RNA Pol IV has antagonistic parent-of-origin effects on Arabidopsis endosperm

Prasad R.V. Satyaki[1] and Mary Gehring[1,2]

[1] Whitehead Institute for Biomedical Research, Cambridge MA 02142

[2] Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139

Corresponding author: M.G. (mgehring@wi.mit.edu)

Abstract

Gene expression in endosperm – a seed tissue that mediates transfer of maternal resources to offspring – is under complex epigenetic control. We show here that plant-specific RNA Polymerase IV mediates parental control of endosperm gene expression. Pol IV is required for the production of small interfering RNAs that typically direct DNA methylation. We compared small RNAs, DNA methylation, and mRNAs in *A. thaliana* endosperm from reciprocal heterozygotes produced by crossing wild-type plants to Pol IV mutants. We find that maternally and paternally acting Pol IV have divergent effects on endosperm with loss of maternal and paternal Pol IV impacting sRNAs and DNA methylation at different genomic sites. Strikingly, maternally and paternally-acting Pol IV have antagonistic impacts on gene expression at some loci, divergently promoting or repressing endosperm gene expression. Antagonistic parent-of-origin effects have only rarely been described and are consistent with a gene regulatory system evolving under parental conflict.

Introduction

Parents instruct zygotic development in viviparous plant and animal species. In flowering plants, parent-of-origin effects on offspring development are observed in an embryo-surrounding seed tissue called the endosperm (Gehring and Satyaki, 2017). Endosperm does not contribute genetic material to the next generation but mediates maternal nutrient transfer to the embryo, coordinates growth between the embryo and maternal tissues, sets seed dormancy and regulates germination, and acts as a nutrient store to support seedling growth (Li and Berger, 2012). Endosperm is typically triploid and develops from the fertilization of a diploid female gamete, called the central cell, by one of two haploid sperm cells that are released by pollen. Violations of the balanced ratio of two maternal to one paternal genomes disrupts normal endosperm development in a parent-of-origin dependent manner (Müntzing, 1936; Milbocker and Sink, 1969; Scott *et al.*, 1998; Stoute *et al.*, 2012; Povilus *et al*., 2018). In some *A. thaliana* accessions*,* crosses between tetraploid mothers and diploid fathers exhibit reduced endosperm proliferation and smaller mature seeds while reciprocal crosses where the fathers are tetraploid (paternal excess crosses) exhibit prolonged endosperm proliferation and larger or aborted seeds. These parent-of-origin effects on endosperm development have been interpreted under the aegis of the parental conflict or kinship model (Haig, 2013). According to this model, when a mother mates with more than one father, the inclusive fitness of the mother is optimal when her resources are equally distributed among her progeny. The inclusive fitness of the father is optimal when his progeny are able to consume more maternal resources than other half-siblings. Such conflicts are postulated to lead to arms races whose impacts may be observed in the molecular machinery mediating parental control. However, our understanding of the impact of conflict on endosperm biology is limited by our incomplete understanding of molecular and genetic mechanisms guiding parental control of endosperm development.

Recent data indicate that mutations in RNA Polymerase IV have effects on reproduction, endosperm, and seed development in multiple species (Erdmann *et al.*, 2017; Grover *et al.*, 2018; Martinez *et al.*, 2018; Kirkbride *et al.*, 2019; Satyaki and Gehring, 2019; Wang *et al.*, 2020). RNA Pol IV functions as part of the RNA directed DNA methylation (RdDM) pathway, in which it produces relatively short, non-coding transcripts that are converted into double stranded RNA by RDR2 (Blevins *et al.*, 2015; Li *et al.*, 2015; Zhai *et al.*, 2015) . These double-stranded RNAs are cleaved into 24nt small RNAs (sRNAs) by DCL3 and single strands are loaded into ARGONAUTE proteins that help target the *de novo* DNA methyltransferase DRM2, which acts in conjunction with RNA Pol V and several other proteins, to methylate DNA (Matzke and

Mosher, 2014). *NRPD1*, which encodes the largest subunit of RNA Pol IV, has roles in endosperm gene dosage control. Endosperm gene expression typically reflects the ratio of two maternally and one paternally inherited genomes, such that for the majority of genes approximately two-thirds of genic transcripts are derived from maternal alleles (Gehring, *et al.*, 2011; Pignatta *et al.*, 2014). A genome-wide survey in *nrpd1* mutant endosperm found that Pol IV is required to maintain the 2:1 maternal to paternal transcript ratio in the endosperm and that loss of Pol IV leads to the mis-regulation of several hundred genes (Erdmann *et al.*, 2017). Additionally, loss of function mutations in *NRPD1* or other members of the RdDM pathway can repress seed abortion in crosses of diploid mothers and tetraploid fathers (Erdmann *et al.*, 2017; Martinez *et al.*, 2018; Satyaki and Gehring, 2019). In *B. rapa,* loss of *NRPD1*, *RDR2*, or *NRPE1* results in high rates of seed abortion due to maternal sporophytic effects (Grover *et al.*, 2018). Loss of Pol IV in both *B. rapa* and in *A. thaliana* also results in smaller seed sizes (Grover *et al.*, 2018) and RNA Pol IV is essential to post-meiotic pollen development in *C. rubella* (Wang *et al.*, 2020).

Molecular data point to the intriguing possibility that mutations in RNA Pol IV have parent-of-origin effects on endosperm. A comparison of sRNAs in wild-type whole seeds (which includes maternal seed coat, endosperm, and embryo) with *NRPD1+/-* endosperm from crosses where the mutation in *NRPD1* was maternally or paternally inherited suggested that loss of maternal *NRPD1* affected more sRNA loci than the loss of paternal *NRPD1* (Kirkbride *et al.*, 2019). Although the comparison of sRNAs from wild-type whole seeds to mutant endosperm in this study makes definitive conclusions hard to draw, it raises the potential question of if and how the loss of *NRPD1* has parent-of-origin effects on sRNA production.

Genetic data also supports a model in which Pol IV action in the maternal and paternal parents have different consequences for the endosperm. In paternal excess interploidy crosses, loss of paternal copies of *NRPD1* or other downstream members of the RdDM pathway repress interploidy seed abortion (Erdmann *et al.*, 2017; Satyaki and Gehring, 2019). By contrast, loss of the maternal copies of *NRPD1* slightly enhances paternal excess interploidy seed abortion (Erdmann *et al.*, 2017; Satyaki and Gehring, 2019). Why this is so remains unclear but suggests that maternal and paternal Pol IV likely have different downstream impacts on gene expression in the endosperm after fertilization. The genome-wide impact of parental Pol IV activity on gene expression in the endosperm is not known.

To examine the impacts of parental Pol IV activity on endosperm in more detail, we examined sRNA and mRNA transcriptomes in wild-type endosperm, *nrpd1* homozygous mutant

endosperm, and *nrpd1* heterozygous endosperm where the mutant allele was inherited from a homozygous mutant mother or father. We also examined methylomes in wild-type and endosperm from the reciprocal heterozygotes. Analysis of these data demonstrate that *NRPD1* has parent-of-origin effects on endosperm, some of which are antagonistic.

Results

Loss of maternal or paternal Pol IV activity impacts small RNAs at distinct sites

To understand the impact of parental Pol IV activity on sRNA production in the endosperm, we first identified Pol IV-dependent sRNAs. Previously we showed that 24nt sRNAs were the predominant sRNA species in endosperm and exhibited a wider distribution over genes and transposable elements (TEs) than in other tissues (Erdmann *et al.*, 2017). We profiled small RNA populations in three replicates of endosperm derived from crosses of L*er* *nrpd1* females pollinated by Col-0 *nrpd1* males (7 days after pollination) and compared them with our previously published sRNA libraries from L*er* x Col-0 endosperm (female parent in cross written first) (Erdmann *et al.*, 2017) (Supplementary Table 1).

We identified 21,131 sRNA peaks in L*er* x Col-0 (wild type) endosperm using ShortStack (Axtell, 2013). 76.9% of these were predominantly populated by 24nt sRNAs, with 1.1%, 0.2%, and 2.2% of peaks dominated by 23, 22, and 21nt sRNAs, respectively. An additional 19.7% of peaks were either dominated by a non-canonical sRNA size or had no predominant size class (Fig. S1A). The majority of sRNAs were dependent on *NRPD1*, with 99% of 24nt sRNA peaks, 94.87% of 22nt sRNA peaks and 70.1% of 21nt sRNA peaks absent in *nrpd1-/-* endosperm (Fig. S1B). An examination of mapped sRNA reads from wild-type and *nrpd1-/-* endosperm showed that numerous Pol IV-dependent sRNAs could be found outside the peaks identified by ShortStack. Therefore, as an alternative analytical approach, we binned sRNAs by size (21 to 24 nt) and calculated read counts overlapping TEs and genes encoding proteins, miRNA, and other ncRNA. We used DESeq2 to separately identify genes and TEs with significant differences in Pol IV-dependent sRNA populations. Loss of RNA Pol IV abolished 21-24 small RNAs at most TEs and genes, while most known miRNAs were not impacted (Supplementary Table 2). 21-23 nt sRNAs were often lost at the same loci as 24 nt sRNAs (Fig. S1C), suggesting that sRNAs of differing sizes arose from the same Pol IV transcript but were likely processed into RNAs shorter than 24 nt by different downstream DICERs or by the exosome components *Atrimmer1* and *2* (Daxinger *et al.*, 2009; Ye *et al.*, 2016). The presence of Pol IV-

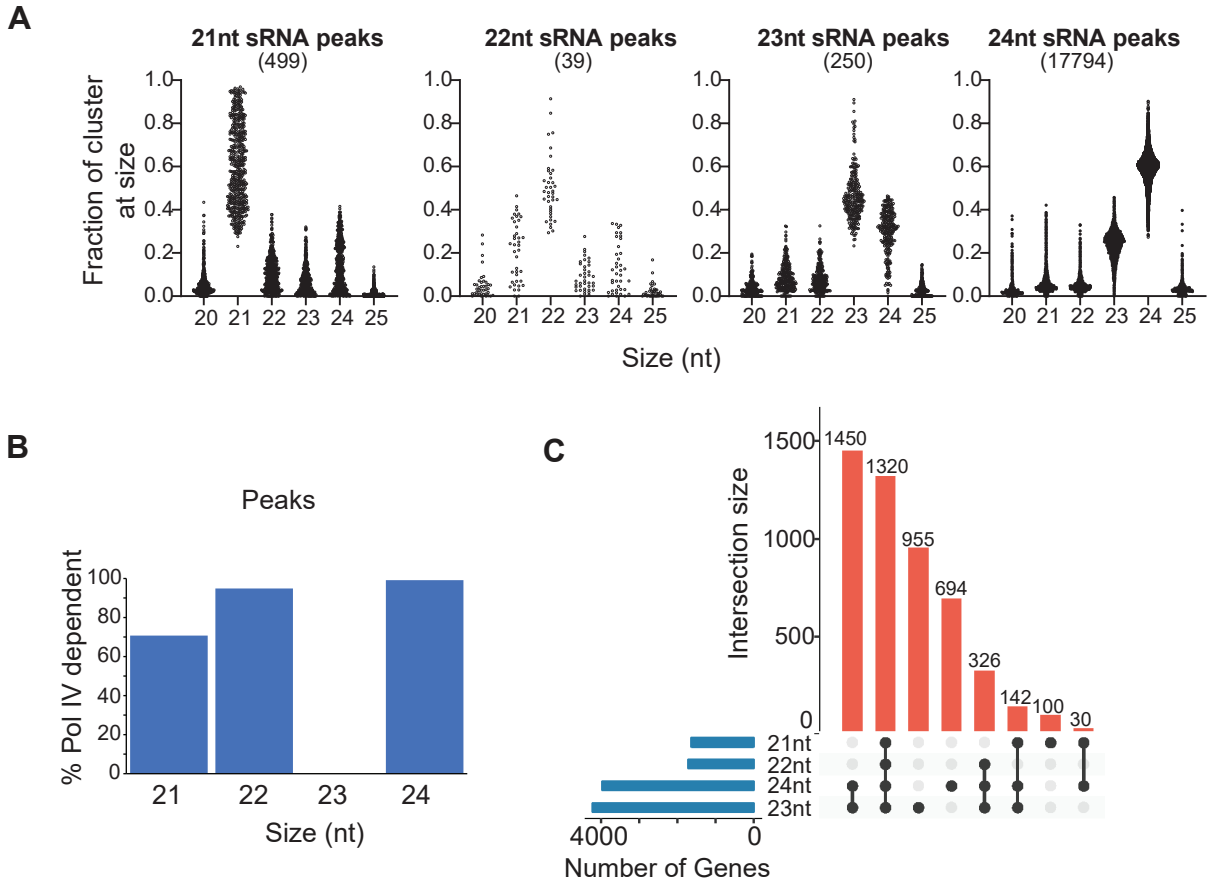


Figure S1. RNA Pol IV is necessary for the production of 21-24nt sRNAs in the endosperm. (A) Size (nt) of all sRNAs in endosperm small RNA peaks dominated by 21, 22, 23, or 24 nt sRNAs. ShortStack was used to call peaks in wild-type (L*er* x Col-0) endosperm. Each peak is grouped into a size class based on the predominant size of the small RNA species in that cluster. Fraction of small RNAs at other sizes are plotted. **(B)** Small RNA peaks of multiple sizes are impacted by loss of *NRPD1*. Genome-wide small RNA coverage over 300bp windows overlapping by 200bp was calculated. DESeq2 was used to identify windows with differentially expressed sRNAs between WT and *nrpd1-/-*. Overlapping windows were merged. Peaks overlapping windows with reduced expression in *nrpd1-/-* were classified as Pol IV-dependent peaks. **(C)** Upset plot shows that the same genes lose sRNAs of all size classes in *nrpd1-/-* endosperm.

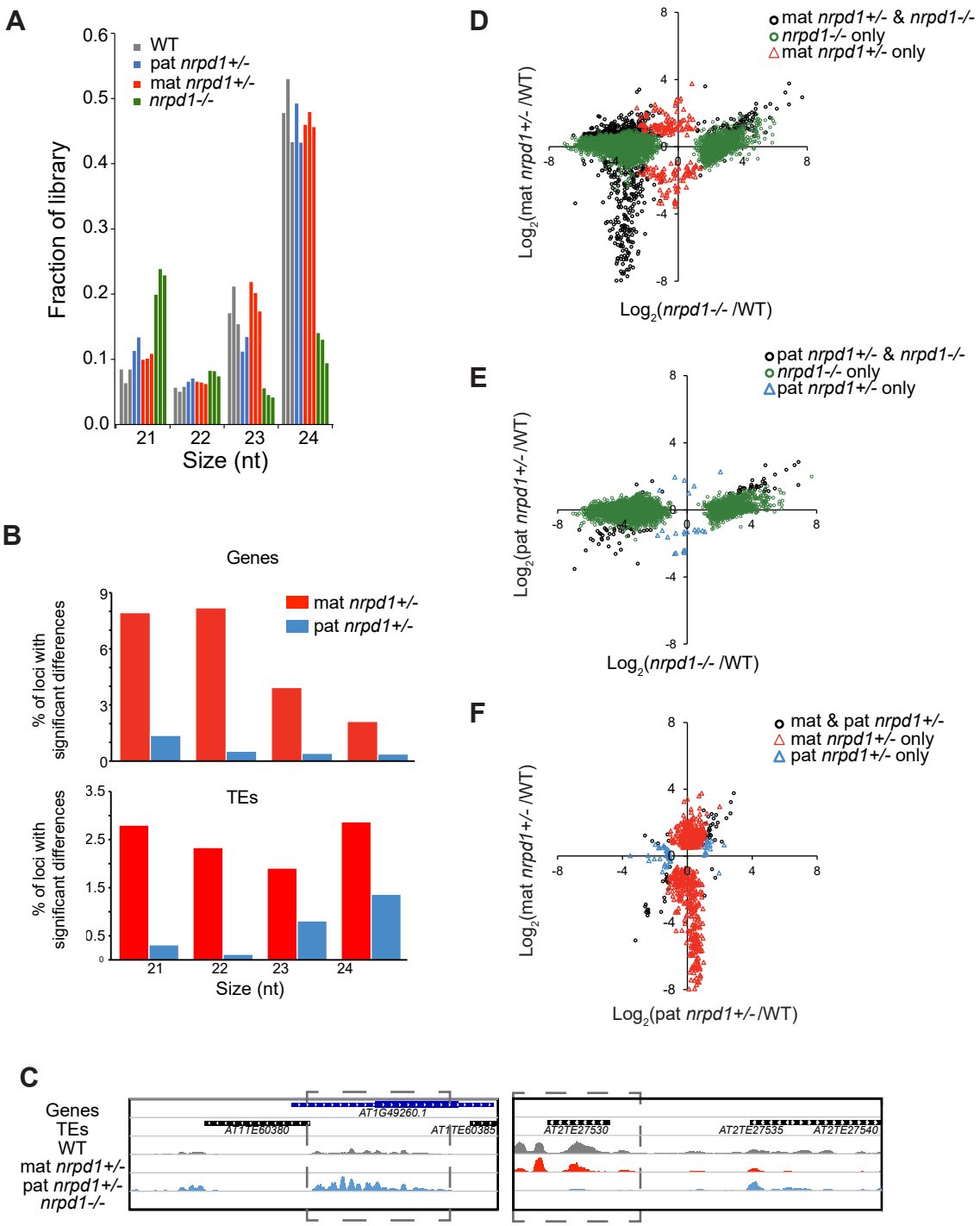


Fig. 1: Impact of loss of maternal, paternal, or both copies of *NRPD1* on endosperm small RNAs. (**A**) Loss of maternal or paternal *NRPD1* does not substantially alter the endosperm small RNA pool. Fraction of aligned small RNA reads in each size class in the indicated genotypes. (**B**) Examination of 21-24 nt sRNAs over genes or TEs shows that inheriting a mutant maternal *nrpd1* allele has a larger impact than inherting a mutant paternal *nrpd1* allele. Percent of loci showing at least a two-fold reduction in sRNA abudance and padj <0.05 according to DESeq2 are indicated in red (mat *nrpd1*+/-) or blue (pat *nrpd1*+/-). Genes and TEs included in this tally have a normalized wild-type read count of five or higher. (**C**) Snapshots of loci with Pol IV-dependent 24 nt sRNAs that show a specific loss of small RNAs in mat (left) or pat (right) *nrpd1*+/- endosperm. (**D-F**) Comparisons of genic 24 nt sRNAs upon loss of maternal, paternal, or both copies of *NRPD1*. Fold change as calculated by DESeq2. Only significant changes (q<0.05) are plotted.

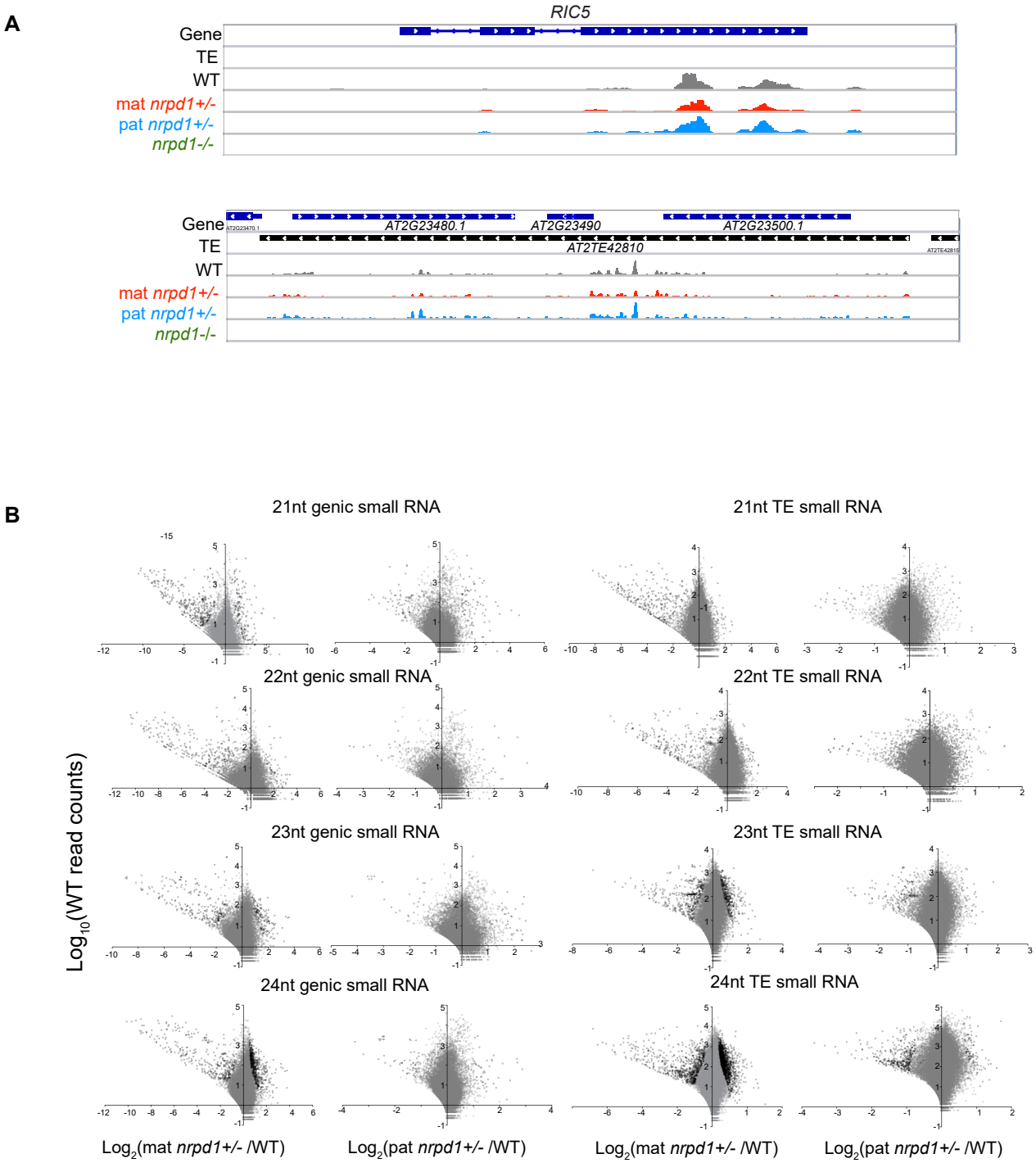


Figure S2: Impact of the loss of maternal and paternal *NRPD1* on the endosperm small RNA population. **(A)** One parent's copy of *NRPD1* is sufficient for 24nt sRNA production from genes and TEs at most loci, here exemplifed by *RIC5* and a VANDAL21 copy. **(B)** Examination of 21-24 sRNA over genes and TEs shows that inheriting a maternal mutation in *NRPD1* has a greater impact than inheriting a paternal mutation in *NRPD1*. Loci with differential sRNA expression were identified using DESeq2. Wild-type (WT) read counts represent average read counts per locus across three replicates. Reads mapping to TE insertions were normalized using genic sRNA expression. Black circles represent $q<0.05$. Gray circles represent $q>0.05$.

dependent 21-23 nt sRNAs in other tissues has recently been reported, indicating this finding is not specific to endosperm (Wu and Zheng, 2019; Panda *et al*., 2020; Wang *et al.*, 2020).

After characterizing Pol IV-dependent sRNAs, we asked how loss of either parent's Pol IV influenced the distribution of Pol IV-dependent sRNAs in *nrpd1* heterozygous endosperm. We sequenced small RNAs from two replicates of L*er* female x Col-0 *nrpd1-*/- male (referred to as pat *nrpd1+/-)* endosperm and three replicates of L*er nrpd1-/-* female x Col-0 male (referred to as mat *nrpd1+/-*) endosperm. Because the endosperm is triploid, in these comparisons there are 3 (wild-type), 2 (pat *nrpd1*+/-), 1 (mat *nrpd1*+/-) and 0 (*nrpd1-*/-) functional *NRPD1* alleles in the endosperm. However, *NRPD1* is a paternally expressed imprinted gene in wild-type L*er* x Col endosperm and the single paternal allele contributes 62% of the *NRPD1* transcript whereas 38% comes from the two maternal alleles (Pignatta *et al.*, 2014). Thus, both heterozygotes should have similar levels of *NRPD1* endosperm transcript. Indeed, mRNA-Seq data shows that *NRPD1* is expressed at 42% of wild-type levels in pat *nrpd1+/-* and at 91% of wild-type levels in mat *nrpd1+/-* (Supplementary Table 6).

We found that the presence of functional *NRPD1* inherited from either parent is sufficient for the biogenesis of nearly wild-type levels of 21, 22, and 24nt sRNAs in endosperm, although pat *nrpd1*+/- exhibited reduced 23 nt sRNA levels (Fig. 1A, Fig. S2). Although the overall sRNA population in the heterozygotes was similar to the wild-type (Fig. 1A), loss of maternal and paternal *NRPD1* had distinct impacts on sRNA at individual loci (Fig. 1B-E, Fig. S2, Table S2-3). We identified genes and transposable element insertions that displayed at least a two-fold change in the abundance of sRNAs in mat or pat *nrpd1+/-* compared to the wild-type (Fig. 1B-F, Fig. S2, Table S2-3). Loss of paternal *NRPD1* caused relatively small reductions in 21-24nt Pol IV sRNAs at a handful of loci, while loss of maternal *NRPD1* had greater impact (Fig. 1B-F, Fig. S2, Table S2-3). Few of the loci with reduced sRNAs were shared between the reciprocal heterozygotes – of 327 24nt sRNA-expressing genic loci that were reduced by more than two-fold in mat *nrpd1*+/-, only 22 were also reduced by two-fold in pat *nrpd1*+/- (Fig. 1F). Moreover, there was no quantitative or correlative relationship between loci affected in mat *nrpd1*+/- and pat *nprd1*+/- (Fig. 1F). Thus, the vast majority of sRNA-producing loci in endosperm only require at least one functional copy of *NRPD1* after fertilization.

Evaluating memory of parental Pol IV activity and endosperm sRNA production

The absence of dramatic differences in sRNAs in heterozygotes could indicate that the alleles inherited from both the wild-type and the *nrpd1-*/- parent produce a wild-type level of

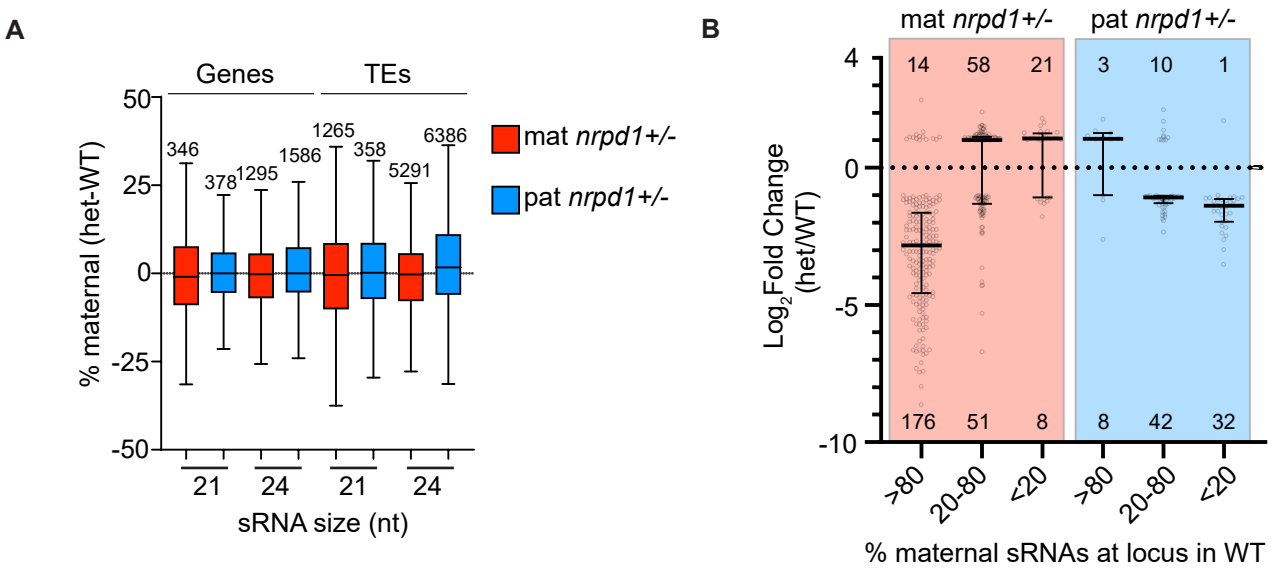


Figure 2: Effects of loss of maternal and paternal Pol IV on the allelic origin of small RNAs. (A) Tukey plot shows that genic and TE small RNAs are bi-allelically produced in heterozygotes. Loci plotted here show similar abundances in wild-type and heterozygotes and have a sum of at least ten allele-specific reads in three wild-type replicates and in heterozygotes. **(B)** Loci losing small RNA in mat *nrpd1*+/- and pat *nrpd1*+/- tend to be maternally and paternally biased in WT. Genes and TE showing differential abundance of 24nt sRNAs in *nrpd1* heterozygotes were grouped into bins by the % of sRNAs produced from the maternal alleles of that locus in WT. Fold-change was calculated by DESeq2.Tukey plot represents fold-change in each group. Circles show fold-change at individual loci. Numbers below and above plot are total number of loci having significantly lower and higher abundance of 24nt sRNAs in *nrpd1*+/- relative to the wild-type.

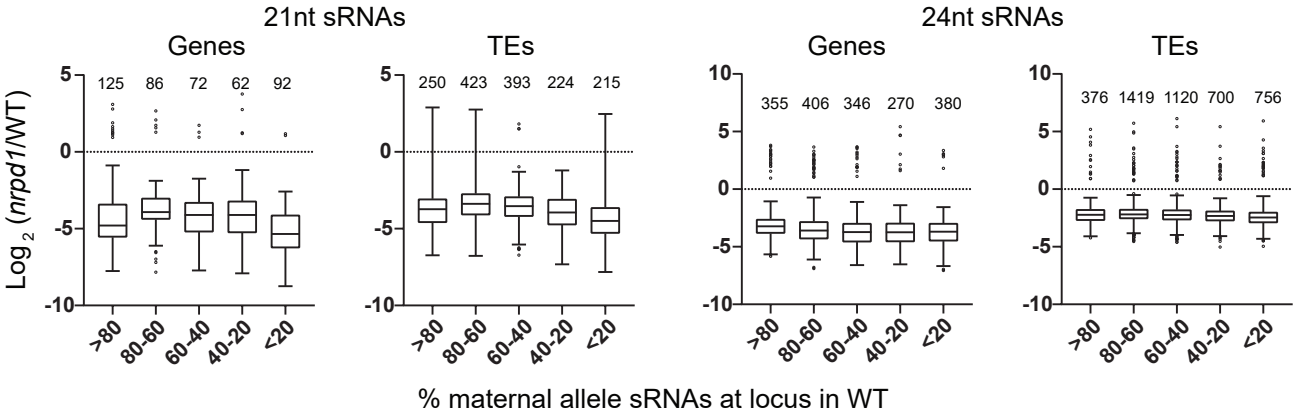


Figure S3: RNA Pol IV-dependent small RNAs arise from both maternal and paternal alleles. SNPs between Col-0 and L*er* were used to identify parental origins of small RNAs arising from genes and transposable elements (TEs). Differentially expressed loci were identified using DESeq2 as described in Figure 1. Loci with a sum of at least ten allele-specific reads in three wild-type L*er* x Col-0 replicates and showing significant differences in 21nt and 24nt sRNAs in Ler *nrpd1* -/- x Col *nrpd1*-/- endosperm were included. Box plots are Tukey plots. Numbers over box plots are number of loci evaluated.

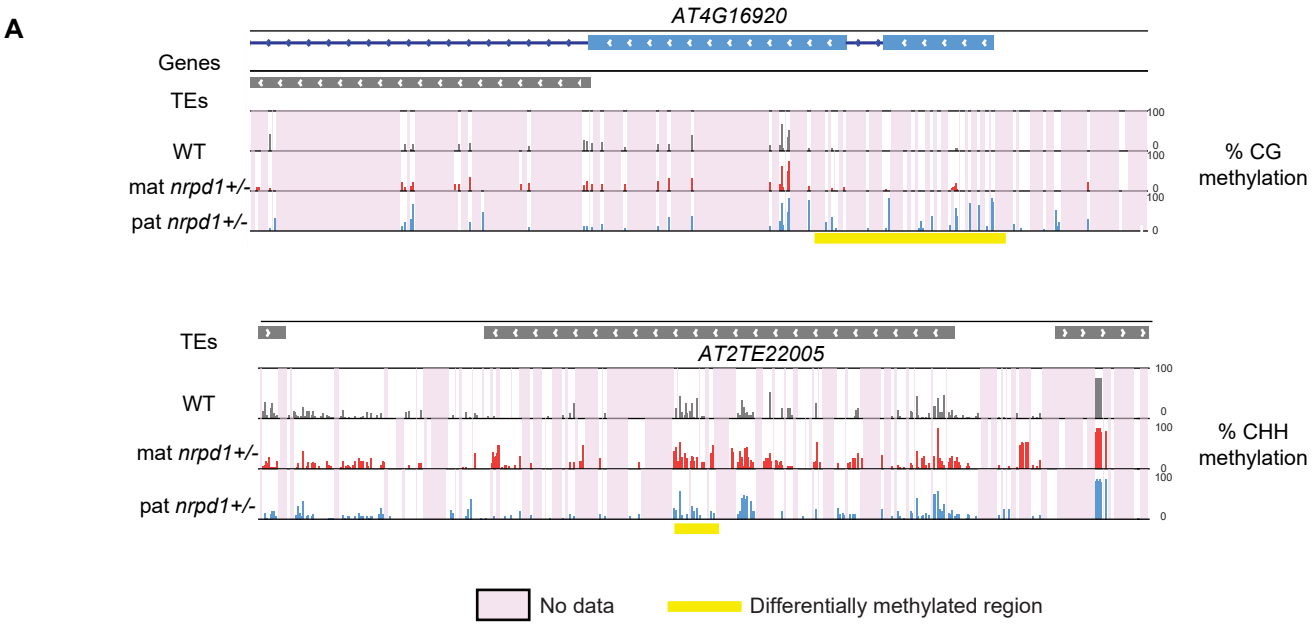
sRNAs after fertilization. This result would be expected for a recessive mutation without parental effects. However, it is known that Pol IV activity at some loci requires prior Pol IV activity (Li *et al.*, 2020). Under such a model, Pol IV activity in the parents before fertilization might be necessary for sRNA production from that parent's allele in the endosperm. Thus, the observed lack of differences in sRNA production at most loci in heterozygous *nrpd1* endosperm could be explained by an upregulation of sRNA production from the alleles inherited from the wild-type parent (i.e. paternal allele sRNAs are upregulated in mat *nrpd1*+/- and maternal allele sRNAs are upregulated in pat *nrpd1*+/- endosperm). To distinguish between these possibilities, we used the SNPs between Col-0 and L*er* to identify the allelic origins of small RNAs in WT and heterozygous endosperm. We first confirmed prior observations that Pol IV sRNAs are biallelically expressed at most loci in endosperm and predominantly expressed from one parental allele, or imprinted, at a several hundred others (Erdmann *et al.*, 2017). Examining sRNAs at genes and TEs, we found that both bi-allelically expressed 21 and 24 nt sRNA loci (defined as between 20% and 80% of sRNAs from maternally-inherited alleles) and those predominantly expressed from one parental allele (>80% or <20% maternal) were Pol IV-dependent (i.e. their accumulation was reduced in *nrpd1*-/- endosperm) (Fig. S3).

To test if sRNA production from alleles inherited from wild-type parents compensated for alleles inherited from an *nrpd1-/-* parent, we first assessed loci that were not mis-regulated in *nrpd1*+/- endosperm. Overall, there were similar contributions from maternal and paternal alleles in mat and pat *nrpd1* heterozygotes compared to wild-type endosperm (Fig. 2A). This suggests that for most loci, endosperm sRNAs are produced from both maternal and paternal alleles regardless of whether the alleles were inherited from a wild-type parent or an *nrpd1*-/- parent. However, we hypothesized that imprinted sRNA loci (Erdmann *et al.*, 2017) might be specifically impacted by loss of parental Pol IV. Consistent with this idea, small RNA loci showing dramatic reductions in abundance in mat *nrpd1+/-* tended to be maternally biased in wild-type endosperm (>80% of sRNAs from the maternally-inherited alleles) (Fig. 2B, leftmost column). Similarly, in pat *nrpd1*+/-, paternally biased small RNA (<20% sRNAs from the maternally-inherited alleles) loci were more impacted (Fig. 2B, rightmost column). These results indicate that parentally-biased sRNA loci in endosperm are dependent on Pol IV activity in the parents and cannot be established post-fertilization. Notably, these sites of Pol IV action are by definition distinct between maternal and paternal parents.

Loss of maternal and paternal Pol IV activity has differing impacts on DNA methylation

Pol IV-dependent sRNAs direct DNA methylation (Onodera *et al.*, 2005; Stroud *et al.*, 2013). The differing impacts of maternal and paternal Pol IV activity on sRNAs in the endosperm suggests that Pol IV activity in the mother and the father might also have different impacts on DNA methylation in the endosperm. To test this model, we performed bisulfite sequencing of two replicates each of WT, mat, and pat *nrpd1+/-* endosperm DNA extracted from seeds at 7 days after pollination. Replicates were merged for each genotype and methylation status extracted using Bismark (Krueger and Andrews, 2011). We then identified differentially methylated regions or DMRs, which were defined as regions in which DNA methylation in CG, CHG and CHH contexts differed between comparisons by at least 30%, 20% and 10%, respectively) using a previously published script (Pignatta *et al*. , 2015).

Loss of Pol IV activity primarily reduces asymmetric CHH methylation (Stroud *et al.*, 2013). We found that heterozygous *nrpd1* endosperm had modest differences in CHH methylation compared to WT, but that loss of paternal *nrpd1* had a greater impact than loss of maternal *nrpd1* (Fig. 3; Table S5). Comparison of mat *nrpd1+/-* and pat *nrpd1+/-* CHH methylation with wild-type endosperm identified 2234 and 2056 DMRs (covering 812.7 Kb and 759.9Kb, respectively) with 50% hypomethylated in mat *nrpd1+/-* and 54.8% hypomethylated in pat *nrpd 1+/-* (Table S5). Less than 50% of the CHH DMRs were shared between the heterozygotes (Fig. 3B). Additionally, a comparison of CHH methylation between mat and pat *nrpd1+/-* identified 2172 DMRs encompassing 799.7Kb of DNA with 1161 DMRs exhibiting higher CHH methylation in mat *nrpd1+/-* (Table S5). To further assess the different impact of mat and pat *NRPD1* on CHH methylation, we examined regions with at least 10% CHH methylation in wild-type and assessed how this methylation changed in mat and pat *nrpd1+/-.* We found that loss of pat *NRPD1* had a more substantial impact on endosperm CHH methylation than loss of mat *NRPD1* (Fig. 3D). Additionally, regions where sRNA accumulation is dependent on paternal inheritance of a wild-type *NRPD1* allele have higher CHH methylation in wild-type endosperm than regions where sRNAs are dependent on maternal *NRPD1* (Fig. 3C). This pattern is consistent with our previous finding that maternally-biased small RNAs are often not associated with methylated DNA in wild-type endosperm (Erdmann et al., 2017). One hallmark of reduced RdDM activity is lower expression of the DNA demethylase *ROS1* (Lei *et al.*, 2015; Williams *et al.*, 2015). We found that *ROS1* expression was significantly reduced in pat *nrpd1*+/- but not mat *nrpd1*+/-, consistent with reduced CHH methylation in pat *nrpd1*+/- (Table S6).



B

CG methylation	pat *nrpd1+/-* < WT (376 DMRs)	pat *nrpd1+/-* > WT (249 DMRs)
mat *nrpd1+/-* < WT (289 DMRs)	21	0
mat *nrpd1+/-* > WT (305 DMRs)	3	14

CHH methylation	pat *nrpd1+/-* < WT (743 DMRs)	pat *nrpd1+/-* > WT (599 DMRs)
mat *nrpd1+/-* < WT (924 DMRs)	335	11
mat *nrpd1+/-* > WT (909 DMRs)	10	129

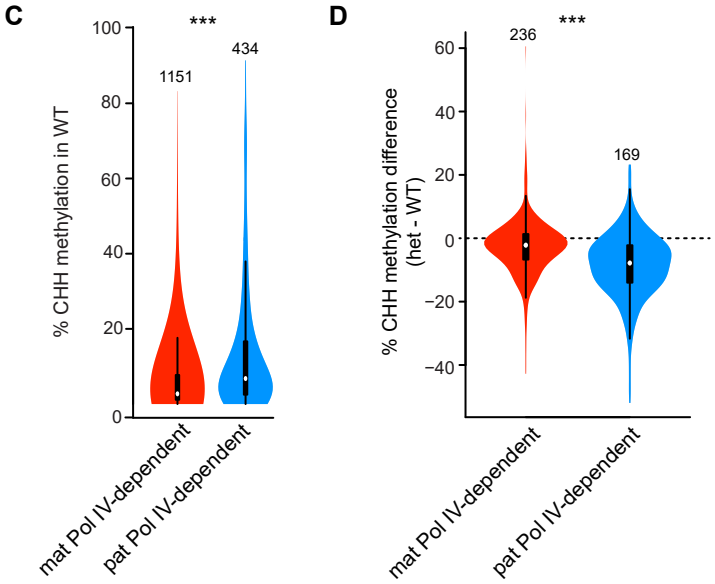


Figure 3: Distinct DNA methylation profiles in *nrpd1* heterozygotes. (A) Browser snapshots show examples of regions where CG and CHH DNA methylation differ between in mat and pat *nrpd1+/-* endosperm. **(B)** Overlap between regions whose CG or CHH methylation changes upon loss of maternal or paternal *NRPD1*. Numbers under genotype represent filtered DMRs with sufficient information in genotypes compared. **(C)** Wild-type CHH methylation at sRNA producing sites that are dependent on maternal or paternal Pol IV. Methylation is significantly higher at paternal Pol IV-dependent sites. **(D)** Effects of parental Pol IV loss on CHH methylation at regions with parental Pol IV-dependent sRNAs and greater than 10% CHH methylation in WT. Red, difference between mat *nrpd1+/-* and WT; blue, difference between pat *nrpd1+/-* and WT. Small RNA producing regions impacted in paternal *nrpd1+/-* have higher losses of CHH methylation. For B and C, CHH methylation was calculated for 300bp sliding windows with a 200 bp overlap. CHH methylation windows overlapping windows losing small RNAs in *nrpd1+/-* endosperm were identified and merged using bedtools; maximum CHH methylation among merged windows was used for violin plot. *** represents a statistically significant difference as calculated by Wilcoxon test ($p<0.001$). Boxplot in the violin plot shows median and inter-quartile range.

Symmetric CG and CHG methylation are typically less affected by loss of *NRPD1* because other mechanisms exist to maintain this type of methylation. Comparison of CHG methylation between wild-type and either heterozygote identified fewer than 100 DMRs and CHG methylation was not investigated further. Both mat and pat *nrpd1+/-* endosperm exhibited changes in CG methylation compared to the wild type (Fig. 3A,B; Table S5). In mat *nrpd1+/-* endosperm, 48.5% of DMRs (of a total 600 DMRs spanning 207 KB) were hypomethylated relative to wild-type while in pat *nrpd1+/-* 60% of DMRs (of a total 707 DMRs spanning 258 KB) were hypomethylated relative to wild-type. Further, we found that few of the sites hypo- or hyper-methylated in the CG context in mat *nrpd1+/-* were shared with those changing methylation state in pat *nrpd1+/-* (Fig. 3B). Consistent with this observation and underscoring the difference between mat and pat *nrpd1+/-*, a comparison of CG methylation in pat and mat *nrpd1+/-* identified 1181 DMRs spanning 428.3 KB with 671 DMRs exhibiting higher methylation in mat *nrpd1+/-* (Table S5). Thus, Pol IV activity in parents has an unequal impact on the endosperm methylomes with paternal Pol IV playing a larger role than maternal Pol IV. Whether these methylation differences are the direct result of Pol IV activity in the endosperm or parents, or whether Pol IV acts indirectly by regulating the expression of other chromatin or DNA methylation modifiers, remains unclear.

Maternal and paternal RNA Pol IV have antagonistic impacts on gene expression

The distinct effects of maternal and paternal Pol IV activity on sRNA and DNA methylation in the endosperm suggested that parental Pol IV might also have differing impacts on endosperm gene expression. To test this, we profiled endosperm gene expression in three replicates each of mat *nrpd1+/-* and pat *nrpd1+/-*, along with appropriate wild-type controls and homozygous mutant *nrpd1* endosperm. Differential expression analyses showed that transcripts from 1791 genes were more abundant and 1455 were less abundant in *nrpd1-/-* compared to wild-type endosperm (Fig. 4; Table S6). Almost 50% of these genes were similarly mis-regulated in mat *nrpd1+/-* (Fig. 4A,B). In contrast, very few genes changed in expression in pat *nrpd1+/-* compared to the wild type (Fig. 4A, B). While the majority of mis-regulated genes in the heterozygotes were not imprinted in wild-type, 15 paternally expressed and 45 maternally expressed imprinted genes were mis-regulated in mat *nrpd1+/-*. Two maternally expressed genes but no paternally expressed genes were mis-regulated in pat *nrpd1+/-* (Table S6). In addition to the difference in the size of the effect, loss of maternal or paternal Pol IV alter the expression of different classes of genes. Panther over-representation tests show that in mat

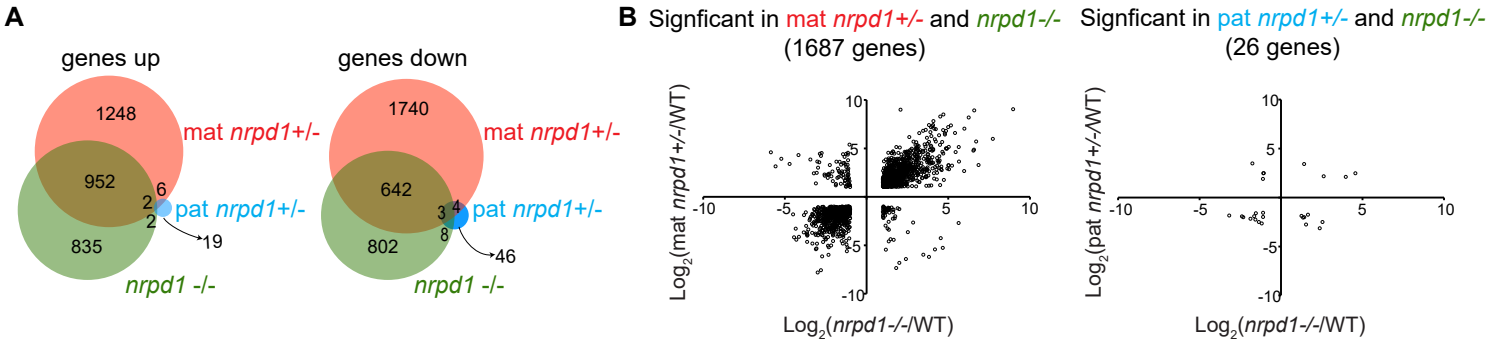


Figure 4: Effect of *NRPD1* mutations on endosperm gene expression. (A) Venn diagrams showing overlap of genes with increased and decreased expression in comparison to wild-type endosperm for the indicated genotypes. **(B)** Scatter plots of genes that are all significantly different (q =/<0.05, Log_2(Fold Change) >/=1 or </=-1) between wild-type and indicated mutants. Fold-change calculated using Cuffdiff.

nrpd1+/-, down-regulated genes are enriched for functions in the cell-cycle while up-regulated genes are enriched for functions in photosynthesis, stress response, and abscisic acid signaling. In pat *nrpd1+/-*, up-regulated genes were enriched for functions in heat stress response while down-regulated genes were enriched for functions in responses to fungi.

Differential expression of a gene between wild-type and *nrpd1-/-* could represent: 1) maternal and paternal effects arising from the loss of *NRPD1* in parents, 2) zygotic effects arising from epistatic interactions between mat *nrpd1-* and pat *nrpd1-*, 3) effects from the loss of all *NRPD1* in the endosperm, or 4) the sum of all three effects. As this study does not examine the effect of knocking-down *NRPD1* specifically in the endosperm, we cannot discriminate the second and third scenarios from one another, but we can detect parental effects. 2988 genes mis-regulated in mat *nrpd1+/-* were not mis-regulated in *nrpd1-/-* endosperm (Fig. 4). We hypothesized that genic mis-regulation in mat *nrpd1+/-* but not *nrpd1-/-* was caused by separate transcriptional effects of maternal and paternal *nrpd1* that were obscured in null mutants. To test this hypothesis, we compared gene expression between mat and pat *nrpd1+/-* (Fig. 5A). We found that 51/90 genes mis-regulated in pat *nrpd1+/-* endosperm were also mis-regulated in mat *nrpd1+/-* endosperm (black circles in Fig. 5A). However, 36 of these 51 genes changed expression in the opposite direction. For example, expression of the gene *SUC2* decreased about four-fold in pat *nrpd1 +/-* endosperm and increased about eight-fold in mat *nrpd1+/-* endosperm (Fig. 5B). If *NRPD1* loss has no zygotic endospermic effect on the expression of these genes, then the mis-regulation observed in *nrpd1-/-* endosperm would be the sum of the parental effects. Indeed, the change in abundance of these genes in *nrpd1-/-* endosperm is close to that predicted by such an additive, antagonistic parental effect model (compare gray and green bars in Fig. 5B). *SUC2* transcript abundance in *nrpd1-/-* changes by 2.7-fold compared to the predicted 2.18-fold change, and other genes show similar effects (Fig. 5B). While the expression of these particular genes showed large effects in both heterozygotes, most genes mis-regulated in mat *nrpd1+/-* did not show a significant change (>2-fold difference in transcript abundance) in pat *nrpd1+/-* endosperm. We therefore hypothesized that mis-regulation of genes in mat *nrpd1+/-* but not *nrpd1-/-* endosperm was due to a small antagonistic effect arising from the loss of pat *NRPD1* in *nrpd1-/-.* Indeed, genes mis-regulated in either mat *nrpd1+/-* or pat *nrpd1+/-* (Fig. 5A) are overall negatively correlated (slope = -0.123). To test this hypothesis further, we evaluated the expression of genes in pat *nrpd1+/-* endosperm that were either mis-regulated in both mat *nrpd1+/-* and *nrpd1-/-* or only in mat *nrpd1 +/-*. We find that transcripts that significantly increased exclusively in mat *nrpd1+/-* had slightly decreased expression in pat *nrpd1+/-* endosperm (Fig. 5D). In contrast, genes that were significantly

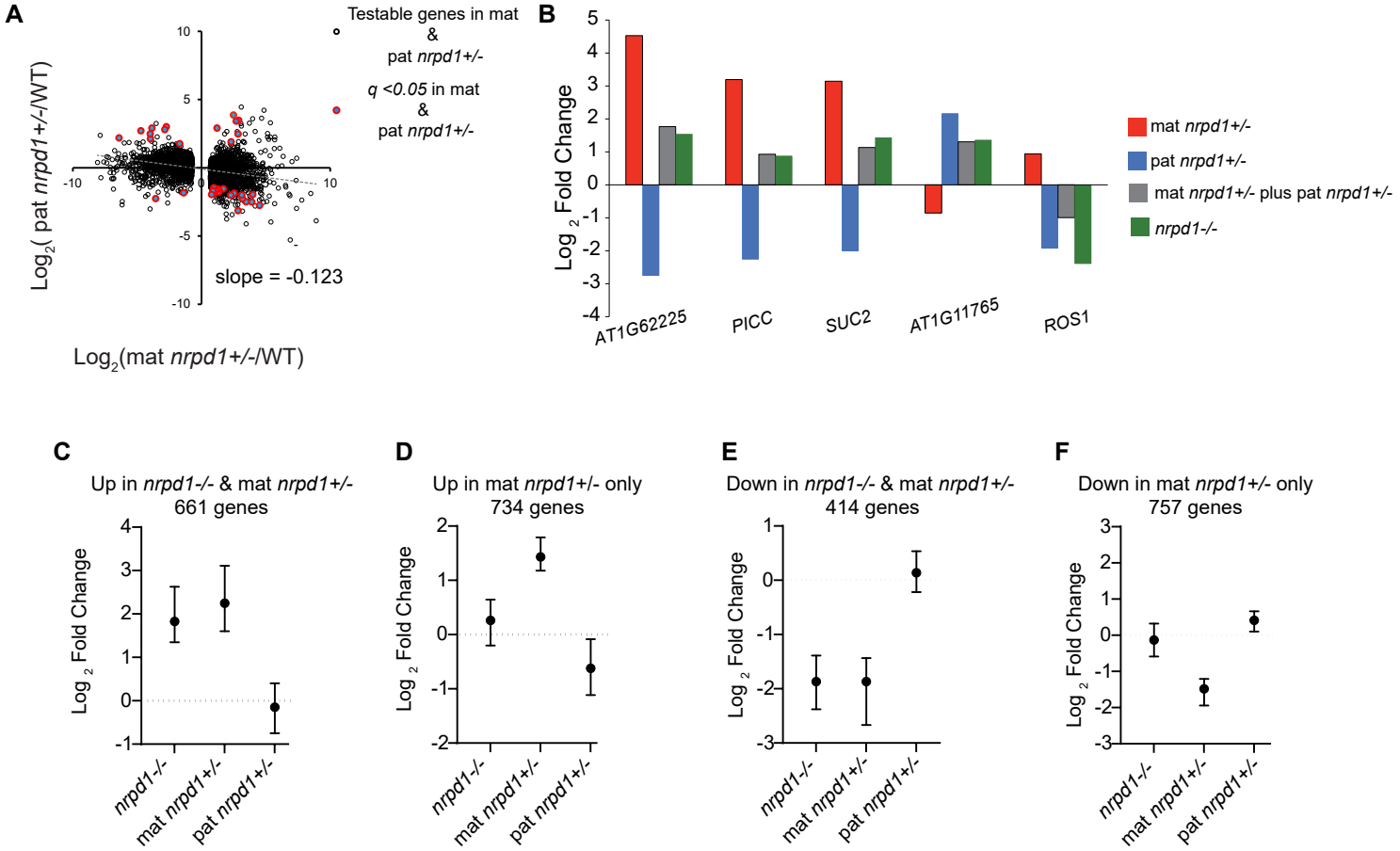


Figure 5. Maternally and paternally acting Pol IV have antagonistic effects on endosperm gene expression. **(A)** Inverse correlation between changes in gene expression in mat and pat *nrpd1*+/- relative to WT. Slope was calculated for all testable genes in comparisons of mat and pat *nrpd1*+/- relative to WT. Genes significantly mis-regulated in both mat and pat *nrpd1*+/- are colored circles. **(B)** Examples of genes that are antagonistically regulated by Pol IV. Gray bars represent mathematical sum of effects observed in mat and pat *nrpd1*+/-. **(C)** Genes up-regulated at least two-fold in both *nrpd1-/-* and mat *nrpd1+/-* do not exhibit mis-regulation in pat *nrpd1+/-*. **(D)** Genes up-regulated only in mat *nrpd1+/-* but not *nrpd1-/-* have decreased expression in pat *nrpd1+/-*. **(E)** Genes down-regulated in both *nrpd1-/-* and mat *nrpd1+/-* are not mis-regulated in pat *nrpd1+/-*. **(F)** Genes down-regulated only in mat *nrpd1+/-* but not in *nrpd1-/-* are overall slightly increased in expression in pat *nrpd1+/-* . Plots C-F show median and inter-quartile range for $\log_2$ fold change in mutant/WT. Fold-change for A-F were calculated by Cuffdiff.

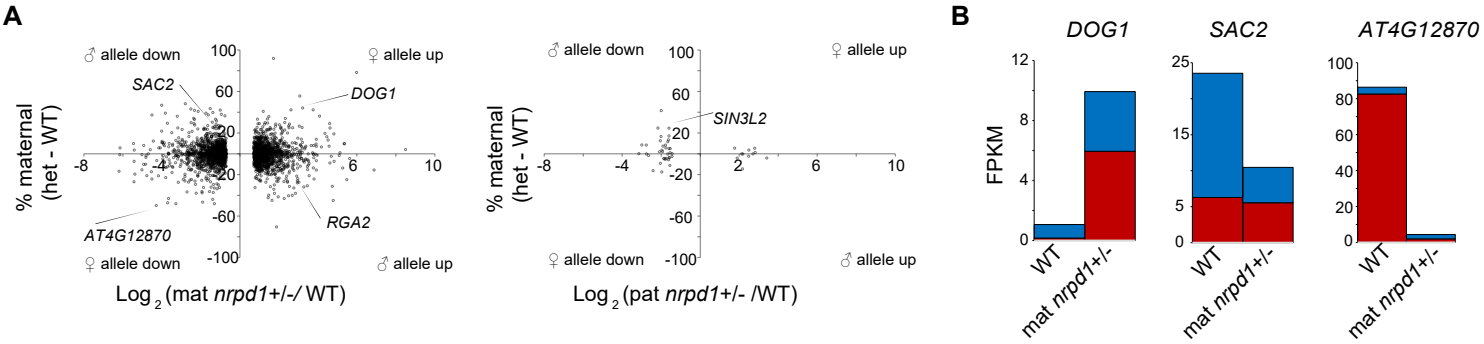


Figure S4: Impact of parental *NRPD1* on maternal and paternal allele contributions to total gene expression. **(A)** Genes were examined to identify those whose expression differences were driven by allele-specific effects. Genes with at least a two-fold, statistically signficant difference in expression between the indicated heterozygote and WT and at least ten allele-specific reads in both genotypes were included. The shift in allelic expression was evaluated by subtracting the % maternal-allele transcripts in WT from the heterozygote. Genes within Col-0 introgressions that remain in L*er nrpd1*-/- plants were excluded from all analyses. **(B)** Examples of genes showing allele-specific impacts upon loss of of maternal Pol IV. FPKM and fold-change in (A) and (B) are from Cuffdiff output.

upregulated in both mat *nrpd1+/-* and *nrpd1-/-* were not affected in pat *nrpd1+/-* (Fig. 5C). Similarly, genes that showed a significant reduction in abundance only in mat *nrpd1+/-* were slightly higher expressed in pat *nrpd1+/-* endosperm (Fig. 5F) while genes with reduced abundance in both mat *nrpd1+/-* and *nrpd1-/-* were not affected in pat *nrpd1+/-* (Fig. 5E). These results are consistent with an antagonistic parent-of-origin effect model for the impact of Pol IV on endosperm transcriptomes.

How does Pol IV acting in the parents affect gene expression in an opposing manner in the endosperm after fertilization? Pol IV might act in parental sporophytes or haploid gametophytes directly at the genes identified as being mis-regulated in heterozygous endosperm. Alternatively, Pol IV's parental activity may control a handful of regulators that in turn control other genes in the endosperm. In the former scenario, gene expression differences in mat *nrpd1+/-* endosperm would be driven predominantly by changes in expression from maternal alleles whereas genes expression differences in pat *nrpd1+/-* endosperm would be driven by changes in expression from paternal alleles. By contrast, biallelic expression changes would be more consistent with the latter scenario, although it would not exclude other explanations. To distinguish between these two possibilities, we utilized SNPs between Col-0 (paternal) and L*er* (maternal) genomes to identify allele-specific mRNA-seq reads. We evaluated the contributions of each parent's alleles in the endosperm for 2372 mis-regulated genes which had at least ten allele-specific reads in wild-type and *nrpd1+/-*. For the majority of genes, mis-regulation in mat *nrpd1+/-* was driven by effects on expression of both maternal and paternal alleles, with some notable exceptions (Fig. S4A). For example, increased expression of *DOG1* in mat *nrpd1+/-* was primarily driven by increased expression from maternal alleles (Fig. S4B). *AT4G12870* was repressed in mat *nrpd1+/-* primarily due to a loss of maternal allele expression (Fig. S4B). In contrast, expression of *SAC2* was primarily repressed in mat *nrpd1+/-* because of decreased expression from the paternal allele (Fig. S4B). Overall, both maternal and paternal alleles made equal contributions to genic mis-regulation in the endosperm. However, 4.7% of down-regulated genes and 5.3% of up-regulated genes showed at least a 20% increase or decrease in maternal allele contribution. This was roughly similar to the contribution of paternal alleles to mis-regulation in mat *nrpd1+/-*. 4.4% of down-regulated genes and 5.3% of up-regulated showed at least a 20% change in paternal allele contribution (Fig. S4A). In pat *nrpd1+/-* , only 8% of down-regulated genes had lower contribution of paternal alleles while both alleles contributed to up-regulation (Fig.S4A). These observations suggest that at most loci the antagonistic parent-of-origin effect of Pol IV in endosperm is not mediated by the direct and inherited effects of Pol IV from before fertilization.

Discussion

We previously demonstrated that Pol IV activity in the father promotes seed abortion in response to extra paternal genomes, whereas Pol IV activity in the mother promotes seed viability in these conditions (Satyaki and Gehring, 2019). Previous observations of sRNA or mRNA at individual genes in diploid endosperm showed that Pol IV function in the mother and the father have different effects on the endosperm (Vu *et al.*, 2013; Kirkbride *et al.*, 2019). These findings suggested that Pol IV has differing, and perhaps even opposing, roles in maternal and paternal parents. In this study, we characterized the effect of maternal and paternal Pol IV activity on the endosperm through genome-wide analyses of transcription, small RNAs, and DNA methylation in balanced endosperm. Our molecular data demonstrate that Pol IV activity in the mother and father have parent-of-origin effects on the endosperm, a subset of which are antagonistic. We found that one parent's copy of *NRPD1* is sufficient for the production of Pol IV-dependent sRNAs at most loci, with a small number of largely non-overlapping loci losing sRNAs upon loss of maternal or paternal *NRPD1*. Pol IV activity in the mother and father also have distinct impacts on the DNA methylation landscape in the endosperm. Endosperm with a paternally inherited *nrpd1* mutation had lower DNA methylation compared with endosperm where the *nrpd1* mutation was maternally inherited. Finally, an interrogation of gene expression shows that loss of maternal Pol IV leads to significant mis-regulation of several hundred genes while loss of paternal Pol IV leads to mis-regulation of only several dozen. A key finding of our study is that genes that are mis-regulated upon loss of maternal *NRPD1* are affected in an opposite manner upon loss of paternal *NRPD1*. Together, our results suggest that maternal and paternal Pol IV are genetically antagonistic and that the major effect on transcription observed in heterozygotes is established before fertilization. These observations are important for understanding both Pol IV's role in reproduction and the genetic architecture underlying parental control of offspring development.

Pol IV, conflict, and the genetic architecture of parental control

Parental conflict theory predicts that in viviparous species, mother and father have antagonistic effects on regulating resource allocation and associated gene expression in offspring. In practice, such effects are difficult to detect and have been infrequently described. Antagonistic parental effects are likely to be balanced in the individuals within an inbred population (like Arabidopsis) and are thus unobservable except in mutants or in hybrids where maternal and paternal effects are out of balance. However, when homozygous mutants are examined, these effects may be missed because they do not cause dramatic developmental

phenotypes or because loss of antagonistic maternal and paternal effects effectively cancel one another out. Thus, reciprocal heterozygotes need to be examined to detect antagonistic parent-of-origin effects. A close examination of our data provides insights into the genetic architecture mediating parental control of offspring development.

A key feature of the regulatory infrastructure that mediates parent-of-origin specific effects on zygotic gene expression is that maternal and paternal alleles need to be distinguished from each other in the zygote (in this case, endosperm is the relevant zygote). In *A. thaliana* endosperm, at many loci maternally inherited alleles are DNA demethylated and marked with H3K27 methylation by Polycomb Repressive Complex2 (PRC2), while paternally inherited alleles remain DNA methylated and have reduced H3K27me3 (Pignatta *et al.*, 2014; Moreno-Romero *et al.*, 2016; Borg *et al.*, 2020). Maternal inheritance of mutations in the PRC2 sub-units *MEA, FIE*, *FIS2* and *MSI1* leads to endosperm defects and seed abortion (Ohad *et al.*, 1996; Chaudhury *et al.*, 1997; Grossniklaus, 1998; Kohler, 2003). Similarly, inheritance of maternal mutations in the DNA demethylase *DME* increases DNA methylation on endosperm maternal alleles and causes seed abortion (Choi *et al.*, 2002). Paternal inheritance of mutations in these genes have no reported effect on endosperm development or gene expression. These results thus argued that the solution to the problem of distinguishing parental alleles from one another after fertilization was to mark maternal and paternal chromosomes with distinct epigenetic modifications. However, this model may not explain all parent-of-origin effects on gene expression, particularly outside of imprinted genes. Our study provides evidence for a distinct model in which the same epigenetic regulator – Pol IV – can mediate both maternal and paternal effects. The only other example of a gene with seemingly antagonistic effects on seeds is the maintenance methyltransferase *MET1*, whose mutation has opposing effects on seed size when inherited maternally or paternally (Xiao *et al.*, 2006).

How does Pol IV in the mother and the father have distinct impacts after fertilization? Pol IV targets can be tissue or developmental stage-specific (Grover *et al.*, 2020) and thus Pol IV may target different genomic regions during male and female gametogenesis. Pol IV could act pre-or post-meiotically in the parental sporophyte (diploid phase of the life cycle), in the gametophyte (haploid phase of life cycle), or post-fertilization in the maternal sporophyte. RT-PCR based examination of dissected synergids and central cells did not detect *NRPD1* transcripts (Vu *et al.*, 2013). This suggests that on the maternal side Pol IV influences endosperm gene expression by acting in the maternal sporophyte or in the female gametophyte

prior to central cell formation. Alternatively, Pol IV could act in the maternal sporophytic integuments/seed coat after fertilization, as the endosperm is developing.

We have shown that a memory of parental Pol IV activity is dispensable for guiding endosperm sRNA production at most loci (except for imprinted sRNAs), but that memory plays an important role in guiding endosperm gene expression. The molecular nature of this memory is unknown. Data from paternal excess interploidy crosses suggests that the molecular identity of Pol IV memory may differ between the maternal and paternal parents. In the father, the genes required for sRNA production (*NRPD1*, *RDR2* and *DCL3*) and the genes required for downstream DNA methylation (*NRPE1/Pol V* and *DRM2*) are both essential for paternal excess seed abortion (Satyaki and Gehring, 2019). In contrast, in the mother, genes required for sRNA production but not for DNA methylation promote paternal excess seed viability (Satyaki and Gehring, 2019). This suggests that DNA methylation or another downstream chromatin mark directed by Pol IV-dependent sRNAs could be the identity of paternally-inherited memory, but is unlikely to be the molecular identity of maternally-inherited memory. This is consistent with the observation that loss of paternal Pol IV has a larger impact on CHH methylation in the endosperm. What would be the nature of maternal DNA methylation-independent memory? Pol IV, like other RNA polymerases (Studitsky *et al.*, 2004), could act as a chromatin remodeler. Or, Pol IV could direct a chromatin modification, produce sRNAs that post-transcriptionally control genes, or control the expression of genes whose products are deposited in the gametes, which in turn sets up a memory to direct gene expression programs in the endosperm after fertilization.

Studies on how resource allocation conflicts between parents impact gene expression have thus far been focused on imprinted genes. However, a handful of studies show the importance of non-imprinted genes in parent-of-origin effects (Mott *et al.*, 2014; Al Adhami *et al.*, 2015). For example, QTL analyses of a heterogeneous mouse stock showed that non-imprinted genes mediate parent-of-origin effects on the offspring's immune system (Mott *et al.*, 2014). Our study describes for the first time a system in which the same epigenetic regulator acts in the mother and the father to antagonistically regulate the same non-imprinted genes in the zygote. Our allele-specific mRNA-seq data shows that loss of Pol IV from one parent can impact alleles inherited from both parents in the endosperm. This suggests that Pol IV does not act directly at antagonistic loci but acts instead by regulating other modifiers of gene expression. Overall, these data suggest that Pol IV is part of a gene regulatory network that is evolving under parental conflict.

How might we interpret these roles for Pol IV in terms of conflicts between parents? A simple interpretation would be that both mother and father use the same weapon – RNA Pol IV – in the conflict over the regulation of some genes. However, these observations can be viewed through a different lens. Parental conflict can be resolved or paused if both parents can modulate the expression level of a gene or the activity of a pathway to an optimum that is tolerable to both parents. Pol IV's role in mediating the antagonistic effects of both parents makes it an ideal system to negotiate optimal gene expression levels. Thus, Pol IV may not be solely an agent of conflict, but also a means to resolving it.

Materials and Methods

Arabidopsis growth conditions, strains and tissue collection

Plants used in this experiment were grown at 22° C in a Conviron chamber on a 16hr light/8hr dark cycle (120 µM light). The *A. thaliana* mutant used in this study was *nrpd1a-4* (*SALK_083051* obtained from ABRC) (Herr *et al.*, 2005) in the Col-0 background. We also utilized *nrpd1a-4* introgressed 4 times into L*er* (Erdmann et al, 2017). Endosperm from approximately 100 seeds (7 days after pollination) from at least three siliques was dissected free of embryos and seed coats and pooled for each replicate as previously described (Gehring et al., 2011). Replicates were collected from crosses that used different individuals as parents.

mRNA, small RNA and DNA isolation and library construction

Large and small sized RNAs were isolated using the RNAqueous micro RNA isolation kit (Thermo Scientific Fisher). Briefly, endosperm dissected from seeds was collected in lysis buffer and then homogenized with an RNAse-free pellet pestle driven by a Kimble motor. Large and small RNA species were isolated and separated using the manufacturer's protocol. The RNA concentration of the larger fraction was measured by Qubit. Small RNA libraries were constructed using the NEXTflex sRNA-seq kit V3 (Biooscientific). Final library amplification was carried out for 25 cycles and the libraries were size selected (135-160bp) using a Pippin Prep (Sage Science). mRNA-seq libraries were constructed using a Smart-Seq2 protocol (Picelli *et al.*, 2014). All libraries were sequenced on the Illumina Hi-Seq 2500 (40bp single-read cycle). DNA for bisulfite sequencing was isolated from dissected endosperm using QiaAMP DNA microkit (QIAGEN 56304). Dissected tissue was incubated overnight in a shaker at 56°C in ATL buffer with Proteinase K. Between 70 and 100ng of endosperm DNA obtained from crosses was subjected to bisulfite treatment using the Methylcode Bisulfite conversion kit (Invitrogen).

Analysis of cytosines from chloroplasts with at least ten sequenced reads showed a conversion rate of greater than 98% for all libraries. Bisulfite converted DNA was used to build libraries with the Pico Methyl-Seq library kit (Zymo Research, D5455). 7 cycles of amplification was used for library construction. All libraries were sequenced on the Illumina Hi-Seq 2500 (60bp paired-end).

Small RNA analysis

Small RNA reads were trimmed with fastq_quality_trimmer (*fastq_quality_trimmer -v -t 20 -l 25*). Cutadapt (Martin, 2011) was used to identify adapter bearing reads of suitable length (*cutadapt -a TGGAATTCTCGGGTGCCAAGG --trimmed-only --quality-base 64 -m 24 -M 40 --max-n 0.5 --too-long-output*). Taking advantage of the random nucleotides on the adapters in NEXTflex kits, we used Prinseq (prinseq-lite-0.20.4) (*prinseq-lite.pl -fastq <infile> -out_format 3 -out_good <filename> -derep 1 -log*) to remove PCR duplicates (Schmieder and Edwards, 2011). Filtered reads were aligned to a genome consisting of concatenated Col-0 TAIR10 and L*er* pseudo-genome (Col-0 genome substituted with L*er* SNPs) using Bowtie (v 1.2.2) *bowtie -v 2 --best -p 8 -5 4 -3 4 --sam <index file> <infile.fq>* (Langmead *et al.*, 2009). Reads mapping to L*er* were lifted over to Col-0 using custom scripts (Erdmann *et al.*, 2017). A custom script assign-to-allele was used to identify reads arising from Col-0 or L*er* alleles (https://github.com/clp90/imprinting_analysis/tree/master/helper_scripts). Aligned reads between 21 and 24nt in length were binned based on size. Bedtools was used to count reads in 300-bp windows with 200-bp overlaps and over annotated genes and TEs from Araport 11. DESeq2 (Love *et al*., 2014) was used to identify features showing differences in small RNA abundance. One complication with using DESeq2 is that the loss of Pol IV-dependent sRNAs at most loci in *nrpd1-/-* leads to an underestimation of wild-type library size by DESeq2, which increases the proportion of false negatives and undercounts the number of Pol IV-dependent sRNA loci. To allay this effect while analyzing genes, we excluded TEs and applied differential expression analysis to just genic and miRNA loci. These non-TE loci also included Pol IV-independent sRNA loci, which provide an estimate of library size. We separately examined TEs using genic sRNA counts to provide an estimate of library size. ShortStack version 3.8.5 (Axtell, 2013) was also used as an orthogonal approach to identify small RNA peaks from bam alignment file output from Bowtie. Parameters chosen for ShortStack included dicermin= 20, dicermax=25 and a mincov of 0.5 rpm.

mRNA-seq analysis

The reads from mRNA-seq were trimmed for quality with "*trim_galore -q 25 --phred64 --fastqc --length 20 --stringency 5*." aligned to the TAIR10 genome using Tophat (v2.1.1) (Kim *et al.*, 2013) using the command *tophat -i 30 -I 3000 --segment-mismatches 1 --segment-length 18 --b2-very-sensitive*. Cuffdiff (v2.1.1) (Trapnell *et al.*, 2013) was used to identify differentially expressed genes.

DNA methylation analysis

Reads from Bisulfite sequencing were trimmed for quality using Trim Galore (https://github.com/FelixKrueger/TrimGalore). Trimmed reads were aligned to the TAIR10 genome using Bismark (Krueger and Andrews, 2011) with parameters set to *-N 1 -L 20 --non_directional*. For this alignment, paired-end reads were treated as single reads. Previously described Bismark methylation extractor and custom scripts (Pignatta *et al.*, 2014; Pignatta *et al*., 2015) were used to determine DNA methylation/base and then methylation was calculated for 300 bp windows that overlapped by 200bp. To be included in analysis, windows needed to have at least three overlapping cytosines and a depth of 6 reads/cytosine. Windows that differed between genotypes by 10% CHH, 20% CHG or 30% CG DNA methylation were identified as differentially methylated. Overlapping windows with differential methylation between genotypes were merged into differentially methylated regions. To increase the robustness of our conclusions, we added two data filtering steps. DNA methylation in the endosperm varies between maternal and paternal alleles and bisulfite sequencing is known to potentially enrich for methylated DNA (Ji *et al.*, 2014). Since we were examining the consequences of loss of *NRPD1* in either parent, we could preferentially lose DNA methylation from one set of alleles. This could lead to lower coverage of one set of parental alleles and lead to faulty measurements of DNA methylation. We therefore limited our analyses to genomic regions in which reads arising from the maternally inherited genome accounted for 67%+/- 15% of total DNA reads (based on the fact that that 2/3 of the DNA in endosperm is maternally-inherited). Next, we identified DMRs between the two replicates for each genotype to mark regions where DNA methylation was variable within the same genotype. These regions were excluded from further analysis

Data Availability

All high-throughput sequencing data will be available in GEO at GSEXXXX.

Acknowledgements

We thank the Whitehead Institute Genome Technology Core for high-throughput sequencing services. This research was supported by NSF MCB 1453459 grant to M.G.

Competing Interests

The authors have no competing interests.

Supplemental Files

Figure S1: RNA Pol IV is necessary for the production of 21-24nt sRNAs in the endosperm.

Figure S2: Impact of the loss of maternal and paternal *NRPD1* on the endosperm small RNA population.

Figure S3: RNA Pol IV-dependent small RNAs arise from both maternal and paternal alleles.

Figure S4: Impact of parental *NRPD1* on maternal and paternal allele contributions to total gene expression.

Table S1: List of sequenced libraries.

Table S2: DESeq2 output for comparison of 21-24nt sRNA over genes between Wild-type and *nrpd1-/-*, maternal *nrpd1+/-*, paternal *nrpd1+/-.*

Table S3: DESeq2 output for comparison of 21-24nt sRNA over transposons between Wild-type and nrpd1-/-, maternal *nrpd1+/-*, paternal *nrpd1+/-.*

Table S4: Bedgraph for windows with differences in 21-24nt sRNA between Wild-type and nrpd1-/-, maternal *nrpd1+/-*, paternal *nrpd1+/-.*

Table S5: Bed files showing regions differentially methylated between wild-type, mat *nrpd1+/-* and pat *nrpd1+/-* endosperm.

Table S6: Cuffdiff output showing genes that are differentially expressed between wild-type, *nrpd1-/-*, mat *nrpd1+/-* and pat *nrpd1+/-.*

References

Al Adhami, H. *et al.* (2015) ‘A systems-level approach to parental genomic imprinting: The imprinted gene network includes extracellular matrix genes and regulates cell cycle exit and

differentiation', *Genome Research*, 25(3), pp. 353–367. doi: 10.1101/gr.175919.114.

Axtell, M. J. (2013) 'ShortStack: Comprehensive annotation and quantification of small RNA genes', *RNA*. doi: 10.1261/rna.035279.112.

Blevins, T. *et al.* (2015) 'Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis', *eLife*, 4. doi: 10.7554/eLife.09591.

Borg, M. *et al.* (2020) 'Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin.', *Nature cell biology*, 22(6), pp. 621–629. doi: 10.1038/s41556-020-0515-y.

Chaudhury, A. M. *et al.* (1997) 'Fertilization-independent seed development in Arabidopsis thaliana.', *Proceedings of the National Academy of Sciences of the United States of America*, 94(8), pp. 4223–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9108133.

Choi, Y. *et al.* (2002) 'DEMETER, a DNA Glycosylase Domain Protein, Is Required for Endosperm Gene Imprinting and Seed Viability in Arabidopsis', *Cell*, 110(1), pp. 33–42. doi: 10.1016/S0092-8674(02)00807-3.

Daxinger, L. *et al.* (2009) 'A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation', *EMBO Journal*. doi: 10.1038/emboj.2008.260.

Erdmann, R. M. *et al.* (2017) 'A Small RNA Pathway Mediates Allelic Dosage in Endosperm', *Cell Reports*. doi: 10.1016/j.celrep.2017.11.078.

Gehring, M., Missirian, V. and Henikoff, S. (2011) 'Genomic analysis of parent-of-origin allelic expression in arabidopsis thaliana seeds', *PLoS ONE*, 6(8). doi: 10.1371/journal.pone.0023687.

Gehring, M. and Satyaki, P. R. (2017) 'Endosperm and imprinting, inextricably linked', *Plant Physiology*. doi: 10.1104/pp.16.01353.

Grossniklaus, U. (1998) 'Maternal Control of Embryogenesis by MEDEA, a Polycomb Group Gene in Arabidopsis', *Science*, 280(5362), pp. 446–450. doi: 10.1126/science.280.5362.446.

Grover, J. W. *et al.* (2018) 'Maternal components of RNA-directed DNA methylation are required for seed development in *Brassica rapa*', *The Plant Journal*, 94(4), pp. 575–582. doi: 10.1111/tpj.13910.

Grover, J. W. *et al.* (2020) 'Abundant expression of maternal siRNAs is a conserved feature of seed development', *Proceedings of the National Academy of Sciences of the United States of*

America. doi: 10.1073/pnas.2001332117.

Haig, D. (2013) 'Kin conflict in seed development: an interdependent but fractious collective.', *Annual review of cell and developmental biology*, 29, pp. 189–211. doi: 10.1146/annurev-cellbio-101512-122324.

Herr, A. J. *et al.* (2005) 'RNA polymerase IV directs silencing of endogenous DNA', *Science*. doi: 10.1126/science.1106910.

Ji, L. *et al.* (2014) 'Methylated DNA is over-represented in whole-genome bisulfite sequencing data', *Frontiers in Genetics*. doi: 10.3389/fgene.2014.00341.

Kim, D. *et al.* (2013) 'TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions', *Genome Biology*. doi: 10.1186/gb-2013-14-4-r36.

Kirkbride, R. C. *et al.* (2019) 'Maternal small RNAs mediate spatial-temporal regulation of gene expression, imprinting, and seed development in Arabidopsis', *Proceedings of the National Academy of Sciences*, 116(7), pp. 2761–2766. doi: 10.1073/pnas.1807621116.

Kohler, C. (2003) 'Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development', *The EMBO Journal*, 22(18), pp. 4804–4814. doi: 10.1093/emboj/cdg444.

Krueger, F. and Andrews, S. R. (2011) 'Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications', *Bioinformatics*. doi: 10.1093/bioinformatics/btr167.

Langmead, B. *et al.* (2009) 'Bowtie: An ultrafast memory-efficient short read aligner. [http://bowtie.cbcb.umd.edu/]', *Genome biology*.

Lei, M. *et al.* (2015) 'Regulatory link between DNA methylation and active demethylation in Arabidopsis', *Proceedings of the National Academy of Sciences*, 112(11), pp. 3553–3557. doi: 10.1073/pnas.1502279112.

Li, J. *et al.* (2020) 'Epigenetic memory marks determine epiallele stability at loci targeted by de novo DNA methylation', *Nature Plants*. doi: 10.1038/s41477-020-0671-x.

Li, J. and Berger, F. (2012) 'Endosperm: Food for humankind and fodder for scientific discoveries', *New Phytologist*, 195(2), pp. 290–305. doi: 10.1111/j.1469-8137.2012.04182.x.

Li, Shaofang *et al.* (2015) 'Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis', *Genome Research*, 25(2),

pp. 235–245. doi: 10.1101/gr.182238.114.

Love, M. I., Huber, W. and Anders, S. (2014) ‘Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2’, *Genome Biology*. doi: 10.1186/s13059-014-0550-8.

Martin, M. (2011) ‘Cutadapt removes adapter sequences from high-throughput sequencing reads’, *EMBnet.journal*. doi: 10.14806/ej.17.1.200.

Martinez, G. *et al.* (2018) ‘Paternal easiRNAs regulate parental genome dosage in Arabidopsis’, *Nature Genetics*. doi: 10.1038/s41588-017-0033-4.

Matzke, M. A. and Mosher, R. A. (2014) ‘RNA-directed DNA methylation: an epigenetic pathway of increasing complexity’, *Nature Reviews Genetics*, 15(6), pp. 394–408. doi: 10.1038/nrg3683.

Milbocker, D. and Sink, K. (1969) ‘EMBRYOLOGY OF DIPLOID × DIPLOID AND DIPLOID × TETRAPLOID CROSSES IN POINSETTIA’, *Canadian Journal of Genetics and Cytology*. doi: 10.1139/g69-070.

Moreno-Romero, J. *et al.* (2016) ‘Parental epigenetic asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm’, *The EMBO Journal*, 35(12), pp. 1298–1311. doi: 10.15252/embj.201593534.

Mott, R. *et al.* (2014) ‘The Architecture of Parent-of-Origin Effects in Mice’, *Cell*, 156(1–2), pp. 332–342. doi: 10.1016/j.cell.2013.11.043.

Müntzing, A. (1936) ‘The Evolutionary Significance of Autopolyploidy’, *Hereditas*. doi: 10.1111/j.1601-5223.1936.tb03204.x.

Ohad, N. *et al.* (1996) ‘A mutation that allows endosperm development without fertilization.’, *Proceedings of the National Academy of Sciences*, 93(11), pp. 5319–5324. doi: 10.1073/pnas.93.11.5319.

Onodera, Y. *et al.* (2005) ‘Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation- dependent heterochromatin formation’, *Cell*. doi: 10.1016/j.cell.2005.02.007.

Panda, K., McCue, A. D. and Slotkin, R. K. (2020) ‘Arabidopsis RNA Polymerase IV generates 21–22 nucleotide small RNAs that can participate in RNA-directed DNA methylation and may regulate genes’, *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2019.0417.

Picelli, S. *et al.* (2014) ‘Full-length RNA-seq from single cells using Smart-seq2’, *Nature*

Protocols, 9(1), pp. 171–181. doi: 10.1038/nprot.2014.006.

Pignatta, D. *et al.* (2014) 'Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting', *eLife*, 3. doi: 10.7554/eLife.03198.

Pignatta, D., Bell, G. and Gehring, M. (2015) 'Whole Genome Bisulfite Sequencing and DNA Methylation Analysis from Plant Tissue', *BIO-PROTOCOL*, 5(4). doi: 10.21769/BioProtoc.1407.

Povilus, R. A., Diggle, P. K. and Friedman, W. E. (2018) 'Evidence for parent-of-origin effects and interparental conflict in seeds of an ancient flowering plant lineage', *Proceedings of the Royal Society B: Biological Sciences*, 285(1872), p. 20172491. doi: 10.1098/rspb.2017.2491.

Satyaki, P. R. V. and Gehring, M. (2019) 'Paternally acting canonical RNA-directed DNA methylation pathway genes sensitize arabidopsis endosperm to paternal genome dosage', *Plant Cell*, 31(7), pp. 1563–1578. doi: 10.1105/tpc.19.00047.

Schmieder, R. and Edwards, R. (2011) 'Quality control and preprocessing of metagenomic datasets', *Bioinformatics*. doi: 10.1093/bioinformatics/btr026.

Scott, R. J. *et al.* (1998) 'Parent-of-origin effects on seed development in Arabidopsis thaliana', *Development*, 125(17), pp. 3329–3341. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9693137.

Stoute, A. I. *et al.* (2012) 'Parental genome imbalance in Brassica oleracea causes asymmetric triploid block', *Plant Journal*. doi: 10.1111/j.1365-313X.2012.05015.x.

Stroud, H. *et al.* (2013) 'Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome', *Cell*, 152(1–2), pp. 352–364. doi: 10.1016/j.cell.2012.10.054.

Studitsky, V. M. *et al.* (2004) 'Chromatin remodeling by RNA polymerases', *Trends in Biochemical Sciences*. doi: 10.1016/j.tibs.2004.01.003.

Trapnell, C. *et al.* (2013) 'Differential analysis of gene regulation at transcript resolution with RNA-seq', *Nature Biotechnology*. doi: 10.1038/nbt.2450.

Vu, T. M. *et al.* (2013) 'RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis.', *Development (Cambridge, England)*, 140(14), pp. 2953–60. doi: 10.1242/dev.092981.

Wang, Z. *et al.* (2020) ‘Polymerase IV Plays a Crucial Role in Pollen Development in Capsella’, *The Plant Cell*, 32(4), pp. 950–966. doi: 10.1105/tpc.19.00938.

Williams, B. P. *et al.* (2015) ‘Methylation-Sensitive Expression of a DNA Demethylase Gene Serves As an Epigenetic Rheostat’, *PLOS Genetics*. Edited by O. Mittelsten Scheid, 11(3), p. e1005142. doi: 10.1371/journal.pgen.1005142.

Wu, W. and Zheng, B. (2019) ‘Intercellular delivery of small RNAs in plant gametes’, *New Phytologist*. doi: 10.1111/nph.15854.

Xiao, W. *et al.* (2006) ‘Regulation of seed size by hypomethylation of maternal and paternal genomes’, *Plant Physiology*. doi: 10.1104/pp.106.088849.

Ye, R. *et al.* (2016) ‘A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in Arabidopsis’, *Molecular Cell*. doi: 10.1016/j.molcel.2015.11.015.

Zhai, J. *et al.* (2015) ‘A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis’, *Cell*, 163(2), pp. 445–455. doi: 10.1016/j.cell.2015.09.032.