Plant Mobile Domain protein-DNA motif modules counteract Polycomb silencing to stabilize gene expression

Thierry Pélissier[1#], Lucas Jarry[2 #], Margaux Olivier[1], Gabin Dajoux[1], Marie-Noëlle Pouch-Pélissier[1], Charles Courtois[1], Julie Descombin[2], Nathalie Picault[2], Guillaume Moissiard[2*], and Olivier Mathieu[1*]

[1] Institute of Genetics Reproduction and Development (iGReD), Université Clermont Auvergne, CNRS-UMR6293, Inserm-U1103, F-63000 Clermont-Ferrand, France.

[2] LGDP-UMR5096, CNRS, Université de Perpignan Via Domitia, F-66000 Perpignan, France.

[#] These authors contributed equally to this work

* Corresponding authors: E-mail: guillaume.moissiard@univ-perp.fr (G.M.); olivier.mathieu@uca.fr (O.M.)

Abstract

In plants and animals, Polycomb group (PcG) proteins are crucial for development, regulating gene expression through H3K27me3 deposition and subsequent gene silencing. While Polycomb silencing target specification is increasingly understood, it remains unclear how certain genes with apparent silencing-attracting features escape this process. Here, we show that the plant mobile domain C (PMD-C) containing proteins MAINTENANCE OF MERISTEMS (MAIN), MAIN-LIKE 1 (MAIL1) and MAIL2 oppose Polycomb silencing at numerous actively transcribed genes in Arabidopsis. Mutations in *MAIN*, *MAIL1* or *MAIL2* result in PcG-dependent ectopic H3K27me3 deposition, often associated with transcriptional repression. We show that MAIL1, which functions in concert with MAIN, and MAIL2 target distinct gene sets and associate with chromatin at specific DNA sequence motifs. We demonstrate that the integrity of these motif sequences is essential for promoting expression and antagonizing H3K27me3 deposition. Our results unveil a novel system opposing Polycomb silencing involving PMD-C protein-DNA motif modules, expanding our understanding of eukaryotic gene regulation mechanisms.

Main

Chromatin-based modulation of gene expression is mediated by enzymatic complexes that catalyze post-translational histone modifications and alter chromatin structure. The evolutionarily conserved Polycomb group (PcG) proteins act as a cellular memory system, mediating epigenetic gene repression. PcG proteins form multi-subunit complexes, primarily Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2), which often work cooperatively to establish and maintain gene repression. PRC2 catalyzes the trimethylation of lysine 27 on histone H3 (H3K27me3), a hallmark generally associated with repressed chromatin, while PRC1 mediates ubiquitination on lysine 121 of H2A (H2AK121ub) in plants (or lysine 119 in animals). In the *Arabidopsis thaliana* genome, these two histone marks frequently colocalize, with PRC1-mediated H2AK121ub being required for the incorporation of H3K27me3 at most genes targeted by PcG-mediated silencing[1].

The dynamic regulation of gene expression necessitates mechanisms to counteract PcG-mediated regulation. In *Arabidopsis*, several factors have been identified that oppose PcG activity. These include several histone methyltransferases of the ARABIDOPSIS TRITHORAX (ATX) and ARABIDOPSIS TRITHORAX RELATED (ATXR) protein families, which catalyze H3K4 methylation, thereby promoting active chromatin states[2–5]. Another important group of proteins counteracting PcG-mediated silencing are H3K27me3 demethylases. In Arabidopsis, these include the Jumonji C domain-containing histone demethylases EARLY FLOWERING 6 (ELF6), RELATIVE OF EARLY FLOWERING 6 (REF6), and JMJ13[6–9]. Additionally, ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 1 (ALP1) and ALP2 have been proposed to interfere with PcG silencing at a small subset of PcG targets by associating with PRC2 and potentially inhibiting its activity[10,11], although mutations in either *ALP* genes do not alter H3K27me3 levels.

We previously reported that two plant-specific proteins, MAINTENANCE OF MERISTEMS (MAIN) and MAIN-LIKE 1 (MAIL1), are important factors for plant development[12–14]. These proteins are characterized by a Plant Mobile Domain of type C (PMD-C) and have been implicated in the regulation of transposable elements and gene expression. However, their precise molecular functions and potential roles in epigenetic regulation remain poorly understood. Here we demonstrate that MAIN, MAIL1 and the

homologous protein MAIL2 counteract PcG-mediated silencing at specific genomic loci by preventing ectopic H3K27me3 incorporation.

Results

Mutations in *NDX* and PRC core components alleviate *mail1*-associated phenotypes

Mutations in components of the MAIL1/MAIN/PP7L complex result in severe developmental defects and transcriptional abnormalities characterized by loss of silencing at several TEs and mis-expression of numerous protein-coding genes[12–15]. To uncover the molecular basis of these defects, we conducted a genetic screen for suppressors of the *mail1* mutant phenotypes. Screening individuals from an ethyl methanesulfonate (EMS) mutagenized population of *mail1* containing the *L5-GUS* transgene[16], we identified a mutant line exhibiting both improved plant development and suppression of L5 silencing release (Fig. 1a-b). Compared to *mail1* plants, the suppressed plants showed faster growth and greener leaves. The traits followed a 1/3 ratio in the F2 progeny from a backcross, indicating a single-locus nuclear recessive causative mutation. Mapping-by-sequencing using an outcross F2 population identified a knock-out allele of *AT4G03090* (*NODULIN HOMEOBOX; NDX*) that we named *ndx-7*, associated with a C to T mutation leading to a premature stop codon (Supplementary Fig. 1a). Crossing *mail1* plants with an independent *ndx-4* T-DNA insertion mutant line and introduction of an *NDX* genomic copy in the *mail1 ndx-7* background confirmed that mutation of *NDX* alleviated *mail1*-associated developmental and silencing defects (Supplementary Fig. 1b-d). Beyond the *L5-GUS* transgene, RT-qPCR analysis revealed strong suppression of *mail1*-transcriptional activation of several endogenous targets in *mail1 ndx-7* plants (Fig. 1c). Consistent with MAIL1's role in a complex with MAIN, introducing *ndx-7* in the *main* background also mitigated *main*-associated developmental defects and GUS upregulation (Supplementary Fig. 1b,d).

NDX is a transcription factor recently found to directly interact with PRC1 core components RING1A and RING1B, regulating a set of ABA-responsive genes[17,18]. Introgression of the *mail1* mutation into a *ring1a ring1b* (*ring1a/1b*) double mutant backgound[18] resulted in a similar reversal of *mail1*-associated phenotypic and transcription defects (Fig. 1b-c). NDX can also interact with VIVIPAROUS1/ABI3-LIKE (VAL1), a transcription factor involved in PRC1 recruitment at the *FLC* gene[19–21]. Crossing *val1* plants,

which appear indistinguishable from WT, with *mail1* or *main* mutant lines resulted in double mutants with very severe developmental phenotype characterized by the apparent absence of functional shoot apical meristem, and unable to initiate post-embryonic differentiation (Supplementary Fig. 1e). This indicates that *ndx* and *ring1a/1b* suppression of *mail1* is independent of *VAL1*.

To assess the genome-wide importance of *NDX* and *RING1A/1B* on *mail1*-associated transcriptional defects, we performed RNA-seq experiments using WT, *ndx*, *ring1a/1b*, *mail1*, *mail1 ndx* and *mail1 ring1a/1b* lines. Using four *mail1* RNA-seq replicates we defined a robust set of *mail1*-specific misregulated targets, primarily consisting of differentially expressed genes and a set of upregulated TEs (Fig. 2a and Supplementary Fig. 2a). The expression of most of these loci was unaffected in *ndx* or *ring1a/1b* mutants (Fig. 2a). Combining either *ndx* or *ring1a/1b* mutations with *mail1* resulted in an overall suppression of TE upregulation (Fig. 2a-b). Similarly, expression levels of *mail1*-upregulated genes tend to significantly return towards WT levels in *mail1 ndx* and *mail1 ring1a/1b* (Fig. 2a-b). Additionally, *mail1*-associated downregulation was suppressed for about 65% of the genes (62 to 68/95) in both mutant combinations (Fig. 2a-b; Supplementary Fig. 2b). These results indicate that *NDX* and *RING1A/1B* are required for the transcriptional misregulation induced in the absence of *MAIL1*.

To determine the impact of mutations in the PRC2 core component CLF or SWN, we generated *mail1 clf*-29 and *mail1 swn*-7 double mutants and performed RNA-seq. CLF and SWN act largely redundantly to catalyze H3K27me3 deposition in vegetative tissues, although certain loci appear to be preferentially targeted by one or other H3K27 methyltransferases[22]. We found that *mail1*-induced transcriptional misregulation at genes was significantly suppressed in *mail1 clf* but not in *mail1 swn* (Supplementary Fig. 2c), although both double mutants showed developmental defects mostly similar to those of *mail1* (Supplementary Fig. 2d). This indicates that mutation of *CLF* tends to suppress *mail1*-induced transcriptional defects, albeit to an insufficient extent to result in an improved phenotype, most likely because of the functional redundancy with SWN[22].

Our results show that *NDX, RING1A/1B,* and to a lesser extent *CLF*, are required for the transcriptional misregulation induced in the absence of *MAIL1,* suggesting a potential link between MAIL1 and PcG-mediated silencing.

MORC1 and MORC2 dampen TE activation in *mail1* suppressor backgrounds

MICRORCHIDIA (MORC) proteins are ATPases required for TE silencing and heterochromatin condensation[23]. We recently linked silencing defects at a subset of TEs to downregulation of MORC1 in *mail1* plants[15]. In both suppressor backgrounds, loss of TE activation correlated with increased *MORC1* transcript accumulation relative to *mail1*, albeit limited in *mail1 ndx* plants (Supplementary Fig. 3a). Interestingly, expression of the closely related *MORC2* gene, also involved in TE silencing[24], showed a four to six fold induction in the presence of *ndx* or *ring1a/1b* mutations (Supplementary Fig. 3a). We tested whether *MORC2* could contribute to the reestablishment of TE silencing by introgressing the *morc2-1* mutation into the *mail1 ndx* background and performing RNA-seq to monitor TE expression. Of the 25 *mail1*-upregulated TEs, at least six showed dependency towards MORC2 for their repression in *mail1 ndx* plants (Supplementary Fig. 3b-c), and we previously showed that overexpressing *MORC1* in the *mail1* background also repressed some *mail1*-activated TEs[15] (Supplementary Fig. 3d). This is consistent with MORC1 and MORC2 acting redundantly in TE silencing[24] and suggests that upregulation of both *MORC1* and *MORC2* contributes to dampening TE activation in *mail1 ndx* and *mail1 ring1a/1b*.

Gene silencing in *mail1* correlates with ectopic H3K27me3 incorporation

In light of the suppressor effect of *ndx* and *ring1a/1b* mutations and the known interaction of NDX and RING1A/1B in PcG-mediated silencing[17,19], we explored a possible link between *mail1*-associated gene misregulation and the PcG-mediated silencing diagnostic mark H3K27me3. Chromatin immunoprecipitation sequencing (ChIP-seq) analyses revealed no overall change in H3K27me3 levels in *mail1* when considering all Arabidopsis genes (Fig. 2c). In *mail1,* upregulated genes showed slight loss of H3K27me3. The addition of the *ndx* mutation had no additional effect, while H3K27me3 levels were further reduced at these genes in *mail1 ring1a/1b* mutants. Despite this reduction, both *ndx* and *ring1a/1b* mutations tend to suppress *mail1*-induced transcriptional activation of these genes (Fig. 2a), suggesting that their misregulation in *mail1* is not directly connected to H3K27me3 variation. Strikingly, gene downregulation in *mail1* correlated with prominent H3K27me3 enrichment, which was largely reversed upon additional *NDX* or *RING1A/1B* depletions (Fig. 2c) that also tended to restore expression of these genes (Fig. 2a). Many genes acquired high levels of H3K27me3 *de novo* in *mail1,* correlating with transcriptional repression, and

this was largely reversed in the *mail1 ndx* and *mail1 ring1a/1b* suppressed lines (Fig. 2d-e and Supplementary Fig. 4a). One such gene was *MORC1*, gaining H3K27me3 in *mail1*, which was completely lost in *mail1 ring1a/1b* (Supplementary Fig. 4a). We also found that its neighbor gene *MORC2* was subject to PcG-mediated silencing in the WT, explaining its upregulation in the presence of *ndx* or *ring1a/1b* mutations (Supplementary Fig. 3a, 4a). Interestingly, H3K27me3-enriched genes in *mail1* included *SPO11-1* (Supplementary Fig.4a), which is mostly expressed in flowers and encodes a topoisomerase-like protein essential for meiotic recombination and fertility[25,26]. Gain of H3K27me3 was maintained in *mail1* immature flowers, correlating with about a 70% decrease in *SPO11-1* transcript accumulation (Supplementary Fig. 4b). Such level of *SPO11-1* downregulation was reported to significantly affect gamete viability[26], and thus likely contributes to *mail1* hypofertility[27].

RING1A and RING1B are largely required for PRC1-mediated H2A ubiquitination[18], playing a key role in recruiting H3K27me3 at numerous loci[1]. This is likely the case at genes gaining H3K27me3 in *mail1* (Supplementary Table 1) as these showed concomitant enrichment in H2AK121ub (Supplementary Fig. 5a). Together, these results suggest that loss of *MAIL1* directly leads to PcG-mediated silencing at many genes.

To search for possible common features associated with the gain of H3K27me3 at genes in *mail1*, we performed DNA motif-based sequence analysis using the 1-kb region surrounding the transcription start site (TSS) of these genes. This revealed significant enrichment for four DNA motifs in about 60 to 70% of the genes (Fig. 2f). The top predicted motif, which we named M1M (for MAIL1/MAIN Motif), extends a previously reported "down" motif found partially enriched among *mail1* and *main*-downregulated genes[13]. Interestingly, the other motifs included telobox and GA-repeat elements, two known polycomb response elements (PREs) bound by transcription factors to tether PRC2 complexes to specific genes[22,28,29]. Using previously published data[30], we found that the TELOMERE-REPEAT BINDING FACTOR (TRB)1, TRB2 and TRB3, which bind telobox elements were loaded in wild-type plants close to the TSS of a significant fraction (31/49) of *mail1* H3K27me3-enriched genes (Fig. 2e and Supplementary Fig. 5b-c). Together, these results suggest that the MAIL1/MAIN complex might promote the expression of some genes by preventing, directly or indirectly, PcG-mediated silencing. Upon MAIL1/MAIN depletion, PRC2 recruitment, through TRB proteins or other factors, requires PRC1 activity to be robustly established, potentially through a positive feedback loop[1].

MAIL1/MAIN complexes are specifically targeted to M1M motifs

To explore the genomic localizations of MAIL1 and MAIN *in vivo*, we used transgenic lines expressing functional FLAG- or MYC-tagged versions of MAIL1 or MAIN driven by their endogenous promoters in their respective mutant backgrounds[13] and performed ChIP-seq. MAIL1 and MAIN largely colocalized and were preferentially enriched at proximal regions of gene promoters (Fig. 3a-d). Noticeably, MAIL1 and MAIN displayed strong enrichment at the promoters of genes that gained H3K27me3 in *mail1* but not at genes showing decreased H3K27me3 levels (Fig. 3e, Supplementary Fig. 6a). In line with MAIL1 and MAIN acting in the same protein complex, *mail1* H3K27me3-enriched genes showed similar transcriptional downregulation and increased H3K27me3 levels in *mail1, main,* and *mail1 main* mutants (Fig. 3f). Furthermore, we searched for DNA motifs enriched at MAIL1/MAIN associated regions and found that 86% of these regions contain the M1M motif (Supplementary Fig. 6b, Fig. 2f), strongly suggesting a functional role for M1M in recruiting MAIL1/MAIN at the proximal promoter of genes (Fig. 2e, 3e).

MAIL2 function is essential to plant development and opposes PcG-mediated silencing at distinct loci

We previously reported that *MAIL1* and *MAIN* are Brassicaceae-specific paralogues of *MAIL2*, which is widely distributed among plant species[13]. The *MAIL2* gene appears to be under strong purifying selection[13] suggesting that it may fulfill an important function. In line with this assumption, the *mail2-1* transfer DNA mutant displayed a very severe developmental phenotype. The SAM was unable to initiate a proper post-embryonic program, with plants rarely developing true leaves and typically arresting growth at the two cotyledons stage (Fig. 4a; Supplementary Fig. 7a-b). Notably, we were unable to obtain homozygous mutants in the progeny segregating from a self-fertilized heterozygous *mail2* CRISPR mutant, suggesting that *mail2-1* may not be a null mutant.

The MAIL1, MAIN, and MAIL2 proteins share a highly homologous PMD-C domain[27] (Supplementary Fig. 8), which is the only identified domain in these proteins. Given the similarity in structure, we hypothesized that MAIL2 might be involved in a protection mechanism against polycomb silencing similar to the one identified for MAIL1/MAIN. To test this hypothesis, we introduced *ndx*, *clf*, or *ring1a/1b* mutations into the *mail2-1* background

through crossing. The resulting double mutants showed varying degrees of phenotypic suppression: *mail2-1 ndx* and *mail2-1 clf* double mutants developed a first pair of leaves before arresting growth, while *mail2-1 ring1a/1b* mutants displayed more extensive development, progressing through the rosette stage and approaching the floral transition (Fig. 4a; Supplementary Fig. 7a). These observations indicate that *ndx*, *clf*, and *ring1a/1b* mutations partially suppress the *mail2-1* developmental phenotype, mirroring their effects on *mail1* and *main* mutants. This suggests a functional similarity between MAIL2 and MAIL1/MAIN in counteracting PcG-mediated silencing, implying that this anti-PcG silencing function may be an ancestral feature of the PMD-C protein family.

We determined MAIL2 genomic localization by ChIP-seq, using a transgenic line expressing a complementing tagged version of MAIL2 (Supplementary Fig. 7c). Our analysis identified 1775 MAIL2-enriched regions throughout the genome, which were largely distinct from MAIL1/MAIN-enriched regions (Supplementary Fig. 9a). However, as in the case of MAIL1 and MAIN, these regions were predominantly associated with genes, particularly their proximal promoters (1059 genes) (Fig. 4b; Supplementary Fig. 9b). Those genes were significantly enriched in the five deciles showing the highest expression level in the WT (Fig. 4c).

We performed DNA motif-based sequence analyses using MAIL2 peaks or the 1kb region surrounding TSS of MAIL2-associated gene promoters and identified a sequence element, which we named MAIL2-Motif (M2M), present in nearly 95% of the peaks and promoter regions (Fig. 4d; Supplementary Fig. 9c-d). Intriguingly, the consensus sequence of M2M resembles that of M1M but in a palindromic version, suggesting M1M may have derived from M2M (Supplementary Fig. 9e), mirroring the evolutionary relationship between the *MAIL1/MAIN* genes and *MAIL2*. MAIL2-associated gene promoters were additionally enriched in GAA-repeats and telobox elements (Fig. 4d; Supplementary Fig. 9c), similar to MAIL1/MAIN-associated genes.

To facilitate exploration of MAIL2 function, we generated hypomorphic *mail2* mutants using artificial microRNAs (amiRNAs) directed against *MAIL2* mRNA (amiR-*mail2*) (Fig. 4a; Supplementary Fig. 7b). While many amiR-*mail2* plants displayed severely affected SAM under *in vitro* selection (Supplementary Fig. 10a) we were able to isolate two amiR-*mail2* lines (amiR-*mail2_4* and amiR-*mail2_5*) that developed further. Homozygous amiR-*mail2_4* and amiR-*mail2_5* mutants showed around 2.5-fold reduction of *MAIL2* mRNA

accumulation (Supplementary Fig. 10b) and exhibited similar developmental phenotypes, characterized by shorter stature, curly leaves, short stems and hypofertility, which were partially suppressed by the introduction of *ring1a/1b* mutations (Fig. 4a; Supplementary Fig. 10c). We confirmed that these phenotypes were caused by *MAIL2* downregulation by restoring MAIL2 function in amiR-*mail2_4* using an amiRNA-insensitive *MAIL2* transgene (Supplementary Fig. 7b, 10d). We evaluated the impact of *MAIL2* downregulation on transcription by performing RNA-seq of amiR-*mail2_4* and amiR-*mail2_5* lines. The transcriptomes of the two lines were highly similar, with differentiated expressed loci almost exclusively comprising genes, either up- (197) or down-regulated (96 genes) (Supplementary Table 2). Genes downregulated in amiR-*mail2* lines were significantly enriched for MAIL2-associated genes, which belong to the five quantiles showing the highest expression level in WT plant (Fig. 4c,e). These genes were distinct from those downregulated in *mail1* (Supplementary Fig. 11a) and introducing amiR-*mail2* in *mail1* background drastically worsen *mail1* plant phenotype (Supplementary Fig. 11b), indicating that MAIL1 and MAIL2 act in different pathways to regulate expression of distinct genes. Gene Ontology term analysis of both sets of downregulated genes did not reveal any strong bias towards a particular biological pathway (Supplementary Fig. 11c). We further found that genes showing MAIL2 binding to their promoter and downregulation in amiR-*mail2* were associated with TRB1/2/3 in WT (Fig. 4f; Supplementary Fig. 11d-e), and specifically gained H3K27me3 upon *MAIL2* downregulation (Fig. 4f-g). We conclude from these results that MAIL2 and MAIL1 are involved in similar pathways antagonizing PcG-silencing to promote expression of largely distinct sets of genes

M1M and M2M are required to promote expression and oppose H3K27me3 deposition

To evaluate the importance of M1M and M2M motifs, we undertook two complementary approaches based on transgenes. In the first approach, we used a transgenic construct containing the promoter of the *SEP3* gene driving *GUS* expression[29]. The *SEP3* promoter contains a telobox that allows PRC2-mediated H3K27me3 deposition through TRB1 binding, resulting in low *GUS* expression[29]. We found that introducing M1M or M2M in the *SEP3* promoter between the telobox and the transgene's TSS significantly enhanced *GUS* expression and decreased H3K27me3 levels (Fig. 5a). In a complementary approach, we created transgenic constructs using promoters of two genes associated with either

MAIL1/MAIN (*KPI-1*) or MAIL2 (*HD2A*) that shows downregulation and H3K27me3 enrichment in *mail1* or *mail2* mutant background, respectively (Fig. 5b). These promoters contained teloboxes and either M1M (KPI-1 promoter) or M2M (HD2A promoter) motifs. We independently transformed WT plants with intact or mutated version of M1M or M2M at the most conserved positions. We found that mutating either M1M or M2M motifs led to significant downregulation of *GUS* expression and increased incorporation of H3K27me3 (Fig. 5c-d). These results demonstrate that the integrity of M1M and M2M sequences is required to promote expression and antagonize H3K27me3-associated repression.

Discussion

Taken together, the findings described here identify MAIL1/MAIN and MAIL2 proteins as gene regulatory factors that antagonize Polycomb silencing to secure gene expression. These proteins associate with different DNA sequence motifs (M1M and M2M) in proximal promoters, constituting specific protein-DNA motif modules that ensure expression of distinct genes, yet through a similar mechanism. While the precise molecular bases of this mechanism remain to be determined, MAIL1/MAIN and MAIL2 likely hinder downstream steps of PRC recruitment rather than preventing binding of PRC-recruiting factors at PREs. This is evidenced by our observation that teloboxes in the promoters of MAIL1/MAIN and MAIL2 targets are bound by TRB proteins, known PRC2-recruiters.

MAIL1 and *MAIN* likely evolved from a duplication of *MAIL2* that occurred specifically in brassicassae[13]. The sequence similarity between M1M and M2M suggests that the palindromic M2M DNA sequence motif possibly evolved into M1M at the same time. *MAIL2* appears to be conserved at least in all eudicot plants[13], opening the possibility that this gene regulatory system involving PMD protein-DNA motif modules may be conserved across a broad range of plant species. Only a fraction of M1M and M2M (about 20-40%) motifs are bound by MAIN/MAIL1 or MAIL2 in the genome (Supplementary Fig. 11f), suggesting that an additional layer of regulation may be involved in targeting these factors to prevent H3K27me3 deposition at numerous actively transcribed genes. Future studies will address this point.

We previously defined four clades of PMD proteins (A1, A2, B and C). PMD-As are encoded by retrotransposons, while PMD-B are found in protein-coding genes. MAIL1, MAIN

and MAIL2 belong to the PMD-C clade. PMD-C domain-containing proteins are synthesized from *bona fide* host genes and, in several monocots and eudicots plant species, can also be encoded by DNA transposons[12] (Supplementary Fig. 12; Supplementary doc. 1). The predicted structures of MAIL1, MAIN and MAIL2 PMD-C domains are highly similar (Supplementary Fig. 8) and, interestingly, share structural homology with the DUF1985 domain of the DNA transposon-encoded VANC proteins[31], which function as an anti-silencing factor antagonizing DNA methylation[32–34]. Recent studies have shown that H3K27me3 plays a role in TE repression in a broad range of eukaryotes, including Arabidopsis, suggesting an ancestral role of Polycomb silencing in controlling TE activity[35,36]. This raises the intriguing possibility that TE-encoded PMD-C proteins may act as Polycomb anti-silencing factors. The potential connection between transposon biology and Polycomb silencing regulation could provide valuable insights into the coevolution of genome defense mechanisms and gene expression control.

Methods

Plant Material. Mutants *mail1-1* (GK-840E05)[27], *main-2* (GK-728H05)[37], *mail2-1* (SAIL_1153_F07), *ndx-4* (WiscDsLox344A04)[38], *ring1a-2* (SAIL_393_E05)*/1b-3* (SAIL_520_B04)[18], *clf-29* (SALK_021003)[39], *swn-7* (SALK_109121)[6] and *val1-2* (SALK_088606C)[21] are all in the Col-0 background. The transgenic L5 line[16] was kindly provided by H. Vaucheret (IJPB, France).

Design of the amiRNA (amiR) construct directed against *MAIL2* transcript was done using the WMD3-Web MicroRNA Designer (http://wmd3.weigelworld.org) and the amiR-*MAIL2* fragment cloned into the BIN-HYG-TX plasmid under the control of p35S promoter.

The *pMAIL1:gMAIL1-FLAG*, *pMAIL1:gMAIL1-MYC*, *pMAIN:gMAIN-FLAG*, *pMAIN:gMAIN-MYC* and *SEP3pro:GUS* constructs were described previously[13,29]. For the *pMAIL2:gMAIL2-9xMYC* construct, two genomic fragments encompassing the *MAIL2* gene (-1834 to +2448 relative to ATG codon; A designated as +1) were PCR-amplified from Col-0 genomic DNA. PCR primers, with several point mutations to *MAIL2* sequence, allowed to reconstitute a *MAIL2* transgene bearing silent mutations and insensitive to amiR-*MAIL2* targeted degradation (Supplementary Fig. 7b); recombination-based technology (NEB, NEBuilder HiFi DNA Assembly Master Mix #2621) was then used to replace *MAIL1*-specific sequences of

pMAIL1:gMAIL1-MYC vector by the *MAIL2* transgene fragment. The *pNDX:gNDX* fragment was PCR amplified from Col-0 genomic DNA and region from -1631 to +5177 relative to ATG codon was cloned into the PMDC123 vector. For *KPI-1:GUS* and *HD2A:GUS* transgenes, proximal gene promoter and 5'UTR amplicons of 459bp and 357bp respectively were fused to *GUS* coding sequence and cloned into the PMDC123 vector. M1M and M2M were mutated using *pKPI-1:GUS* or *pHD2A:GUS* as matrix and sequence modifications are described on Fig. 5. Insertion of M1M (-321 to -306 relative to ATG) or M2M (-460 to -423 relative to ATG) in *proSEP3:GUS* was done by inverse PCR.

All transgenic constructs were introduced in Col-0 plants through agroinfiltration by floral dipping[40] and T1 transformants were isolated following *in vitro* culture (Murashige and Skoog medium, 1% sucrose) with adequate antibiotic selection. Introgression in diverse genetic backgrounds was obtained by crossing.

Plants were grown on soil (60% relative humidity) or *in vitro* under long-day conditions (16 h light, 8 h dark) at 23 °C.

Mutagenesis and mapping. Ethyl methanesulfonate (EMS) mutagenesis of the L5/*mail1-1* mutant line was done as previously described[12] . To isolate suppressors of the *mail1* mutation, a double screen was performed based on: 1/ visual amelioration of plant phenotype 2/ GUS staining to confirm L5 transgene resilencing ; GUS assay was done by dissecting one leaf per M2 plant candidate and leaves were incubated at 37 °C for 24 h in 400µg ml^{-1} 5-bromo-4-chloro-3-indolyl-b-D-glucuronic acid, 10 mM EDTA, 50 mM sodium phosphate buffer pH 7.2, 0.2% triton X-100. The *ndx-7* mutant allele was identified following mapping-by-sequencing using genomic DNA from a pool of 47 individual F2 mutant plants segregating from a cross with Ler

RNA isolation and analyses. Total RNA was extracted from aerial tissues of 12-day-old plants or immature flower buds using RNAzol RT (RN190, Molecular Research Center, Inc). For real time RT-qPCR, 0.5 µg of total RNA was DNase treated and reverse transcribed using the PrimeScript RT reagent kit with gDNA eraser (Perfect real time) (TaKaRa) in a final volume of 20 µl and one microlitre of cDNA was used for subsequent amplification using SYBR Premix Ex Taq II (Tli RnaseH Plus) (TaKaRa) and Real-Time LightCycler 480 system (Roche) in a final reaction volume of 15 µl. Data were normalized to a reference gene and

analyzed according to the $2^{-\Delta\Delta Ct}$ method. Means and standard errors of the mean were calculated from independent biological samples. Primers used are listed in Supplementary Table 3.

For RNA-seq, DNase and purification were performed using RNA Clean & Concentrator-5 kit (Zymo Research). Libraries were generated from about one µg total RNA and sequenced as PE100 or PE150 reads on a DNBSEQ platform at Beijing Genomics Institute (Hong Kong).

ChIP-seq. About two grams of aerial two-week-old rosettes were collected and incubated 15 min under vacuum in 25 ml of freshly prepared cross-linking buffer (10 mM phosphate buffer pH 7, 100 mM NaCl, 0.1 M sucrose, 1% formaldehyde). 2 M glycine was added to stop cross-link reaction (0.125 M final) and incubated an additional 5 min under vacuum. Tissues were then washed three times with 10 mM phosphate buffer pH 7, 100 mM NaCl, 0.1 M sucrose, pooled in 50 ml falcon tube and snap frozen in liquid nitrogen. Cross-linked tissues were ground into fine powder by adding 7 ceramic balls to falcon tube and vortexing, and resuspended into 25 ml of extraction Buffer 1 (0.4 M Sucrose, 10 mM Tris-HCl pH 8, 10 mM MgCl2, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1X complete EDTA-free cocktail inhibitor (Roche)) at 4°C. Alternatively, collected tissues were ground into fine powder with liquid nitrogen and samples were resuspended with 25 ml Hepes Buffer containing 1% formaldehyde (50 mM Hepes pH 8, 1 M sucrose, 5 mM KCl, 5 mM MgCl2, 0.6% triton X-100, 0.4 mM PMSF, 1% formaldehyde, 1X complete EDTA-free cocktail inhibitor), and incubated for 10 min at RT. Crosslinking was stopped with 2 M glycine (0.125 M final), inverting several times. The lysates were filtered through one layer of Miracloth and the nuclei were collected by centrifuge at 4°C with 4000 rpm for 20 min. The nuclei were resuspended with extraction buffer 2 (0.25 M Sucrose, 10 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 1% triton X-100, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1X complete EDTA-free cocktail inhibitor), centrifuge with 14000 g at 4°C for 10 min, and then resuspended with 400 µl extraction buffer 3 (1.7 M Sucrose, 10 mM Tris-HCl pH 8, 0.15% Triton X-100, 2 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1X complete EDTA-free cocktail inhibitor). The nuclei suspension was loaded over a cushion of 400 µl extraction buffer 3 and microtubes were centrifugated with 14000 g for 1 h at 4°C. For chromatin shearing using Bioruptor UCD-200 (Diagenode), nuclei were resuspended in 250 µl of lysis buffer (50 mM Tris-HCl pH 8, 10 mM

EDTA pH 8, 1% SDS, 1X complete EDTA-free cocktail inhibitor) and 25 cycles of 25 s 'ON', 1 min 15 s 'OFF were applied at 4°C. Following centrifugation, chromatin was split into two fractions of 80 µl and 9 vol. of ChIP dilution buffer added (1.1% Triton X-100, 16.7 mM Tris-HCl pH 8, 1.2 mM EDTA pH 8, 167 mM NaCl, 1X complete EDTA-free cocktail inhibitor) before proceeding to preclearing and IP as described hereafter. For Covaris fragmentation, nuclei were resuspended with 1 ml of covaris buffer (20 mM Tris pH 8, 2 mM EDTA, 0.1% SDS, 1X complete EDTA-free cocktail inhibitor) for 30 min at 4°C on a turning wheel. The lysate was transferred into Covaris milliTube Glass 1 ml AFA Fiber and chromatin was mainly sheared to 200-400 bp fragments using S220 Focused-ultrasonicator with the following conditions: for histone marks analysis, PIP175, CBP200, DF20, T 5°C, 5 min; for MAIL2-MYCtag, PIP140, CBP200, DF10, T 5°C, 6 min. Sheared chromatin was centrifuged with 21000 g for 10 min at 4°C and the supernatant was split into two fractions of 440µl. After adding 0.5 volume of covaris dilution buffer (20 mM Tris-HCL pH 8, 2 mM EDTA, 0.1% SDS, 450 mM NaCl, 3% Triton X-100, 1X complete EDTA-free cocktail inhibitor), chromatin was precleared with 30 µl of magnetic Protein A and Protein G Dynabeads (Invitrogen) for 1 h at 4°C on a turning wheel. Twenty µl was saved as input and the remaining supernatant was then incubated at 4°C overnight with specific antibodies (5 µg of H3K27me3 Ab, Diagenode C15410069 or Millipore 07-449; 10 µl H2AUb Ab, Cell Signaling Technology ref.8240; 10 µl Myc-Tag Ab, Cell Signaling Technology ref.2276 or ref.2278; 3.5 µl monoclonal ANTI-FLAG M2 Ab, Sigma F1804-1MG). Next, 40 µl of magnetic Protein A and Protein G Dynabeads were added and the mixture was rotated at 4°C for 3-4h. The beads were washed for 5 min rotating at 4°C with low salt solution twice (0.1% SDS; 1% Triton X-100, 2 mM EDTA pH 8, 20 mM Tris-HCl pH 8, 150 mM NaCl), high salt solution (0.1% SDS; 1% Triton X-100, 2 mM EDTA pH 8, 20 mM Tris-HCl pH8, 500 mM NaCl), LiCl solution (0.25 M LiCl, 1% IGEPAL CA630, 1% deoxycholic acid, 1 mM EDTA, 10 mM Tris-HCl pH8) and TE solution (10 mM Tris-HCl pH 8, 1 mM EDTA pH 8). A second TE wash was performed after transferring beads to a new DNA low-bind Eppendorf microtube. The chromatin was eluted with 250 µl elution buffer twice (1% SDS, 0.1 M NaHCO3) at 65°C for 15 min on a thermomixer with shaking (950 rpm) and subjected to reverse crosslinking by adding 51 µl of reverse crosslink mix (2 M NaCl, 0.1 M EDTA pH 8, 0.4 M Tris-HCl pH 8, 0.4 mg/ml proteinaseK) for 3 h at 50°C then overnight at 65°C. Following digestion with 10 µg RNaseA for 30 min at RT°, DNA was purified using ChIP

DNA clean and concentrator kit (Zymo ZD5205). For MAIL2-MYC, five independent IP experiments (equivalent 10 g tissues) were pooled.

For immature flower buds, three tubes of about 0.15 g material were harvested and snap-frozen in liquid nitrogen. Following tissue disruption using Qiagen TissueLyser, samples were resuspended in 1.8 ml Hepes Buffer containing 1% formaldehyde, pooled and further processed as described above with covaris fragmentation at 5°C, PIP140, CBP200, DF10 for 6 min. IP was then performed using 4 µg of antibody to H3K27me3 (Diagenode, C15410069). Libraries were generated and sequenced as PE100 reads on a DNBSEQ platform at Beijing Genomics Institute (Hong Kong).

ChIP-qPCR. Three tubes of about 0.15 g of aerial 12-day-old rosettes were snap frozen in liquid nitrogen and processed as described above for floral buds, except that nuclei were resuspended in 130 µl covaris buffer and sonicated using microtube-130 AFA Fiber. Each IP was performed using 2.1 µg of H3K27me3 antibody. (Diagenode, C15410069) and one of 15 µl purified DNA was used for real-time PCR amplification using SYBR Premix Ex Taq II (Tli RnaseH Plus) (TaKaRa) and Real-Time LightCycler 480 system (Roche) in a final reaction volume of 15 µl. Primers used are listed in Supplementary Table 3.

RNA-seq analysis. Reads were trimmed and filtered for quality and adapter contamination using Trim Galore (v.0.6.7, Babraham Institute) and aligned to the *Arabidopsis* reference genome (TAIR10) genome using STAR (v.2.7.9a)[39]. Reads aligning equally well to more than one position in the reference genome were discarded, and probable PCR duplicates were removed using MarkDuplicates from the Picard Tools suite[42]. Sample tracks were generated using deeptools (v.3.5.1)[43] bamCoverage with the option normalizeUsing CPM and binSize 1. Subsequent read counting for each gene and TE was performed using featureCounts (v.2.0.3)[44] using the Araport11 genome annotation. Annotated TEs overlapping strongly (> 80%) with an annotated TE gene were considered TE genes, and the TE annotation was discarded. Differential expression analysis was performed using DESeq2 (v.1.36.0)[45] with absolute fold-change ≥ 4 and Benjamini-Hochberg adjusted *p*-value < 0.05 were considered differentially expressed. For DEseq2 analysis of amiR-*mail2* hypomorphic lines, criteria of absolute fold change ≥2 and Benjamini-Hochberg adjusted *p*-value < 0.05 were retained. All plots were generated using ggplot2 (v.3.4.4).

ChIP-seq analysis. Reads were trimmed and filtered using Trim Galore (v.0.6.7, Babraham Institute) or Trimmomatic (v.0.39)[46], and aligned to the TAIR10 *Arabidopsis* reference genome using bowtie2 (v.2.4.2)[47], allowing only uniquely mapped reads with perfect matches. PCR duplicates were discarded using MarkDuplicates from the Picard Tools suite or markdup from Sambamba (v.0.8.2)[48]. Sample tracks were generated using deeptools (v.3.5.1)[43] bamCoverage with the option normalizeUsing RPGC and binSize 1. Peaks calling was performed with MACS2 (v.2.2.7.1)[49] using corresponding inputs as a background and --nomodel and –extsize 200 options or MACS3 (v.3.0.0b1)[49] using --model and –call-summits, q-value < 0.01, and peaks called in the anti-MYC or anti-FLAG WT controls were removed. MAIL1/MAIN bound-regions are designed as the summit of peaks plus 100bp on each side.

Motif discovery. DNA sequences of MAIL1/MAIN and MAIL2 ChIP-seq peaks and of gene proximal promoter regions (+/- 500 pb surrounding TSS) were extracted using bedtools (v.2.31.1) getfasta and analyzed by STREME v5.5.5[50] with the following parameters: streme --verbosity 1 --oc streme_out -dna --minw 6 --maxw 16 --order 2 --bfile ./background --seed 0 --align center --time 4740 --totallength 4000000 --evalue --thresh 0.05, or by MEME v5.5.5[51] with the following parameters: meme -oc meme_out -objfun classic -mod anr -csites 10000 -nmotifs 10 -minw 6 -maxw 15 -markov_order 0 -dna -revcomp -evt 0.05. Genomic distribution of M1M and M2M identified from ChIP peaks was determined using FIMO v5.5.5[52] with *P* value match < 1e-05 and default parameters. Genomic localization of telobox elements was determined using FIMO v5.5.5[52] with p-value match < 1e-05 and default parameters and the consensus sequence "AAACCCTAR".

Gene ontology analysis. GO term enrichment analysis was performed using ShinyGO 0.80[53]. P-value was derived from hypergeometric distribution, adjusted using the FDR (false discovery rate) method. Results with a minimum *P*-value of < 5e-02 were reported; FDR < 1e-05 is considered as really significant.

Phylogenetic analysis. Candidate MuDR-encoded PMD sequences were identified through tBlastN analysis of Repbase database[54] using the MAIL2 PMD-C region as query (Supplementary doc. 1). The "PhyML+SMS/OneClick" pipeline[55] (https://ngphylogeny.fr/)

with default parameters was then used to build a phylogenetic tree based on MuDR-associated PMD and including Arabidopsis gene-encoded PMD-B and -C sequences. Resulting tree was exported using the online Interactive Tree Of Life (ITOL) tool. PMD sequences used to build the tree are reported in Supplementary Table 4.

Data availability. The data supporting the findings of this study are available within the article and its Supplementary Information. High throughput sequencing data has been deposited in the Gene Expression Omnibus (GEO) database and can be accessed with the accession number GSEXXXXXX. All data are available from the corresponding author upon reasonable request.

References

1. Zhou, Y., Romero-Campero, F. J., Gómez-Zambrano, Á., Turck, F. & Calonje, M. H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. *Genome Biol.* **18**, 13–69 (2017).
2. Chen, L.-Q. *et al.* ATX3, ATX4, and ATX5 Encode Putative H3K4 Methyltransferases and Are Critical for Plant Development. *Plant Physiol.* **174**, 1795–1806 (2017).
3. Alvarez-Venegas, R. *et al.* ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. *Curr. Biol. CB* **13**, 627–637 (2003).
4. Tamada, Y., Yun, J.-Y., Woo, S. chul & Amasino, R. M. ARABIDOPSIS TRITHORAX-RELATED7 Is Required for Methylation of Lysine 4 of Histone H3 and for Transcriptional Activation of FLOWERING LOCUS C. *Plant Cell* **21**, 3257–3269 (2009).
5. Berr, A. *et al.* Arabidopsis SET DOMAIN GROUP2 Is Required for H3K4 Trimethylation and Is Crucial for Both Sporophyte and Gametophyte Development. *Plant Cell* **22**, 3232–3248 (2010).
6. Lu, F., Cui, X., Zhang, S., Jenuwein, T. & Cao, X. Arabidopsis REF6 is a histone H3 lysine 27 demethylase. *Nat. Publ. Group* **43**, 715–719 (2011).
7. Cui, X. *et al.* REF6 recognizes a specific DNA sequence to demethylate H3K27me3 and regulate organ boundary formation in Arabidopsis. *Nat. Publ. Group* (2016).
8. Yan, W. *et al.* Dynamic and spatial restriction of Polycomb activity by plant histone demethylases. *Nat. Plants* **4**, 681–689 (2018).
9. Antunez-Sanchez, J. *et al.* A new role for histone demethylases in the maintenance of plant genome integrity. *eLife* **9**, (2020).
10. Liang, S. C. *et al.* Kicking against the PRCs – A Domesticated Transposase Antagonises Silencing Mediated by Polycomb Group Proteins and Is an Accessory Component of Polycomb Repressive Complex 2. *PLoS Genet.* **11**, e1005660-26 (2015).
11. Velanis, C. N. *et al.* The domesticated transposase ALP2 mediates formation of a novel Polycomb protein complex by direct interaction with MSI1, a core subunit of Polycomb Repressive Complex 2 (PRC2). *PLoS Genet.* **16**, e1008681 (2020).
12. Ikeda, Y. *et al.* Arabidopsis proteins with a transposon-related domain act in gene silencing. *Nat. Commun.* **8**, 1–10 (2017).

13. Nicolau, M. *et al.* The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in Arabidopsis thaliana. *PLoS Genet.* **16**, e1008324 (2020).
14. Luxán Hernández, C. *et al.* PP7L is essential for MAIL1-mediated transposable element silencing and primary root growth. *Plant J.* **8**, 15115–15122 (2020).
15. Jarry, L. *et al.* Plant mobile domain proteins ensure Microrchidia 1 expression to fulfill transposon silencing. *Life Sci. Alliance* **6**, e202201539 (2023).
16. Morel, J. B., Mourrain, P., Béclin, C. & Vaucheret, H. DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. *Curr. Biol. CB* **10**, 1591–1594 (2000).
17. Zhu, Y. *et al.* The Arabidopsis noduliN homeobox factor AtNDX interacts with AtRiNG1a/b and negatively regulates abscisic acid signaling[Open]. *Plant Cell* **32**, 703–721 (2020).
18. Li, J., Wang, Z., Hu, Y., Cao, Y. & Ma, L. Polycomb Group Proteins RING1A and RING1B Regulate the Vegetative Phase Transition in Arabidopsis. *Front. Plant Sci.* **8**, 867 (2017).
19. Mikulski, P. *et al.* VAL1 acts as an assembly platform co-ordinating co-transcriptional repression and chromatin regulation at Arabidopsis FLC. *Nat. Commun.* **13**, 5542 (2022).
20. Baile, F., Merini, W., Hidalgo, I. & Calonje, M. EAR domain-containing transcription factors trigger PRC2-mediated chromatin marking in Arabidopsis. *Plant Cell* **33**, 2701–2715 (2021).
21. Yuan, L. *et al.* The transcriptional repressors VAL1 and VAL2 recruit PRC2 for genome-wide Polycomb silencing in Arabidopsis. *Nucleic Acids Res.* **49**, 98–113 (2021).
22. Shu, J. *et al.* Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in Arabidopsis seedlings. *Plant Direct* **3**, e00100 (2019).
23. Moissiard, G. *et al.* MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* **336**, 1448–1451 (2012).
24. Moissiard, G. *et al.* Transcriptional gene silencing by Arabidopsis microrchidia homologues involves the formation of heteromers. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7474–7479 (2014).
25. Stacey, N. J. *et al.* Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *Plant J.* **48**, 206–216 (2006).
26. Xue, M. *et al.* The Number of Meiotic Double-Strand Breaks Influences Crossover Distribution in Arabidopsis. *Plant Cell* **30**, 2628–2638 (2018).
27. Ühlken, C., Horvath, B., Stadler, R., Sauer, N. & Weingartner, M. MAIN-LIKE1 is a crucial factor for correct cell division and differentiation in Arabidopsis thaliana. *Plant J.* **78**, 107–120 (2014).
28. Xiao, J. *et al.* Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nat. Publ. Group* **49**, 1546–1552 (2017).
29. Zhou, Y. *et al.* Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. *Nat. Publ. Group* **50**, 638 (2018).
30. Wang, M. *et al.* Arabidopsis TRB proteins function in H3K4me3 demethylation by recruiting JMJ14. *Nat. Commun.* **14**, 1736 (2023).
31. Borredá, C., Leduque, B., Colot, V. & Quadrana, L. Transposable element products, functions, and regulatory networks in Arabidopsis. 2024.04.02.587720 Preprint at https://doi.org/10.1101/2024.04.02.587720 (2024).
32. Hosaka, A. *et al.* Evolution of sequence-specific anti-silencing systems in Arabidopsis. *Nat. Commun.* **8**, 2161 (2017).

33. Sasaki, T. *et al.* Fast co-evolution of anti-silencing systems shapes the invasiveness of Mu-like DNA transposons in eudicots. *EMBO J.* **41**, e110070 (2022).
34. Sasaki, T. *et al.* Arms race between anti-silencing and RdDM in noncoding regions of transposable elements. *EMBO Rep.* **n/a**, e56678 (2023).
35. Hisanaga, T. *et al.* The Polycomb repressive complex 2 deposits H3K27me3 and represses transposable elements in a broad range of eukaryotes. *Curr. Biol.* **33**, 4367-4380.e9 (2023).
36. Hure, V. *et al.* Alternative silencing states of Transposable Elements in Arabidopsis. 2024.03.16.585326 Preprint at https://doi.org/10.1101/2024.03.16.585326 (2024).
37. Wenig, U. *et al.* Identification of MAIN, a factor involved in genome stability in the meristems of Arabidopsis thaliana. *Plant J.* **75**, 469–483 (2013).
38. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J. & Dean, C. R-Loop Stabilization Represses Antisense Transcription at the Arabidopsis FLC Locus. *Science* **340**, 619–621 (2013).
39. Xu, L. & Shen, W.-H. Polycomb Silencing of KNOX Genes Confines Shoot Stem Cell Niches in Arabidopsis. *Curr. Biol.* **18**, 1966–1971 (2008).
40. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for -mediated transformation of. *Plant J.* **16**, 735–743 (1998).
41. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
42. Broad Institute. Picard Toolkit. Broad Institute, GitHub repository http:// broadinstitute.github.io/picard/. (2019).
43. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
44. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
45. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
47. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
48. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
49. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, 1–9 (2008).
50. Bailey, T. L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834–2840 (2021).
51. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
52. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinforma. Oxf. Engl.* **27**, 1017–1018 (2011).
53. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinforma. Oxf. Engl.* **36**, 2628–2629 (2020).
54. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
55. Lemoine, F. *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-

specialists. *Nucleic Acids Res.* **47**, W260–W265 (2019).
56. Varadi, M. *et al.* AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375 (2024).

Acknowledgments

We thank F. Turck (Max Planck Institute for Plant Breeding Research) for kindly providing the *SEP3pro:GUS* construct. Work in the Mathieu laboratory was supported by CNRS, Inserm, Université Clermont Auvergne core funding, a grant from the iSITE CAP2025 (to M.O.) and a grant from the Agence Nationale de la Recherche (ANR-23-CE20-0012 to O.M.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

T.P., L.J., G.M. and O.M. conceived the study. T.P., L.J., M.O., G.D., M.-N.P.-P., C.C., J.D., N.P., G.M. and O.M. conducted laboratory experiments. T.P., L.J., G.M. and O.M interpreted the data. T.P. and O.M. drafted the manuscript. T.P., L.J., G.M. and O.M. edited the manuscript. G.M and O.M. coordinated the research. The authors read and approved the final manuscript.

Additional information

Competing interests: The authors declare that they have no conflict of interest.

Figure legends

Fig. 1. NDX and PRC1 RING1A/1B components suppress *mail1*-induced defects. (**a**) Histochemical staining for GUS activity in leaves from plants of the indicated genotypes. Scale bar is 500 µm. (**b**) Developmental phenotype of three-week-old WT, *mail1*, *mail1 ndx-7*, *mail1 ring1a/1b*, *ndx-7* and *ring1a/1b* seedlings. Scale bar is 1 cm. (**c**) Expression analysis of *mail1*-overexpressed loci by quantitative RT–PCR in the indicated genotypes. Transcript levels are normalized to *ACT2* and further normalized to *mail1*; values represent means from three biological replicates ± s.e.m. Asterisks mark statistically significant differences from the *mail1* mutant (unpaired, two-sided Student's t-test, * $P < 0.05$; ** $P < 0.01$).

Fig. 2. *mail1*-associated TE and gene misregulation is dependent on NDX or RING1A/1B function. (**a**) Heatmaps of RNA-seq expression z-score computed for *mail1* upregulated TEs and misregulated genes across the indicated genotypes (left). Means of two biological replicates were computed. Boxplots with the same data expressed in RPKM are shown on the right. Asterisks mark statistically significant differences from the *mail1* mutant (Wilcoxon rank sum test with continuity correction; two-sided, ** $P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$). (**b**) Genome browser tracks showing mRNA profiles (CPM) in the indicated genotypes at *mail1*-misregulated loci. (**c**) Metagene plots of H3K27me3 accumulation (RPGC) in the indicated genotypes at *mail1*-upregulated, *mail1*-downregulated genes compared to all Arabidopsis genes. One of the two biological replicates is shown (**d**) H3K27me3 level variations in *mail1* compared to WT at *mail1*-downregulated genes. (**e**) Representative genome browser views of RNA-seq (CPM) and H3K27me3 ChIP-seq (RPGC) profiles in the indicated genotypes at H3K27me3-enriched genes in *mail1*. MAIL1-FLAG and MAIN-FLAG associated patterns are shown together with TRB1-FLAG data (GEO-GSM6895939 retrieved from[30]). Predicted Mail1 (M1M) and telobox motifs are indicated. (**f**) STREME software prediction of most significantly enriched DNA motifs within the 1kb region surrounding TSS of H3K27me3-enriched genes in *mail1*. Percentage of regions exhibiting each motif is indicated.

Fig. 3. MAIL1 and MAIN are mainly co-targeted to M1M at gene proximal promoters. (**a**) Heatmaps showing MAIN and MAIL1 ChIPseq signals (Log2 FC / input) over MAIL1 (n = 660) and MAIN (n = 1064) peaks respectively. Peaks were aligned at their center position. (**b**) Overlap between MAIL1 and MAIN peaks. (**c**) Diagram illustrating the genomic distribution of MAIL1/MAIN common peaks (n = 626). (**d**) Metagene plot showing MAIL1 and MAIN enrichment at MAIL1/MAIN co-bound genes (n = 670). Co-bound genes displayed MAIL1/MAIN peak overlaping over the 200 bp region upstream TSS (proximal promoter targeted genes) or encompassed at least 50% of the peak within the TSS - TES region (body targeted genes). (**e**) Metaplots showing MAIL1, MAIN and M1M enrichment at promoter region of *mail1*-H3K27me3 enriched genes (left panel). An equivalent random set of genes is shown as control (right panel). (**f**) transcript and H3K27me3 levels, expressed in RPKM, at *mail1*-H3K27me3 enriched genes in the indicated genotypes. The effect of genotype was verified with a Wilcoxon rank sum test with continuity correction; two-sided (ns. not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$).

Fig. 4. MAIL2 mainly associates to gene promoters and protects numerous actively transcribed genes from H3K27me3-mediated Pcg silencing (**a**) Photos of three-week-old seedlings of indicated genotype. Scale bar is 1 cm. (**b**) Diagram of MAIL2 peaks genomic distribution. (**c**) Arabidopsis genes were split in ten deciles based on their WT expression level (RPKM; bottom diagram). Upper diagram showing distribution of gene with MAIL2 located at proximal promoter (i.e. MAIL2 peak overlap over the 200 bp region upstream TSS) and of an equivalent random set of genes. Asterisks mark statistically significant differences from an equal representation in each decile (Fisher's exact test, *P*-value < 1e-04). (**d**) Top 3 of significantly enriched DNA motifs, predicted by STREME, within the 1kb region surrounding TSS of genes harboring MAIL2 peak at proximal promoter (n = 1059). Percentage of promoter regions exhibiting each motif is shown. (**e**) Overlap between MAIL2-associated promoter genes and misregulated genes in amiR-*mail2* lines. Fold change enrichment compared to random distribution and p-value from hypergeometric tests are indicated. (**f**) Representative genome browser views of RNA-seq (CPM) and H3K27me3 ChIP-seq (RPGC) profiles in the indicated genotypes at MAIL2-associated promoter genes downregulated in amiR-*mail2* background. MAIL2-MYC associated patterns are shown together with TRB1-FLAG data (GEO-GSM6895939 retrieved from[30]) and predicted MAIL2

(M2M) and telobox motif positions. (**g**) Box plots showing transcript and H3K27me3 levels (RPKM) in WT and amiR-*mail2* lines at MAIL2-associated promoter genes downregulated (n = 39) or not (n = 1020) in hypomorphic amiR-*mail2* background. The effect of genotype was verified with a Wilcoxon rank sum test with continuity correction; two-sided (ns. not significant; *** $P < 0.001$; **** $P < 0.0001$).

Fig. 5. Promoter-associated M1M and M2M are functionally relevant to safeguard genes from PcG-mediated silencing. (a) Schematic representation of the *proSEP3:GUS* original transgene and derivatives following insertion of one M1 or two M2 motifs. Box plots showing relative GUS expression and H3K27me3 accumulation associated to different transgene combinations. For each transgene, data represent two technical replicates on 4 independent transformed lines. Quantitative RT-PCR analysis of *GUS* expression was normalized to *ACT2*. H3K27me3 ChIP-qPCR is reported as % input. Asterisks mark statistically significant differences from *pSEP3:GUS* transgene (unpaired, two-sided Student's t-test, * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$). (**b**) Genome browser views of RNA-seq (CPM) and H3K27me3 ChIP-seq (RPGC) profiles at pKPl-1 and pHD2A loci in the indicated genotypes. MAIL1 and MAIL2-associated patterns are shown together with TRB1-FLAG data (GEO-GSM6895939 retrieved from[30]) and predicted M1M, M2M and telobox motifs. (**c**) Schematic representation of *pPKL1:GUS* transgenes. Most conserved M1M motifs were mutated (m1m) and sequence modifications are shown under the scheme. Telobox motif is also reported in blue. Relative GUS expression and H3K27me3 accumulation were performed as described in (a) except that nine independent lines were analysed for each transgene. (**d**) Schematic representation of *pHD2A:GUS* transgenes showing localization of intact (M2M) and mutated (m2m) motifs. Telobox motifs are indicated in blue and sequence modifications are reported under the scheme. Molecular analyses were performed as described in (a) except that nine independent lines were analysed for each transgene.

Supplementary Fig. 1. *mail1*- and *main*-associated phenotypes are suppressed in *ndx* but not in *val1* mutant backgrounds. (a) Representation of *NDX* describing the insertion (*ndx-4*) and point mutation (*ndx-7*; this study) alleles. C to T mutation for *ndx-7* created an early STOP codon. Position is indicated relative to the transcription start site (+1). (**b**)

Developmental phenotype of 3-week-old WT, *mail1*, *main*, *ndx-7*, *mail1 ndx-7*, *mail1 ndx-4* and *main ndx-7* seedlings. Scale bar is 1 cm. (**c**) Photos of two-week-old seedlings of indicated genotype. Scale bar is 1 cm. (**d**) Histochemical staining for GUS activity in leaves from plants of the indicated genotypes. Scale bar is 500 µm. (**e**) Developmental phenotypes observed for *mail1 val1* and *main val1* seedlings showing abnormal differentiated structures emerging from apical meristem.

Supplementary Fig.2. *clf* and *swn* mutations poorly suppress *mail1*-associated molecular and developmental defects. (a) Fraction and number of TEs and genes significantly up- or down-regulated in *mail1* seedlings versus WT (Log2 FC ≥ 2; Padj < 0.05). (**b**) Venn diagrams showing overlaps of genes downregulated in *mail1*, *mail1 ndx* and *mail1 ring1a/1b* mutant backgrounds. (**c**) Transcript accumulation (in RPKM) at *mail1*-upregulated (left panel) and *mail1*-downregulated (right panel) genes in the indicated genotypes. Means of two biological replicates were computed. Statistical significances of *mail1 clf* or *mail1 swn* values compared to *mail1* are indicated (Wilcoxon rank sum test with continuity correction; two-sided; n.s. non-significant; ** $P < 0.01$). (**d**) Photos of three-week-old seedlings of indicated genotype. Scale bar is 1 cm.

Supplementary Fig.3. *MORC2* upregulation in *mail1 ndx* plants participates to TE resilencing. (**a**) Genome browser views of RNA-seq (CPM) profiles at *MORC1* and *MORC2* loci in the indicated genotypes. Two biological replicates are shown. Barplots below show *MORC1* and *MORC2* mean transcript level in the indicated genotypes relative to *mail1* (set to 1). (**b**) Heatmaps of RNA-seq expression z-score computed for *mail1*-upregulated (-up) TEs across the indicated genotypes. Means of two biological replicates were computed. TEs showing significant release of expression in *mail1 ndx morc2* compared to *mail1 ndx* (Log2 FC > 2) are indicated in bold. (**c**) RNA-seq genome browser views at two *mail1*-up TEs in the indicated genotypes. (**d**) Venn diagram analyses showing the overlap between *mail1*-up TEs, *mail1*-up TEs suppressed in *mail1 ndx* due to increased MORC2 expression and *mail1*-up TEs suppressed upon ectopic *UBQ1::MORC1* expression[15].

Supplementary Fig. 4. MAIL1 complexes prevent H3K27me3 incorporation at different loci (**a**) Genome browser views of RNA-seq (CPM) and H3K27me3 ChIP-seq (RPGC) profiles at

mail1 H3K27me3-enriched genes in the indicated genotypes. MAIL1-FLAG and MAIN-FLAG associated patterns are shown together with TRB1-FLAG data (GEO-GSM6895939 retrieved from[30]). Predicted Mail1 (M1M) and telobox motifs are indicated. (**b**) Genome browser view of H3K27me3 levels at *SPO11-1* gene in WT and *mail1* immature flowers (top). Transcript levels measured by RT-qPCR (bottom) are normalized to *ACT2* and further normalized to WT; values represent means from four biological replicates ± s.e.m and statistically significant difference is indicated (unpaired, two-sided Student's t-test, * $P < 0.05$)

Supplementary Fig. 5. Ectopic H3K27me3 deposition is detected at TRBs-enriched genes and correlates with increased levels of H2AUb incorporation. (**a**) Metagene plots of H3K27me3 and H2Aub accumulation (in RPGC) in the indicated genotypes at *mail1* H3K27me3 -enriched genes compared to an equivalent random set of genes or to all Arabidopsis genes. (**b**) Metagene plots showing TRB1/2/3 levels (data from[30]) at *mail1* K3K27me3-enriched genes or at an equivalent random set of genes. (**c**) Plot heatmaps illustrating TRB1/2/3 enrichment (data from[30]) at TSS of *mail1* H3K27me3-enriched genes.

Supplementary Fig. 6. MAIL1 and MAIN mainly co-localize at M1M-enriched gene TSS. (**a**) MAIL1 and MAIN signals (LOG2 FC vs input) at *mail1* H3K27me3-enriched or depleted genes. (**b**) Sequence logo of a highly enriched MEME-predicted motif among the 626 MAIL1/MAIN peaks.

Supplementary Fig. 7. *mail2-1* phenotypic defects are partially suppressed by *ring1a/1b* mutations. (**a**) Developmental phenotypes associated to two-week-old seedlings of indicated genotype. (**b**) Schematic representation of *MAIL2* gene showing t-DNA *mail2-1* localization and amiR-*MAIL2* targeted sequence. Synonymous substitutions creating amiR-*MAIL2* insensitive *gMAIL2-9xMYC* transgene are indicated in red and corresponding amino acids are shown. Positions are given relative to the transcription start site (+1). (**c**) Two-week-old phenotype of WT, *mail2-1* and *gMAIL2-1-MYC* complemented *mail2-1* seedlings. Scale bar is 1 cm.

Supplementary Fig. 8. MAIL1, MAIN and MAIL2 PMD-C domains displays highly similar 3D structures. Structure predictions of MAIL1, MAIN and MAIL2 were recovered from the

Alphafold Protein Structure Database[56] and, individual and merged 3D representation of PMD-specific parts are shown. UniProt IDs and position of PMD-C domains within the proteins are indicated below the structures.

Supplementary Fig. 9. MAIL2 targets a distinct set of genes through recognition of a specific DNA sequence motif. (**a**) Overlap between MAIL1/MAIN and MAIL2-associated peaks. (**b**) Metagene plot showing mean MAIL2-MYC vs input signal at Arabidopsis gene and TE loci. (**c**) Most significantly enriched motifs at MAIL2 peaks (n = 1775) predicted by STREME analysis. Relative representation of each motif is reported. (**d**) Centrimo plot showing strong enrichment of M2M over MAIL2 peaks, within a 45 bp central region delimited by the dashed lines (e-value = 8.7e-143). The 200bp region surrounding the peak summits is shown. (**e**) M1M homology to half of the palindromic M2M consensus sequence is highlighted by the dashed box.

Supplementary Fig. 10. Phenotype associated to amiR-*mail2* lines. (**a**) Examples of developmental phenotype recovered from 10-day-old *in vitro* growing plants. (**b**) *MAIL2* transcript accumulation in WT and amiR-*mail2_4* and *_5* lines, detected using primers framing the amiRNA-targeted cutting site. (**c**) Photos of five-week-old WT plants compared to amiR-*mail2_4* mutant plants containing a single transgene locus at heterozygous (he) or homozygous (ho) state. (**d**) Two-week-old phenotypes of WT, amiR-*mail2* and *gMAIL2-1-9xMYC*-complemented amiR-*mail2* seedlings. Scale bar is 1 cm.

Supplementary Fig. 11. MAIL1/MAIN and MAIL2 pathways target distinct sets of genes. (**a**) Venn diagrams showing no overlap between *mail1*- and amiR-*mail2*- downregulated genes. (**b**) Developmental phenotypes observed upon MAIL2 depletion in *mail1* mutant background. Scheme of the cross is illustrated with photos of two-week-old plants. Scale bar is 1cm. Zoomed-in pictures of *mail1* F2 progeny expressing or not the amiR-*MAIL2* transgene are shown. (**c**) Results of GO enrichment term analysis for biological processes using the sets of *mail1*- and amiR-*mail2*- downregulated genes. P-value (FDR) range from low (< 5e-02) to high (< 1e-05) statistical significance. (**d**) Metagene plots (top) and heatmaps (bottom) illustrating TRBs (data from[30]) and MAIL2 enrichment at MAIL2-promoter associated genes downregulated in amiRNA-*mail2* lines. For TRBs, only TRB1 data

were represented on the heatmap. (**e**) Venn Diagram showing overlap between MAIL2 and TRB1 peaks at proximal promoter region of amiRNA-*mail2* downregulated genes. (**f**) Venn diagrams showing overlap of M1M and M2M loci with MAIL1 and MAIL2 peaks respectively.

Supplementary Fig. 12. PMD-C encoded by MuDR transposons are widely distributed in angiosperms. A simplified phylogenetic tree was contructed using *Arabidopsis* gene-encoded PMD-B and PMD-C, and selected MuDR-encoded domains identified through tBlastN analysis of Repbase database[54] with MAIL2-specific PMD-C sequence as query (Supplementary doc. 1). Basal angiosperm *Amborella trichopoda*, monocot and dicot species are depicted in blue, orange and green, respectively. MuDR-encoded PMD-C were identified in *Amborella* and monocot species as previously reported[12], but were also found in dicotyledon species distributed across several Eudicots orders. Bootstrap value is indicated only for the node separating PMD-B and -C clades. Scale bar indicates one substitution/site. PMD sequences used to build the tree are reported in Supplementary Table 4. Abbreviations are: ATr *Amborella trichopoda*, Cba *Capsicum baccatum*, Gar *Gossypium arboretum*, GM *Glycine max*, Ipa *Ilex paraguariensis*, MT *Medicago truncatula*, OS *Oryza Sativa*, SeTo *Senna tora*, STu *Solanum tuberosum*, Tae *Triticum aestivum*, VV *Vitis Vinifera*, ZM *Zea mays*.

Supplementary Table 1. List of H3K27me3-enriched genes in *mail1* compared to WT background.

Supplementary Table 2. Lists of differentially expressed loci in amiR-*mail2*_4 and _5 mutant lines.

Supplementary Table 3. List of qPCR primers used in this study.

Supplementary Table 4. List of PMD sequences used for phylogenetic analyses.

Supplementary doc. 1. Best results of TBLASTN search on Repbase[54] using MAIL2 PMD-C domain as query. Only results with e-values < 1e-10 were reported. Dicot species : ArHy *Arachis hypogaea*, AvMa *Avicennia marina*, Cba *Capsicum baccatum*, Daca *Daucus carota*, Gar *Gossypium arboretum*, GR *Gossypium raimondii*, GM *Glycine max*, Ipa *Ilex*

paraguariensis, LU *Linum usitatissimum*, MT *Medicago truncatula*, NS *Nicotiana sylvestris*, SeTo *Senna tora*, STu *Solanum tuberosum*, VV *Vitis Vinifera*. Monocot species : DRot *Dioscorea cayennensis subsp. Rotundata*, OS *oryza Sativa*, SBi *Sorghum bicolor*, SHS *Saccharum hybrid cultivar*, Sit *Setaria italica*, Tae *Triticum aestivum*, ZM *Zea mays*. Basal angiosperm species: ATr *Amborella trichopoda*.

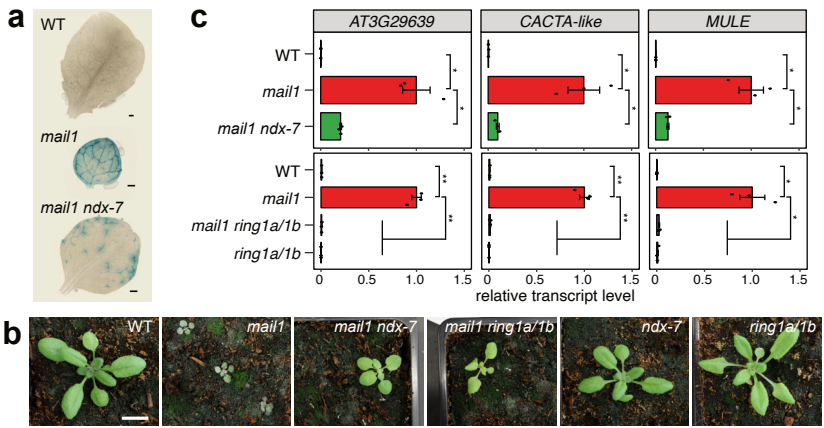
a
WT
mail1
mail1 ndx-7
c
AT3G29639
CACTA-like
MULE
WT
mail1
mail1 ndx-7
WT
mail1
mail1 ring1a/1b
ring1a/1b
0
0.5
1.0
1.5
relative transcript level
b
WT
mail1
mail1 ndx-7
mail1 ring1a/1b
ndx-7
ring1a/1b

Fig. 1

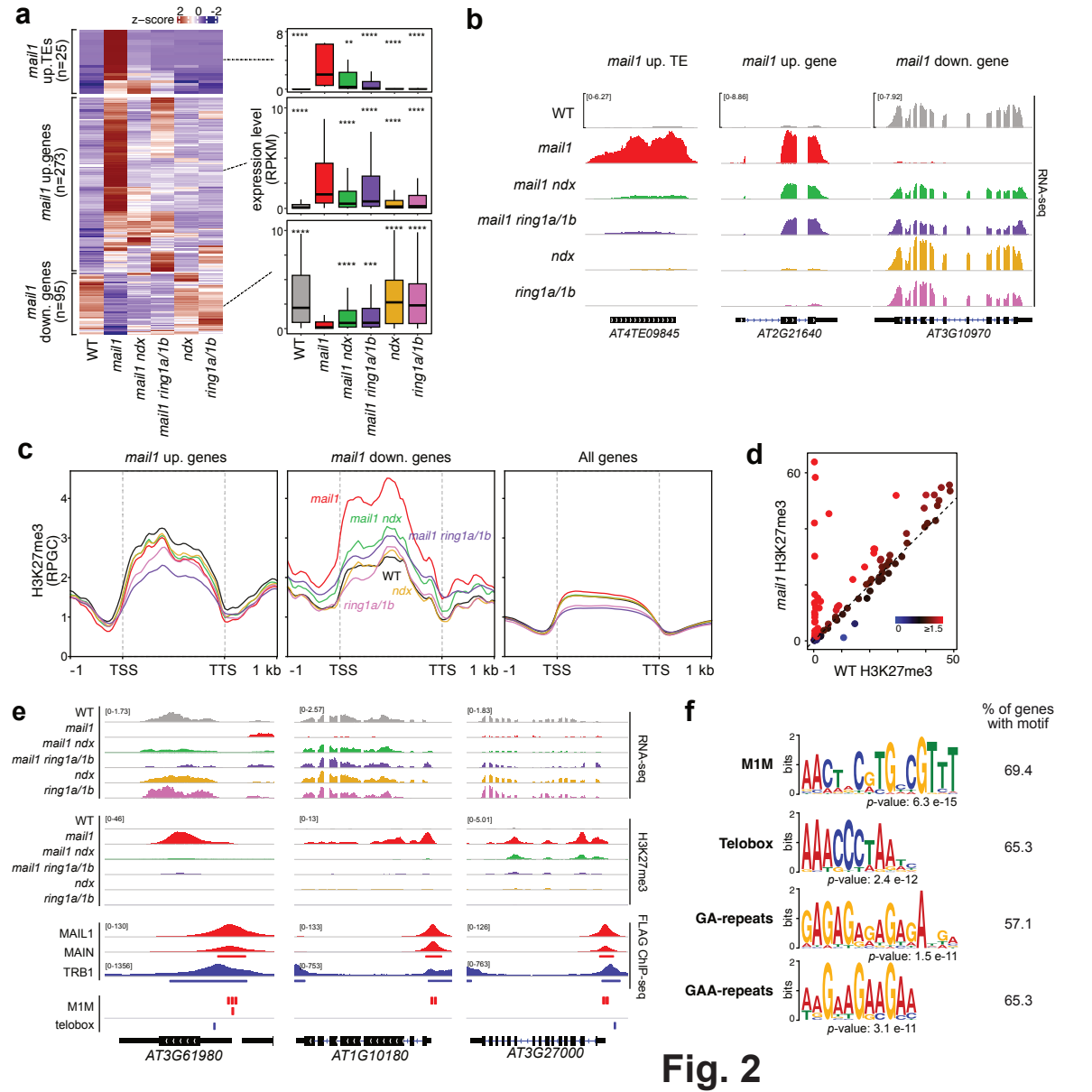
a
z-score
2
0
-2
mail1 up.TEs (n=25)
mail1 up.genes (n=273)
mail1 down. genes (n=95)
expression level (RPKM)
WT
mail1
mail1 ndx
mail1 ring1a/1b
ndx
ring1a/1b
b
mail1 up. TE
mail1 up. gene
mail1 down. gene
RNA-seq
AT4TE09845
AT2G21640
AT3G10970
c
mail1 up. genes
mail1 down. genes
All genes
H3K27me3 (RPGC)
TSS
TTS
1 kb
d
mail1 H3K27me3
WT H3K27me3
e
RNA-seq
H3K27me3
FLAG ChIP-seq
MAIL1
MAIN
TRB1
M1M
telobox
AT3G61980
AT1G10180
AT3G27000
f
% of genes with motif
M1M
69.4
p-value: 6.3 e-15
Telobox
65.3
p-value: 2.4 e-12
GA-repeats
57.1
p-value: 1.5 e-11
GAA-repeats
65.3
p-value: 3.1 e-11
bits

Fig. 2

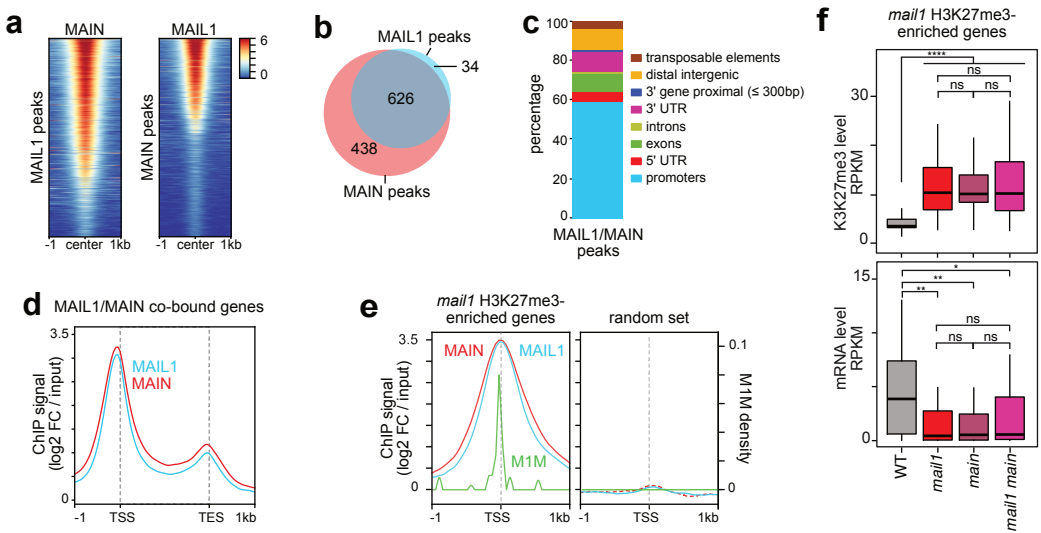
a
MAIN
MAIL1
MAIL1 peaks
MAIN peaks
6
0
-1 center 1kb
-1 center 1kb
b
MAIL1 peaks
34
626
438
MAIN peaks
c
percentage
100
80
60
40
20
0
transposable elements
distal intergenic
3' gene proximal (≤ 300bp)
3' UTR
introns
exons
5' UTR
promoters
MAIL1/MAIN peaks
d
MAIL1/MAIN co-bound genes
ChIP signal (log2 FC / input)
3.5
0
MAIL1
MAIN
-1 TSS TES 1kb
e
mail1 H3K27me3-enriched genes
random set
ChIP signal (log2 FC / input)
3.5
0
MAIN
MAIL1
M1M
0.1
0
M1M density
-1 TSS 1kb
-1 TSS 1kb
f
mail1 H3K27me3-enriched genes
K3K27me3 level RPKM
30
0
****
ns
ns
ns
mRNA level RPKM
15
0
**
**
*
ns
ns
ns
WT
mail1-
main-
mail1 main-

Fig. 3

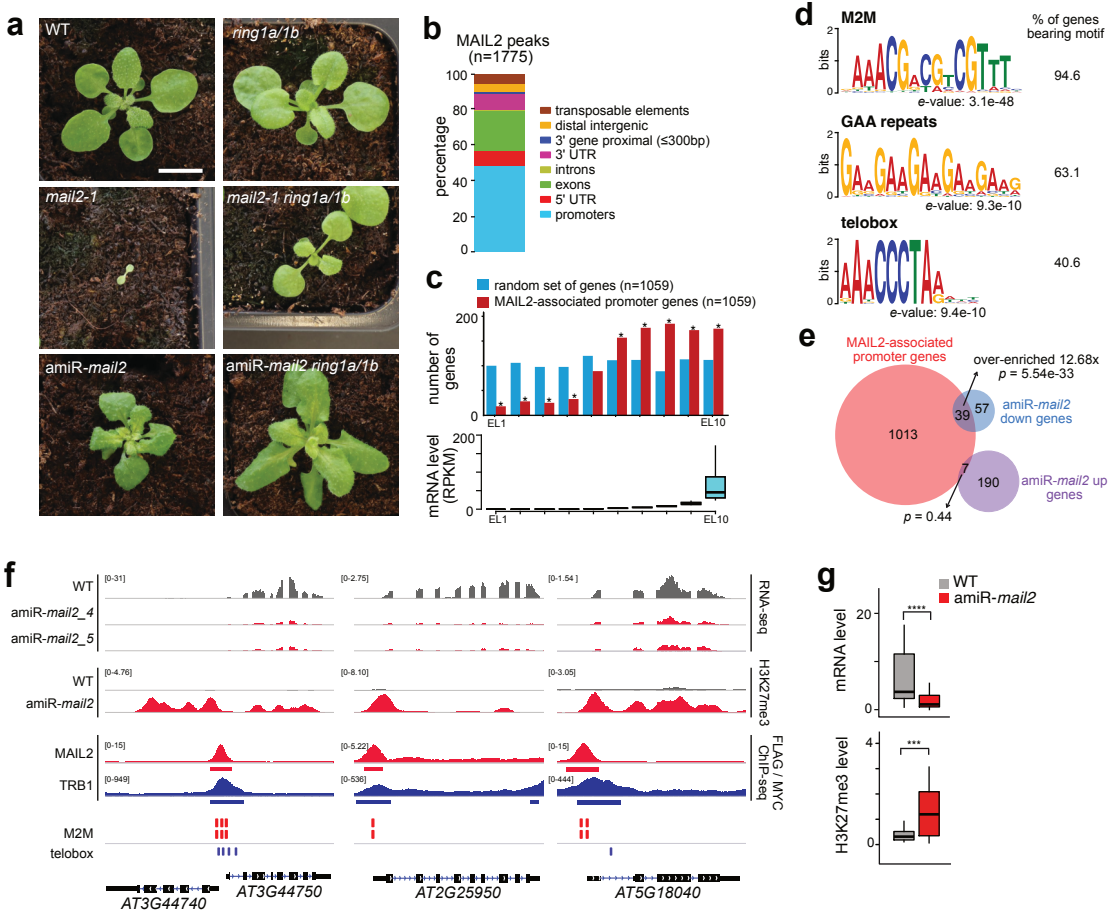
a
WT
ring1a/1b
mail2-1
mail2-1 ring1a/1b
amiR-mail2
amiR-mail2 ring1a/1b
b
MAIL2 peaks
(n=1775)
percentage
transposable elements
distal intergenic
3' gene proximal (≤300bp)
3' UTR
introns
exons
5' UTR
promoters
c
random set of genes (n=1059)
MAIL2-associated promoter genes (n=1059)
number of genes
mRNA level (RPKM)
EL1
EL10
d
M2M
% of genes bearing motif
94.6
e-value: 3.1e-48
GAA repeats
63.1
e-value: 9.3e-10
telobox
40.6
e-value: 9.4e-10
bits
e
MAIL2-associated promoter genes
over-enriched 12.68x
p = 5.54e-33
1013
39
57
amiR-mail2 down genes
7
190
amiR-mail2 up genes
p = 0.44
f
WT
amiR-mail2_4
amiR-mail2_5
RNA-seq
WT
amiR-mail2
H3K27me3
MAIL2
TRB1
FLAG / MYC ChIP-seq
M2M
telobox
AT3G44740
AT3G44750
AT2G25950
AT5G18040
g
WT
amiR-mail2
mRNA level
H3K27me3 level

Fig. 4

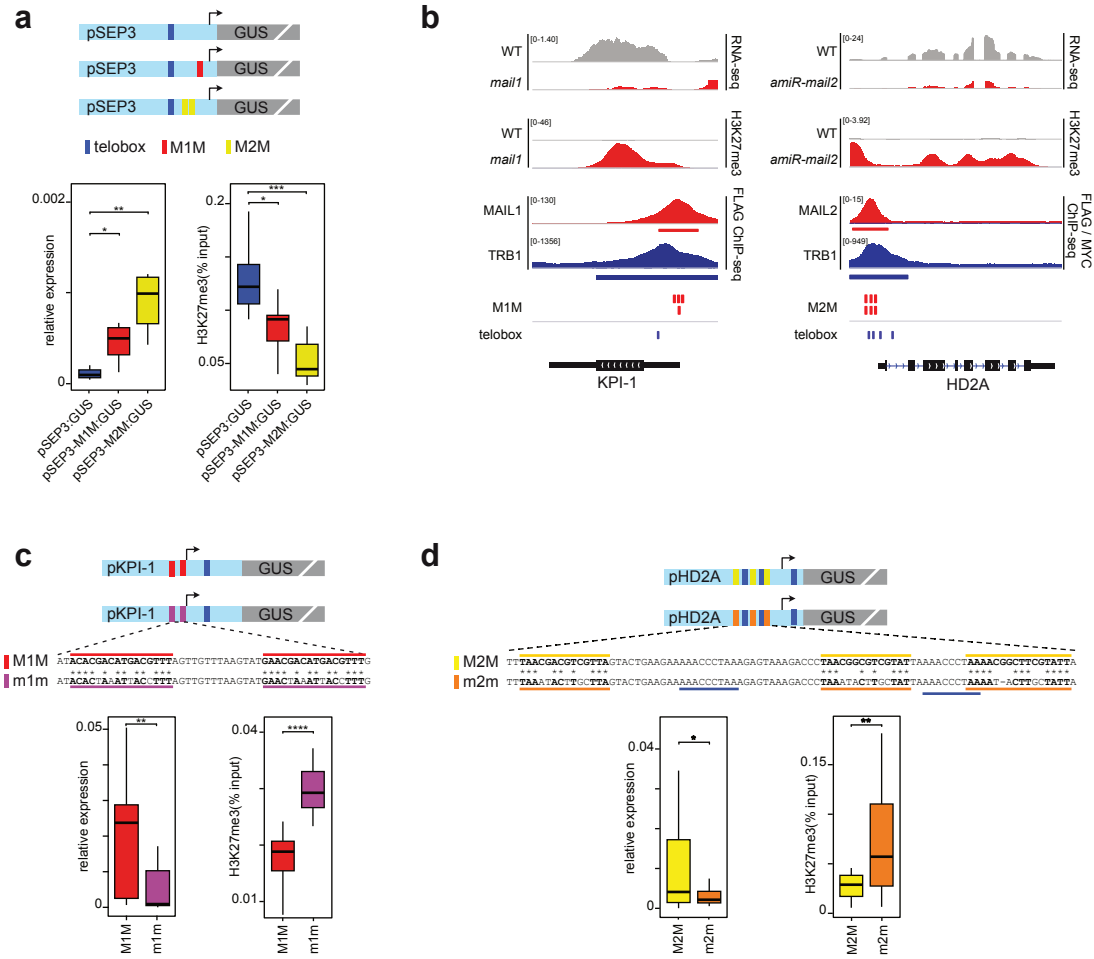
a
pSEP3
GUS
telobox
M1M
M2M
relative expression
H3K27me3(% input)
pSEP3:GUS
pSEP3-M1M:GUS
pSEP3-M2M:GUS
b
WT
mail1
amiR-mail2
RNA-seq
H3K27me3
MAIL1
MAIL2
TRB1
FLAG / MYC ChIP-seq
KPI-1
HD2A
c
pKPI-1
M1M
m1m
d
pHD2A
M2M
m2m

Fig. 5