396503	-28220	
AT4G34850	yes		-35101	-15669	5734	-548297	-35114	
AT4G35420		yes	-2252	-2245	-5998	-773917	-2234	yes
AT4G35720	yes	yes	4389	-63092	-1089	779626	130	yes
AT4G36380	yes	yes	-12580	-27123	0	516479	21483	yes
AT4G37160	yes	yes	-641	-33400	-2667	211672	4176	yes
AT4G37900		yes	-134	-2719	0	-103578	12973	
AT4G38780	yes	yes	10596	8152	-3511	138615	71371	yes
AT4G39110		yes	0	-19175	-6812	24525	7181	
AT4G39340	yes	yes	5915	69038	10706	-35701	-12230	
AT4G39720		yes	-3225	3397	6414	-172464	-3193	yes
AT4G40090	yes	yes	-8354	-47186	-41	-323410	-47139	
AT5G01330		yes	1521	-18312	-10544	19	-18620	yes
AT5G01570		yes	-26646	33588	-477	-78389	27264	
AT5G01700	yes	yes	-10010	-3094	-9775	83174	-9527	
AT5G02020		yes	5511	6458	5594	-28053	5477	

AT5G02170	yes	yes	32767	-8898	-670	-69868	-33034	
AT5G02540		yes	5826	56379	-3678	-210559	11380	
AT5G02750	yes		39872	6120	-476	-262129	-37595	yes
AT5G03230	yes	yes	1601	-24366	1601	-411549	-3315	
AT5G03552		yes	32074	0	-449	-539055	129144	yes
AT5G04960	yes		25365	0	2796	347695	-4308	yes
AT5G04970	yes	yes	22293	0	0	344623	-6546	yes
AT5G05490	yes	yes	-21153	-36229	3451	182395	-4707	
AT5G05500	yes	yes	-26614	-41690	2324	181268	-10168	
AT5G05880		yes	-510	2781	-6677	40149	-38049	
AT5G06165	yes		28332	-13712	-650	-47379	2856	
AT5G06520	yes	yes	9809	-5634	922	46452	-5151	yes
AT5G06640		yes	12755	-32032	0	0	29572	yes
AT5G06820	yes	yes	-4855	57783	10188	-70326	-40270	
AT5G07010		yes	10730	0	0	79596	10694	yes
AT5G07060	yes	yes	1343	-5388	3375	59739	-4767	yes
AT5G07190		yes	18873	-2906	645	17031	1429	yes
AT5G07200		yes	12205	-8952	-4207	10363	-3254	yes
AT5G07322	yes	yes	-23408	28991	1588	-44644	-23020	
AT5G07380		yes	11669	6491	0	36625	11637	
AT5G07475	yes	yes	8496	5838	4301	7083	-16046	yes
AT5G07571	yes		-19029	-24844	-608	0	-47733	yes
AT5G07690	yes		-3127	10551	0	-45106	-98065	yes
AT5G07700	yes		-6688	7516	-3287	-48667	-101626	yes
AT5G08600		yes	-6787	27905	0	-384529	57559	
AT5G09480	yes	yes	5533	-3524	0	-549256	5407	yes
AT5G09520		yes	-250	-10902	0	-556634	-393	
AT5G09750		yes	4621	4621	5389	-624743	4725	yes
AT5G09805		yes	5444	-15293	-4942	-645560	-15219	yes
AT5G09840	yes		-4938	-27072	0	-657339	17330	
AT5G09970		yes	63	813	293	-710287	61	yes
AT5G10040	yes		17445	-27716	13879	-742065	17443	
AT5G10130	yes	yes	-8873	24089	-216	728895	-8900	yes
AT5G10580		yes	22125	760	7812	551352	22129	yes
AT5G11620		yes	-10662	62556	-376	161277	-214	yes
AT5G11930		yes	4412	7004	0	55778	-29541	yes
AT5G12280	yes		0	-74012	-10424	-55682	-24701	yes
AT5G12380	yes	yes	-36871	53034	1655	-93105	-62124	
AT5G13150	yes	yes	3358	65294	38	28113	-6530	
AT5G13170	yes	yes	-2560	56918	0	19737	-14743	yes
AT5G13330		yes	-1818	-31430	213	37284	7940	yes
AT5G13580	yes	yes	30298	106169	-806	-43278	-44493	
AT5G13910	yes	yes	12129	-2825	0	-155036	-17351	yes
AT5G13990	yes	yes	12649	-20631	3022	-187154	-6069	yes
AT5G14130		yes	-5868	13407	-12695	-231290	1133	
AT5G15022	yes		21112	-9287	947	-65346	-52538	yes
AT5G15160		yes	-6111	-479	-2763	-120375	14252	
AT5G15190	yes	yes	3384	-12033	24	-131929	3323	
AT5G15700	yes		-523	-530	5755	0	-501	
AT5G15890		yes	28002	-17449	0	25943	27952	
AT5G15900	yes		25280	-19353	-476	23221	25230	
AT5G16640	yes		0	-868	-879	161670	-31277	
AT5G16960	yes	yes	0	14394	5855	48777	16790	
AT5G17260		yes	0	33880	7724	-25743	23388	yes
AT5G18020		yes	0	0	-1884	2116	2423	
AT5G18030		yes	0	0	322	0	162	
AT5G18040		yes	535	-186	-2172	0	-1720	

AT5G18050		yes	0	0	1698	0	1746	
AT5G18060		yes	0	0	256	0	304	
AT5G18065		yes	-177	441	-74	0	-179	
AT5G18080		yes	0	0	-5236	-5154	-5341	
AT5G18633	yes	yes	0	23794	7552	0	1656	
AT5G18810		yes	42527	-34223	-1641	-49770	24575	
AT5G19170		yes	0	1780	0	-3512	0	yes
AT5G19800	yes	yes	1113	-1464	0	-54217	10962	
AT5G19970	yes		-45133	7440	-5870	-111119	2759	
AT5G20860	yes	yes	-180	43489	-1708	-15396	-191	yes
AT5G20940	yes		8462	15865	8332	-42433	-6128	
AT5G21150		yes	-1697	-23250	-7689	-130745	-1640	yes
AT5G22410	yes		2823	2823	-10111	-7710	2736	yes
AT5G23155	yes		5665	-4393	-3233	22677	5690	
AT5G23190		yes	-2233	-13392	10160	11872	-2242	yes
AT5G23980	yes		1223	80	-5701	45829	77	yes
AT5G24070	yes	yes	-5274	0	2785	11161	-502	yes
AT5G24100	yes		-378	-378	-4832	0	-317	yes
AT5G24240		yes	-6962	-2769	0	-5612	-3943	yes
AT5G24280	yes	yes	10686	19251	2674	-26048	846	
AT5G24580		yes	-5140	-6392	-6392	-4870	-5159	yes
AT5G24910	yes		1601	17284	-6767	0	-1382	yes
AT5G25045	yes	yes	0	22484	-7521	0	-12815	
AT5G25110	yes	yes	0	5700	0	6600	-327	yes
AT5G25160	yes	yes	2330	-10999	0	-1408	-2358	yes
AT5G25230	yes	yes	1031	1029	2775	0	1506	yes
AT5G26230	yes	yes	1092	2745	1658	657	1458	
AT5G26310	yes	yes	-1981	-7731	1305	-1050	415	yes
AT5G26660	yes		3735	-5638	-725	6203	3726	yes
AT5G26790		yes	-7566	3812	18667	0	-393	
AT5G27220		yes	312	-3747	0	0	548	yes
AT5G27230	yes	yes	-265	269	0	0	-193	
AT5G27889		yes	-3339	-3336	-1275	0	-1138	yes
AT5G27890		yes	2166	-3539	-1478	0	-1341	yes
AT5G27895		yes	0	1301	0	0	-3759	yes
AT5G28415		yes	0	-9247	-1637	0	-3193	
AT5G35190	yes	yes	-589	666	0	0	-582	yes
AT5G35660	yes		-566	-24384	0	0	-491	
AT5G35932		yes	-382	2012	1037	0	-251	
AT5G35935	yes	yes	0	-130	0	0	-489	
AT5G35940		yes	0	3860	-5539	0	1923	yes
AT5G36110		yes	2161	2161	-5204	0	2987	yes
AT5G37072		yes	4242	-4123	581	-4739	-3251	
AT5G37690		yes	-1295	1599	-1234	1845	0	yes
AT5G38230	yes		0	-2358	-1082	-182	-2227	yes
AT5G38320	yes		3823	9791	-4249	0	-7616	yes
AT5G38970	yes	yes	-917	-8221	-7679	0	-4475	yes
AT5G39240		yes	1943	560	1905	0	-1384	yes
AT5G39520		yes	-2504	30258	-3307	-14541	-2536	
AT5G39610		yes	-1455	-2565	-1011	0	-2301	yes
AT5G39840	yes		-507	339	368	0	-470	
AT5G39860	yes	yes	-343	-376	-1628	-4123	-318	yes
AT5G40230		yes	16638	-23257	6351	-42290	-22849	
AT5G40630	yes	yes	3602	3602	3993	357	3530	yes
AT5G42510	yes	yes	-16286	-79	-77	35988	-1069	yes
AT5G42840	yes		0	21553	-5126	26160	-3719	yes
AT5G43230	yes		-1549	-1548	-6939	-21954	-1477	yes

AT5G43290	yes	yes	-4715	-8063	-6469	1599	-1239	yes
AT5G43530		yes	-7281	-8487	0	0	-828	
AT5G43860	yes	yes	6333	48520	0	0	6288	
AT5G44210		yes	-709	-9126	1969	75872	5966	
AT5G44610		yes	2559	1331	0	-18845	924	
AT5G44980	yes	yes	0	-132	6126	0	2536	yes
AT5G45070		yes	0	-15611	4502	17137	-15231	yes
AT5G45150		yes	-207	23541	0	-33789	10154	
AT5G45210		yes	0	-7077	-2408	-68869	8266	yes
AT5G45240	yes		-3638	0	0	-87054	-6607	yes
AT5G46500	yes		-379	-2232	0	85801	-555	yes
AT5G46890		yes	5890	-8527	-6295	0	5884	
AT5G46900		yes	2404	-11932	-9700	0	2398	yes
AT5G47450	yes	yes	0	-42029	1735	-34038	30596	yes
AT5G47610	yes	yes	-20580	34771	-7815	72871	-20589	
AT5G48090		yes	-4008	-3766	-3780	-18840	-4066	yes
AT5G48390		yes	24699	-39135	4491	-94815	-4263	
AT5G48850	yes		15286	-7228	5049	21073	1741	yes
AT5G48950		yes	-12391	47716	-8520	-4721	-13454	yes
AT5G49080	yes		986	2342	0	-49352	17591	yes
AT5G49270	yes		-3940	17342	-1221	-130477	-3910	yes
AT5G49340	yes	yes	-9021	11252	0	-165916	-8949	
AT5G50335	yes	yes	21691	8170	8506	154653	8373	
AT5G50760	yes		-4278	3956	3535	0	-4262	yes
AT5G50790		yes	11867	-6624	-460	-5589	2099	yes
AT5G50820	yes	yes	191	27194	1955	10779	75	yes
AT5G50940		yes	16097	-16439	0	-24681	17121	
AT5G51174	yes		1374	302	6651	-96039	-7037	
AT5G51330		yes	-1941	-1207	-2255	-158809	-1907	
AT5G51390		yes	5616	-3253	2224	-178070	5436	
AT5G51440	yes		-4428	-5461	-112	-191266	-6864	
AT5G52000	yes	yes	0	-10636	9160	39143	-18125	
AT5G52290		yes	6968	17645	5600	-33317	32546	
AT5G53230	yes	yes	0	-23905	-1702	-32718	-14267	yes
AT5G53660	yes		27	-8852	2063	50766	0	
AT5G53700	yes	yes	485	-15459	2413	44025	-4744	
AT5G53895	yes		-19688	-34841	-7399	-23856	-43120	
AT5G53950		yes	11029	21557	1421	-45203	38577	yes
AT5G53980	yes		103	10631	-3610	-57282	27651	
AT5G54190		yes	9203	17758	-2021	22964	8988	
AT5G54230		yes	3040	-5510	1996	0	2867	yes
AT5G54240	yes	yes	-1120	879	-2140	0	-1079	yes
AT5G54370	yes	yes	741	-26150	1778	-50270	17423	yes
AT5G54470	yes	yes	-4925	27558	3194	-89291	-3505	
AT5G54585	yes	yes	-13318	-106	-4704	63269	-25740	yes
AT5G54950	yes		24838	12473	20571	-60275	12258	
AT5G55110	yes	yes	-21000	-14508	1653	-118313	-771	yes
AT5G56320	yes	yes	0	858	0	35691	-6642	yes
AT5G56970		yes	0	-2706	-2706	96220	0	yes
AT5G57090	yes	yes	-17348	24934	-518	39753	24565	yes
AT5G57340		yes	2853	11394	8793	-63970	11318	yes
AT5G57530	yes	yes	0	-473	-134	-138205	8195	
AT5G57625	yes	yes	-14517	-14518	6045	-175734	1293	yes
AT5G58010	yes	yes	-23316	20032	26156	-321429	-50646	
AT5G58400		yes	-6703	-1147	-1169	428236	47672	yes
AT5G58610		yes	-24191	44204	-5259	344211	-24172	yes
AT5G58660		yes	-38887	31269	5657	331276	-38868	yes

AT5G58860		yes	-3885	-18373	6733	267215	-2415	yes
AT5G58910		yes	-1379	5957	-1477	243583	-1323	yes
AT5G59240	yes	yes	4325	-3107	-18973	131531	-76866	yes
AT5G59350	yes	yes	-4266	-8728	-5061	93051	59887	
AT5G59390		yes	-3517	-22880	7053	78113	44949	yes
AT5G59660		yes	0	37105	0	0	-32388	
AT5G59780		yes	27222	-4999	0	-34229	27269	
AT5G59930	yes		0	53340	3634	-86558	-15927	yes
AT5G60250	yes	yes	0	38284	-9803	-204190	58751	yes
AT5G60260		yes	-1443	36848	-13079	-207466	57315	yes
AT5G60310	yes	yes	0	23151	-25656	-220043	43618	
AT5G60520	yes		-10340	-6112	2251	-283702	11055	yes
AT5G60530	yes	yes	8304	-8349	0	-285939	8318	
AT5G60630	yes		338	-6047	-6050	-321682	396	
AT5G61550	yes		0	36572	-6611	-700134	-44203	
AT5G62340	yes	yes	32154	-62304	807	838096	14196	yes
AT5G62430	yes		-2331	97718	2300	801566	-19794	
AT5G63090		yes	0	-10144	0	561527	-4073	yes
AT5G63390	yes	yes	-29381	-49591	-15301	479909	-23813	
AT5G63450	yes		29973	44207	12479	461927	29927	yes
AT5G63660	yes	yes	5109	-14940	571	386438	-14693	yes
AT5G64170		yes	-14437	-14437	0	195754	-3867	
AT5G64190		yes	-21403	-21403	153	191235	-10833	
AT5G64870	yes		-1184	-20114	47	-51408	4516	yes
AT5G65030	yes	yes	-7047	17050	-2046	-97567	-6998	
AT5G65170	yes		-25308	27413	-247	-163044	-3275	yes
AT5G65850		yes	0	-55470	2610	-468184	-55415	yes
AT5G66280	yes	yes	5267	6017	5497	-598036	5246	yes
AT5G66562	yes		-13628	14486	-8468	-689240	14547	
AT5G66607	yes		-998	-448	67	-705301	-232	
AT5G67060	yes	yes	19855	-16327	-16322	-887998	19850	yes
AT5G67390		yes	2399	2765	2830	1000100	2362	
AT5G67400	yes	yes	-1457	-1834	-2460	1000100	-1457	
AT5G67430		yes	-16916	-3180	14753	1000100	-16916	yes
AT5G67620	yes	yes	4794	-57635	617	1000100	8704	yes
AT1TE09820	yes		0	-39343	-1879	-217597	0	
AT1TE13240	yes	yes	18349	-9130	0	37437	-9679	yes
AT1TE13245	yes		18124	-9622	0	37212	-10171	yes
AT1TE14870		yes	-66	-1287	-1286	-167512	-17	
AT1TE15090		yes	22325	21647	0	-235455	0	
AT1TE17095	yes	yes	-3947	66781	0	-185618	-3936	
AT1TE23980		yes	-636	-18600	-228	-65953	-47519	
AT1TE34650	yes	yes	0	-15169	-3	-13223	0	
AT1TE42205	yes		0	7074	177	0	0	
AT1TE53125	yes		7792	-9302	-2531	0	6563	
AT1TE58700	yes	yes	0	2252	-472	0	4397	yes
AT1TE59475		yes	-11217	-15875	20	-67325	-16366	
AT1TE76350	yes		-266	331	3679	-29164	-952	yes
AT1TE80470		yes	-2645	4416	7847	-60082	3022	yes
AT1TE80680	yes		492	2634	792	-120740	-9239	yes
AT1TE93275		yes	2557	0	12453	14046	4682	
AT2TE06390	yes		0	0	0	0	0	
AT2TE06955		yes	0	-20255	0	0	0	yes
AT2TE26680	yes		0	10	0	0	0	
AT2TE28180		yes	3754	-25611	0	0	3706	
AT2TE28345	yes		-3661	3865	0	0	4240	
AT2TE45020	yes	yes	0	12328	0	-66580	696	yes

AT2TE45275	yes		134	-5947	-468	45225	-4988	
AT2TE47040		yes	-1530	16203	691	-45230	1289	yes
AT2TE57460	yes	yes	61	-2101	-7246	115711	3441	yes
AT2TE57465	yes	yes	13	-2482	-7627	115663	3393	yes
AT3TE17230	yes	yes	-6004	-106137	0	182071	-6168	
AT3TE18040		yes	0	-12370	-1541	0	316	
AT3TE18045	yes	yes	995	-19382	4824	0	3172	
AT3TE18060	yes	yes	17	-24352	1608	0	0	
AT3TE30965	yes	yes	0	-12395	-3979	430391	-12368	yes
AT3TE30975		yes	0	-16888	100	425553	-16861	yes
AT3TE40735	yes		-416	3004	0	0	512	
AT3TE46175	yes		-193	9451	-328	0	0	
AT3TE46185	yes		69	8248	-1233	0	-625	
AT3TE64730		yes	0	7028	3410	0	0	
AT3TE66330	yes		0	2576	-2934	-7443	18596	
AT3TE66770	yes	yes	8337	11573	-418	-7828	-235	
AT3TE71565	yes	yes	0	-20614	10296	5496	0	
AT3TE86990		yes	-6888	-2359	-92	-28031	-6904	yes
AT3TE91870		yes	0	2024	5965	-3700	-9217	yes
AT3TE94195	yes	yes	0	4013	-74	14124	3956	yes
AT4TE06710		yes	0	-12261	0	0	-9496	yes
AT4TE10335	yes		0	4283	-133	0	14707	
AT4TE11300	yes		0	6972	9900	-12283	0	
AT4TE11395	yes		0	1257	0	9942	-1839	
AT4TE19585	yes		0	1779	0	0	0	
AT4TE21110	yes		0	-3782	665	0	0	
AT4TE23775		yes	0	0	0	0	0	
AT4TE27915	yes		0	-497	-528	0	0	yes
AT4TE50620	yes		2643	48378	0	0	-31578	
AT4TE61380		yes	1147	23922	2442	228337	26206	
AT4TE63095	yes		-514	-57908	0	-85854	-475	
AT5TE02230		yes	40704	6952	-519	-262172	-37638	yes
AT5TE22500	yes		0	24091	7849	0	1953	
AT5TE29250		yes	0	-2038	-2057	58021	-2028	yes
AT5TE31200	yes	yes	0	22235	-7318	0	-12612	
AT5TE36080		yes	72	2993	1141	0	-3593	yes
AT5TE36085		yes	0	2276	424	0	-5754	yes
AT5TE36090		yes	0	2236	384	0	-6427	yes
AT5TE37815		yes	0	-9000	-1390	0	-2946	yes
AT5TE47605	yes	yes	814	811	0	0	-1095	yes
AT5TE50235		yes	0	0	104	0	0	
AT5TE50260	yes	yes	0	-16	0	0	-375	
AT5TE55265	yes		0	-2889	-1613	-713	-2758	yes
AT5TE64925		yes	3060	1832	0	-19120	1425	
AT5TE71590	yes		-1434	3305	0	-49764	18554	yes
AT5TE71595	yes		1276	2632	0	-50431	17881	yes
AT5TE86640		yes	27253	-6109	0	-35339	27300	
AT5TE87550	yes	yes	8505	8727	0	-287306	8519	

Table 2.6 Genes and that are more than 2 fold down-regulated in *hpr1-5* or *tex1-5* and their association with repeats, H3K9me2, RdDM target loci and H3K27me3.

ID	Down-regulated in *hpr1*	Down-regulated in *tex1*	Distance to dispersed repeat	Distance to inverted repeat	Distance to tandem repeat	Distance to H3K9me2	Distance to sde4 and nrpe1 CHH DMR	H3K27 me3 target
AT1G02660	yes	yes	10517	16117	12612	-136703	12519	
AT1G02800	yes		-7801	-13570	0	-177964	-417	
AT1G02920		yes	-17307	-17307	-230	-223405	10829	
AT1G03020		yes	-833	-56857	-1249	-262955	-808	yes
AT1G03055		yes	-12644	47398	1176	-274766	-12619	
AT1G03400	yes		0	-20262	-11348	-407495	-20248	
AT1G03457	yes		-14492	24680	0	-425536	24575	
AT1G05310	yes		-27058	-23089	0	-103991	-27044	yes
AT1G07000	yes	yes	4784	4355	-3331	-267541	81475	yes
AT1G07135		yes	7506	-24226	0	-307275	43141	
AT1G07620	yes	yes	-3978	-2418	2719	183129	-3891	
AT1G08310		yes	-30218	15821	5111	-82932	-30148	
AT1G08930	yes	yes	-26097	105738	0	-67676	-72766	
AT1G09500		yes	4629	63517	-4000	-260910	-3000	yes
AT1G10550	yes	yes	16764	37759	-4798	298538	16712	
AT1G13750		yes	7071	-8602	2361	-299290	-29295	
AT1G14250		yes	-7835	-47939	0	225871	-7853	yes
AT1G14520		yes	-1125	74371	23436	127048	-943	
AT1G14880	yes	yes	18768	-87444	516	-28818	-30236	yes
AT1G16110		yes	0	-125698	13823	157113	10309	
AT1G16260		yes	0	91603	1746	115891	-27680	
AT1G16390		yes	11653	49053	7274	73341	52637	yes
AT1G17610	yes	yes	1105	-57541	3222	-39071	16443	
AT1G18382	yes		-23209	-56324	-10995	-106578	-52667	
AT1G19020	yes	yes	4867	18621	-2626	17706	5903	
AT1G19380	yes		2096	-2455	-967	4422	5276	
AT1G20690		yes	-12274	-10994	-6438	-82960	-35539	
AT1G20691		yes	-12420	-11140	-6584	-83106	-35685	
AT1G20780	yes	yes	-14663	9782	276	115690	6816	yes
AT1G21100		yes	0	12379	-2204	-10574	3872	
AT1G21250		yes	0	-15648	0	-63001	-44567	
AT1G22530	yes		911	2191	0	0	1458	
AT1G22570	yes	yes	-52	-51	-114	0	-21	yes
AT1G22740	yes		2968	-4687	3262	-62046	-4701	
AT1G23110	yes	yes	2383	33268	-3398	-25966	2869	
AT1G23830		yes	-3271	33238	-860	23074	-13761	
AT1G24145	yes	yes	-1264	-1784	0	-45089	-1228	
AT1G24147	yes	yes	-2683	-3203	-1252	-46508	426	yes
AT1G25400		yes	-14332	7456	368	8766	14958	
AT1G25530		yes	-24453	-43800	2082	-25982	-36745	yes
AT1G26730	yes	yes	0	-4512	-4063	-61741	-4336	
AT1G27730		yes	-18346	-12442	-521	24500	-21042	
AT1G28230		yes	0	0	2474	43031	0	yes
AT1G29720		yes	0	45598	461	-22728	21327	yes
AT1G30040	yes	yes	-2772	-8885	4548	0	369	

AT1G30420	yes	yes	-24071	2077	13754	-38605	-23617	
AT1G31540		yes	0	366	-868	0	-753	
AT1G31690	yes	yes	3747	-42531	2929	-19182	3733	yes
AT1G33600		yes	1042	-4663	2715	0	1037	yes
AT1G33610		yes	-2980	-180	-1077	-390	-4811	yes
AT1G33760	yes	yes	1907	-9393	-1229	987	1802	yes
AT1G33790	yes		0	0	0	0	-1065	yes
AT1G34630	yes		-269	-39111	0	0	-201	
AT1G35140	yes	yes	809	16961	-565	0	-598	yes
AT1G35210		yes	-161	-158	-227	0	-153	
AT1G35350	yes	yes	0	0	-1441	0	-652	
AT1G35515		yes	330	7864	3424	0	235	
AT1G35516		yes	822	8356	3916	0	727	
AT1G35710		yes	-316	-9181	0	0	-241	
AT1G45191	yes		0	6645	0	-1827	0	yes
AT1G48090		yes	0	519	-4097	-12031	556	
AT1G49032		yes	-2134	-6566	-4732	14063	-2169	
AT1G49990		yes	-1373	-24497	0	120421	-19529	
AT1G50040	yes	yes	4138	4146	1790	91572	4152	
AT1G51270	yes	yes	0	577	564	-18531	15528	yes
AT1G52290	yes	yes	0	12303	0	-83427	-3534	
AT1G56240	yes	yes	-1077	2375	-1594	-69044	-1033	yes
AT1G56242	yes	yes	-1246	2137	-1763	-69213	-1202	yes
AT1G56510		yes	0	0	0	-34697	0	yes
AT1G57630		yes	0	-3109	-3983	0	-3040	
AT1G57980	yes	yes	-2296	15638	7837	2919	3195	
AT1G57990	yes	yes	-4569	13523	5722	804	1080	
AT1G58235		yes	-2517	11114	4058	24608	1104	
AT1G58370		yes	0	-54374	-2292	0	-1085	yes
AT1G58561		yes	0	-118465	-2123	0	1554	yes
AT1G61260		yes	-241	-1046	1711	101647	-266	
AT1G62290		yes	-4402	-21088	0	-11762	-15812	yes
AT1G63260	yes	yes	2152	19660	-11161	0	7896	
AT1G63350		yes	0	-1654	20105	-24765	11969	yes
AT1G63710	yes		-7570	45312	3928	-162008	6358	yes
AT1G63750	yes	yes	0	23891	-5536	-180678	0	yes
AT1G63860	yes	yes	0	4141	0	218795	4095	
AT1G65390	yes	yes	0	67478	1236	-6359	-1179	yes
AT1G65486	yes		434	11803	1994	3494	394	yes
AT1G66090	yes	yes	0	622	353	-159548	137	yes
AT1G66160	yes	yes	444	3393	-782	137046	6892	
AT1G66345		yes	1774	2013	2029	36301	1747	
AT1G67470	yes		0	2152	-11285	-102389	1224	
AT1G69890	yes	yes	10507	-26838	1358	41064	11284	yes
AT1G70580		yes	0	-29485	3326	-231420	-29533	
AT1G72060	yes		226	31508	-757	-737235	-248	
AT1G72180		yes	12154	-8858	1819	-781608	19270	
AT1G72360	yes	yes	-532	-54465	-5001	-860595	-508	
AT1G72416	yes	yes	-9644	-71487	-7777	-877617	-17530	
AT1G72520		yes	-5379	107750	-25315	-927303	21246	
AT1G72790		yes	-8424	23705	660	1000100	31601	
AT1G72800		yes	-10893	22004	0	1000100	29900	yes
AT1G72910		yes	0	-1178	-4670	1000100	3456	
AT1G72920		yes	0	-3492	-6984	1000100	1509	
AT1G75310		yes	-10057	54417	12004	264040	-1605	
AT1G75960	yes	yes	0	-2016	8697	10290	926	
AT1G76600		yes	2054	37689	1588	-204380	2055	

AT1G80840	yes	yes	-64448	0	0	-896558	41501	
AT2G01180	yes	yes	13234	2539	1144	-31665	1097	
AT2G01300	yes	yes	287	-25	-38	-75822	271	yes
AT2G04450	yes		362	944	565	13	-121	
AT2G07688		yes	-20914	470	-20651	0	-119271	
AT2G13790	yes	yes	0	-705	-559	0	-421	
AT2G14560	yes	yes	0	0	-2566	0	0	yes
AT2G15040		yes	288	13840	0	0	-4424	
AT2G15042		yes	0	9768	-396	0	763	
AT2G15292		yes	2387	540	3265	-4971	3167	
AT2G15390		yes	0	1439	1250	0	-2561	yes
AT2G16660	yes	yes	2459	0	-3180	0	2475	
AT2G17120	yes		-147	-24863	-908	0	-148	
AT2G17280		yes	2203	6104	-3349	18535	-8529	
AT2G18050		yes	-237	-4114	-2664	-54185	-1124	yes
AT2G18660		yes	0	4363	9310	-36715	0	
AT2G18690		yes	3862	794	2139	-43477	3773	yes
AT2G19800	yes		1626	-3417	-279	24972	1592	yes
AT2G20870		yes	10966	-17998	2100	16538	15288	yes
AT2G21200	yes	yes	0	16040	7437	-11693	-3907	
AT2G22880		yes	-4486	-28405	-5730	-75744	-4520	
AT2G23100	yes	yes	0	-8465	0	19191	-21	yes
AT2G23680	yes	yes	2647	21196	2448	-17574	2645	
AT2G24160		yes	0	1180	1195	87446	7388	
AT2G24600	yes	yes	5079	652	-672	-30467	11076	
AT2G24850	yes	yes	-2838	0	6651	6962	-2931	yes
AT2G25735	yes	yes	-1451	-11255	-60	-86439	-1161	
AT2G26010		yes	3870	489	0	-198727	-10549	yes
AT2G26355		yes	0	2201	-5150	88396	-3309	
AT2G26530		yes	886	13603	0	25702	871	
AT2G26560	yes	yes	-6609	2026	-1077	14125	2494	yes
AT2G27080		yes	-7631	-45559	-15887	-155838	-4634	yes
AT2G27402		yes	-15674	27668	-1459	-315430	15538	
AT2G29280	yes	yes	322	-4579	-229	0	-28690	yes
AT2G29460		yes	-11932	-45226	-3536	-40288	-3537	
AT2G31010	yes	yes	-9982	5103	0	25442	11172	
AT2G31730	yes		-2620	-6415	-9373	-52472	-6069	
AT2G31880		yes	-14161	-41886	0	-119798	3123	
AT2G32030	yes	yes	214	1263	222	-197372	665	
AT2G32290	yes	yes	-52707	-9095	-4708	-279451	371	
AT2G32680	yes	yes	0	3591	0	-424673	3586	yes
AT2G33570	yes	yes	-26229	20672	15114	186555	-26324	
AT2G33580		yes	-28629	18063	12505	183946	-28724	
AT2G34510	yes		6862	-18378	0	-116661	6821	
AT2G35290	yes	yes	-1794	-49458	7866	-170166	-513	yes
AT2G36390	yes		-2238	13123	-21501	-572991	3708	
AT2G37025	yes	yes	14706	-40310	13403	454064	14695	
AT2G37540	yes	yes	-14322	-83752	-741	252202	-17869	yes
AT2G38470		yes	3736	10592	0	-70159	6134	
AT2G40750		yes	-17278	788	788	335542	-26655	
AT2G41990	yes	yes	1693	22966	-3769	-172696	1763	yes
AT2G42040	yes		533	6673	680	-188473	470	yes
AT2G42540		yes	-31295	-40040	-2345	-356471	-35033	yes
AT2G42580	yes	yes	16351	21009	-20028	-374154	-52716	
AT2G43920	yes		-51881	927	-2992	464452	-22909	yes
AT2G44500	yes	yes	10593	-1172	-1172	279010	11010	
AT2G46880	yes	yes	2151	32020	0	-596657	2554	yes

AT2G47130		yes	-4388	-25755	-6872	-681620	3761	
AT3G01290	yes	yes	1629	24262	-6200	1000100	16178	
AT3G02140		yes	36142	16435	3264	779399	36062	
AT3G02390		yes	6233	30904	929	676463	6226	
AT3G02670	yes	yes	19510	429	0	591311	-15681	yes
AT3G03020		yes	247	37587	-12413	483938	78301	
AT3G04210	yes	yes	0	-917	0	57868	-41664	yes
AT3G05320		yes	11238	-13601	3	38935	56605	
AT3G05660		yes	0	-148994	0	-40897	40499	
AT3G06370	yes		-1803	-6085	4262	24265	-1797	yes
AT3G07040	yes	yes	9951	-17251	-2384	-252801	22276	
AT3G07195	yes	yes	-27906	-27907	-5301	-315741	-27893	
AT3G08860	yes	yes	11260	20028	-765	107203	-47960	yes
AT3G09830		yes	-10494	-24054	-10109	92325	-17895	
AT3G10200	yes	yes	20642	-36908	-6	-34316	44512	
AT3G11010		yes	0	-19914	0	-327204	24517	
AT3G11040		yes	-3734	-29561	-1041	-336851	14888	
AT3G11420	yes	yes	-4031	5206	-4129	211333	-4076	
AT3G11840	yes	yes	4693	2963	-7493	67646	17043	
AT3G12110		yes	0	-18875	3158	-43365	14002	
AT3G12520	yes	yes	-1668	14930	-4443	-153338	-1617	
AT3G13437		yes	8799	103	-6073	-86960	76141	
AT3G16030	yes		-1740	-18036	-1823	-151571	-1710	
AT3G19620		yes	-19215	-75324	-5219	-974726	-5514	
AT3G21950		yes	688	39874	-20294	63523	29106	yes
AT3G22231	yes	yes	2488	0	0	52	-424	yes
AT3G22235	yes	yes	-2582	0	146	0	686	yes
AT3G23110		yes	0	9137	0	-8452	-2265	
AT3G23440		yes	2791	-6725	-16	-29103	-6759	
AT3G23550	yes	yes	-21262	0	2348	-73018	1903	yes
AT3G25240		yes	-5804	-27240	-9329	33386	4315	yes
AT3G27960		yes	-5553	2475	-2713	11096	-10981	
AT3G28160	yes	yes	0	-6332	-2972	0	-1057	
AT3G28340		yes	-2973	2795	-235	-21960	2748	
AT3G29000		yes	3759	-3895	0	0	3786	
AT3G29370	yes		4903	6560	-3597	8687	292	yes
AT3G29590	yes	yes	0	-738	-854	0	-1718	yes
AT3G33000		yes	-1305	-262	-2447	0	-5978	
AT3G45860	yes	yes	0	-7962	421	0	-8002	yes
AT3G45970	yes	yes	188	-402	2864	-1750	4992	
AT3G46110	yes	yes	427	421	1090	-13984	374	
AT3G47340	yes	yes	-591	7141	-1480	0	-585	
AT3G47348		yes	2351	2337	-9559	0	1853	
AT3G47350		yes	0	0	9717	0	0	
AT3G47570		yes	722	-4862	747	-67059	771	
AT3G48650	yes	yes	1238	-1185	-2565	-24795	2951	
AT3G50350		yes	-18975	-18763	-7614	-5447	-17003	
AT3G50930		yes	0	-3134	9019	-44226	-727	
AT3G52400		yes	894	917	1161	11391	859	
AT3G55710	yes	yes	9659	9654	-13215	-1193	5854	yes
AT3G55980		yes	-5831	-12851	-6262	-65289	-5796	
AT3G56200		yes	-834	-15707	245	-139023	23573	
AT3G56710		yes	683	-49600	562	-295744	735	
AT3G57260		yes	-57823	383	-7228	143181	-57718	yes
AT3G57470		yes	-12703	857	5346	57027	1287	
AT3G59080		yes	13361	-26075	0	229688	43980	
AT3G59310		yes	9562	17393	-2660	143893	-38247	

AT3G59350	yes	yes	0	6616	-635	133116	33708	
AT3G60420		yes	12120	31060	-6729	40195	28883	
AT3G61380		yes	-6963	-30196	2998	0	-33883	
AT3G62150	yes	yes	15965	-18147	-4927	105059	-30756	
AT3G62570	yes		13989	-6994	0	-11173	-24886	
AT3G62740	yes	yes	0	41604	8399	-80104	-93817	
AT4G00300		yes	1715	0	-4956	-71333	1601	yes
AT4G00955	yes	yes	2193	1	1	0	-1119	
AT4G00970	yes	yes	-292	-2589	-3999	0	-238	
AT4G01525	yes		148	-1251	1	0	2302	yes
AT4G01750		yes	583	-594	12	-87143	5610	
AT4G02330	yes		0	-5563	-5571	-6967	-5414	yes
AT4G02970	yes		-278	-17365	-3070	-1616	-14600	
AT4G06746		yes	534	-1337	-1337	0	-4463	
AT4G08040	yes		483	483	-3835	0	-4613	yes
AT4G09420	yes	yes	-585	652	-748	0	12709	
AT4G11900	yes	yes	-52	-2389	328	77676	-42	yes
AT4G13520		yes	-798	13259	-3988	15727	-895	
AT4G13570		yes	10307	0	3131	0	3818	
AT4G14365	yes	yes	6625	-8502	608	-4159	-4265	
AT4G14400		yes	0	28841	0	9514	0	yes
AT4G15233		yes	0	19122	0	30618	11242	
AT4G15660	yes	yes	-8988	546	546	-11991	-9007	yes
AT4G16563		yes	-9092	37502	1717	-9043	4477	yes
AT4G17030		yes	-5983	-20957	7387	-4729	-5887	yes
AT4G17470	yes	yes	-5026	-717	-7306	-713	-4960	
AT4G17490		yes	8027	-10782	2243	-10778	3033	
AT4G18205		yes	12199	23669	-433	-18105	-23059	
AT4G19520		yes	0	0	0	-75045	221	yes
AT4G19530		yes	0	-214	0	-87520	0	yes
AT4G19820	yes		-8883	2153	2171	128986	-7837	yes
AT4G20780		yes	4002	4739	609	-11486	3330	
AT4G21326	yes		0	19083	11356	7499	11247	
AT4G21870	yes		-19503	-9806	2284	71127	-19569	
AT4G22880		yes	-685	-41422	-1109	-157460	-669	
AT4G23170		yes	0	-635	-827	-287785	-19394	
AT4G23190	yes	yes	0	0	56	-293724	-25333	
AT4G23200	yes	yes	-3641	0	480	-298072	-29681	yes
AT4G23220	yes	yes	643	-3951	-5512	-306648	-38257	
AT4G23260	yes		0	0	11229	-320046	-51655	
AT4G24450	yes		-28380	55550	-762	683134	-28327	
AT4G24570	yes	yes	10849	10697	-1306	638281	10850	
AT4G25110	yes	yes	2090	58644	0	435669	-42419	yes
AT4G25810	yes	yes	868	-6229	-4309	196047	-4144	yes
AT4G26070	yes		-9538	12322	4747	105998	-7750	
AT4G26690		yes	-29342	34190	-7694	-122414	-37035	
AT4G28085	yes	yes	2784	-72619	-4508	316573	-71771	
AT4G28490	yes	yes	-2765	-1092	-1105	195976	-32058	
AT4G29050		yes	0	-3247	-5404	10788	10921	
AT4G29610	yes		-59165	0	-11345	-117298	27971	yes
AT4G29740	yes		12711	-47893	281	-166115	0	yes
AT4G29780		yes	343	55301	657	-179554	500	
AT4G30270	yes		-5203	-35856	-3837	156773	-5193	
AT4G30280	yes	yes	-11819	-42472	-10453	150318	-11809	yes
AT4G30570	yes		-2709	-6548	7582	45414	-17978	
AT4G31000	yes	yes	19181	-2109	-2780	-112477	-36560	
AT4G31800	yes	yes	12716	12716	9221	167131	-9502	

AT4G31870	yes	yes	-11512	-11512	-910	140543	-11506	yes
AT4G32790		yes	1068	8549	1351	239158	1013	
AT4G33810		yes	36301	-818	1384	-153308	-11925	yes
AT4G35110	yes		-2712	4307	-2917	-652446	-6563	
AT4G36648	yes	yes	-30939	-18709	-209	426066	-30929	
AT4G37380		yes	0	-110735	258	134858	42462	
AT4G38400	yes		3682	-71627	1598	-260284	-140744	
AT4G39250	yes	yes	13653	-68149	-2707	-14029	8843	
AT5G01732	yes		-22123	-15207	1503	70199	8274	
AT5G03350	yes	yes	1709	39615	15873	-457709	-49475	
AT5G03995	yes		-9587	10556	-342	-719518	-9584	
AT5G04020	yes	yes	-13715	2314	0	-723646	-13712	
AT5G04340		yes	-9276	51292	357	594558	-34732	
AT5G05300	yes	yes	-806	-2870	2451	241254	48890	yes
AT5G05430	yes		-2222	-17298	-3904	204961	12597	yes
AT5G05600		yes	1524	-84143	1722	136925	1461	
AT5G07770	yes		0	-1027	350	-72984	80498	
AT5G09860	yes		5024	-34637	2067	-664904	8367	
AT5G10380	yes	yes	252	2549	5110	632610	42187	
AT5G10760		yes	26003	42053	-14442	499216	-27143	yes
AT5G14545		yes	-919	-14350	2461	102268	46063	
AT5G15960	yes	yes	6623	6619	8183	4564	6573	yes
AT5G17220	yes	yes	-9697	52406	936	-8880	-13961	
AT5G20230	yes	yes	16284	2996	217	-190075	16471	yes
AT5G22500	yes		0	-2437	0	28980	0	yes
AT5G24030	yes	yes	317	328	15992	24368	316	yes
AT5G24530		yes	-2609	6677	0	-13220	-9734	
AT5G24740		yes	13905	735	8916	12412	21397	
AT5G24770	yes	yes	1384	-1782	-315	0	8876	yes
AT5G25250		yes	-257	-255	-3165	0	-228	yes
AT5G25440	yes	yes	3952	-18830	-4611	-24642	3942	
AT5G25930		yes	-1029	6094	0	0	-895	
AT5G26270	yes	yes	1771	18733	-405	0	-2976	yes
AT5G27420		yes	11976	19380	1549	11846	11939	
AT5G28630		yes	-1319	-13768	0	0	-1332	yes
AT5G35480		yes	1318	3289	1820	0	-11105	
AT5G36000		yes	1803	1861	-290	0	1708	
AT5G36910		yes	-625	10840	-8176	0	-618	yes
AT5G36925	yes	yes	357	7346	-12172	0	-1714	yes
AT5G36930		yes	-749	960	8006	0	570	
AT5G38210		yes	316	-534	6280	0	-1506	
AT5G38212		yes	705	1345	6669	0	2659	
AT5G39020	yes	yes	0	25871	3611	0	3742	yes
AT5G40380		yes	4429	2652	-1839	-12825	-4673	yes
AT5G40450		yes	485	3948	1346	-45703	449	
AT5G40780	yes	yes	-14742	-850	-2554	-37186	29519	yes
AT5G41740		yes	0	-5619	0	4297	-5527	yes
AT5G41750		yes	0	-10902	-797	0	-10810	
AT5G41800	yes	yes	8863	10766	-3621	0	10409	
AT5G42250	yes	yes	943	3229	-492	0	10480	yes
AT5G42800	yes	yes	-1454	-10622	-2932	-32739	9282	yes
AT5G42825		yes	0	-18445	174	34694	2458	
AT5G42980		yes	3792	7348	-7435	-21965	3788	
AT5G42990		yes	1716	5272	-8815	-23345	1712	
AT5G44420		yes	-2222	41093	-8745	0	-5470	yes
AT5G44430		yes	-4098	39184	-10621	0	-7346	yes
AT5G44520		yes	-160	12164	-2927	-14777	-7588	

AT5G44568	yes		-512	0	-1338	0	-4002	
AT5G45340	yes	yes	-10568	-499	2522	115617	10109	
AT5G46450		yes	0	14549	0	104042	14354	
AT5G48450		yes	4418	46955	-12874	-115544	4536	
AT5G48540		yes	882	12699	1312	-151717	880	
AT5G49640		yes	-3268	-31545	0	142018	-5664	
AT5G51190		yes	-3013	-4179	1286	-100687	-11685	
AT5G52050		yes	16715	20858	-5791	19619	20757	
AT5G52320		yes	0	7517	1395	-47942	22418	yes
AT5G54060	yes	yes	0	-5477	-6558	77407	4031	yes
AT5G54380		yes	0	-27959	0	-52079	14320	
AT5G54490		yes	10108	21099	2664	-96391	-10605	
AT5G54610	yes	yes	10549	-9346	-13944	53435	-34980	yes
AT5G54710	yes	yes	451	36787	-4103	9460	500	yes
AT5G54720	yes	yes	0	34338	-8724	7011	-96	
AT5G55570		yes	-1288	-2003	367	114110	-576	
AT5G55930		yes	1190	21259	0	-12772	-9516	yes
AT5G56130		yes	-51325	-13101	1353	-82539	-47083	
AT5G57123		yes	1687	2015	1824	16834	1646	
AT5G57220		yes	0	-23632	-10832	-25724	-21141	yes
AT5G57560	yes	yes	-1278	-7282	-2727	-145014	1518	
AT5G58120		yes	0	-12343	828	-355376	24733	
AT5G59130		yes	0	-12298	-1553	161544	-44499	yes
AT5G59670	yes	yes	0	31432	1901	0	-38195	
AT5G60780	yes	yes	-1464	12834	562	-402871	-1440	yes
AT5G61010		yes	1243	-29095	0	-505873	-10942	
AT5G61560		yes	-1811	31725	-11359	-704882	-48951	
AT5G63160	yes		5620	4893	-14990	537015	25415	
AT5G66053		yes	-1055	66964	2499	-538531	47993	
AT5G66110	yes		-14596	52859	-2994	-552072	33888	
AT5G66400	yes	yes	17872	-33499	0	-640147	17858	yes
AT5G66790	yes	yes	2848	16269	25159	-787012	2769	
AT1TE02360		yes	-29321	32364	-13650	-291443	-29296	
AT1TE04835		yes	0	-20808	0	-31121	-874	
AT1TE11950	yes		-25211	-97985	-1765	106514	89088	
AT1TE14655		yes	-2395	-6878	-2030	-96480	12433	
AT1TE25675		yes	967	-851	-623	0	941	
AT1TE69995	yes	yes	-392	-1434	-1227	-40469	-262	yes
AT1TE70530		yes	6902	-4181	-6450	-7999	9615	
AT1TE71765		yes	47	-116302	0	0	-1413	yes
AT1TE71770		yes	0	-116530	-188	0	-1641	yes
AT1TE71775		yes	0	-117984	-1642	0	1412	yes
AT1TE80025	yes		892	12261	2452	3952	852	yes
AT1TE80815	yes	yes	0	1444	1175	-161182	959	yes
AT1TE89775		yes	-388	-4353	-7845	1000100	1611	
AT2TE01080	yes	yes	-1081	-12972	5505	-17301	0	
AT2TE20520	yes		-1292	17070	1338	0	-905	
AT2TE25295	yes	yes	249	1	-3729	0	460	yes
AT2TE27080		yes	2472	625	3350	5750	3252	
AT2TE27700		yes	-1192	931	1463	0	-1292	
AT2TE35695	yes		3757	-2710	0	27103	-1090	yes
AT2TE42065	yes		-13	-10390	0	19781	-1946	yes
AT2TE48800	yes	yes	-310	10252	-445	22351	-271	
AT2TE50950		yes	-15695	28152	-1480	-315451	16022	
AT2TE52400		yes	19723	14514	-2138	423300	43904	yes
AT2TE57395	yes	yes	-1364	11715	-12	130152	6382	
AT2TE72160		yes	5751	12607	364	-70386	8149	

AT3TE02890		yes	688	38028	-12878	484379	78742	
AT3TE34080	yes		240	12571	0	-20236	276	yes
AT3TE43605	yes	yes	734	-6370	-3010	0	-1095	
AT3TE43610	yes	yes	13	-6695	-3335	0	-1420	
AT3TE43615	yes	yes	0	-7385	-4025	0	-2110	
AT3TE44620	yes		-147	3401	-3333	10240	-3263	
AT3TE94450	yes		-13916	-37296	-4300	-41475	-55188	
AT4TE03410	yes		148	-1251	1	0	2302	yes
AT4TE10620	yes	yes	0	4546	1112	0	-3337	
AT4TE12660	yes		1319	-26322	1319	0	-2744	
AT4TE18395		yes	-417	-1296	773	0	-253	
AT4TE31060		yes	0	-2174	-2174	81052	0	
AT4TE32800		yes	2648	0	-1763	37488	2707	
AT4TE42070		yes	-10204	37623	1838	-10155	4598	yes
AT4TE43210	yes	yes	0	-29	-3278	8500	4973	yes
AT4TE48760		yes	0	-888	-3861	-88194	640	yes
AT5TE38735	yes	yes	-1581	-14030	0	0	-1594	yes
AT5TE48720	yes	yes	1172	3143	1674	0	-11325	
AT5TE49540		yes	986	982	-2955	0	1012	
AT5TE53790	yes		-8271	-1968	2511	-14250	-3183	yes
AT5TE58300		yes	-1005	-8353	68	0	-1002	yes
AT5TE67880		yes	0	257	250	89750	62	yes
AT5TE72580		yes	-3369	-31646	0	141804	-5765	
AT5TE75530	yes		0	-10737	960	192708	-6489	
AT5TE80010	yes	yes	146	36482	-7226	9155	195	
AT5TE80020		yes	-53	34938	-8845	7611	-217	
AT5TE83620		yes	2546	2541	-1940	-94084	-14801	yes

Conclusion

In my thesis research, I have worked on two independent research projects to study the transcriptional and post-transcriptional regulations of gene expression in *Arabidopsis*. The transcript levels of some genes are post-transcriptionally regulated by miRNAs and therefore affected by the accumulation level of miRNAs. One goal of my research was to study the molecular mechanisms underlying miRNA degradation. I identified a nucleotidyl transferase HESO1 that functions in the degradation of miRNAs through a reverse genetic approach. I studied the molecular function of HESO1 and clarified the relationship between small RNA tailing and degradation in *Arabidopsis hen1* mutant. Another goal of my research was to identify novel players involved in the DNA methylation pathways through a forward genetic approach. Using two luciferase reporter lines that were regulated by DNA methylation, I isolated mutants with reduced luminescence and cloned three genes encoding subunits of the THO complex. I found that the THO complex mutants did not function in the regulation of DNA methylation. I identified genes that were differentially expressed in the THO complex mutants and characterized some features of the genes.

Mechanistic study of small RNA degradation in *Arabidopsis*

I identified a gene *HESO1* that functions in small RNA tailing and degradation from a family of nucleotidyl transferases in *Arabidopsis*. I found that a mutation in *HESO1* partially suppresses the developmental defects of *hen1* by reducing the degradation level of miRNAs. HESO1 functions in small RNA degradation by tailing unmethylated small RNAs. Additionally, with reduced tailing level of miRNAs in *hen1 heso1* compared to *hen1*, the 3'-end truncation of miRNAs was not obviously affected, suggesting that a nuclease distinct from the one that

mediated miRNA 3'-end truncation degraded the tailed miRNAs. This work uncovered the small RNA tailing mechanism during the degradation process of small RNAs.

Characterization of the THO complex mutants

A forward genetic screen was conducted to identify potential players involved in the DNA methylation pathways using two luciferase reporter lines which were regulated by DNA methylation. Three mutant lines with reduced luminescence were isolated and cloned and the disrupted genes encoded TEX1, HPR1 and THO5A, three subunits of a conserved nuclear protein complex THO, respectively. Although the reporter lines were regulated by DNA methylation, I found that mutations in the THO complex genes did not affect DNA methylation levels at either the *LUC* promoter or endogenous loci. I profiled the transcriptomes of *tex1* and *hpr1* and found that genes that were in close proximity to dispersed repeats, inverted repeats were enriched among the down-regulated genes in *tex1* and *hpr1*. Genes with H3K27me3 were also enriched among the up-regulated and down-regulated genes in *tex1* and *hpr1*. Although the THO complex was not likely to regulate DNA methylation as expected, this work identified genes that were regulated by THO complex in a genome-scale and provided insights into the potential target of the THO complex.

Appendix A
Isolation of a low luminescence mutant from *YJ11-3F* and gene cloning

From *YJ11-3F* screening, I isolated a mutant named *YY-1170* with reduced luminescence compared to the reporter line *YJ11-3F*. The *YY-1170* mutant has decreased *LUC* transcript levels (Figure A.1A).

The mutated site was identified by map-based cloning and whole genome high-throughput sequencing. A mutation at *AT3G06290* (*SAC3B*) was identified (Figure A.1(C)). The *AT3G06290* gene encodes a yeast TREX-2 subunit SAC3.

The *YJ11-3F* and *YY-1170* seedlings were treated with 5-aza-2' deoxycytidine, a chemical that inhibits DNA methylation. After treatment, *YY-1170* plants showed decreased levels of luminescence compared to *YJ11-3F*, indicating *LUC* expression was repressed through a DNA methylation-independent mechanism in *YY-1170*.

Whole genome bisulfite sequencing was performed with *YJ11-3F* and *YY-1170*. Few DMRs were identified (Table A.1) in *YY-1170* compared to *YJ11-3F*. This indicats that *sac3b-3* mutation does not have a significant effect on DNA methylation.

The TREX-2 complex is a conserved complex in eukaryotes consisted of THP1, SAC3B, CDC31 and SUS1 in eukaryotes. It associates with the nuclear pore complex (NPC) in yeast and human and mediates mRNA export [1] [2]. In *Arabidopsis*, the SAC3B protein is located at nuclear periphery [3], and it interacts with the yeast THP1 homolog AtTHP1 and the CDC31

homolog AtCEN1 and 2 *in vitro* [3]. Thus, it is likely that the mutation in *SAC3B* compromises *LUC* transcript accumulation in *YJ11-3F* through a defect in mRNA export.

References

1. Fischer, T., Strasser, K., Racz, A., Rodriguez-Navarro, S., Oppizzi, M., Ihrig, P., Lechner, J., and Hurt, E. (2002). The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. The EMBO journal *21*, 5843-5852.

2. Umlauf, D., Bonnet, J., Waharte, F., Fournier, M., Stierle, M., Fischer, B., Brino, L., Devys, D., and Tora, L. (2013). The human TREX-2 complex is stably associated with the nuclear pore basket. Journal of cell science *126*, 2656-2667.

3. Lu, Q., Tang, X., Tian, G., Wang, F., Liu, K., Nguyen, V., Kohalmi, S.E., Keller, W.A., Tsang, E.W., Harada, J.J., et al. (2010). Arabidopsis homolog of the yeast TREX-2 mRNA export complex: components and anchoring nucleoporin. The Plant journal : for cell and molecular biology *61*, 259-270.

Figures and Tables

Figure A.1 Mutation in *YY-1170* decreases *LUC* expression in a DNA methylation-independent manner

(A) Relative transcript levels from the *LUC* transgene in *YJ11-3F* and *YY-1170* as determined by real-time RT-PCR. The transcript levels of *LUC* were normalized to *UBQ5*. Compared to *YJ11-3F*, *YY-1170* showed decreased *LUC* transcript levels.

(B) Luminescence of *YJ11-3F* and *YY-1170* with or without 5-aza-2’ deoxycytidine treatment. *YY-1170* seedlings showed lower luminescence with or without treatment compared to *YJ11-3F*.

(C) Gene structure of *SAC3B* and the mutation site. Boxes represent exons. Gray and white boxes indicate coding regions and UTRs, respectively. The T-DNA insertion site is indicated by the arrow.

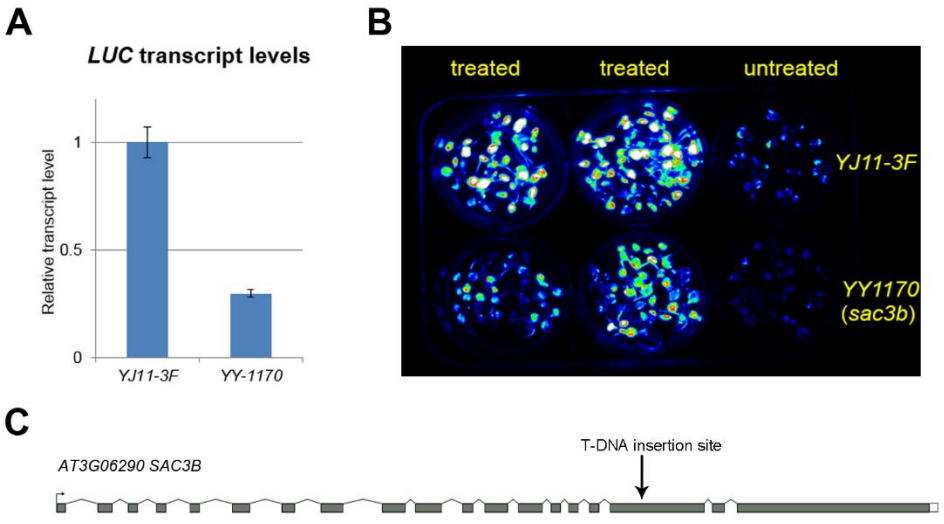

A
LUC transcript levels
Relative transcript level
1
0.5
0
YJ11-3F
YY-1170
B
treated
treated
untreated
YJ11-3F
YY1170
(sac3b)
C
AT3G06290 SAC3B
T-DNA insertion site

Table A.1 Number of differentially methylated regions in *YY-1170* (*sac3b*)

	hypermethylated DMR	hypomethylated DMR
CHH	25	82
CHG	2	5
CG	41	42

Appendix B
Construction of bisulfite sequencing libraries

I constructed bisulfite sequencing libraries for 74 samples including lines mentioned in Chapter II and mutant lines from my co-workers in the lab. The basic information of each library is listed in Table B.1. Sequencing data could be obtained from http://illumina.ucr.edu/ht.

Table B.1 Bisulfite sequencing libraries

Ecotype	Genotype	index #	Lane #	Index sequence	Project ID	Reference line for analysis	Mutated gene	Comment
Col	9-7-2_r2	2	5	CGATGT	468	Ctrl		
Col	28-6L_r2	4	5	TGACCA	468	9-7-2_r2	*TEX1*	
Col	115-1H_r2	5	5	ACAGTG	468	9-7-2_r2	*AGO4*	
Col	86-1H_r2	6	5	GCCAAT	468	9-7-2_r2	*DRD1*	
Col	hpr1 ago4_r2	2	6	CGATGT	468	9-7-2_r2	*HPR1 AGO4*	
Col	10-34L_r2	4	6	TGACCA	468	9-7-2_r2	*AT3G04490*	
Col	P254R_r2	6	6	ACAGTG	468	YJ_r2	*HSP20/LIL*	
Col	L1_r2	12	6	CTTGTA	468	9-7-2_r2	*HSP20/LIL*	
L*er*	WT-Ler_r2	2	8	CGATGT	471	Ctrl		
L*er*	pwr-1_r2	4	8	TGACCA	471	WT-Ler_r2	*PWR*	
L*er*	top1a_r2	5	8	ACAGTG	471	WT-Ler_r2	*TOP1A*	
Col	s112625_r2	6	8	GCCAAT	471	Col_r2	*TOP1A*	
Col	Col_r2	5	1	ACAGTG	502	Ctrl		
Col	sde4_r2	12	1	CTTGTA	502	Col_r2	*NRPD1*	
Col	tho5tho5l_r2	18	1	GTCCGC	502	Col_r2	*THO5A*	*tho5a-1* mutant crossed with *tho5b* mutant *salk_122387*, *THO5B*(*AT1G45233*) transcript was not affected
Col	nrpe1_r2	19	1	GTGAAA	502	Col_r2	*NRPE1*	
Col	hpr1 drd1_r2	4	2	TGACCA	502	9-7-2_r2	*HPR1 DRD1*	
Col	nua-3_r2	5	2	ACAGTG	502	Col_r2	*NUA*	
Col	11-63H_r2	6	2	GCCAAT	502	9-7-2_r2	*TAF6*	
Col	pwr-2_r2	18	2	GTCCGC	502	Col_r2	*PWR*	
Col	115-1Htop1a_r2	2	3	CGATGT	502	9-7-2_r2	*AGO4 TOP1A*	
Col	P204R_r2	6	3	GCCAAT	502	YJ_r2	*HPR1*	
Col	YJ_r2	18	3	GTCCGC	502	Ctrl		
Col	P100R_r2	19	3	GTGAAA	502	YJ_r2	*NUA*	
Col	SB2_r2	2	5	CGATGT	502	YJ (lin)_r2		overexpressing *SSB2*
Col	RH2_r2	5	5	ACAGTG	502	YJ (lin)_r2		overexpressing *RNH2*
Col	YJ (lin)_r2	6	5	GCCAAT	502	Ctrl		
Col	RH1_r2	19	5	GTGAAA	502	YJ (lin)_r2		overexpressing

								RNH1
Col	963DMSO_r2	2	6	CGATGT	502	963_r3		
Col	KU_r2	4	6	TGACCA	502			
Col	963_CPT_r2	7	6	CAGATC	502	963DMSO_r2		
Col	KU-CPT_r2	19	6	GTGAAA	502			
Col	SB3_r2	4	7	TGACCA	502	YJ (lin)_r2		overexpressing *SSB3*
Col	RH1_r3	12	7	CTTGTA	502	YJ_r3		overexpressing *RNH1*
Col	SB2_r3	13	7	AGTCAA	502	YJ_r3		overexpressing *SSB2*
Col	RH2_r3	18	7	GTCCGC	502	YJ_r3		overexpressing *RNH2*
Col	Col_r3	6	8	GCCAAT	502	Ctrl		
Col	s093159_r3	7	8	CAGATC	502	Col_r3	*AT3G04490*	
Col	sde4_r3	13	8	AGTCAA	502	Col_r3	*NRPD1*	
Col	86-1H_r3	16	8	CCGTCC	502	9-7-2_r3	*DRD1*	
Col	SUVH1_r4	4	1	TGACCA	503	YJ_r4	*SUVH1*	
Col	yy1170_r4	7	1	CAGATC	503	YJ_r4	*SAC3B*	
Col	963_r3	16	1	CCGTCC	503	Ctrl		
Col	yy1170_r3	19	1	GTGAAA	503	YJ_r3	*SAC3B*	
Col	YJ_r3	4	2	TGACCA	503	Ctrl		
Col	10-34L_r3	6	2	GCCAAT	503	9-7-2_r3	*AT3G04490*	
Col	nrpe1_r3	14	2	AGTTCC	503	Col_r3	*NRPE1*	
Col	hpr1 drd1_r3	19	2	GTGAAA	503	9-7-2_r3	*HPR1 DRD1*	
Col	9-7-2_r3	5	3	ACAGTG	503	Ctrl		
Col	28-6L_r3	6	3	GCCAAT	503	9-7-2_r3	*TEX1*	
Col	115-1Htop1a_r3	7	3	CAGATC	503	9-7-2_r3	*AGO4 TOP1A*	
Col	hpr1ago4_r3	12	3	CTTGTA	503	9-7-2_r3	*HPR1 AGO4*	
Col	P204R_r3	5	5	ACAGTG	503	YJ_r3	*HPR1*	
Col	tho5tho5l_r3	12	5	CTTGTA	503	Col_r3	*THO5A*	*tho5a-1* mutant crossed with *tho5b* mutant *salk_122387*, *THO5B*(*AT1G45233*) transcript was not affected
Col	SB3_r3	14	5	AGTTCC	503	YJ_r3		overexpressing SSB3
Col	SUVH1_r3	18	5	GTCCGC	503	YJ_r3	*SUVH1*	
Col	YJ_r4	6	6	GCCAAT	503	Ctrl		
Col	11-36H_r3	7	6	CAGATC	503	9-7-2_r3	*TAF6*	
Col	115-1H_r3	18	6	GTCCGC	503	9-7-2_r3	*AGO4*	
Col	s100012_r3	19	6	GTGAAA	503	Col_r3	*TEX1*	

Col	Col_r4	12	1	CTTGTA	538	Ctrl		
Col	pwr-2_r4	14	1	AGTTCC	538	Col_r4	*PWR*	
Col	s112625_r4	16	1	CCGTCC	538	Col_r4	*TOP1A*	
L*er*	Ler_r3	13	2	AGTCAA	538	Ctrl		
L*er*	top1a_r3	18	2	GTCCGC	538	Ler_r3	*TOP1A*	
Col	nua-3_r4	15	2	ATGTCA	538	Ler_r3	*NUA*	
Col	Col_r5	6	2	GCCAAT	586	Ctrl		
Col	rdd_r5	13	2	AGTCAA	586	Col_r5	*ROS1* *DML2* *DML3*	
Col	YJ11-3F_r5	16	2	CCGTCC	586	Ctrl		
Col	P254R_r5	19	2	GTGAAA	586	YJ11-3F_r5	*HSP20/ LIL*	
Col	Col_r6	6		GCCAAT	612	Ctrl		
Col	Col_r7	12		CTTGTA	612	Ctrl		
Col	pwr-2_r6	13		AGTCAA	612	Col_r6	*PWR*	
Col	pwr-2_r7	19		GTGAAA	612	Col_r7	*PWR*	

Appendix C
R-loop profiling in *Arabidopsis*

As mentioned in Chapter II, the THO complex in yeast functions in preventing RNA-DNA hybrid (R-loop) formation during transcription. In *tho* mutant strains, R-loops are formed at some transcribed loci and result in transcription associated hyper-recombination. The formation of R-loops could diminish Pol II processivity. Thus, preventing R-loop formation or stabilizing R-loops may regulate transcript accumulation. In order to study whether the THO complex affects R-loop formation or not in *Arabidopsis*, I tried to profile R-loop accumulation in Col, *tex1-5* and *hpr1-5* by DNA immunoprecipitation (DIP).

Test of antibody specificity

At an R-loop generating locus, both R-loop and single-stranded DNA (ssRNA) exist. Thus we obtained two hybridoma cell lines from ATCC (American Type Culture Collection) that express an R-loop antibody (HB-8730) and a single-stranded DNA antibody (HB-69), respectively. I purified the two antibodies from ascites that were produced from the two hybridoma cell lines.

In order to test the specificities of the two antibodies, I used a pGEM-T Easy vector with *35S* promoter sequence insertion (pGEM-T-35S) to prepare different nucleic acid antigens for immunoblotting. The pGEM-T-35S plasmid was purified by the Zymo plasmid miniprep kit. Double stranded DNA (T7/SP6-dsDNA) was generated using PCR with T7 and SP6 primers and the PCR product was purified with Zymo DNA Clean & Concentrator™-5 kit. Single-stranded DNA (SP6-ssDNA) was generated by PCR with SP6 primer and linearized vector as templates and the PCR product was purified with the Zymo DNA Clean & Concentrator™-5 kit following

the ssDNA purification protocol. Single-stranded T7-RNA was generated by *in vitro* transcription by T7 RNA polymerase using a linearized pGEM-T-35S vector DNA as the templates. The product was heat inactivated and treated with Dnase I to remove the template DNA. Then the RNA was purified by phenol/chloroform extraction and isopropanol precipitation. *In vitro* transcribed RNA using SP6 RNA polymerase (SP6-RNA) was prepared through the same procedure. Double-stranded RNA (T7/SP6-dsRNA) was generated by mixing T7-RNA and SP6-RNA with a molar ratio of 1:1 followed by heat denaturation and subsequent slow cooling to room temperature. The R-loop sample (T7/SP6-R-loop) was generated by mixing SP6-ssDNA and T7-RNA with a molar ratio of 1:1 followed by heat denaturation and subsequent slow cooling to room temperature.

The above-described different types of nucleic acids were spotted on a nitrocellulose membrane followed by UV crosslinking. Then the membrane was baked at 80°C for 30 min.

The prepared membranes were immunoblotted using the R-loop antibody and the ssDNA antibody through a conventional western blotting procedure and signals were detected with the Amersham ECL Prime Western Blotting Detection Reagent.

As indicated in Figure C.1 (left panel), the R-loop antibody could detect R-loops (left panel, 1~4). It also interacted with ssDNA (left panel, 9) and dsRNA (left panel, 21) with lower efficiency. Although it also interacted with plasmid DNA (left panel 5 and 6), the signal was likely resulted from R-loops in the plasmid DNA rather than dsDNA since the dsDNA sample generated from PCR (left panel 13~16) was not detected by the antibody. Considering that ssDNA often co-exists with R-loop and dsRNA cannot be detected by sequencing without reverse transcription, the R-loop antibody could be used to specifically immunoprecipitate R-loops for R-loop profiling.

On the other hand, the ssDNA antibody could detect ssDNAs (right panel, 9~12). It also recognizes R-loops (right panel, 1~3) and dsDNAs (right panel 13 and 14). The signal detected from R-loop samples could resulted either from R-loops or from free ssDNAs that were not hybridized to ssRNA. However, the cross-reaction of ssDNA antibody with dsDNAs may introduce nonspecific signals from dsDNAs in ssDNA profiling.

R-loop immunoprecipitation in *Arabidopsis*

Based on the specificities of the two antibodies, I proceeded with immunoprecipitation only with the R-loop antibody.

First, nuclei was isolated from seedlings and nucleic acid was purified from the nuclei by phenol/chloroform extraction and precipitation. The purified nucleic acid was fragmented by sonication. One nucleic acid sample was also treated with Rnase H that specifically degrades RNAs from R-loops. Input nucleic acid, sonicated nucleic acid from two biological replicates as well as an Rnase H treated sample were tested for R-loop accumulation using the R-loop antibody. As shown in Figure C.2B, R-loops can be detected from total nucleic acid samples (sample 1 and 2). R-loop signals were greatly reduced by Rnase H treatment (sample 5), suggesting that R-loops exist in *Arabidopsis* nucleic acids. In addition, R-loops can be detected from sonication fragmented nucleic acids with similar intensity (Figure C.2 sample 3 and 4), suggesting that sonication does not significantly disrupt R-loop structure. However, it remains to be determined whether R-loops at some loci are sensitive to sonication or not. I also tested fragmentation by restriction digestion. The nucleic acid sample was digested by XbaI and MboI (Figure C.2C). The fragmented sample was purified by AMPure XP beads. As indicated in Figure C.2D, the R-loops were not significantly affected by restriction digestion and can be retained through AMPure beads purification.

I performed R-loop precipitation with nucleic acids extracted from Col, *tex1-5*, and *hpr1-5*. Three biological replicates were prepared. Nucleic acid samples from replicate 1 and 2 were fragmented by sonication and nucleic acid samples from replicate 3 were fragmented by restriction digestion. In replicate 1, DNA was not pre-cleared before immunoprecipitation. In replicate 2 and 3, DNA was pre-cleared with protein G beads before immunoprecipitation. Samples without antibodies during IP were used as negative controls in replicates 1 and 2. Samples treated with Rnase H prior to IP were used as negative controls in replicate 3. DNA sequencing libraries were constructed using the immunoprecipitated products and sequenced by high-throughput sequencing.

Preliminary data of R-loop IP libraries

Libraries from replicates 1 and 2 have been sequenced and the reads numbers are listed in table C.1. In replicate 1, the *hpr1* sample has a much smaller number of reads than the other two genotypes, and the numbers of reads in replicate 2 are more even among the three genotypes. The background signal is lower in replicate 2 compared with replicate 1 (Figure C.3A). From genome browser views (Figure C.3B), R-loop signals can be observed at expressed genes or genes that do not show mRNA reads from mRNA-seq. Some intergenic regions also show R-loop accumulation, suggesting that transcription of non-coding RNAs also leads to R-loop accumulation at some loci.

In replicate 2, R-loop accumulation show similar patterns of distribution among Col, *hpr1* and *tex1* in the genome for most loci (data not shown). Further quantitative analysis is required to identify loci with different levels of R-loop accumulation in *hpr1* and *tex1* compared to Col.

Discussion

So far, Genome-wide R-loop profiling has not been reported in plants. Our preliminary data on R-loop profiling suggests that the R-loop antibody specifically discriminates R-loops from other types of nucleic acids and can be used to profile R-loops using purified nucleic acids.

From the sequencing data of replicate 2, we have an overview of R-loop accumulation in *Arabidopsis*. R-loops originate from transcription. From the genome browser, we have observed that R-loops are often found at transcribed genes or regions very close to the transcribed region of genes. However, the abundance of R-loops are not determined by the abundance of mRNAs. Some genes with low mRNA levels may have R-loop accumulation while at some highly expressed genes, the accumulation of R-loops is hardly observed. R-loops are also found at some intergenic regions, indicating that transcription of non-coding RNAs may also results in R-loop formation.

From this preliminary study, we found that most R-loops are relatively stable and can be retained during sonication. However, replication 3 was prepared in which nucleic acids were fragmented by restriction digestion. This different fragmentation approach may help us to identify potential R-loop loci that are not very stable. In addition, Rnase H-treated samples were used as negative controls to prevent the identification of false positive signals. Thus, data from replicate 3 may help us to further decide whether these optimizations may improve the sensitivity and specificity of the R-loop IP.

Figure C.1 Specificities of the R-loop antibody and the ssRNA antibody

The membranes were spotted with different types of nucleic acids and immunoblotted using the R-loop antibody and the ssDNA antibody. The samples are listed as follows:

	100ng	20ng	10ng	2ng
T7/SP6-R-loop	1	2	3	4
pGEM-T-35S plasmid DNA	5	6	7	8
SP6-ssDNA	9	10	11	12
T7/SP6-dsDNA	13	14	15	16
T7-ssRNA	17	18	19	20
T7/SP6-dsRNA	21	22	23	24

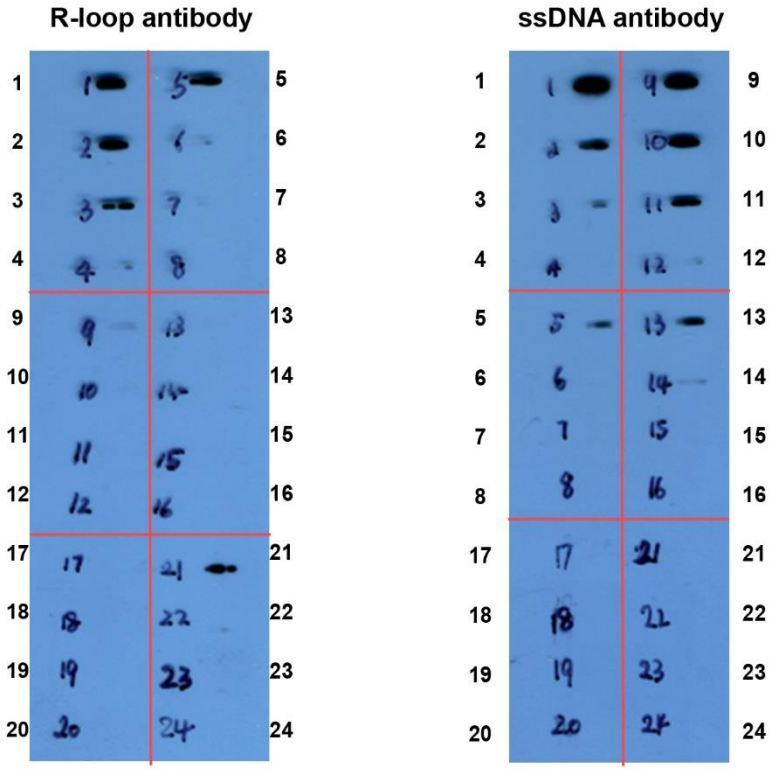

R-loop antibody
ssDNA antibody

Figure C.2 Detection of R-loops from *Arabidopsis* nucleic acids and test of fragmentation conditions.

(A) Fragmentation of nucleic acids by sonication. Samples 1 and 2 were input samples before sonication. Samples 3 and 4 were fragmented by sonication. Sample 5 was treated with RNase H.

(B) Dot blot of nucleic acids using the R-loop antibody. Numbers correspond to samples in Figure (A). An *in vitro* annealed R-loop sample was spotted as a positive control (indicated by +).

(C) Fragmentation of nucleic acids by restriction digest. Nucleic acids were fragmented by MboI and XbaI.

(D) Dot blot of nucleic acids using R-loop antibody. Samples are indicated in the figure. The bottom left sample corresponds to the sample in Figure (C)

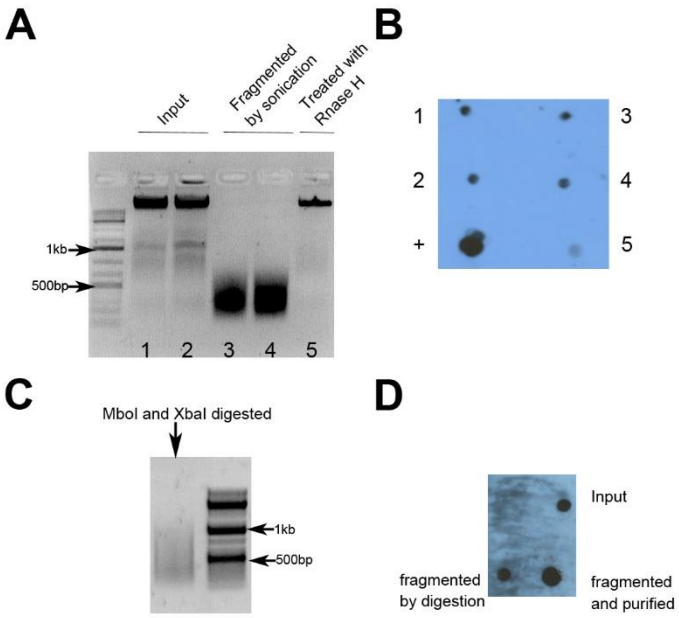

A
Input
Fragmented by sonication
Treated with Rnase H
1kb
500bp
1 2 3 4 5
B
1
2
+
3
4
5
C
MboI and XbaI digested
1kb
500bp
D
Input
fragmented by digestion
fragmented and purified

Figure C.3 Genome browser views of R-loop IP sequencing libraries.

(A) Genome browser views of R-loop sequencing from replicates 1 and 2. Gene models are shown in the bottom track. The background signal is lower in samples from replicate 2.

(B) Genome browser screen shots of Col mRNA-seq (replicate 1) in the top track and Col R-loop sequencing (replicate 2) in the middle track. Gene models are shown in the bottom track.

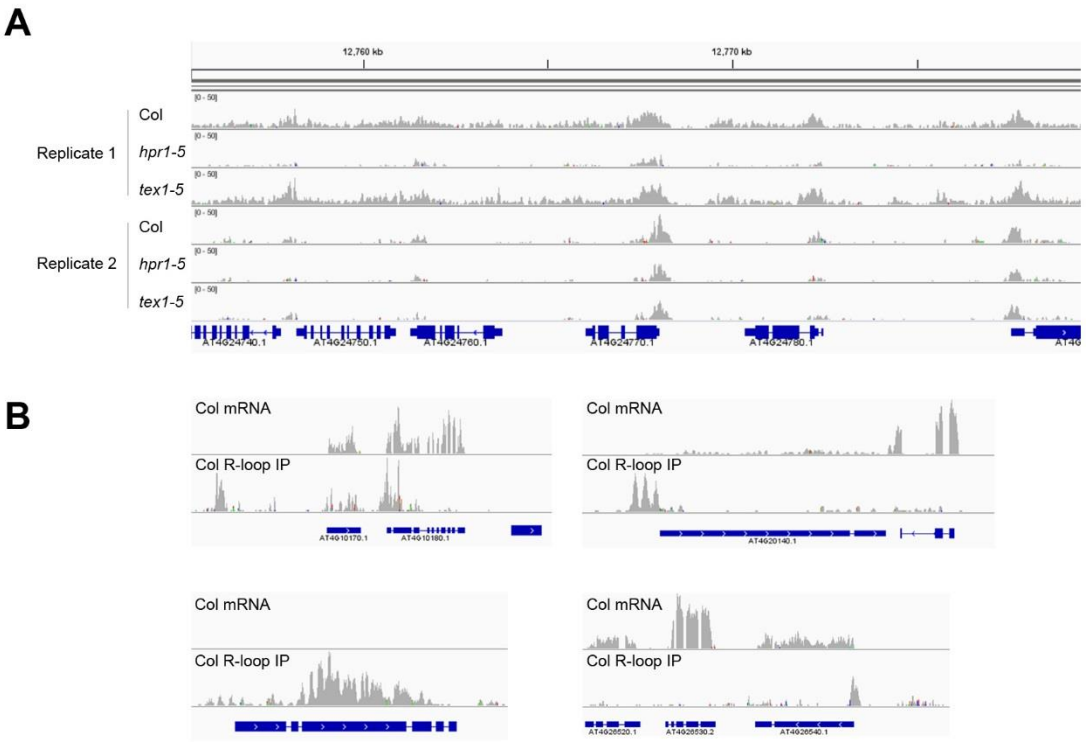

A
12,760 kb
12,770 kb
Replicate 1
Col
hpr1-5
tex1-5
Replicate 2
Col
hpr1-5
tex1-5
B
Col mRNA
Col R-loop IP
Col mRNA
Col R-loop IP
Col mRNA
Col R-loop IP
Col mRNA
Col R-loop IP

Table C.1 Read numbers for R-loop IP libraries

Raw reads that passed through illumna filters were mapped to the *Arabidopsis* Tair 10 genome with up to 2 nucleotide mismatches. Sibling reads and reads that were mapped to multiple genome loci were removed to get the processed reads.

	Replicate1			Replicate 2		
	Col_IP	*hpr1*_IP	*tex1*_IP	Col_IP	*hpr1*_IP	*tex1*_IP
Raw reads	25,279,384	15,709,067	33,266,128	5,761,880	6,074,387	9,948,726
Processed reads	9,435,027	1,941,851	11,219,656	2,368,398	1,336,384	1,571,964