Across the Lifespan of Long Non-Coding RNA Transcripts: Molecular Mechanisms of RNA Polymerase V Initiation and Transcript Degradation

Adriana N. Coke

Undergraduate Honors Thesis in Cellular Molecular Biology
University of Michigan
April 2020

Table of Contents

Abstract 2

Background 3

Introduction 3

RNA Polymerase V and its Role in Gene and Transposon Silencing 4

Unknowns about RNA Polymerase V: Initiation of Transcription 6

Unknowns about RNA Polymerase V: Transcript Degradation 8

Results 12

Locus-Specific Validation of non-RdDM Pol V Transcription 12

Effects of RNA Degradation Pathway Mutants on Pol V Transcript Levels 13

Discussion 17

Pervasive RNA Polymerase V Transcription and de novo RNA-directed DNA Methylation 17

Possible Mechanisms of RNA Polymerase V Transcript Degradation 19

Conclusion 22

Materials and Methods 23

Tables 25

Acknowledgements 27

References 28

Abstract

RNA Polymerase V (Pol V) is a plant-specific RNA Polymerase whose transcription is needed for silencing. In *Arabidopsis thaliana,* Pol V transcribes long non-coding RNA, which acts as a scaffold for the recruitment of siRNA and various protein components of RNA-directed DNA Methylation (RdDM)[1,2]. RdDM silences specific regions of the genome, mostly targeting transposable elements[3]. Despite the importance of Pol V transcripts for this process, very little is known about their initiation or degradation. In this study, preliminary tests are conducted which provide evidence for a new hypothesis whereby Pol V initiates transcription indiscriminately, therefore transcribing pervasively throughout the genome. Additionally, various potential routes for the degradation of Pol V transcripts are explored and ruled out, suggesting that the true mechanism remains unexplored at this time. Taken together, these results provide important insight into the regulation of RdDM locus-specificity, as well as present an increasingly interesting view of transcription by Pol V.

Background

Introduction

The central dogma of biology, first stated by Francis Crick in 1957, states that DNA contains genetic information in its primary structure, and is transcribed into RNA, which is then translated into protein[4]. There are many exceptions and caveats to this initial model, the realizations of which have enhanced our view of the fundamental processes governing life in countless ways. Some examples include viruses and transposons performing reverse transcription to multiply within host genomes[5], RNA-dependent RNA polymerases generating double-stranded RNA for RNA interference mechanisms[6], and various types of non-coding RNAs which are never translated into proteins, instead acting as gene regulatory molecules or even as catalytic ribozymes[7].

DNA-dependent RNA polymerases hold an integral role in the central dogma, as the enzymes that transcribe DNA into RNA. Different organisms have various types of RNA polymerases, each with their own specific purposes. In eukaryotes, RNA Polymerase I transcribes the majority of ribosomal RNAs, RNA Polymerase II transcribes the precursors to mRNA and many small RNAs, and RNA Polymerase III transcribes tRNAs, some ribosomal RNAs, and some types of small RNAs[8]. Many instances of transcription by RNA polymerases, as we may come to expect given the examples outlined above, extend far away from the “norm” of the central dogma. In fact, only RNA Polymerase II fulfills the role of transcribing RNA which is meant to be translated into protein, in some sense making the exceptions more of the rule.

One such exception to the role of transcription outlined in the central dogma is the presence of two unique RNA polymerases in plants, RNA Polymerases IV (Pol IV) and V (Pol V). These polymerases evolved specifically to participate in a mechanism of gene and transposon silencing called RNA-directed DNA Methylation (RdDM)[9]. Interestingly, this process utilizes transcription in order to cause transcriptional silencing. The RNA transcribed by Pol IV is used to create small double-stranded RNA fragments, called small interfering RNAs (siRNAs), which are analogous to small RNAs in other RNA interference processes[10]. On the other hand, Pol V

transcribes long non-coding RNA which, upon base pairing with complementary siRNA, acts as a scaffold for the binding of various RdDM protein components[11]. This scaffold function is interesting and relatively unique for RNA, especially given the fact that plants have evolved an entirely specific RNA polymerase to perform this single function exclusively[9].

Many fascinating questions remain unanswered about Pol V, including how and if it is initiated in a locus-specific manner, how long its transcripts are and how much this length varies, how exactly the polymerase evolved compared to other RNA polymerases, if and how its transcripts are processed during or after transcription, and how its transcripts are degraded. Answering questions such as these goes beyond helping us understand how RdDM functions in plants; the study of Pol V transcription presents an interesting opportunity to understand the true variety of the function of RNA and transcription beyond the "norm" presented by the central dogma.

RNA Polymerase V and its Role in Gene and Transposon Silencing

RNA-directed DNA methylation (RdDM) primarily uses RNA Polymerases IV and V to transcriptionally silence transposons in plants[1]. Transposons, also known as "jumping genes", are mobile genetic elements, meaning they are stretches of DNA which move themselves around their host genome. There are many different types of transposons, and they have been reported across the tree of life, from bacteria to animals to plants[8]. Additionally, transposons and repetitive DNA make up a large fraction of many genomes and are thought to contribute significantly to genome size and structure[8,12]. Transposons evolved due to selective pressures on their own survival in genomes, similar to viruses. While most transposons have a neutral or only small negative effect on their host genomes, some insertions can have large, often harmful effects on gene expression or genome stability[13]. In humans, for example, there are more than 120 transposon insertions that result in disease[14]. Transposons can have these effects by directly inserting themselves into genes or gene regulatory sequences[12], or even by causing major chromosomal rearrangements or breakages[15].

In order to combat these types of deleterious effects, many organisms have evolved mechanisms to block transposon movement; these mechanisms generally involve transcriptional

or post-transcriptional silencing of transposon expression[13]. Post-transcriptional silencing of transposons often involves splicing of transposon RNA by Argonaute proteins; for example, transposable element RNA is cleaved into piRNAs by PIWI or AUB in *Drosophila*, MILI and MIWI2 in mice, and PRG-1 in yeast[13]. Transcriptional silencing, on the other hand, prevents transcription of transposon DNA through the addition of heterochromatic marks to DNA or histones. Interestingly, the same mechanisms employed to transcriptionally silence transposons can also silence nearby genes, since both transposons and normal gene expression relies on transcription by RNA Polymerase II[8]. As a result, another potentially negative effect of transposon insertion can be the transcriptional silencing of genes near the insertion site[5]. Some examples of transcriptional gene silencing mechanisms which block transposon movement are the piRNA-PIWI system in *Drosophila*, which generates heterochromatin through H3K9 trimethylation[16], and the piRNA-MIWI2 system in mammals, which deposits *de novo* DNA methylation[17]. Additionally, RNA-directed DNA Methylation (RdDM) is responsible for the transcriptional silencing of transposons in plants.

RdDM transcriptionally silences transposons by depositing *de novo* methylation on DNA. Unlike similar mechanisms in other species, RdDM works via two plant-specific RNA polymerases, Pol IV and Pol V. Pol IV transcripts are processed into small interfering RNAs (siRNAs) that interact with ARGONAUTE 4 (AGO4)[18]. The siRNA-AGO4 complex goes on to interact with Pol V long non-coding RNA, which plays the role of a scaffold for protein binding at loci designated for transcriptional silencing[19]. The interaction between the siRNA-AGO4 complex and a Pol V transcript results in the recruitment of various other proteins, most notably DRM2, a *de novo* methyltransferase[11,20]. DRM2 deposits DNA methylation marks which prevent transcription by RNA Polymerase II, thus blocking gene expression and transposon mobility[20,21].

Pol V is a particularly important component of RdDM because its transcription, at least in part, is what dictates where on the genome RdDM will occur. When the siRNA-AGO4 complex associates with Pol V transcripts, the siRNA is thought to base-pair with the Pol V scaffold transcript[22], suggesting that in order for RdDM to occur, both Pol IV and Pol V must transcribe the same genomic region at around the same time. As a result, understanding the transcription initiation of these two RNA polymerases is likely paramount to understanding how RdDM

specifically targets transposons for silencing. Besides initiation, understanding how RNA involved in RdDM is degraded may also be important for understanding regulation of RdDM, since it is possible that degradation of RNA is what prevents RdDM from occurring at regions where it should not.

Unknowns about RNA Polymerase V: Initiation of Transcription

One important open question about Pol V is how its transcription is initiated. Understanding this is important to understanding how RdDM only targets specific regions of the genome. Without locus-specificity, the RdDM machinery could silence important genes that should remain active, or it could miss regions of the genome that ought to be silenced, such as transposable elements. Besides better understanding RdDM, understanding the differences between Pol V and other RNA polymerases may give us important insight into how all of these polymerases function. In this case, understanding how Pol V transcription causes transcriptional silencing at specific regions may help us to better understand how specific genes are targeted for active transcription by other RNA polymerases.

For most DNA-dependent RNA polymerases, specific DNA sequence motifs are what recruit the polymerase through various transcription factors, and transcription follows this recruitment[8]. However, there seems to be no DNA sequence motif which is conserved across areas where Pol V transcribes[19]. An alternative mechanism has been proposed which suggests that specific patterns of DNA methylation, not DNA sequence, may cause Pol V recruitment and thus transcription initiation at specific loci. This recruitment is thought to take place through the so-called DDR complex, made up of DRD1, DMS3, and RDM1[3]. Pol V transcription has been shown to depend on DRD1 and DMS3[23,24], and RDM1 is a DNA-binding protein[19]; this suggests the DDR complex may be involved in Pol V recruitment to DNA at specific loci. Two other proteins, SUVH2 and SUVH9, have been shown to interact with the DDR complex, as well as recognizing and binding cytosines which have been methylated by MET1, a maintenance methyltransferase[25]. Taken all together, it is possible that the DDR complex, along with SUVH2 and SUVH9, recognize specific DNA methylation patterns and then recruit Pol V to these genomic regions[19]. Since RdDM itself establishes *de novo* DNA methylation marks on DNA, this

hypothesis may suggest the possibility of a positive feedback loop whereby DNA methylation results in more recruitment of Pol V, which results in more DNA methylation, and so on[25].

There are several reasons why this hypothesis is not completely satisfying. For one, it begs the question about how the positive feedback loop of RdDM versus DNA methylation begins in the first place[26]. Additionally, evidence that DNA methylation pattern alone explains Pol V initiation has been unconvincing because no specific pattern has been shown to be necessary and sufficient for Pol V transcription[27]. In fact, unpublished results from our lab reveal that Pol V transcription occurs outside of methylated regions, suggesting that Pol V transcription does not require DNA methylation at all.

These unpublished results are from RNA immunoprecipitation and high throughput sequencing (RIP-seq) experiments performed with antibodies against NRPE1, the largest subunit of Pol V[28]. This experiment reveals, with very high sensitivity, the sequences of RNAs pulled down with Pol V. These RNA sequencing reads can then be mapped to the *Arabidopsis* genome, revealing all of the locations where Pol V transcription occurs. Upon mapping of the sequencing reads to the genome, this experiment revealed many new locations where Pol V transcribes. In fact, the results suggest that Pol V transcribes in low amounts throughout the majority of the genome, including at many places where RdDM does not occur and where DNA methylation is sparse or non-existent. From this data, we hypothesize that Pol V may initiate transcription indiscriminately, thus transcribing pervasively throughout the genome. This hypothesis explains the lack of a conserved sequence or pattern of DNA methylation to explain Pol V transcription because transcription would not depend on either of these.

So far, the hypothesis for pervasive Pol V transcription is only based on RIP-seq data. Without independent validation of these results, they remain speculative. One way to independently test this hypothesis is through a locus-specific confirmation that Pol V transcripts are present in low amounts at regions where RdDM clearly does not occur. The preliminary results of an independent locus-specific confirmation are presented here.

Unknowns about RNA Polymerase V: Transcript Degradation

Another major unknown about Pol V is how its transcripts are degraded. RNA transcripts must be degraded for cells to recycle ribonucleotides and to prevent unwanted effects from transcript accumulation[29]. In the case of mRNA, for example, accumulation of transcripts when RNA degradation is disrupted may be expected to result in excessive translation, but often we see the opposite: Accumulated transcripts are turned into small double-stranded RNAs through RNA interference pathways, thus causing unwanted post-transcriptional silencing of genes[22]. As a result, carefully regulated mRNA degradation pathways are extremely important to maintain the correct amounts of mRNA in cells. Since Pol V transcripts are extensively involved in transcriptional gene silencing through RdDM, their accumulation may cause RdDM to occur in excess, which would result in a waste of energy as well as the potential silencing of genes that should be left active. This potential function of Pol V transcript degradation as a negative regulator of RdDM is particularly interesting.

Many pathways for RNA degradation are specialized and carefully regulated to specifically degrade RNA depending on the needs of the cell[29]. Given this and the importance of Pol V transcripts for RdDM, it is likely that Pol V transcript degradation is specific and carefully regulated as well. In general, there are two major routes for RNA degradation in eukaryotes: 5’-3’ degradation is mostly conducted by XRN family exonucleases[30], and 3’-5’ degradation is mostly conducted by the exosome[31].

XRN family proteins are 5’-3’ exonucleases involved in the degradation and processing of various types of RNA, including non-coding RNAs such as rRNA and miRNA[30]. XRN family proteins are found broadly in eukaryotes; homologs include the yeast Rat1, *Drosophila* PACMAN, and human XRN2[30]. In *Arabidopsis*, there are three XRN family proteins: XRN2, XRN3, and XRN4. XRN2 and XRN3 are both located in the nucleus, whereas XRN4 is found in the cytoplasm[32]. All three exonucleases have specialized roles in RNA degradation, although enhanced phenotypes and substrate accumulation in double and triple mutants suggest they do have some overlaps in function: XRN2 and XRN3 have been shown to be involved in rRNA processing and degradation, XRN3 is involved in degradation of non-coding Pol II transcripts, XRN4 degrades specific mRNAs during the plant stress response, and all three XRNs are

involved in miRNA pathways by degrading either miRNAs themselves or the products of miRNA-mediated mRNA cleavage[32]. Besides RNA degradation, there is also evidence in yeast and *Arabidopsis* that XRN3 is involved in transcriptional termination of Pol II, possibly by degrading RNA co-transcriptionally and essentially pushing Pol II off of its DNA template[33,34].

XRN2 and XRN3 were selected as candidates to test for effects on Pol V transcripts. The 5'-3' directionality in exonuclease activity exhibited by these proteins is especially relevant to the question of Pol V transcript degradation because it would allow degradation or processing of Pol V transcripts to occur while they are still associated with Pol V. This degradation could even potentially occur co-transcriptionally, which would be especially interesting because this would suggest a very direct regulatory impact on RdDM. Since XRN4 is located in the cytoplasm and is primarily involved in stress-related mRNA degradation, and Pol V transcripts are not thought to be exported into the cytoplasm, it is less likely to be involved in Pol V transcript degradation. XRN2 and XRN3 are also good choices for their known widespread roles in non-coding RNA degradation and processing.

FIERY1 (FRY1) is a key regulator of XRN exonucleases in plants: XRN exonuclease activity is inhibited by 3'-phosphoadenosine 5'-phosphate (PAP), and FRY1 dephosphorylates PAP into AMP and Pi, thus promoting XRN exonuclease activity[35,23]. Taking advantage of this, the *fry1-6* mutant is extensively used in the study of XRN exonucleases, as it mimics a triple *XRN2/XRN3/XRN4* knockdown[32]. Another 5'-3' exonuclease which will be discussed next, DXO1, is also inhibited by PAP and thus positively regulated by FRY1[36]. As a result of its effects on various exonucleases of interest, FRY1 was also selected as a good candidate to study for any effects on Pol V transcript levels.

Another 5'-3' exonuclease in *Arabidopsis* which is suspected to be involved in the degradation and processing of RNA is DXO1[37]. In yeast, the DXO1 homolog Rai1 has been shown to be involved in rRNA processing and degradation[38]. Additionally, Rai1 and the human homolog DXO have been shown to modify incorrectly-capped 5' ends of mRNA transcripts to target them for degradation by Rai1/DXO itself or by other 5'-3' exonucleases[39]. These incorrectly capped mRNAs generally have either tri-phosphates or di-phosphates at their 5' ends, which other exonucleases such as XRN family proteins cannot degrade, and which standard

decapping proteins cannot remove. DXO1 and its homologs turn this unusual 5' end into a monophosphate, which can then be degraded by standard mechanisms[39].

DXO1 is also a strong candidate for Pol V transcript degradation for various reasons. Most obviously, it has 5'-3' exonuclease activity, which, as explained previously, is especially important for this study due to the potential for 5'-3' exonucleases to have a regulatory effect on RdDM. Additionally, the roles that DXO1 plays in other species such as yeast is promising and suggests that it may have a role in non-coding RNA degradation in *Arabidopsis* that is so far undiscovered. In fact, very little is known so far about the role of DXO1 exonuclease activity in *Arabidopsis*[37], and in a sense this is promising since it has not yet been investigated for its effects on Pol V. Lastly, very little is known about the 5' ends of Pol V transcripts[19], and DXO1 is known to remove unusual 5' ends, such as those of incorrectly capped mRNAs, as described above. Since it is possible that Pol V transcript 5' ends are also usual, this makes DXO1 a good candidate as well.

Yet another pathway for RNA degradation across eukaryotic species is the exosome. The exosome is a multisubunit protein complex involved in RNA processing and degradation through 3'-5' exonuclease activity[31]. The exosome is extensively involved in rRNA processing, regulation of mRNA turnover, "quality control" via degradation of many kinds of aberrant RNA transcripts, and even degradation of several non-coding RNAs[31,40]. The exosome core is a nine-subunit protein complex, conserved between archaea and eukarya[41]. In archaea, this complex has exonuclease activity itself, but in yeast and humans this activity is lost, and the exosome core instead acts as an essential structural scaffold for its associated exonucleases[41]. In yeast, these exosome-associated exonucleases are Rrp6 and Rrp44[41]. Interestingly, the plant exosome core seems to have maintained its exonuclease activity, in contrast with other eukaryotes[42]; nevertheless, the plant exosome core also has associated additional exonucleases analogous to the yeast proteins. In *Arabidopsis*, the exosome-associated exonucleases are the Rrp44-like protein RRP44A and the Rrp6-like proteins RRP6L1, 2, and 3[43,44]. The *Arabidopsis* RRP6L proteins have been shown to have specialized roles in 3' RNA degradation and processing, each with varying RNA substrates and subcellular locations[44]. In addition to associated exonucleases, the helicase cofactor Mtr4 is also required for efficient exosome activity

and substrate recognition in yeast[45]. In *Arabidopsis*, there are two helicase cofactors associated with the exosome, MTR4 and HEN2. Each cofactor seems to direct the exosome to different RNA substrates: MTR4 is associated only with rRNA degradation and processing, whereas HEN2 is associated with degradation of many non-coding RNAs and incorrectly spliced mRNAs[46], making it more promising to check for potential effects on Pol V transcripts than the former helicase.

The wide variety of RNAs degraded by the exosome and its associated exonucleases makes this complex a good candidate to check for effects on Pol V transcript levels. In fact, Zhang et. al. reported in 2014 that RRP6L1 may process Pol V RNA transcripts in order to increase their stability at RdDM loci[47]. This group found that RRP6L1 stabilizes Pol V transcript levels and increases occupancy of Pol V on DNA. However, it is unclear exactly how RRP6L1 may affect Pol V transcripts mechanistically, and the question of how Pol V transcripts are degraded remains unanswered from these results. RRP6L1 does not degrade Pol V transcripts; it seems to stabilize them.

Mutants of the core exosome and RRP44A are lethal[40,43], making this degradation pathway difficult to study as a whole. To get around this problem, individual mutants of exosome-associated exonucleases such as the RRP6L proteins can be used. Alternatively, RNAi knockdown lines can also be generated to study the effects of a reduction in core exosome activity[40,43]. For this study, null mutants of RRP6L2, RRP6L3, and HEN2 were selected as a first look at whether the exosome is involved in Pol V transcript degradation.

Results

Locus-Specific Validation of non-RdDM Pol V Transcription

As discussed previously, unpublished data from our lab suggests that Pol V transcribes in low amounts throughout the genome, leading us to propose a hypothesis whereby Pol V initiates transcription indiscriminately and transcribes pervasively. Because this data comes from only one type of experiment, namely RIP-seq, an independent test of this finding is needed to confirm or deny the presence of Pol V transcripts at various non-RdDM loci. To provide validation of this result, I performed reverse transcription and quantitative polymerase chain reaction (RT-qPCR). Reverse transcription from total RNA makes this assay fully independent from RIP-seq, which relies on immunoprecipitation of Pol V-associated RNAs and high throughput sequencing. Extensive experimental optimization resulted in the use of strand-specific RT primers to maximize the signal to noise ratio of this challenging experiment.

In order to measure Pol V transcript levels in regions where RdDM does not occur, I selected 23 specific loci using the lab's RIP-seq data. To ensure these regions had detectable amounts of Pol V transcription, all selected loci had a higher number of reads in wildtype (Col-0) than in *nrpe1*, a knockout mutant of the Pol V largest subunit NRPE1 in which no Pol V transcripts should be present[48,49]. To ensure RdDM was not occuring at these regions, loci were selected which had a relatively low number of reads in wildtype and no DNA methylation. The 23 selected loci were named PVS1-23 (primer sequences can be found in Table 1).

Of these 23 selected loci, 13 have been tested so far using strand-specific RT-qPCR. IGN5A and IGN22, two previously identified loci where RdDM is known to occur[2,27], were used as positive controls to ensure amplification of Pol V transcripts is successful with this method (Figure 1). All values were normalized to an *ACTIN2* reference gene and shown relative to wildtype amplification. Of the 13 tested PVS loci, 9 primers showed greater amplification in wildtype than in *nrpe1* (Figure 1). This may suggest a successful amplification of Pol V transcripts at these loci because any Pol V transcription in wildtype is expected to be eliminated in *nrpe1*. Successful detection of Pol V transcripts at all tested loci is not to be expected because of the low transcriptional activity of Pol V at non-RdDM regions. Following from this, the loci

where Pol V transcripts were not identified may have a signal to noise ratio which is too low, and this is a limitation in sensitivity of the method used for this experiment.

The results from this experiment so far are preliminary because only one replicate was conducted. Further replicates are necessary to confirm that the amplification detected in wildtype is statistically significant compared to *nrpe1*, thus confirming that it is in fact due to Pol V transcripts. However, the results so far do seem to support the hypothesis that Pol V transcribes in low amounts at identified non-RdDM regions because Pol V transcripts appear to be detected at many of the selected loci.

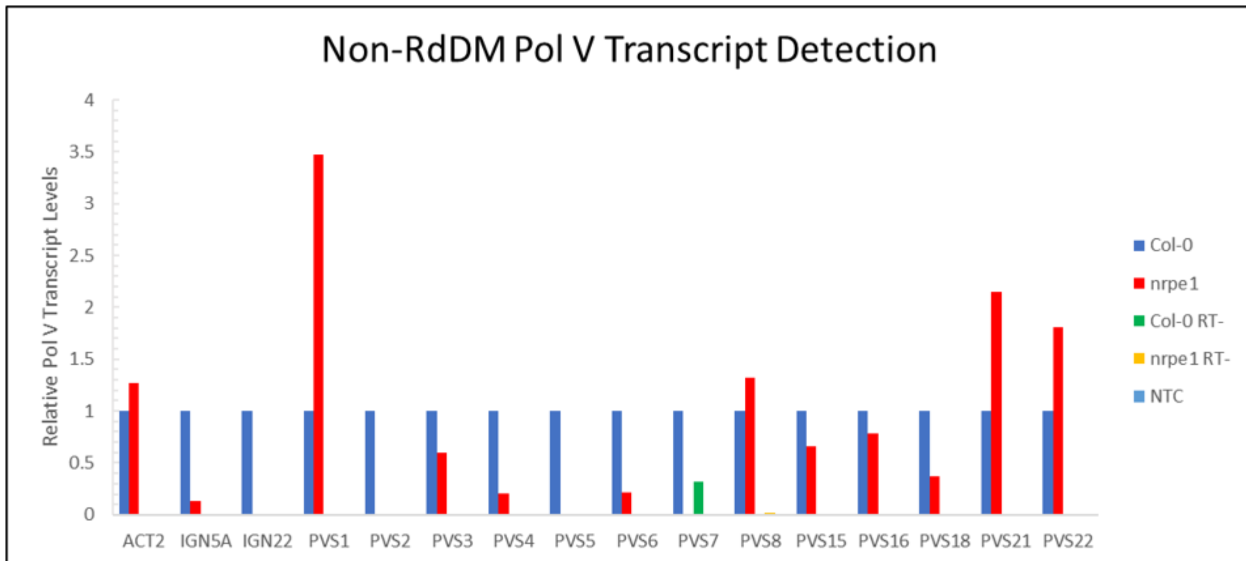


Figure 1: Locus-Specific Validation of non-RdDM Pol V Transcription. Strand-specific RT-qPCR was conducted in Col-0 and *nrpe1* at selected RdDM (IGN5A, IGN22) and non-RdDM (PVS) loci. Results are normalized against an *ACTIN2* (*ACT2*) reference gene and calculated relative to Col-0 wildtype. RT- negative controls, where the reverse transcriptase enzyme was omitted from the reaction, are shown for Col-0 and *nrpe1*. A no template control (NTC) was also added as a negative control for qPCR.

Effects of RNA Degradation Pathway Mutants on Pol V Transcript Levels

Besides transcription initiation, it remains a mystery how Pol V transcripts are degraded and whether this degradation occurs before, during, or only after RdDM. In order to determine which RNA exonucleases may affect Pol V transcript levels, I performed reverse transcription and quantitative polymerase chain reaction (RT-qPCR) in various mutant backgrounds of *Arabidopsis* (mutant and seed details can be found in Table 2). This experiment reveals whether

various proteins affect the total levels of Pol V transcripts at selected loci. An increase in transcript levels in a mutant compared to wildtype may suggest that the exonuclease which is absent in the mutant normally degrades Pol V transcripts. All of the selected mutants were tested at various loci where RdDM is known to occur; the IGN loci were previously identified[2,27,50], and the MD and POLV loci were newly identified for this study (primer sequences can be found in Table 1). For these experiments, Pol V transcript levels were normalized to a 25S rRNA reference gene, and calculated relative to wildtype (Col-0) transcript levels. To ensure that the transcripts amplified by this method are indeed generated by Pol V, *nrpe1,* a knockout mutant of the Pol V largest subunit, was again used as a negative control.

I first tested two XRN family mutants, *xrn2-3* and *xrn3-8*. While *xrn2-3* is a null allele[51], *xrn3-8* is an RNAi knockdown of *XRN3* since null alleles of this gene are lethal[34,51]. Both the *xrn2-3* and *xrn3-8* mutants did not affect the levels of Pol V transcripts at most tested loci (Figure 2A). Slight increases in Pol V transcript levels were detected at the IGN22 locus in *xrn2-3* and the POLV15 locus in *xrn3-8* (Figure 1A); however, these isolated minor changes are unlikely to be biologically relevant. I conclude that XRN2 and XRN3 are not the main exonucleases responsible for the degradation of Pol V transcripts.

I next tested the mutant *dxo1-2,* which contains a null allele of *DXO1*[37]. The *dxo1-2* mutant also did not affect Pol V transcript levels at most tested loci; however, slight increases were detected at IGN29 and POLV5 (Figure 2B). These small effects are also unlikely to be biologically relevant, leaving me to conclude that DXO1 is not the main exonuclease responsible for Pol V transcript degradation either.

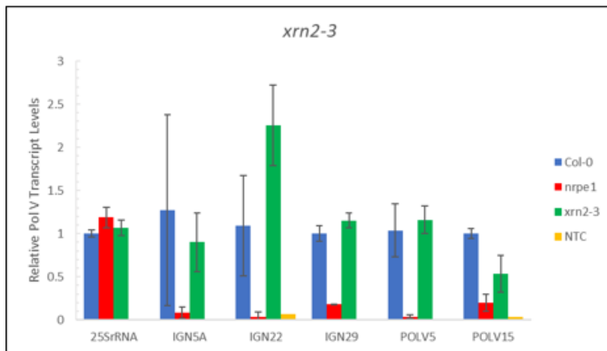
The effects of FRY1 were invested next, through the knockout mutant *fry1-6*[52]. Pol V transcript levels were unchanged at most loci in the *fry1-6* mutant, but some small increases were evident at IGN22 and POLV5 (Figure 2B). I conclude that FRY1 has no biologically significant effect on Pol V transcript levels. Since the *fry1-6* mutant shows decreased activity of XRN family proteins and DXO1[30,36], this result also supports the conclusion that neither XRN2, XRN3, nor DXO1 is the main exonucleases involved in the degradation of Pol V transcripts.

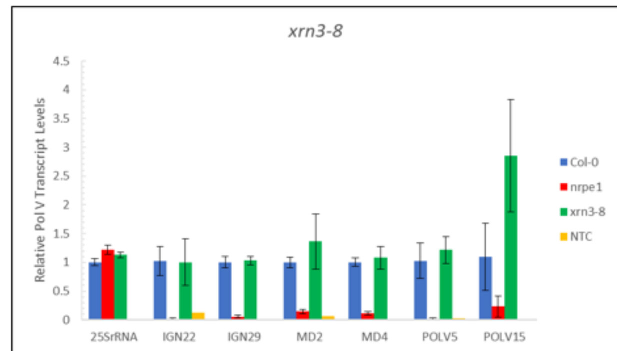
Lastly, I tested the effects of the exosome-associated proteins RRP6L2, RRP6L3, and HEN2. The knockout mutants *rrp6L2-3*, *rrp6L3*, and *hen2-4* were used[44,53,54]. Both *rrp6L2-3* and

rrp6L3 caused a small increase in Pol V transcripts at the loci IGN29 and POLV5, and a decrease in Pol V transcript levels at POLV2 (Figure 2C). No significant change was detected at other tested loci (Figure 2C), and the changes that were detected were minor, so it is still unlikely that these effects are biologically relevant. In the *hen2-4* mutant, Pol V transcript levels are unchanged at most loci, but small increases were detected at IGN22 and IGN29 (Figure 2C). I conclude that the exosome-associated proteins RRP6L2, RRP6L3, and HEN2 are not mainly responsible for the degradation of Pol V transcripts.

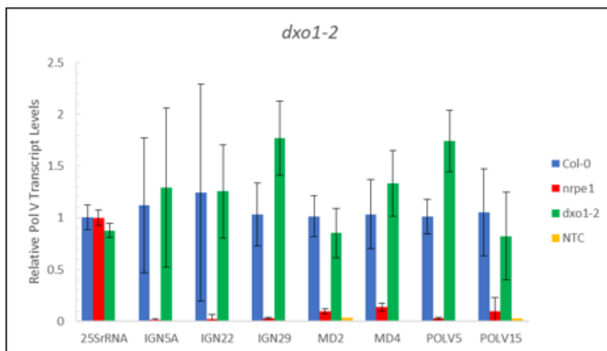
Figure 2: Effects of RNA Degradation Pathway Mutants on Pol V Transcript Levels. (Shown on the following page.) RT-qPCR was conducted in various mutants of *Arabidopsis* at select RdDM loci. Results are normalized against a 25S rRNA reference gene and calculated relative to Col-0. A no template control (NTC) was added as a negative control for qPCR. Pol V transcript amplification is evident by the strong reduction in transcript levels in *nrpe1* compared to Col-0 wildtype. A) Pol V transcript levels in XRN family exonuclease mutants *xrn2-3* and *xrn3-8* are shown next to Col-0 and *nrpe1*. B) Pol V transcript levels in DXO1 exonuclease mutant *dxo1-2* and FRY1 mutant *fry1-6* are shown next to Col-0 and *nrpe1*. C) Pol V transcript levels in exosome-associated protein mutants *rrp6L2-3*, *rrp6L3*, and *hen2-4* are shown next to Col-0 and *nrpe1*.

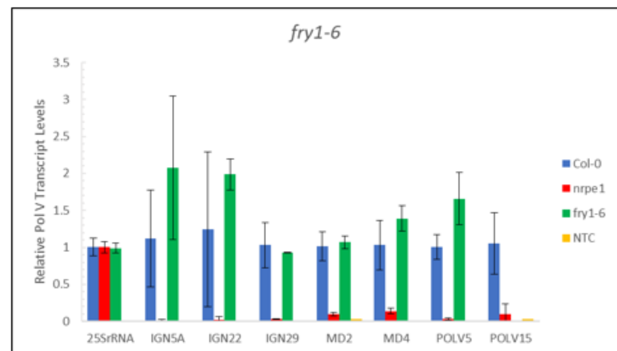
A



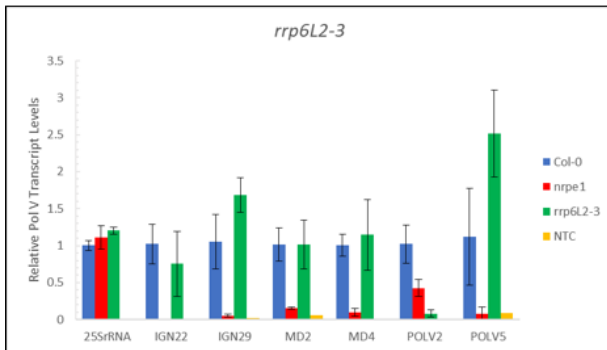


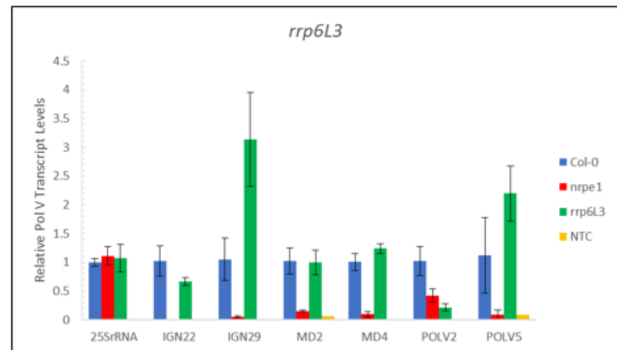
B

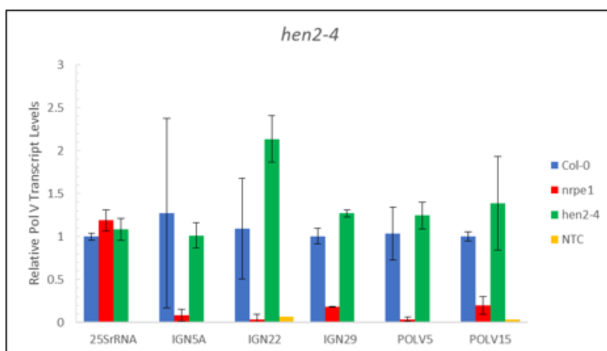




C







Discussion

Pervasive RNA Polymerase V Transcription and* de novo *RNA-directed DNA Methylation

The locus-specific validation of Pol V transcription at non-RdDM regions preliminarily supports the hypothesis that Pol V transcription is pervasive throughout the genome (Figure 1). This initial experiment was conducted without replicates in order to determine whether the method used would be sufficient to amplify Pol V transcripts at non-RdDM loci, where copy numbers are expected to be extremely low. As a result of this lack of replicates, these results are very preliminary at this time; however, if its results are confirmed upon replication, the conclusions from this experiment will be very important for our understanding of the initiation of RdDM at specific regions.

Previously, Pol V recruitment and initiation were thought to play a major role in determining where RdDM will occur[19]; however, if Pol V transcription occurs throughout the genome, not only at RdDM regions, then Pol V transcription itself cannot explain the initiation of RdDM. There must be another factor which causes RdDM to initiate at specific sites of Pol V transcription. One possibility is that siRNA determines where RdDM will occur. In this case, siRNA base pairing with complementary Pol V transcripts could cause the recruitment of other RdDM components, thus initiating transcriptional silencing in a locus-specific manner. The siRNAs involved in RdDM are generally produced from Pol IV transcripts[1,55]; however, it has been proposed that Pol II transcripts from active transposons can also be turned into siRNA for use in the RdDM pathway[19,21]. It is therefore possible that active transposons are targeted for *de novo* transcriptional silencing by RdDM through siRNA association with Pol V transcripts.

Since our unpublished data reveals that Pol V transcript levels are significantly higher at regions where RdDM occurs than at regions where it does not, there is also the question of how transcript levels vary between these regions. One possibility for this is the aforementioned positive feedback loop between DNA methylation and recruitment of Pol V by SUVH2, SUVH9, and the DDR complex[3,25]. In this case, once siRNA base pairing with Pol V transcripts initiates *de novo* DNA methylation at a specific region through RdDM, then more Pol V is recruited to

the newly methylated region, resulting in more RdDM, and so on. This feedback loop hypothesis, while compelling, requires further study if it is to be confirmed and understood.

Besides siRNA alone, another possibility for how RdDM is initiated at some Pol V transcribed regions but not others could be through regulation of Pol V transcripts by RNA degradation pathways. These pathways could potentially regulate, in a locus-specific manner, the amount of Pol V transcript or the turnover time before transcripts are degraded. In this way, Pol V transcripts may be degraded at regions where RdDM should not occur, and left to multiply at regions where RdDM should occur. Regulation of this degradation could occur in either direction; one possibility is that degradation occurs by default, but it is inhibited at specific regions where RdDM needs to occur. At this point, the increased amount of Pol V transcripts could be a determining factor of RdDM initiation by making it more likely that siRNA and other RdDM components will find and bind to Pol V scaffold transcripts. This hypothesis would also explain why more Pol V transcripts are found at RdDM regions. No direct evidence has been found in favor of this hypothesis as of now; however, very little is known about Pol V transcript degradation at this time, and this possibility further supports the importance of research to determine how Pol V transcripts are degraded and how this degradation may affect RdDM.

While its function cannot yet be fully understood, pervasive transcription by Pol V may serve the role of making unsilenced regions across the genome competent for transcriptional silencing. The availability of Pol V transcripts at these unsilenced regions may allow the rapid initiation of RdDM if needed, whether by siRNA production and base pairing, inhibition of Pol V transcript degradation, or another mechanism. This presents a potential novel function of pervasive transcription, which is often supposed to be little more than "transcriptional noise" caused by flexible transcription initiation[56].

Once the preliminary experiments explained here are validated through subsequent replicates, several next steps ought to be taken to better understand the pervasive transcription of Pol V. These could include *in vitro* studies to confirm sequence and methylation-independent transcription initiation of Pol V, as well as structural analysis to understand how Pol V initiates transcription indiscriminately. Additionally, various experiments ought to be conducted to determine whether siRNA, regulation of transcript degradation, or some other mechanism is

responsible for the initiation of RdDM and the increased levels of Pol V transcripts seen at RdDM regions.

Possible Mechanisms of RNA Polymerase V Transcript Degradation

The results from the RNA degradation mutant analysis indicate that XRN family proteins, DXO1, and selected exosome-associated proteins are not solely responsible for the degradation of Pol V transcripts (Figure 2). This result, though negative, is important for furthering our understanding of Pol V transcript degradation because it rules out several potential options for the proteins involved in this degradation. While some limited locus-specific effects were observed in all mutants tested, these effects are not consistent or strong enough to make any conclusions about their biological significance. Additionally, since this experiment measures changes in total transcript level, it is impossible to say whether these small effects are due to RNA degradation specifically or some other regulation of RNA levels, direct or indirect. It is also important to note that, in order to completely rule out involvement of the tested proteins, a genome-wide analysis of changes in Pol V transcript levels ought to be conducted because a locus-specific validation such as the one conducted here may not reveal certain genome-wide trends; for instance, changes in Pol V transcript levels at non-RdDM regions has not yet been investigated for these mutants.

Additionally, it is also possible that certain proteins investigated have redundant roles in Pol V transcript degradation, so that single mutants of these proteins do not reveal their true involvement. Redundancy is a common theme amongst RNA degradation pathways[29]. One example of this occurrence is between XRN2 and XRN3, where a double mutant of these two exonucleases can reveal phenotypes seen in neither single mutant, presumably because the two have overlapping functions, so that in a single mutant the remaining exonuclease makes up for the absence of the other[52]. The use of the *fry1-6* mutant does provide some evidence that hidden redundancy is not a problem in this case for the XRN proteins and DXO1, since it functions as a partial knockdown of XRN2, XRN3, and DXO1[30,36]. Still, to rule out this hypothesis completely, various combinations of mutants would need to be generated and tested.

Continuing the search for how Pol V transcripts are degraded may be useful for understanding how RdDM is regulated. Besides identifying the proteins involved in this process, it will be very important to determine whether degradation occurs before, during, or only after RdDM, as this will help us to understand the potential role of transcript degradation in regulation. Additionally, it will become important to determine whether or not transcript degradation varies between RdDM and non-RdDM loci. As was explained previously, it is possible that Pol V transcript degradation is what determines where RdDM is initiated, and if this is the case, degradation would vary considerably between regions where RdDM does or not occur. Even if this hypothesis is not correct, another connection between the studies of Pol V initiation and transcript degradation is simply a practical one: If a mutant is identified in which Pol V transcript levels are increased at all the regions where Pol V transcribes, then this mutant would be extremely useful for revealing genome-wide Pol V transcription and for further studying this phenomenon. In studies of the exosome, for instance, many new non-coding RNAs, previously hidden by rapid degradation, were revealed after a genome-wide analysis of exosome mutants was conducted in *Arabidopsis*[40]. Alternatively, Pol V transcript degradation may be important for making sure RdDM does not occur in excess by simply limiting the lifespan of Pol V transcripts, similar to how specific degradation of miRNAs has been suggested to regulate their activity levels[57].

Besides gaining a better understanding of RdDM, research into Pol V transcript degradation may also be useful for understanding how the involved RNA degradation pathway is regulated and specific to certain RNA molecules. This is especially interesting in the case of Pol V because, if our hypothesis for indiscriminate initiation and pervasive transcription is correct, then Pol V transcripts do not have specific sequence motifs, so it will be very interesting to see how they may be specifically recognized by degradation machinery. In some cases, specific sequences are used to determine the turnover rate of target RNA molecules; for example, sequence elements of mRNA 3'-UTRs are often recognized by RNA-binding proteins which stabilize the transcripts or target them for degradation[58]. Additionally, as expressed previously, little is known about Pol V transcript 5' ends or how they are processed[19], and characteristics of the ends of RNAs are often used to target them for degradation as well[29]. An obvious example is

that decapping enzymes recognize the 5'-methylguanosine caps specific to mRNAs, and the removal of these caps targets mRNAs for degradation by 5’-3’ exonucleases such as XRN family proteins[39]. As a result of these unknowns about Pol V transcripts, much remains to be explained in terms of how Pol V transcripts may be specifically targeted for degradation. As of now, it is still a large open question in the field how RNA degradation pathways target specific transcripts in general, especially in the case of non-coding RNAs, which often do not have the same characteristic ends of mRNAs[29]. Understanding how this process works in the case of Pol V transcripts may provide important insight which could be helpful for understanding the specificity of RNA degradation more generally.

The next step that ought to be taken in the search for the pathway which degrades Pol V transcripts is to investigate the effects of various other RNA degradation proteins which were left out of this study so far. Proteins that ought to be investigated next include subunits of the core exosome and the exosome-associated proteins RRP44A and MTR4. The exosome core subunits and RRP44A were left out of this study so far because of the lethality of their null alleles[40,43], and MTR4 was deemed less likely to be involved in non-coding RNA degradation than other exosome-associated proteins, but the involvement of the exosome cannot be ruled out just yet. In fact, while the effects of RRP6L2-3 and RRP6L3 reported here were determined to be too limited to have biological significance, the mutants of these two proteins did affect the same RdDM loci in a limited manner (Figure 2C), which may be a subtle indication that the exosome has some effect on Pol V transcript levels. More tests are needed to investigate this idea. To accomplish this, RNAi knockdown lines of RRP44A and subunits of the exosome core have been generated in *Arabidopsis* which could be useful[40,43]. The use of these knockdown lines would more definitely rule out involvement of the exosome than solely studying the effects of exosome-associated proteins, and it is therefore a worthy idea for an immediate next step.

In addition to further study of the exosome and its associated proteins, the effects of endonucleases on Pol V transcript levels ought to be carefully studied as well. Endonucleases are often involved in RNA processing[59], and how Pol V transcripts may be processed is, like their degradation, also largely unknown[19]. Endonucleases are also often needed for RNA degradation[29]; for example, Argonaute proteins are endonucleases often involved in small

RNA-mediated decay of mRNAs[60]. AGO4, the Argonaute protein primarily involved in RdDM[18], has even been proposed to cut Pol V transcripts, possibly as a part of the RdDM mechanism[61]. Still, it is unknown whether AGO4 slicing, if it does occur, is in any way related to Pol V transcript degradation, so further studies are warranted.

Finally, an RNA degradation pathway specific to Pol V transcript degradation may exist which has not yet been well studied. A genetic screen may be helpful in identifying novel proteins which affect Pol V transcript levels, whether directly or indirectly.

Conclusion

The research on Pol V discussed here may suggest some unexpected functions for long-known processes. The sequence of Pol V transcripts is obviously important for RdDM, since complementary siRNA base pairs with Pol V transcripts, and this may be what determines the locus-specificity of the mechanism[19]; however, sequence may not be important at all for Pol V initiation. If it is the case that Pol V initiates transcription indiscriminately, then the sequences of its transcripts are neither specific nor conserved. In this case, the lack of a conserved sequence may actually be useful for Pol V by allowing it to transcribe a greater portion of the genome, surveilling for regions in need of silencing. This surveillance function of Pol V presents a novel role for an RNA polymerase, and for the long non-coding RNA it produces, which would exist without any sequence specificity. Additionally, while they concern opposite ends of the lifespan of Pol V transcripts, transcript initiation and RNA degradation may be connected in various ways. The degradation of Pol V transcripts may play a regulatory role in RdDM, possibly by determining its locus-specificity or by regulating its intensity more generally. In any case, the possibility of a connection between transcriptional gene silencing, pervasive transcription of non-coding RNA, and RNA degradation is fascinating and important. Since mechanisms of transcriptional gene silencing and RNA degradation pathways are widely conserved between kingdoms, insights gained from the research conducted on Pol V, including that about pervasive transcription and the potential regulatory role of RNA degradation on transcriptional gene silencing, may turn out to be widely applicable outside of plants as well.

Materials and Methods

Plant Material

Arabidopsis thaliana for these experiments was grown in soil under long day (16h light, 8h dark) conditions. The following mutants were used in this study: *nrpe1* (SALK_029919)[48], *xrn2-3* (SALK_114258)[51,62], *xrn3-8* (RNAi knockdown)[34,51], *dxo1-2* (SALK_032903)[37,62], *fry1-6* (SALK_020882)[52], *rrp6L2-3* (SALK_011429)[53], *rrp6L3* (SALK_122492)[44,62], *hen2-4* (SALK_091606C)[54]. Additional details about mutants and seeds can be found in Table 2.

RNA Isolation

RNA was isolated from 2.5-3 week old *Arabidopsis thaliana* leaf tissue. Tissue was frozen in liquid nitrogen and then RNA was extracted using TRI Reagent (Sigma-Aldrich) according to the manufacturer's protocol.

Random Primer Reverse Transcription (RT)

10ug RNA was treated with Turbo DNase for 30min at 25°C. The following reagents were included per reaction: 1.5uL Turbo DNase, 10x Turbo DNase Buffer, 0.6uL Ribolock, and H_2O to 15uL. 3uL of 25mM EDTA was added after 30min to stop the reaction, and samples were kept at 65°C for 10min. Next, samples were split into experimental samples and negative reverse transcription controls. 0.4ug Random Primers, 1uL 10mM dNTP mix, and 2.6uL H_2O were added per reaction. Samples were left at 65°C for another 5min, after which they were moved immediately to ice.

Reverse transcription was then performed using SuperScript III. Per reaction, the following reagents were added to the above mixtures: 4uL 5x First Strand Buffer, 1uL 0.1M DTT, 1uL Ribolock, and 1uL Superscript III for experimental samples or 1uL H_2O for negative controls. Samples were kept at 25°C for 5min for annealing, 50°C for 60min for elongation, and 70°C for 15min for denaturation. cDNA was then diluted by ⅕, and 1uL was used per qPCR reaction.

Random primer reverse transcription was performed with three biological replicates per genotype.

Strand-Specific Reverse Transcription (RT)

4ug RNA was treated with Turbo DNase for 30min at 25°C. The following reagents were included per reaction: 1.5uL Turbo DNase, 10x Turbo DNase Buffer, 0.6uL Ribolock, and H_2O to 15uL. 3uL of 25mM EDTA was added after 30min to stop the reaction, and samples were kept at 65°C for 10min. Next, 0.4uL of 50mM strand-specific primer, 2uL 10mM dNTP mix, and 5.6uL H_2O were added per reaction. Samples were left at 65°C for another 5min, after which they were moved immediately to ice. At this point, samples were split into experimental and negative controls for reverse transcription.

Reverse transcription was then performed using SuperScript III. Per reaction, the following reagents were added to the above mixtures, same as the above conditions: 4uL 5x First Strand Buffer, 1uL 0.1M DTT, 1uL Ribolock, and 1uL Superscript III for experimental samples or 1uL H_2O for negative controls. Samples were kept at 55°C for 30min for elongation and then 70°C for 15min for denaturation. cDNA was not diluted, and 6uL was used per qPCR reaction.

Strand-specific reverse transcription was performed so far in one biological replicate.

Quantitative Polymerase Chain Reaction (qPCR)

qPCR was performed using Platinum Taq Polymerase and corresponding buffer from Thermo-Fisher. Per 25ul reaction, the following reagents were included: 2.5uL 10x Taq Buffer, 1.2uL 50mM $MgCl_2$, 0.5uL 10mM dNTP mix, 0.25uL 1:400 diluted SYBR Green, 0.2uL 25uM Primer mix, 0.1uL Platinum Taq Polymerase, cDNA as indicated above, and H_2O to 25uL. The following cycling conditions were used: initial denaturation at 95°C for 4min; then 40 cycles of denaturation at 95°C for 15sec, annealing at 60°C for 30sec, and elongation at 72°C for 45sec. Melt curve analysis was then conducted in 0.5°C increments, from 65°C to 95°C. qPCR was performed with two or three technical replicates per sample.

Tables

Table 1: qPCR Primer Sequences

Locus	Strand	Primer Sequence
ACTIN2	Forward	GAGAGATTCAGATGCCCAGAAGTC
	Reverse	TGGATTCCAGCAGCTTCCA
25S_rRNA	Forward	TTTCGCGTATCGGCATGATC
	Reverse	AGCCGTTCGTTTGCATGTTC
IGN5A	Forward	TACATGAAGAAAGCCCAAACC
	Reverse	ATTGGGCCGAATAACAGC
IGN22	Forward	GCCCCACAAATGGAAGAGGTCT
	Reverse	TCGTGACCGGAATAATTAAATGG
IGN29	Forward	CGTTTGTTTATGTAGGGCGAAAG
	Reverse	TAAAACTTTTCCCGCCAACCA
MD2	Forward	ATCCGTTTTCGTAATAGCAAGGCCCATCCC
	Reverse	GATCAGGTTGGATCGGGTCGGATAAAGTAG
MD4	Forward	AGGAGCGGTGGACCTGAATCTTTGAGAATG
	Reverse	TTTGAAACAATCGAGGTGCATATGTAGATCTCC
POLV2	Forward	GCCTCATGCCCAAAACACACA
	Reverse	TGAAGTTTGTGGAGACGACACG
POLV5	Forward	AGGATACAAGTGTCGAACTCCGTT
	Reverse	TCCAAGCAGTCTTCTCTTCTTCGT
POLV15	Forward	GCCTCATGCCCAAAACACACA
	Reverse	TGAAGTTTGTGGAGACGACACG
PVS1	Forward	TGTGTGAAAATTGACGGTATGAAGATG
	Reverse	CCACAAGATATGATTTTCGCAGATGTATAAAA
PVS2	Forward	CACGAAGAAGAGAAGACTGCTTGGA
	Reverse	AAGAGATAGATGAGTGAAATTTTGGAGATAGATTG
PVS3	Forward	TTCATCCTTTTTCGTCTTTCAGATATTCA
	Reverse	CGGGTTTGGTGTTGCTCCTATTT
PVS4	Forward	GACACCGCTGCAAAAGAACGAC
	Reverse	AGAGATGCTCTTAAACGTTGATTGAAGTAGAG

PVS5	Forward	GCGACGAAAGGAGAGAGGATATGTTT
	Reverse	AAGTACCAATAGAGACAGAGTTGTGATGTTCC
PVS6	Forward	CGTCGAAAAGCATAACGAGTTCAAAG
	Reverse	CCGGCCCAATAACTTAAAACCAATAGA
PVS7	Forward	CTATTCAAAATGAAACCTGCGAAAACA
	Reverse	CAAAAACACTTCTCATTCAAAAACACAATAACA
PVS8	Forward	CCACACTTTCAGAGACGGACAGGA
	Reverse	TTACCAAAAGCTAATAAATCGTAAGCCACA
PVS15	Forward	CAAGCCTCTCAAATTATCCATCTCAATTCT
	Reverse	GACGTGGCTGACTCTTATCTGTTGTG
PVS16	Forward	AAGCCCTCACACTATTCAAAGACCAA
	Reverse	TGATTGAGACAAATACAAACGGAATAGTAAGG
PVS18	Forward	CGGAACATCATCACACACAGCA
	Reverse	GGCACGTCTAGGGTACGATCCA
PVS21	Forward	CCATTGTAACGAAGGCCCACTTTCT
	Reverse	CAAGGAAATAAAAACGACAAAATGAAAAATC
PVS22	Forward	AAAAGAAAGCTGAGTTAGCAACAAGTGC
	Reverse	GGGTGGGTATGCGACAAAAAGA

Table 2:* Arabidopsis thaliana *Mutant and Seed Details

Name	Type	Line	Seeds Retrieved From
nrpe1 (nrpdb1-11)	Null allele	SALK_029919	Arabidopsis Biological Resource Center
xrn2-3	Null allele	SALK_114258	Prof. Joanna Kufel (University of Warsaw)
xrn3-8	RNAi knockdown	N/A	Prof. Joanna Kufel (University of Warsaw)
dxo1-2	Null allele	SALK_032903	Prof. Joanna Kufel (University of Warsaw)
fry1-6	Null allele	SALK_020882	Prof. Joanna Kufel (University of Warsaw)
rrp6L2-3	Null allele	SALK_011429	Arabidopsis Biological Resource Center
rrp6L3	Null allele	SALK_122492	Arabidopsis Biological Resource Center
hen2-4	Null allele	SALK_091606C	Prof. Joanna Kufel (University of Warsaw)

Acknowledgements

I would like to offer my sincere thanks to Professor Andrzej Wierzbicki for supporting my research over the past two years. His guidance has undoubtedly improved the way I read, write, and think about science, and my experience in his lab has been the most rewarding aspect of my undergraduate career.

This opportunity would not have been possible without the other members of Wierzbicki Lab as well. I especially wish to thank Masayuki Tsuzuki, Ph.D., who was a great mentor to me during my first year in the lab, and who played an instrumental role in teaching me how to conduct science independently. I also extend my gratitude to the other past and present members of the lab, for their continuous support and for cultivating a fantastic work environment: M. Hafiz Rothi, Miguel Palomar, Ph.D., Shriya Sethuraman, Jan Kuciński, Ph.D., Sho Fuji, Ph.D., Zoë Anderson, and Angelica Previero.

I also wish to thank Professor Erik Nielsen and Professor Laura Buttitta for generously agreeing to read and evaluate this thesis. Thanks also to Professor Joanna Kufel for providing seeds for many of the mutants used in this study.

Finally, I would like to acknowledge the members of my 2019 Honors Summer Fellowship cohort, all of whom have given me extremely helpful practice with discussing my research, unwavering support, and great friendship.

References

1. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408 (2014).
2. Wierzbicki, A. T., Haag, J. R. & Pikaard, C. S. Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell* **135**, 635–648 (2008).
3. Zhong, X. *et al.* DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* **19**, 870–875 (2012).
4. Crick, F. On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–163 (1958).
5. Hughes, S. H. Reverse Transcription of Retroviruses and LTR Retrotransposons. 40 (2019).
6. Mello, C. C. & Conte, D. Revealing the world of RNA interference. *Nature* **431**, 338–342 (2004).
7. Jimenez, R. M., Polanco, J. A. & Lupták, A. Chemistry and Biology of Self-Cleaving Ribozymes. *Trends Biochem. Sci.* **40**, 648–661 (2015).
8. Griffiths, A., Wessler, S., Carroll, S. & Doebley, J. *Introduction to Genetic Analysis*. (W. H. Freeman and Company, 2015).
9. Tucker, S. L., Reece, J., Ream, T. S. & Pikaard, C. S. Evolutionary History of Plant Multisubunit RNA Polymerases IV and V: Subunit Origins via Genome-Wide and Segmental Gene Duplications, Retrotransposition, and Lineage-Specific Subfunctionalization. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 285–297 (2010).
10. Kuo, H. Y., Jacobsen, E. L., Long, Y., Chen, X. & Zhai, J. Characteristics and processing of Pol IV-dependent transcripts in Arabidopsis. *J. Genet. Genomics* **44**, 3–6 (2017).
11. Böhmdorfer, G. *et al.* RNA-directed DNA methylation requires stepwise binding of silencing factors to long non-coding RNA. *Plant J.* **79**, 181–191 (2014).
12. Feschotte, C. & Pritham, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu. Rev. Genet.* **41**, 331–368 (2007).
13. Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **20**, 417–431 (2019).
14. Hancks, D. C. & Kazazian, H. H. Roles for retrotransposon insertions in human disease. *Mob. DNA* **7**, 9 (2016).
15. Huang, J. T. & Dooner, H. K. Macrotransposition and Other Complex Chromosomal Restructuring in Maize by Closely Linked Transposons in Direct Orientation. *Plant Cell* **20**, 2019 (2008).
16. Russell, S., Stalker, L. & LaMarre, J. PIWIs, piRNAs and Retrotransposons: Complex battles during reprogramming in gametes and early embryos. *Reprod. Domest. Anim.* **52**, 28–38 (2017).
17. Molaro, A. & Malik, H. S. Hide and seek: how chromatin-based pathways silence retroelements in the mammalian germline. *Curr. Opin. Genet. Dev.* **37**, 51–58 (2016).

18. Qi, Y. *et al.* Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**, 1008–1012 (2006).
19. Wendte, J. M. & Pikaard, C. S. The RNAs of RNA-directed DNA methylation. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1860**, 140–148 (2017).
20. Cao, X. & Jacobsen, S. E. Role of the Arabidopsis DRM Methyltransferases in De Novo DNA Methylation and Gene Silencing. *Curr. Biol.* **12**, 1138–1144 (2002).
21. Sigman, M. J. & Slotkin, R. K. The First Rule of Plant Transposable Element Silencing: Location, Location, Location. *Plant Cell* **28**, 304 (2016).
22. Wierzbicki, A. T., Ream, T. S., Haag, J. R. & Pikaard, C. S. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* **41**, 630–634 (2009).
23. Kanno, T. *et al.* A SNF2-like protein facilitates dynamic control of DNA methylation. *EMBO Rep.* **6**, 649–655 (2005).
24. Kanno, T. *et al.* A structural-maintenance-of-chromosomes hinge domain–containing protein is required for RNA-directed DNA methylation. *Nat. Genet.* **40**, 670–675 (2008).
25. Johnson, L. M. *et al.* SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**, 124–128 (2014).
26. Cuerda-Gil, D. & Slotkin, R. K. Non-canonical RNA-directed DNA methylation. *Nat. Plants* **2**, 16163 (2016).
27. Wierzbicki, A. T. *et al.* Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. *Genes Dev.* **26**, 1825–1836 (2012).
28. Pontier, D. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev.* **19**, 2030–2040 (2005).
29. Houseley, J. & Tollervey, D. The Many Pathways of RNA Degradation. *Cell* **136**, 763–776 (2009).
30. Nagarajan, V. K., Jones, C. I., Newbury, S. F. & Green, P. J. XRN 5′→3′ exoribonucleases: Structure, mechanisms and functions. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1829**, 590–603 (2013).
31. Schmid, M. & Jensen, T. H. The exosome: a multipurpose RNA-decay machine. *Trends Biochem. Sci.* **33**, 501–510 (2008).
32. Kurihara, Y. Activity and roles of Arabidopsis thaliana XRN family exoribonucleases in noncoding RNA pathways. *J. Plant Res.* **130**, 25–31 (2017).
33. Chalamcharla, V. R., Folco, H. D., Dhakshnamoorthy, J. & Grewal, S. I. S. Conserved factor Dhp1/Rat1/Xrn2 triggers premature transcription termination and nucleates heterochromatin to promote gene silencing. *Proc. Natl. Acad. Sci.* 201522127 (2015) doi:10.1073/pnas.1522127112.
34. Krzyszton, M. *et al.* Defective XRN3-mediated transcription termination in Arabidopsis affects the expression of protein-coding genes. *Plant J.* **93**, 1017–1031 (2018).

35. Kurihara, Y. *et al.* Surveillance of 3′ Noncoding Transcripts Requires FIERY1 and XRN3 in Arabidopsis. *G3 GenesGenomesGenetics* **2**, 487 (2012).
36. Yun, J.-S. *et al.* Molecular mechanism for the inhibition of DXO by adenosine 3′,5′-bisphosphate. *Biochem. Biophys. Res. Commun.* **504**, 89–95 (2018).
37. Kwasnik, A. *et al.* Arabidopsis DXO1 links RNA turnover and chloroplast function independently of its enzymatic activity. 14.
38. Fang, F. Rat1p and Rai1p function with the nuclear exosome in the processing and degradation of rRNA precursors. *RNA* **11**, 1571–1578 (2005).
39. Zhai, L. & Xiang, S. mRNA quality control at the 5' end. *J. Zhejiang Univ. Sci. B* **15**, 438–443 (2014).
40. Chekanova, J. A. *et al.* Genome-Wide High-Resolution Mapping of Exosome Substrates Reveals Hidden Features in the Arabidopsis Transcriptome. *Cell* **131**, 1340–1353 (2007).
41. Tomecki, R., Drazkowska, K. & Dziembowski, A. Mechanisms of RNA Degradation by the Eukaryotic Exosome. *ChemBioChem* **11**, 938–945 (2010).
42. Sikorska, N., Zuber, H., Gobert, A., Lange, H. & Gagliardi, D. RNA degradation by the plant RNA exosome involves both phosphorolytic and hydrolytic activities. *Nat. Commun.* **8**, 2162 (2017).
43. Kumakura, N., Otsuki, H., Tsuzuki, M., Takeda, A. & Watanabe, Y. Arabidopsis AtRRP44A Is the Functional Homolog of Rrp44/Dis3, an Exosome Component, Is Essential for Viability and Is Required for RNA Processing and Degradation. *PLoS ONE* **8**, e79219 (2013).
44. Lange, H. *et al.* Degradation of a Polyadenylated rRNA Maturation By-Product Involves One of the Three RRP6-Like Proteins in Arabidopsis thaliana. *Mol. Cell. Biol.* **28**, 3038–3044 (2008).
45. Schmid, M. & Jensen, T. H. The Nuclear RNA Exosome and Its Cofactors. in *The Biology of mRNA: Structure and Function* (eds. Oeffinger, M. & Zenklusen, D.) 113–132 (Springer International Publishing, 2019). doi:10.1007/978-3-030-31434-7_4.
46. Lange, H. *et al.* The RNA Helicases AtMTR4 and HEN2 Target Specific Subsets of Nuclear Transcripts for Degradation by the Nuclear Exosome in Arabidopsis thaliana. *PLoS Genet.* **10**, e1004564 (2014).
47. Zhang, H. *et al.* An Rrp6-like Protein Positively Regulates Noncoding RNA Levels and DNA Methylation in Arabidopsis. *Mol. Cell* **54**, 418–430 (2014).
48. Pontes, O. *et al.* The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. *Cell* **126**, 79–92 (2006).
49. Böhmdorfer, G. *et al.* Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *eLife* **5**, e19092 (2016).
50. Zhu, Y., Rowley, M. J., Böhmdorfer, G. & Wierzbicki, A. T. A SWI/SNF Chromatin-Remodeling Complex Acts in Noncoding RNA-Mediated Transcriptional Silencing. *Mol. Cell* **49**, 298–309 (2013).
51. Zakrzewska-Placzek, M., Souret, F. F., Sobczyk, G. J., Green, P. J. & Kufel, J. Arabidopsis

thaliana XRN2 is required for primary cleavage in the pre-ribosomal RNA. *Nucleic Acids Res.* **38**, 4487–4502 (2010).

52. Gy, I. *et al. Arabidopsis* FIERY1, XRN2, and XRN3 Are Endogenous RNA Silencing Suppressors. *Plant Cell* **19**, 3451–3461 (2007).
53. Shin, J.-H. *et al.* The Role of the Arabidopsis Exosome in siRNA–Independent Silencing of Heterochromatic Loci. *PLoS Genet.* **9**, e1003411 (2013).
54. Western, T. L., Cheng, Y., Liu, J. & Chen, X. HUA ENHANCER2, a putative DExH-box RNA helicase, maintains homeotic B and C gene expression in Arabidopsis. 28 (2016).
55. Blevins, T. *et al.* Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *eLife* **4**, e09591 (2015).
56. Jensen, T. H., Jacquier, A. & Libri, D. Dealing with Pervasive Transcription. *Mol. Cell* **52**, 473–484 (2013).
57. Rüegger, S. & Großhans, H. MicroRNA turnover: when, how, and why. *Trends Biochem. Sci.* **37**, 436–446 (2012).
58. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **8**, 113–126 (2007).
59. Calvin, K. & Li, H. RNA-splicing endonuclease structure and function. *Cell. Mol. Life Sci.* **65**, 1176–1185 (2008).
60. Liu, J. *et al.* Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science* **305**, 1437 (2004).
61. Liu, W. *et al.* RNA-directed DNA methylation involves co-transcriptional small-RNA-guided slicing of polymerase V transcripts in Arabidopsis. *Nat. Plants* **4**, 181–188 (2018).
62. Alonso, J. M. *et al.* Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana. *Science* **301**, 653–657 (2003).