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

62 counteract PcG-mediated silencing at specific genomic loci by preventing ectopic H3K27me3
63 incorporation.
64

65 Results

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

67 Mutations in components of the MAIL1/MAIN/PP7L complex result in severe developmental
68 defects and transcriptional abnormalities characterized by loss of silencing at several TEs and
69 mis-expression of numerous protein-coding genes[12–15]. To uncover the molecular basis of
70 these defects, we conducted a genetic screen for suppressors of the *mail1* mutant
71 phenotypes. Screening individuals from an ethyl methanesulfonate (EMS) mutagenized
72 population of *mail1* containing the *L5-GUS* transgene[16], we identified a mutant line exhibiting
73 both improved plant development and suppression of L5 silencing release (Fig. 1a,b).
74 Compared to *mail1* plants, the suppressed plants showed faster growth and greener leaves.
75 The traits followed a 1/3 ratio in the F2 progeny from a backcross, indicating a single-locus
76 nuclear recessive causative mutation. Mapping-by-sequencing using an outcross F2
77 population identified a knock-out allele of *AT4G03090* (*NODULIN HOMEOBOX; NDX*) that we
78 named *ndx-7*, associated with a C to T mutation leading to a premature stop codon
79 (Supplementary Fig. 1a). Crossing *mail1* plants with an independent *ndx-4* T-DNA insertion
80 mutant line and introduction of an *NDX* genomic copy in the *mail1 ndx-7* background
81 confirmed that mutation of *NDX* alleviated *mail1*-associated developmental and silencing
82 defects (Supplementary Fig. 1b-d). Beyond the *L5-GUS* transgene, RT-qPCR analysis revealed
83 strong suppression of *mail1*-transcriptional activation of several endogenous targets in *mail1*
84 *ndx-7* plants (Fig. 1c). Consistent with MAIL1's role in a complex with MAIN, introducing *ndx-*
85 *7* in the *main* background also mitigated *main*-associated developmental defects and GUS
86 upregulation (Supplementary Fig. 1b,d).

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

93 RING1A and RING1B. Crossing *val1* plants, which appear indistinguishable from WT, with
94 *mail1* or *main* mutant lines resulted in double mutants with very severe developmental
95 phenotype characterized by the apparent absence of functional shoot apical meristem, and
96 unable to initiate post-embryonic differentiation (Supplementary Fig. 1e). This indicates that
97 *ndx* and *ring1a/1b* suppression of *mail1* is independent of *VAL1*.

98 To assess the genome-wide importance of *NDX* and *RING1A/1B* on *mail1*-associated
99 transcriptional defects, we performed RNA-seq experiments using WT, *ndx*, *ring1a/1b*, *mail1*,
100 *mail1 ndx* and *mail1 ring1a/1b* lines. To account for the variability we observed in *mail1* plant
101 phenotypic severity -possibly due to subtle differences in growth conditions- we used four
102 RNA-seq replicates to define a robust set of *mail1*-specific misregulated targets, including
103 differentially expressed genes and a subset of upregulated TEs (Fig. 2a and Supplementary
104 Fig. 2a). Gene ontology (GO) analysis revealed that 29% of *mail1*-upregulated genes (79 out
105 of 273) are linked to hormone and stress pathways (Supplementary Fig. 2b), consistent with
106 *mail1* enhancing plant sensitivity to environmental cues. The expression of most misregulated
107 loci was unaffected in *ndx* or *ring1a/1b* mutants (Fig. 2a). Combining either *ndx* or *ring1a/1b*
108 mutations with *mail1* resulted in an overall suppression of TE upregulation (Fig. 2a,b).
109 Similarly, expression levels of *mail1*-upregulated genes tend to significantly return towards
110 WT levels in *mail1 ndx* and *mail1 ring1a/1b* (Fig. 2a,b). Additionally, *mail1*-associated
111 downregulation was suppressed for about 65% of the genes (62 to 68/95) in both mutant
112 combinations (Fig. 2a,b and Supplementary Fig. 2c). These results indicate that *NDX* and
113 *RING1A/1B* are required for the transcriptional misregulation induced in the absence of
114 *MAIL1*.

115 To determine the impact of mutations in the PRC2 core component CLF or SWN, we
116 generated *mail1 clf*-29 and *mail1 swn*-7 double mutants and performed RNA-seq. CLF and
117 SWN act largely redundantly to catalyze H3K27me3 deposition in vegetative tissues, although
118 certain loci appear to be preferentially targeted by one or other H3K27 methyltransferases[22].
119 We found that *mail1*-induced transcriptional misregulation at genes was significantly
120 suppressed in *mail1 clf* but not in *mail1 swn* (Supplementary Fig. 2d), although both double
121 mutants showed developmental defects mostly similar to those of *mail1* (Supplementary Fig.
122 2e). This indicates that mutation of *CLF* tends to suppress *mail1*-induced transcriptional
123 defects, albeit to an insufficient extent to result in an improved phenotype, most likely
124 because of the functional redundancy with SWN[22].

125 Our results show that *NDX, RING1A/1B,* and to a lesser extent *CLF*, are required for the
126 transcriptional misregulation induced in the absence of *MAIL1,* suggesting a potential link
127 between MAIL1 and PcG-mediated silencing.
128
129 **MORC1 and MORC2 dampen TE activation in *mail1* suppressor backgrounds**
130 MICRORCHIDIA (MORC) proteins are ATPases required for TE silencing and heterochromatin
131 condensation[23]. We recently linked silencing defects at a subset of TEs to downregulation of
132 MORC1 in *mail1* plants[15]. In both suppressor backgrounds, loss of TE activation correlated
133 with increased *MORC1* transcript accumulation relative to *mail1*, albeit limited in *mail1 ndx*
134 plants (Supplementary Fig. 3a). Interestingly, expression of the closely related *MORC2* gene,
135 also involved in TE silencing[24], showed a four to six fold induction in the presence of *ndx* or
136 *ring1a/1b* mutations (Supplementary Fig. 3a). We tested whether *MORC2* could contribute to
137 the reestablishment of TE silencing by introgressing the *morc2-1* mutation into the *mail1 ndx*
138 background and performing RNA-seq to monitor TE expression. Of the 25 *mail1*-upregulated
139 TEs, at least six showed dependency towards MORC2 for their repression in *mail1 ndx* plants
140 (Supplementary Fig. 3b,c), and we previously showed that overexpressing *MORC1* in the *mail1*
141 background also repressed some *mail1*-activated TEs[15] (Supplementary Fig. 3d). This is
142 consistent with MORC1 and MORC2 acting redundantly in TE silencing[24] and suggests that
143 upregulation of both *MORC1* and *MORC2* contributes to dampening TE activation in *mail1 ndx*
144 and *mail1 ring1a/1b*.
145
146 **Gene silencing in *mail1* correlates with ectopic H3K27me3 incorporation**
147 In light of the suppressor effect of *ndx* and *ring1a/1b* mutations and the known interaction of
148 NDX and RING1A/1B in PcG-mediated silencing[17,19], we explored a possible link between
149 *mail1*-associated gene misregulation and the PcG-mediated silencing diagnostic mark
150 H3K27me3. Chromatin immunoprecipitation sequencing (ChIP-seq) analyses revealed no
151 overall change in H3K27me3 levels in *mail1* when considering all Arabidopsis genes (Fig. 2c).
152 Likewise, the *ndx* mutation did not significantly alter H3K27me3 distribution over genes or
153 TEs (Fig. 2c and Supplementary Fig. 4). In *mail1,* upregulated genes showed slight loss of
154 H3K27me3. The addition of the *ndx* mutation had no additional effect, while H3K27me3 levels
155 were further reduced at these genes in *mail1 ring1a/1b* mutants (Fig. 2c and Supplementary
156 Fig. 5a). Despite this reduction, both *ndx* and *ring1a/1b* mutations tend to suppress *mail1*-

157 induced transcriptional activation of these genes (Fig. 2a), suggesting that their misregulation
158 in *mail1* is not directly connected to H3K27me3 variation. Strikingly, gene downregulation in
159 *mail1* correlated with prominent H3K27me3 enrichment, which was largely reversed upon
160 additional *NDX* or *RING1A/1B* depletions (Fig. 2c and Supplementary Fig. 5a) that also tended
161 to restore expression of these genes (Fig. 2a). Many genes acquired high levels of H3K27me3
162 *de novo* in *mail1* (Fig. 2d), correlating with transcriptional repression (Supplementary Fig. 5b),
163 and this was largely reversed in the *mail1 ndx* and *mail1 ring1a/1b* suppressed lines (Fig. 2e
164 and Supplementary Fig. 5c). One such gene was *MORC1*, gaining H3K27me3 in *mail1*, which
165 was completely lost in *mail1 ring1a/1b* (Supplementary Fig. 5c). We also found that its
166 neighbor gene *MORC2* was subject to PcG-mediated silencing in the WT, explaining its
167 upregulation in the presence of *ndx* or *ring1a/1b* mutations (Supplementary Fig. 3a,5c).
168 Interestingly, H3K27me3-enriched genes in *mail1* included *SPO11-1* (Supplementary Fig. 5c),
169 which is mostly expressed in flowers and encodes a topoisomerase-like protein essential for
170 meiotic recombination and fertility[25,26]. Gain of H3K27me3 was maintained in *mail1* immature
171 flowers, correlating with about a 70% decrease in *SPO11-1* transcript accumulation
172 (Supplementary Fig. 5d). Such level of *SPO11-1* downregulation was reported to significantly
173 affect gamete viability[26], and thus likely contributes to *mail1* hypofertility[27].

174 RING1A and RING1B are largely required for PRC1-mediated H2A ubiquitination[18],
175 playing a key role in recruiting H3K27me3 at numerous loci[1]. This is likely the case at genes
176 gaining H3K27me3 in *mail1* (Supplementary Table 1) as these showed concomitant
177 enrichment in H2AK121ub (Supplementary Fig. 6a,b). Together, these results suggest that loss
178 of *MAIL1* directly leads to PcG-mediated silencing at many genes.

179 To search for possible common features associated with the gain of H3K27me3 at
180 genes in *mail1*, we performed a DNA motif-based sequence analysis on the 1-kb region
181 surrounding the transcription start site (TSS) of these genes. Four conserved DNA motifs were
182 detected in 60-70% of the genes (Supplementary Fig. 6c), but only the top-scoring motif was
183 significantly enriched compared to the control set (Fig. 2f). We named this motif M1M (for
184 MAIL1/MAIN Motif), which notably extends a previously reported "down" motif partially
185 enriched among *mail1*- and *main*-downregulated genes[13]. Interestingly, the other motifs
186 included telobox and GA-repeat elements, two known polycomb response elements (PREs)
187 bound by transcription factors to tether PRC2 complexes to specific genes[22,28,29]. Using
188 previously published data[30], we found that the TELOMERE-REPEAT BINDING FACTOR (TRB)1,

189 TRB2 and TRB3, which bind telobox elements were loaded in wild-type plants close to the TSS
190 of a significant fraction (31/49) of *mail1* H3K27me3-enriched genes (Fig. 2e and
191 Supplementary Fig. 6d,e). Together, these results suggest that the MAIL1/MAIN complex
192 might promote the expression of some genes by preventing, directly or indirectly, PcG-
193 mediated silencing. Upon MAIL1/MAIN depletion, PRC2 recruitment, through TRB proteins or
194 other factors, requires PRC1 activity to be robustly established, potentially through a positive
195 feedback loop[1].
196
197 **MAIL1/MAIN complexes are specifically targeted to M1M motifs**
198 To explore the genomic localizations of MAIL1 and MAIN *in vivo*, we performed ChIP-seq using
199 transgenic lines expressing functional FLAG- or MYC-tagged versions of MAIL1 or MAIN under
200 their native promoters in their respective mutant backgrounds[13]. MAIL1 and MAIN largely
201 colocalized and were predominantly enriched at proximal promoter regions of genes (Fig. 3a-
202 d). Notably, they bound approximately 68% of the genes that gain H3K27me3 in *mail1*, but
203 showed no enrichment at genes with reduced H3K27me3 levels (Fig. 3e and Supplementary
204 Fig. 7a). These H3K27me3-enriched genes showed similar transcriptional downregulation and
205 H3K27me3 accumulation in *mail1*, *main*, and *mail1 main* mutants, consistent with MAIL1 and
206 MAIN acting in the same regulatory complex (Fig. 3f). Importantly, motif analysis revealed
207 that approximately 86% of MAIL1/MAIN-bound regions harbor the M1M motif
208 (Supplementary Fig. 7b), strongly suggesting a functional role for M1M in recruiting
209 MAIL1/MAIN to proximal promoter regions (Fig. 2e, 3e).
210
211 **NDX and RING1A are enriched at MAIL1/MAIN binding regions**
212 To further explore how mutations of NDX may suppress *mail1*, we sought to determine the
213 genomic localization of NDX. We generated transgenic lines expressing functional FLAG- or
214 MYC-tagged NDX in *ndx-7* and *mail1 ndx* mutant backgrounds (Supplementary Fig. 8a,b).
215 However, ChiP-seq experiments using anti-MYC or anti-FLAG antibodies were unsuccessful,
216 potentially due to low NDX protein abundance. Quality control analysis indicated that recently
217 published NDX ChIP-seq data[31] were not suitable for further use. We therefore turned to a
218 CUT&Tag approach, which provided good quality data only for the *ndx-7* background. Both
219 NDX-MYC and FLAG-NDX showed similar genomic distributions, with clear enrichment at
220 genes, particularly in promoter regions upstream of the TSSs, and depletion at TEs, especially

221 those located in pericentromeric heterochromatin (Supplementary Fig. 8c,d). This distribution
222 is consistent with NDX belonging to a large family of transcription factors[32,33]. Furthermore,
223 re-analysis of published RING1A ChIP-seq data[34] revealed substantial colocalization with NDX,
224 consistent with their previously reported physical interaction[17,19] (Supplementary Fig. 8e,f).
225 Notably, both NDX and RING1A were enriched at MAIL1/MAIN binding sites (Supplementary
226 Fig. 8e), suggesting that MAIL1/MAIN may function to prevent the recruitment of
227 downstream PcG-silencing components at their target genes.
228
229 **MAIL2 function is essential to plant development and opposes PcG-mediated silencing at**
230 **distinct loci**
231 We previously reported that *MAIL1* and *MAIN* are Brassicaceae-specific paralogues of *MAIL2*,
232 which is widely distributed among plant species[13]. The *MAIL2* gene appears to be under strong
233 purifying selection[13] suggesting that it may fulfill an important function. In line with this
234 assumption, the *mail2-1* transfer DNA mutant displayed a very severe developmental
235 phenotype. The shoot apical meristem (SAM) was unable to initiate a proper post-embryonic
236 program, with plants rarely developing true leaves and typically arresting growth at the two
237 cotyledons stage (Fig. 4a and Supplementary Fig. 9a,b). Notably, we were unable to obtain
238 homozygous mutants in the progeny segregating from a self-fertilized heterozygous *mail2*
239 CRISPR mutant, suggesting that *mail2-1* may not be a null mutant.
240 The MAIL1, MAIN, and MAIL2 proteins share a highly homologous PMD-C domain[27]
241 (Supplementary Fig. 10), which is the only identified domain in these proteins. Given the
242 similarity in structure, we hypothesized that MAIL2 might be involved in a protection
243 mechanism against polycomb silencing similar to the one identified for MAIL1/MAIN. To test
244 this hypothesis, we introduced *ndx*, *clf*, or *ring1a/1b* mutations into the *mail2-1* background
245 through crossing. The resulting mutants showed varying degrees of phenotypic suppression:
246 *mail2-1 ndx* and *mail2-1 clf* double mutants developed a first pair of leaves before arresting
247 growth, while *mail2-1 ring1a/1b* mutants displayed more extensive development,
248 progressing through the rosette stage and approaching the floral transition (Fig. 4a and
249 Supplementary Fig. 9a). These observations indicate that *ndx*, *clf*, and *ring1a/1b* mutations
250 partially suppress the *mail2-1* developmental phenotype, mirroring their effects on *mail1* and
251 *main* mutants. This suggests a functional similarity between MAIL2 and MAIL1/MAIN in

252 counteracting PcG-mediated silencing, implying that this anti-PcG silencing function may be
253 an ancestral feature of the PMD-C protein family.

254 We determined the genomic localization of MAIL2 by ChIP-seq using a transgenic line
255 expressing a functional, tagged version of MAIL2 (Supplementary Fig. 9c). This analysis
256 identified 1775 MAIL2-enriched regions across the genome, largely distinct from those bound
257 by MAIL1/MAIN (Supplementary Fig. 11a), yet similarly associated with elevated NDX and
258 RING1A occupancy (Supplementary Fig. 8e). As in the case of MAIL1 and MAIN, these regions
259 were predominantly associated with genes, particularly within proximal promoter regions
260 (1059 genes) (Fig. 4b and Supplementary Fig. 11b). These genes were significantly
261 overrepresented in the five deciles showing the highest expression level in the WT (Fig. 4c).

262 We performed DNA motif-based sequence analyses using MAIL2 peaks or the 1kb
263 region surrounding TSS of MAIL2-associated gene promoters and identified a conserved
264 sequence element, which we named MAIL2-Motif (M2M), present in nearly 95% of both the
265 peaks and promoter regions (Fig. 4d and Supplementary Fig. 11c-e). Intriguingly, the M2M
266 consensus resembles that of the M1M motif but in a palindromic configuration, suggesting
267 M1M may have derived from M2M (Supplementary Fig. 11f), and mirroring the evolutionary
268 relationship between the *MAIL1/MAIN* genes and *MAIL2*. As observed for MAIL1/MAIN-
269 target genes, GAA-repeats and telobox elements were also present in a substantial fraction
270 of MAIL2-associated gene promoters (Supplementary Fig. 11c).

271 To facilitate exploration of MAIL2 function, we generated hypomorphic *mail2* mutants
272 using artificial microRNAs (amiRNAs) targeting *MAIL2* mRNA (amiR-*mail2*) (Fig. 4a and
273 Supplementary Fig. 9b). While many amiR-*mail2* plants displayed severely compromised SAM
274 development under *in vitro* selection (Supplementary Fig. 12a) we were able to isolate two
275 amiR-*mail2* lines (amiR-*mail2_4* and amiR-*mail2_5*) that developed further. Both homozygous
276 lines showed around 2.5-fold reduction in *MAIL2* mRNA levels (Supplementary Fig. 12b) and
277 similar developmental phenotypes, including shorter stature, curly leaves, short stems and
278 hypofertility, which were partially suppressed by the introduction of *ring1a/1b* mutations
279 (Fig. 4a and Supplementary Fig. 12c). We confirmed that these phenotypes were caused by
280 *MAIL2* downregulation by restoring MAIL2 function in amiR-*mail2_4* using an amiRNA-
281 insensitive *MAIL2* transgene (Supplementary Fig. 9b, 12d). We evaluated the impact of *MAIL2*
282 downregulation on transcription by performing RNA-seq of amiR-*mail2_4* and amiR-*mail2_5*
283 lines. The transcriptomes of the two lines were highly similar, with differentiated expressed

284 loci almost exclusively comprising genes, either up- (197) or down-regulated (96 genes)
285 (Supplementary Table 2). Downregulated genes were significantly enriched for MAIL2-
286 associated genes and were predominantly found in the top five expression quantiles in WT
287 plants (Fig. 4c,e). These genes showed no enrichment for specific biological pathway and were
288 distinct from those downregulated in *mail1* mutants (Supplementary Fig. 13a). Introduction
289 of amiR-*mail2* into the *mail1* background drastically worsened the *mail1* plant phenotype
290 (Supplementary Fig. 13b), indicating that MAIL1 and MAIL2 function in parallel pathways
291 regulating distinct gene sets. In amiR-*mail2*, H3K27me3 distribution was also affected, and a
292 large fraction of protein-coding genes (75 out of 102) gaining this mark were also bound by
293 MAIL2 (Supplementary Fig. 13c). We further found that genes showing MAIL2 binding to their
294 promoter and downregulation in amiR-*mail2* were associated with TRB1/2/3 in WT plants
295 (Fig. 4f and Supplementary Fig. 13d,e), and specifically gained H3K27me3 upon *MAIL2*
296 downregulation (Fig. 4e,f). We conclude from these results that MAIL2 and MAIL1 are
297 involved in similar pathways antagonizing PcG-silencing to promote expression of largely
298 distinct sets of genes
299
300 **M1M and M2M are required to promote expression and oppose H3K27me3 deposition**
301 To assess the functional relevance of M1M and M2M motifs, we employed two
302 complementary approaches based on transgenes. First, we used a transgenic construct in
303 which the promoter of the *SEP3* gene drives *GUS* expression[29]. The *SEP3* promoter contains a
304 telobox that allows PRC2-mediated H3K27me3 deposition through TRB1 binding, resulting in
305 low *GUS* expression[29]. Inserting either M1M or M2M in the *SEP3* promoter between the
306 telobox and the transgene's TSS significantly enhanced *GUS* expression and reduced
307 H3K27me3 levels (Fig. 5a). In a complementary approach, we generated transgenic lines using
308 the promoters of two endogenous genes: *KPI-1*, a MAIL1/MAIN target, or *HD2A,* a MAIL2
309 target, both of which are transcriptionally downregulated and enriched for H3K27me3 in their
310 respective mutant backgrounds (Fig. 5b). These promoters naturally contain teloboxes and
311 either M1M (KPI-1 promoter) or M2M (HD2A promoter) motifs. We independently
312 transformed WT plants with constructs carrying either the native promoter or mutated
313 version of M1M or M2M at the most conserved positions. We found that mutating either
314 M1M or M2M motif led to significant downregulation of *GUS* expression and increased
315 H3K27me3 deposition (Fig. 5c,d). These results demonstrate that the integrity of M1M and

316 M2M sequences is required to promote expression and antagonize H3K27me3-associated
317 repression.
318

319 Discussion

320 We previously reported that MAIL1, MAIN, and PP7L form a regulatory complex that primarily
321 controls gene expression in Arabidopsis[12–15], yet the molecular mechanisms remained
322 unclear. Here, we show that mutations in NDX or in the PRC1 components RING1A and
323 RING1B alleviate the developmental and molecular defects associated with MAIL1 depletion,
324 revealing an unexpected functional antagonism between the MAIL1/MAIN pathway and
325 Polycomb silencing.
326

327 NDX, a homeodomain-containing protein, interacts with RING1A/B and stabilizes
328 transcriptional states at euchromatic genes[17,19,35]. It was also recently implicated in
329 influencing siRNA accumulation and DNA methylation, predominantly at pericentromeric
330 TEs[31]. While altered DNA methylation can lead to ectopic H3K27me3 deposition at TEs [36,37],
331 we did not detect global changes in H3K27me3 patterns in *ndx* mutants (Supplementary Fig.
332 4). We observed reciprocal co-enrichment of NDX and RING1A/B at their binding sites
333 (Supplementary Fig. 8e,f), consistent with their demonstrated physical interaction[17,19].
334 Notably, NDX enrichment was detected at protein-coding genes rather than TEs
335 (Supplementary Fig. 8d), contrasting with prior reports of its localization at centromeric and
336 pericentromeric regions[31]. Moreover, NDX and RING1A are co-enriched at MAIL1/MAIN and
337 MAIL2-binding sites (Supplementary Fig. 8e), suggesting that these loci are poised to recruit
338 PRC2. We propose a model whereby MAIL1/MAIN and MAIL2 maintain gene expression by
339 limiting PcG activity at these loci. Upon depletion of MAIL1 or MAIL2, RING1A and NDX may
340 promote PRC2 recruitment, leading to H3K27me3 deposition and gene repression. Additional
341 cofactors such as TRB proteins, known PRC2 recruiters, may further contribute to this process
342 at specific targets.
343

344 Interestingly, although VAL1 also recruits PRC1 and PRC2 and cooperates with NDX at
345 *FLC*[19–21], *val1* mutations do not suppress *mail1* defects but instead exacerbate them
346 (Supplementary Fig. 1e). This suggests that different PRC1/2 recruitment pathways do not

347 contribute equally to MAIL1/MAIN-dependent gene regulation. The aggravated phenotype of
348 *mail1 val1* double mutants may reflect misregulation of distinct gene subsets affected in each
349 background (Supplementary Fig. 14a), although genome-wide binding of VAL1[21] includes
350 many MAIL1/MAIN targets (Supplementary Fig. 14b), indicating possible context-dependent
351 cooperation.
352
353 Our findings establish MAIL1/MAIN and MAIL2 as key regulators that counteract
354 Polycomb-mediated repression to safeguard gene expression. Despite targeting distinct
355 genes via specific promoter motifs (M1M and M2M), these PMD-C proteins likely operate
356 through a shared molecular mechanism. Structural modeling and DNA-binding predictions
357 suggest that the PMD-C domains of MAIL1, MAIN and MAIL2 share a conserved concave
358 surface, a structural feature commonly associated with DNA-binding domains[38], that includes
359 residues predicted to mediate DNA contact (Supplementary Fig. 15). Moreover, the MAIL1
360 PMD-C domain shows structural similarity to the DUF1985 domain of the transposon-
361 encoded anti-silencing protein VANC[39], a known DNA-binding factor[40–42]. This raises the
362 possibility that PMD-C proteins may directly bind DNA, although this interaction remains to
363 be experimentally demonstrated. Alternatively, additional DNA- or chromatin-binding factors
364 may be required to guide or stabilize the interaction of PMD-C with DNA.
365
366 We previously linked TE reactivation in *mail1* and *main* mutants to reduced expression
367 of the silencing effector MORC1[15]. Here, we show that *MORC1* is among the genes protected
368 from PcG repression by MAIL1/MAIN. In *mail1 ndx* and *mail1 ring1a/b* mutants, TEs are
369 largely re-silenced, likely due to restored expression of MORC1 along with upregulation of its
370 paralog MORC2 (Fig. 2b and Supplementary Fig. 3). Many genes are upregulated in *mail1*,
371 possibly due to the downregulation of transcriptional repressors. However, we did not
372 identify any clear candidate. *MORC2* is unlikely to be involved, as its expression remains very
373 low in both *mail1* and WT (Supplementary Fig. 3a), and *morc2* mutants show no gene
374 deregulation[24]. *MORC1* is also most likely not responsible, as the genes upregulated in *mail1*
375 and *morc1* are largely distinct (Supplementary Fig. 14c), and *MORC1* overexpression in *mail1*
376 has minimal impact[15]. Alternatively, gene upregulation in *mail1* may result from changes in
377 higher-order chromatin architecture associated with loss of MAIL1 function[12].
378

379 *MAIL1* and *MAIN* likely evolved from a duplication of *MAIL2* that occurred specifically
380 within the Brassicaceae lineage[13]. The sequence similarity between the M1M and M2M
381 motifs suggests that the palindromic M2M motif may have evolved into M1M at the same
382 time. *MAIL2* appears conserved at least in all eudicot plants[13], raising the possibility that this
383 PMD protein-DNA motif regulatory module may be conserved across a broad range of plant
384 species. Only a subset (about 20-40%) of M1M and M2M motifs are bound by MAIN/MAIL1
385 or MAIL2 (Supplementary Fig. 13f), suggesting that additional regulatory layers restrict their
386 binding to specific active genes. Future studies will address this point.
387
388 We previously defined four clades of PMD proteins (A1, A2, B and C). PMD-As are
389 encoded by retrotransposons, while PMD-B are found in protein-coding genes. MAIL1, MAIN
390 and MAIL2 belong to the PMD-C clade, which comprises protein synthesized from *bona fide*
391 host genes as well as DNA transposon-encoded proteins from various monocots and eudicots
392 plant species[12] (Supplementary Fig. 16 and Supplementary doc. 1). The predicted structures
393 of MAIL1, MAIN and MAIL2 PMD-C domains are highly similar (Supplementary Fig. 10) and, as
394 mentioned above, share interesting structural homology with the DUF1985 domain of the
395 DNA transposon-encoded VANC proteins[39], which function as an anti-silencing factor
396 antagonizing DNA methylation [40–42]. Recent studies indicate that H3K27me3 plays a role in TE
397 repression in a broad range of eukaryotes, including Arabidopsis, suggesting an ancestral role
398 of Polycomb silencing in controlling TE activity[43,44]. This raises the intriguing possibility that
399 TE-encoded PMD-C proteins also function as Polycomb anti-silencing factors, a hypothesis we
400 are currently testing. Such a connection between transposon biology and Polycomb silencing
401 regulation could provide valuable insights into the coevolution of genome defense and gene
402 expression control mechanisms.
403

404 ## Methods

405 **Plant Material.** Mutants *mail1-1* (GK-840E05)[27], *main-2* (GK-728H05)[45], *mail2-1*
406 (SAIL_1153_F07), *ndx-4* (WiscDsLox344A04)[35], *ring1a-2* (SAIL_393_E05)/*1b-3*
407 (SAIL_520_B04)[18], *clf-29* (SALK_021003)[46], *swn-7* (SALK_109121)[6] and *val1-2*
408 (SALK_088606C)[21] are all in the Col-0 background. The transgenic L5 line[16] was kindly provided
409 by H. Vaucheret (IJPB, France).

410 Design of the amiRNA (amiR) construct directed against *MAIL2* transcript was done using the
411 WMD3-Web MicroRNA Designer (http://wmd3.weigelworld.org) and the amiR-*MAIL2*
412 fragment cloned into the BIN-HYG-TX plasmid under the control of p35S promoter.
413 The *pMAIL1:gMAIL1-FLAG*, *pMAIL1:gMAIL1-MYC*, *pMAIN:gMAIN-FLAG*, *pMAIN:gMAIN-MYC*
414 and *SEP3pro:GUS* constructs were described previously[13,29]. For the *pMAIL2:gMAIL2-9xMYC*
415 construct, two genomic fragments encompassing the *MAIL2* gene (-1834 to +2448 relative to
416 ATG codon; A designated as +1) were PCR-amplified from Col-0 genomic DNA. PCR primers,
417 with several point mutations to *MAIL2* sequence, allowed to reconstitute a *MAIL2* transgene
418 bearing silent mutations and insensitive to amiR-*MAIL2* targeted degradation (Supplementary
419 Fig. 9b); recombination-based technology (NEB, NEBuilder HiFi DNA Assembly Master Mix
420 #2621) was then used to replace *MAIL1*-specific sequences of *pMAIL1:gMAIL1-MYC* vector by
421 the *MAIL2* transgene fragment. The *pNDX:gNDX* fragment was PCR amplified from Col-0
422 genomic DNA and region from -1631 to +5177 relative to ATG codon was cloned into the
423 PMDC123 vector. This *pNDX:gNDX* construct was used to produce a Nter-FLAG version of
424 NDX by inserting three repeats of the FLAG tag in frame with the second amino acids of the
425 NDX coding sequence. A Cter-9xMYC tag version was obtained through a recombination-
426 based approach similar to the one described for generating the *pMAIL2:gMAIL2-9xMYC*. For
427 *KPI-1:GUS* and *HD2A:GUS* transgenes, proximal gene promoter and 5'UTR amplicons of 459bp
428 and 357bp respectively were fused to *GUS* coding sequence and cloned into the PMDC123
429 vector. M1M and M2M were mutated using *pKPI-1:GUS* or *pHD2A:GUS* as matrix and
430 sequence modifications are described on Fig. 5. Insertion of M1M (-321 to -306 relative to
431 ATG) or M2M (-460 to -423 relative to ATG) in *proSEP3:GUS* was done by inverse PCR.
432 All transgenic constructs were introduced in plants through agroinfiltration by floral dipping[47]
433 and T1 transformants were isolated following adequate antibiotic selection.
434 Plants were grown on soil (60% relative humidity) or *in vitro* under long-day conditions (16 h
435 light, 8 h dark) at 23 °C.
436
437 **Mutagenesis and mapping.** Ethyl methanesulfonate (EMS) mutagenesis of the L5/*mail1-1*
438 mutant line was done as previously described[12] . To isolate suppressors of the *mail1* mutation,
439 a double screen was performed based on: 1/ visual amelioration of plant phenotype 2/ GUS
440 staining to confirm L5 transgene resilencing ; GUS assay was done by dissecting one leaf per
441 M2 plant candidate and leaves were incubated at 37 °C for 24 h in 400µg ml^{-1} 5-bromo-4-

442 chloro-3-indolyl-b-D-glucuronic acid, 10 mM EDTA, 50 mM sodium phosphate buffer pH 7.2,
443 0.2% triton X-100. The *ndx-7* mutant allele was identified following mapping-by-sequencing
444 using genomic DNA from a pool of 47 individual F2 mutant plants segregating from a cross
445 with Ler
446
447 **RNA isolation and analyses.** Total RNA was extracted from aerial tissues of 12-day-old plants
448 or immature flower buds using RNAzol RT (RN190, Molecular Research Center, Inc). For real
449 time RT-qPCR, 0.5 µg of total RNA was DNase treated and reverse transcribed using the
450 PrimeScript RT reagent kit with gDNA eraser (Perfect real time) (TaKaRa) in a final volume of
451 20 µl and one microlitre of cDNA was used for subsequent amplification using SYBR Premix Ex
452 Taq II (Tli RnaseH Plus) (TaKaRa) and Real-Time LightCycler 480 system (Roche) in a final
453 reaction volume of 15 µl. Data were normalized to a reference gene and analyzed according
454 to the $2^{-\Delta\Delta Ct}$ method. Means and standard errors of the mean were calculated from
455 independent biological samples. Primers used are listed in Supplementary Table 3.
456 For RNA-seq, DNase and purification were performed using RNA Clean &
457 Concentrator-5 kit (Zymo Research). Libraries were generated from about one µg total RNA
458 and sequenced as PE100 or PE150 reads on a DNBSEQ platform at Beijing Genomics Institute
459 (Hong Kong).
460
461 **ChIP-seq.** About two grams of aerial two-week-old rosettes were collected and incubated 15
462 min under vacuum in 25 ml of freshly prepared cross-linking buffer (10 mM phosphate buffer
463 pH 7, 100 mM NaCl, 0.1 M sucrose, 1% formaldehyde). 2 M glycine was added to stop cross-
464 link reaction (0.125 M final) and incubated an additional 5 min under vacuum. Tissues were
465 then washed three times with 10 mM phosphate buffer pH 7, 100 mM NaCl, 0.1 M sucrose,
466 pooled in 50 ml falcon tube and snap frozen in liquid nitrogen. Cross-linked tissues were
467 ground into fine powder by adding 7 ceramic balls to falcon tube and vortexing, and
468 resuspended into 25 ml of extraction buffer 1 (0.4 M Sucrose, 10 mM Tris-HCl pH 8, 10 mM
469 MgCl2, 5 mM β-mercaptoethanol, 0.1 mM PMSF, 1X complete EDTA-free cocktail inhibitor
470 (Roche)) at 4°C. Alternatively, collected tissues were ground into fine powder with liquid
471 nitrogen and samples were resuspended with 25 ml Hepes buffer containing 1%
472 formaldehyde (50 mM Hepes pH 8, 1 M sucrose, 5 mM KCl, 5 mM MgCl2, 0.6% triton X-100,
473 0.4 mM PMSF, 1% formaldehyde, 1X complete EDTA-free cocktail inhibitor), and incubated

474 for 10 min at RT. Crosslinking was stopped with 2 M glycine (0.125 M final), inverting several
475 times. The lysates were filtered through one layer of Miracloth and the nuclei were collected
476 by centrifuge at 4°C with 4000 rpm for 20 min. The nuclei were resuspended with extraction
477 buffer 2 (0.25 M Sucrose, 10 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 1% triton X-100, 5 mM β-
478 mercaptoethanol, 0.1 mM PMSF, 1X complete EDTA-free cocktail inhibitor), centrifuge with
479 14000 g at 4°C for 10 min, and then resuspended with 400 µl extraction buffer 3 (1.7 M
480 Sucrose, 10 mM Tris-HCl pH 8, 0.15% Triton X-100, 2 mM $MgCl_2$, 5 mM β-mercaptoethanol,
481 0.1 mM PMSF, 1X complete EDTA-free cocktail inhibitor). The nuclei suspension was loaded
482 over a cushion of 400 µl extraction buffer 3 and microtubes were centrifugated with 14000 g
483 for 1 h at 4°C. For chromatin shearing using Bioruptor UCD-200 (Diagenode), nuclei were
484 resuspended in 250 µl of lysis buffer (50 mM Tris-HCl pH 8, 10 mM EDTA pH 8, 1% SDS, 1X
485 complete EDTA-free cocktail inhibitor) and 25 cycles of 25 s 'ON', 1 min 15 s 'OFF were applied
486 at 4°C. Following centrifugation, chromatin was split into two fractions of 80 µl and 9 vol. of
487 ChIP dilution buffer added (1.1% Triton X-100, 16.7 mM Tris-HCl pH 8, 1.2 mM EDTA pH 8, 167
488 mM NaCl, 1X complete EDTA-free cocktail inhibitor) before proceeding to preclearing and IP
489 as described hereafter. For Covaris fragmentation, nuclei were resuspended with 1 ml of
490 covaris buffer (20 mM Tris pH 8, 2 mM EDTA, 0.1% SDS, 1X complete EDTA-free cocktail
491 inhibitor) for 30 min at 4°C on a turning wheel. The lysate was transferred into Covaris
492 milliTube Glass 1 ml AFA Fiber and chromatin was mainly sheared to 200-400 bp fragments
493 using S220 Focused-ultrasonicator with the following conditions: for histone marks analysis,
494 PIP175, CBP200, DF20, T 5°C, 5 min; for MAIL2-MYCtag, PIP140, CBP200, DF10, T 5°C, 6 min.
495 Sheared chromatin was centrifuged with 21000 g for 10 min at 4°C and the supernatant was
496 split into two fractions of 440µl. After adding 0.5 volume of covaris dilution buffer (20 mM
497 Tris-HCL pH 8, 2 mM EDTA, 0.1% SDS, 450 mM NaCl, 3% Triton X-100, 1X complete EDTA-free
498 cocktail inhibitor), chromatin was precleared with 30 µl of magnetic Protein A and Protein G
499 Dynabeads (Invitrogen) for 1 h at 4°C on a turning wheel. Twenty µl was saved as input and
500 the remaining supernatant was then incubated at 4°C overnight with specific antibodies (5 µg
501 of H3K27me3 Ab, Diagenode C15410069 or Millipore 07-449; 10 µl H2AUb Ab, Cell Signaling
502 Technology ref.8240; 10 µl Myc-Tag Ab, Cell Signaling Technology ref.2276 or ref.2278; 3.5 µl
503 monoclonal ANTI-FLAG M2 Ab, Sigma F1804-1MG). Next, 40 µl of magnetic Protein A and
504 Protein G Dynabeads were added and the mixture was rotated at 4°C for 3-4h. The beads
505 were washed for 5 min rotating at 4°C with low salt solution twice (0.1% SDS; 1% Triton X-

506 100, 2 mM EDTA pH 8, 20 mM Tris-HCl pH 8, 150 mM NaCl), high salt solution (0.1% SDS; 1%
507 Triton X-100, 2 mM EDTA pH 8, 20 mM Tris-HCl pH8, 500 mM NaCl), LiCl solution (0.25 M LiCl,
508 1% IGEPAL CA630, 1% deoxycholic acid, 1 mM EDTA, 10 mM Tris-HCl pH8) and TE solution (10
509 mM Tris-HCl pH 8, 1 mM EDTA pH 8). A second TE wash was performed after transferring
510 beads to a new DNA low-bind Eppendorf microtube. The chromatin was eluted with 250 µl
511 elution buffer twice (1% SDS, 0.1 M NaHCO3) at 65°C for 15 min on a thermomixer with
512 shaking (950 rpm) and subjected to reverse crosslinking by adding 51 µl of reverse crosslink
513 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
514 then overnight at 65°C. Following digestion with 10 µg RNaseA for 30 min at RT°, DNA was
515 purified using ChIP DNA clean and concentrator kit (Zymo ZD5205). For MAIL2-MYC, five
516 independent IP experiments (equivalent 10 g tissues) were pooled.
517 For immature flower buds, three tubes of about 0.15 g material were harvested and snap-
518 frozen in liquid nitrogen. Following tissue disruption using Qiagen TissueLyser, samples were
519 resuspended in 1.8 ml Hepes Buffer containing 1% formaldehyde, pooled and further
520 processed as described above with covaris fragmentation at 5°C, PIP140, CBP200, DF10 for 6
521 min. IP was then performed using 4 µg of antibody to H3K27me3 (Diagenode, C15410069).
522 Libraries were generated and sequenced as PE100 reads on a DNBSEQ platform at Beijing
523 Genomics Institute (Hong Kong).
524
525 **ChIP-qPCR.** Three tubes of about 0.15 g of aerial 12-day-old rosettes were snap frozen in
526 liquid nitrogen and processed as described above for floral buds, except that nuclei were
527 resuspended in 130 µl covaris buffer and sonicated using microtube-130 AFA Fiber. Each IP
528 was performed using 2.1 µg of H3K27me3 antibody. (Diagenode, C15410069) and one of 15
529 µl purified DNA was used for real-time PCR amplification using SYBR Premix Ex Taq II (Tli
530 RnaseH Plus) (TaKaRa) and Real-Time LightCycler 480 system (Roche) in a final reaction
531 volume of 15 µl. Primers used are listed in Supplementary Table 3.
532
533 **CUT&Tag.** About 0.15g of WT, *ndx-7*::gNDX-9xMYC or *ndx-7*::g3xFLAG-NDX immature flower
534 buds was snap frozen in liquid nitrogen, grounded to a fine powder using Qiagen TissueLyser
535 followed by a light *in vitro* cross-linking for 2 min at RT with 1.8ml ChIP extraction buffer 1
536 containing 0.2% formaldehyde. Cross-linking was stopped with 122.5µl of 2M glycine rotating
537 for 10 min at RT. Lysate was filtrated through two layers of Miracloth and centrifugated for

538 20 min at 3000 g and 4°C. Nuclei were resuspended into 1ml of ChIP extraction buffer 2 and
539 centrifugated with 12000 g at 4°C for 10 min; this step was repeated for a total of 3 washes
540 until the nuclei pellet appears totally white. The resulting pellet was resuspended in 250 µl
541 ChIP extraction buffer 2, then loaded over a cushion of 250 µl ChIP extraction buffer 3 and
542 centrifugated with 16000 g for 1 h at 4°C. Nuclei were resuspended in 1 ml Nuclear Extraction
543 (NE) buffer (20 mM Hepes pH 7.9, 10 mM KCl , 0.1% triton X-100, 20% glycerol, 0.5 mM
544 Spermidine, 1X complete EDTA-free cocktail inhibitor), incubated for 10 min on ice followed
545 with 3 min centrifugation for 600 g at 4°C. The extract was resuspended in 500 µl NE buffer
546 and centrifuged 3 min for 600 g at 4°C. The nuclei fraction was finally resuspended in 500 µl
547 NE buffer and nuclei concentration was estimated by counting using Neubauer chamber. Each
548 CUT&Tag reaction was then performed on 400 000 nuclei essentially as described in the
549 Epicypher CUTANA CUT&Tag protocol (v1.7) and using 0.5µl monoclonal Myc-Tag Ab (Cell
550 Signaling Technology ref.2276) or 0.5 µl monoclonal ANTI-FLAG M2 Ab (Sigma F1804-1MG).
551 Libraries were amplified with 18 PCR cycles and > 75 bp DNA fragments were cleaned through
552 two rounds of purification using 1.3x, then 0.8x AMPure Beads (Beckman Coulter A63880).
553 Libraries were pooled and sequenced as PE100 reads on a DNBSEQ platform at Beijing
554 Genomics Institute (Poland).
555
556 **RNA-seq analysis.** Reads were trimmed and filtered for quality and adapter contamination
557 using Trim Galore (v.0.6.7, Babraham Institute) and aligned to the *Arabidopsis* reference
558 genome (TAIR10) genome using STAR (v.2.7.9a)[39]. Reads aligning equally well to more than
559 one position in the reference genome were discarded, and probable PCR duplicates were
560 removed using MarkDuplicates from the Picard Tools suite[49]. Sample tracks were generated
561 using deeptools (v.3.5.1)[50] bamCoverage with the option normalizeUsing CPM and binSize 1.
562 Subsequent read counting for each gene and TE was performed using featureCounts
563 (v.2.0.3)[51] using the Araport11 genome annotation. Annotated TEs overlapping strongly (>
564 80%) with an annotated TE gene were considered TE genes, and the TE annotation was
565 discarded. Differential expression analysis was performed using DESeq2 (v.1.36.0)[52] with
566 absolute fold-change ≥ 4 and Benjamini-Hochberg adjusted *p*-value < 0.05 were considered
567 differentially expressed. For DEseq2 analysis of amiR-*mail2* hypomorphic lines, criteria of
568 absolute fold change ≥2 and Benjamini-Hochberg adjusted *p*-value < 0.05 were retained. All
569 plots were generated using ggplot2 (v.3.4.4).

570

571 **ChIP-seq analysis.** Reads were trimmed and filtered using Trim Galore (v.0.6.7, Babraham
572 Institute) or Trimmomatic (v.0.39)[53], and aligned to the TAIR10 *Arabidopsis* reference genome
573 using bowtie2 (v.2.4.2)[54], allowing only uniquely mapped reads with perfect matches. PCR
574 duplicates were discarded using MarkDuplicates from the Picard Tools suite or markdup from
575 Sambamba (v.0.8.2)[55]. Sample tracks were generated using deeptools (v.3.5.1)[50]
576 bamCoverage with the option normalizeUsing RPGC and binSize 1. Peaks calling was
577 performed with MACS2 (v.2.2.7.1)[56] using corresponding inputs as a background and --
578 nomodel and –extsize 200 options or MACS3 (v.3.0.0b1)[56] using --model and –call-summits,
579 q-value < 0.01, and peaks called in the anti-MYC or anti-FLAG WT controls were removed.
580 MAIL1/MAIN bound regions are designed as the summit of peaks plus 100bp on each side.
581

582 **CUT&Tag data processing.** Raw paired-end reads were trimmed using Trim Galore (v.0.6.7)
583 with default parameters. Trimmed reads were aligned to the TAIR10 reference genome using
584 Bowtie2 (v.2.4.2 in --very-sensitive-local mode, with insert size constraints (-I 10 -X 700), and
585 with --no-mixed and --no-discordant options to restrict alignments to properly paired reads.
586 Reads mapping to organellar genomes (chloroplast or mitochondria) were filtered out, and
587 only uniquely mapping reads with a mapping quality ≥ 10 were retained. BAM files were then
588 sorted and indexed and insert size distributions were computed using Picard's
589 *CollectInsertSizeMetrics*. For sample normalization, exogenous *E. coli* DNA carried by the pA-
590 Tn5 transposase was used as a spike-in control. Clean reads were aligned to the *E. coli* genome
591 using Bowtie2, and a scaling factor was computed for each sample as the ratio between the
592 number of *E. coli* reads in the sample and the smallest *E. coli* read count observed among all
593 samples. These scaling factors were used to normalize read coverage when generating signal
594 tracks (bigWig format) using *bamCoverage* of deepTools (v.3.5.1), with a bin size of 5 bp.
595

596 **Motif discovery.** DNA sequences of MAIL1/MAIN and MAIL2 ChIP-seq peaks and of gene
597 proximal promoter regions (+/- 500 pb surrounding TSS) were extracted using bedtools
598 (v.2.31.1) getfasta. Conserved motifs were identified using STREME v5.5.5[57] and the shuffled
599 input sequences option with the following parameters: streme --verbosity 1 --oc streme_out
600 -dna --minw 6 --maxw 16 --order 2 --bfile ./background --seed 0 --align center --time 4740 --
601 totallength 4000000 --evalue --thresh 0.05, or using MEME v5.5.5[58] with the Classic mode

602 option and the following parameters: meme -oc meme_out -objfun classic -mod anr -csites
603 10000 -nmotifs 10 -minw 6 -maxw 15 -markov_order 0 -dna -revcomp -evt 0.05. Differentially
604 enriched DNA motifs were recovered using the MEME DE mode option and an equivalent
605 control set; relative enrichment ratio was evaluated with the SEA program of the MEME suite.
606 Genomic distribution of M1M and M2M identified from ChIP peaks was determined using
607 FIMO v5.5.5[59] with *P* value match < 1e-05 and default parameters. Genomic localization of
608 telobox elements was determined using FIMO v5.5.5[59] with p-value match < 1e-05 and default
609 parameters and the consensus sequence "AAACCCTAR".
610
611 **Gene ontology analysis.** GO term enrichment analysis was performed using ShinyGO 0.82[60]
612 and the *Arabidopsis thaliana* Araport11 annotation. *P*-value was derived from
613 hypergeometric distribution, adjusted using the FDR (false discovery rate) method. Results
614 with a *P*-value < 1e-05 were considered as significant.
615
616 **Protein structure and DNA-binding potential prediction.** The structures of the PMD-C
617 domains of MAIL1, MAIN and MAIL2 were modeled using AlphaFold 3 [61]. Potential DNA-
618 binding residues of MAIN, MAIL1 and MAIL2 were predicted using the geometry-aware
619 binding site predictor GPSite [62]. For each PMD-C model, prediction scores were mapped onto
620 the corresponding residues for visualization using UCSF ChimeraX v. 1.9 [63].
621
622 **Phylogenetic analysis.** Candidate MuDR-encoded PMD sequences were identified through
623 tBlastN analysis of Repbase database[64] using the MAIL2 PMD-C region as query
624 (Supplementary doc. 1). The "PhyML+SMS/OneClick" pipeline[65] (https://ngphylogeny.fr/)
625 with default parameters was then used to build a phylogenetic tree based on MuDR-
626 associated PMD and including Arabidopsis gene-encoded PMD-B and -C sequences. Resulting
627 tree was exported using the online Interactive Tree Of Life (ITOL) tool. PMD sequences used
628 to build the tree are reported in Supplementary Table 4.
629
630 **Data availability.** The data supporting the findings of this study are available within the article
631 and its Supplementary Information. High throughput sequencing data has been deposited in
632 the Gene Expression Omnibus (GEO) database and can be accessed with the accession

633 number GSE278560 (accessible using the token *ktghyqoafluttcz* during the review process).

634 All data are available from the corresponding author upon reasonable request.

635

636 **References**

637 1. Zhou, Y., Romero-Campero, F. J., Gómez-Zambrano, Á., Turck, F. & Calonje, M.
638 H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and
639 PRC2 activity. *Genome Biol.* **18**, 13–69 (2017).
640 2. Chen, L.-Q. *et al.* ATX3, ATX4, and ATX5 Encode Putative H3K4
641 Methyltransferases and Are Critical for Plant Development. *Plant Physiol.* **174**, 1795–1806
642 (2017).
643 3. Alvarez-Venegas, R. *et al.* ATX-1, an Arabidopsis homolog of trithorax, activates
644 flower homeotic genes. *Curr. Biol. CB* **13**, 627–637 (2003).
645 4. Tamada, Y., Yun, J.-Y., Woo, S. chul & Amasino, R. M. ARABIDOPSIS
646 TRITHORAX-RELATED7 Is Required for Methylation of Lysine 4 of Histone H3 and for
647 Transcriptional Activation of FLOWERING LOCUS C. *Plant Cell* **21**, 3257–3269 (2009).
648 5. Berr, A. *et al.* Arabidopsis SET DOMAIN GROUP2 Is Required for H3K4
649 Trimethylation and Is Crucial for Both Sporophyte and Gametophyte Development. *Plant*
650 *Cell* **22**, 3232–3248 (2010).
651 6. Lu, F., Cui, X., Zhang, S., Jenuwein, T. & Cao, X. Arabidopsis REF6 is a histone H3
652 lysine 27 demethylase. *Nat. Publ. Group* **43**, 715–719 (2011).
653 7. Cui, X. *et al.* REF6 recognizes a specific DNA sequence to demethylate H3K27me3
654 and regulate organ boundary formation in Arabidopsis. *Nat. Publ. Group* (2016).
655 8. Yan, W. *et al.* Dynamic and spatial restriction of Polycomb activity by plant histone
656 demethylases. *Nat. Plants* **4**, 681–689 (2018).
657 9. Antunez-Sanchez, J. *et al.* A new role for histone demethylases in the maintenance of
658 plant genome integrity. *eLife* **9**, (2020).
659 10. Liang, S. C. *et al.* Kicking against the PRCs – A Domesticated Transposase
660 Antagonises Silencing Mediated by Polycomb Group Proteins and Is an Accessory
661 Component of Polycomb Repressive Complex 2. *PLoS Genet.* **11**, e1005660-26 (2015).
662 11. Velanis, C. N. *et al.* The domesticated transposase ALP2 mediates formation of a
663 novel Polycomb protein complex by direct interaction with MSI1, a core subunit of
664 Polycomb Repressive Complex 2 (PRC2). *PLoS Genet.* **16**, e1008681 (2020).
665 12. Ikeda, Y. *et al.* Arabidopsis proteins with a transposon-related domain act in gene
666 silencing. *Nat. Commun.* **8**, 1–10 (2017).
667 13. Nicolau, M. *et al.* The plant mobile domain proteins MAIN and MAIL1 interact with
668 the phosphatase PP7L to regulate gene expression and silence transposable elements in
669 Arabidopsis thaliana. *PLoS Genet.* **16**, e1008324 (2020).
670 14. Luxán Hernández, C. *et al.* PP7L is essential for MAIL1-mediated transposable
671 element silencing and primary root growth. *Plant J.* **8**, 15115–15122 (2020).
672 15. Jarry, L. *et al.* Plant mobile domain proteins ensure Microrchidia 1 expression to
673 fulfill transposon silencing. *Life Sci. Alliance* **6**, e202201539 (2023).
674 16. Morel, J. B., Mourrain, P., Béclin, C. & Vaucheret, H. DNA methylation and
675 chromatin structure affect transcriptional and post-transcriptional transgene silencing in
676 Arabidopsis. *Curr. Biol. CB* **10**, 1591–1594 (2000).
677 17. Zhu, Y. *et al.* The Arabidopsis noduliN homeobox factor AtNDX interacts with
678 AtRiNG1a/b and negatively regulates abscisic acid signaling[Open]. *Plant Cell* **32**, 703–721
679 (2020).

680 18. Li, J., Wang, Z., Hu, Y., Cao, Y. & Ma, L. Polycomb Group Proteins RING1A and
681 RING1B Regulate the Vegetative Phase Transition in Arabidopsis. *Front. Plant Sci.* **8**, 867
682 (2017).
683 19. Mikulski, P. *et al.* VAL1 acts as an assembly platform co-ordinating co-
684 transcriptional repression and chromatin regulation at Arabidopsis FLC. *Nat. Commun.* **13**,
685 5542 (2022).
686 20. Baile, F., Merini, W., Hidalgo, I. & Calonje, M. EAR domain-containing transcription
687 factors trigger PRC2-mediated chromatin marking in Arabidopsis. *Plant Cell* **33**, 2701–2715
688 (2021).
689 21. Yuan, L. *et al.* The transcriptional repressors VAL1 and VAL2 recruit PRC2 for
690 genome-wide Polycomb silencing in Arabidopsis. *Nucleic Acids Res.* **49**, 98–113 (2021).
691 22. Shu, J. *et al.* Genome-wide occupancy of histone H3K27 methyltransferases CURLY
692 LEAF and SWINGER in Arabidopsis seedlings. *Plant Direct* **3**, e00100 (2019).
693 23. Moissiard, G. *et al.* MORC family ATPases required for heterochromatin
694 condensation and gene silencing. *Science* **336**, 1448–1451 (2012).
695 24. Moissiard, G. *et al.* Transcriptional gene silencing by Arabidopsis microrchidia
696 homologues involves the formation of heteromers. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7474–
697 7479 (2014).
698 25. Stacey, N. J. *et al.* Arabidopsis SPO11-2 functions with SPO11-1 in meiotic
699 recombination. *Plant J.* **48**, 206–216 (2006).
700 26. Xue, M. *et al.* The Number of Meiotic Double-Strand Breaks Influences Crossover
701 Distribution in Arabidopsis. *Plant Cell* **30**, 2628–2638 (2018).
702 27. Ühlken, C., Horvath, B., Stadler, R., Sauer, N. & Weingartner, M. MAIN-LIKE1 is a
703 crucial factor for correct cell division and differentiation in Arabidopsis thaliana. *Plant J.* **78**,
704 107–120 (2014).
705 28. Xiao, J. *et al.* Cis and trans determinants of epigenetic silencing by Polycomb
706 repressive complex 2 in Arabidopsis. *Nat. Publ. Group* **49**, 1546–1552 (2017).
707 29. Zhou, Y. *et al.* Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via
708 TRB factors in Arabidopsis. *Nat. Publ. Group* **50**, 638 (2018).
709 30. Wang, M. *et al.* Arabidopsis TRB proteins function in H3K4me3 demethylation by
710 recruiting JMJ14. *Nat. Commun.* **14**, 1736 (2023).
711 31. Karányi, Z. *et al.* NODULIN HOMEOBOX is required for heterochromatin
712 homeostasis in Arabidopsis. *Nat. Commun.* **13**, 5058 (2022).
713 32. Mukherjee, K., Brocchieri, L. & Bürglin, T. R. A comprehensive classification and
714 evolutionary analysis of plant homeobox genes. *Mol. Biol. Evol.* **26**, 2775–2794 (2009).
715 33. Christensen, R. G. *et al.* Recognition models to predict DNA-binding specificities of
716 homeodomain proteins. *Bioinformatics* **28**, i84–i89 (2012).
717 34. Su, X.-M. *et al.* ALFIN-like proteins link histone H3K4me3 to H2A ubiquitination
718 and coordinate diverse chromatin modifications in Arabidopsis. *Mol. Plant* **18**, 130–150
719 (2025).
720 35. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J. & Dean, C. R-Loop
721 Stabilization Represses Antisense Transcription at the Arabidopsis FLC Locus. *Science* **340**,
722 619–621 (2013).
723 36. Deleris, A. *et al.* Loss of the DNA Methyltransferase MET1 Induces H3K9
724 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to
725 Transposons in Arabidopsis thaliana. *PLoS Genet.* **8**, e1003062-13 (2012).
726 37. Rougée, M. *et al.* Polycomb mutant partially suppresses DNA hypomethylation-
727 associated phenotypes in Arabidopsis. *Life Sci. Alliance* **4**, (2021).
728 38. Barbi, M. & Paillusson, F. Chapter Seven - Protein–DNA Electrostatics: Toward a
729 New Paradigm for Protein Sliding. in *Advances in Protein Chemistry and Structural Biology*

(ed. Karabencheva-Christova, T.) vol. 92 253–297 (Academic Press, 2013).
39. Borredá, C., Leduque, B., Colot, V. & Quadrana, L. Transposable element products, functions, and regulatory networks in Arabidopsis. *bioRxiv* 2024.04.02.587720 (2024) doi:10.1101/2024.04.02.587720.
40. Hosaka, A. *et al.* Evolution of sequence-specific anti-silencing systems in Arabidopsis. *Nat. Commun.* **8**, 2161 (2017).
41. 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).
42. Sasaki, T. *et al.* Arms race between anti-silencing and RdDM in noncoding regions of transposable elements. *EMBO Rep.* **n/a**, e56678 (2023).
43. 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).
44. Hure, V. *et al.* Alternative silencing states of transposable elements in Arabidopsis associated with H3K27me3. *Genome Biol.* **26**, 11 (2025).
45. 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).
46. Xu, L. & Shen, W.-H. Polycomb Silencing of KNOX Genes Confines Shoot Stem Cell Niches in Arabidopsis. *Curr. Biol.* **18**, 1966–1971 (2008).
47. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for -mediated transformation of. *Plant J.* **16**, 735–743 (1998).
48. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
49. Broad Institute. Picard Toolkit. Broad Institute, GitHub repository http:// broadinstitute.github.io/picard/. (2019).
50. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
51. 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).
52. 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).
53. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
54. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
55. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
56. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, 1–9 (2008).
57. Bailey, T. L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834–2840 (2021).
58. 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).
59. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinforma. Oxf. Engl.* **27**, 1017–1018 (2011).
60. 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).
61. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
62. Yuan, Q., Tian, C. & Yang, Y. Genome-scale annotation of protein binding sites via

language model and geometric deep learning. *eLife* **13**, (2024).
63. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci. Publ. Protein Soc.* **32**, e4792 (2023).
64. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
65. Lemoine, F. *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* **47**, W260–W265 (2019).

Acknowledgments

We thank F. Turck (Max Planck Institute for Plant Breeding Research) for kindly providing the *SEP3pro:GUS* construct. We are also grateful to S. Marquardt (Department of Plant and Environmental Sciences, University of Copenhagen) for his help in setting up the CUT&Tag assay. 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 grants from the Agence Nationale de la Recherche (ANR-20-CE12-0009 & ANR-23-CE20-0012 to O.M.). Work in the Moissiard laboratory was supported by CNRS and University of Perpignan Via Domitia (UPVD) core funding, and by grants from the Agence Nationale de la Recherche (ANR-23-CE20-0012 to G.M). This study was also supported by "Laboratoires d'Excellences" (LAbEX) AGRO 2011-LABX-002 (under I-Site Muse framework) coordinated by Agropolis Foundation (ID 2101-009 to G.M), region Occitanie PhD grant to L.J, and by the "Laboratoires d'Excellences (LABEX)" TULIP (ANR-10-LABX-0041)" and "École Universitaire de Recherche (EUR)" TULIP-GS (ANR-18-EURE-0019).

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.

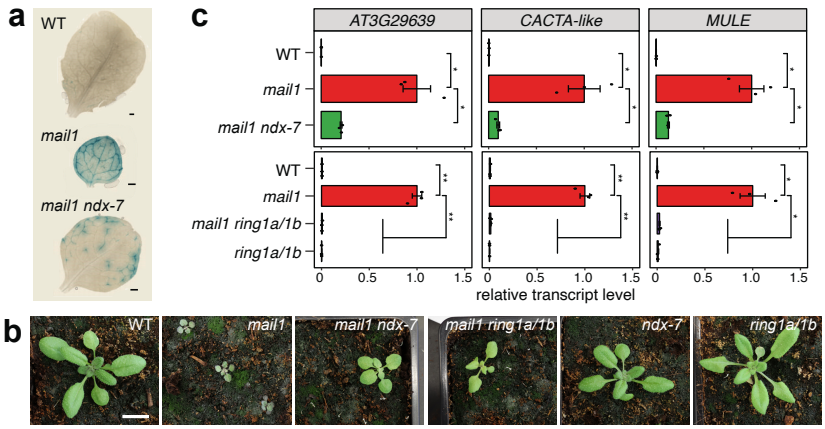

Fig. 1. NDX and PRC1 RING1A/1B components suppress *mail1*-induced defects. (**a**) Histochemical staining for GUS activity in leaves from plants of the indicated genotypes. Scale bar is 500 µm. (**b**) Developmental phenotype of three-week-old WT, *mail1*, *mail1 ndx-7*, *mail1 ring1a/1b*, *ndx-7* and *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, * $P < 0.05$; ** $P < 0.01$).

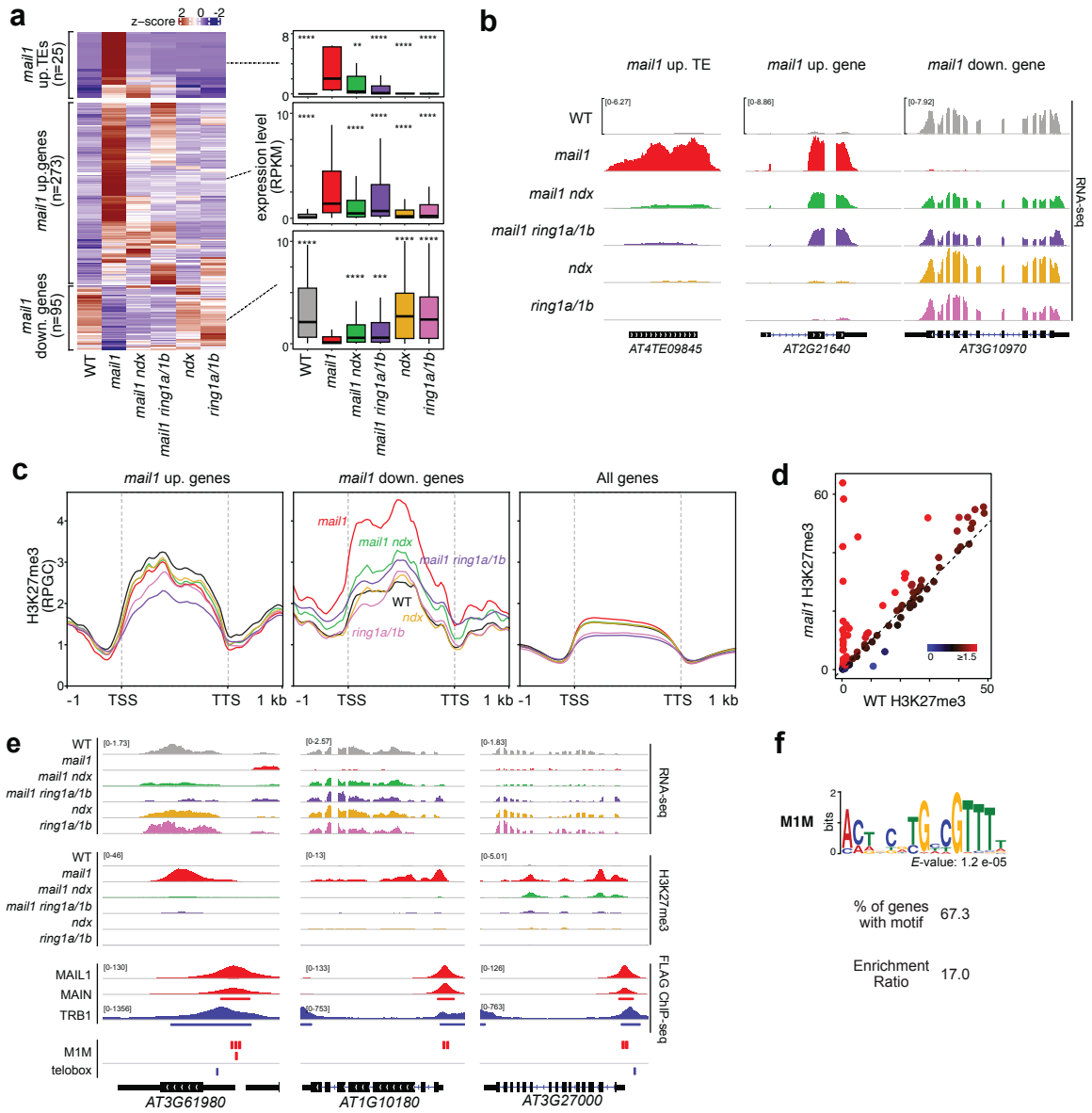

Fig. 2. *mail1*-associated TE and gene misregulation is dependent on NDX or RING1A/1B function. (**a**) Heatmaps of RNA-seq expression z-score computed for *mail1* upregulated TEs and misregulated genes across the indicated genotypes (left). Means of two biological replicates were computed. Boxplots with the same data expressed in RPKM are shown on the right. Center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. 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**) MEME prediction of a differentially enriched DNA motif within the 1kb region surrounding TSS of H3K27me3-enriched genes in *mail1*. A control promoter set was randomly selected from genes with similar H3K27me3 and expression patterns in wild type plants (ranging from 0 to 3 RPKM and 0.1 to 15 RPKM, respectively). Percentage of regions exhibiting the M1M motif and relative enrichment ratio are indicated.

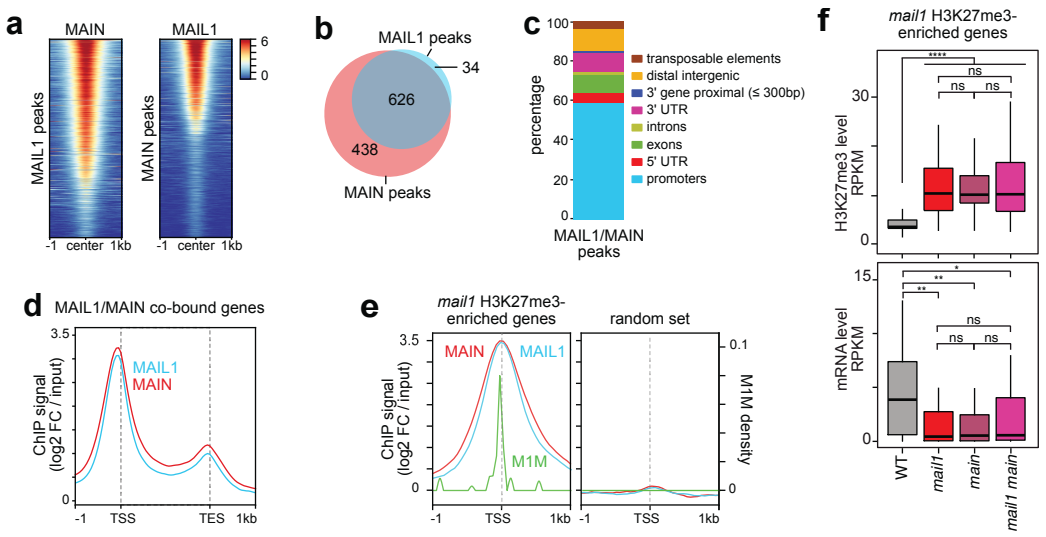

Fig. 3. MAIL1 and MAIN are mainly co-targeted to M1M at gene proximal promoters. (**a**) Heatmaps showing MAIN and MAIL1 ChIPseq signals (Log2 FC / input) over MAIL1 (n = 660) and MAIN (n = 1064) peaks respectively. Peaks were aligned at their center position. (**b**) Overlap between MAIL1 and MAIN peaks. (**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 overlapping 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 (n=49; 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. Center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. The effect of genotype was verified with a Wilcoxon rank sum test with continuity correction; two-sided (ns. not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$).

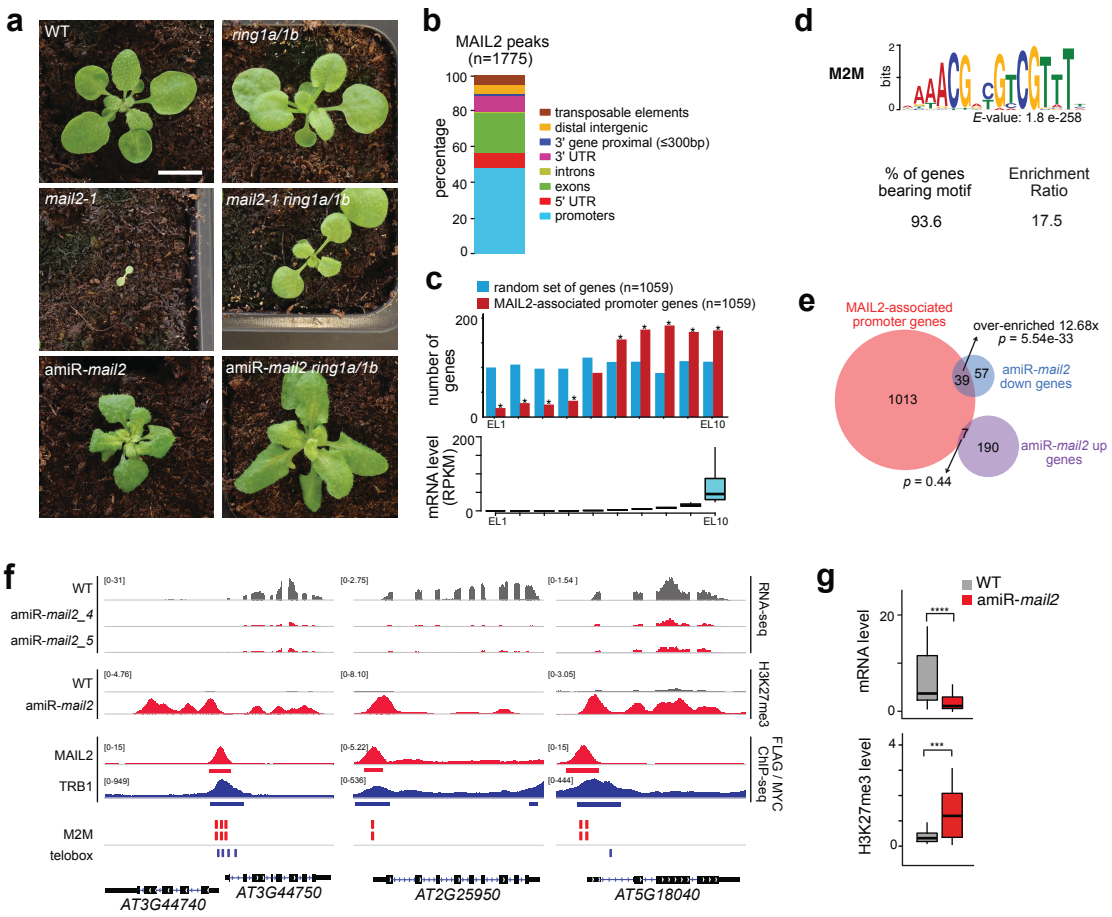

Fig. 4. MAIL2 mainly associates to gene promoters and protects numerous actively transcribed genes from H3K27me3-mediated PcG silencing (**a**) Photos of three-week-old seedlings of indicated genotype. Scale bar is 1 cm. (**b**) Genomic features associated with MAIL2 peaks. (**c**) Distribution of MAIL2-associated promoter genes in relation to their expression level in WT plants. A random set of the same number of genes is shown for comparison. Asterisks mark statistically significant differences from an equal representation in each decile (Fisher's exact test, *P*-value < 1e-04). All Arabidopsis genes were split in ten deciles based on their WT expression level (RPKM; bottom diagram). Center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. (**d**) Identification of a highly enriched motif within the 1kb region surrounding TSS of genes harboring MAIL2 peak at proximal promoter (n = 1059). Analysis was carried out using MEME in differential enrichment mode and an equivalent control set randomly selected among genes ranging in the five deciles with the highest expression level in the WT. Percentage of regions exhibiting the M2M motif and relative enrichment ratio are indicated. (**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) in hypomorphic amiR-*mail2* background. Center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. The effect of genotype was verified with a Wilcoxon rank sum test with continuity correction; two-sided (ns. not significant; *** $P < 0.001$; **** $P < 0.0001$).

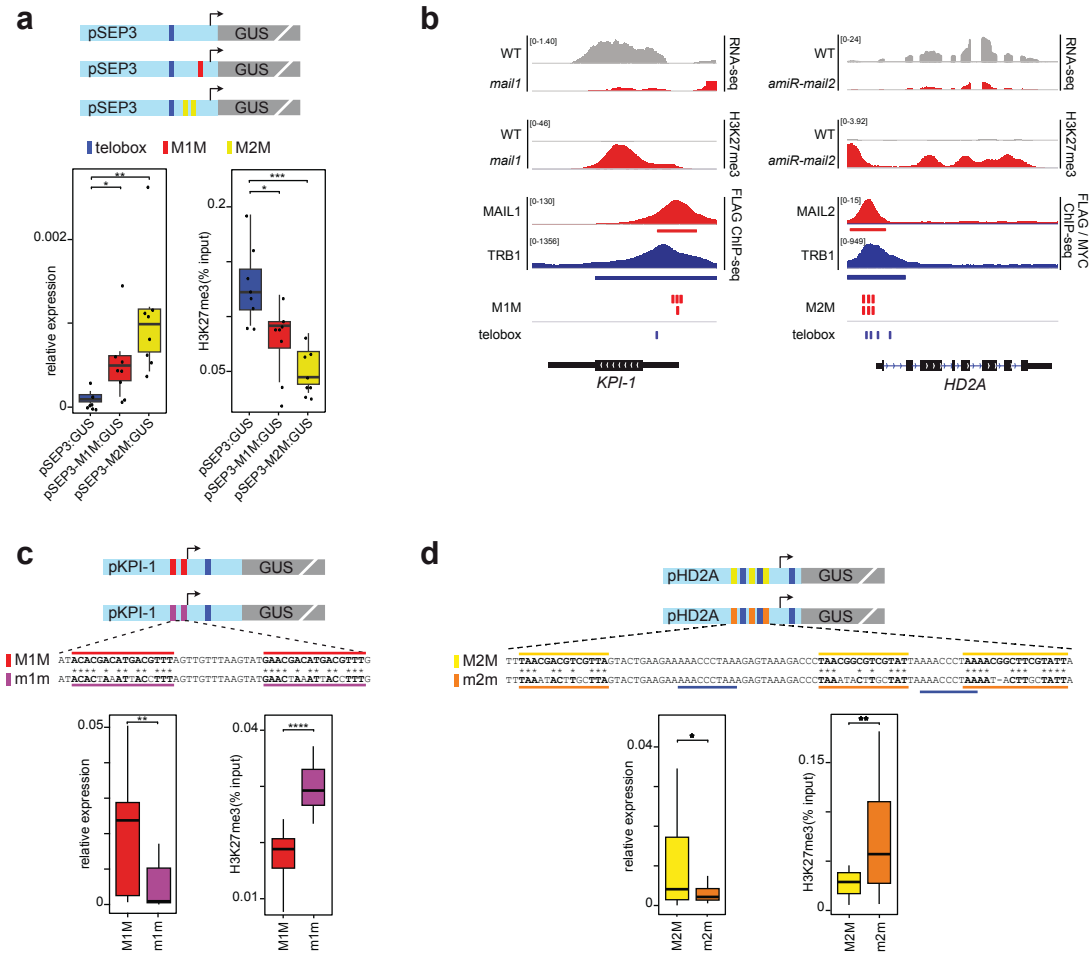

Fig. 5. Promoter-associated M1M and M2M are functionally relevant to safeguard genes from PcG-mediated silencing. (a) Schematic representation of the *proSEP3:GUS* original transgene and derivatives following insertion of one M1 or two M2 motifs. Box plots showing relative GUS expression and H3K27me3 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 *KPI-1* and *HD2A* loci in the indicated genotypes. MAIL1 and MAIL2-associated patterns are shown together with TRB1-FLAG data (GEO-GSM6895939

909 retrieved from[30]) and predicted M1M, M2M and telobox motifs. (**c**) Schematic representation
910 of *pPKL1:GUS* transgenes. Most conserved M1M motifs were mutated (m1m) and sequence
911 modifications are shown under the scheme. Telobox motif is also reported in blue. Relative
912 GUS expression and H3K27me3 accumulation were performed as described in (a) except that
913 nine independent lines were analysed for each transgene. (**d**) Schematic representation of
914 *pHD2A:GUS* transgenes showing localization of intact (M2M) and mutated (m2m) motifs.
915 Telobox motifs are indicated in blue and sequence modifications are reported under the
916 scheme. Molecular analyses were performed as described in (a) except that nine independent
917 lines were analysed for each transgene. In box plots, center lines show the medians; box limits
918 indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from
919 the 25th and 75th percentiles.

920 **Supplementary Table 1. List of H3K27me3-enriched genes (Log2 FC > 1; *P*-value < 0.01) in**
921 ***mail1* compared to WT background.**
922
923 **Supplementary Table 2. Lists of differentially expressed loci in amiR-*mail2*_4 and _5 mutant**
924 **lines.**
925
926 **Supplementary Table 3. List of qPCR primers used in this study.**
927
928 **Supplementary Table 4. List of PMD sequences used for phylogenetic analyses.**
929
930 **Supplementary Table 5. DNA-binding prediction scores for amino acid residues of MAIN,**
931 **MAIL1 and MAIL2.**
932
933 **Supplementary doc. 1. Best results of TBLASTN search on Repbase[64] using MAIL2 PMD-C**
934 **domain as query.** Only results with e-values < 1e-10 were reported. Dicot species: ArHy
935 *Arachis hypogaea*, AvMa *Avicennia marina*, Cba *Capsicum baccatum*, Daca *Daucus carota*, Gar
936 *Gossypium arboretum*, GR *Gossypium raimondii*, GM *Glycine max*, Ipa *Ilex paraguariensis*, LU
937 *Linum usitatissimum*, MT *Medicago truncatula*, NS *Nicotiana sylvestris*, SeTo *Senna tora*, STu
938 *Solanum tuberosum*, VV *Vitis Vinifera*. Monocot species: DRot *Dioscorea cayennensis subsp.*
939 *Rotundata*, OS *oryza Sativa*, SBi *Sorghum bicolor*, SHS *Saccharum hybrid cultivar*, Sit *Setaria*
940 *italica*, Tae *Triticum aestivum*, ZM *Zea mays*. Basal angiosperm species: ATr *Amborella*
941 *trichopoda*.