Paternally-acting canonical RNA-directed DNA methylation pathway genes sensitize Arabidopsis endosperm to paternal dosage

P.R. V. Satyaki[1], Mary Gehring[1,2*]

[1]Whitehead Institute for Biomedical Research, [2]Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02142

* author for correspondence: mgehring@wi.mit.edu

Abstract

Seed development is sensitive to parental dosage, with excess maternal or paternal genomes creating reciprocal phenotypes. Paternal genomic excess results in extensive endosperm proliferation without cellularization and eventual seed abortion. We previously showed that loss of the *RNA POL IV* gene *nrpd1* in tetraploid fathers represses seed abortion in paternal excess crosses. Here we show genetically that RNA-directed DNA methylation (RdDM) pathway activity in the paternal parent is sufficient to determine the viability of paternal excess seeds. The status of the RdDM pathway in paternal excess endosperm does not impact seed viability. Comparison of endosperm transcriptomes, DNA methylation, and small RNAs from balanced and paternal excess endosperm demonstrates that paternal excess seed abortion is unlikely to be dependent on either transposable element or imprinted gene mis-regulation. We suggest instead that loss of paternal RdDM modulates expression at a small subset of genes and desensitizes endosperm to paternal excess. Finally, using allele-specific transcription data, we present evidence of a transcriptional buffering system that up-regulates maternal alleles and represses paternal alleles in response to excess paternal genomic dosage. These findings prompt reconsideration of models for dosage sensitivity in endosperm.

Introduction

The endosperm of flowering plants is an essential tissue for seed viability. It is most commonly triploid, formed when the diploid central cell is fertilized by a haploid sperm. In many flowering plants, endosperm proceeds through the early phases of proliferative development as a syncytium before differentiating into three sub-types: micropylar, peripheral and chalazal endosperm (Li and Berger, 2012). Cellularization is a critical step in endosperm development, after which cell division slows and eventually ceases (Hehenberger et al., 2012; Li and Berger, 2012). In some species, like Arabidopsis, the endosperm is almost completely degraded at seed maturation as nutrients are assimilated and stored in the embryo (Li and Berger, 2012), whereas in other species, like grasses, it is persistent and mobilized later during seed germination. The endosperm serves several important functions in addition to mediating resource transfer from the mother to the growing embryo or germinating seedling (Li and Berger, 2012). Endosperm signals to the seed coat to promote its proliferation and allow accommodation of the growing offspring (Figueiredo et al., 2016). It is required for embryonic development and influences seed dormancy and germination (Fiume and Fletcher, 2012; Piskurewicz et al., 2016; Yan et al., 2014). Through these activities, endosperm influences seed size.

Balance between the maternal and paternal genomes in endosperm is important for normal endosperm development and, consequently, seed development and viability. Violations of the 2:1 maternal:paternal genome ratio lead to developmental defects, although the extent of violations tolerated varies widely across species (Cooper, 1951; Esen and Soost, 1973; Håkansson and Ellerström, 1950; Milbocker and Sink, 1969; Muntzing, 1936; Povilus et al., 2018; Scott et al., 1998; Stoute et al., 2012). In Arabidopsis, increased maternal genome dosage (maternal excess) leads to premature cellularization of the endosperm and the formation of smaller seeds (Scott et al., 1998). By contrast, increased paternal genome dosage (paternal excess) in crosses between diploid mothers and hexaploid fathers leads to a failure of endosperm cellularization, prolonged cell proliferation, and seed abortion (Scott et al., 1998). However, there is intra-specific variation in the levels of seed abortion observed in paternal excess

crosses. For example, whereas tetraploid Col-0 induces seed abortion when pollinating diploid mothers, tetraploid C24 and Cvi do not (Dilkes et al., 2008; Lu et al., 2012; Piskurewicz et al., 2016; Scott et al., 1998). These observations leave open the question of what determines the threshold between seed lethality and viability in paternal excess crosses.

Components in the endosperm, in the gametophytes, or in the parental sporophyte have been proposed to be responsible for endosperm dosage sensitivity in interploid crosses. Some of the elements that determine the critical threshold may be linked to interactions between the endosperm and maternal genotype. In one model, the imbalance is between the paternal dose and the female gametophyte (Birchler, 2014; von Wangenheim and Peterson, 2004). A second model involves interactions between the paternal excess endosperm and the diploid seed coat inherited from the diploid mother (Muntzing, 1936). The ability of maternal sporophytic mutations in the flavonoid pathway to repress paternal excess seed abortion is consistent with this model (Dilkes et al., 2008; Doughty et al., 2014).

Other models suggest that endosperm ploidy incompatibilities are caused by epigenetic abnormalities. In normal triploid endosperm, the maternal and paternal alleles of some genes are regulated epigenetically. Maternal chromosomes are DNA hypomethylated relative to paternal chromosomes due to the activity of the DNA demethylase *DME* (Gehring et al., 2009; Hsieh et al., 2009; Ibarra et al., 2012). In Arabidopsis, DNA hypomethylation is typically found at sequences derived from transposable elements (TEs) in euchromatin (Gehring et al., 2009; Ibarra et al., 2012). In other tissues, DNA methylation is established at these sequences through the RNA-directed DNA methylation (RdDM) pathway. During this process, short non-coding RNAs generated by RNA POL IV are converted into double-stranded RNAs by RDR2. These are then subsequently diced by DCL3 into 24 nt small RNAs (sRNAs), loaded into an ARGONAUTE complex (usually AGO4 or AGO6), and then interact with a non-coding RNA transcribed by another polymerase, RNA POL V, to direct DNA methylation by the *de novo* methyltransferase DRM2. In the endosperm, DNA methylation represses expression of some genes and promotes the expression of others (Satyaki and Gehring,

2017). Another important epigenetic regulatory pathway in endosperm is H3K27 trimethylation, mediated by POLYCOMB REPRESSIVE COMPLEX 2 (PRC2). H3K27me3 is biased toward maternal alleles and associated with their repression (Moreno-Romero et al., 2016).

It has been proposed that a cause of interploid seed abortion is the mis-regulation of transposable elements (TEs) due to epigenetic alterations (Martienssen, 2010). Under this untested model, a maternally deposited sRNA dose is insufficient to silence the doubled number of paternally inherited TEs. Another model is influenced by the parental conflict theory (Haig and Westboy, 1991). The theory argues that in a polyandrous system in which the mother provisions resources for her offspring from various fathers, genes restricting resource allocation to any one seed and favoring equitable distribution of resources across all the progeny become predominantly maternal in expression (maternally expressed genes or MEGs). On the other hand, genes promoting resource allocation and larger seed production are predominantly paternally expressed (paternally expressed genes or PEGs). This type of allele-specific expression is an epigenetic phenomenon referred to as gene imprinting. The maternal alleles of PEGs are often associated with DNA hypomethylation in flanking sequences and H3K27me3 and 24 nt small RNAs over coding regions (Erdmann et al., 2017; Moreno-Romero et al., 2016). According to these ideas, an excess dose of paternal chromosomes leads to over-expression of PEGs, genes that promote endosperm proliferation, eventually leading to seed abortion. Consistent with this model, PEGs were reported to be over-expressed in paternal excess seeds (Kradolfer et al., 2013; Schatlowski et al., 2014) and multiple PEG mutants repress seed abortion in conditions of paternal genomic excess (Huang et al., 2017; Jiang et al., 2017; Kradolfer et al., 2013; Wolff et al., 2015).

We recently proposed another model based on discoveries about the function of *RNA POL IV* in normal triploid endosperm (Erdmann et al., 2017). Expression of 24 nt sRNAs in *Arabidopsis thaliana* endosperm is paternally biased overall and concentrated in pericentromeric heterochromatin, but a subset of sRNAs are associated with the maternal alleles of genes (Erdmann et al., 2017). We showed that loss of paternal

NRPD1 (encoding the largest subunit of RNA POL IV) in balanced triploid endosperm increased the maternal fraction of the transcriptome, suggesting that normally *RNA POL IV* represses maternal genome dosage, probably via the production of 24 nt small RNAs. We also found that paternally inherited mutations in *NRPD1* repress seed abortion caused by excess paternal genomes (i.e. 2N Col x 4N Col), leading to the hypothesis that the increased maternal fraction of the transcriptome in *nrpd1* mutant endosperm compensates for increased paternal genomic dosage. In this model, the loss of *RNA POL IV*-dependent sRNA production in the endosperm is essential for viability (Erdmann et al., 2017).

Finally, the "easiRNA" model argues that *POL IV* and *RDR6* act together in a non-canonical pathway in the male gametophyte to produce easiRNAs whose concentration scales with ploidy (Borges et al., 2018; Martinez et al., 2018). easiRNAs are defined as 21-22 nt sRNAs produced from the processing of TE mRNAs (Creasey et al., 2014). It is proposed that these easiRNAs are transmitted during fertilization from sperm to the central cell, where they inhibit the *POL IV-RDR2* complex from making canonical 24 nt sRNAs in the developing endosperm (Borges et al., 2018; Martinez et al., 2018). In this model, easiRNAs are lost from sperm when the paternal copy of *NRPD1* is mutated. The loss of paternal easiRNAs allows the restoration of a functional RdDM pathway in endosperm using maternal copies of *NRPD1*. The restored RdDM pathway was proposed to repress excess paternal dosage and restore seed viability.

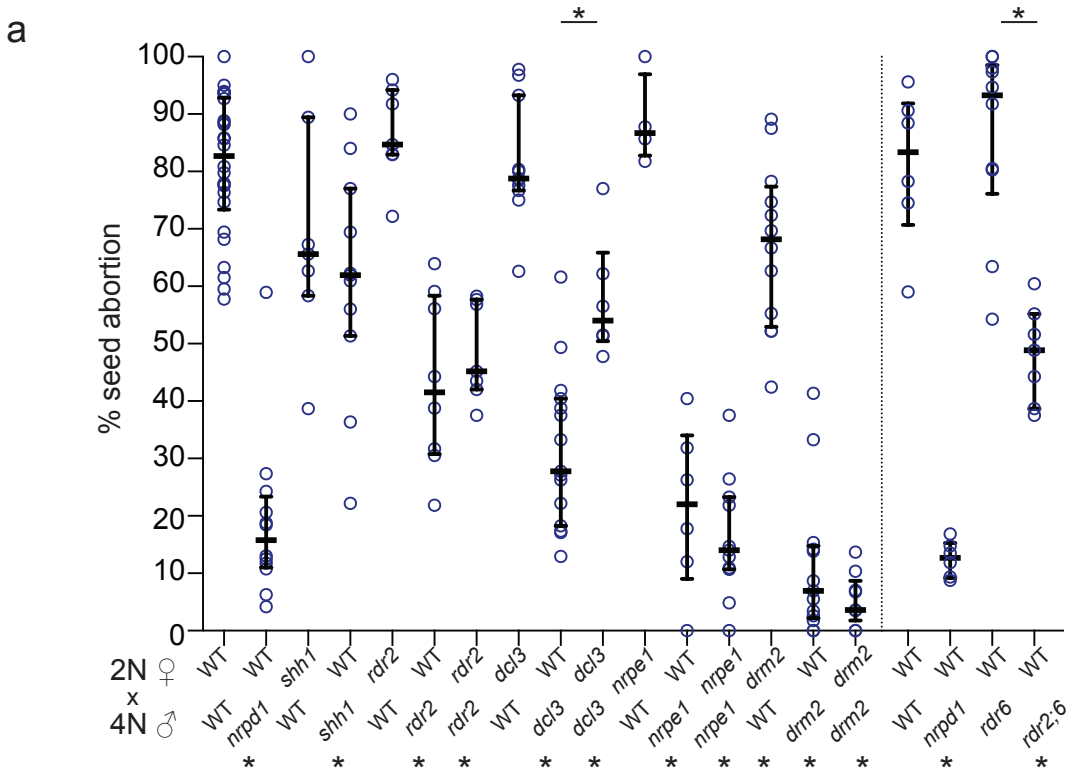
To test among these models and further understand the nature of intolerance or tolerance to paternal genomic excess, we tested the genetic and molecular contributions to interploidy seed abortion and repression. We created tetraploid mutants for members of the canonical and non-canonical small RNA pathways to test the genetic requirements for seed abortion in paternal excess crosses. We also profiled sRNAs, DNA methylation, and gene expression in balanced (2Nx2N), lethal paternal excess (2Nx4N), and viable paternal excess (2Nx4N *nrpd1*) endosperm. In contrast to previous reports, we found that genes of the canonical RdDM pathway were necessary in the male parent for seed abortion induced by paternal genomic excess. We also observed that the status of the canonical RdDM pathway in paternal excess endosperm

is not a determinant of seed viability. Through transcriptomic profiling we found that there were extensive gene and transposable element expression changes associated with lethal paternal genomic excess but that only a small fraction were ameliorated in viable paternal excess endosperm. Further, allele-specific analyses showed that an excess dose of the paternal genome caused mis-regulation of both maternal and paternal allele transcription, with differential affects at specific loci. This analysis also revealed the signatures of a potential buffering system that attempts to rebalance transcription in conditions of paternal excess by repressing paternal alleles and activating maternal alleles. Our observations show that endosperm development can tolerate surprisingly large variations in gene expression and forces a reassessment of what factors contribute to the dosage that determines seed viability

Results

Paternal loss of multiple RdDM pathway genes is sufficient for suppression of interploid seed abortion

In crosses between diploid females and tetraploid males, tetraploid *nrpd1* mutant fathers repress seed abortion, whereas diploid *nrpd1* mothers have no effect (Erdmann et al., 2017; Martinez et al., 2018). We tested if other members of the canonical RdDM pathway behaved genetically in the same manner. Using colchicine-induced tetraploid mutants of *shh1, rdr2, dcl3, nrpe1,* and *drm2,* we examined seeds obtained from crosses between 2N wild-type (Col) mothers and 4N (Col) fathers that were either wild-type or mutant for one of the RdDM pathway components (Figure 1). Mature seeds were scored as normal, aborted, or abnormal and multiple independent crosses were analyzed. Typically, seed abortion in each set of crosses varied by between 20 and 30%, regardless of whether parents were wild-type or mutant (Figure 1A). For example, in wild-type paternal excess crosses (2N Col-0 x 4N Col-0) seed abortion ranged from 57.8%-100%, with the mean seed abortion at 81.1%. Therefore, we considered only those mutants capable of substantially enhancing seed viability across multiple crosses as being true repressors of seed abortion.



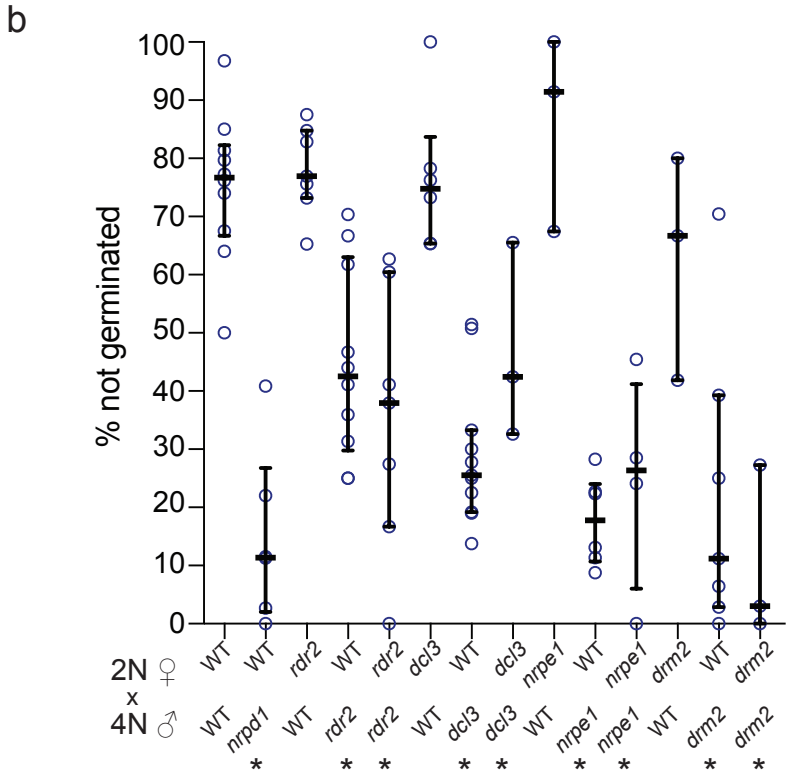
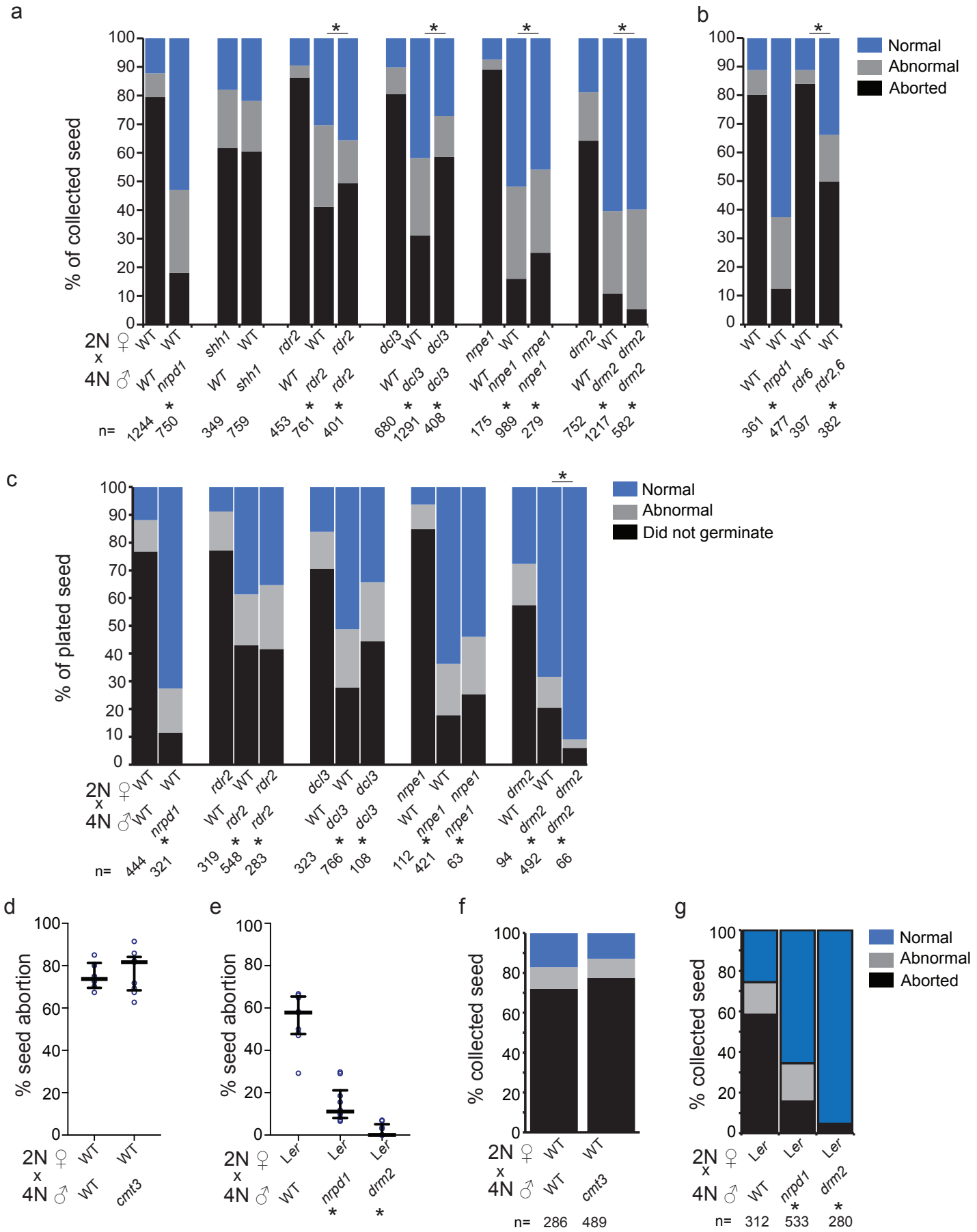


Figure 1: Loss of paternal RdDM genes but not *RDR6* represses seed abortion in paternal excess crosses.
a) Each circle represents seed abortion rate in one cross and represents multiple siliques from a single inflorescence. **b)** Each circle represents the percent of seeds that failed to germinate from each scored collection of seed. Failure to germinate was defined as the inability to produce either a radicle or a hypocotyl. Bars show median and interquartile range. * at bottom represents statistically significant difference ($p<0.05$) in comparisons between indicated cross and cross between wild-type (WT) diploid (2N) Col-0 mothers and wild-type tetraploid (4N) Col-0 fathers. * at top represents statistically significant differences ($p<0.05$) between crosses indicated. Statistical significance calculated by Wilcoxon test.



Supplemental Figure 1: Loss of paternal RdDM genes but not *RDR6* and *CMT3* represses seed abortion in paternal excess crosses. a,b,f,g) Histograms depicting percentage of seed collected that was either aborted, abnormal or normal. **c)** Histograms depicting the germination status of collected seed. Seeds grouped under "normal" produced both true leaves and roots. Abnormally germinating seeds germinated but failed to produce a radicle or had abnormal cotyledons. Seeds that failed to produce either a radicle or a hypoctyl are grouped under "did not germinate. **d)** Loss of CMT3 does not repress paternal excess lethality. Wilcox test did not show statistically significant differences between crosses in which the tetraploid father is either wild-type or *cmt3* mutant. **e)** Interploidy crosses with Ler mothers are similar to crosses with Col-0 mothers. Wilcox test showed statistically significant differences between crosses. For **d** and **e**, circles represent a single cross and includes multiple siliques from a single infloroscence. Bars represent median and interquartile range. * at bottom represents statistically significant difference in seed abortion or germination ($p<0.05$) in comparisons between indicated cross and cross between wild-type diploid mothers and wild-type tetraploid fathers.
Statistical significance calculated by Chi-Squared test. N represents number of seeds/seedlings counted in each cross. WT = Col-0.

All paternally inherited RdDM pathway mutations, except *shh1*, substantially repressed seed abortion and promoted the production of fully developed seeds capable of germination (Figure 1, Supplemental Figure 1). SHH1 is required for RNA POL IV recruitment at a subset of its target sites (Law et al., 2013), suggesting that SHH1-independent activity of RNA POL IV is important for seed abortion. Mutations in *nrpe1* (encoding the largest subunit of RNA POL V) and *drm2* mirrored *nrpd1* in their substantial repression of seed abortion; mean seed abortion among examined seeds was 18.9% for *nrpd1,* 21.4% for *nrpe1* and 11.3% for *drm2* (Figure 1A, Supplemental Figure 1). *rdr2* and *dcl3* mutations repressed seed abortion to a lesser extent, suggesting that other *RDR* and *DCL* paralogs are partially redundant with *RDR2* and *DCL3*, as has been previously shown (Figure 1, Supplemental Figure 1C) (Gasciolli et al., 2005; Stroud et al., 2013). For all mutants the seed abortion ratios were reflected in the percentage of seeds that germinated on plates (Figure 1B, Supplemental Figure 1). Tetraploid *drm2* and *nrpd1* mutants also suppressed interploidy seed abortion when crossed with wild-type diploid L*er* mothers (Erdmann et al., 2017) (Supplemental Figure 1E,G), indicating the effect is not specific to Col mothers. RNA POL IV has also been proposed to function in partnership with RDR6 in a non-canonical RdDM pathway (Martinez et al., 2018). We therefore tested if tetraploid fathers mutant for *rdr6* could suppress paternal excess seed abortion or if tetraploid *rdr2 rdr6* double mutants additively suppressed seed abortion in paternal excess crosses (Figure 1A, Supplemental Figure 1B,G). A paternally inherited *rdr6* mutation did not repress seed abortion and did not act additively with *rdr2* (Figure 1A, Supplemental Figure 1G). We also tested whether mutations in the CHG methyltransferase *CMT3* could suppress interploidy seed abortion when inherited through the male – no effect was observed (Supplemental Figure 1D,F).

In contrast to repression of seed abortion by paternally inherited loss of function mutations in RdDM pathway genes, most loss-of-function mutations inherited through the diploid mother did not repress seed abortion in paternal excess crosses with wild-type tetraploid fathers (Figure 1, Supplemental Figure 1A-C). An exception was maternal loss of *DRM2*, which resulted in a statistically significant repression in seed abortion, although the magnitude of the repression was small (mean percentage of

aborted seed was 66.9% for *drm2* compared to 81.1% for WT) and did not phenocopy the extensive reduction in seed abortion observed upon the loss of paternal RdDM pathway components (Figure 1, Supplemental Figure 1A-C). It was previously reported that suppression of paternal excess seed abortion also occurs when both parents are mutant for *NRPD1* (Erdmann et al., 2017; Martinez et al., 2018). We observed the same effect for all other genes tested: *rdr2*, *dcl3*, *nrpe1*, and *drm2* (Figure 1, Supplemental Figure 1A-C). In sum, these genetic results indicate 1) that paternal loss of the canonical RdDM pathway is sufficient to suppress seed abortion caused by paternal genomic excess and 2) that a genetically complete canonical RdDM pathway is not required in endosperm itself for suppression of seed abortion.

Massive gene mis-regulation in lethal and viable paternal excess endosperm

To determine what genes or processes were associated with interploidy seed lethality and its genetic suppression, we examined the transcriptional effects of doubling paternal dosage in the endosperm. We performed mRNA-seq to high depth on endosperm from three biological replicates of balanced crosses (L*er* x Col), lethal paternal excess (L*er* x 4N Col) and viable paternal excess (L*er* x 4N Col *nrpd1*) (Figure 2, Supplemental Table 1). Surprisingly, comparisons of the three transcriptomes by principal component analysis indicated that lethal and viable paternal excess endosperm were more similar to each other than to balanced endosperm (Figure 2A). About a third of the transcriptome was significantly mis-expressed (2 fold or greater at $q < 0.05$) in lethal paternal-excess endosperm compared to balanced endosperm: 4,054 genes were more highly expressed and 3,855 genes decreased in expression (Figure 2B, Supplemental Dataset 1). GO analyses (Supplemental Dataset 2) showed that genes with lower transcript abundance in lethal paternal excess relative to balanced endosperm were enriched for those encoding light harvesting proteins, proteins in glucose and starch metabolism, hormone responses, response to abiotic stimuli, and genes involved in cell wall organization. Whereas the predicted consequences of many of these changes remain unclear, the decreased expression of cell wall genes is consistent with the failure in endosperm cellularization that has been previously described in lethal paternal excess crosses (Wolff et al., 2015). GO analyses of up-

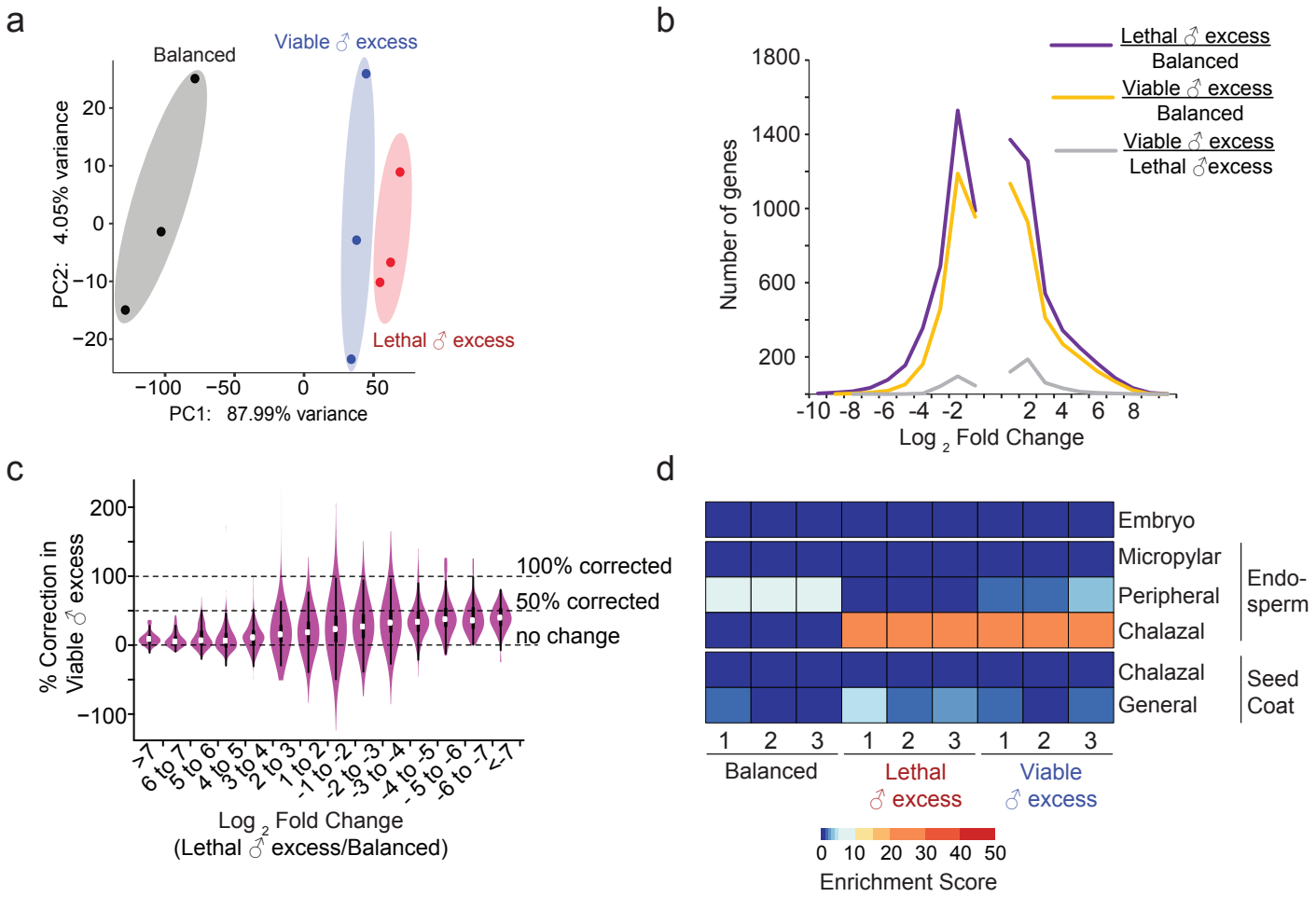


Figure 2: Lethal and viable paternal excess endosperm are transcriptionally more similar to each other than to balanced endosperm. a) PCA plot of read counts for genes from biological replicate mRNA-Seq samples. **b)** Plot of number of genes differentially expressed in comparisons of balanced endosperm with both lethal (purple) and viable (yellow) ♂ excess endosperm. Only 614 genes were differentially expressed between viable and lethal ♂ excess endosperm (gray). **c)** Correction value in viable paternal excess endosperm for each gene that was called as being significantly differentially expressed in comparisons of balanced and lethal ♂ excess endosperm. The value was calculated as % Correction= 100-(((log$_2$ (Viable/Balanced)/log$_2$ (Lethal/Balanced))*100). A value of 100% indicates that the gene, which was mis-regulated in lethal ♂ excess, was not differentially expressed in viable ♂ excess relative to balanced endosperm. A value of 0% represents similar mis-regulation in both lethal and viable ♂ excess relative to balanced endosperm. Fold change values and significance for fold-change for **b** and **c** were calculated using CUFFDIFF. **d)** Lethal paternal excess endosperm was enriched for chalazal endosperm gene expression; viable paternal excess endosperm showed both chalazal and peripheral markers. Tissue enrichment for each biological replicate is shown.

Supplemental Table 1: Details of sequencing libraires

RNA-Seq

Genotype	Mapped reads	Mapping efficiency
Lethal paternal excess replicate 1	123133930	91.60%
Lethal paternal excess replicate 2	116724380	92.30%
Lethal paternal excess replicate 3	105331054	89.30%
Viable paternal excess replicate 1	88644757	92%
Viable paternal excess replicate 2	99177965	92.70%
Viable paternal excess replicate 3	93705685	81.70%

small RNA sequencing

Genotype	Reads remaining after processing	Mapped reads	Mapping efficiency
Lethal paternal excess replicate 1	4636843	3777633	81.47%
Lethal paternal excess replicate 2	15358892	13311163	86.67%
Viable paternal excess replicate 1	19840772	17098180	86.18%
Viable paternal excess replicate 2	20129312	17518163	87.03%

Methylome

Genotype	Mapped read pairs (unique only)	Mapping efficiency*	Reads remaining after deduplication by bismark
Balanced replicate 1	9495700	41.70%	7096231
Balanced replicate 2	14406400	44%	10176085
Lethal paternal excess replicate 1	10075611	49.60%	6850969
Lethal paternal excess replicate 2	14320244	49.50%	9505864
Viable paternal excess replicate 1	20027548	46.80%	13152774
Viable paternal excess replicate 2	12051462	44.30%	8589294

Note that biological replicates for the methylomes were merged for analysis

** Mapping efficiency refers to % of uniquely mapping paired end reads*

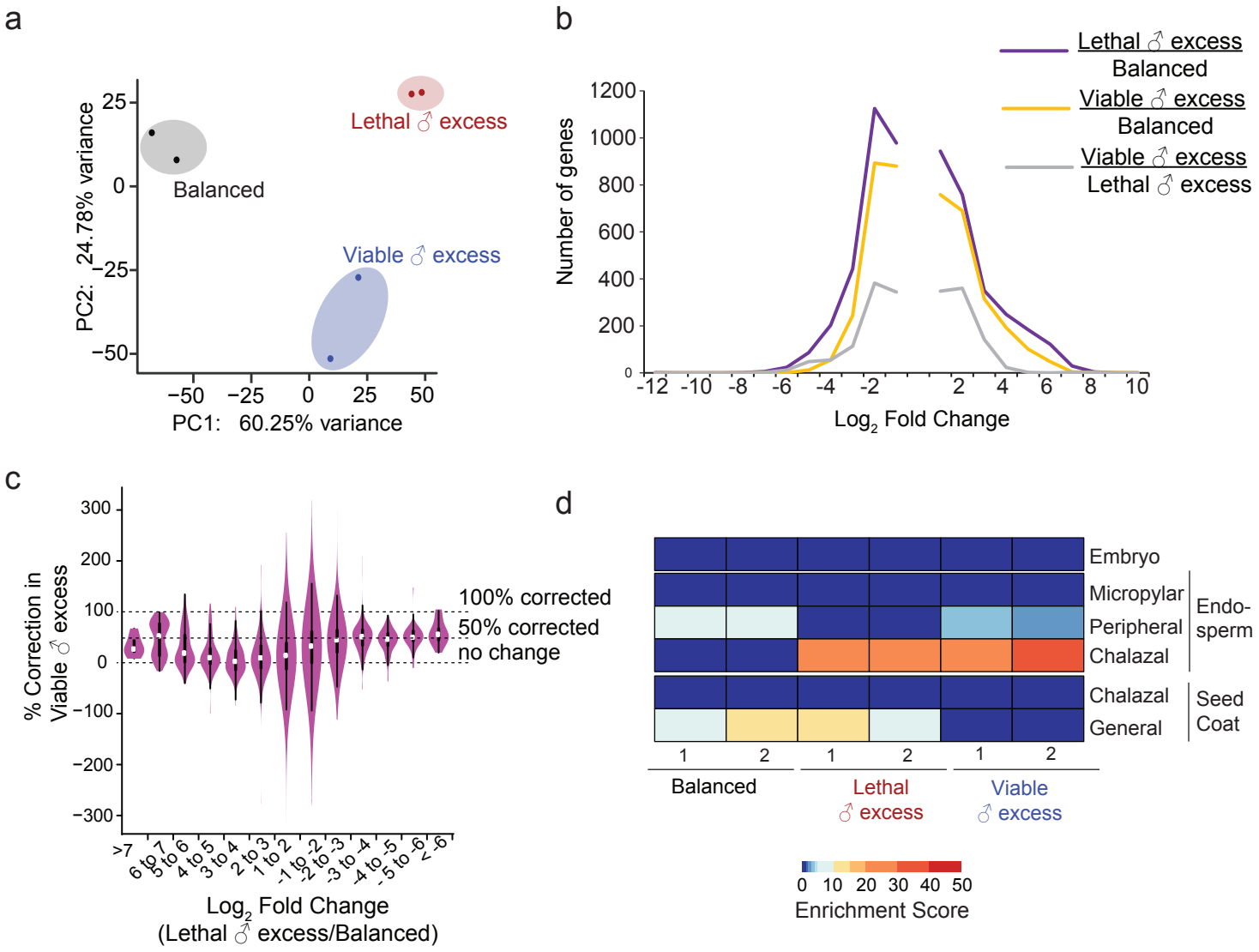
regulated genes identified enrichment for genes encoding proteins involved in protein deneddylation, ribosome biogenesis, DNA replication, chromosome segregation, and cell-cycle genes. The increased expression of these genes is consistent with increased cell proliferation in paternal excess endosperm (Scott et al., 1998; Tiwari et al., 2010). Compared to balanced endosperm, viable paternal excess endosperm also showed a similar quantitative change in gene expression: 3150 genes increased in expression and 2845 decreased (Figure 2B, Supplemental Dataset 1). Many of same genes were mis-regulated in viable and lethal paternal excess endosperm compared to balanced endosperm. Compared with lethal paternal excess, only 188 genes had reduced transcript abundance in viable paternal excess endosperm and 426 genes were more highly expressed (Figure 2B, Supplemental Dataset 1). To further test if the viability of paternal excess seeds was predicated on the transcriptome coming nearer to the dosage of balanced endosperm, we examined the extent to which gene expression was "corrected" in viable paternal excess endosperm (Figure 2C). Most genes with increased expression in lethal paternal excess endosperm were not corrected, but a small subset of genes with decreased expression in lethal paternal excess were moderately corrected in viable paternal excess (Figure 2C).

Several important developmental regulators were differentially expressed among balanced endosperm and lethal and viable paternal excess endosperm. The abundance of *CLE8*, a gene encoding an inhibitor of endosperm differentiation (Fiume and Fletcher, 2012), was highest in lethal paternal excess endosperm, followed by viable paternal excess, and then balanced endosperm (Supplemental Dataset 1). The *MINI3-CKX2* gene circuit influences seed size (Li et al., 2013); both genes exhibited increased expression in paternal excess endosperm (Supplemental Dataset 1). The expression of these genes remained elevated in viable paternal excess endosperm relative to balanced endosperm but had lower abundance relative to lethal paternal excess endosperm (Supplemental Dataset 1). *AGL*s, which encode MADS Box transcription factors, were mis-regulated in lethal and viable paternal excess endosperm (Supplemental Dataset 1). *AGL62*, a known inhibitor of cellularization (Kang et al., 2008), was expressed 32-fold higher in lethal paternal excess endosperm relative to balanced endosperm and 4-fold lower in viable paternal excess relative to lethal

paternal excess (Supplemental Dataset 1). The mis-regulation of *CKX2*, *MINI3* and the *AGLs* are consistent with gene expression data from whole siliques or seeds derived from paternal excess crosses (Kradolfer et al., 2013; Tiwari et al., 2010).

We further analyzed the transcriptomic data using a tissue enrichment tool to assess if the abundances of markers enriched in chalazal, peripheral and micropylar endosperm were altered in paternal excess endosperm (Schon and Nodine, 2017). Paternal excess endosperm adopted a gene expression program characteristic of chalazal endosperm (Figure 2D), consistent with phenotypic analysis of lethal paternal excess seeds (Martinez et al., 2018; Scott et al., 1998; Wolff et al., 2015). Viable paternal excess endosperm transcriptomes also showed increased chalazal endosperm marker gene expression. However, there was also slightly elevated peripheral endosperm marker gene expression relative to lethal paternal excess endosperm. These results suggest that lethal and viable paternal excess endosperm differ similarly from balanced endosperm in gene expression and subsequent developmental programs.

We also reanalyzed published endosperm transcriptome data from lethal and viable paternal excess endosperm generated using *osd1* and *osd1 nrpd1* mutations (Martinez et al., 2018). Consistent with our data, we found that both L*er* x Col *osd1* (lethal paternal excess) and L*er* x Col *osd1 nrpd1* (viable paternal excess) showed elevated expression of chalazal markers (Supplemental Figure 2D). Endosperm from L*er* x Col *osd1 nrpd1* also showed extensive genic mis-regulation (Supplemental Figure 2B). However, the percentage of the transcriptome that was mis-regulated was lower than in our viable paternal excess endosperm datasets (Figure 2B). Additionally, the extent of gene expression correction in viable paternal excess endosperm was also lower in our datasets compared to theirs (Figure 2C, Supplemental Figure 2C). These discrepancies could stem from multiple differences between our experiments. We created paternal excess endosperm using tetraploid fathers while Martinez et al. created paternal excess via the use of the *osd1* mutation, which causes omission of the second meiotic division and thus generates diploid pollen. The *osd1* mutation was backcrossed into Col from another accession (d'Erfurth et al., 2009; Martinez et al.,



Supplemental Figure 2: Analysis of gene expression differences in ♂ excess endosperm derived using the *osd1* mutation (data from Martinez et al., 2018). a) PCA plot shows that the transcriptomes of L*er* x Col *osd1* (lethal paternal excess) and L*er* x Col *osd1 nrpd1* (viable paternal excess) are more similar to each other than to L*er* x Col (balanced).
b) Frequency distribution of all gene expression changes between balanced, lethal ♂ excess and viable ♂ excess. Fold change was calculated using CUFFDIFF. All genes whose output FC was greater than 1 or less than -1 shown.
c) Correction value was calculated for genes that are statistically significantly different in comparisons of balanced and lethal ♂ excess endosperm. The value, calculated as % Correction= 100-((log$_2$ (viable/balanced))/(log$_2$ (lethal/balanced))*100). Value of 100% indicates that a gene misregulated in lethal ♂ excess is not misregulated in viable ♂ excess. 0% represents similar misregulation in both lethal and viable ♂ excess. Fold change values and significance were calculated using CUFFDIFF.
d) Both L*er* x Col *osd1* and L*er* x Col *osd1 nrpd1* endosperm are enriched for chalazal endospermmarker gene expression. Analysis was performed using the tissue enrichment test (Schon and Nodine, 2017) on gene expression measured by the HT-Seq count command. Tissue enrichment per biological replicate shown.

2018) and interploidy cross seed abortion is sensitive to the genetic background of both parents (Lu et al., 2012; Piskurewicz et al., 2016; Scott et al., 1998). Our endosperm data had more replicates, higher mappable read depth, and lower levels of seed coat contamination (Supplemental Figure 2, Supplemental Table 1, (Martinez et al., 2018)). There might also be biological differences, as yet unclear, caused by a diploid parent generating diploid sperm as compared to a tetraploid parent generating diploid sperm.

Mis-regulation of imprinted genes characterizes both lethal and viable paternal excess endosperm

Imprinted gene mis-regulation in the endosperm has been suggested as a culprit for the developmental catastrophe of paternal excess crosses (Gutierrez-Marcos et al., 2003; Haig and Westboy, 1991) . We found that imprinted genes are disproportionately more likely to be upregulated than all genes in the genome (N-1 chi-square test, $p= 3x10^{-4}$ for MEGs and $p<10^{-4}$ for PEGs) (Figure 3). Of 43 previously identified Col-L*er* PEGs and 130 Col-L*er* MEGs (Pignatta et al., 2014), 25 PEGs (58%) and 27 MEGs (19%) displayed at least a two-fold increase in transcript abundance in lethal paternal excess endosperm (Figure 3A). Additionally, two PEGs and 18 MEGs were down-regulated. The vast majority of these imprinted genes remained mis-regulated in viable paternal excess endosperm (Figure 3B,C). Only two MEGs were differentially expressed in comparisons of viable and lethal paternal excess endosperm – *JLO* transcript abundance increased 2.6 fold whereas *GSR1* transcript abundance decreased 2.4 fold (Figure 3C, Supplemental Dataset 1). Similarly, only one PEG, *AT4G20800*, had lower transcript abundance in viable paternal excess endosperm compared to lethal paternal excess endosperm (Figure 3C, Supplemental Dataset 1). These results suggest that broad misregulation of imprinted genes in endosperm is unlikely to be the cause of endosperm dysfunction and seed lethality in interploidy crosses.

It was previously demonstrated that mutations in a subset of PEGs or their interactors can suppress paternal excess seed abortion when inherited paternally (Wolff et al., 2015; Huang et al., 2017; Kradolfer et al., 2013; Jiang et al., 2017). We found that while several of these genes indeed have increased transcript abundance in lethal

paternal excess relative to balanced endosperm, they remain mis-regulated in viable paternal excess (Supplemental Figure 3). This observation suggests that seed viability brought about by loss of paternal *nrpd1* is independent of gene expression normalization in endosperm of genes whose paternal loss also represses seed abortion.

In summary, these gene expression data suggest that the expression of only a small number of genes distinguishes lethal paternal excess endosperm from viable paternal excess endosperm and that correction of gene expression to levels observed in balanced endosperm, even for imprinted genes, is not necessary for paternal excess seed viability.

TE mis-regulation is similar in lethal and viable paternal excess seeds

Mis-regulation of transposable elements (TEs) in paternal excess endosperm has been speculated as a cause of seed abortion (Castillo and Moyle, 2012; Martienssen, 2010). We compared TE expression levels in endosperm by mapping mRNA-seq reads to 375 consensus TE sequences in REPBASE (Figure 3D-F) (Bao et al., 2015). In paternal excess endosperm, 73 families showed a statistically significant change in transcript abundance and 40 were upregulated by 5-fold or more, including the transpositionally active family *ONSEN/ATCOPIA78* (Figure 3D, Supplemental Dataset 3). The most upregulated TE families were *ATLINE1-10A* and *ATGP7*, both of which increased by at least 52-fold (Figure 3D). Additionally, 26 TE families were downregulated in paternal excess crosses. *TAG1*, a transpositionally active TE that is present only in the maternally inherited L*er* genome (Tsay et al., 1993), was the most repressed, nearly 56-fold (Figure 3D). Like for genes, viable paternal excess endosperm displayed a TE expression profile similar to lethal paternal excess endosperm (Figure 3E,F, Supplemental Dataset 3). Only 14 TE families exhibited decreased transcript abundance in viable paternal excess endosperm relative to lethal endosperm (Figure 3F, Supplemental Dataset 3). Indeed, 19 TE families exhibited increased transcript abundance in viable paternal excess crosses relative to lethal crosses (Figure 3f, Supplemental Dataset 3). Together, these results demonstrate that while paternal genomic excess leads to extensive TE mis-regulation in endosperm,

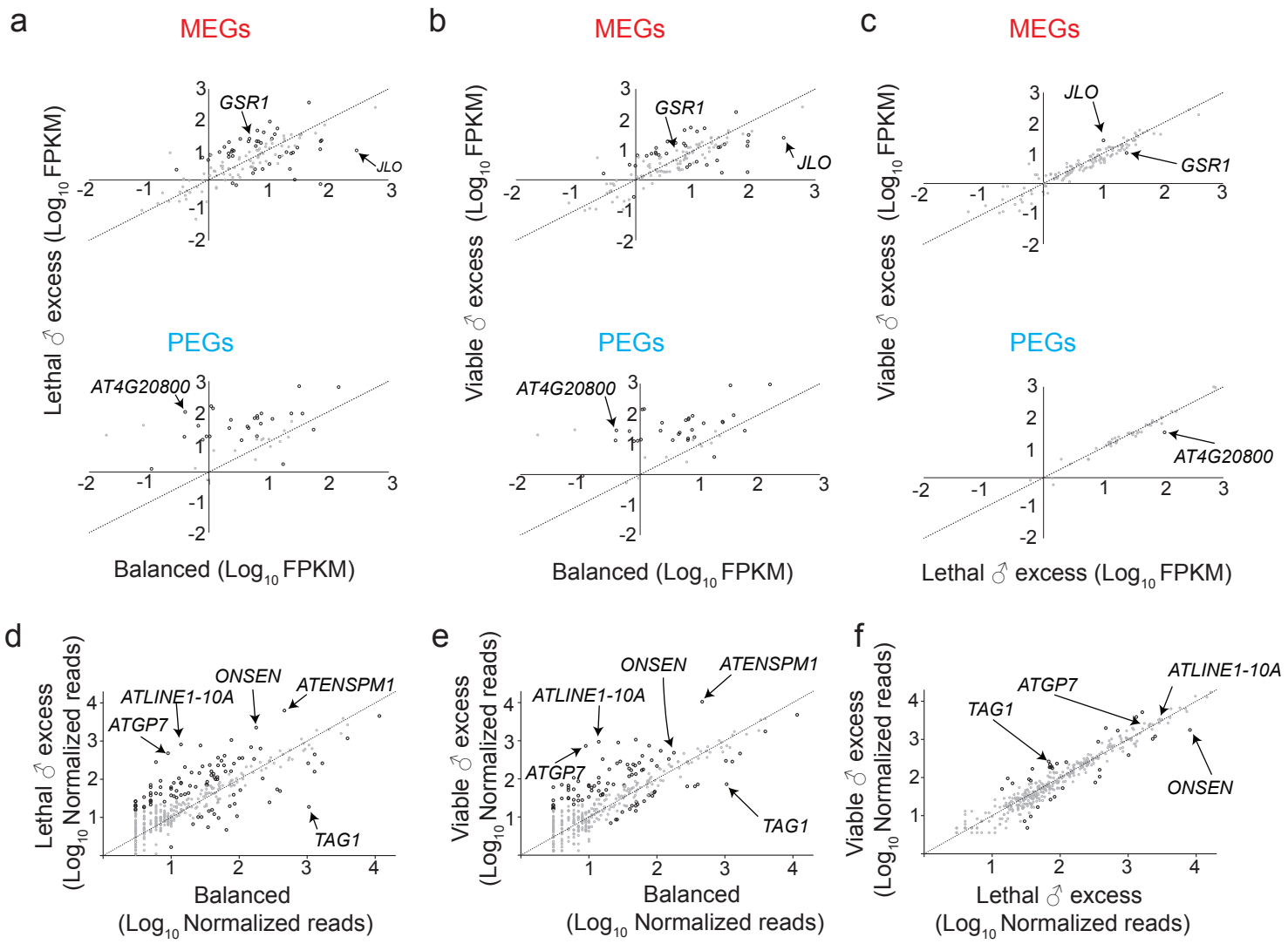
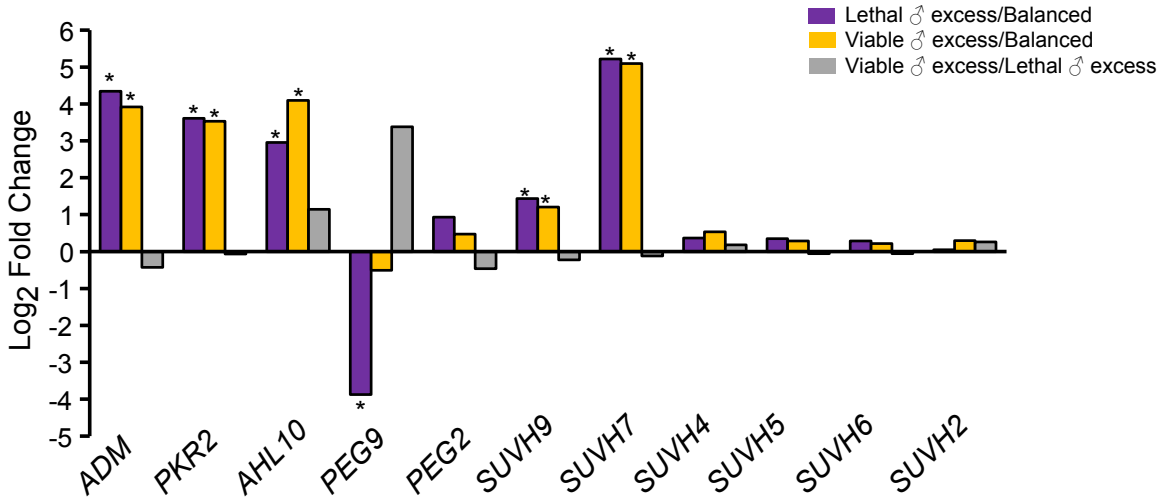


Figure 3: **Imprinted genes and transposons are mis-regulated in both lethal and viable ♂ excess endosperm. a-c)** Expresson of Col-L*er* imprinted genes in endosperm. FPKM is normalized expression; statistical significance of difference in abundance calculated by CUFFDIFF. q<0.05 represented by black circles. q>0.05 represented by gray circles. **d-f)** Expression from transposable elements is elevated in lethal and viable ♂ excess endosperm. RNA-Seq reads were mapped to consensus sequences from REPBASE. Black circles represent TEs with significant differences in transcript abundances according to DEGSEQ. Gray circles represent TEs without significant differences in transcript abundances.



Supplemental Figure 3: Suppression of paternal excess lethality by loss of *NRPD1* does not require normalization of genes implicated in interploid seed abortion. Expression level of genes shown to repress seed abortion in *osd1*-based paternal excess crosses is not repressed by loss of *NRPD1*. Fold change and statistical significance (represented by *) calculated by CUFFDIFF.

broad TE mis-regulation itself is unlikely to be the cause of seed abortion induced by paternal genomic excess.

The RdDM pathway is attenuated in paternal excess endosperm

A number of proteins functioning in the small RNA pathway as well as those that establish or maintain CG, CHG and CHH methylation were differentially expressed between balanced, lethal, and viable paternal excess endosperm (Figure 4). The differential expression of genes encoding members of the RdDM pathway is particularly noteworthy. *RDR2*, *AGO4*, *DRM2*, and subunits of *RNA POL IV* and *V*, were all significantly down-regulated in lethal paternal excess endosperm (Figure 4A, Supplemental Dataset 1). Consistent with these findings, the expression of the 5-methylcytosine DNA glycosylase *ROS1*, whose expression is directly promoted by RdDM (Williams et al., 2015), was also reduced (Figure 4A). With the exception of *RDR2* and *ROS1*, these genes remained down-regulated in viable paternal excess endosperm relative to balanced endosperm (Figure 4A). Thus, in both viable and lethal paternal excess endosperm the expression of members of the RdDM pathway is reduced. We therefore tested if sRNA production and DNA methylation were impacted in paternal excess endosperm. We sequenced endosperm small RNAs from two replicates of lethal and viable paternal excess endosperm (Supplemental Table 1) and compared it to previously published sRNA data from balanced endosperm (Erdmann et al., 2017). Previous evaluations of small RNAs in paternal excess crosses have been performed on whole seeds, not endosperm (Lu et al., 2012; Martinez et al., 2018). We assessed the overall functionality of small RNA production in the endosperm by examining small RNA size profiles. The endosperm 24 nt sRNA population was attenuated in lethal paternal excess endosperm, but still represented the most abundant size class of endosperm small RNAs (Figure 4B). In viable paternal excess endosperm, the proportion of 24 nt sRNAs was comparable to that of balanced crosses (Figure 4B). Genic and TE associated sRNAs in lethal and viable paternal excess showed a pattern similar to that of bulk sRNAs, but genic sRNAs were more impacted than TE associated sRNAs (Supplemental Figure 4A-B). These observations suggest that the capacity to

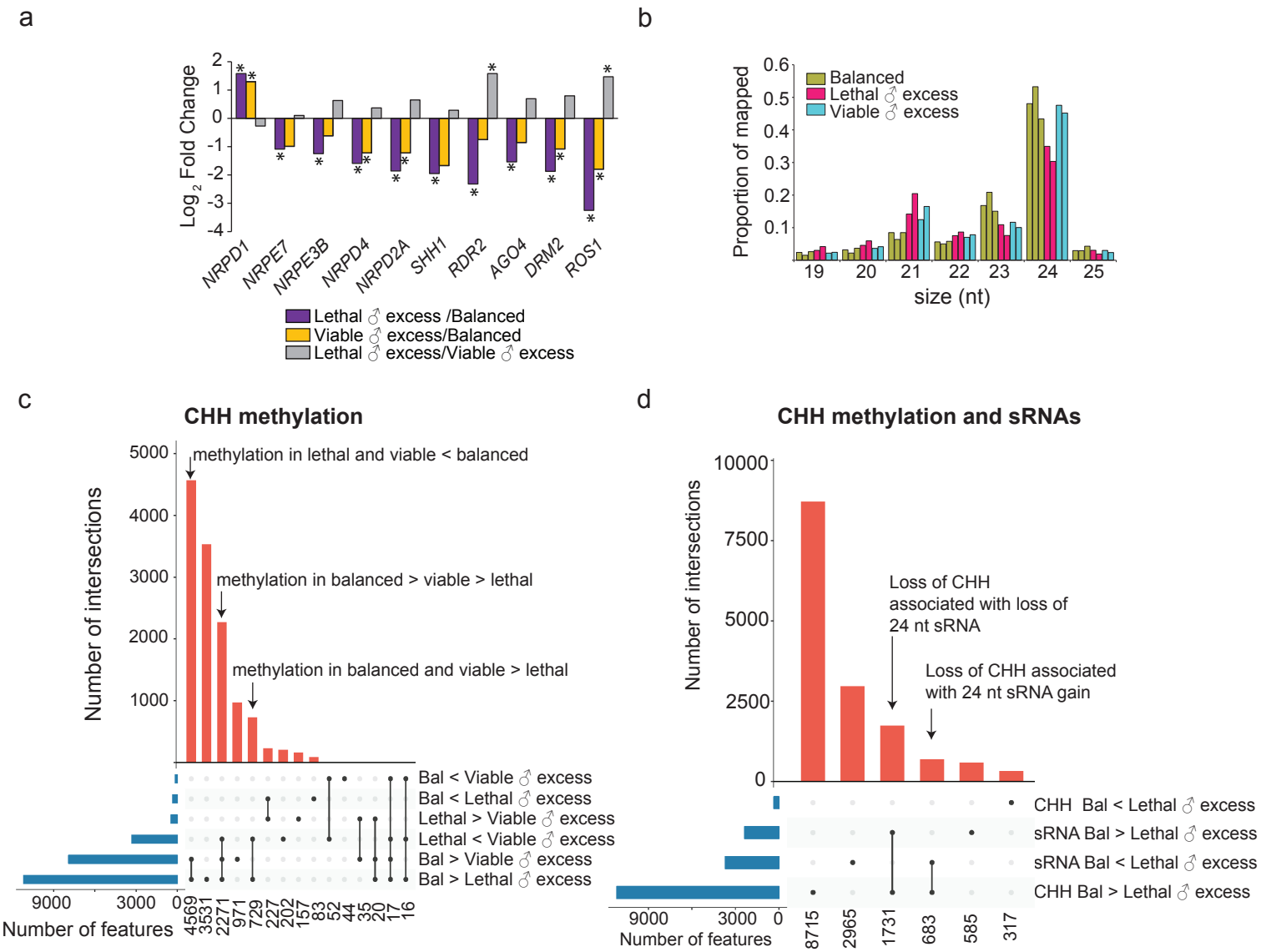
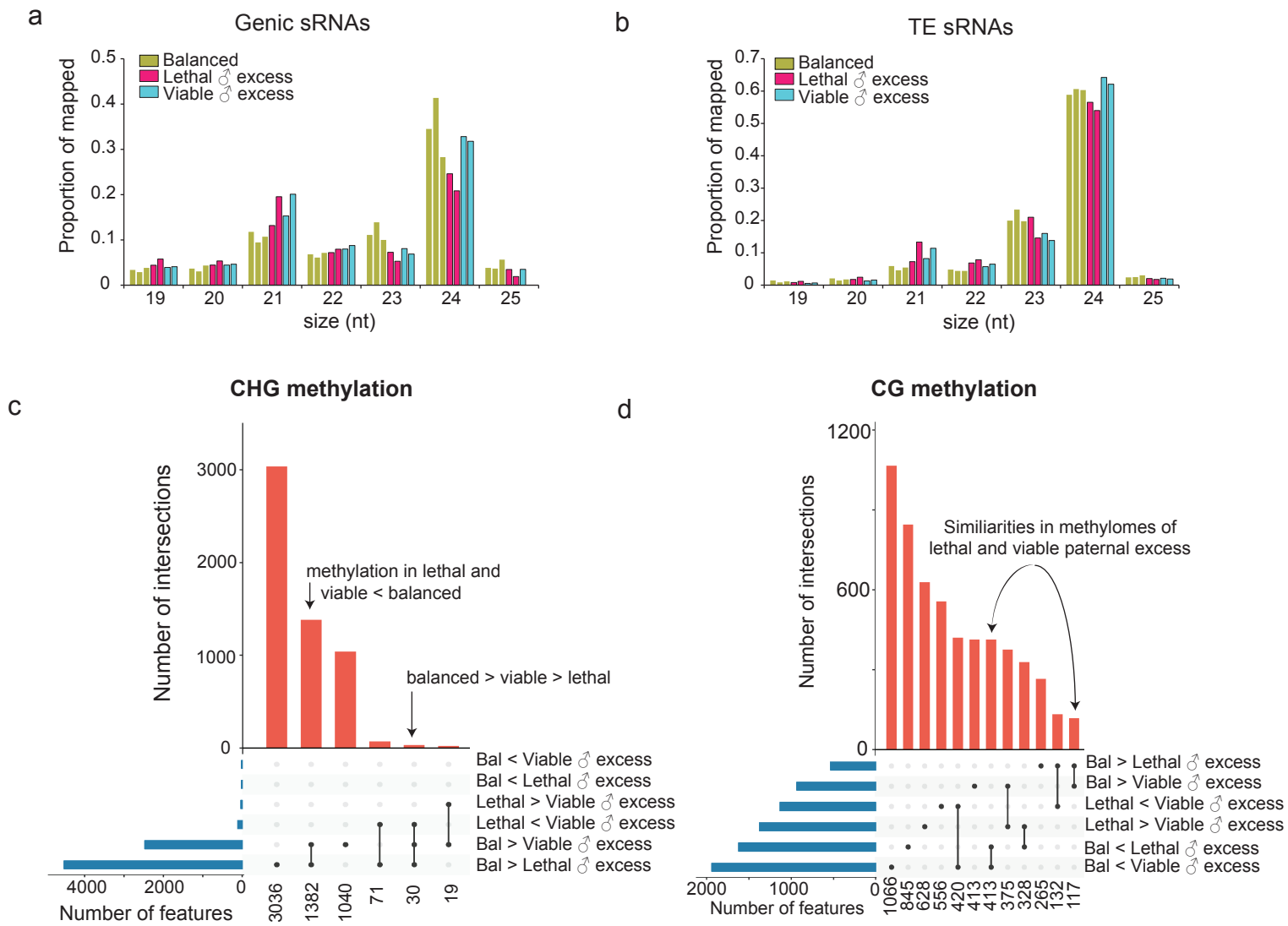


Figure 4: Canonical RdDM pathway function in endosperm affected by paternal excess. a) Genes encoding RdDM components are down-regulated in lethal and viable ♂ excess endosperm. *ROS1* expression is a read-out of RdDM activity and reflects differential activity of RdDM. * represents statistically significantly different gene expression. **b)** Small RNA production is impacted in paternal excess endosperm. Size profiles of sRNA reads mapped to TAIR10 genome for three replicates of balanced endosperm and two replicates each of lethal and viable paternal excess endosperm. **c)** CHH methylation losses in paternal excess endosperm. Upset plot shows intersections of CHH DMRs obtained from comparisons of balanced, lethal and viable paternal excess endosperm. Lethal and viable paternal excess endosperm share a significant proportion of regions that are hypomethylated relative to balanced endosperm. A smaller subset of DMRs lose more methylation in lethal relative to viable. **d)** Loss of CHH methylation is associated with loss of 24 nt sRNAs. Upset plot shows the relationship between changes in sRNA abundance and CHH methylation levels. A subset of CHH DMRs are associated with loss of sRNAs. A smaller subset is associated with gains in 24 nt sRNAs. In **c** and **d**, Upset plots are used to compare different datasets. Number of features represents the total elements in each set. Sets being compared are marked by black dots connected together by a line in the matrix below the histogram. Number of intersections represents the size of the intersect. A black dot not linked by a line to any other set represent unique elements in that set. Numbers below the matrix represent the number of elements associated with the relationship in the matrix above.



Supplemental Figure 4. Small RNA and DNA methylation in balanced, lethal and viable paternal excess endosperm. a) Genic sRNAs are reduced in lethal paternal excess. Size profiles of sRNA reads overlapping genes for three replicates of balanced endosperm, two replicates of lethal paternal excess and two replicates of viable paternal excess endosperm. Sum of all reads overlapping genes was used as denominator to determine proportion of small RNA at each size. **b)** TE insertion sRNAs are reduced in lethal paternal excess. Analysis as in a). **c)** CHG methylation losses in paternal excess endosperm. Upset plot shows intersections of CHG DMRs from comparisons of methylomes of balanced, lethal and viable paternal excess endosperm. A large subset of regions with reduced CHG methylation in lethal and viable paternal excess endosperm relative to balanced endosperm are shared. Only a small number of regions in viable paternal excess endosperm have levels of CHG methylation intermediate between lethal paternal excess and balanced endosperm. **d)** CG methylation changes in paternal excess endosperm. Upset plot shows that a subset of the regions with increased CG methylation relative to balanced endosperm is shared by both lethal and viable paternal excess endosperm. For **c** and **d**, differences in methylation were calculated genome-wide for 300 bp sliding windows with 200 bp overlaps. Windows with fewer than 6 mapped reads or with fewer than 3 overlapping cytosines in both genotypes were excluded. Windows called as significantly different were at least 20% for CHG and 30% for CG. Significance of difference was calculated by F.E.T with a Bonferroni-Hochberg correction ($p<0.01$).

produce 24 nt sRNAs is relatively normal in viable paternal excess endosperm, presumably because of restored *RDR2* expression (Figure 4A-B).

A hallmark of the 24 nt sRNA pathway is non-CG methylation. We performed whole genome bisulfite sequencing to profile the methylomes of balanced, lethal, and viable paternal excess endosperm (Supplemental Table 1). To identify regions with altered CHG/CHH methylation, we divided the genome into 300 bp windows with 200 bp overlaps. We then identified windows that differed between genotypes by at least 10% for CHH methylation and 20% for CHG methylation. CHH and CHG methylation were slightly higher in viable paternal excess endosperm relative to lethal paternal excess endosperm, but in both they were drastically reduced relative to balanced endosperm (Figure 4C, Supplemental Figure 4C, Supplemental Datasets 4 & 5). Regions with changes in non-CG methylation in both lethal and viable paternal excess were linked to differences in 24 nt sRNA levels (Figure 4D). Regions with reduced CHH methylation significantly overlapped with regions with lowered sRNA levels in lethal paternal excess endosperm (Figure 4D).

We also assessed if CG methylation was impacted by identifying regions that differed between genotypes by at least 30%. Many changes in methylation relative to balanced endosperm were shared by both lethal and viable paternal excess endosperm (Supplemental Figure 4D, Supplemental Dataset 6). Overall, slightly more sites gained rather than lost CG methylation in both lethal and viable paternal excess endosperm relative to balanced endosperm (Supplemental Figure 4D, Supplemental Dataset 6). Increases in CG methylation at some sites might be the consequence of reduced *ROS1* expression (Figure 4A).

A subset of gene expression changes are associated with changes in sRNAs and/or DNA methylation

Are the changes in sRNAs and DNA methylation associated with the widespread changes in gene expression (Figure 5)? Production of sRNAs associated with gene bodies was extensively altered by paternal excess (Figure 5A, Supplemental Dataset 7). Using DESeq2, we identified genes at which 21 and 24 nt sRNAs varied in abundance by at least two fold. More genes lost 21 and 24 nt sRNAs in both lethal and viable

paternal excess endosperm (Figure 5A,F, Supplemental Dataset 7). The parallel behavior of 21 and 24 nt sRNAs in genes led us to test if changes in 21 and 24 nt sRNA populations were correlated. This was indeed the case – the R^2 value for differences between lethal paternal excess and balanced was 0.90, 0.87 for viable paternal excess and balanced, and 0.92 for lethal and viable paternal excess (Figure 5C).

Small RNAs can also act as part of the post-transcriptional gene silencing pathway (Borges and Martienssen, 2015). We therefore examined if sRNA levels varied with gene expression levels. Genes whose mRNA levels were increased in lethal paternal excess also had higher levels of 21 nt sRNAs, whereas genes that had lower mRNA abundance in lethal paternal excess had lower 21 nt sRNA levels (R^2 = 0.48, Figure 5D). These results suggest that 21 nt sRNA and mRNA levels are coupled. We did not find a similar genome-wide association similar between 24 nt sRNAs and mRNA levels (R^2 = 0.06, Figure 5D,F). This was puzzling since we found that genes with changes in 21 nt sRNA populations showed correlated changes in 24 nt sRNA populations (Figure 5C,F). We therefore tested if 24 nt sRNA levels at genes where 21/24 nt sRNA levels were coupled also correlated with mRNA levels. This relationship could be assessed for 98 genes – the mRNA levels at these loci were indeed correlated with 24 nt sRNA levels (R^2= 0.35, Figure 5E). These observations suggests that 24 nt sRNA at these loci are not the siRNA species associated with canonical RdDM and is consistent with increased 24 nt sRNAs in a sub-set of regions with lower CHH methylation in lethal paternal excess endosperm (Figure 5F).

It has been recently proposed that TE-associated 21 nt sRNAs inhibit the production of 24 nt sRNAs in the endosperm (Martinez et al., 2018). To test this hypothesis, we extended our analyses to TE insertions. TE insertions gained 21 and 24 nt sRNAs in lethal paternal excess endosperm relative to both balanced endosperm and viable paternal excess endosperm (Figure 5B, Supplemental Dataset 8). Changes in the abundance of TE-associated 21 and 24 nt sRNAs exhibited correlated changes between balanced, lethal, and viable paternal excess endosperm (R^2 for differences between balanced and lethal paternal excess = 0.81, between balanced and viable paternal excess = 0.86, between lethal and viable = 0.17, Figure 5C). These

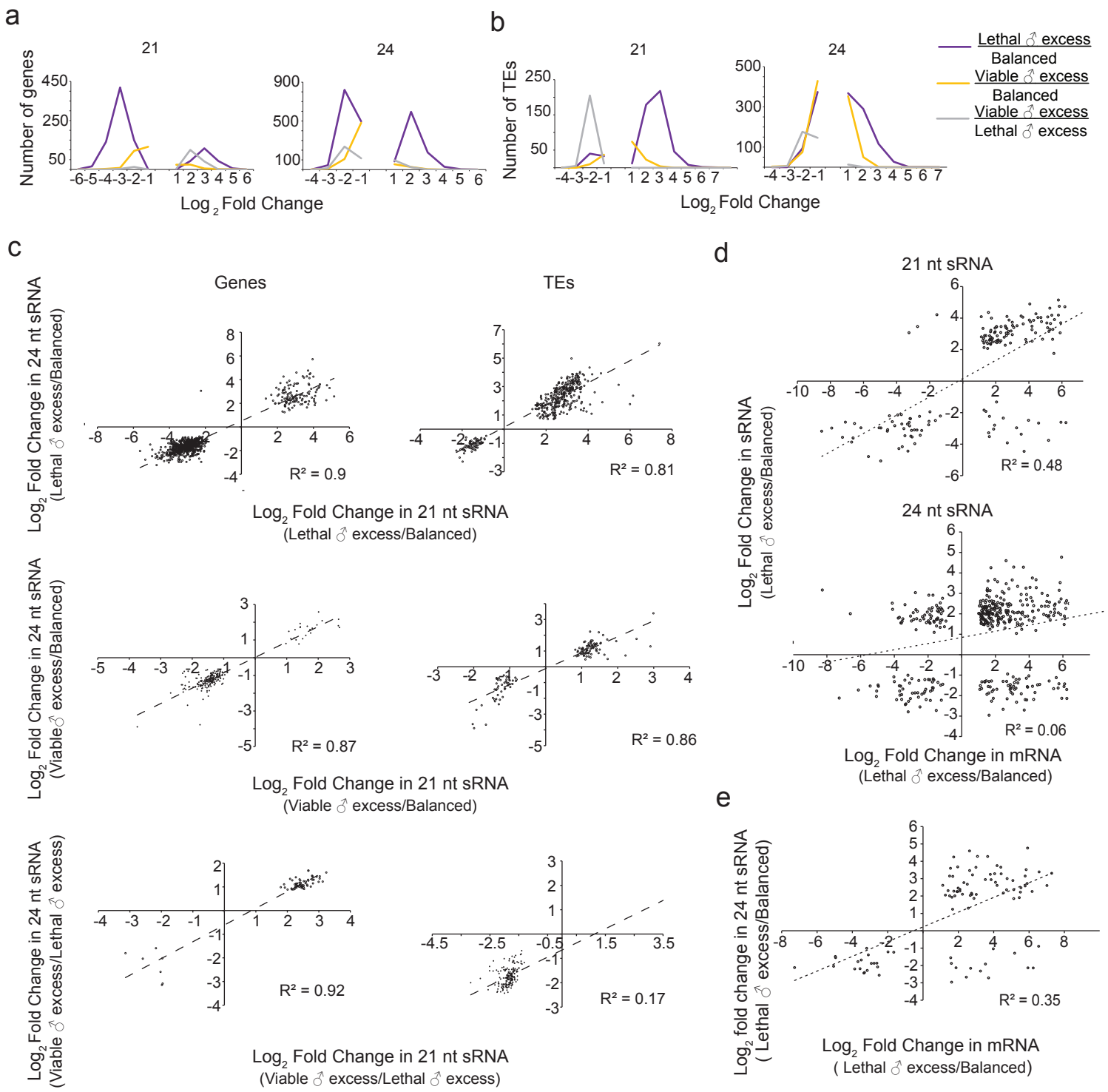

a
21
24
Number of genes
Log 2 Fold Change
b
Number of TEs
Log2 Fold Change
Lethal ♂ excess
Balanced
Viable ♂ excess
Balanced
Viable ♂ excess
Lethal ♂ excess
c
Genes
TEs
Log2 Fold Change in 24 nt sRNA (Lethal ♂ excess/Balanced)
Log2 Fold Change in 21 nt sRNA (Lethal ♂ excess/Balanced)
R² = 0.9
R² = 0.81
Log2 Fold Change in 24 nt sRNA (Viable ♂ excess/Balanced)
Log2 Fold Change in 21 nt sRNA (Viable ♂ excess/Balanced)
R² = 0.87
R² = 0.86
Log2 Fold Change in 24 nt sRNA (Viable ♂ excess/Lethal ♂ excess)
Log2 Fold Change in 21 nt sRNA (Viable ♂ excess/Lethal ♂ excess)
R² = 0.92
R² = 0.17
d
21 nt sRNA
24 nt sRNA
Log2 Fold Change in sRNA (Lethal ♂ excess/Balanced)
R² = 0.48
R² = 0.06
Log2 Fold Change in mRNA (Lethal ♂ excess/Balanced)
e
Log2 fold change in 24 nt sRNA (Lethal ♂ excess/Balanced)
R² = 0.35
Log2 Fold Change in mRNA (Lethal ♂ excess/Balanced)
f
Number of intersections
21nt and 24 nt both down
mRNA and 24 nt both up: MET1, PKR2, SDG21, AGL33,38,86,91
mRNA up, 24 nt down: VIM3, AGL64, AGL23
mRNA down, 24 nt up: CHR38, EIN2, MEE57, ARF4
mRNA, 21 nt, and 24 nt up: HDG3, HDG8
21 nt and 24 nt both up
mRNA up; 21 nt and 24 nt down: ARF15, 20, 21, 22, 23, MYB3R2
Balanced < Lethal ♂ excess
Balanced > Lethal ♂ excess
21
Balanced < Lethal ♂ excess
Balanced > Lethal ♂ excess
24
Balanced > Lethal ♂ excess
Balanced < Lethal ♂ excess
mRNA
Number of features
3702 3663 707 522 496 185 143 69 62 61 59 41 37 36 28 26 24

Figure 5: Small RNA changes associated with TEs and genes. Frequency distribution of significant changes in sRNA abundance at **a)** genes and **b)** transposable element insertions. Fold-change and significance calculated using DESeq2. Read counts for TAIR10 TEs generated using bedtools. **c)** Changes in 21 and 24 nt sRNA abundance are correlated for genes and TEs. Fold change calculated from DESeq2 was plotted for both 21 and 24 nt sRNA abundance for TE insertions and genes. All loci that showed significant differences in 24nt sRNA levels and 21nt sRNAs are shown. **d)** Changes in 21 nt sRNA but not 24 nt sRNAs at genes are correlated with changes in mRNA abundance. Genes included in this analysis showed statistically signficant differences in both mRNA (according to CUFFDIFF) and sRNA (according to DESeq2). **e)** A subset of genes that show correlated changes in 21 and 24 nt sRNA levels also show a correlation between mRNA and 24 nt sRNA. In this analysis, differences in 21 and 24 nt sRNA abundance as well as mRNA abundance were significant. **f)** Upset plot also identifies different relationships in the abundances of 21 nt sRNAs, 24 nt sRNAs and mRNAs arising from genes in balanced and lethal paternal excess endosperm.

observations are not consistent with the proposal that the pathways producing 21 and 24 nt sRNAs act antagonistically in the endosperm (Martinez et al., 2018).

We also examined potential relationships between DNA methylation and gene expression differences between balanced and lethal paternal excess endosperm. While all DNA methylation changes are unlikely regulate gene expression, it could be an important regulator of gene expression changes at some loci. Because DNA methylation can be involved in both activation and repression of gene expression in the endosperm (Satyaki and Gehring, 2017), we identified all DMRs that were 2 kb upstream, within, or 1 kb downstream of a gene whose transcript abundance was statistically different between balanced endosperm and lethal paternal excess endosperm. We identified 856 genes associated with CG DMRs, 564 with CHG DMRs and 1734 DMRs with CHH DMRs (Supplemental Figure 5). Genes identified as potentially influenced by changes in CG methylation included *FWA*, whose expression is known to be repressed by CG methylation in the 5' region of the gene. *FWA* expression was highest in lethal paternal excess endosperm, slightly less in viable paternal excess and the lowest in balanced endosperm. *FWA* CG methylation levels tracked gene expression in an anti-correlated manner. An exemplar locus for a potential relationship between non-CG methylation and gene expression was *ARF15* (Figure 6). *ARF15* expression was not detected in balanced endosperm, was high in lethal paternal excess endosperm and was intermediate in viable paternal excess endosperm (Figure 6B). *ARF15's* gene body was covered with non-CG methylation in balanced endosperm (Figure 6A). Non-CG methylation is likely to have a repressive role at *ARF15*; non-CG methylation was almost completely lost in lethal paternal excess endosperm but was present at intermediate levels in viable paternal excess (Figure 6A). Another exemplar locus for the potential influence of DNA methylation on gene expression was the cytokinin oxidase *CKX3* (Figure 6C-D). There was a patch of non-CG methylation in its first intron (Figure 6C). This methylation was lost in lethal paternal excess and was restored to intermediate levels in viable paternal excess (Figure 6D). *CKX3* transcript levels were inversely associated with DNA methylation (Figure 6C-D).

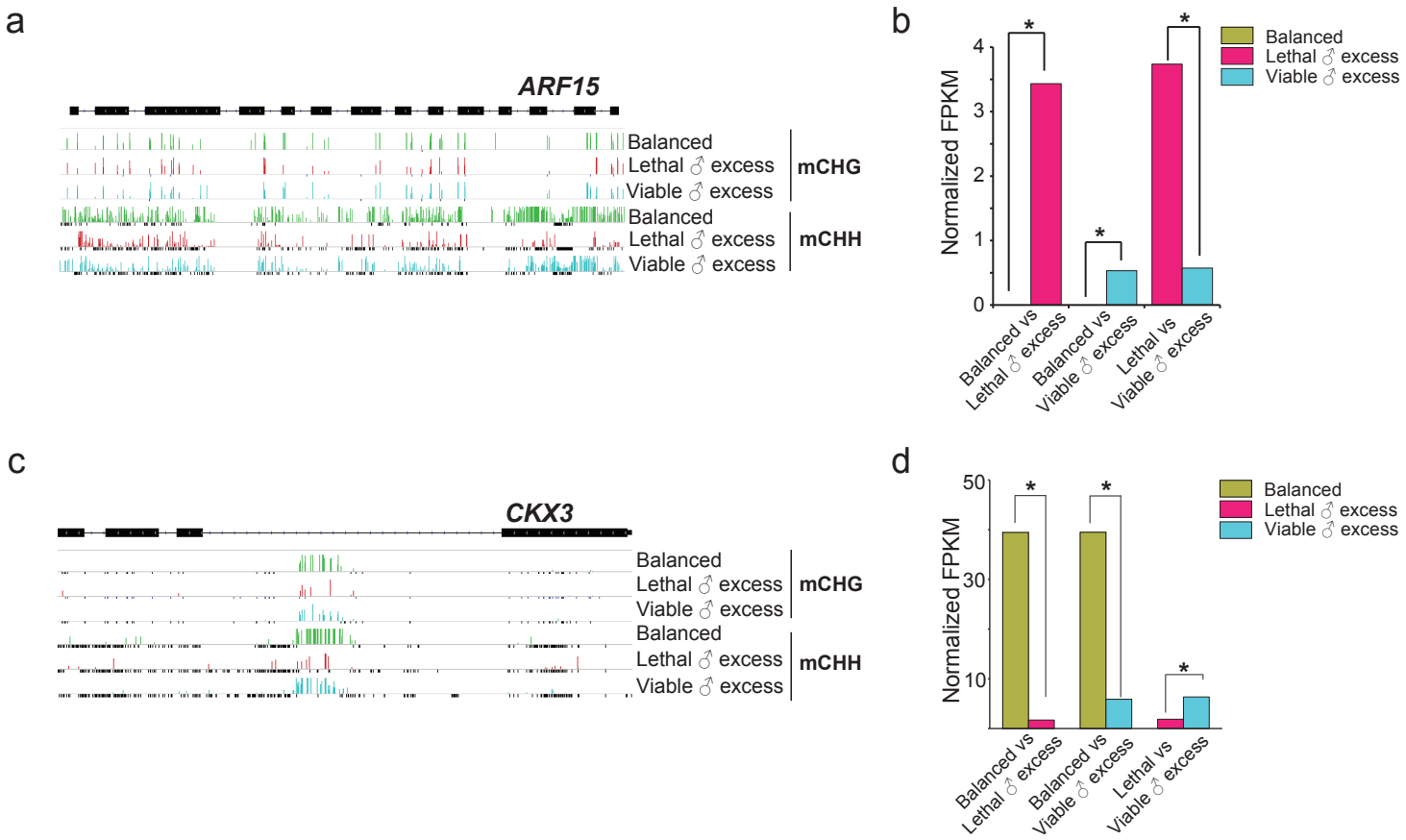
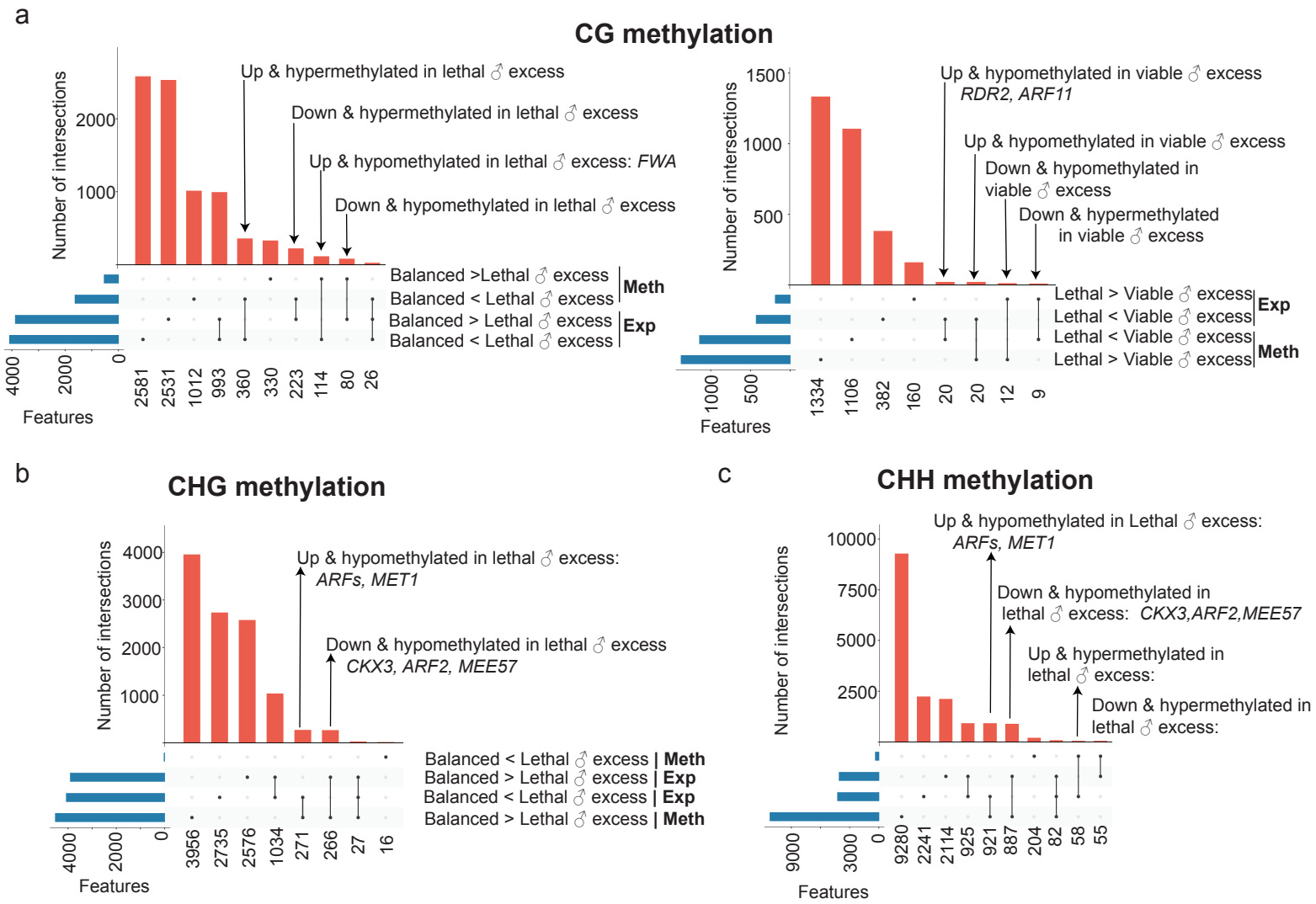


Figure 6: DNA methylation alterations associated with gene expression differences at two exemplar loci. a) CHH and CHG methylation at *ARF15* is reduced by paternal excess. **b)** *ARF15* expression is increased in paternal excess. Expression in viable paternal excess is intermediate between lethal paternal excess and balanced endosperm. **c)** CHH and CHG methylation in the first *CKX3* intron are reduced by paternal excess. **d)** *CKX3* expression is decreased by paternal excess. Expression in viable paternal excess is intermediate between lethal paternal excess and balanced endosperm. In **a** and **c**, black bars below axis represent cytosines with data but no methylation. In **b** and **d**, * indicates $q < 0.05$ as determined by CUFFDIFF analysis.



Supplemental Figure 5: Association between DNA methylation and differentially expressed genes. a) Association between CG methylation and up or down regulation of gene expression. “Meth” represents methylation and “Exp” represents gene expression. **b)** Association between CHG methylation and genic misregulation. **c)** Association between CHH methylation and genic misregulation. In **a-c**, DMRs 2 KB upstream of a gene, within the gene, and 1 KB downstream of a gene were considered to be associated.

Paternal excess differentially affects the expression of maternally and paternally inherited alleles

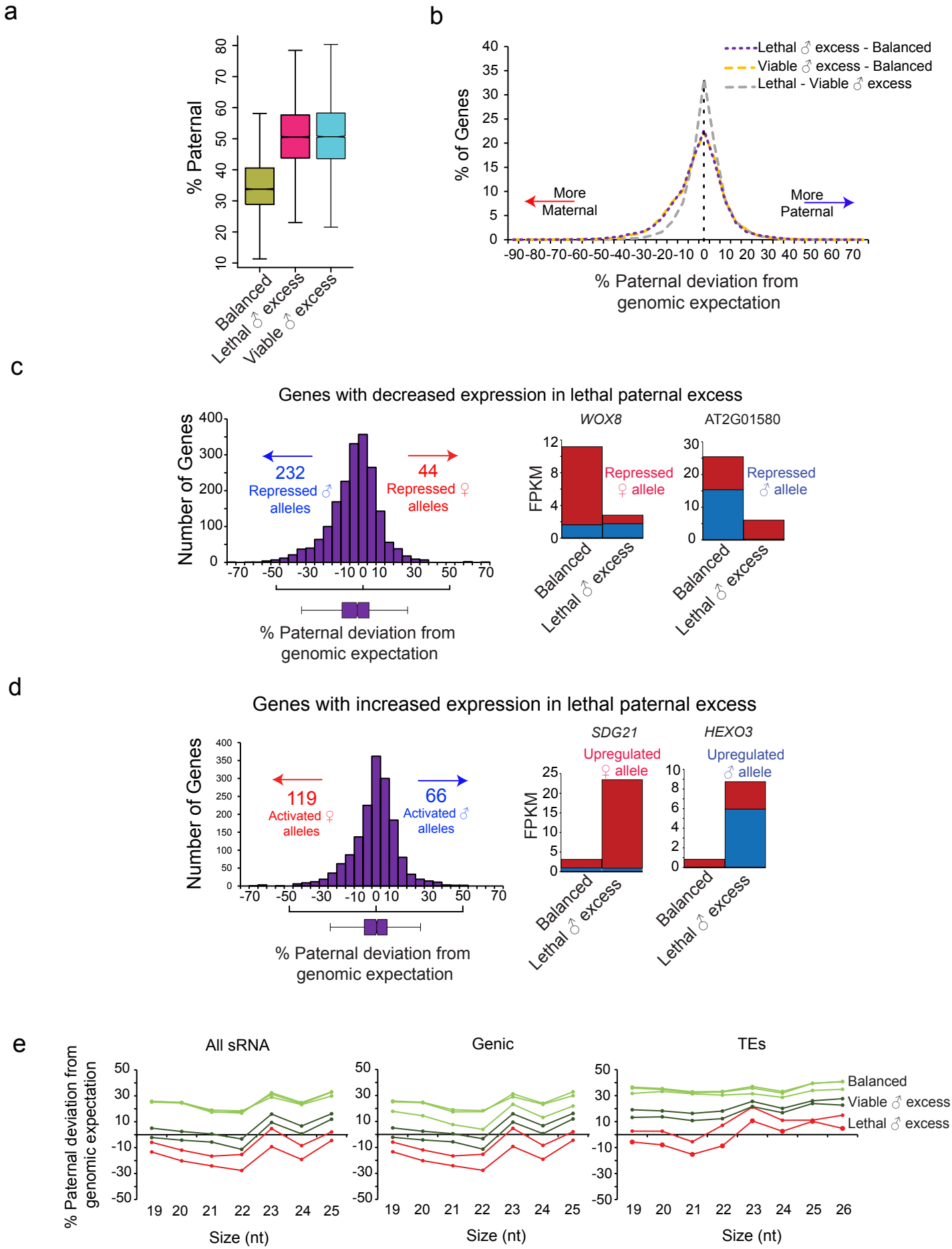
Total gene expression levels are derived from the contribution of maternally and paternally inherited alleles. Based on the effect of *nrpd1* mutations on allele-specific expression in balanced endosperm, we previously proposed that 4N *nrpd1* might suppress paternal excess lethality by increasing transcriptional dosage from maternal alleles for many genes (Erdmann et al., 2017). To test this hypothesis, we evaluated allele-specific expression in balanced, lethal, and viable paternal excess endosperm for genes with at least 50 allele-specific reads in both balanced and paternal excess crosses (Figure 7) (Supplemental Dataset 9). Balanced endosperm has a 2:1 ratio of maternal:paternal genomes and paternal excess endosperm has a 2:2 ratio of maternal:paternal genomes. Mirroring the genomic ratios, the median paternal fraction for genes in balanced endosperm was 33.8%, in lethal paternal excess endosperm it was 50.1%, and in viable paternal excess it was 50.6% (Figure 7A). To assess allelic mis-regulation of specific genes, we normalized transcriptomic fractions to genomic fractions in each cross by calculating deviation from paternal genomic contributions in balanced crosses (% paternal-33%) and paternal excess crosses (% paternal-50%) for each gene. Contrary to our expectation, there was a small but detectable *decrease* in paternal contribution relative to genomic expectation for lethal paternal excess crosses (*two-sided D'Agostino's* K^2 *test, skew = -0.58*, *p< 2.2e-16* ; Figure 7B, purple line, Supplemental Dataset 9). A similar bias was observed in viable paternal excess endosperm (*two-sided D'Agostino's* K^2 test, *skew = -0.53*, *p <2.2e-16;* Figure 7B, yellow line, Supplemental Dataset 9*)*. Thus, in viable relative to lethal paternal excess endosperm there were few changes in overall allelic bias (*two-sided D'Agostino's* K^2 test, *skew = 0.0443*, *p = 0.05887;* Figure 7B, gray line, Supplemental Dataset 9).

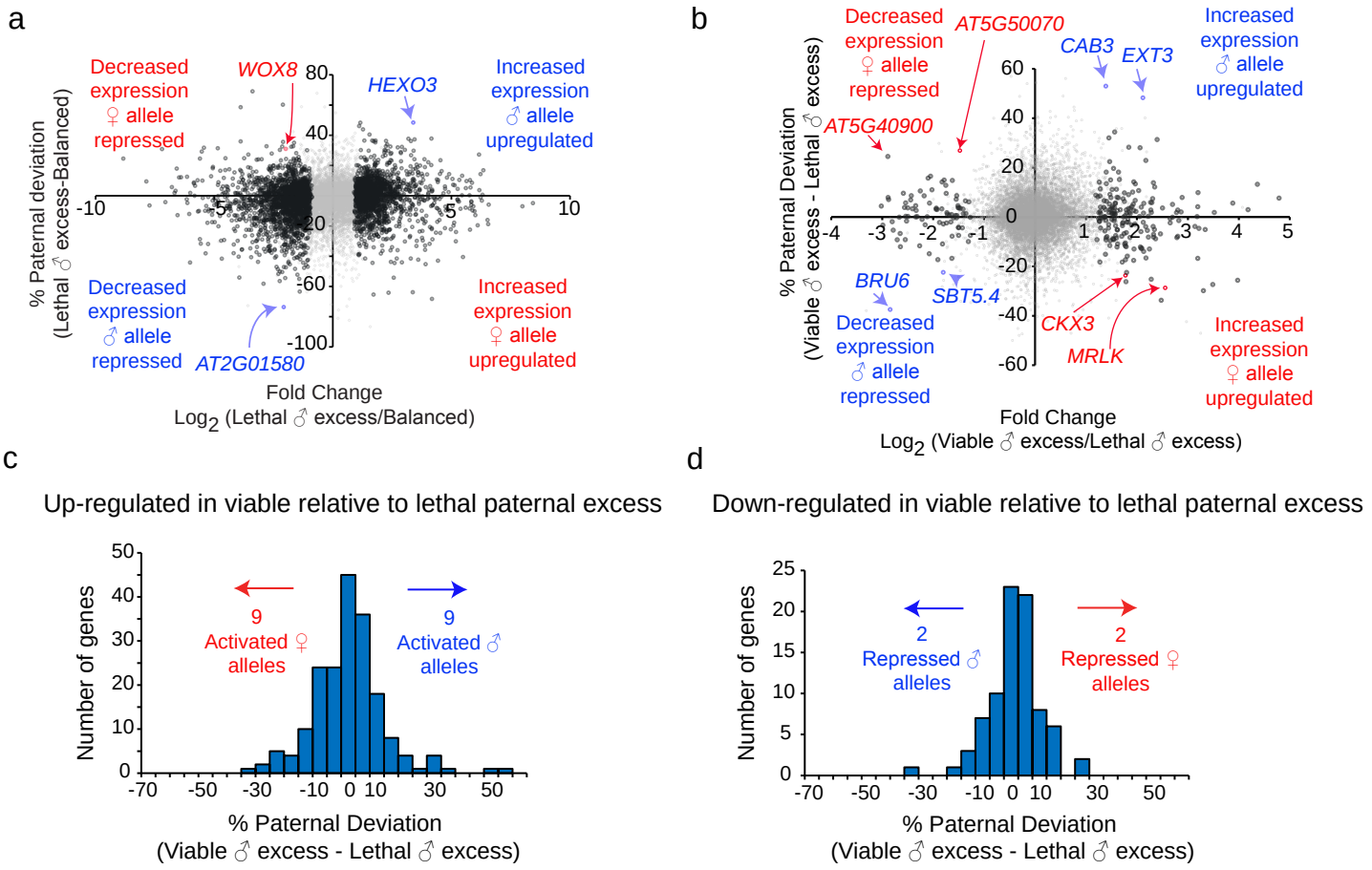
To explore the contribution of allelic mis-regulation to genic mis-regulation, we examined shifts in paternal deviation from genomic expectation for genes up-regulated and down-regulated in lethal paternal excess endosperm compared to balanced endosperm (Figure 7C-D, Supplemental Figure 6A). Down-regulated genes were more heavily influenced by loss of paternal allele contributions (*two-sided D'Agostino's* K^2

test, skew for down-regulated genes =-0.4506, p-value= 1.243e-16; Figure 7C*).* 12.1% of down-regulated genes with sufficient allele-specific reads showed at least 20% decrease in paternal bias, indicating increased repression of paternal alleles, whereas 2.3% showed at least a 20% increase in paternal bias, indicating increased repression of maternal alleles (Figure 7C). Genes with increased expression in lethal paternal excess endosperm were more impacted by increases in maternal allele contributions (*skew for up-regulated genes= -0.52084, p-value= 6.61e-16*; Figure 7D). 7.4% showed at least a 20% decrease in paternal deviation, indicating increased maternal allele expression, and 4.1% showed at least a 20% increase in paternal bias, indicating increased paternal allele expression (Figure 7D).

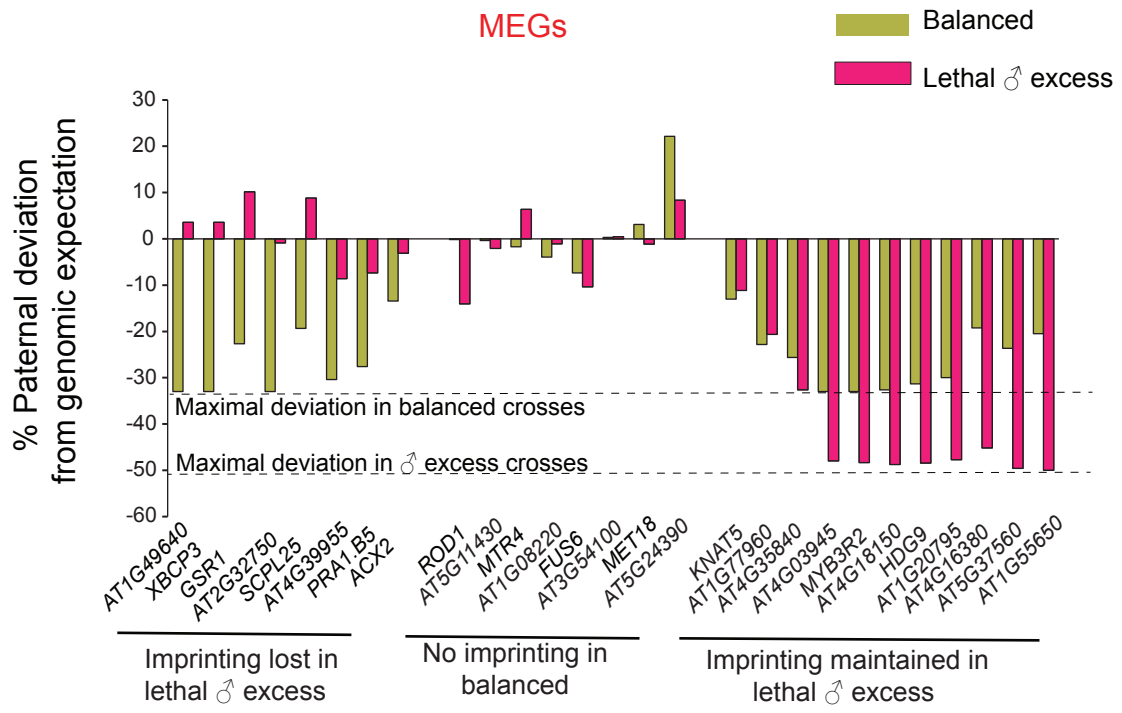
Comparison of allelic contributions between viable and lethal endosperm revealed large differences at a very limited number of genes (Supplemental Figure 6B-D). Among up-regulated genes, increased paternal allele contributions were observed at 9 loci, including *CAB3* and *EXT3* (Supplemental Figure 6B-C, Supplemental Dataset 9)*.* Increased maternal allele contributions were observed at 9 loci, including *MRLK* and *CKX3* (Supplemental Figure 6B-C , Supplemental Dataset 9)*.* Among down-regulated genes, only two genes, *SBT5.4* and *BRU6*, showed strong evidence for a repression of paternal allele expression (Supplemental Figure 6B,D). Maternal allele repression among genes down-regulated in viable endosperm relative to inviable endosperm was detected only at two genes, *AT5G50070* and *AT5G40900* (Supplemental Figure 6B,D)*.*

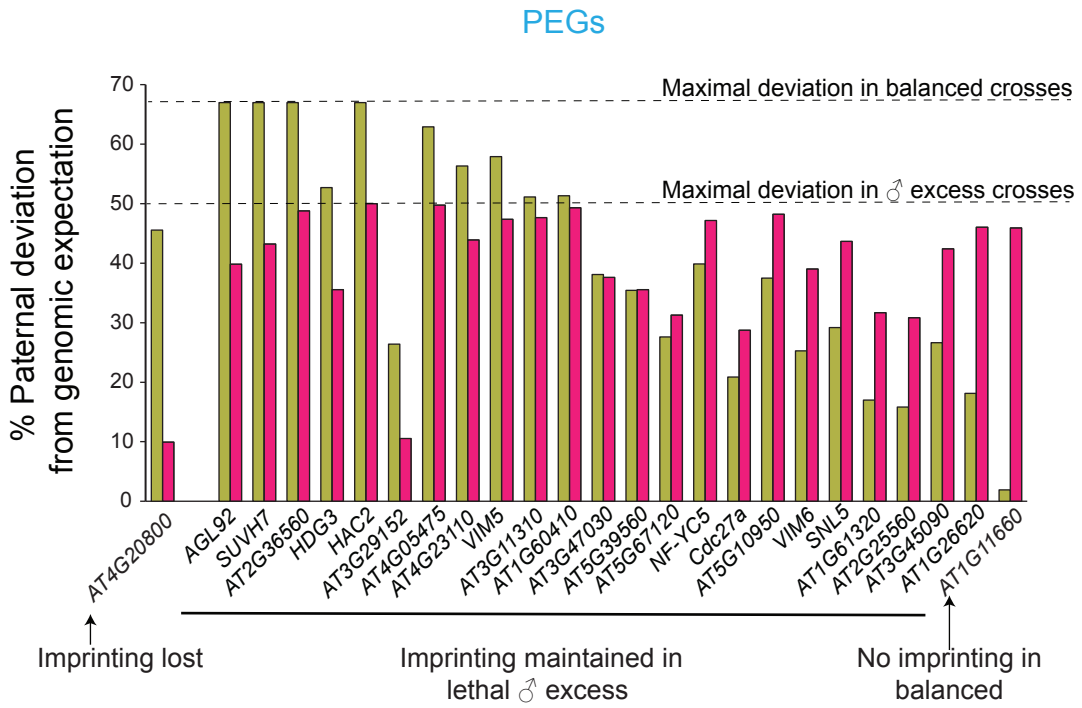
We also tested if the increased expression of imprinted genes in paternal excess endosperm is related to a breakdown of imprinting (Supplemental Figure 7). In comparisons of balanced endosperm and lethal paternal excess, loss of imprinting was apparent among MEGs. Among 27 MEGs that were upregulated in paternal excess endosperm, paternal deviation increased for 8 genes, which were bi-allelically expressed in paternal-excess crosses (Supplemental Figure 7). However, strongly maternally biased expression was maintained at 11 MEGs (Supplemental Figure 7). PEGs mostly retained their paternal biases – of 27 mis-regulated PEGs, examination of allele-specific reads identified 13 genes with reduced paternal bias (Supplemental Figure 7). However, most of these changes were small, with the exception of


a
% Paternal
Balanced
Lethal ♂ excess
Viable ♂ excess
b
% of Genes
Lethal ♂ excess - Balanced
Viable ♂ excess - Balanced
Lethal - Viable ♂ excess
More Maternal
More Paternal
% Paternal deviation from genomic expectation
c
Genes with decreased expression in lethal paternal excess
Number of Genes
232 Repressed ♂ alleles
44 Repressed ♀ alleles
% Paternal deviation from genomic expectation
WOX8
AT2G01580
FPKM
Repressed ♀ allele
Repressed ♂ allele
Balanced
Lethal ♂ excess
d
Genes with increased expression in lethal paternal excess
Number of Genes
119 Activated ♀ alleles
66 Activated ♂ alleles
% Paternal deviation from genomic expectation
SDG21
HEXO3
Upregulated ♀ allele
Upregulated ♂ allele
FPKM
Balanced
Lethal ♂ excess
e
All sRNA
Genic
TEs
% Paternal deviation from genomic expectation
Size (nt)
Balanced
Viable ♂ excess
Lethal ♂ excess



Supplemental Figure 6: Allelic contributions to gene expression differences. a) Contribution of allelic mis-regulation to differential gene expression in lethal ♂ excess. Scatter plot of relationship between change in paternal deviation and change in gene expression between balanced and lethal paternal excess. **b)** Contribution of allelic mis-regulation to differential gene expression between lethal and viable ♂ excess. Scatter plot of relationship between change in paternal deviation and change in gene expression between balanced and lethal paternal excess. In **a** and **b**, black circles represent genes called as being significantly different by CUFFDIFF; gray cricles represent genes that are not significantly different. **c)** Allele-specific changes in genes upregulated in viable relative to lethal paternal excess (N=189). **d)** Allele specific changes among genes downregulated in viable paternal excess relative to lethal paternal excess. (N=83) In **a-d**, genes with more than 50 allele specific reads were considered for analysis. In **c** and **d**, genes with at least a 20% shift in allelic balanced between viable and lethal ♂ excess were considered to have a modified allelic contribution.





Supplemental Figure 7: Allelic bias of some imprinted genes is lost in lethal paternal excess. Changes in paternal deviation for differentially expressed MEGs and PEGs. Maximal deviation for a MEG in balanced endosperm and paternal excess endosperm is -33 and -50 respectively. Maximal deviation for PEGs in balanced endosperm and paternal excess endosperm is 67 and 50 respectively. Our data did not show imprinting in balanced endosperm at a small subset of imprinted loci previously shown to be imprinted in balanced endosperm.

Figure 7: Allelic contributions to expression differences between balanced and lethal paternal excess endosperm. a) Paternal fraction in genic transcripts. Boxplot represents all genes with at least 50 allele specific reads in indicated genotype. Number of genes in balanced = 10,388, lethal ♂ excess =11,709, viable ♂ excess =11,430. For balanced endosperm, a total of five replicate libraries were analyzed (Erdmann et al., 2017). **b)** Paternal excess endosperm transcriptome is maternally skewed. Frequency distribution plot of % paternal deviation for each gene. Paternal deviation was calculated for genes with at least 50 allele specific reads in each pair of genotypes being compared. Number of genes in lethal ♂ excess - balanced = 9,550, viable ♂ excess - balanced = 9,639, viable ♂ excess - lethal ♂ excess = 10,919. **c)** Decreased paternal allele contribution at a larger proportion of genes with decreased expression in lethal paternal excess compared to balanced. N = 1910 genes. Impacts of allele-specific changes on gene expression at two examples, *WOX8* and AT2G01580, are shown. **d)** Increased maternal allele activation at a larger proportion of genes with increased expression in lethal paternal excess. N= 1608 genes. Impacts of allele specific changes on gene expression at two examples, *SDG21* and *HEXO3*, are shown. Genes analyzed for **c** and **d** were detected as being significantly different in gene expression by CUFFDIFF. A gene with an allelic shift of at least 20% was considered to have a modified allelic balance. Boxplots represent median values for paternal deviation from genomic expectation. For gene-specific histograms, maternal allele is red and paternal allele is blue. **e)** sRNA populations are increasingly maternally biased in paternal excess endosperm. The paternal fraction of sRNA populations from viable paternal excess endosperm is intermediate between lethal paternal excess and balanced endosperm. Data from three replicates of balanced endosperm and two replicates each of lethal and viable paternal excess endosperm are plotted here. sRNA reads mapped to TAIR10 genome were first split based on size and then into Col and L*er* reads and reads using SNPs. % of paternal reads, and deviation from genomic expectation at each size was calculated.

AT4G20800, which was bi-allelically expressed (Supplemental Figure 7). These observations indicate that loss of imprinting contributes to imprinted gene over-expression (Figure 3) at only a subset of imprinted genes.

The allele specific changes in gene expression (Figure 7A-D) led us to test if sRNAs change in an allele-specific manner. The sRNA pool in balanced endosperm is overall paternally biased (Erdmann et al., 2017) (Figure 7E). Like we found for mRNA, for most sRNA size classes, lethal paternal excess endosperm was more maternally biased than expected from the ratio of maternal and paternal genomes (Figure 7E). For example, in lethal paternal excess, 24 nt sRNAs were 13.8% more maternal than expected given the 2:2 ratio of maternal to paternal genomes. This indicates, counterintuitively, that extra paternal genomes are associated with decreased small RNAs from paternal genomes (or increased sRNAs from maternal genomes). sRNAs in viable paternal excess endosperm are slightly more paternal than lethal paternal excess endosperm, although to a lesser extent than is observed in balanced endosperm (Figure 7E), consistent with increased functionality of the RdDM pathway in viable paternal excess endosperm (Figure 4A).

In summary, from these analyses we conclude that increasing paternal genome copy number can impact the expression of only maternal or only paternal alleles for a subset of loci. Thus, maternally and paternally inherited alleles contribute unequally to genome-wide transcriptome mis-regulation. However, overall, we did not find evidence for our hypothesis that a shift toward maternal allele expression is associated with repression of paternal excess seed abortion (Erdmann et al., 2017), as *both* lethal and viable paternal excess endosperm have a higher fraction of maternal transcripts than expected from parental genomic ratios.

DISCUSSION

Transcriptional effects and sensitivity to genomic dosage

We have shown that an extra copy of the paternal genome induces regulatory changes at both maternal and paternal alleles and drives massive changes in gene expression in the endosperm (Figures 2, 3, 7). Our data allow evaluation of a number of proposals regarding the transcriptional changes at genes or TEs that could result in interploidy seed lethality. Inspired by models of transposon-induced dysgenesis in *Drosophila* leading to atrophied ovaries (Kelleher, 2016), it was speculated that an imbalance between sRNAs proposed to be deposited by a diploid mother and the TE load in a tetraploid father could lead to TE mis-regulation, which would then trigger seed abortion (Castillo and Moyle, 2012; Martienssen, 2010). We examined TE transcript levels in the endosperm of paternal excess seeds for the first time and found that TEs were mis-regulated in both lethal and viable paternal excess (Figure 3D-F). The observation that viable seeds have high TE transcript abundance, while perhaps surprising, is not unprecedented. Arabidopsis mutants like *met1*, which have high levels of TE expression (Oberlin et al., 2017; Zilberman et al., 2007), also produce viable seed (Xiao et al., 2006). This leads us to conclude that broad TE mis-regulation in endosperm is not a determinant of seed viability.

Transcription of imprinted genes is potentially a strong candidate for the critical gene dosage difference separating lethality and viability (Haig and Westboy, 1991; Lafon-Placette and Köhler, 2015; Wolff et al., 2015). Accordingly, we find that both MEGs and PEGs are mis-regulated in lethal paternal excess endosperm (Figure 3A-C). The coordinated up-regulation of PEGs is particularly striking (Figure 3A-C) and might be linked to a handful of paternally imprinted transcription factors that act as master regulators of other imprinted genes. Supporting this idea, co-regulation of imprinted genes has been previously described in mice (Varrault et al., 2006), transcription factors are enriched among Arabidopsis PEGs (Pignatta et al., 2014), and the loss of the imprinted PEG homeodomain transcription factor *HDG3* leads to the mis-regulation of a subset of imprinted genes in balanced endosperm (Pignatta et al., 2018). Additionally, many PEGs showed a non-linear increase in gene expression upon a doubling of paternal genome dosage, which could be explained by positive auto-regulatory loops (Mileyko et al., 2008). Such an auto-regulatory loop, albeit inhibitory, exists for the MEG *MEDEA* (Baroux et al., 2006; Gehring et al., 2006; Jullien et al., 2006). Under such a

model, a handful of PEG transcription factors could represent the additional paternal chromosomal copy. However, imprinted genes remain mis-regulated in viable paternal excess endosperm (Figure 3B,C) raising the possibility that imprinted gene mis-regulation might only be a consequence of paternal excess but not itself be the primary trigger for seed lethality.

An alternative is that non-imprinted genes act additively with imprinted genes to contribute to seed viability and that the loss of paternal *nrpd1* promotes seed viability by affecting a non-imprinted subset of genes. The exact identities of the genes whose dosage separates a viable seed from a lethal seed remain unclear, but likely constitutes only about 614 genes or less, representing 1.65% of the genome (Figure 2B-,C, Supplemental Dataset 1). These 614 differentially expressed genes includes several encoding proteins involved in phytohormone signaling and developmental regulators – all potentially good candidates for determining seed viability (Supplemental Dataset 1). *CKX2* is an exemplar for a candidate gene that could potentially convert chromosomal dosage information into a developmental cue that can influence seed viability. *CKX2* positively regulates seed size and its expression tracks the relative dosage of maternal and paternal chromosomes (Li et al., 2013; Tiwari et al., 2010) (Supplemental Dataset 1). It is repressed in maternal excess seeds and upregulated In lethal paternal excess endosperm (Li et al., 2013; Tiwari et al., 2010) (Supplemental Dataset 1). In viable paternal excess endosperm, *CKX2* expression is lower than in lethal paternal excess and higher than in balanced endosperm (Supplemental Dataset 1). Together, our results indicate that the transcriptional alterations associated with increased paternal chromosomal dosage is comparable in both lethal and viable paternal excess endosperm (Figure 2). Instead, the loss of paternal *NRPD1* seems to render the endosperm insensitive to increased paternal dosage by modulating the expression of a relatively small number of genes.

A dosage buffering mechanism in endosperm

The process of seed development is remarkably plastic. Analysis of developmental markers shows that both viable and lethal paternal excess endosperm exhibit similarly elevated expression of chalazal endosperm marker genes (Figure 2D),

consistent with increased chalazal endosperm seen in cross-sections of paternal excess seeds (Martinez et al., 2018; Scott et al., 1998; Wolff et al., 2015). Importantly, even wild-type paternal excess crosses produce some viable seed (Figure 1, Supplemental Figure 1). Our results suggest that transcriptional buffering, operational in both lethal and viable paternal excess endosperm, is one feature that could contribute to the ability of the seed to withstand the addition of an entire extra genome. In buffering, expression per copy of a gene can be increased or decreased based on the number of copies of the gene relative to that of the rest of the genome (Birchler and Veitia, 2012; Zhang et al., 2010). We observed the signatures of buffering in paternal excess transcriptomes (Figure 7). Decreased paternal transcript levels contributed predominantly to genic down-regulation whereas increases in the levels of transcripts from maternal alleles contributed predominantly to genic up-regulation (Figure 7C-D). The differential activity of maternal and paternal alleles makes the transcriptome more maternal than expected from parental dosage and constitutes a push back against excess paternal genomes. Many mechanisms have been either proposed or identified to explain such buffering phenomena in other systems. In animals, in what represents a specialized case of aneuploidy, dosage compensation mediated by chromatin modifiers acts to equalize expression between sex chromosomes and autosomes (Disteche, 2016). Our previous and current results suggest small RNAs in endosperm also regulate gene dose. We have previously shown that RNA POL IV, which generates the primary transcripts for 24 nt sRNA production, is required in balanced endosperm for mild repression of maternal alleles (which outnumber paternal alleles). In paternal excess endosperm, the expression of genes encoding subunits of *RNA POL IV* as well as other downstream members of the RdDM pathway are decreased, presumably as a mechanism of increasing maternal transcriptome fraction (Figure 4A). Other sRNA species may also play a role in buffering gene dose in the endosperm. 21 nt sRNAs are involved in post-transcriptional gene silencing. At a subset of loci, the change in abundance of 21 nt sRNAs between balanced and lethal paternal excess endosperm is correlated with mRNA expression, suggesting that these 21 nt sRNAs are produced from processing of mRNAs. Additionally, we also identified some sites where 24 nt sRNAs were increased concurrent with a reduction in CHH methylation (Figure 4D). This suggests that these

sRNAs might be distinct from the 24 nt sRNAs that guide DNA methylation. Instead, they are likely involved in a non-canonical post-transcriptional gene silencing pathway that involves *DCL3*. This would be analogous to previous findings that 24 nt sRNA species increase with increased mRNA abundance of the retrotransposon *EVADE* (Marí-Ordóñez et al., 2013). In summary, our data highlights potential mechanisms that allow seeds to compensate for increased paternal dosage.

Mechanism of seed abortion repression by paternal RdDM pathway

The precise molecular mechanism by which paternal action of the RdDM pathway modulates endosperm gene expression and influences interploidy seed viability remains unknown. There are at least three potentially heritable epigenetic effectors that paternal RdDM pathway genes could employ to influence developmental outcomes after fertilization – sRNAs, DNA methylation, and histone modifications. Because of reasons explained in the next section, we argue that sRNAs transmitted from the father to the endosperm are unlikely to be part of the answer to this question. We instead favor a model in which RdDM pathway genes direct covalent modifications to chromatin. Little is known about how or if RdDM-dependent histone modifications are established in the parent. While we cannot rule out a potential role for histones, several observations suggests that DNA methylation is a strong candidate for the *RNA POLIV*-dependent epigenetic mark that is essential for paternal excess seed abortion. We have shown that several key members of the RdDM pathway, from *NRPD1* to *DRM2*, are required for paternal excess seed abortion. This is consistent with the argument that RNA POL IV is acting via the canonical RdDM pathway in the paternal parent to regulate the endosperm transcriptome after fertilization, resulting in seed abortion in paternal excess crosses. Consistent with this model, it has been shown that loss of *NRPD1* or *DRM2* in the father can impact gene expression in the endosperm of balanced crosses (Erdmann et al., 2017; Vu et al., 2013) and that *DRM2* establishes methylation patterns at a subset of genomic sites in the male germline (Walker et al., 2018).

An important unanswered question about this DNA methylation-based model is why a presumed loss of DNA methylation in the genome donated by the tetraploid

father leads to a repression of seed abortion. Some studies suggest that tetraploidization itself induces DNA methylation changes (Baubec et al., 2010; Mittelsten Scheid et al., 2003; Yu et al., 2010; Zhang et al., 2015). In one scenario, DNA methylation changes in tetraploid pollen could cause a change in gene expression in the endosperm. In tetraploids without RdDM, this methylation change, as well as the consequent gene expression differences in endosperm, could revert back to levels seen in balanced endosperm. An alternative possibility is that loss of RdDM may lead to an altered methylation state not found in either balanced or lethal paternal excess crosses. This ectopic paternal methylation state could then dictate a gene expression pattern that represses seed abortion.

Small RNAs and models for repression of paternal excess

Two separate hypotheses revolving around sRNAs have been postulated to explain how loss of *NRPD1* in tetraploid fathers, but not diploid mothers, represses seed abortion in paternal excess crosses (Erdmann et al., 2017; Martinez et al., 2018). The easiRNA model stands on four key conclusions from the data of Martinez et al (Martinez et al., 2018). First, a non-canonical pathway dependent on *POLIV* and *RDR6* functions in pollen to make easiRNAs that scale with paternal chromosomal dosage. Second, the transfer of excess easiRNAs into the endosperm inhibits RdDM. Third, the absence of these sRNAs in *nrpd1* mutant pollen permits the RdDM pathway to function in the endosperm. Fourth, gene expression is normalized in viable paternal excess endosperm. Our data do not support these conclusions. First, we failed to obtain viable paternal excess seeds with 4N *rdr6-15* (Figure 1A, Supplemental Figure 1B). This result suggests that any *RDR6*-produced sRNAs are not linked to paternal excess seed abortion. Consistent with this, paternally produced siRNA854, which is dependent on *DCL2/4* and *RDR6*, is needed for paternal excess seed viability (McCue et al., 2012; Wang et al., 2018). Next, we find that *NRPD1*, *RDR2, DCL3, NRPE1* and *DRM2,* all members of the canonical RdDM pathway, restore paternal excess seed viability (Figure 1, Supplemental Figure 1A,C). Additionally, unlike the relatively low levels of viable seed produced by mutations in non-RdDM genes such as *RDR6, DCL2, DCL4*, *AGO2* and *miRNA845* (Borges et al., 2018; Martinez et al., 2018), loss of *NRPE1* and *DRM2*

phenocopy *NRPD1* in the large magnitude of repression of seed abortion (Figure 1, Supplemental Figure 1). The cause of the discrepancy between the studies remains unclear but may be linked to the use of *osd1* mutation to create conditions of paternal excess. Paternal excess cross viability is sensitive to parental genetic backgrounds (Piskurewicz et al., 2016; Scott et al., 1998). The *osd1* mutation was originally isolated in the Nossen ecotype and has since been back-crossed into the Col ecotype (d'Erfurth et al., 2009; Martinez et al., 2018). Creation of double mutants with *osd1* and a second mutation could create lines with different combinations of introgressions from each genetic background. Similar discrepancies have been observed in other studies. In the case of *PKR2, osd1 pkr2* mutants did not repress seed abortion while 4N *pkr2* showed reduced seed abortion (Huang et al., 2017; Wolff et al., 2015). The next conclusion drawn by the easiRNA model is that the production of sRNAs as well as DNA methylation is lost in paternal excess endosperm because easiRNAs, transferred from the pollen, act to inhibit the catalytic activity of RNA POL IV-RDR2 complex. Thus, the loss of easiRNAs and their inhibitory capability is what allows the reestablishment of RdDM in viable paternal excess endosperm. Correlated increases in some small RNA species in both tetraploid pollen and lethal paternal excess whole seeds has been used to argue that that these small RNA species have been transferred from pollen to the central cell at fertilization (Martinez et al., 2018). Our data however suggests that the increases in these sRNA species are likely to be dependent on increased expression from maternal chromosomes (Figure 7E), suggesting that they are produced *de novo* in the endosperm and not transferred from the pollen. The cause of the difference in conclusions between our study and those in Martinez et al may be linked to the fact that our sRNA libraries were from purified endosperm and the libraries in Martinez et al were from whole seeds, which includes the seed coat and embryo. Additionally, we find that the observed attenuation of RdDM activity is likely the result of the coordinate down-regulation of the genes encoding the members of the RdDM pathway in lethal paternal excess (Figure 4A). The higher levels of RdDM in viable paternal excess relative to lethal paternal excess is likely a result of the slightly higher expression of RdDM genes (Figure 4A). This result negates the need to invoke an inhibitory role for easiRNAs to explain why the RdDM pathway is attenuated in paternal excess endosperm. Finally,

759 the easiRNA model also suggests that the RdDM pathway's recovery in the viable
760 paternal excess endosperm is essential for seed viability. However, loss of both
761 paternal and maternal copies of *NRPD1, RDR2, DCL3, NRPE1*, and *DRM2* in paternal
762 excess crosses did not diminish the seed viability observed upon loss of the paternal
763 copy alone (Figure 1, Supplemental Figure 1) (Erdmann et al., 2017; Martinez et al.,
764 2018), indicating that the activity of a genetically complete canonical RdDM pathway in
765 the endosperm is not necessary for suppression of paternal excess seed abortion.

766 We previously showed that maintaining a 2:1 ratio of maternal to paternal allele
767 transcripts in balanced endosperm was dependent on *NRPD1*; in its absence the
768 expression of many genes shifted more maternal (Erdmann et al., 2017). Thus, we
769 suggested that the loss of *NRPD1* repressed paternal excess seed abortion by raising
770 the maternal fraction of the transcriptome, thus balancing excess paternal gene dosage
771 (Erdmann et al., 2017). A central tenet of our model was that the reduction in the RNA
772 POL IV sRNA pathway in endosperm (caused by loss of paternal *NRPD1*) was essential
773 for seed viability. However, we find that lethal paternal excess endosperm already
774 phenocopies some aspects of balanced *nrpd1* mutant endosperm – there is reduced
775 functionality of the RdDM pathway (Figure 4) and, accordingly, expression is shifted
776 more maternal (Figure 7B). This maternal shift is retained (and not increased) in viable
777 paternal excess endosperm (Figure 7B) and thus is unlikely to be the determinant
778 between seed lethality and viability. Additionally, heterozygous NRPD1/*nrpd1* viable
779 paternal excess endosperm has a *more* functional RNA POL IV sRNA pathway than
780 *NRPD1/NRPD1* lethal paternal excess endosperm, as measured by expression of
781 genes encoding sRNA pathway components, 24 nt sRNA levels, and CHH methylation
782 levels (Figure 4). We also previously suggested that the ability of paternal *nrpd1* to
783 repress seed abortion was likely linked to *NRPD1*'s status as a PEG in endosperm
784 (Erdmann et al., 2017). In this work, we find that the loss of paternal copies of *RDR2,*
785 *DCL3, NRPE1,* and *DRM2* also represses seed abortion (Figure 1, Supplementary
786 Figure 1). These genes are not imprinted and are expressed at similar levels from both
787 alleles (Pignatta et al., 2014). Furthermore, we find that the repression of seed lethality
788 in paternal excess crosses is not enhanced when both paternal and maternal copies of
789 RdDM pathway genes are lost (Figure 1, Supplementary Figure 1). These results

indicate that these mutations behave as paternal effect loss-of-function mutations, although our genetic results cannot distinguish whether the repression is a sporophytic or gametophytic effect. Abrogation of the RNA POLIV pathway is needed only in the father to promote seed viability and its status in the endosperm is not an essential determinant of seed life or death.

Methods

Arabidopsis strains, tetraploid production, and tissue collection

Mutants used in this study were: *nrpd1a-4*, *rdr2-1*, *dcl3-1*, *nrpd1b-11*, *drm 2-2*, *rdr6-15* (CS879578), *rdr2-1 rdr6-15* (CS66111) and *cmt3-11t* (CS16392). Tetraploids were generated by applying 0.25% colchicine in 0.2% Silwet to the apices of 2-3 week old diploid plants. Progeny from treated plants were screened by flow cytometry to identify tetraploids. For crosses, wild-type diploid Col-0 or L*er* buds were emasculated and pollination was carried out two days later with diploid or tetraploid pollen in Col-0 strain background. To assess seed abortion, siliques were harvested after drying and seed examined under a dissecting microscope. To assess germination rates, seeds were sterilized in 2% PPM (Plant Cell Technologies) for three days at 4°C and then plated on 0.5X MS/Phytagar media. To obtain purified endosperm, approximately one hundred seeds at 7 DAP from at least three siliques were dissected away from embryo and seed coat for each replicate as previously described (Gehring et al., 2009, 2011) For lethal paternal excess seeds, endosperm was collected only from seeds with arrested embryonic growth. For viable paternal excess seeds, endosperm was collected only from seeds where embryonic growth was progressing normally.

mRNA-Seq library prep and transcriptome analyses

Long RNA was isolated as previously described (Erdmann et al 2017) and mRNA-seq libraries were constructed at the Whitehead Institute Genome Technology Core using the SMARTerUltra-lowPOLYA-V4 kit. Libraries were sequenced on a 40 base, single read cycle. Sequence data were filtered for quality with " *trim_galore -q 25 --phred64 --fastqc --length 20 --stringency 5*". Filtered reads were aligned to the TAIR10 genome

using "*tophat -i 30 -I 3000 --solexa1.3-quals -p 5 -g 5 --segment-mismatches 1 --segment-length 18 --b2-very-sensitive*" (Kim et al., 2013). We used CUFFDIFF with default settings and the ARAPORT11 genome annotation to calculate changes in gene expression and their statistical significance. Genes with q-value <0.05 were considered to be significantly different. We used a custom script (assign_to_allele.py; https://github.com/clp90/imprinting_analysis/tree/master/helper_scripts) and SNPs between Col-0 and L*er* to identify allele specific reads. Allele specific reads per gene were assessed using "*htseq-count -s no –m union*" and the ARAPORT11 annotation. To assess TE transcript levels, reads were aligned to the REPBASE consensus sequence (Jurka et al., 2005) using "*bowtie -v 2 -m 3 --best --strata -p 5 --phred64"* and reads mapping to each TE family were summed. *META1, ATHILA6A* and *ATGP8* were precluded from further analyses because of mapping artifacts. Differential abundance of TE mapping reads were calculated using Fisher's exact test (p<0.05) option in DEGseq (Wang et al., 2010).

Tissue enrichment test

Expression per gene was measured using HT-Seq count and analyses were performed using the seed tissue enrichment test (Schon and Nodine, 2017). Time point was set to bent cotyledon.

Small RNA-Seq library preparation and analyses

RNA was isolated from manually dissected endosperm using the RNAqueous micro kit (Ambion). Small RNA was obtained as previously described (Erdmann et al, 2017). Libraries were built using the NEXTflex sRNA-seq kit v3 (Bioo Scientific). Final library amplification was carried out for 24 cycles; resultant libraries were size selected (135-160 bp) with a pippin prep. 40 base single read sequencing was carried out an Illumina HiSeq 2500. Sequencing reads were trimmed for quality with "*fastq_quality_trimmer -v -t 20 -l 25*". Reads were further filtered and adapter containing reads were retained using "*cutadapt -a TGGAATTCTCGGGTGCCAAGG --trimmed-only --quality-base 64 -m 24 -M 40 --max-n 0.5"*. The reads from libraries produced by the NEXTflex library prep kit include four random nucleotides at both 5' and 3' ends of the reads. Taking advantage of these tags, we were able to remove PCR duplicates using "*prinseq-lite –fastq <infile>*

-out_format 3 –out_good <outfile> -derep 1". Reads were aligned to TAIR10 with "*bowtie -v 1 --best -5 4 -3 4 --sam --phred64-quals or --phred33-quals"* . We used a custom script (assign_to_allele.py; https://github.com/clp90/imprinting_analysis/tree/master/helper_scripts) to identify allele-specific reads. Regions with differential sRNA levels were identified by counting reads in 300 bp windows with 200 bp overlaps. DESeq2 (Love et al., 2014) was used to identify windows with differential abundance of sRNAs. Overlapping windows with increased sRNA were merged. Upset plots (Lex et al., 2014) were created with the *intersectBed* command implemented using the Upset Intervene package (Khan and Mathelier, 2017). To identify genes with differences in 21 and 24 nt sRNAs, mapped reads were separated based on size. 21 and 24 nt sRNAs mapping to genes were counted using the ARAPORT11 annotation and "*htseq-count -s no –m union*". To identify TE insertions with differences in 21 and 24 nt sRNAs, we first identified TE insertions that did not overlap with genes and used "*coverageBed -counts*" command from Bedtools suite (Quinlan and Hall, 2010) to count reads mapping to TE insertions. We identified genes and TEs with differential abundance in 21 and 24 nt sRNA using DESeq2.

Methylome library preparation and analyses

Genomic DNA was isolated from manually dissected endosperm using the QiaAMP DNA microkit (QIAGEN 56304); dissected tissue was incubated in ATL buffer and proteinase K at 56°C overnight on a shaking incubator. 80-100 ng of DNA obtained from these protocols was subjected to bisulfite treatment using Methylcode Bisulfite Conversion kit (Invitrogen). Bisulfite libraries were constructed from these materials using Illumina's Truseq DNA methylation kit. Bisulfite libraries were sequenced on Illumina's HISeq 2500 in a paired end configuration. Reads were filtered for quality with "*trim_galore --phred64 --fastqc --stringency 5 --length 15 --paired --clip_R1 2 --clip_R2 2".* The reads were then aligned to TAIR10 using *"bismark -N 1 -L 20".* Duplicate reads were removed using Bismark. Bismark methylation extractor and custom scripts previously described (Pignatta et al., 2014, 2015) were used to obtain per base methylation. Briefly, differences in methylation were calculated genome-wide for 300 bp

sliding windows with 200 bp overlaps. To be included, windows had at least 3 overlapping cytosines in both genotypes with a read depth of at least 6 reads per cytosine. Windows called as significantly different were at least 10% different between tested genotypes for CHH methylation, 20% for CHG methylation and 30% for CG methylation. Significance of difference was calculated by F.E.T with a Bonferroni-Hochberg correction ($p<0.01$).

Data Access

Whole genome bisulfite sequencing data, small RNA sequencing and mRNA sequencing data are deposited in NCBI GEO under accession XXXX.

Figure Legends

Figure 1: **Loss of paternal RdDM genes but not *RDR6* represses seed abortion in paternal excess crosses. a)** Each circle represents seed abortion rate in one cross and represents multiple siliques from a single infloroscence. **b)** Each circle represents the percent of seeds that failed to germinate from each scored collection of seed. Failure to germinate was defined as the inability to produce either a radicle or a hypocotyl. Bars show median and interquartile range. * at bottom represents statistically significant difference ($p<0.05$) in comparisons between indicated cross and cross between wild-type (WT) diploid (2N) Col-0 mothers and wild-type tetraploid (4N) Col-0 fathers. * at top represents statistically significant differences ($p<0.05$) between crosses indicated. Statistical significance calculated by Wilcoxon test.

Figure 2: Lethal and viable paternal excess endosperm are transcriptionally more similar to each other than to balanced endosperm. a) PCA plot of read counts for genes from biological replicate mRNA-Seq samples. **b)** Plot of number of genes differentially expressed in comparisons of balanced endosperm with both lethal (purple) and viable (yellow) ♂ excess endosperm. Only 614 genes were differentially expressed between viable and lethal ♂ excess endosperm (gray). **c)** Correction value in viable paternal excess endosperm for each gene that was called as being significantly differentially expressed in comparisons of balanced and lethal ♂ excess endosperm.

The value was calculated as % Correction= 100-(((log2 (Viable/Balanced)/log2 (Lethal/Balanced))*100). A value of 100% indicates that the gene, which was mis-regulated in lethal ♂ excess, was not differentially expressed in viable ♂ excess relative to balanced endosperm. A value of 0% represents similar mis-regulation in both lethal and viable ♂ excess relative to balanced endosperm. Fold change values and significance for fold-change for b and c were calculated using CUFFDIFF. **d)** Lethal paternal excess endosperm was enriched for chalazal endosperm gene expression; viable paternal excess endosperm showed both chalazal and peripheral markers. Tissue enrichment for each biological replicate is shown.

Figure 3: Imprinted genes and transposons are mis-regulated in both lethal and viable ♂ excess endosperm. a-c) Expresson of Col-L*er* imprinted genes in endosperm. FPKM is normalized expression; statistical significance of difference in abundance calculated by CUFFDIFF. q<0.05 represented by black circles. q>0.05 represented by gray circles. d-f) Expression from transposable elements is elevated in lethal and viable ♂ excess endosperm. RNA-Seq reads were mapped to consensus sequences from REPBASE. Black circles represent TEs with significant differences in transcript abundances according to DEGSEQ. Gray circles represent TEs without significant differences in transcript abundances.

Figure 4: Canonical RdDM pathway function in endosperm affected by paternal excess. a) Genes encoding RdDM components are down-regulated in lethal and viable ♂ excess endosperm. *ROS1* expression is a read-out of RdDM activity and reflects differential activity of RdDM. * represents statistically significantly different gene expression. **b)** Small RNA production is impacted in paternal excess endosperm. Size profiles of sRNA reads mapped to TAIR10 genome for three replicates of balanced endosperm and two replicates each of lethal and viable paternal excess endosperm. **c)** CHH methylation losses in paternal excess endosperm. Upset plot shows intersections of CHH DMRs obtained from comparisons of balanced, lethal and viable paternal excess endosperm. Lethal and viable paternal excess endosperm share a significant proportion of regions that are hypomethylated relative to balanced endosperm. A smaller subset of DMRs lose more methylation in lethal relative to viable. **d)** Loss of

CHH methylation is associated with loss of 24 nt sRNAs. Upset plot shows the relationship between changes in sRNA abundance and CHH methylation levels. A subset of CHH DMRs are associated with loss of sRNAs. A smaller subset is associated with gains in 24 nt sRNAs. In **c** and **d**, Upset plots are used to compare different datasets. Number of features represents the total elements in each set. Sets being compared are marked by black dots connected together by a line in the matrix below the histogram. Number of intersections represents the size of the intersect. A black dot not linked by a line to any other set represent unique elements in that set. Numbers below the matrix represent the number of elements associated with the relationship in the matrix above.

Figure 5: Small RNA changes associated with TEs and genes. Frequency distribution of significant changes in sRNA abundance at **a)** genes and **b)** transposable element insertions. Fold-change and significance calculated using DESeq2. Read counts for TAIR10 TEs generated using bedtools. **c)** Changes in 21 and 24 nt sRNA abundance are correlated for genes and TEs. Fold change calculated from DESeq2 was plotted for both 21 and 24 nt sRNA abundance for TE insertions and genes. All loci that showed significant differences in 24nt sRNA levels and 21nt sRNAs are shown. **d)** Changes in 21 nt sRNA but not 24 nt sRNAs at genes are correlated with changes in mRNA abundance. Genes included in this analysis showed statistically signficant differences in both mRNA (according to CUFFDIFF) and sRNA (according to DESeq2). **e)** A subset of genes that show correlated changes in 21 and 24 nt sRNA levels also show a correlation between mRNA and 24 nt sRNA. In this analysis, differences in 21 and 24 nt sRNA abundance as well as mRNA abundance were significant. **f)** Upset plot also identifies different relationships in the abundances of 21 nt sRNAs, 24 nt sRNAs and mRNAs arising from genes in balanced and lethal paternal excess endosperm.

Figure 6: DNA methylation alterations associated with gene expression differences at two exemplar loci. a) CHH and CHG methylation at *ARF1*5 is reduced by paternal excess. b) *ARF15* expression is increased in paternal excess. Expression in viable paternal excess is intermediate between lethal paternal excess and balanced endosperm. c) CHH and CHG methylation in the first *CKX3* intron are reduced by

paternal excess. d) *CKX3* expression is decreased by paternal excess. Expression in viable paternal excess is intermediate between lethal paternal excess and balanced endosperm. In **a** and **c**, black bars below axis represent cytosines with data but no methylation. In **b** and **d**, * indicates q < 0.05 as determined by CUFFDIFF analysis.

Figure 7: Allelic contributions to expression differences between balanced and lethal paternal excess endosperm. a) Paternal fraction in genic transcripts. Boxplot represents all genes with at least 50 allele specific reads in indicated genotype. Number of genes in balanced = 10,388, lethal ♂ excess =11,709, viable ♂ excess =11,430. For balanced endosperm, a total of five replicate libraries were analyzed (Erdmann et al., 2017). **b)** Paternal excess endosperm transcriptome is maternally skewed. Frequency distribution plot of % paternal deviation for each gene. Paternal deviation was calculated for genes with at least 50 allele specific reads in each pair of genotypes being compared. Number of genes in lethal ♂ excess - balanced = 9,550, viable ♂ excess - balanced = 9,639, viable ♂ excess - lethal ♂ excess = 10,919. **c)** Decreased paternal allele contribution at a larger proportion of genes with decreased expression in lethal paternal excess compared to balanced. N = 1910 genes. Impacts of allele-specific changes on gene expression at two examples, *WOX8* and *AT2G01580*, are shown. **d)** Increased maternal allele activation at a larger proportion of genes with increased expression in lethal paternal excess. N= 1608 genes. Impacts of allele specific changes on gene expression at two examples, *SDG21* and *HEXO3*, are shown. Genes analyzed for **c** and **d** were detected as being significantly different in gene expression by CUFFDIFF. A gene with an allelic shift of at least 20% was considered to have a modified allelic balance. Boxplots represent median values for paternal deviation from genomic expectation. For gene-specific histograms, maternal allele is red and paternal allele is blue. **e)** sRNA populations are increasingly maternally biased in paternal excess endosperm. The paternal fraction of sRNA populations from viable paternal excess endosperm is intermediate between lethal paternal excess and balanced endosperm. Data from three replicates of balanced endosperm and two replicates each of lethal and viable paternal excess endosperm are plotted here. sRNA reads mapped to TAIR10

genome were first split based on size and then into Col and L*er* reads and reads using SNPs. % of paternal reads, and deviation from genomic expectation at each size was calculated.

Acknowledgements

This research was supported by NSF MCB CAREER Award 1453459 to M.G.

Author Contributions

P.R.V.S and M.G. conceived of and designed the study, P.R.V.S. performed experiments and analyzed data, and P.R.V.S. and M.G. wrote the manuscript.

References

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA *6*, 11.

Baroux, C., Gagliardini, V., Page, D.R., and Grossniklaus, U. (2006). Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis. Genes Dev. *20*, 1081–1086.

Baubec, T., Dinh, H.Q., Pecinka, A., Rakic, B., Rozhon, W., Wohlrab, B., von Haeseler, A., and Scheid, O.M. (2010). Cooperation of Multiple Chromatin Modifications Can Generate Unanticipated Stability of Epigenetic States in Arabidopsis[C][W][OA]. Plant Cell *22*, 34–47.

Birchler, J.A. (2014). Interploidy hybridization barrier of endosperm as a dosage interaction. Front. Plant Sci. *5*.

Birchler, J.A., and Veitia, R.A. (2012). Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. Proc. Natl. Acad. Sci. *109*, 14746–14753.

Borges, F., and Martienssen, R.A. (2015). The expanding world of small RNAs in plants. Nat. Rev. Mol. Cell Biol. *16*, 727–741.

Borges, F., Parent, J.-S., Ex, F. van, Wolff, P., Martínez, G., Köhler, C., and Martienssen, R.A. (2018). Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in Arabidopsis. Nat. Genet. *50*, 186.

Castillo, D.M., and Moyle, L.C. (2012). Evolutionary Implications of Mechanistic Models of TE-Mediated Hybrid Incompatibility.

Cooper, D.C. (1951). Caryopsis Development Following Matings Between Diploid and Tetraploid Strains of Zea Mays. Am. J. Bot. *38*, 702–708.

Creasey, K.M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B.C., and Martienssen, R.A. (2014). miRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*. Nature *508*, 411–415.

Dilkes, B.P., Spielman, M., Weizbauer, R., Watson, B., Burkart-Waco, D., Scott, R.J., and Comai, L. (2008). The Maternally Expressed WRKY Transcription Factor TTG2 Controls Lethality in Interploidy Crosses of Arabidopsis. PLOS Biol. *6*, e308.

Disteche, C.M. (2016). Dosage compensation of the sex chromosomes and autosomes. Semin. Cell Dev. Biol. *56*, 9–18.

Doughty, J., Aljabri, M., and Scott, R.J. (2014). Flavonoids and the regulation of seed size in Arabidopsis. Biochem. Soc. Trans. *42*, 364–369.

Erdmann, R.M., Satyaki, P.R.V., Klosinska, M., and Gehring, M. (2017). A Small RNA Pathway Mediates Allelic Dosage in Endosperm. Cell Rep. *21*, 3364–3372.

d'Erfurth, I., Jolivet, S., Froger, N., Catrice, O., Novatchkova, M., and Mercier, R. (2009). Turning Meiosis into Mitosis. PLOS Biol. *7*, e1000124.

Esen, A., and Soost, R.K. (1973). Seed Development in Citrus with Special Reference to 2x X 4x Crosses. Am. J. Bot. *60*, 448–462.

Figueiredo, D.D., Batista, R.A., Roszak, P.J., Hennig, L., and Köhler, C. (2016). Auxin production in the endosperm drives seed coat development in Arabidopsis. ELife *5*, e20542.

Fiume, E., and Fletcher, J.C. (2012). Regulation of Arabidopsis Embryo and Endosperm Development by the Polypeptide Signaling Molecule CLE8[C][W]. Plant Cell *24*, 1000–1012.

Gasciolli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. (2005). Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. Curr. Biol. CB *15*, 1494–1500.

Gehring, M., Huh, J.H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA Glycosylase Establishes MEDEA Polycomb Gene Self-Imprinting by Allele-Specific Demethylation. Cell *124*, 495–506.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009). Extensive Demethylation of Repetitive Elements During Seed Development Underlies Gene Imprinting. Science *324*, 1447–1451.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in Arabidopsis thaliana seeds. PloS One *6*, e23687.

Gutierrez-Marcos, J.F., Pennington, P.D., Costa, L.M., and Dickinson, H.G. (2003). Imprinting in the endosperm: a possible role in preventing wide hybridization. Philos. Trans. R. Soc. Lond. B Biol. Sci. *358*, 1105–1111.

Haig, and Westboy (1991). Genomic imprinting in endosperm: its effect on seed development in crosses between species, and its implications for the evolution of apomixis. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *333*, 1–13.

Håkansson, A., and Ellerström, S. (1950). SEED DEVELOPMENT AFTER RECIPROCAL CROSSES BETWEEN DIPLOID AND TETRAPLOID RYE. Hereditas *36*, 256–296.

Hehenberger, E., Kradolfer, D., and Köhler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. Development *139*, 2031–2039.

Hsieh, T.-F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D. (2009). Genome-Wide Demethylation of Arabidopsis Endosperm. Science *324*, 1451–1454.

Huang, F., Zhu, Q.-H., Zhu, A., Wu, X., Xie, L., Wu, X., Helliwell, C., Chaudhury, A., Finnegan, E.J., and Luo, M. (2017). Mutants in the imprinted PICKLE RELATED 2 gene suppress seed abortion of fertilization independent seed class mutants and paternal excess interploidy crosses in Arabidopsis. Plant J. Cell Mol. Biol. *90*, 383–395.

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.-F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., et al. (2012). Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. Science *337*, 1360–1364.

Jiang, H., Moreno-Romero, J., Santos-González, J., Jaeger, G.D., Gevaert, K., Slijke, E.V.D., and Köhler, C. (2017). Ectopic application of the repressive histone modification H3K9me2 establishes post-zygotic reproductive isolation in Arabidopsis thaliana. Genes Dev.

Jullien, P.E., Katz, A., Oliva, M., Ohad, N., and Berger, F. (2006). Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis. Curr. Biol. CB *16*, 486–492.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. *110*, 462–467.

Kang, I.-H., Steffen, J.G., Portereiko, M.F., Lloyd, A., and Drews, G.N. (2008). The AGL62 MADS Domain Protein Regulates Cellularization during Endosperm Development in Arabidopsis. Plant Cell *20*, 635–647.

Kelleher, E.S. (2016). Reexamining the P-Element Invasion of Drosophila melanogaster Through the Lens of piRNA Silencing. Genetics *203*, 1513–1531.

Khan, A., and Mathelier, A. (2017). Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics *18*, 287.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kradolfer, D., Wolff, P., Jiang, H., Siretskiy, A., and Köhler, C. (2013). An imprinted gene underlies postzygotic reproductive isolation in Arabidopsis thaliana. Dev. Cell *26*, 525–535.

Lafon-Placette, C., and Köhler, C. (2015). Epigenetic mechanisms of postzygotic reproductive isolation in plants. Curr. Opin. Plant Biol. *23*, 39–44.

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

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. IEEE Trans. Vis. Comput. Graph. *20*, 1983–1992.

Li, J., and Berger, F. (2012). Endosperm: food for humankind and fodder for scientific discoveries. New Phytol. *195*, 290–305.

Li, J., Nie, X., Tan, J.L.H., and Berger, F. (2013). Integration of epigenetic and genetic controls of seed size by cytokinin in Arabidopsis. Proc. Natl. Acad. Sci. *110*, 15479–15484.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lu, J., Zhang, C., Baulcombe, D.C., and Chen, Z.J. (2012). Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. Proc. Natl. Acad. Sci. U. S. A. *109*, 5529–5534.

Marí-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing *de novo* silencing of an active plant retrotransposon. Nat. Genet. *45*, 1029–1039.

Martienssen, R.A. (2010). Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of Arabidopsis. New Phytol. *186*, 46–53.

Martinez, G., Wolff, P., Wang, Z., Moreno-Romero, J., Santos-González, J., Conze, L.L., DeFraia, C., Slotkin, R.K., and Köhler, C. (2018). Paternal easiRNAs regulate parental genome dosage in Arabidopsis. Nat. Genet. *50*, 193.

McCue, A.D., Nuthikattu, S., Reeder, S.H., and Slotkin, R.K. (2012). Gene Expression and Stress Response Mediated by the Epigenetic Regulation of a Transposable Element Small RNA. PLOS Genet. *8*, e1002474.

Milbocker, D., and Sink, K. (1969). Embryology of diploid × diploid and diploid × tetraploid crosses in poinsettia. Can. J. Genet. Cytol. *11*, 598–601.

Mileyko, Y., Joh, R.I., and Weitz, J.S. (2008). Small-scale copy number variation and large-scale changes in gene expression. Proc. Natl. Acad. Sci. *105*, 16659–16664.

Mittelsten Scheid, O., Afsar, K., and Paszkowski, J. (2003). Formation of stable epialleles and their paramutation-like interaction in tetraploid *Arabidopsis thaliana*. Nat. Genet. *34*, 450–454.

Moreno-Romero, J., Jiang, H., Santos-González, J., and Köhler, C. (2016). Parental epigenetic asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm. EMBO J. *35*, 1298–1311.

Muntzing, A. (1936). The Evolutionary Significance of Autopolyploidy. Hereditas *XXI*.

Oberlin, S., Sarazin, A., Chevalier, C., Voinnet, O., and Marí-Ordóñez, A. (2017). A genome-wide transcriptome and translatome analysis of Arabidopsis transposons identifies a unique and conserved genome expression strategy for Ty1/Copia retroelements. Genome Res. *27*, 1549–1562.

Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W., and Gehring, M. (2014). Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. ELife *3*, e03198.

Pignatta, D., Bell, G., and Gehring, M. (2015). Whole Genome Bisulfite Sequencing and DNA Methylation Analysis from Plant Tissue. BIO-Protoc. *5*.

Pignatta, D., Novitzky, K., Satyaki, P.R.V., and Gehring, M. (2018). A variably imprinted epiallele impacts seed development. PLOS Genet. *14*, e1007469.

Piskurewicz, U., Iwasaki, M., Susaki, D., Megies, C., Kinoshita, T., and Lopez-Molina, L. (2016). Dormancy-specific imprinting underlies maternal inheritance of seed dormancy in Arabidopsis thaliana. ELife *5*, e19573.

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. Proc. Biol. Sci. *285*.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Satyaki, P.R.V., and Gehring, M. (2017). DNA methylation and imprinting in plants: machinery and mechanisms. Crit. Rev. Biochem. Mol. Biol. *52*, 163–175.

Schatlowski, N., Wolff, P., Santos-González, J., Schoft, V., Siretskiy, A., Scott, R., Tamaru, H., and Köhler, C. (2014). Hypomethylated Pollen Bypasses the Interploidy Hybridization Barrier in Arabidopsis[C][W]. Plant Cell *26*, 3556–3568.

Schon, M.A., and Nodine, M.D. (2017). Widespread Contamination of Arabidopsis Embryo and Endosperm Transcriptome Data Sets. Plant Cell *29*, 608–617.

Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. Development *125*, 3329–3341.

Stoute, A.I., Varenko, V., King, G.J., Scott, R.J., and Kurup, S. (2012). Parental genome imbalance in Brassica oleracea causes asymmetric triploid block: Parent-of-origin effects in Brassica oleracea. Plant J. no-no.

Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell *152*, 352–364.

Tiwari, S., Spielman, M., Schulz, R., Oakey, R.J., Kelsey, G., Salazar, A., Zhang, K., Pennell, R., and Scott, R.J. (2010). Transcriptional profiles underlying parent-of-origin effects in seeds of Arabidopsis thaliana. BMC Plant Biol. *10*, 72.

Tsay, Y.F., Frank, M.J., Page, T., Dean, C., and Crawford, N.M. (1993). Identification of a mobile endogenous transposon in Arabidopsis thaliana. Science *260*, 342–344.

Varrault, A., Gueydan, C., Delalbre, A., Bellmann, A., Houssami, S., Aknin, C., Severac, D., Chotard, L., Kahli, M., Le Digarcher, A., et al. (2006). Zac1 Regulates an Imprinted Gene Network Critically Involved in the Control of Embryonic Growth. Dev. Cell *11*, 711–722.

Vu, T.M., Nakamura, M., Calarco, J.P., Susaki, D., Lim, P.Q., Kinoshita, T., Higashiyama, T., Martienssen, R.A., and Berger, F. (2013). RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis. Dev. Camb. Engl. *140*, 2953–2960.

Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D., and Feng, X. (2018). Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. Nat. Genet. *50*, 130.

Wang, G., Jiang, H., Del Toro de León, G., Martinez, G., and Köhler, C. (2018). Sequestration of a Transposon-Derived siRNA by a Target Mimic Imprinted Gene Induces Postzygotic Reproductive Isolation in Arabidopsis. Dev. Cell *46*, 696-705.e4.

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinforma. Oxf. Engl. *26*, 136–138.

von Wangenheim, K.-H., and Peterson, H.-P. (2004). Aberrant endosperm development in interploidy crosses reveals a timer of differentiation. Dev. Biol. *270*, 277–289.

Williams, B.P., Pignatta, D., Henikoff, S., and Gehring, M. (2015). Methylation-Sensitive Expression of a DNA Demethylase Gene Serves As an Epigenetic Rheostat. PLOS Genet. *11*, e1005142.

Wolff, P., Jiang, H., Wang, G., Santos-González, J., and Köhler, C. (2015). Paternally expressed imprinted genes establish postzygotic hybridization barriers in Arabidopsis thaliana. ELife *4*.

Xiao, W., Brown, R.C., Lemmon, B.E., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). Regulation of Seed Size by Hypomethylation of Maternal and Paternal Genomes. Plant Physiol. *142*, 1160–1168.

Yan, D., Duermeyer, L., Leoveanu, C., and Nambara, E. (2014). The Functions of the Endosperm During Seed Germination. Plant Cell Physiol. *55*, 1521–1533.

Yu, Z., Haberer, G., Matthes, M., Rattei, T., Mayer, K.F.X., Gierl, A., and Torres-Ruiz, R.A. (2010). Impact of natural genetic variation on the transcriptome of autotetraploid Arabidopsis thaliana. Proc. Natl. Acad. Sci. *107*, 17809–17814.

Zhang, J., Liu, Y., Xia, E.-H., Yao, Q.-Y., Liu, X.-D., and Gao, L.-Z. (2015). Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. Proc. Natl. Acad. Sci. U. S. A. *112*, E7022-7029.

Zhang, Y., Malone, J.H., Powell, S.K., Periwal, V., Spana, E., MacAlpine, D.M., and Oliver, B. (2010). Expression in Aneuploid Drosophila S2 Cells. PLOS Biol. *8*, e1000320.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. *39*, 61–69.

Supplemental Data Files

Supplemental Figure 1: Paternal loss of RdDM genes but not *RDR6* and *CMT3* represses seed abortion in paternal excess crosses

Supplemental Figure 2: Analysis of gene expression differences in paternal excess endosperm derived using the *osd1* mutation

Supplemental Figure 3: Rescue by loss of NRPD1 does not require normalization of genes implicated in interploid seed abortion

Supplemental Figure 4: Small RNA and DNA methylation in Balanced, Lethal and Viable paternal excess endosperm

Supplemental Figure 5: Association between DNA methylation changes and gene expression

Supplemental Figure 6: Allelic contributions to gene expression differences

Supplemental Figure 7: Allelic bias of some imprinted genes is lost in Lethal paternal excess

Supplemental Table 1: Details of sequencing libraries

Supplemental Dataset 1: CUFFDIFF output comparing gene expression in balanced, lethal and viable paternal excess endosperm

Supplemental Dataset 2: Gene ontology analysis of genes that are differentially expressed between balanced, lethal and viable paternal excess endosperm

Supplemental Dataset 3: DEGseq output comparing transposon transcript levels between balanced, lethal and viable paternal excess endosperm

Supplemental Dataset 4: Differentially methylated regions in the CHH context in comparisons between balanced, lethal and viable paternal excess endosperm

Supplemental Dataset 5: Differentially methylated region in the CHG context in comparisons between balanced, lethal and viable paternal excess endosperm

Supplemental Dataset 6: Differentially methylated region in the CG context in comparisons between balanced, lethal and viable paternal excess endosperm

Supplemental Dataset 7: DESeq output for differential abundance of 21 and 24 nt genic sRNA between balanced, lethal and viable paternal excess endosperm

Supplemental Dataset 8: DESeq output for differential abundance of 21 and 24 nt TE sRNA between balanced, lethal and viable paternal excess endosperm

Supplemental Dataset 9: Allele-specific mRNA read counts and differences in allelic contributions between balanced, lethal and viable paternal excess endosperm