



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Mobile small RNAs regulate genome-wide DNA methylation

Citation for published version:

Lewsey, MG, Hardcastle, TJ, Melnyk, CW, Molnar, A, Valli, A, Urich, MA, Nery, JR, Baulcombe, DC & Ecker, JR 2016, 'Mobile small RNAs regulate genome-wide DNA methylation', *Proceedings of the National Academy of Sciences (PNAS)*, vol. 113, no. 6, pp. E801-810. <https://doi.org/10.1073/pnas.1515072113>

Digital Object Identifier (DOI):

[10.1073/pnas.1515072113](https://doi.org/10.1073/pnas.1515072113)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Proceedings of the National Academy of Sciences (PNAS)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



1 **BIOLOGICAL SCIENCES: Plant Biology**

2
3
4 **Mobile small RNAs regulate genome-wide DNA methylation**

5
6 Mathew G. Lewsey^{1,2,3}, Thomas J. Hardcastle^{3,4}, Charles W. Melnyk^{4,5}, Attila Molnar^{4,6}, Adrián
7 Valli⁴, Mark A. Urich¹, Joseph R. Nery¹, David C. Baulcombe^{4,7*} and Joseph R. Ecker^{1,2,7,8*}

8
9 ¹Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, California
10 92037, USA

11 ²Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037,
12 USA

13 ³Co-first author

14 ⁴Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA,
15 UK

16 ⁵Present address: Sainsbury Laboratory, University of Cambridge, Bateman Street, Cambridge
17 CB2 1LR, UK

18 ⁶Present address: School of Biological Sciences, The University of Edinburgh, Institute of
19 Molecular Plant Sciences, Mayfield Road, Edinburgh EH9 3JR, UK

20 ⁷Co-corresponding author

21 ⁸Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California
22 92037, USA

23
24 *Correspondence: ecker@salk.edu (J.R.E.), dcb40@cam.ac.uk (D.C.B.)

25
26 **Keywords: RNA-directed DNA methylation, plant grafting, transposable element, small**
27 **RNA, transcriptional gene silencing**

30 **Abstract**

31

32 RNA silencing at the transcriptional and post-transcriptional levels regulates endogenous gene
33 expression, controls invading transposable elements (TEs) and protects the cell against viruses.
34 Key components of the mechanism are small RNAs (sRNAs) of 21-24 nucleotides (nt) that guide
35 the silencing machinery to their nucleic acid targets in a nucleotide sequence-specific manner.
36 Transcriptional gene silencing is associated with 24 nt sRNAs and RNA-directed DNA
37 methylation (RdDM) at cytosine residues in three DNA sequence contexts (CG, CHG and CHH).
38 We previously demonstrated that 24 nt sRNAs are mobile from shoot to root in *Arabidopsis*
39 *thaliana* and confirmed that they mediate DNA methylation at three sites in recipient cells. In this
40 study we extend that finding by demonstrating that RdDM of thousands of loci in root tissues is
41 dependent upon mobile sRNAs from the shoot and that mobile sRNA-dependent DNA
42 methylation occurs predominantly in non-CG contexts. Mobile sRNA-dependent non-CG
43 methylation is largely dependent on the DRM1/DRM2 RdDM pathway but independent of the
44 CMT2/3 DNA methyltransferases. Specific superfamilies of TEs, including those typically found in
45 gene-rich euchromatic regions, lose DNA methylation in a mutant lacking 22 to 24 nt sRNAs
46 (*dicer-like 2, 3, 4* triple mutant). Transcriptome analyses identified a small number of genes
47 whose expression in roots is associated with mobile sRNAs and connected to DNA methylation
48 directly or indirectly. Finally, we demonstrate that sRNAs from shoots of one accession move
49 across a graft union and target DNA methylation *de novo* at normally unmethylated sites in the
50 genomes of root cells from a different accession.

51

52 **Significance statement**

53 Small RNAs (sRNAs) of 24 nucleotides are associated with transcriptional gene silencing by
54 targeting DNA methylation to complementary sequences. We demonstrated previously sRNAs
55 move from shoot to root, where they regulate DNA methylation of three endogenous transposable
56 elements (TEs). However, the full extent of root DNA methylation dependent on mobile sRNAs
57 was unknown. We demonstrate that DNA methylation at thousands of sites depends upon mobile
58 sRNAs. These sites are associated with TE superfamilies found in gene rich regions of the
59 genome, which lose methylation selectively in an sRNA-deficient mutant. If the TEs were able to
60 reactivate they could cause genome instability and altered gene expression patterns, with
61 negative effects on the plant. Consequently, mobile sRNAs may defend against these TEs.

62

63

64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118

\body

Introduction

RNA silencing in plants and animals is a process that controls gene expression at both the transcriptional and post-transcriptional levels (1). In plants small non-coding RNAs (sRNAs), 21 to 24 nucleotides (nt) in length (2, 3), direct the RNA silencing machinery to target nucleic acids in a sequence-specific manner (4). The 21/22 nt sRNAs are primarily associated with mRNA cleavage and are involved in post-transcriptional gene silencing (PTGS) (3). The 24 nt sRNAs are primarily associated with RNA-directed DNA methylation (RdDM) and transcriptional gene silencing (2, 5). However, a recent study suggests that both 21 and 24 nt sRNAs are involved in deposition of DNA methylation (6). It is proposed that the 21 nt sRNAs may establish DNA methylation, whereas the 24 nt species are involved in its amplification and maintenance (7, 8).

RdDM involves methylation of cytosine residues in CG, CHG and CHH sequence contexts (where H denotes any base except G) (9, 10). It is closely associated with repressive chromatin marks at target loci (11) and blocks gene transcription when present in promoter regions (12). RdDM maintains genome integrity by repression of transposable element (TE) activity, as well as contributing to environmental and developmental regulation of gene expression (4, 8, 13-17). The methylation status of DNA is heritable through both meiosis and mitosis, allowing it to persistently alter gene expression (18-20).

Initial establishment of RdDM involves cleavage of double-stranded RNA by DICER-LIKE (DCL) proteins to form 21 nt to 24 nt sRNAs, which load into ARGONAUTE (AGO) proteins (4, 21). These nucleoprotein complexes target chromatin-associated scaffold transcripts in a sequence-specific manner (22-24). The chromatin-bound complexes then recruit DOMAINS REARRANGED METHYLTRANSFERASES 1 and 2 (DRM1 and DRM2) that methylate DNA in CG, CHG and CHH sequence contexts (8, 25). A complex set of maintenance mechanisms ensures persistence of established DNA methylation through cell division and even between generations. Most of these mechanisms are independent of RNA and involve epigenetic histone marks. The VARIANT IN METHYLATION (VIM) family proteins 1, 2 and 3 and DNA METHYLTRANSFERASE 1 (MET1) maintain efficiently CG context methylation, resulting in near complete methylation of target sequences (26-28). Non-CG context methylation (i.e. CHG, CHH) is maintained by a self-reinforcing loop involving KRYPTONITE family enzymes (29). These proteins recognize non-CG context methylated DNA and methylate lysine 9 of adjacent histone H3 (H3K9me2). CHROMOMETHYLASE (CMT) proteins CMT2 and CMT3 bind H3K9me2 and methylate adjacent non-CG sites (30) of the newly replicated DNA. Redundancy exists between target sites of CMT2 and CMT3, but their predominant functions are to maintain CHH and CHG context methylation respectively (31). The activity of CMT2 is substantially less efficient than that of MET1 so that CMT2 target sites exhibit variable levels of DNA methylation. CMT3 efficiency is intermediate between MET1 and CMT3. Both CMT2 and CMT3 typically target long TEs and gene distal TEs (31).

There are also RNA-dependent mechanisms to maintain DNA methylation that involve plant-specific DNA-dependent RNA polymerases IV and V (POL IV and POL V). These polymerases are recruited to chromatin by methyl-DNA-binding proteins SHH1 and SUVH2 and 9 (32). POL IV produces precursor RNAs that are processed into 24 nt sRNAs, whereas POL V produces chromatin bound scaffold transcripts at sites of DNA methylation (33). Together, they ensure maintenance of CG, CHG and CHH context DNA methylation through an AGO-dependent mechanism similar to the mechanism that establishes methylation, in which an AGO-sRNA recruits DRM1 and DRM2 to maintain non-CG DNA methylation (8, 25). The target sites of the DRM1/DRM2 DNA methylation maintenance pathway are largely non-overlapping with those of CMT2 and CMT3, and tend to be short, gene proximal TEs and the edges of long TEs (31, 34).

RdDM can operate cell-to-cell and systemically due to translocation of 23-24 nt sRNAs from shoots to roots (35). In our previous studies we confirmed that these mobile 23-24 nt sRNAs

119 target RdDM and TGS at one transgene and RdDM at three endogenous TEs (35-37). Depletion
120 of shoot sRNAs corresponded with reduction of 23-24 nt sRNAs in wild-type roots, indicating
121 shoot-derived sRNAs contribute to the total root sRNA population (36). In this study we
122 investigate the extent to which mobile sRNAs mediate genome-wide RdDM. Our approach, as
123 before, is to analyze sRNA and DNA methylation in roots of grafted plants that are defective for
124 the production of the 24 nt sRNA species associated with RdDM. The sRNAs in these grafted
125 plants move predominantly from shoot to root following source-sink gradients (36) and, by
126 grafting different genotypes as shoots, we identify changes in DNA methylation and gene
127 expression in the roots that are dependent on mobile sRNAs.

128
129 We show that mobile sRNAs influence genomic DNA methylation at thousands of loci, and that
130 the affected loci are predominantly associated with transposons of specific classes. A very small
131 number of protein coding genes were influenced by this mobile RdDM. The mobile sRNA-
132 dependent DNA methylation is associated with the DRM1/DRM2 RdDM pathway but not the
133 CMT2/3 DNA methyltransferase pathway. Furthermore, we demonstrate that mobile sRNAs
134 unique to one accession established DNA methylation *de novo* in unmethylated regions of the
135 genome of a second accession.

136

137 Results

138

139 Identification of DNA methylation loci targeted by mobile sRNAs

140

141 The primary aims of our study were to determine how many genomic loci in *Arabidopsis* may be
142 targeted by RdDM due to the direct action of mobile sRNAs (Fig. 1) and to compare these loci
143 with other sRNA targeted and cytosine methylated regions throughout the genome. We reasoned
144 that mobile sRNAs that direct RdDM may be associated with specific features of the genome and
145 depend upon specific genes in the RdDM pathway. In addition, we aimed to determine how many
146 of these directly-affected loci might influence gene expression. It could be, for example, that DNA
147 methylation, leading to TGS, of a transposon target of mobile sRNA affects expression of an
148 adjacent gene.

149

150 In previous work we compared sRNA populations in roots of WT *Arabidopsis thaliana* (accessions
151 Col-0, termed Col, and C24) and a *dicer-like 2 dicer-like 3 dicer-like 4* triple mutant (Col
152 background, termed *dcl234*) that had been grafted to C24, Col and *dcl234* shoots in order to
153 identify loci producing mobile sRNAs in the *Arabidopsis* genome. The *dcl234* triple mutant is
154 unable to produce 22 to 24 nt sRNAs associated with RNA silencing (23). We reasoned that any
155 22 to 24 nt sRNAs present in *dcl234* roots grafted to WT shoots must have moved from the
156 grafted WT shoot (36). Loci similarly represented by sRNA reads in C24/C24, C24/Col, Col/Col
157 and C24/*dcl234* but not *dcl234/dcl234* datasets (notation shoot/root, with underline indicating the
158 analyzed plant part) could be confidently assigned as producing mobile sRNA, since the WT
159 shoot complements the inability of *dcl234* roots to produce sRNA. Contrastingly, the sRNAs from
160 loci that were absent in C24/*dcl234* and *dcl234/dcl234* grafts and present in grafts with Col or
161 C24 roots were interpreted as being dependent on *DCL2,3,4* but not mobile. Loci where sRNAs
162 were present in all datasets indicate sRNAs produced by DCL1 or in a DCL-independent manner.

163

164 We incorporated genome-wide MethylC-seq data in our analyses and applied the above
165 reasoning to classify the genome-wide patterns of DNA methylation affected by graft
166 transmissible sRNAs. We used five genotype combinations in two-way grafts to define six locus
167 classes (A-F) by their combinations of DNA methylation and sRNA representation across graft
168 combinations (Fig. 1, where high and low sRNA abundance are indicated by dark or light blue,
169 respectively, and high and low DNA methylation levels by dark or light red, respectively; also SI
170 Appendix Figs. S1, S2). Three classes of these were of primary interest:

171

- 172 A. “Direct” loci correspond to overlapping regions of cytosine methylation and sRNAs
173 present in *C24/C24*, *C24/Col*, *Col/Col* and *C24/dcl234* but not *dcl234/dcl234*. This model
174 is consistent with direct targeting of DNA methylation by mobile sRNAs.
175 B. “Indirect” loci show the same pattern of DNA methylation between grafts as “Direct” loci
176 but these were not associated with sRNAs. Here we infer that there are undetectable
177 levels of mobile sRNAs or that the change in methylation is due to indirect effects of a
178 mobile signal.
179 C. “De novo methylated” loci correspond to overlapping regions of sRNAs and cytosine
180 methylation present in *C24/C24*, *C24/Col* and *C24/dcl234*, but not in *dcl234/dcl234* or
181 *Col/Col*. These loci likely correspond to sRNAs produced uniquely by C24 that target
182 DNA methylation to previously unmethylated regions of the Col genome.
183

184 The class A (direct) and C (*de novo*) loci because these are most likely to be associated with the
185 mobile sRNA. We compared their characteristics with those of class B (indirect) loci that are
186 associated with a mobile signal but, by definition, do not correspond with detectable candidate
187 sRNAs. Three classes of loci (D-F) not associated with regulation of DNA methylation by mobile
188 sRNA were also identified and are described in the Supplemental Results.
189

190

191 **Mobile sRNA and DNA methylation**

192

193 We developed a statistical framework that could be applied to sRNA and MethylC-seq data to
194 accurately assess the models described in Fig. 1 (see Supplemental Materials and Methods
195 sections 1, 3 and 4 for detailed explanation). Our previous sRNA abundance data (36) were
196 reanalyzed using this framework to ensure consistency with analyses of the MethylC-seq
197 datasets generated in the current study (Table 1). DNA methylation associated with mobile sRNA
198 was examined separately by sequence context (CG, CHG, CHH), because the genetic
199 requirements for each are distinct.
200

201 Our analyses revealed that non-CG DNA methylation is abundant at class A (direct) loci; we
202 identified 13 CG, 398 CHG and 401 CHH-context class A loci with a false discovery rate (FDR) of
203 5% (Table 1, Dataset S1). These included the three loci we identified previously as possessing
204 mobile sRNA associated DNA methylation (36). The association between regions of sRNA and
205 DNA methylation that corresponded to the class A loci model was statistically significantly more
206 frequent than expected by chance (Table 1, Z-scores; CG 154.69, CHG 77.123, CHH 84.250;
207 $p < 0.05$). Furthermore, class A loci of all three DNA methylation contexts overlapped statistically
208 significantly (SI Appendix Fig. S3). Observation of co-localized DNA methylation in all three
209 contexts is characteristic of RdDM. Fewer than 2.5% of the class A loci overlapped with regions
210 identified as variable over 30 generations in *Arabidopsis* (18) and thus it is unlikely that the results
211 are due to spontaneous epiallelic variation.
212

213 We investigated the genes potentially regulating DNA methylation at class A loci by analysis of
214 MethylC-seq data from 86 *Arabidopsis* gene silencing mutant lines (31). We examined the
215 effects of mutation of a subset of genes with known roles in RdDM (*dcl2/3/4*, *met1*, *svh4/5*,
216 *ddm1*, *svh4/5/6*, *vim1/2/3*, *cmt2*, *cmt3*, and WT control) on DNA methylation at class A loci. We
217 focused on non-CG methylation patterns because very few CG-context class A loci were
218 identified. Mutation of *drm1/2* caused the strongest reduction in non-CG context methylation in
219 class A loci (Fig. 2, SI Appendix Fig. S4). The other mutations examined showed weaker effects,
220 of which *cmt2* and *cmt3* caused the least changes in DNA methylation; *cmt3* reduced CHG
221 context DNA methylation to levels intermediate between WT and *drm1/drm2* but had no effect in
222 the CHH context, whilst *cmt2* had no effect in either context. DNA methylation was reduced to
223 levels intermediate between WT and *drm1/2* in both sequence contexts by all others of these
224 mutations. It should be noted that the MethylC-seq data from gene silencing mutant lines were
225 generated using leaves of three-week old plants, whereas our MethylC-seq data were generated
226 using roots of three-week old plants. We considered it appropriate to compare the two datasets
227 because there is little evidence for tissue-specific variation in DNA methylation in *Arabidopsis*. We

228 examined the similarity of DNA methylation in these two tissues by making comparisons between
229 the likelihoods of methylation at loci identified as methylated in root tissue of the Col/Col graft
230 from our dataset and in the three-week old Col-0 (WT) leaf tissue of Stroud and colleagues (31).
231 DNA methylation was broadly consistent between the two tissues (Dataset S2, SI Appendix Fig.
232 S5). Additionally, a prior study demonstrated that DNA methylomes from two tissues in each of
233 eleven *Arabidopsis* accessions cluster by accession, not tissue type (38).

234 **DNA methylation under indirect regulation by the mobile signal**

235
236 Class B (“indirect”) loci show the same pattern of DNA methylation between grafts as do class A
237 (“direct”) loci. However, they are distinguished by the lack of associated sRNA. We found that
238 class B and class A loci had extremely similar characteristics. Hundreds of class B loci were
239 detected almost exclusively in non-CG contexts (13 CG, 868 CHG, 1029 CHH; Table 1, Dataset
240 S1). These loci were unambiguously identified and located in the genome. We found fewer than
241 0.5% of the class B loci overlapped with regions previously identified as spontaneously variable
242 between parent and offspring plants, suggesting our results are not a consequence of this
243 phenomenon (18).

244
245 The distribution of the class B loci across DNA methylation contexts mirrors that of the class A
246 loci (13 CG, 398 CHG and 401 CHH), though substantially more class B loci were identified. We
247 focused on non-CG class B loci since so few CG loci were identified. The association between
248 regions lacking sRNAs and possessing DNA methylation, corresponding to the class B loci
249 model, was lower than would be expected by chance (Table 1; CHG, $p=0$, $Z=-92.603$; CHH, $p=0$,
250 $Z=-105.48$). This pattern suggests that methylation loci regulated by a mobile signal are
251 negatively associated with regions lacking in small RNAs.

252
253 The class A and B loci were also influenced similarly by mutations in RdDM pathway genes (31)
254 (Fig. 2, SI Appendix Fig. S4). Mutation of *drm1/2* caused the strongest reduction in non-CG
255 context DNA methylation (Fig. 2, SI Appendix Fig. S4). All other mutations examined reduced
256 non-CG context methylation at class B loci except *cmt2*. Notably, *cmt3* had the next weakest
257 effect, reducing CHG context DNA methylation but not affecting CHH context methylation.

258
259 In summary, there was very little to distinguish class A and class B loci, suggesting they are
260 ultimately under the regulation of the same mobile signal. In principle this signal could be an as
261 yet unidentified secondary regulatory factor dependent on the DCL234-dependent sRNA at the
262 class A loci. Alternatively it could be caused by the very low levels of DCL234-dependent sRNA
263 that fail to pass the significance thresholds in our model (SI Appendix Figs. S6-8), or which are
264 not well sequenced.

265
266 We used an independent statistical approach to estimate the total number of class A and B loci in
267 the dataset, as a complement to the results above (see Supplemental Materials and Methods
268 section 5 for detailed explanation). This approach evaluates all loci that may exist without
269 specifying genomic coordinates. We estimate that there exist 72 CG (59-85 with 95%
270 confidence), 1557 CHG (1506-1608 with 95% confidence) and 2238 CHH (2170-2303 with 95%
271 confidence) loci of class A, and 526 CG (485-568 with 95% confidence), 5699 CHG (5597-5802
272 with 95% confidence) and 10943 CHH (10785-11099 with 95% confidence) loci of class B by this
273 approach. From these data we conclude that thousands of DNA methylation loci found
274 predominantly in the non-CG context may be regulated by the mobile signal.

275 **DNA methylation regulated by the mobile signal associates with specific features of the genome**

276
277 Non-CG context class A loci were significantly associated with TEs and promoters containing
278 TEs, but not with promoters that do not contain TEs (Fig. 3A). Contrastingly, they were
279 significantly depleted in genes, coding regions and 5' UTRs. CG context class A loci were
280 significantly associated with TEs only. Further analysis showed that class A loci associated
281
282
283

284 significantly with many superfamilies of TE (Fig. 3B). RATHs, LINE and SINE superfamily TEs
285 were targeted more strongly than other TEs, most clearly so by the non-CG context class A loci.
286 TEs are broadly grouped according to their replication strategies as Type 1 (retroelements) or
287 Type 2 (DNA elements), and the targeted superfamilies are all short, gene-proximal Type 1 TEs.
288 Class A loci were targeted most highly by 23-24 nt mobile sRNAs, eliminated in *dcl234/dcl234*
289 grafts (Fig. 3C). These results indicate that DNA methylation directly targeted by the shoot-root
290 mobile sRNA is associated with specific TE superfamilies.

291
292 Non-CG class B loci were also associated significantly with TEs and promoters containing TEs
293 (Fig. 4A, SI Appendix Fig S2), as were non-CG class A loci. They were associated with many
294 different TE superfamilies, most strongly so with certain Type 1 retroelement superfamilies
295 (RATHs, LINE, SINE). They also showed clear associations with specific retroelement
296 superfamilies (Type 2 TEs) including DNA Mariner, Pogo and HAT.

297
298 The genome features associated with class A and B loci might either be specific to these locus
299 classes and to the mobile signal or they could be driven by dependence of the loci upon DCL234-
300 derived sRNAs. To address this point we compared the features associated with class A and B
301 loci to all features that lose DNA methylation in *dcl234/dcl234* grafts. The features that lost DNA
302 methylation in *dcl234/dcl234* grafts were highly similar to those associated with class A and B loci
303 (SI Appendix Fig. S9). The data indicate a specific subset of TEs lose DNA methylation in the
304 *dcl234* mutant. Furthermore, they suggest that dependency upon DCL234 drives the features
305 associated with class A and B loci rather than these features being specific to the mobile signal.

306 307 **The influence of the mobile signal on gene expression**

308
309 We examined whether DNA methylation regulated by the mobile signal may influence gene
310 expression. Root transcriptomes of the graft combinations described by Fig. 1 and SI Appendix
311 Fig S1 (excluding C24/Col) were profiled by RNA-seq. A total of 23 transcripts were identified as
312 significantly differentially regulated in a manner that indicated association with the mobile signal
313 (Datasets S3A, S3B). These included transcripts coding for proteins involved in cell wall
314 formation, membrane transport and regulation of osmotic potential. Three of the genes encoding
315 these were associated with DNA methylation regulated either directly or indirectly by the mobile
316 signal; AT2G01880 - PURPLE ACID PHOSPHATASE 7, AT2G36490 - REPRESSOR OF
317 SILENCING 1 (*ROS1*), AT3G43270 - pectin methyltransferase. We measured the number of
318 differentially expressed genes between C24/Col and *dcl234/dcl234* roots as a control. Twenty-
319 seven transcripts were identified as significantly differentially regulated in roots of *dcl234*
320 compared to WT roots (Datasets S4A, S4B). Differential expression of *ROS1* (AT2G36490.1) was
321 confirmed in roots of *dcl234/dcl234* versus C24/Col and C24/Col by quantitative real time PCR
322 (Q-RT-PCR, SI Appendix Fig. S10). These findings demonstrate that loss of *DCL2*, *DCL3* and
323 *DCL4* causes differential expression of a small number of genes, consistent with data in previous
324 reports (39).

325
326 We investigated whether transcripts dependent upon *DCL234* (Datasets S4A, S4B) or associated
327 with the mobile signal (Datasets S3A, S3B) might also depend upon *DRM1/2*. This was done by
328 identifying transcripts differentially regulated between WT and the *drm1/2* mutant using published
329 RNA-seq data. These data were generated from leaves of three-week old plants in parallel with
330 the MethylC-seq data discussed by Fig. 3 and SI Appendix Fig. S4 (31). Eighty-six transcripts
331 were significantly differentially regulated between WT and *drm1/2* plants (Dataset S5). The
332 transcript of *ROS1* (AT2G36490) was shared between these and transcripts associated with the
333 mobile signal (Dataset S3B). The transcript AT1G53480.1 was dependent upon *DRM1/2* (Dataset
334 S5) and *DCL234* (Dataset S4A). We attribute the low degree of overlap between *DRM1/2*-
335 dependent and mobile signal associated transcripts to tissue specificity of gene expression. Our
336 sRNA-seq and MethylC-seq experiments were conducted on roots of three-week old plants. As
337 discussed above, comparison between DNA methylomes from different tissues is appropriate
338 because there is little evidence for tissue specific variation in DNA methylation. However,
339 transcript abundance differs more between tissues. This was demonstrated by comparison of

340 RNA-seq data from two tissues of multiple *Arabidopsis* accessions, where transcript abundance
341 correlated more highly between like tissues of different accessions than different tissues of one
342 accession (38).

343 344 **Epialleles are transmitted across accessions by mobile sRNAs**

345
346 There is natural variation in RdDM and its associated sRNAs in *Arabidopsis* and the term
347 epiallele describes genomic locations where methylation differs between accessions (38, 40, 41).
348 We investigated whether sRNAs produced uniquely by C24, and not by Col, could direct DNA
349 methylation *de novo* at previously unmethylated sites in the Col genome (Fig. 1; class C loci).
350 Two class C loci were unambiguously identified in the CHG DNA methylation context and four in
351 the CHH context (Dataset S1). Three distinct sites of *de novo* RdDM existed after taking account
352 for overlap of these loci between DNA methylation contexts (Fig. 5). None of these loci had been
353 previously identified as spontaneously variable between parent and offspring plants (18). These
354 results indicate that exogenous sRNAs supplied by shoots can target *de novo* DNA methylation at
355 unmethylated sites in the genome of root cells, thereby transmitting epiallelic states from one
356 *Arabidopsis* accession to another.

357 358 **Discussion**

359
360 In this study we have characterized the genome-wide distribution of 23-24 nt shoot/root mobile
361 sRNAs and identified regions of DNA methylation that they target directly (class A loci). We also
362 identified regions of DNA methylation dependent indirectly upon mobile sRNAs (class B loci). We
363 found that loss of mobile sRNAs generated in the shoots disrupts DNA methylation at thousands
364 of sites in the roots. The class B (indirect) loci were most numerous, but the two classes were
365 otherwise indistinguishable. Both were almost entirely in the non-CG (CHG/CHH) DNA
366 methylation context. The class A and B loci were also significantly associated with the same TE
367 superfamilies. We found that these TE superfamilies also lost DNA methylation in the *dcl234*
368 mutant, which is deficient in 22 to 24 nt sRNAs, indicating that the feature association of class A
369 and B loci was driven by the dependency of mobile sRNAs on DCL234. Our results suggest the
370 function of mobile sRNAs is to reinforce silencing of these TEs. Furthermore, we have provided
371 mechanistic insights into the specific RdDM pathways regulating mobile sRNA-dependent
372 methylation. Data can be visualized in our interactive genome browser
373 (http://neomorph.salk.edu/mobile_methylome.php).

374
375 Our data confirm that mobile sRNA-dependent DNA methylation requires *DRM1* and *DRM2*, key
376 components of the 24 nt RdDM pathway. There was only limited dependence upon *CMT3* and
377 *CMT2*, both of which belong to a distinct DNA methylation maintenance pathway. Furthermore,
378 class A and B loci were associated with the shortest superfamilies of TE (Figs. 2, 4, SI Appendix
379 Table S1) including Type 1 retroelements, such as R1 elements (E1, E2, E3), SINEs and
380 LINEs, as well as certain Type 2 DNA elements (Mariner, Pogo, RC Helitron). The DNA
381 methylation pathways containing *DRM1/2* and *CMT2/3* are distinct, but overlapping, and both
382 deposit non-CG methylation (8, 34). The *DRM1/2* pathway methylates small TEs and TE edges,
383 whereas the *CMT2/3* pathway is responsible for methylation of long TEs (31, 34). Our data are in
384 concordance with this pattern and they establish a relationship of mobile sRNA-dependent
385 methylation with a specific RNA silencing mechanism.

386
387 The direct and indirect (class A and B) loci are essentially identical according to our data, except
388 that no mobile sRNAs were identified associated with indirect class B loci. These findings lead to
389 three hypotheses for further investigation. Firstly, that the mobile sRNAs from the shoot regulate
390 an unidentified secondary signal in the roots. This secondary signal would then directly regulate
391 DNA methylation at the B loci. Secondly, that mobile sRNAs can direct the RNA silencing
392 machinery to sites that they match imperfectly (i.e. with which they have mismatches in sequence
393 homology). We required perfect matching between sRNAs and genomic sequence when
394 identifying sRNA target loci, but it is possible that a certain degree of mismatch between mobile
395 sRNAs and their targets is either tolerated or required in certain circumstances. However,

396 allowing mismatches during sRNA mapping (and therefore in targeting) permits individual sRNAs
397 to target multiple sites due to their inherently short sequence, many of which are presumably
398 spurious (42). Therefore, detailed analysis of individual methylation regions and thorough
399 experimental validation is required to test this hypothesis. Moreover, it appears unlikely the
400 conclusions of our study would be altered by the outcome of such experiments, because we find
401 the characteristics of class A (direct) and class B (indirect) loci to be essentially identical. The
402 third hypothesis is that these loci are targeted by mobile sRNAs of very low abundance, which
403 could not be detected in our sRNA-seq libraries.
404

405 Class A and B loci were substantially more numerous than the other classes (Fig. 1, Table 1, SI
406 Appendix Table S2 and Fig. S1). The class D, E and F loci described respectively loci not
407 methylated by mobile sRNAs, not remethylated by mobile sRNAs, and loci whose methylation
408 was independent of DCL234-derived sRNAs. Class F loci were the least numerous. This
409 observation suggests there may be few mobile sRNA-targeted loci at which that sRNA does not
410 regulate DNA methylation.
411

412 sRNAs of 24 nt length were proportionally the most common sRNA species detected in our
413 sRNA-seq experiments and demonstrated the clearest shoot-root transmission (Fig. 2) (36). This
414 study and another indicate that 21-22 nt sRNAs are also mobile and associated with DNA
415 methylation (37). These data are consistent with previous proposals of a role for the 21-22 nt size
416 class in RdDM (6, 7, 37, 43). Furthermore, the Type 1 retroelements that we find to be the
417 predominant targets of mobile sRNA-dependent DNA methylation produce RNA intermediate
418 replication stages, making it possible that if transcriptionally active they are also targets of PTGS
419 via mobile 21 nt sRNAs (44). However, fewer 21-22 nt sRNAs were mobile than 24 nt sRNAs,
420 suggesting 24 nt sRNAs are the primary shoot-root signal. The design of our experiment did not
421 permit us to completely eliminate the possibility that “mobile” 21 nt sRNAs are produced in the
422 roots. This is because the *dcl234* mutant contains functional DCL1, which generates 21 nt
423 sRNAs, and the *dcl234* mutant eliminates their production only partially (23, 45-47). Detailed
424 investigation of their contribution would involve generation of additional series of *Arabidopsis*
425 mutant grafts and is complicated by the lethality of a *dcl1* null plant (48).
426

427 Although mobile sRNAs have minimal effects on gene expression in *Arabidopsis*, we predict that
428 their influence would be much greater in TE rich genomes. In such genomes, including those of
429 many common crops, mobile sRNAs may be an important mechanism of genome defense. The
430 TE superfamilies targeted by mobile sRNAs are enriched in euchromatic regions, typically within
431 c. 500 nt of the nearest gene (49, 50). TEs inserted near genes can have dramatic effects on
432 gene expression (15, 50). Expression of a small number of transcripts was differentially regulated
433 in our study, consistent with previous observations. (39). However, the *A. thaliana* genome
434 contains substantially fewer TEs than related outcrossing species, such as *A. lyrata* (51, 52).
435 Furthermore, TEs in the *A. thaliana* genome show a lower rate of active transposition (51). This
436 suggests that TEs in the *A. thaliana* genome may have relatively less influence on gene
437 expression than TEs in outcrossing species, which would result in fewer mobile sRNA-dependent
438 regulated genes. Nonetheless, those genes that did exhibit mobile sRNA-dependent regulation
439 had diverse functions, indicating they may have significant influence in the correct conditions.
440 Moreover, mobile sRNAs are able to move into both meristematic and meiotically active tissues,
441 where they can alter DNA methylation and gene expression (37, 53). In these tissues it is
442 essential to protect genome stability by repression of TEs so that gametes and developing organs
443 are not harmed (54, 55).
444

445 Grafting is routinely used in agriculture to combine rootstocks and shoots (scions) with desirable
446 characteristics, such as for grapevine, apples and tomatoes (56, 57). We have demonstrated that
447 mobile sRNAs regulate patterns of DNA methylation genome wide, and that expression of
448 specific genes exhibits mobile regulation. Base-resolution methylomes are being actively
449 generated in multiple species, including crops and non-model plants (38, 58-60). Clear diversity in
450 the epigenomes of closely related subspecies and accessions has been observed from these
451 data (38, 58). With our demonstration that site-specific transmission of epiallele states from one

452 accession to another can be achieved by grafting and by *de novo* methylation of unmethylated
453 DNA it is likely that at least some effects of grafting are due to the movement of RNA. Our
454 findings also indicate that DNA methylomes may provide a potential new resource to growers
455 who employ grafting. They could consider potential modification of gene expression patterns in
456 sink tissues *via* sRNA transmission from source tissues. The mobile sRNAs can alter DNA
457 methylation in germ line tissue (37) so that DNA methylation patterns altered by grafting may be
458 heritable by progeny plants (37, 61).
459

460 In summary, we have shown that transmission of sRNAs from shoots to roots of *Arabidopsis*
461 regulates DNA methylation at thousands of sites genome wide. Mobile sRNA-dependent
462 methylation is predominantly in the non-CG context (CHG and CHH), and is associated with short
463 Type 1 retroelements found in gene rich regions of the genome. We confirm that deposition of
464 mobile sRNA-dependent methylation is dependent upon the *DRM1* and *DRM2* RdDM pathway,
465 and largely independent of the *CMT2* and *CMT3* methylation pathway. Our conclusions underpin
466 future research into why plants possess a system for communicating methylation status from
467 shoot to root tissues.

468

469 **Materials and Methods**

470

471 See also Supplemental Materials and Methods.

472

473

473 **Plant Materials and Grafting**

474

475 Plants of *Arabidopsis thaliana* accessions Col-0 and C24 were used in our experiments, as well
476 as a previously described *dcl2-1/dcl3-1/dcl4-2* triple mutant (Col-0 background) (23). The C24
477 plants had a GFP transgene silenced by a partial GFP inverted-repeat, termed GxGF-IR, as
478 previously described (36). *Arabidopsis* plants were grown under 10 h supplemental fluorescent
479 lighting at 20°C, on vertical plates of 0.8% agar, 1/2 Murashige and Skoog media, pH 5.7.
480 Micrografting was conducted seven days after germination as previously described (57, 62). Plant
481 tissue was harvested five weeks after grafting, taking care to separate root from shoot and
482 exclude tissue 0.3-0.5 cm surrounding the graft junction. Two independent bioreplicates were
483 conducted for all RNA-seq samples, and two to four for each MethylC-seq sample (SI Appendix
484 Table S3).

485

486

486 **MethylC-seq and RNA-seq Library Construction**

487

488 DNA for the MethylC-seq libraries was prepared from 3-4 pooled plants per sample using the
489 Genra Puregene Tissue Kit (Qiagen) according to manufacturer's instructions. MethylC-seq
490 libraries were constructed as described previously (63). RNA-seq libraries were prepared from
491 total RNA extracted as previously described (36). RNA-seq libraries were generated with the
492 ScriptSeq RNA-seq Library Preparation Kit (Epicentre) according to the manufacturer's
493 instructions.

494

495

495 **Sequencing operations**

496

497 MethylC-seq and RNA-seq libraries were sequenced using an Illumina HiSeq 2000 and 2500
498 platforms respectively according to the manufacturer's instructions. MethylC-seq libraries were
499 sequenced for 101 cycles and RNA-seq libraries for 100 cycles. Image analyses and basecalling
500 were carried out using the manufacturer supplied software and standard parameters.

501

502

502 **Sequencing analyses**

503

504

505

See Supplemental Materials and Methods for detailed description of analytical methods and software. In brief, sRNA data were aligned to the TAIR9 reference genome requiring perfect

506 matching. When analyzing MethylC-seq data we used YAMA (Yet Another Methylome Aligner),
507 based upon Bowtie 2, to map reads to the genome (source code available at
508 <https://github.com/tjh48/YAMA>). Mapping statistics of libraries are given in SI Appendix Table S3.
509 Methylation loci were then identified independently in each context (CG, CHG, CHG) using the
510 segmentSeq R package (64), accounting for nonconversion rates in each sample (65).
511 Subsequently, loci exhibiting various models of differential methylation were identified using the
512 empirical Bayesian methods for analysis of methylation data (64) with the baySeq R package (66,
513 67). Average methylation profiles were calculated by dividing defined regions into an equal
514 number of windows and calculating methylation across these windows. Gene annotations were
515 retrieved from the *Arabidopsis* Information Portal (<https://www.araport.org>) (68) and transposon
516 annotations from The *Arabidopsis* Information Resource (69). Estimations of the numbers of loci
517 expected to fit a given model were calculated using the sum of the posterior likelihoods for that
518 model and confidence intervals were calculated by simulating ten thousand sets of true and false
519 positives based on the posterior likelihoods. Unambiguous lists of specific differentially
520 represented loci were identified by applying a false discovery rate of lower than 0.05 to the model,
521 which in general represents a more conservative analysis. Associations between DMRs, genome
522 features and sRNAs were assessed using a block-bootstrap method (70). Intersections between
523 loci and genome features were normalized for abundance of loci and features, the sizes of loci
524 and features and how loci and features might cluster within the genome when plotting. We did so
525 by calculating the number of features overlapped by the DNA methylation/sRNA loci per
526 megabase total feature length per megabase total locus length. This approach gave a visually
527 accurate representation of significance of association. Independent measurements of statistical
528 significance based upon the block-bootstrap analyses are presented beneath the bars also.
529 Promoters were defined as 2000 nt preceding a TSS. Data were deposited in the European
530 Nucleotide Archive (accession [E-MTAB-3473](https://www.ebi.ac.uk/ena/record/E-MTAB-3473))
531

532 Q-RT-PCR

533
534 Transcript abundance was assessed in roots of independent graft plants according as described
535 previously (37). Fold changes in abundance of test transcripts relative to the *ACTIN 2* transcript
536 (AT3G18780), a stable control (71), were calculated using delta delta C_T methodology (72, 73).
537 Primer sequences (5'-3') were; GGATATTGTAGCAAGCCACAG (*ROS1* forward),
538 CAAAGTTCATCTGGAGAAAGAC (*ROS1* reverse), GCCATCCAAGCTGTTCTCTC (*ACTIN 2*
539 forward), CCCTCGTAGATTGGCACAGT (*ACTIN 2* reverse).
540

541 Acknowledgements

542
543 We thank Matthew D. Schultz and Yupeng He for assistance with analyses, Huaming Chen for
544 assistance with web browser development, and Roberto Solano, Robert J. Schmitz, Taiji
545 Kawakatsu, Chongyuan Luo, Ryan Lister, Ronan C. O'Malley and Lindsay Robinson for helpful
546 discussions. MGL was funded by an EU Marie Curie FP7 International Outgoing Fellowship
547 (#252475). Work in JRE's laboratory is funded by the Gordon and Betty Moore Foundation
548 (GBMF 3034), and the National Science Foundation (MCB-1024999). JRE is an investigator of
549 the Howard Hughes Medical Institute. CWM was funded by a Clare College (Cambridge, UK)
550 Junior Research Fellowship. Work in the Baulcombe laboratory was supported by The Gatsby
551 Charitable Foundation, the EU FP7 Collaborative Project grant AENEAS, and the ERC Advanced
552 Investigator grant ERC-2013-AdG 340642. DCB is the Royal Society Edward Penley Abraham
553 Research Professor.
554

555 References

- 556
557 1. Martínez de Alba AE, Elvira-Matelot E, & Vaucheret H (2013) Gene silencing in plants: a
558 diversity of pathways. *Biochim. Biophys. Acta* 1829(12):1300-1308.
559 2. Hamilton A, Voinnet O, Chappell L, & Baulcombe D (2002) Two classes of short
560 interfering RNA in RNA silencing. *EMBO J* 21(17):4671-4679.

- 561 3. Hamilton AJ & Baulcombe DC (1999) A Species of Small Antisense RNA in
562 Posttranscriptional Gene Silencing in Plants. *Science* 286(5441):950-952.
- 563 4. Baumberger N & Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer
564 that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A*
565 102(33):11928-11933.
- 566 5. Qi Y, *et al.* (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-
567 directed DNA methylation. *Nature* 443(7114):1008-1012.
- 568 6. Nuthikattu S, *et al.* (2013) The initiation of epigenetic silencing of active transposable
569 elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Phys*
570 162(1):116-131.
- 571 7. Bond DM & Baulcombe DC (2015) Epigenetic transitions leading to heritable, RNA-
572 mediated de novo silencing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*
573 112(3):917-922.
- 574 8. Kim MY & Zilberman D (2014) DNA methylation as a system of plant genomic immunity.
575 *Trends Plant Sci* 19(5):320-326.
- 576 9. Bernstein BE, Meissner A, & Lander ES (2007) The mammalian epigenome. *Cell*
577 128(4):669-681.
- 578 10. Henderson IR & Jacobsen SE (2007) Epigenetic inheritance in plants. *Nature*
579 447(7143):418-424.
- 580 11. Lister R, *et al.* (2008) Highly integrated single-base resolution maps of the epigenome in
581 Arabidopsis. *Cell* 133(3):523-536.
- 582 12. Chan SW-L-L, *et al.* (2004) RNA silencing genes control de novo DNA methylation.
583 *Science* 303(5662):1336.
- 584 13. Dowen RH, *et al.* (2012) Widespread dynamic DNA methylation in response to biotic
585 stress. *Proc Natl Acad Sci U S A* 109(32):E2183-2191.
- 586 14. Zhong S, *et al.* (2013) Single-base resolution methylomes of tomato fruit development
587 reveal epigenome modifications associated with ripening. *Nat Biotechnol* 31(2):154-159.
- 588 15. Gent JI, *et al.* (2013) CHH islands: de novo DNA methylation in near-gene chromatin
589 regulation in maize. *Genome Res* 23(4):628-637.
- 590 16. Zilberman D, *et al.* (2004) Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA
591 methylation triggered by inverted repeats. *Curr Biol* 14(13):1214-1220.
- 592 17. Zilberman D, Cao X, & Jacobsen SE (2003) ARGONAUTE4 control of locus-specific
593 siRNA accumulation and DNA and histone methylation. *Science* 299(5607):716-719.
- 594 18. Schmitz RJ, *et al.* (2011) Transgenerational epigenetic instability is a source of novel
595 methylation variants. *Science* 334(6054):369-373.
- 596 19. Becker C & Weigel D (2012) Epigenetic variation: origin and transgenerational
597 inheritance. *Curr Opin Plant Biol* 15(5):562-567.
- 598 20. Richards EJ (2006) Inherited epigenetic variation--revisiting soft inheritance. *Nat Rev*
599 *Gen* 7(5):395-401.
- 600 21. Papp I, *et al.* (2003) Evidence for Nuclear Processing of Plant Micro RNA and Short
601 Interfering RNA Precursors. *Plant Phys* 132(3):1382-1390.
- 602 22. Wierzbicki AT, Ream TS, Haag JR, & Pikaard CS (2009) RNA polymerase V transcription
603 guides ARGONAUTE4 to chromatin. *Nat Gen* 41(5):630-634.
- 604 23. Henderson IR, *et al.* (2006) Dissecting Arabidopsis thaliana DICER function in small RNA
605 processing, gene silencing and DNA methylation patterning. *Nat Gen* 38(6):721-725.
- 606 24. Mosher RA, Schwach F, Studholme D, & Baulcombe DC (2008) PolIVb influences RNA-
607 directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad*
608 *Sci U S A* 105(8):3145-3150.
- 609 25. Zhong X, *et al.* (2014) Molecular mechanism of action of plant DRM de novo DNA
610 methyltransferases. *Cell* 157(5):1050-1060.
- 611 26. Woo HR, Dittmer TA, & Richards EJ (2008) Three SRA-Domain Methylcytosine-Binding
612 Proteins Cooperate to Maintain Global CpG Methylation and Epigenetic Silencing in
613 Arabidopsis. *PLoS Gen* 4(8):e1000156.
- 614 27. Chan SW-L, Henderson IR, & Jacobsen SE (2005) Gardening the genome: DNA
615 methylation in Arabidopsis thaliana. *Nat Rev Gen* 6(5):351-360.

- 616 28. Kim J, Kim JH, Richards EJ, Chung KM, & Woo HR (2014) Arabidopsis VIM proteins
617 regulate epigenetic silencing by modulating DNA methylation and histone modification in
618 cooperation with MET1. *Mol Plant* 7(9):1470-1485.
- 619 29. Law JA & Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation
620 patterns in plants and animals. *Nat Rev Gen* 11(3):204-220.
- 621 30. Du J, *et al.* (2012) Dual binding of chromomethylase domains to H3K9me2-containing
622 nucleosomes directs DNA methylation in plants. *Cell* 151(1):167-180.
- 623 31. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, & Jacobsen SE (2013)
624 Comprehensive analysis of silencing mutants reveals complex regulation of the
625 Arabidopsis methylome. *Cell* 152(1-2):352-364.
- 626 32. Johnson LM, *et al.* (2014) SRA- and SET-domain-containing proteins link RNA
627 polymerase V occupancy to DNA methylation. *Nature*.
- 628 33. Haag JR & Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of
629 non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol* 12(8):483-492.
- 630 34. Stroud H, *et al.* (2014) Non-CG methylation patterns shape the epigenetic landscape in
631 Arabidopsis. *Nat Struct Mol Biol* 21(1):64-72.
- 632 35. Melnyk CW, Molnar A, & Baulcombe DC (2011) Intercellular and systemic movement of
633 RNA silencing signals. *EMBO J* 30(17):3553-3563.
- 634 36. Molnar A, *et al.* (2010) Small silencing RNAs in plants are mobile and direct epigenetic
635 modification in recipient cells. *Science* 328(5980):872-875.
- 636 37. Melnyk CW, Molnar A, Bassett A, & Baulcombe DC (2011) Mobile 24 nt small RNAs
637 direct transcriptional gene silencing in the root meristems of Arabidopsis thaliana. *Curr*
638 *Biol* 21(19):1678-1683.
- 639 38. Schmitz RJ, *et al.* (2013) Patterns of population epigenomic diversity. *Nature*.
- 640 39. Laubinger S, *et al.* (2010) Global effects of the small RNA biogenesis machinery on the
641 Arabidopsis thaliana transcriptome. *Proc Natl Acad Sci U S A* 107(41):17466-17473.
- 642 40. Bond DM & Baulcombe DC (2014) Small RNAs and heritable epigenetic variation in
643 plants. *Trends Cell Biol* 24(2):100-107.
- 644 41. Groszmann M, *et al.* (2011) Changes in 24-nt siRNA levels in Arabidopsis hybrids
645 suggest an epigenetic contribution to hybrid vigor. *Proc Natl Acad Sci U S A* 108(6):2617-
646 2622.
- 647 42. McCormick KP, Willmann MR, & Meyers BC (2011) Experimental design, preprocessing,
648 normalization and differential expression analysis of small RNA sequencing experiments.
649 *Silence* 2(1):2.
- 650 43. Dunoyer P, *et al.* (2010) An endogenous, systemic RNAi pathway in plants. *EMBO J*
651 29(10):1699-1712.
- 652 44. Fedoroff NV (2012) Presidential address. Transposable elements, epigenetics, and
653 genome evolution. *Science* 338(6108):758-767.
- 654 45. Howell MD, *et al.* (2007) Genome-wide analysis of the RNA-DEPENDENT RNA
655 POLYMERASE6/DICER-LIKE4 pathway in Arabidopsis reveals dependency on miRNA-
656 and tasiRNA-directed targeting. *Plant Cell* 19(3):926-942.
- 657 46. Gascioli V, Mallory AC, Bartel DP, & Vaucheret Hae (2005) Partially redundant functions
658 of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting
659 siRNAs. *Curr Biol* 15(16):1494-1500.
- 660 47. Dunoyer P, Himber C, & Voinnet O (2005) DICER-LIKE 4 is required for RNA
661 interference and produces the 21-nucleotide small interfering RNA component of the
662 plant cell-to-cell silencing signal. *Nat Gen* 37(12):1356-1360.
- 663 48. Golden TA, *et al.* (2002) SHORT INTEGUMENTS1/SUSPENSOR1/CARPEL FACTORY,
664 a Dicer homolog, is a maternal effect gene required for embryo development in
665 Arabidopsis. *Plant Phys* 130(2):808-822.
- 666 49. Lenoir A, *et al.* (2001) The evolutionary origin and genomic organization of SINES in
667 Arabidopsis thaliana. *Mol. Biol. Evol.* 18(12):2315-2322.
- 668 50. Wang X, Weigel D, & Smith LM (2013) Transposon variants and their effects on gene
669 expression in Arabidopsis. *PLoS Gen* 9(2):e1003255.

670 51. de la Chaux N, Tsuchimatsu T, Shimizu KK, & Wagner A (2012) The predominantly
671 selfing plant *Arabidopsis thaliana* experienced a recent reduction in transposable element
672 abundance compared to its outcrossing relative *Arabidopsis lyrata*. *Mob DNA* 3(1):2.
673 52. Zhang X & Wessler SR (2004) Genome-wide comparative analysis of the transposable
674 elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. *Proc Natl*
675 *Acad Sci U S A* 101(15):5589-5594.
676 53. Zhang W, *et al.* (2014) Graft-transmissible Movement of Inverted Repeat - induced siRNA
677 Signals into Flowers. *Plant J* 80(1):106-21.
678 54. Ibarra CA, *et al.* (2012) Active DNA demethylation in plant companion cells reinforces
679 transposon methylation in gametes. *Science* 337(6100):1360-1364.
680 55. Hsieh TF, *et al.* (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science*
681 324(5933):1451-1454.
682 56. Melnyk CW & Meyerowitz EM (2015) Plant grafting. *Curr Biol* 25(5):R183-188.
683 57. Melnyk Charles W, Schuster C, Leyser O, & Meyerowitz Elliot M (2015) A Developmental
684 Framework for Graft Formation and Vascular Reconnection in *Arabidopsis thaliana*. *Curr*
685 *Biol* 25(10):1306-1318.
686 58. Li X, *et al.* (2012) Single-base resolution maps of cultivated and wild rice methylomes and
687 regulatory roles of DNA methylation in plant gene expression. *BMC Genomics* 13:300.
688 59. Schmitz RJ, *et al.* (2013) Epigenome-wide inheritance of cytosine methylation variants in
689 a recombinant inbred population. *Genome Res* 23(10):1663-1674.
690 60. Takuno S & Gaut BS (2013) Gene body methylation is conserved between plant
691 orthologs and is of evolutionary consequence. *Proc Natl Acad Sci U S A* 110(5):1797-
692 1802.
693 61. Wu R, *et al.* (2013) Inter-species grafting caused extensive and heritable alterations of
694 DNA methylation in Solanaceae plants. *PLoS one* 8(4):e61995.
695 62. Turnbull CGN, Booker JP, & Leyser HMO (2002) Micrografting techniques for testing
696 long-distance signalling in *Arabidopsis*. *Plant J* 32(2):255-262.
697 63. Urich MA, Nery JR, Lister R, Schmitz RJ, & Ecker JR (2015) MethylC-seq library
698 preparation for base-resolution whole-genome bisulfite sequencing. *Nature Protoc*
699 10(3):475-483.
700 64. Hardcastle TJ (2015) Discovery of methylation loci and analyses of differential
701 methylation from replicated high-throughput sequencing data. *bioRxiv*.
702 65. Team RDC (2008) R: A language and environment for statistical computing. R
703 Foundation for Statistical Computing.
704 66. Hardcastle TJ & Kelly KA (2013) Empirical Bayesian analysis of paired high-throughput
705 sequencing data with a beta-binomial distribution. *BMC Bioinformatics* 14:135.
706 67. Hardcastle TJ (2015) Generalised empirical Bayesian methods for discovery of
707 differential data in high-throughput biology. *Bioinformatics*.
708 68. Consortium IAI (2012) Taking the next step: building an *Arabidopsis* information portal.
709 *Plant Cell* 24(6):2248-2256.
710 69. Huala E, *et al.* (2001) The *Arabidopsis* Information Resource (TAIR): a comprehensive
711 database and web-based information retrieval, analysis, and visualization system for a
712 model plant. *Nucleic acids research* 29(1):102-105.
713 70. Bickel PJ & Freedman DA (1981) Some Asymptotic Theory for the Bootstrap. *Ann.*
714 *Statist.* 9(6):1196-1217.
715 71. Havecker ER, Wallbridge LM, Fedito P, Hardcastle TJ, & Baulcombe DC (2012)
716 Metastable differentially methylated regions within *Arabidopsis* inbred populations are
717 associated with modified expression of non-coding transcripts. *PLoS One* 7(9):e45242.
718 72. Yuan JS, Wang D, & Stewart CN (2008) Statistical methods for efficiency adjusted real-
719 time PCR quantification. *Biotechnol J* 3(1):112-123.
720 73. Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-
721 time quantitative PCR and the 2^{-Delta Delta C(T)} Method. *Methods* 25(4):402-408.
722
723
724
725

726 **Figure Legends**

727

728 **Figure 1. Genomic loci in *Arabidopsis* roots were classified according to their combination**
729 **of sRNA and DNA methylation status.** Grafts were made between shoots and roots of various
730 *Arabidopsis* genotypes (combinations indicated at top, denoted shoot/root). Target loci of sRNAs
731 and DNA methylation were identified by analyzing MethylC-seq and sRNA-seq data from roots of
732 all graft combinations. Each locus classification was defined by a specific combination of sRNA
733 and DNA methylation (mC) levels across the five graft combinations. "+" denotes a relatively high
734 level of mC or sRNA, whilst "-" denotes a relatively low level. Classifications were designated A-C
735 and are indicated on the left. See also SI Appendix Fig. S1.

736

737 **Figure 2. DNA methylation of mobile sRNA associated loci (class A, direct, and B, indirect)**
738 **is reduced most strongly by *drm1/2* mutation.** The plots show the mean proportion of DNA
739 methylation across class A and B loci in mutants of RNA-directed DNA methylation (RdDM)
740 pathway components and wild-type (WT) plants. The legend on the right indicates which color
741 represents each mutant. Data are shown separately for non-CG (CHG and CHH) sequence
742 contexts, which had the strongest association with mobile sRNAs. The proportion of methylation
743 was calculated per base (between 0 and 1, unmethylated to fully methylated) for all loci within a
744 class using published methylC-seq data from RdDM mutants (31). Loci were then normalized to
745 same size and the mean proportion of DNA methylation calculated across them. The profiles of
746 the mean proportion of methylation across the size-normalized loci are plotted between the
747 dashed vertical lines (indicated by the solid black bar labeled loci). Mean proportion of
748 methylation in flanking DNA, 4kb upstream and downstream of the loci, are indicated to the left
749 and right of the dashed vertical lines, respectively. Total number of loci assessed is given in
750 parentheses at the top right corner of each plot.

751

752 **Figure 3. Mobile sRNA directly targets DNA methylation at promoters containing**
753 **transposable elements.** (A) CHG and CHH (non-CG) context class A loci are enriched in TEs
754 and in promoters containing TEs, but are depleted in genes and coding sequences (CDS). Very
755 few CG context loci were identified. (B) Non-CG context class A loci preferentially target classes
756 of Type 1 retroelement, indicated by significant enrichment. They also, less strongly, target some
757 Type 2 DNA element classes. (C) The class A loci are targeted by graft-transmissible 23-24 nt
758 sRNAs (the known mobile class). Significance levels; # = $0 < p < 10^{-5}$, + = $10^{-4} < p < 10^{-5}$, * = 10^{-3}
759 $< p < 10^{-4}$, ! = $10^{-2} < p < 10^{-3}$, blue = under-represented, red = over-represented (relative to
760 background). (A) CDS – coding sequence, TE – transposable element, UTR – untranslated
761 region. (A, B) Y-axis units normalize the feature/locus overlap by both sum of genome feature
762 size and sum of methylation locus size, which permits comparison between columns. These units
763 are number of annotated features per total megabase (MB) of named feature per total MB of
764 methylation in locus class. Numbers of loci in each DNA methylation context shown within
765 parentheses in the legend. (C) Loci were normalized to same size on graph, indicated by dark bar
766 on position axis, flanked by sRNA coverage 4kb upstream and downstream of locus. Numbers of
767 loci assessed are given in parentheses.

768

769 **Figure 4. Loci where DNA methylation is indirectly targeted by mobile sRNAs (class B)**
770 **exhibit similar characteristics to loci that are directly targeted (class A).** (A) Non-CG (CHG
771 and CHH) context B loci are significantly enriched in TEs and in promoters containing TEs, but
772 are depleted in promoters that do not contain TEs, genes and coding sequences (CDS). Very few
773 CG context loci were identified. (B) Non-CG context B loci target most classes of TEs, but show
774 most significant association with classes of Type 1 retroelement. They also, less strongly, target
775 some Type 2 DNA element classes. Y-axis units normalize the feature/locus overlap by both sum
776 of feature size and sum of methylation locus size, which permits comparison between columns.
777 These units are number of annotated features per total megabase (MB) of named feature per
778 total MB of methylation in locus class. Significance levels; # = $0 < p < 10^{-5}$, + = $10^{-4} < p < 10^{-5}$, * =
779 $10^{-3} < p < 10^{-4}$, ! = $10^{-2} < p < 10^{-3}$, blue = under-represented, red = over-represented (relative to
780 background). TE – transposable element, UTR – untranslated region. Numbers of loci assessed
781 are given in parentheses.

782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800

Figure 5. C24-derived sRNAs can target DNA methylation *de novo* at unmethylated regions of the Col-0 genome. Three such *de novo* loci are shown in AnnoJ genome browser screenshots. The screenshots display sRNA reads and methylated cytosine residues (mC) for graft combinations C24/C24, Col/Col and C24/Col. Results from two independent biological replicates, suffixed 1 and 2, are shown for each graft combination. Note the groups of sRNA reads present only in C24/C24 and C24/Col, which correlate with the presence of mC. Red and green sRNAs indicate that they map to the Watson (W) and Crick (C) strands, respectively. Gold, blue and pink mC positions indicate methylation in the CG, CHG and CHH context, respectively.