Title: **Pseudouridine guides germline small RNA transport and epigenetic inheritance.**

Authors: Rowan P Herridge[1, ‡*], Jakub Dolata[1,4*], Valentina Migliori[2,3*], Cristiane de Santis Alves[1*], Filipe Borges[1], Frédéric Van Ex[1], Ann Lin[1], Mateusz Bajczyk[4], Tommaso Leonardi[2,5], Alan Hendrick[6], Andrea J Schorn[7], Tony Kouzarides[2**], and Robert A Martienssen[1**]

Affiliations:

[1]Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA

[2]The Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK

[3]present address: Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

[4] Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznan, Poland

[5]Center for Genomic Science of IIT@SEMM, Instituto Italiano di Tecnologia (IIT), 20139 Milan, Italy

[6]Storm Therapeutics, Ltd., Moneta Building (B280), Babraham Research Campus, Cambridge CB22 3AT, UK

[7]Cold Spring Harbor Laboratory, Cold Spring Harbor NY11724

[‡]Present address: Department of Biochemistry, University of Otago, Dunedin, NZ

[*]these authors contributed equally

[**]For correspondence: t.kouzarides@gurdon.cam.ac.uk, martiens@cshl.edu

Summary:

Epigenetic modifications that arise during plant and animal development, such as DNA and histone modification, are mostly reset during gamete formation, but some are inherited from the germline including those marking imprinted genes[1,2]. Small RNAs guide these epigenetic modifications, and some are also inherited by the next generation[3,4]. In *C. elegans*, inherited small RNA precursors have poly (UG) tails[5], but how inherited small RNAs are distinguished in other animals and plants is unknown. Pseudouridine (Ψ) is the most abundant RNA modification but has not been explored in small RNAs. Here, we develop novel assays to detect Ψ in short RNA sequences, demonstrating its presence in mouse and *Arabidopsis* microRNAs and their precursors. We also detect substantial enrichment in germline small RNAs, namely epigenetically activated siRNAs (easiRNAs) in *Arabidopsis* pollen, and piwi-interacting piRNAs in mouse testis. In pollen, pseudouridylated easiRNAs are localized to sperm cells, and we found that *PAUSED/HEN5 (PSD)*, the plant homolog of Exportin-t, interacts genetically with Ψ and is required for transport of easiRNAs into sperm cells from the vegetative nucleus. We further show that Exportin-t is required for the triploid block: chromosome dosage-dependent seed lethality that is epigenetically inherited from pollen. Thus, Ψ has a conserved role in marking inherited small RNAs in the germline.

One-Sentence Summary: Pseudouridine marks germline small RNAs in plants and mammals, impacting epigenetic inheritance via nuclear transport.

22

23 The germlines of both plants and metazoans have abundant small RNA, many of which are cell-
24 type specific. In the mammalian germline, spermatids accumulate 26-28 nt PIWI-interacting
25 RNAs (piRNAs) which silence TEs, while spermatocytes accumulate 29-30nt piRNA that also
26 target genes[6]. In plants, the male germline terminates in mature pollen, which contains two cell
27 types, the pollen grain (or vegetative cell), and two sperm cells enclosed within its cytoplasm.
28 Plants do not have piRNAs, but instead 21-22 nt epigenetically activated siRNAs (easiRNAs) are
29 produced from TEs in the vegetative nucleus (VN) and translocated to the sperm cells[7-9].
30 Furthermore, 24 nt siRNAs from some DNA transposons are translocated into developing
31 meiocytes in the anther from the surrounding layer of somatic cells (the tapetum)[8]. Both types of
32 small RNA are produced by the plant-specific RNA polymerase, Pol IV[8,10,11] and mediate
33 epigenetic inheritance after fertilization [10,12]. In maize, 21 nt and 24 nt "phasiRNAs" are also
34 transported from epidermal and tapetal cells into the meiocytes, demonstrating that these
35 mechanisms may be conserved[13]. Both miRNAs and siRNAs in plants are modified by 2'-O-
36 methylation, while in animals only piRNAs have this modification[14].

37

38 We set out to determine whether pseudouridine (Ψ) was present in small RNAs. Ψ is the most
39 common non-canonical base in RNA but has not been explored in small RNAs due to difficulties
40 in detection. High-throughput techniques to detect pseudouridylation in longer RNAs
41 specifically label Ψ with N-cyclohexyl-N'-(2-morpholinoethyl)-carbodiimide-metho-*p*-
42 toluenesulfonate (CMCT) to block reverse transcriptase (RT) during RNA-seq library
43 preparation[15-18]. However, in the case of small RNA, RT blocks would result in truncated
44 sequences too short to map back to the genome. More recently, anti-Ψ antibody binding[19], as

45 well as bisulphite treatment at neutral pH[20] achieve similar results. In mouse, long primary
46 miRNA precursor transcripts (pri-miRNAs) have been shown to contain Ψ [21,22], so we initially
47 tested for the presence of Ψ in small RNAs from mouse NIH/3T3 cells. Following
48 immunoprecipitation with a Ψ-specific antibody, we could robustly detect Ψ using a combination
49 of mass spectrometry (Extended Data Fig.1a) and dotblots (Extended Data Fig. 1b) using *in vitro*
50 synthesized pseudouridylated transcripts as a positive control. Using this antibody, we then
51 performed pseudouridine immunoprecipitation (Ψ-IP) and sequencing of small RNAs from
52 NIH/3T3 cells before and after knockdown of the Ψ synthase PUS1 (Extended Data Fig. 1c-e;
53 Supplementary Dataset S3). We detected Ψ in several miRNAs (Extended Data Fig. 1c, d) four
54 of which were strongly impacted in the *PUS1* knockdown (Fig. 1e). Members of the let-7 family
55 were heavily pseudouridylated (Extended Data Fig. 1d), but unaffected by *PUS1* knockdown,
56 consistent with let-7 precursor interaction with DKC1 and TruB1 instead[21,22].

57

58 In search of factors that might modify small RNAs in plants, we performed a yeast two-hybrid
59 screen for interactors of *Arabidopsis* ARGONAUTE 3 (AGO3) using a cDNA library of
60 transcripts from *Arabidopsis* flowers[23]. The screen (Methods) identified a potential interactor of
61 AGO3, namely H/ACA RIBONUCLEOPROTEIN COMPLEX SUBUNIT 2 (NHP2), and we
62 confirmed the interaction by bimolecular fluorescence complementation in phase-separated
63 nucleolar bodies (Fig. 1a). NHP2 plays a key role in RNA pseudouridylation via the
64 pseudouridine synthase Dyskerin (DKC1), often, though not always, via snoRNA guides [24-26].
65 NHP2 was also in close proximity with other AGO proteins, notably AGO4, which has been
66 found in nucleolar bodies and in related Cajal bodies (Extended Data Fig. 1f)[27].
67 Immunoprecipitation of GFP-tagged NHP2 followed by mass spectrometry revealed interactions

68 with other subunits of the Dyskerin complex (DKC1 and NOP10) as well as components of the
69 RNAi machinery, including AGO1, 2, 4, 5, 9 and 10 (Fig1b). Immunoprecipitation of GFP-
70 tagged NHP2 followed by RNA sequencing revealed snoRNAs, snRNAs and several pri-
71 miRNAs that co-immunoprecipitated with NHP2 either directly or as part of a larger complex
72 (Supplementary Dataset S2; Supp Extended Data Fig. 1g). In order to determine whether pri-
73 miRNAs were pseudouridylated in plants, as they are in animals, we then performed Ψ-IP and
74 sequencing on mRNAs from *Arabidopsis* flower buds (collectively called the inflorescence).
75 Several pri-miRNA transcripts were enriched, and there was a significant ($p<0.01$) overlap with
76 those enriched by immunoprecipitation with GFP-tagged NHP2 (Extended Data Fig. 1h;
77 Supplementary Dataset S4).
78
79 Encouraged by these results, we tested three approaches to detect Ψ directly in small RNAs in
80 *Arabidopsis* (Methods; Extended Data Fig.2 a): (i) Ψ-IP followed by small RNA sequencing, (ii)
81 CMC treatment of small RNA libraries following 5′-adapter ligation, to block RT and deplete
82 small RNAs that contain Ψ, and (iii) reverse transcription of CMC-treated small RNAs in the
83 presence of Mn^{2+}, which causes deletions and mutations at Ψ residues instead of a reverse
84 transcription stop[28]. To determine the efficacy of these techniques, we examined coverage at
85 known pseudouridylation sites in rRNA and tRNA. Ψ at UNΨAR motifs is directed by PUS7
86 and was detected by CMC-depletion in WT but not in *pus7* mutants (Extended Data Fig. 2b, c).
87 Ψ at position 55 in tRNA depends on PUS10 (Extended Data Fig 2b). tRNA fragments
88 containing Ψ-55 were IP-enriched and CMC-depleted, contained mismatches in CMC/Mn^{2+}-
89 treated libraries, and were ablated in *pus10* mutants (Extended Data Fig 2b, d, e). Predicted sites
90 of rRNA pseudouridylation were also enriched by Ψ-IP, depleted by CMC treatment, and

91 mismatched in CMC/Mn^{2+} libraries (Extended Data Fig2f-l). Thus, all three methods robustly
92 detected Ψ in small RNA fragments derived from tRNA and rRNA.
93
94 We then applied our assays to detect Ψ in mature small RNAs from *Arabidopsis* inflorescence
95 and detected more than forty pseudouridylated mature miRNAs (Fig. 1d), of which four
96 (miR159b-3p, miR403-5p, miR447a-3p and miR833a-5p) were detected using all three assays
97 (Fig. 1e; Supplementary Datasets S5-7). Examination of *pus7* and *pus10* mutant datasets
98 (Extended Data Fig.2 c, d) revealed that these tRNA modification enzymes had no detectable
99 effect on Ψ-IP enrichment or CMC-depletion of miRNA. Instead, we expected that DYSKERIN
100 (DKC1) might be responsible, given the enrichment for pri-miRNA in pulldowns with the
101 Dyskerin co-factor NHP2 (Extended Data Fig 1b). As mutants of NHP2 and other members of
102 the DKC complex are lethal[29], we rescued *dkc1* knockout mutations with an embryo-specific
103 DKC1 cassette (Extended Data Fig. 3a), which restored about 10% of wild-type *DKC1*
104 expression in the leaves (Extended Data Fig. 3b) and allowed seeds to germinate into viable but
105 infertile plants (Extended Data Fig. 3c). Several (44) miRNAs were significantly enriched by Ψ-
106 IP in WT plants of which 12 lost Ψ enrichment in *dkc1-2* leaves and another 15 required *DKC1*
107 to a lesser extent (<75% Ψ enrichment). Overall, Ψ levels in miRNA were significantly
108 decreased in *dkc1-2* ($P<0.05$; Extended Data Fig. 3d). We conclude that just as some miRNAs
109 are pseudouridylated by PUS1 in the mouse (Extended Data Fig. 1h), some are pseudouridylated
110 by DKC1 in Arabidopsis, but in both cases multiple pseudouridine synthases must be responsible
111 for miRNA pseudouridylation, just as they are for mRNA pseudouridylation in plants, yeast and
112 mammals[30].
113

114 Ψ-detection by mismatch sequencing revealed that pseudouridylation in *Arabidopsis* miRNAs
115 (Fig. 1g) and in siRNAs (Extended Data Fig. 4 a, b) was prevalent at the 5′ nucleotide, which is
116 frequently uridine in miRNAs and contributes to AGO loading preference[31-33]. As pri-miRNAs
117 associate with NHP2 and contain Ψ (Extended Data Fig. 1b), we predicted that Ψ in mature
118 miRNAs was deposited prior to processing of pri-miRNAs into small RNAs by the endonuclease
119 DICER-LIKE1 (DCL1). cDNA reverse transcribed from pri-miR159b in the presence of CMC
120 and Mn^{2+} revealed extensive mismatches (pseudouridylation) at the 5′ dicing site (25%Ψ), which
121 was exacerbated in a *hyponastic leaves1 (hyl1)* mutant (75%Ψ) that accumulates pri-miRNAs
122 due to loss of the microprocessor component R2D2 (Extended Data Fig. 4c). Ψ was also detected
123 at 5′ and 3′ sites in mature miR159b, suggesting that pseudouridylated pri-miRNAs are processed
124 into mature miRNAs (Extended Data Fig. 4c). We confirmed the presence of Ψ in mature
125 miR159b on Northern blots, where CMCT-treatment induces a mobility shift (Fig. 1h).
126 Intriguingly, it has recently been shown that miR159b is inherited from sperm cells by the
127 fertilized endosperm, where it impacts seed development[4]. As miR159b is heavily
128 pseudouridylated (Fig.1h), this suggests that pseudouridylated small RNA may be inherited from
129 pollen.

130

131 We therefore applied Ψ-IP, CMC and CMC/Mn^{2+} treatments to small RNA isolated from pollen.
132 We found that known sites of pseudouridylation in 18S rRNA were robustly detected in rRNA
133 fragments in pollen by all three techniques (Extended Data Fig. 5a-d). In both CMC and Ψ-IP
134 datasets, many more miRNAs in pollen were pseudouridylated than in flower buds, including
135 miR159b (Fig. 2a; Supplementary Datasets S7-9). This effect was even more dramatic for
136 siRNAs, which were heavily pseudouridylated in pollen (Fig.2b). Pseudouridylated siRNA (Ψ-

137 siRNA) were associated with most annotated TE families, although there were notable
138 exceptions (Copia41, Copia 36, AtGP1 and AtGP2). Next, we tested whether Ψ-miRNA and Ψ-
139 siRNA were enriched in sperm cells or associated with AGO proteins. We purified sperm cells
140 from total pollen by FACS and immunoprecipitated AGO proteins from total pollen, before
141 making small RNA sequencing libraries (Fig.2c). To determine whether pseudouridylation of
142 small RNA correlates with their localization, we compared pseudouridylation of pollen miRNAs
143 to sperm cell localization, revealing some association (Fig. 2d)[34]. Pseudouridylation of siRNA,
144 on the other hand, was strongly associated with sperm cell localization ($R^2 = 0.88$). AGO 2, 5, 6,
145 7 and 9 are all expressed in mature pollen, where they are mostly restricted to sperm cells[35],
146 including AGO5 and AGO9 which are localized to the perinuclear space (Fig. 1e,f). We
147 performed small RNA sequencing after immunoprecipitation (RIP-seq) with anti-AGO9
148 antibodies (Methods) and compared the abundance of siRNA in each TE family by sequencing
149 following Ψ-IP, along with published RIP-seq data from AGO1 and AGO5[36]. In contrast to Ψ-
150 miRNA, which were relatively weakly associated (Fig.2 g-i), Ψ-siRNA were strongly associated
151 with sperm cell localization, and with small RNA in AGO5, AGO9 and AGO1 pulldowns (Fig.
152 2j-l).

153

154 Pseudouridylation has been implicated in cytoplasmic localization of mRNA and tRNA
155 suggesting a role for Ψ in small RNA transport[17,37]. Pol IV-dependent 21-22 nt easiRNAs in
156 pollen are also mobile, and move into sperm cells from the VN[7-9]. To further investigate a role
157 for Ψ in inter-cellular mobility of small RNAs, we probed the function of Ψ in sperm cell
158 localization. Since our *dkc1* mutant was completely sterile (Extended Data Fig. 3c), we could not
159 use this to investigate the function of Ψ in pollen. Additionally, other PUS enzymes likely have

160 considerable redundancy, as twenty exist in Arabidopsis[30,38]. Instead, we took advantage of the
161 fact that pseudouridylation of tRNAs by Pus1^{+} contributes to nuclear export in yeast, and
162 genetically interacts with the tRNA export factor Loss of Suppression 1 (Los1^{+}), which encodes
163 Exportin-t[37]. Exportin-t binds the upper minihelix of L-shaped tRNA, consisting of the acceptor-
164 stem and the TψC-loop, and terminating at Ψ55 and C56[39]. This minihelix includes the 22nt 3'
165 tRNA fragment with a 5' Ψ and 3'CCA, resembling a 21-22nt small RNA duplex[40].
166 Intriguingly, the *Arabidopsis* homolog of *Los1*$^{+}$, which is known as *PAUSED/HEN5 (PSD)*, was
167 previously identified in genetic screens for mobile small RNA phenotypes, along with mutants in
168 *ARGONAUTE7* (*zippy*) and *HASTY (hst*). *HST* encodes Exportin-5 and has recently been shown
169 to be required for long-range transport of artificial miRNA in Arabidopsis[41-44]. We therefore
170 investigated whether *psd* mutants could be used as a proxy for the function of Ψ in siRNA, as
171 mutants are viable and fertile.

172

173 First, we demonstrated that the *psd* mutation in *Arabidopsis* is synthetically lethal with *pus7* in a
174 manner similar to *los1*$^{-}$ and *pus1*$^{-}$ from yeast. We found that *pus7-/-*; *psd+/-* double mutants
175 were semi-sterile (Extended Data Fig. 6a). In contrast, *pus7+/-; psd-/-* plants had fertile pollen,
176 but had 25% aborted seed compared with less than 5% in *psd* alone (Extended Data Fig. 6b, c).
177 Viable double homozygous seeds were not recovered from either genotype upon self-
178 fertilization. These results indicated that tRNA require *PUS7* for modification in diploid cells
179 before meiosis (*pus7-/-*), and subsequently require *PSD* for nuclear export in haploid pollen
180 (from *psd+/-* parents). Next, we performed Ψ-IP and CMC depletion on small RNA from WT
181 and *psd* mutant pollen as well as WT and mutant inflorescence (flower buds) and found robust
182 detection of pseudouridylation sites in rRNA (Extended Data Fig.5 a-b, e-f). Of those miRNAs

183 that were significantly pseudouridylated in the inflorescence, only miR403 lost Ψ in the mutant
184 (Extended Data Fig. 5g), while miRNAs overall were unaffected (Extended Data Fig. 5h).
185 Similarly, siRNAs from TEs were mostly unmodified in the inflorescence, and were unaffected
186 by *psd* (Extended Data Fig. 5i). In pollen, however, miRNAs were significantly depleted of Ψ in
187 *psd* relative to WT (Fig. 3a). easiRNA accumulation depends on both PolIV and miR845
188 (miR845a and miR845b), the most abundant pollen miRNAs, which target hundreds of LTR
189 retrotransposons at their reverse transcription primer binding site[10,36]. siRNAs from miR845
190 targets were heavily pseudouridylated, which depended on *psd* (Fig. 3a), as were Pol IV-
191 dependent 21-22 nt and 23-24 nt siRNAs (Fig. 3b).

192

193 In pollen, RdRM components are prominently expressed in the VN[45], but easiRNAs accumulate
194 in sperm cells[7-9]. To determine if Ψ affected siRNA transport from VN into sperm cells, we
195 isolated sperm cells from mutant *psd-13* (Exportin-t), mutant *hst-6* (Exportin-5), and wild-type
196 pollen by FACS and compared the abundance of siRNAs and miRNAs (Fig. 3c, d). Analysis
197 revealed that 21-22nt easiRNAs were reduced in *psd* and elevated in *hst* (Fig. 3c). In contrast,
198 miRNAs were sharply reduced in *hst* sperm cells as expected, but unaffected by *psd* (Fig.3d)
199 consistent with the dependence of miRNA but not siRNA accumulation on Exportin-5[44,46].
200 Strikingly, 21-22nt easiRNA that depend on PolIV, were strongly depleted from sperm cells in
201 *psd* but not in *hst* mutants (Fig.3e, Extended Data Fig. 7a-f), consistent with transport being
202 dependent on Ψ. While overall levels of miRNA in sperm cells were unaffected by *psd* (Fig. 3d),
203 individual miRNA including miR159 and miR403 significantly lost Ψ in pollen (Fig. 3f). A
204 likely explanation is that miRNA that lose Ψ in *psd* mutants can use the *HST* pathway instead to
205 accumulate in sperm cells. One prediction of this model is that Ψ-miRNA might be able to use

206 the PSD pathway in *hst* mutants. This prediction was confirmed, as miRNA without 5' terminal
207 Ψ were lost in sperm cells from *hst* mutants, while those with terminal Ψ were retained,
208 suggesting PSD was sufficient for their transport (Fig. 3g).

209

210 easiRNAs from pollen target imprinted genes in Arabidopsis[10,12]. Imprinted genes in plants are
211 expressed in the seed from either the maternal or the paternal genome and are known as
212 maternally expressed and paternally expressed genes (MEGs and PEGs), respectively.
213 *DEMETER* (*DME*) encodes a DNA glycosylase that is expressed in the endosperm and in the
214 pollen VN, where it removes DNA methylation from TEs upstream and downstream of
215 imprinted genes resulting in parent-of-origin expression[47]. We therefore examined Ψ in siRNAs
216 from DME targets and imprinted genes. We detected Ψ in 21-22 nt easiRNAs and 24 nt siRNAs
217 from DME targets (Fig.4a), and from TEs surrounding maternally- and paternally-expressed
218 imprinted genes (MEGs and PEGs) (Extended Data Fig. 7g, h). Pseudouridylation of siRNAs
219 from upstream and downstream of DME targets was lost in *psd-13* mutant pollen (Fig. 4a), and
220 there was a striking depletion of easiRNAs from the promoters of DME target genes in *psd*
221 sperm cells relative to WT (Fig.4b-c). Ψ loss was more significant in MEGs rather than PEGs
222 ($p<0.05$; Extended Data Fig. 7a) and also affected sperm cell localization of siRNAs from both
223 MEGs and PEGs, with 20-22 nt siRNAs depleted and 23-25 nt siRNAs enriched in *psd-13* sperm
224 cells compared to wild-type ($p<0.001$; Extended Data Fig. 7b).

225

226 Pol IV-dependent 21-22 nt easiRNAs and miR845 mediate the triploid block, a dose-dependent
227 epigenetic phenomenon whereby *Arabidopsis* seeds abort during development, if they inherit an
228 excess of paternal genomes from diploid instead of haploid pollen [10-12]. As 21-22 nt Pol IV-

229 dependent easiRNAs were lost from sperm cells in *psd-13* pollen, this suggested that *psd-13*
230 might rescue triploid seed abortion[10]. As expected, we observed ~90% seed abortion when WT
231 Columbia plants were crossed with *omission of second division 1 (osd1)* mutants, which have
232 diploid pollen (Fig. 4d-g). However, when *osd1*; *psd-13* diploid pollen was applied, this resulted
233 in only ~10% seed abortion, similar to when *osd1*; *nrpd1a* diploid pollen was applied, which is
234 deficient in the large subunit of Pol IV (NRPD1). Importantly, levels of miR845a and miR845b
235 were not affected in *psd-13* mutant pollen, and were not responsible for rescuing the triploid
236 block (Extended Data Fig. 7i)[10]. Exportin-t is therefore essential for the triploid block and for
237 pseudouridylation of easiRNAs.

238

239 Thus *PAUSED/EXPORTIN-T* is required for transport of pseudouridylated easiRNAs, and of
240 pseudouridylated miRNAs into sperm cells from the VN, while *HASTY/EXPORTIN-5* transports
241 unmodified miRNAs (Fig. 4h). Pseudouridylated easiRNAs lose Ψ in *psd* mutants, presumably
242 because they fail to bind AGO proteins if they are not exported to sperm cells (Fig. 2e-h; Fig. 4i)
243 and are degraded. Alternatively, nucleo-cytoplasmic transport might be required for
244 pseudouridylation. This novel function for *PAUSED/EXPORTIN-T* is pollen-specific, as
245 pseudouridylated miRNAs and easiRNAs are mostly restricted to the male germline (Extended
246 Data Fig. 5g-i). It is noteworthy that tRNA have been found to carry a strong signal for
247 intercellular transport in plants, so that mRNA-tRNA fusions are mobile[48]. It is tempting to
248 speculate that Ψ and *PSD* may be involved there as well. Cytoplasmic connections between VN
249 and sperm cells are thought to be involved in transport [7,9], consistent with a role for
250 nucleocytoplasmic exportin. AGO5 is found in these "male germ units"[34], and AGO5 is also
251 required at least in part for the triploid block[36]. AGO9 has a similar localization (Fig. 2e, f) and

252 has previously been associated with siRNA movement in pollen[49,50]. We therefore examined
253 AGO9 localization using an mCherry N-terminal fusion and found that AGO9 was first localized
254 in the VN at the bicellular stage, but then moved to the sperm cells by the tricellular stage (Fig.
255 4i). This localization is consistent with a role for AGO9 before, and after siRNA transport as
256 previously proposed[49,51].

257

258 TE-derived siRNAs in the plant germline have similar functions to piRNAs in mammalian
259 germline, and so we asked if high levels of Ψ were conserved in piRNAs. We used Ψ-IP as well
260 as CMC depletion to detect Ψ in small RNAs from 8-week *post-partum* mouse testis. As
261 expected, 3'-tRNA fragments (tRFs) were enriched by Ψ-IP (Extended Data Fig. 8a) and
262 depleted by CMC (Extended Data Fig. 8b). Importantly many piRNAs from TEs, including from
263 LINE and LTR retroelements, were strongly enriched for Ψ (Extended Data Fig. 8c, d). Up to
264 3% of piRNAs were differentially enriched for Ψ, though many more were immunoprecipitated
265 and hence potentially pseudouridylated (Extended Data Fig. 8d). No individual piRNA cluster
266 was significantly enriched or depleted, suggesting sequence- rather than locus- specificity
267 (Extended Data Fig. 8e, f). 3'-tRF that were depleted by CMC treatment were 22nt in size as
268 expected for 3'tRF containing Ψ-55 from mature tRNA (Extended Data Fig.8e) while piRNA
269 were mostly 30 nt in size as expected (Extended Data Fig.8f). The same small RNA sizes and
270 classes were detected by Ψ-IP (Extended Data Fig.8g, h). Thus, not only 3'tRFs, but also
271 piRNAs and miRNAs in mammals are pseudouridylated. It remains to be seen if they also
272 depend on Ψ for transport by Exportin-t. Intriguingly, 5'tRFs are also pseudouridylated, and
273 were reported to be inherited suggesting a similar mechanism in mammalian sperm[52,53]. Why are
274 germline small RNAs so heavily modified in both plants and mammals? An intriguing possibility

275 is that modifications of RNA in the germline may avoid viral surveillance systems after
276 fertilization, which could otherwise recognize inherited small RNA as "non-self"[54].
277
278
279 **Materials and methods**
280 Cell lines and tissues
281 NIH/3T3 (ATCC, CRL-1658) were cultured in DMEM (Invitrogen), supplemented with 10%
282 fetal bovine serum (FBS) and 1% penicillin/streptomycin/glutamine. Cells were obtained from
283 ATCC and tested negative for mycoplasma contamination. Testes were harvested from
284 C57Bl/6 mice at 8 weeks *post-partum* and flash frozen in liquid nitrogen.
285
286 Plant growth conditions
287 *Arabidopsis* plants were grown under long-day conditions at 22 °C. Seeds were always surface-
288 sterilized with sodium hypochlorite, sown on Murashige and Skoog medium and stratified for 3
289 d at 4 °C. Seedlings were transplanted to soil 2 weeks after germination and grown under long-
290 day conditions at 22 °C. The following mutants were used: pus7-1 (SALK_098773), pus10¬-1
291 (SALK_091360), nrpd1a-3 (SALK_128428). Homozygous mutants were identified using PCR
292 with primers specific to the predicted site of T-DNA integration (Supplementary Table S1). psd-
293 13 and hst-6 alleles were previously identified[41,55]. The osd1-1 mutant was kindly provided by
294 Raphael Mercier.
295

296 Yeast two hybrid screening.

297 The two-hybrid screen was performed by Hybrigenics (Paris, France) using an Arabidopsis
298 thaliana Col-0 library (ATFO) prepared from young flowers, using an in-house cDNA clone of
299 AGO3 as bait. 7 candidate proteins were tested by bimolecular fluorescence complementation,
300 leading to the identification of AT5G08180, which encodes the H/ACA ribonucleoprotein
301 complex subunit 2 (NHP2), as a high confidence interactor. This was subsequently confirmed by
302 Mass Spectrometry.

303

304 Bi-molecular fluorescence complementation

305 BiFC was performed using pBiFC-2in1 vectors[56]. AGO3 and NHP2 were fused in all
306 orientations (NN, CN, NC, CC; only NN and CC complemented), while other AGO proteins
307 were fused only in NN configuration. Tobacco leaves were infiltrated essentially as previously
308 described[57]. Briefly, agrobacterium cultures harboring pBiFC vectors were grown overnight
309 before subculturing in 50 mL of LB broth with antibiotics and 20 µM acetosyringone. Cultures
310 were pelleted by centrifugation, and resuspended in infiltration buffer (10 mM MES, 10 mM
311 MgCl2, 100 µM acetosyringone). Leaves of ~3 weeks old Nicotiana benthamiana plants were
312 infiltrated on the abaxial surface using a syringe without a needle. Three days post-infiltration
313 leaf discs were removed and imaged using a Zeiss LSM780 confocal microscope. Primers for
314 cloning AGO proteins and NHP2 are detailed in Supplementary Table S1.

315

316 Construction of viable *DYSKERIN* (*dkc1*) mutants in Arabidopsis

317 Dyskerin (*dkc1*) viable mutants were generated as follows. *AtNAP57 (DKC1)* genomic sequence
318 was amplified and cloned into the pENTR vector. The *ABI3* promoter sequence was amplified
319 and inserted using the NotI restriction site. *pABI3:AtNAP57* cassette was transferred to the
320 pGWB501 vector using the Gateway cloning system and used for Agrobacterium-mediated,
321 floral dip Arabidopsis transformation of *nap57*/+ (SALK_031065) heterozygous plants.
322 Homozygous transformant lines were obtained by three rounds of Hygromycin resistance
323 selection followed by verification of *nap57/nap57* genetic background by PCR.

324

325 Co-immunoprecipitation and mass-spectrometry analyses

326 For each co-IP, 0.6 g of tissue from GFP-NHP2 plants was used. The powder was extracted for
327 30 min in 1.5 ml of lysis buffer (50 mM Tris HCl pH 8.0, 50 mM NaCl, 1% Triton X-100 and 1×
328 cOmplete™ ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor (Roche). After
329 removal of cell debris by centrifugation (twice 10 min, 16 000 g, 4°C) the cleared supernatants
330 were incubated for 30 min with anti-GFP antibodies coupled to magnetic microbeads (µMACS
331 GFP isolation Kits, Miltenyi). Beads were loaded on magnetized MACS separation columns
332 equilibrated with lysis buffer and washed four times with 300 µl washing buffer (50 mMTris
333 HCl pH 7.5, 0.1% Triton X-100). Samples were eluted in 100 µl of pre-warmed elution buffer
334 (Milteny). Control IPs were performed using Col-0 plants and GFP antibodies. Eluted proteins
335 were digested with sequencing-grade trypsin (Promega). Peptide mixtures were separated using
336 Nano-Ultra Performance Liquid Chromatography coupled to Orbitrap Velos mass spectrometer
337 (Thermo) at IBB PAS, Warsaw. Data were searched against the TAIR10 database. Peptides were
338 identified with the Mascot algorithm (Matrix Science, London, UK). The total number of
339 MS/MS fragmentation spectra was used to quantify each protein from three replicates. The

340 statistical analysis based on spectral counts was performed using a homemade R package that
341 calculates fold change and p values using the quasi-likelihood negative binomial generalized log-
342 linear model implemented in the edgeR package[58]. P-value was adjusted using the Benjamini–
343 Hochberg method from the stats R package. The size factor used to scale samples was calculated
344 according to the DESeq2 normalization method[59].

345

346 miRNA precursor pull-down with NHP2-GFP

347 Anti-GFP antibodies were used to immunoprecipitate NHP2-GFP from three independent
348 transgenic lines expressing GFP-NHP2 under native promoter and WT plants as a negative
349 control. RNA was converted into cDNA libraries using NEBNext Ultra II Directional RNA
350 Library Prep Kit for Illumina. Libraries were sequenced with Nextseq HO 100bp paired end.
351 Adapters and unpaired reads were removed using Trimmomatic and files were analyzed using
352 Salmon software. Obtained TPM values for each transcript were adjusted with 0.1 or 0.01 for
353 snRNA/snoRNA and pri-/pre-miRNA respectively to avoid dividing by zero in the case of non-
354 detected transcripts. IP values were normalized against input, and Fold Change was calculated
355 for GFP-NHP2 vs. negative control (GFP IP in WT plants). Reproducibly enriched RNAs in
356 NHP2 IP had FC $\geq$ 1.2 in at least two lines.

357

358 RNA extraction of mammalian cells

359 Cells were lysed in Qiazol (QIAGEN) and total RNA was purified using the miRNEasy mini kit
360 (QIAGEN) according to the manufacturer's instructions. Tissues were lysed in Qiazol
361 (QIAGEN) and total RNA was purified using Tissuelyser II (QIAGEN), as per the

362 manufacturer's instructions. Small RNAs (<200nt) size-fractionated using RNA Clean &
363 Concentrator 5 column kits (Zymo), as per the manufacturer's instructions. Testes of 8 weeks old
364 mice were grinded under liquid nitrogen and RNA was extracted using Trizol (Thermo Fisher
365 Scientific) according to the manufacturer's instructions but using 80% EtOH during the wash
366 step.

367

368 Dot-blot

369 For dot-blot analysis, input RNA or RNA immunoprecipitated with either anti-Ψ or isotypic non-
370 specific antibodies was spotted onto a nitrocellulose membrane and UV cross-linked at 254 nm
371 (120 mJ/cm2). The membranes were blocked in Denhart's solution (1% Ficoll, 1%
372 polyvinylpyrrolidone, 1% bovine serum albumin; Thermo Fisher) for 1 h at room temperature
373 and incubated with Ψ antibody for 1 h at room temperature. Signal was detected using HRP
374 conjugated secondary antibodies and ECL (GE Healthcare) and developed on a Chemidoc MP
375 machine (BioRad).

376

377 Pseudouridine Immunoprecipitation

378 Small RNAs were isolated from 0.01-0.1 mg total RNA, first by separating <200 nt RNAs with
379 the Zymo clean and concentrator kit, and subsequently gel extracting RNA 20-30 nt on a 15%
380 polyacrylamide gel. Small RNAs were eluted from the crushed gel slice overnight in 500 µL
381 RNA elution buffer (0.3M NaCl, 0.5 mM EDTA). RNA was resuspended in 50 µL water,
382 typically containing ~4-40 ng/µL RNA. Resuspended RNA was mixed with 200 µL IPP buffer
383 (150 mM NaCl, 0.1% NP-40, and 10 mM Tris-HCl [pH 7.4] + 10 U/mL RNAse inhibitor) and 5

384 μg of mouse monoclonal pseudouridine antibody (D347-3, MBL intl.). Samples were incubated
385 for 2 hours at 4C with rotation before adding 50 μL sheep anti-mouse IgG dynabeads washed in
386 IPP buffer (Invitrogen). The unbound fraction was retained, and the beads were washed 4 times
387 with IPP buffer. To prepare RNA from the unbound fraction, 400 μL binding buffer from the
388 Zymo clean and concentrator kit was added, and RNA was run through the column according to
389 manufacturer's instructions. To elute bound plant siRNA from the beads, 150 μL Trizol was
390 added and incubated for 2 minutes before placing tubes on a magnetic rack. Trizol was
391 transferred to a new tube and 30 μL chloroform was added and mixed thoroughly. Samples were
392 centrifuged for 10 minutes at maximum speed and the aqueous phase was pipetted into a new
393 tube with an equal volume of 100% ethanol. Mouse small RNAs were eluted using 6.7 mM Ψ
394 triphosphate in IPP buffer with RNase-OUT (Thermo Fisher Scientific) for 30 minutes at 37C.
395 Samples were then run through a Zymo Clean and Concentrator kit according to manufacturer's
396 instructions, and eluted in 8 μL. Libraries were prepared using NEBNext small RNA sequencing
397 kit (New England Biolabs) or NextFlex small RNA sequencing kit (Perkin Elmer) according to
398 manufacturer's instructions, with the same temperature modifications as the CMC-treated
399 samples. Libraries were pooled and run on NextSeq 500, HiSeq 4000 or MiSeq high-throughput
400 sequencing systems (Illumina).

401

402 Nucleoside analysis of RNA

403 RNA samples were incubated overnight with a 2x nuclease mix containing 62.5 units of
404 Benzonase (Sigma Aldrich), 5 units of Antarctic Phosphatase (NEB) and 10 mU/μL of
405 phosphodiesterase I (PDEI) from Crotalus adamanteus venom (Sigma Aldrich) made up in a 5X
406 digest buffer composed of 20 mM Tris-HCl (pH 8), 20 mM MgCl2 and 100 mM NaCl. Samples

407 were then further purified to remove protein by filtration at 14,000 rpm for 30 minutes through
408 Amicon® Ultra-0.5 spin columns (Sigma Aldrich) which have a 30 kDa molecular weight cutoff
409 (56). Samples were then analysed using a Q Exactive HF Orbitrap mass spectrometer (Thermo
410 Scientific). Briefly, 2ul of digested sample was loaded onto a 100Å, 1.8 µm, 2.1 mm X 100 mm
411 ACQUITY UPLC HSS T3 Column (Waters) in an aqueous buffer consisting of 0.1% MS-grade
412 formic acid in water: acetonitrile 98:2 at a constant flow of 0.3ml/min. A gradient of organic
413 solvent was used to facilitate separation and elution as follows: from 2 to 8 minutes the
414 percentage of organic solvent (0.1% formic acid in acetonitrile) increased from 2% to 10%.
415 From 8 to 13 minutes the gradient increased from 10% to 98% organic solvent, where it was held
416 for 1 minute before returning to 2% organic to re-equilibrate the column for the next sample.
417 Samples were read on a full scan positive mode from 100-600 m/z and data were analysed using
418 Xcalibur software (Thermo Scientific). All nucleosides were compared to exogenous standards
419 which allowed the interpolation of concentrations of sample nucleosides. Isobaric nucleosides
420 were differentiated on the basis of their retention times as compared with pure commercial
421 standards. As an example, Ψ has a retention time of 1.55 minutes in this protocol, whereas
422 uridine elutes at 3.05 minutes.

423

424 RT-qPCR

425 Cells were lysed in Qiazol (QIAGEN) and total RNA was purified using the miRNEasy mini kit
426 (QIAGEN) according to the manufacturer's instructions. For mRNA detection, 1 µg of purified
427 total RNA was reverse transcribed using the high-capacity cDNA reverse transcription kit
428 (Applied Biosystems). For specific miRNA quantification, we size-fractionated small RNAs (<
429 200nt, containing pre-miRNAs) using RNA Clean & Concentrator 5 column kits (Zymo), as per

430 the manufacturer's instructions. miRNAs were reverse transcribed with miScript II RT kit
431 (QIAGEN). Primers were designed to anneal on the 5p or 3p arm of the miRNA of interest
432 (Supplementary Table S1). miRNAs were quantified using Fast SybrGreen PCR mastermix
433 (Applied Biosystems) according to the manufacturer's instructions.

434

435 CMC-treatment and small RNA sequencing

436 For each pair of libraries 4 µg of Arabidopsis inflorescence RNA, prepared using Zymo
437 Quickzol RNA kit, was ligated to the 3′ adapter supplied in the NEB NextSeq kit in a 40 µL
438 reaction. After ligation samples were divided in two and subjected to CMC or mock treatment.
439 80 µL of 0.5M CMC in BEU buffer (7 M urea, 4 mM EDTA, 50 mM bicine pH 8) was added to
440 20 µL sample and incubated at 37C for 30 minutes. Mock treated samples were incubated with
441 80 µL of BEU buffer. After CMC/mock treatment 25 µL 3M sodium acetate, 3 µL glycogen and
442 1 mL 100% ethanol was added and samples were incubated for 1 hour at -80C. Samples were
443 centrifuged at 4°C for 30 minutes and the resulting pellets were washed with 80% ethanol,
444 followed by another 1 hour at -80C. Samples were centrifuged at 4C for 30 minutes and the
445 ethanol removed; pellets were allowed to dry for 5-10 minutes before addition of 50 µL freshly
446 prepared 0.05 M carbonate buffer (pH 10.4) containing 4 mM EDTA. Samples were incubated at
447 37C for 3 hours and subsequently cleaned using Zymo RNA clean and concentrator kit, eluting
448 in a final volume of 25.5 µL. After cleaning, libraries were prepared according to instructions in
449 the NEBNext small RNA sequencing kit (New England Biolabs) or Nextflex Small RNA-Seq
450 Kit v3 (Perkin Elmer), beginning with annealing of the RT primer, with some modifications.
451 Primer annealing was performed at 65C for 5 minutes, followed by 37C for 15 minutes and 25°C
452 for 15 minutes. cDNA synthesis was carried out at 42C. After PCR amplification libraries were

453 cleaned using Qiagen PCR cleanup kit and run on a 6% acrylamide gel. Bands of ~140 nt were
454 cut from the gel, crushed, and eluted for 2 hours in DNA elution buffer followed by ethanol
455 precipitation as outlined in the manufacturer's instructions. Libraries were pooled and run on
456 NextSeq or MiSeq high-throughput sequencing systems (Illumina) at the Cold Spring Harbor
457 Laboratory sequencing facility.

458

459 CMC/Mn2+ treatment and sequencing

460 For high-throughput siRNA mismatch analysis 2 µg total Arabidopsis RNA was treated as
461 described above for CMC-siRNA sequencing libraries with some modifications. Nextflex Small
462 RNA-Seq Kit v3 (Perkin Elmer) was used for all samples. The RT step was modified to utilize
463 SuperScript II (Thermo Fisher) and a custom 5X First-Strand Buffer containing MnCl2 (250 mM
464 Tris-HCl, pH 8.3; 375 mM KCl; 30 mM $MnCl_2$). For pri-miRNA mismatch analysis, 10 µg of
465 total RNA was treated in the same way as for siRNA mismatch analysis. Obtained cDNA was
466 used for amplification with pri-miRNA specific primers. PCR products were gel-purified and
467 cloned into pGEM-T Easy (Promega) for Sanger sequencing. The same number of clones was
468 sequenced for corresponding CMC+ and CMC- samples.

469

470 Northern Blotting

471 50 µg total RNA was treated with 0.5 M CMC in BEU buffer (typically 25 µL RNA in 200 uL
472 buffer). Samples were cleaned and prepared in the same way as for small RNA sequencing,
473 except that RNA <200 nt was prepared using Zymo RNA Clean and Concentrator to enrich small
474 RNA in a very low volume. Samples were briefly denatured at 65C for two minutes (higher

475 temperatures should be avoided as these can remove the CMC adducts) before loading on a 15%
476 denaturing polyacrylamide TBE gel. CMC+/CMC- were treated identically, except for addition
477 of CMC. Untreated RNA was loaded directly on the gel after denaturation. Gels were run until
478 the bromophenol blue band reached the bottom. RNA was transferred to Hybond NX membrane
479 using a semi-dry electroblotter for 30 minutes. Crosslinking was performed with EDC[60].
480 Membranes were blocked using DIG easy-hyb buffer (Roche) followed by incubation with 10
481 pmol/mL 5′/3′ DIG labeled LNA probe (Exiqon) in easy-hyb buffer at 65C overnight. Probes
482 were detected using Roche DIG wash and block buffer set, anti-DIG antibody and CDP star
483 reagent according to manufacturer's instructions.

484

485 Immunoprecipitation of AGO9

486 Immunoprecipitation of AGO9 from Arabidopsis pollen was performed as previously described
487 (35). Pollen was collected from open Arabidopsis flowers as described above, prior to grinding
488 in liquid nitrogen and homogenizing in 5 ml g–1 of extraction buffer for 1 h at 4C (50mM Tris-
489 HCl pH 7.5, 150mM NaCl, 10% Glycerol, 0.2% NP-40, 5mM MgCl2, 5mM DTT, containing
490 one tablet per 10 ml of protease inhibitor cocktail, Roche). Cell debris was removed by
491 centrifugation at 12 Krpm at 4C for 20 min and the supernatant was recovered. Protein extract
492 was pre-cleared by incubation with 10 µl Protein A Dynabeads (Invitrogen) at 4C for 30 min.
493 Supernatant was recovered by centrifugation at 12 Krpm at 4C for 1min. Precleared extracts
494 were incubated with AGO9 antibody (1:100 dilution) and 30 µl of Protein A Dynabeads at 4C
495 overnight. The immunoprecipitate was pelleted and washed three times (15 min each in rotator
496 shaker) in extraction buffer. Small RNAs from the purified AGO9 complex were isolated using
497 Ambion columns. Small RNA was resolved on a 12.5% denaturing PAGE 7M-Urea gel, and

498 stained with SYBR-gold (Invitrogen). For cloning, gel slices within the range of 18–30 nt were
499 excised, and the RNAs were eluted and purified as described above. Small RNAs were
500 sequenced by Fasteris (Switzerland).

501

502 Flow cytometry and staining of Arabidopsis pollen

503 Flow cytometry and Alexander staining was performed as previously reported[34]. Pollen was
504 collected in 1.5mL eppendorf tubes by vortexing open flowers in pollen extraction buffer (PEB,
505 10 mM CaCl2, 2 mM MES, 1 mM KCl, 1% H3BO3, 10%) for 3 min 47, followed by filtration
506 through a 30um mesh (Partec/Sysmex) and centrifugation at 5,000g for 1 min. Pollen was
507 suspended in 50ul of PEB, immediately frozen in liquid nitrogen and stored at -80C until RNA
508 extraction. Open flowers were collected into a 50 ml falcon tube, frozen in liquid nitrogen, and
509 stored at -80C for storage. The pollen enriched fraction was isolated with the procedure
510 described in Borges et al. (2012) using Sperm Extraction Buffer (SEB). Intact pollen grains were
511 sorted with fluorescence-activated cell sorting (FACS) using a S3e Cell Sorter Bio-Rad. To the
512 obtained pollen sample in 1.5 ml tube, 100 µl of acid-washed glass beads (425–600 µm, Sigma)
513 was added and vortexed continuously at medium speed for 1 minute in order to break mature
514 pollen grains. The sample was transferred to a new tube and the volume of the sample was made
515 up to 1 ml with SEB. 1 µl of Sybr Green was added for staining. The sample was ready for
516 FACS for sperm cells sorting according to Borges, et al. [34]. RNA from the sperm cells fraction
517 was isolated with Trizol, treated with DNase I, and used for small RNA library preparation
518 (NextFlex Small RNA Kit, Perkin Elmer). Pollen staining was done following Alexander's
519 staining protocol[61]. Only viable pollen grains take up the stain.

520

521 Bioinformatic analysis of rRNA coverage

522 Adapters were trimmed from reads, converted to fasta and collapsed using fastx toolkit.
523 Trimmed reads between 15-30nt in length were mapped to the Arabidopsis genome (TAIR10)
524 using bowtie[62] allowing mismatches and multi-mapping. Coverage across the rRNA locus on
525 chromosome 2 (position 3706-9521) was calculated using bedtools [63]. Coverage within a sample
526 was normalized to the total amount of coverage across the locus, and the log2 fold-change
527 between paired samples (IP/Unbound, CMC/Mock) was calculated and averaged between
528 technical replicates. Sites of predicted pseudouridylation were taken from published datasets and
529 the average log2 fold change surrounding these sites was calculated[64,65].

530

531 Bioinformatic analysis of CMC/Mn2+-treated Arabidopsis samples

532 For adapter trimming and size selection of reads (18-42 nt) Cutadapt was used[66]. Terminal
533 deletions in CMC+ samples were detected using DeAnnIso[67] or miRMOD[68]. For internal
534 deletions detection reads were aligned to the Arabidopsis genome (TAIR10) using BWA[69] and
535 further processed using SAMtools[70] bam-readcount and AWK homemade scripts.

536

537 Bioinformatic analysis of Arabidopsis siRNA/miRNA

538 Adapters were trimmed from reads, converted to fasta and collapsed using fastx toolkit.
539 Trimmed reads between 20-25nt in length were mapped to the Arabidopsis genome (TAIR10)
540 using bowtie allowing mismatches and discarding unmapped reads. Individual sequences with an
541 average of >10 reads per sample were selected and analysed using DESeq2[59], treating
542 CMC+/Mock or IP/Unbound fractions as pairs. Mature miRNA sequences from miRBase (v21)

543 were extracted from these datasets. Sequences mapping to tRNA, snoRNA, snRNA or rRNA
544 were removed for comparison with AGO-bound sequences from published datasets. UNUAR-
545 containing and 3′CCA tRFs were identified by extracting reads mapping to tRNAs and selecting
546 those containing the UNUAR motif or 3′CCA; these sequences were extracted from the DESeq2
547 output for plotting. Significance of the overlap between CMC, IP and CMC/Mn2+ results was
548 calculated by simulating 1x10e6 outcomes and calculating the probability of the observed
549 overlap with a custom script.

550

551 Bioinformatic analysis of mouse miRNA

552 The mouse small RNA Sequencing data from NIH/3T3 cells were analysed with the SeqImp
553 pipeline configured with the mm10 assembly of the Mus musculus genome, the transcriptome
554 annotation from Ensembl (v89) and the miRbase annotation of microRNAs (v21). The pipeline
555 was executed with the default mouse miRNA configuration and default parameters, specifying
556 the sequence of the 3' adapter used for library preparation (AGATCGGAAGAGCACACGTC).
557 The miRNA quantifications obtained with SeqImp were further analyzed in R/Bioconductor with
558 the DESeq2 package[59]. Briefly, counts were library-size normalized using the median ratio
559 method as implemented in the estimateSizeFactors function. Variance was estimated with the
560 estimateDispersions function and statistical testing for enrichment performed with the Wald test
561 as implemented in the nbinomWaldTest function. P-values were corrected for multiple
562 hypothesis testing with the Benjamini-Hochberg procedure.

563

564 Analysis of TE-derived siRNA

565 Collapsed reads 18-30 nt long were mapped to the TAIR10 genome allowing multimapping, but
566 no mismatches using Bowtie. Quantification of siRNA mapping to each TE was performed as
567 described previously[10]. Reads mapping to TE families were aggregated and the average log2
568 fold-change between paired samples was calculated. The mean of all technical replicates
569 (generated by separate CMC/mock or IP/unbound library preparations on the same RNA) for any
570 given biological replicate was taken prior to calculating the mean of all biological replicates for
571 each tissue (inflorescence/pollen) and treatment (CMC or IP) combination.

572

573 Metaplots

574 Uncollapsed reads 20-25nt in length were aligned to the Arabidopsis (TAIR10) genome using
575 Bowtie allowing mismatches but only a single mapping event per read. Mapping files from the
576 same treatment (ie. CMC+, CMC-, IP, Unbound) were merged after normalization (as above) to
577 enable the production of a single representative plot. Coverage was calculated using bedtools
578 (60). Metaplots were produced using the Versatile Aggregate Profiler[71] using a window size of
579 25, and a smoothing window of 6 for TE analysis, and a window size of 50 and smoothing
580 window of 12 for DME target analysis. For comparison between psd-13 and wild-type, metaplots
581 of three biological replicates were created and the mean enrichment was plotted.

582

583 Bioinformatic analysis of mouse testis small RNA

584 Reads were clipped for Illumina Truseq (AGATCGGAAGAGCACACGTCTGAAC) or Perkin
585 Elmer Nextflex Small RNA adapter sequences (N(4)TGGAATTCTCGGGTGCCAAGG) using
586 Cutadapt. Reads were required to be 14-44nt and have a quality score of at least Q20 over 90%

587 of their length. tRNA fragment sequences (tRFs) were analyzed as described[72] using all tRNA
588 sequences included in the current UCSC repeat annotation
589 (ftp://hgdownload.soe.ucsc.edu/goldenPath/mm10/database/rmskOutCurrent.txt.gz). Reads that
590 were not tRFs were mapped to the UCSC mm10 genome sequence using Bowtie 2, filtered for 0-
591 3 mismatches (to accommodate mismatches due to RNA modifications) using SAMtools and
592 BAMtools, and intersected with the UCSC repeat annotation using BEDtools. Repeat and non-
593 repeat reads were further intersected with mm10 mouse piRNA clusters
594 (https://www.smallrnagroup.uni-mainz.de/piCdb/) to result in piRNA reads that overlap (i)
595 transposon sequences, (ii) structural RNAs or (iii) neither. Raw counts of individual piRNA
596 sequences were analyzed for pseudouridine enrichment (enriched in IP, depleted in CMC-treated
597 libraries) using DESeq2[59]. Similarly, raw counts per piRNA cluster were analyzed using
598 DESeq2, but each cluster was required to have at least an average of 100 raw reads (per IP or per
599 CMC-untreated sample). Read counts that were used as input for DESeq2 analysis were from
600 three replicates of each condition (input, IP; CMC-untreated, CMC-treated). Log2 fold changes
601 were plotted against -log10(adjusted p-value) using R.

602

603 **Acknowledgments:** The authors would like to thank J.S. Parent and M. Donoghue for
604 discussions and observations, Joe Simorowski, Uma Ramu and Alex Oruci for technical
605 assistance, and Tim Mulligan for plant care.

606

607 **Funding:** This work was supported by the Howard Hughes Medical Institute, and by grants from
608 the National Institutes of Health (GM067014 and GM76396), the National Science Foundation
609 Plant Genome Research Program, and the Robertson Research Foundation (to R.A.M.). Work in

610 the Kouzarides laboratory is supported by a grant from Cancer Research UK (grant reference
611 RG96894), in addition to benefiting from core support from the Welcome Trust (WT203144)
612 and Cancer Research UK (grant reference C6946/A24843). V.M. was funded by a Kay Kendall
613 Leukemia Fund project grant (grant reference RG88664) and Cancer Research UK (grant
614 reference RG96894). J.D was funded by a grant from the Polish National Science Centre
615 (2020/39/D/NZ1/01918). The authors acknowledge assistance from the Cold Spring Harbor
616 Laboratory Shared Resources, which are funded in part by the Cancer Center (Support Grant
617 5PP30CA045508).

618

619 **Author contributions:** R.P.H., J.D., C.S.A., V.M., T.K. and R.A.M. conceived and designed
620 experiments. R.P.H., V.M., J.D., C.S.A., A.L, F.V., and A.H. performed the experiments. R.P.H.,
621 J.D., T.L. and A.J.S. performed bioinformatic analysis. F.B. provided code and advice for
622 transposon analysis. R.A.M. and R.P.H. wrote the manuscript with input from V.M., J.D.,
623 C.S.A., A.J.S., T.L. and T.K.

624

625 **Competing interests:** T.K. is a co-founder of Abcam Plc and Storm Therapeutics Ltd,
626 Cambridge, UK. A.H. is an employee of Storm Therapeutics Ltd, Cambridge, UK.
627 **Data and materials availability:** Data has been deposited in GEO under accession numbers
628 GSE229986; GSE230359;GSE230228; GSE222751.

629

630

631 **Supplementary Materials**

632 Supplemental Table S1

633 Supplemental Data S1 to S9

634

635

636 **References**

637 1 Kohler, C., Wolff, P. & Spillane, C. Epigenetic mechanisms underlying genomic
638 imprinting in plants. *Annu Rev Plant Biol* **63**, 331-352, doi:10.1146/annurev-arplant-
639 042811-105514 (2012).

640 2 Skvortsova, K., Iovino, N. & Bogdanovic, O. Functions and mechanisms of epigenetic
641 inheritance in animals. *Nat Rev Mol Cell Biol* **19**, 774-790, doi:10.1038/s41580-018-
642 0074-2 (2018).

643 3 Zhang, Y. *et al.* Dnmt2 mediates intergenerational transmission of paternally acquired
644 metabolic disorders through sperm small non-coding RNAs. *Nat Cell Biol* **20**, 535-540,
645 doi:10.1038/s41556-018-0087-2 (2018).

646 4 Zhao, Y. *et al.* Clearance of maternal barriers by paternal miR159 to initiate endosperm
647 nuclear division in Arabidopsis. *Nat Commun* **9**, 5011, doi:10.1038/s41467-018-07429-x
648 (2018).

649 5 Shukla, A. *et al.* poly(UG)-tailed RNAs in genome protection and epigenetic inheritance.
650 *Nature* **582**, 283-288, doi:10.1038/s41586-020-2323-8 (2020).

651 6 Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific
652 class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199-202,
653 doi:10.1038/nature04917 (2006).

654 7 Slotkin, R. K. *et al.* Epigenetic reprogramming and small RNA silencing of transposable
655 elements in pollen. *Cell* **136**, 461-472, doi:10.1016/j.cell.2008.12.038 (2009).

656 8 Long, J. *et al.* Nurse cell–derived small RNAs define paternal epigenetic inheritance in
657 Arabidopsis. **373**, eabh0556 (2021).

658 9 Martinez, G., Panda, K., Köhler, C. & Slotkin, R. K. Silencing in sperm cells is directed
659 by RNA movement from the surrounding nurse cell. *Nat Plants* **2**, 1-8 (2016).

660 10 Borges, F. *et al.* Transposon-derived small RNAs triggered by miR845 mediate genome
661 dosage response in Arabidopsis. *Nat Genet* **50**, 186-192, doi:10.1038/s41588-017-0032-5
662 (2018).

663 11 Panda, K., McCue, A. D. & Slotkin, R. K. Arabidopsis RNA Polymerase IV generates
664 21-22 nucleotide small RNAs that can participate in RNA-directed DNA methylation and

665 may regulate genes. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190417,
666 doi:10.1098/rstb.2019.0417 (2020).

667 12 Martinez, G. *et al.* Paternal easiRNAs regulate parental genome dosage in Arabidopsis.
668 *Nat Genet* **50**, 193-198, doi:10.1038/s41588-017-0033-4 (2018).

669 13 Zhai, J. *et al.* Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic
670 phasiRNAs in maize anthers. *Proc Natl Acad Sci U S A* **112**, 3146-3151,
671 doi:10.1073/pnas.1418918112 (2015).

672 14 Gainetdinov, I. *et al.* Terminal modification, sequence, length, and PIWI-protein identity
673 determine piRNA stability. *Mol Cell* **81**, 4826-4842 e4828,
674 doi:10.1016/j.molcel.2021.09.012 (2021).

675 15 Carlile, T. M. *et al.* Pseudouridine profiling reveals regulated mRNA pseudouridylation
676 in yeast and human cells. *Nature* **515**, 143-146, doi:10.1038/nature13802 (2014).

677 16 Li, X. *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian
678 transcriptome. *Nat Chem Biol* **11**, 592-597, doi:10.1038/nchembio.1836 (2015).

679 17 Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated
680 pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162,
681 doi:10.1016/j.cell.2014.08.028 (2014).

682 18 Lovejoy, A. F., Riordan, D. P. & Brown, P. O. Transcriptome-wide mapping of
683 pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. *PLoS*
684 *One* **9**, e110799, doi:10.1371/journal.pone.0110799 (2014).

685 19 Martinez Campos, C. *et al.* Mapping of pseudouridine residues on cellular and viral
686 transcripts using a novel antibody-based technique. *RNA* **27**, 1400-1411,
687 doi:10.1261/rna.078940.121 (2021).

688 20 Dai, Q. *et al.* Quantitative sequencing using BID-seq uncovers abundant pseudouridines
689 in mammalian mRNA at base resolution. *Nat Biotechnol* **41**, 344-354,
690 doi:10.1038/s41587-022-01505-w (2023).

691 21 Kurimoto, R. *et al.* The tRNA pseudouridine synthase TruB1 regulates the maturation of
692 let-7 miRNA. **39**, e104708 (2020).

693 22 Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA
694 processing by Lin28. *Science* **320**, 97-100, doi:10.1126/science.1154040 (2008).

695 23 Chen, C. *et al.* Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated
696 Arabidopsis male meiocytes. *BMC plant biology* **10**, 1-14 (2010).

697 24 Gurha, P., Joardar, A., Chaurasia, P. & Gupta, R. Differential roles of archaeal box
698 H/ACA proteins in guide RNA-dependent and independent pseudouridine formation.
699 *RNA biology* **4**, 101-109 (2007).

700 25 Lafontaine, D. L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M. & Tollervey,
701 D. The box H+ ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase.
702 *Genes and development* **12**, 527-537 (1998).

703 26 Watkins, N. J. *et al.* Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative
704 RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif
705 snoRNPs and constitute a common bipartite structure. *RNA* **4**, 1549-1568 (1998).

706 27 Li, C. F. *et al.* An ARGONAUTE4-containing nuclear processing center colocalized with
707 Cajal bodies in Arabidopsis thaliana. *Cell* **126**, 93-106, doi:10.1016/j.cell.2006.05.032
708 (2006).

709 28 Zhou, K. I. *et al.* Pseudouridines have context-dependent mutation and stop rates in high-
710 throughput sequencing. *RNA Biol* **15**, 892-900, doi:10.1080/15476286.2018.1462654
711 (2018).

712 29 Kannan, K., Nelson, A. D. & Shippen, D. E. Dyskerin is a component of the Arabidopsis
713 telomerase RNP required for telomere maintenance. *Mol Cell Biol* **28**, 2332-2341,
714 doi:10.1128/MCB.01490-07 (2008).

715 30 Sun, L. *et al.* Transcriptome-wide analysis of pseudouridylation of mRNA and non-
716 coding RNAs in Arabidopsis. *J Exp Bot* **70**, 5089-5600, doi:10.1093/jxb/erz273 (2019).

717 31 Havecker, E. R. *et al.* The Arabidopsis RNA-directed DNA methylation argonautes
718 functionally diverge based on their expression and interaction with target loci. *Plant Cell*
719 **22**, 321-334, doi:10.1105/tpc.109.072199 (2010).

720 32 Mi, S. *et al.* Sorting of small RNAs into Arabidopsis argonaute complexes is directed by
721 the 5' terminal nucleotide. *Cell* **133**, 116-127, doi:10.1016/j.cell.2008.02.034 (2008).

722 33 Montgomery, T. A. *et al.* Specificity of ARGONAUTE7-miR390 interaction and dual
723 functionality in TAS3 trans-acting siRNA formation. *Cell* **133**, 128-141,
724 doi:10.1016/j.cell.2008.02.033 (2008).

725 34 Borges, F. *et al.* FACS-based purification of Arabidopsis microspores, sperm cells and
726 vegetative nuclei. *Plant Methods* **8**, 1-9 (2012).

727 35 Jullien, P. E., Schroder, J. A., Bonnet, D. M. V., Pumplin, N. & Voinnet, O. Asymmetric
728 expression of Argonautes in reproductive tissues. *Plant Physiol* **188**, 38-43,
729 doi:10.1093/plphys/kiab474 (2022).

730 36 Oliver, C. *et al.* The miRNome function transitions from regulating developmental genes
731 to transposable elements during pollen maturation. *Plant Cell* **34**, 784-801,
732 doi:10.1093/plcell/koab280 (2022).

733 37 Grosshans, H., Lecointe, F., Grosjean, H., Hurt, E. & Simos, G. Pus1p-dependent tRNA
734 pseudouridinylation becomes essential when tRNA biogenesis is compromised in yeast. *J*
735 *Biol Chem* **276**, 46333-46339, doi:10.1074/jbc.M107141200 (2001).

736 38 Xie, Y. *et al.* Genome-Wide Identification and Expression Analysis of Pseudouridine
737 Synthase Family in Arabidopsis and Maize. *Int J Mol Sci* **23**, doi:10.3390/ijms23052680
738 (2022).

739 39 Cook, A. G., Fukuhara, N., Jinek, M. & Conti, E. Structures of the tRNA export factor in
740 the nuclear and cytosolic states. *Nature* **461**, 60-65, doi:10.1038/nature08394 (2009).

741 40 Schorn, A. J. & Martienssen, R. Tie-Break: Host and Retrotransposons Play tRNA.
742 *Trends Cell Biol* **28**, 793-806, doi:10.1016/j.tcb.2018.05.006 (2018).

743 41 Hunter, C. A., Aukerman, M. J., Sun, H., Fokina, M. & Poethig, R. S. PAUSED encodes
744 the Arabidopsis exportin-t ortholog. *Plant Physiol* **132**, 2135-2143 (2003).

745 42 Li, J. & Chen, X. PAUSED, a putative exportin-t, acts pleiotropically in Arabidopsis
746 development but is dispensable for viability. *Plant Physiol* **132**, 1913-1924 (2003).

747 43 Garcia, D., Collier, S. A., Byrne, M. E. & Martienssen, R. A. Specification of leaf
748 polarity in Arabidopsis via the trans-acting siRNA pathway. *Curr Biol* **16**, 933-938,
749 doi:10.1016/j.cub.2006.03.064 (2006).

750 44 Brioudes, F. *et al.* HASTY, the Arabidopsis EXPORTIN5 ortholog, regulates cell-to-cell
751 and vascular microRNA movement. *The EMBO Journal*, e107455 (2021).

752 45 Calarco, J. P. *et al.* Reprogramming of DNA methylation in pollen guides epigenetic
753 inheritance via small RNA. *Cell* **151**, 194-205, doi:10.1016/j.cell.2012.09.001 (2012).

754 46 Park, M. Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H. & Poethig, R. S. Nuclear
755 processing and export of microRNAs in Arabidopsis. *Proc Natl Acad Sci U S A* **102**,
756 3691-3696, doi:10.1073/pnas.0405570102 (2005).

757 47 Khouider, S. *et al.* Male fertility in Arabidopsis requires active DNA demethylation of
758 genes that control pollen tube function. *Nat Commun* **12**, 410, doi:10.1038/s41467-020-
759 20606-1 (2021).

760 48 Zhang, W. *et al.* tRNA-Related Sequences Trigger Systemic mRNA Transport in Plants.
761 *Plant Cell* **28**, 1237-1249, doi:10.1105/tpc.15.01056 (2016).

762 49 Wu, W. *et al.* Heterochromatic silencing is reinforced by ARID1-mediated small RNA
763 movement in Arabidopsis pollen. *New Phytol* **229**, 3269-3280, doi:10.1111/nph.16871
764 (2021).

765 50 Olmedo-Monfil, V. *et al.* Control of female gamete formation by a small RNA pathway
766 in Arabidopsis. *Nature* **464**, 628-632, doi:10.1038/nature08828 (2010).

767 51 Yelina, N. E. *et al.* Putative Arabidopsis THO/TREX mRNA export complex is involved
768 in transgene and endogenous siRNA biosynthesis. *Proc Natl Acad Sci U S A* **107**, 13948-
769 13953, doi:10.1073/pnas.0911341107 (2010).

770 52 Sharma, U. *et al.* Biogenesis and function of tRNA fragments during sperm maturation
771 and fertilization in mammals. *Science* **351**, 391-396, doi:10.1126/science.aad6780 (2016).

772 53 Chen, Q. *et al.* Sperm tsRNAs contribute to intergenerational inheritance of an acquired
773 metabolic disorder. *Science* **351**, 397-400, doi:10.1126/science.aad7977 (2016).

774 54 Lin, T. Y., Mehta, R. & Glatt, S. Pseudouridines in RNAs: switching atoms means
775 shifting paradigms. *FEBS Lett* **595**, 2310-2322, doi:10.1002/1873-3468.14188 (2021).

776 55 Bollman, K. M. *et al.* HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates
777 phase change and morphogenesis. *Development* **130**, 1493-1504, doi:10.1242/dev.00362
778 (2003).

779 56 Grefen, C. & Blatt, M. R. A 2in1 cloning system enables ratiometric bimolecular
780 fluorescence complementation (rBiFC). *Biotechniques* **53**, 311-314,
781 doi:10.2144/000113941 (2012).

57 Li, X. Infiltration of Nicotiana benthamiana protocol for transient expression via Agrobacterium. *Bio-protocol*, e95-e95 (2011).

58 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

59 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

60 Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of small RNA. *Nat Protoc* **3**, 1077-1084, doi:10.1038/nprot.2008.67 (2008).

61 Hedhly, A., Vogler, H., Eichenberger, C. & Grossniklaus, U. Whole-mount Clearing and Staining of Arabidopsis Flower Organs and Siliques. *J Vis Exp*, doi:10.3791/56441 (2018).

62 Langmead, B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics* **Chapter 11**, Unit 11 17, doi:10.1002/0471250953.bi1107s32 (2010).

63 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

64 Marker, C. *et al.* Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant Arabidopsis thaliana. *Current Biology* **12** (2002).

65 Piekna-Przybylska, D., Decatur, W. A. & Fournier, M. J. The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic acids research* **36**, D178-D183 (2007).

66 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, 3 doi:10.14806/ej.17.1.200 (2011).

67 Zhang, Y. *et al.* DeAnnIso: a tool for online detection and annotation of isomiRs from small RNA sequencing data. *Nucleic Acids Res* **44**, W166-175, doi:10.1093/nar/gkw427 (2016).

68 Kaushik, A., Saraf, S., Mukherjee, S. K. & Gupta, D. miRMOD: a tool for identification and analysis of 5' and 3' miRNA modifications in Next Generation Sequencing small RNA data. *PeerJ* **3**, e1332, doi:10.7717/peerj.1332 (2015).

69 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

70 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

71 Coulombe, C. *et al.* VAP: a versatile aggregate profiler for efficient genome-wide data representation and discovery. *Nucleic Acids Res* **42**, W485-493, doi:10.1093/nar/gku302 (2014).

820 72 Schorn, A. J., Gutbrod, M. J., LeBlanc, C. & Martienssen, R. LTR-Retrotransposon
821 Control by tRNA-Derived Small RNAs. *Cell* **170**, 61-71 e11,
822 doi:10.1016/j.cell.2017.06.013 (2017).

823

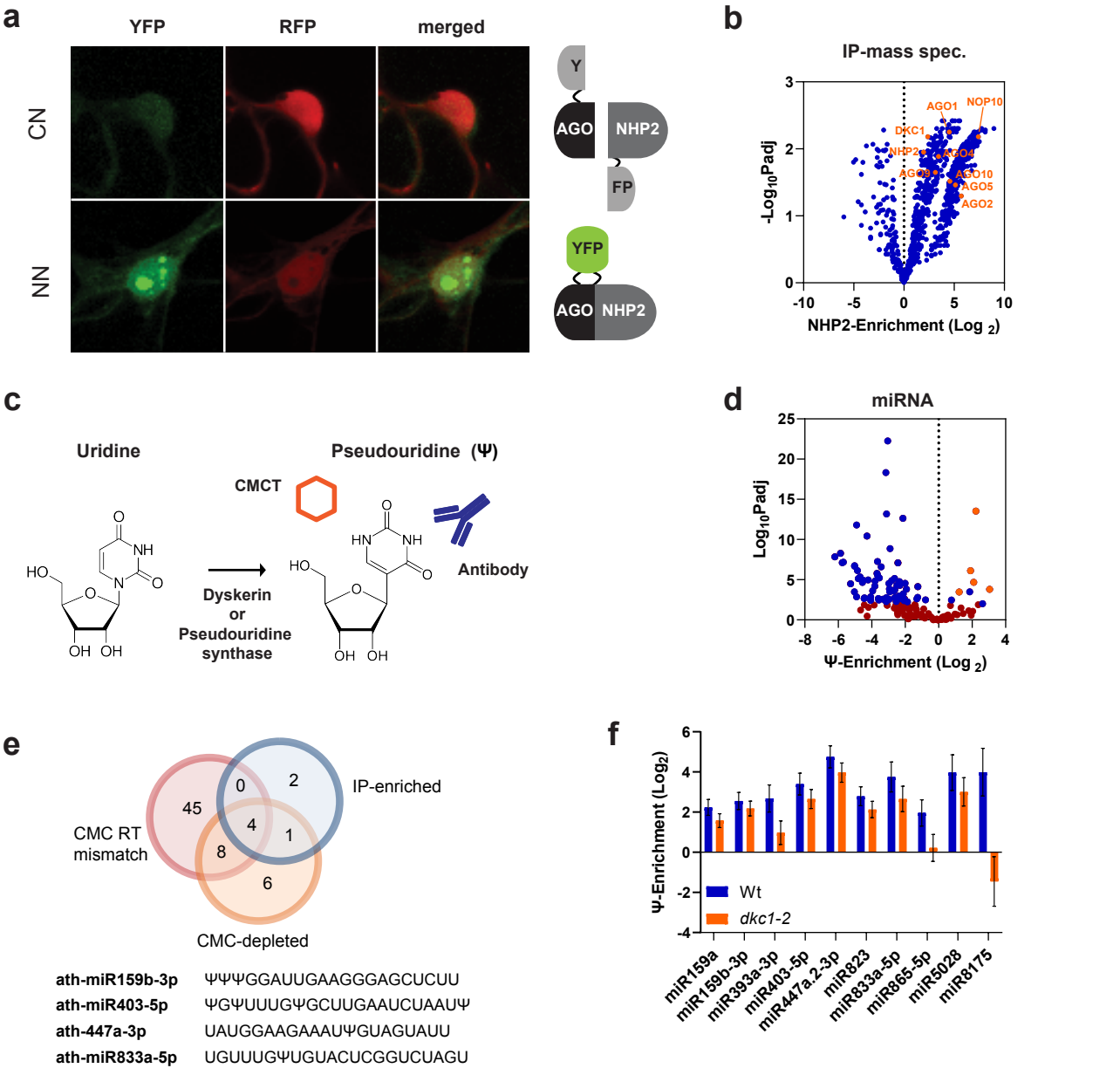

a
YFP
RFP
merged
CN
NN
Y
AGO
NHP2
FP
YFP
AGO
NHP2
b
IP-mass spec.
-Log10Padj
NHP2-Enrichment (Log 2)
AGO1
NOP10
DKC1
AGO10
AGO5
AGO2
c
Uridine
Pseudouridine (Ψ)
CMCT
Antibody
Dyskerin
or
Pseudouridine
synthase
d
miRNA
Log10Padj
Ψ-Enrichment (Log 2)
e
IP-enriched
CMC RT
mismatch
CMC-depleted
45
0
2
4
1
8
6
ath-miR159b-3p ΨΨΨGGAUUGAAGGGAGCUCUU
ath-miR403-5p ΨGΨUUUGΨGCUUGAAUCUAAUΨ
ath-447a-3p UAUGGAAGAAAUΨGUAGUAUU
ath-miR833a-5p UGUUUGΨUGUACUCGGUCUAGU
f
Ψ-Enrichment (Log2)
Wt
dkc1-2
miR159a
miR159b-3p
miR393a-3p
miR403-5p
miR447a.2-3p
miR823
miR833a-5p
miR865-5p
miR5028
miR8175
g
Deletion rate (%)
Position from 5' end
Position from 3' end
h
None
Ψ21
Ψ10
CMC
synthetic Ψ
spike-in
+ CMC
-CMC
Untreated
ath-miR159b

Figure 1

824 **Figure 1. microRNAs are pseudouridylated in plants and mammals.** (**a**) Bimolecular
825 fluorescence complementation confirming interaction of AGO3 with the Dyskerin co-factor NHP2
826 in Arabidopsis (Methods). Split YFP was fused to the C- and N-, or N- and N-terminals of AGO3
827 and NHP2. Reconstituted YFP signifies interaction. p35S::RFP acts as a transformation control.
828 (**b**) Volcano plot showing proteins co-purified with GFP-NHP2 as compared to IPs performed in
829 wild-type plants. H/ACA ribonucleoprotein complex subunits and AGO family proteins are
830 depicted in orange (p-value adjusted by Benjamini-Hochberg correction). **(c)** Structure of uridine
831 and pseudouridine (Ψ), enzymes involved in catalysis, and methods of detection with specific
832 antibodies and chemical modification with CMCT (Methods; Supplemental Fig. 2). (**d**) Volcano
833 plot of miRNAs enriched by Ψ-IP compared to unbound fractions from *Arabidopsis* flower buds
834 (DESeq2). Blue, $p<0.01$; orange, both Ψ-IP-enriched and depleted from libraries following CMC-
835 treatment ($p<0.01$); dark red, not significant. (**e**) Venn diagram showing overlap of miRNAs
836 detected in flower buds by each technique. Predicted Ψ sites in miRNAs detected by all 3
837 techniques. (**f**) miRNAs enriched by Ψ-IP in WT and *dkc1* mutants in Arabidopsis flower buds (**g**)
838 Rate of deletion in CMC/Mn^{2+} treated libraries at each site of miRNAs aligned at the 5′ (left panel)
839 and 3′ (right panel) ends. (**h**) Northern blots of synthetic miR159 RNA pseudouridylated at position
840 10 and 21 (upper) and flower bud small RNA (lower) treated (+) or mock treated (-) with CMC,
841 and probed with miR159b to reveal mobility shifts.

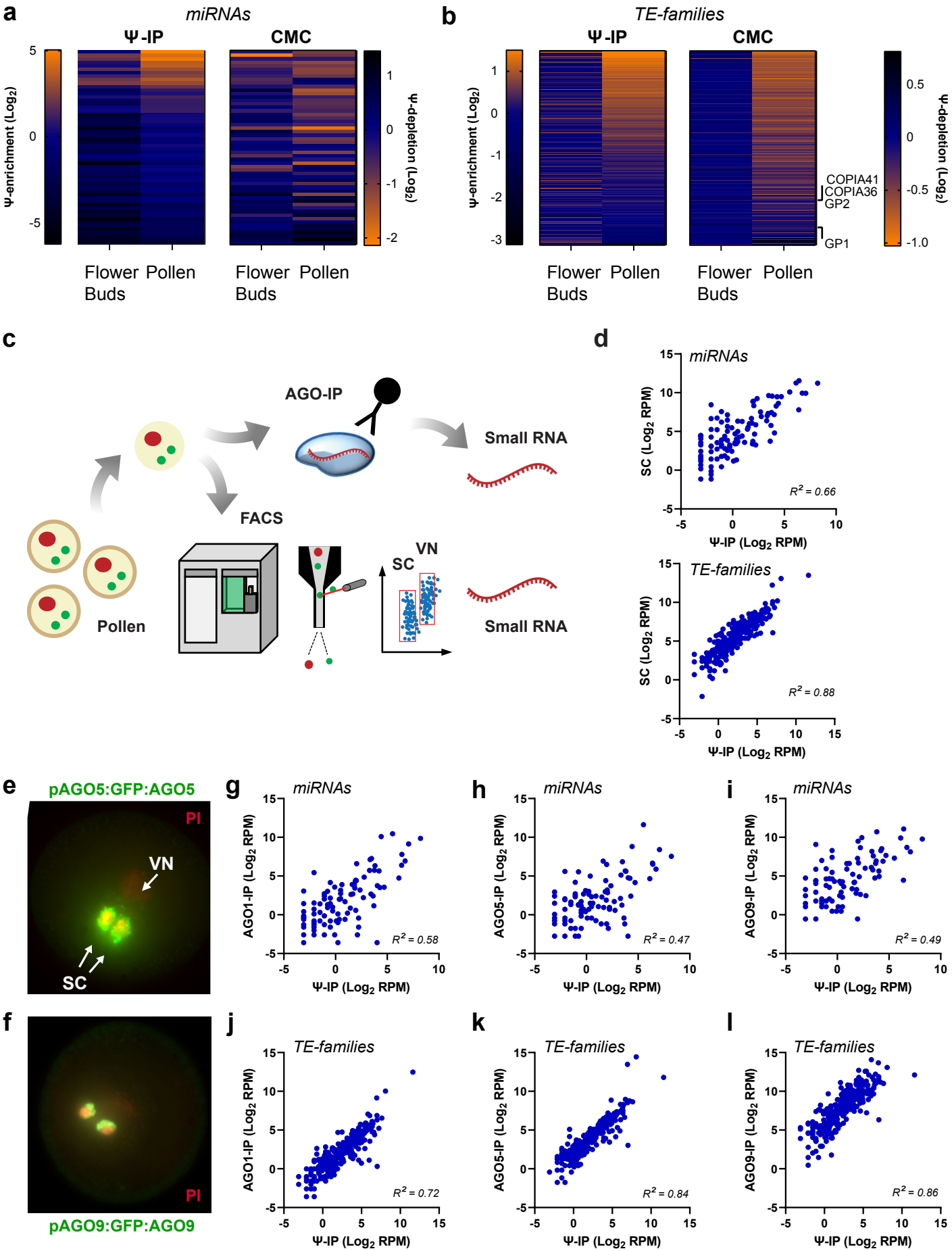

a
miRNAs
Ψ-IP
CMC
Ψ-enrichment (Log2)
Ψ-depletion (Log2)
Flower Buds
Pollen
b
TE-families
COPIA41
COPIA36
GP2
GP1
c
AGO-IP
Small RNA
FACS
Pollen
SC
VN
d
SC (Log2 RPM)
Ψ-IP (Log2 RPM)
R2 = 0.66
R2 = 0.88
e
pAGO5:GFP:AGO5
PI
VN
SC
f
pAGO9:GFP:AGO9
g
AGO1-IP (Log2 RPM)
R2 = 0.58
h
AGO5-IP (Log2 RPM)
R2 = 0.47
i
AGO9-IP (Log2 RPM)
R2 = 0.49
j
R2 = 0.72
k
R2 = 0.84
l
R2 = 0.86

Figure 2

842 **Fig. 2 Pollen small interfering RNAs are highly pseudouridylated and loaded into**
843 **ARGONAUTE proteins in sperm cells.** (**a**) Heatmaps of individual miRNA Ψ-enrichment in
844 flower buds and pollen by Ψ-IP (Log_2 IP/Unbound) and by CMC-mediated depletion from small
845 RNA libraries (Log_2 CMC+/CMC-) assessed by DESeq2. (**b**) Heatmaps of siRNA Ψ-enrichment
846 from individual TE families in flower buds and pollen by Ψ-IP and CMC-depletion as in (a). Small
847 RNAs from most TE families were enriched for Ψ in pollen, except for those highlighted. (**c)**
848 Small RNAs were sequenced from sperm cells isolated from pollen grains by FACS, and by RNA
849 immunoprecipitation (RIP-seq) with antibodies to ARGONAUTE (AGO) proteins. (**d**)
850 Association of pseudouridylation determined by Ψ-IP in pollen with sperm-cell abundance by
851 FACS for miRNAs and siRNAs from TE families. (**e)** GFP::AGO5 fusion protein expressed from
852 *AGO5* promoter (*pAGO5*) in mature pollen is found in the perinuclear space around sperm cells
853 (SC). DAPI was used to stain the vegetative nucleus (VN). Reproduced, with permission, from
854 (*12*) (**f)** AGO9::GFP fusion protein expressed from the *AGO9* promoter (*pAGO9*) in mature pollen.
855 (**g-i**) Association of pseudouridylation determined by Ψ-IP with AGO1 (**g, j**), AGO5 (**h, k**), and
856 AGO9 (**i, l**) binding measured by RIPseq for miRNAs (**g-i**) and siRNAs from TE families (**j-l**).
857 Pearson R^2 correlations are indicated.
858

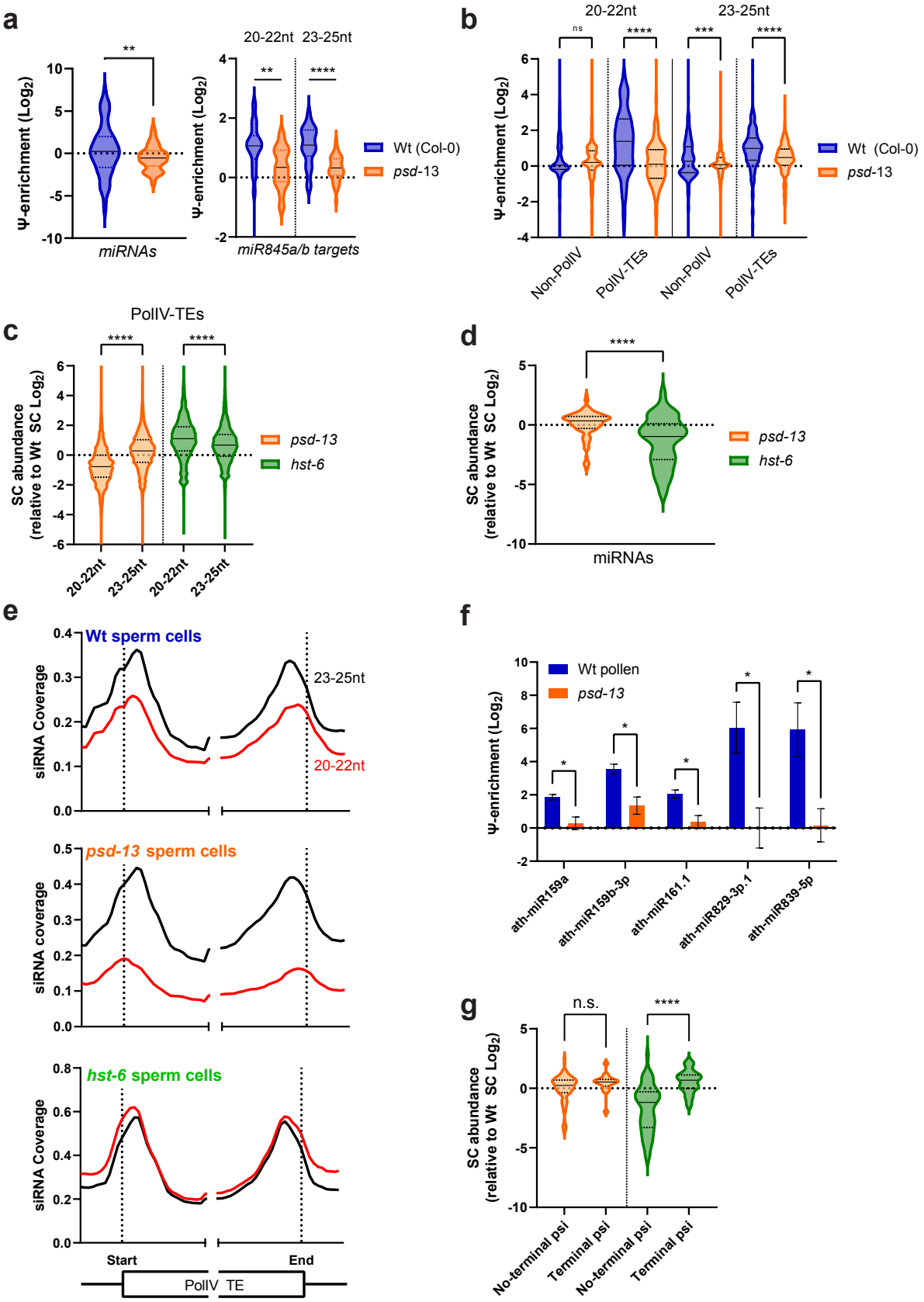

a
Ψ-enrichment (Log2)
miRNAs
20-22nt 23-25nt
miR845a/b targets
Wt (Col-0)
psd-13
b
20-22nt
23-25nt
Non-PolIV
PolIV-TEs
c
PolIV-TEs
SC abundance (relative to Wt SC Log2)
psd-13
hst-6
d
miRNAs
e
Wt sperm cells
psd-13 sperm cells
hst-6 sperm cells
siRNA Coverage
23-25nt
20-22nt
Start
End
PolIV TE
f
Wt pollen
ath-miR159a
ath-miR159b-3p
ath-miR161.1
ath-miR829-3p.1
ath-miR839-5p
g
n.s.
No-terminal psi
Terminal psi

Figure 3

859 **Fig. 3 Transport of pseudouridylated siRNAs into sperm cells depends on *PAUSED***
860 **(Exportin-t).** (**a**) Ψ-enrichment of miRNAs and some of their targets were determined by Ψ-IP
861 and small RNA sequencing in *psd-13* mutant pollen relative to wild-type. Ψ-enrichment of the
862 targets of miR845 were strongly dependent on *PSD* (paired t-test, miRNAs; one-way ANOVA,
863 miR845 targets) (**b**) Ψ-enrichment of Pol IV-dependent siRNA from "PolIV TEs" and other TEs
864 (non-PolIV) were determined by Ψ-IP and small RNA sequencing in *psd-13* mutant pollen relative
865 to wild-type ($p<0.01$, one-way ANOVA based on ranks). siRNA from PolIV TEs was strongly
866 dependent on *PSD.* (**c**) siRNA abundance was determined by sequencing small RNA from FACS
867 sorted sperm cells, and was reduced in *psd*, but increased in *hst* mutants relative to WT. Significant
868 differences between 20-22nt and 23-25nt siRNAs were observed (one-way ANOVA). (**d**) miRNA
869 abundance from FACS sorted sperm cells was reduced in *hst*, but not in *psd* mutant sperm cells
870 relative to wild type (one-way ANOVA). (**e**) Metaplots of 20-22 nt (red) and 23-25 nt (black)
871 siRNA abundance across "PolIV-TEs" from wild-type, *psd-13* and *hst-6* sperm cells sorted by
872 FACS. 250 bp up- and down-stream regions as well as 500 bp from each end of the TE are shown.
873 (**f**) Histogram of miRNA Ψ-enrichment in wild-type and *psd-13* pollen for those miRNA
874 dependent on *PSD* for pseudouridylation (student's t-test $p<0.05$; n = 3 biological replicates). (**g**)
875 Abundance of miRNA with and without terminal Ψ in *psd* and *hst* mutants relative to wildtype in
876 sperm cells sorted by FACS. miRNA without terminal Ψ were lost in *hst*, but not in *psd* sperm
877 cells ($p<0.001$). Asterisks indicate $p<0.05$ (*), 0.01 (**), 0.001 (***) and 0.0001(****).

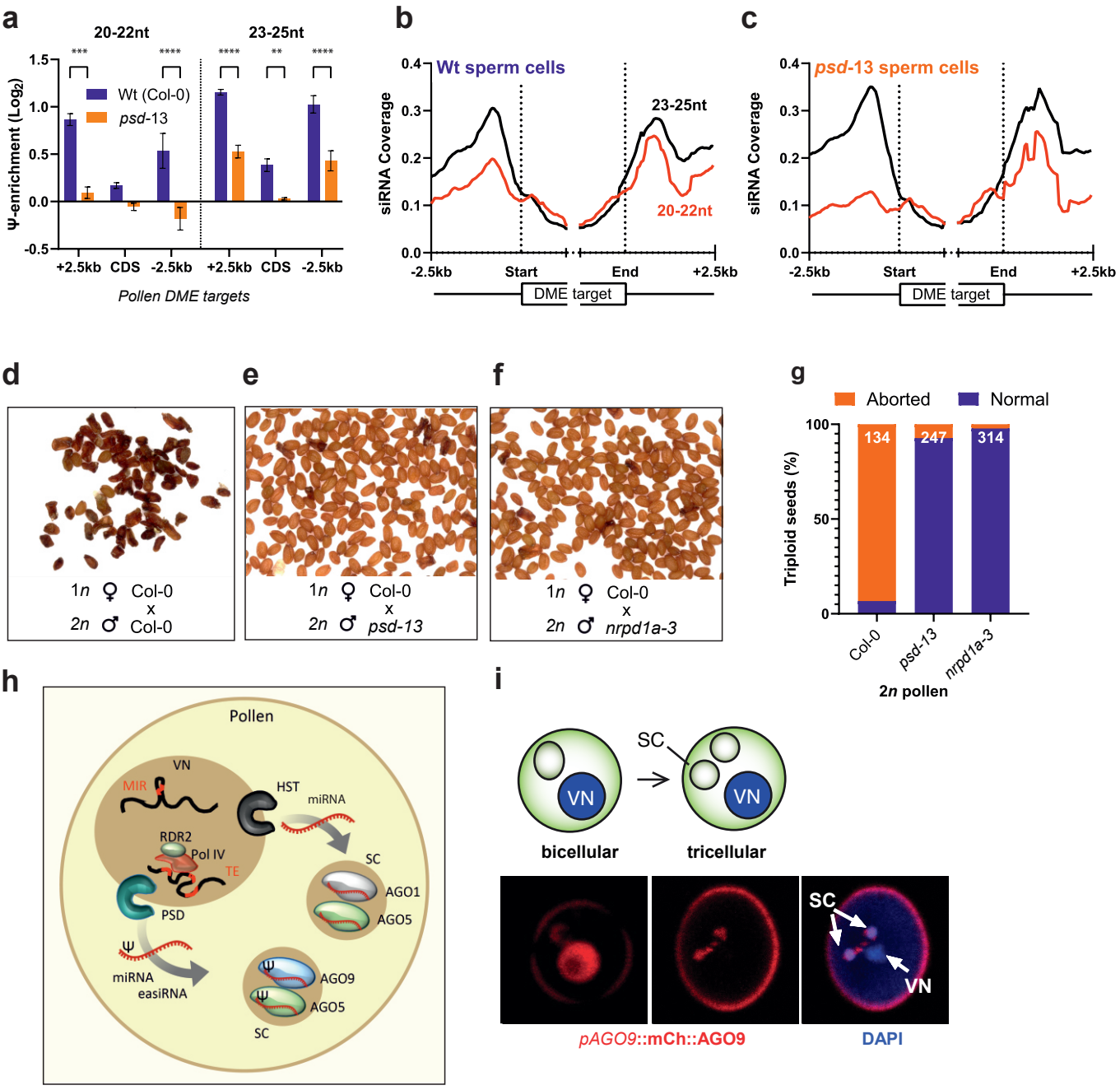
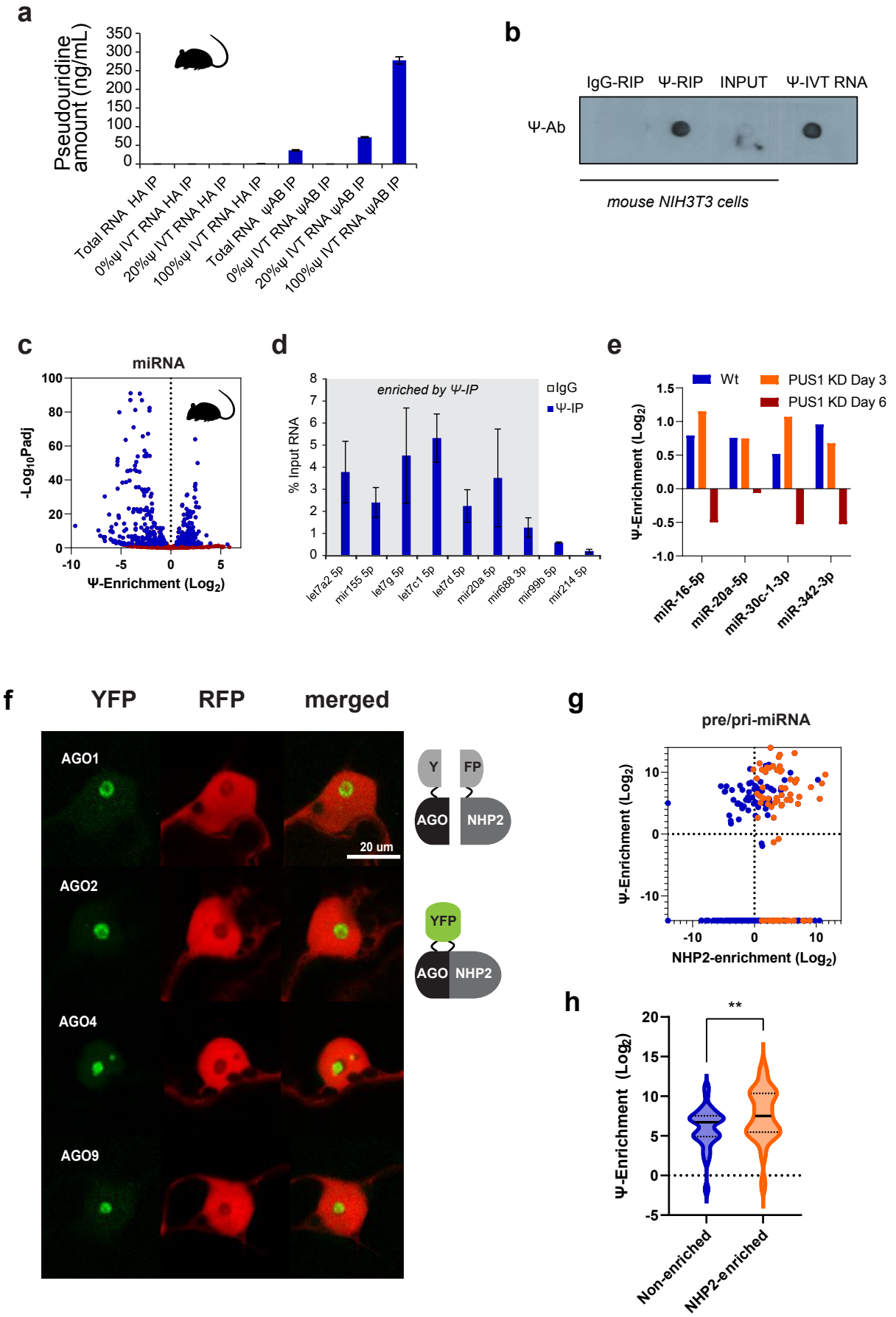

a
20-22nt
23-25nt
Ψ-enrichment (Log2)
Wt (Col-0)
psd-13
+2.5kb CDS -2.5kb +2.5kb CDS -2.5kb
Pollen DME targets
b
Wt sperm cells
siRNA Coverage
23-25nt
20-22nt
-2.5kb Start End +2.5kb
DME target
c
psd-13 sperm cells
siRNA Coverage
-2.5kb Start End +2.5kb
DME target
d
1n ♀ Col-0
x
2n ♂ Col-0
e
1n ♀ Col-0
x
2n ♂ psd-13
f
1n ♀ Col-0
x
2n ♂ nrpd1a-3
g
Aborted
Normal
Triploid seeds (%)
134
247
314
Col-0
psd-13
nrpd1a-3
2n pollen
h
Pollen
VN
MIR
HST
miRNA
RDR2
Pol IV
TE
SC
AGO1
AGO5
PSD
Ψ
miRNA
easiRNA
AGO9
AGO5
SC
i
SC
VN
VN
bicellular
tricellular
SC
VN
pAGO9::mCh::AGO9
DAPI

Figure 4

878 **Fig. 4. Pseudouridine and Exportin-t mediate epigenetic inheritance and the triploid block.**
879 (**a**) Ψ-enrichment (Log_2 Ψ-IP/unbound) of 20-22 nt siRNAs and 23-25 nt siRNAs surrounding
880 imprinted genes targeted by *DEMETER (DME),* in wild-type (blue) and *psd-13* (orange) pollen
881 Asterisks indicate student's t-test; $p<0.05$ (*), 0.01 (**), 0.001 (***) and 0.0001(****). (**b, c**)
882 sperm cell abundance of 20-22 nt siRNAs (red) and 23-25 nt siRNAs (black) surrounding
883 imprinted genes targeted by *DEMETER (DME),* in wild-type (**b**) and *psd-13* (**c**) pollen. (**d, e, f**)
884 Images of representative seeds from crosses between wild-type Col-0 female and (**d**) *osd1* diploid
885 pollen, (**e**) double mutant *osd1*; *psd-13* diploid pollen or (**f**) double mutant *osd1; nrpd1a-1* diploid
886 pollen. *NRPD1*a encodes the large subunit of Pol IV which is required for the triploid block (**g**)
887 Frequency of aborted seeds from the same crosses. (**h**) Model for transport of small RNA during
888 pollen development. As the pollen matures 21/22 nt easiRNAs and 24nt siRNAs are produced by
889 Pol IV in the vegetative nucleus and pseudouridylated along with some miRNAs. pseudouridylated
890 miRNAs and easiRNAs are exported via PSD (Exportin-t) into sperm cells, where they are loaded
891 onto AGO1, AGO5 and AGO9, while unmodified miRNAs are transported via HST (Exportin-5)
892 and loaded onto AGO5 and AGO1. Pseudouridylated sperm cell easiRNAs mediate dose-
893 dependent lethality (the triploid block) by targeting maternally expressed imprinted genes in the
894 seed. (**i**) mCherry::AGO9 fusion protein was expressed from the *AGO9* promoter (*pAGO9*) in
895 developing pollen. DAPI was used to stain nucleic acids (blue). AGO9 is present in the VN at the
896 binuclear stage, but is enriched in the sperm cells in mature pollen.

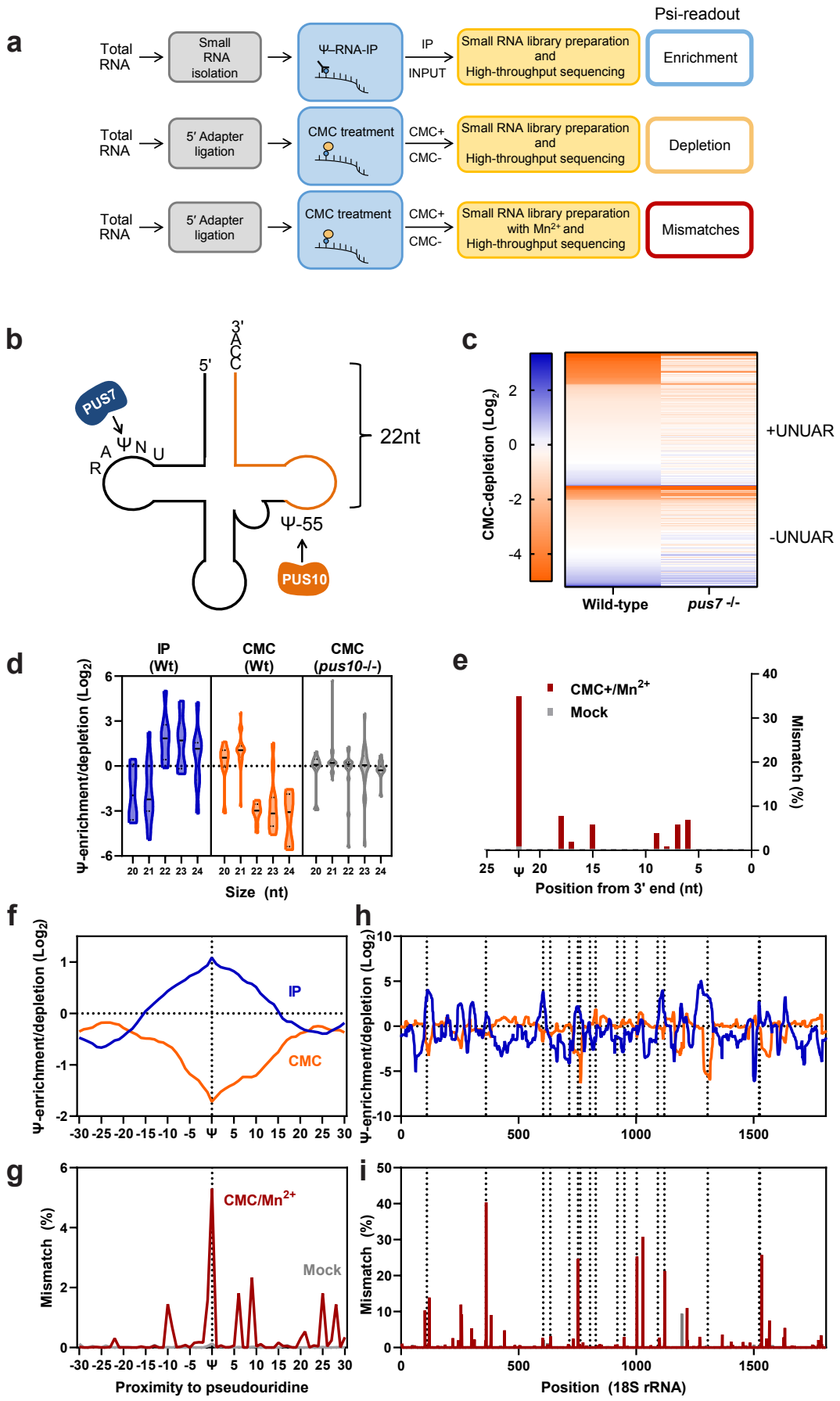
897 **Extended Data.**



Extended Data Figure 1

898 **Extended Data Figure 1. miRNA from plants and mammals are pseudouridylated.**

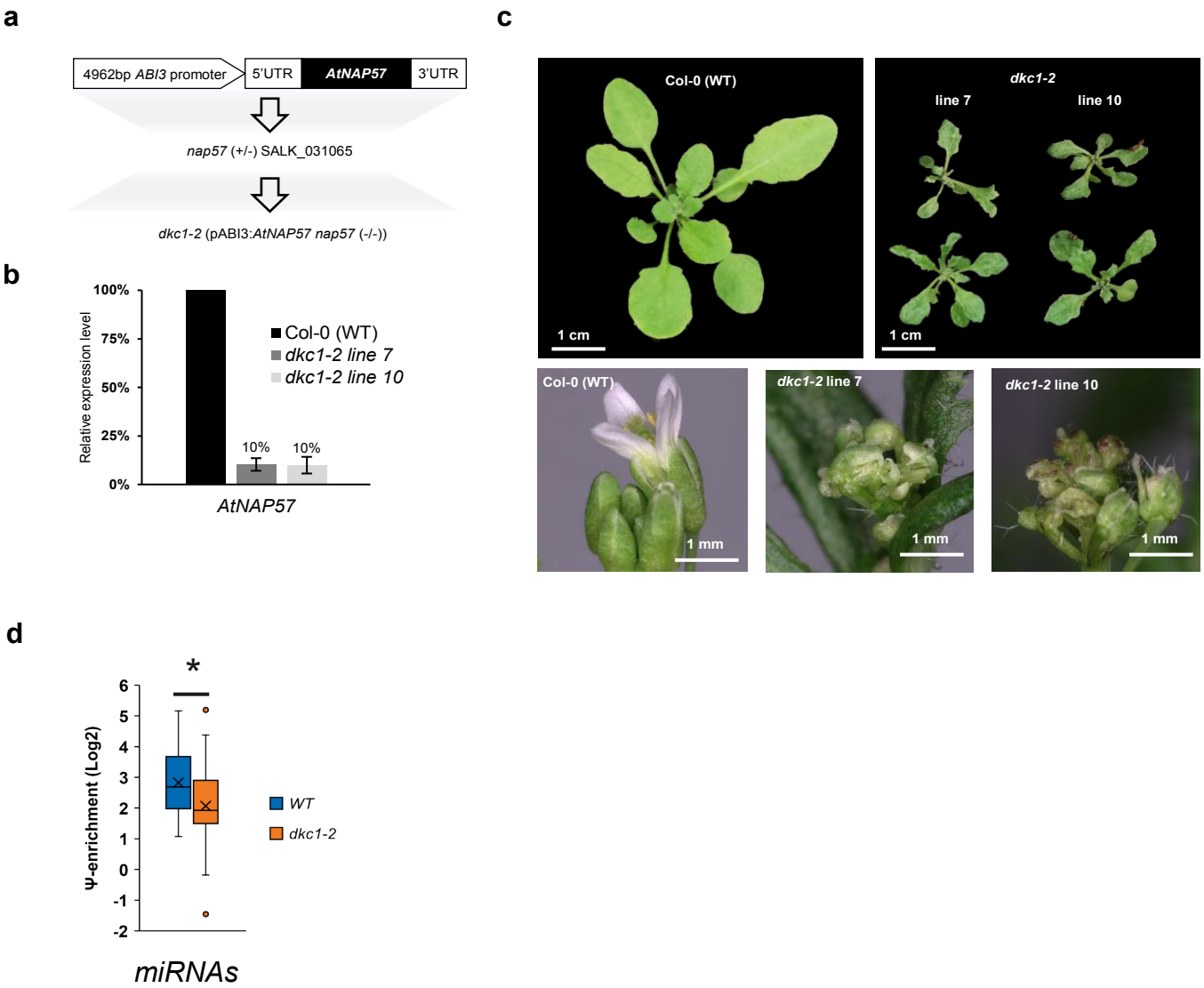
899 (**a**) Immunoprecipitation with anti-Ψ antibody enriches for Ψ-containing RNAs made by in vitro
900 transcription (IVT) as determined by mass spectrometry (n = 3 biological replicates, error bars ±
901 SD). (**b**) Immuno-dot blot of 100ng total input RNA or 100ng RNA immunoprecipitated with
902 anti-Ψ antibody or control immunoglobulin G (IgG). 100ng of 100% Ψ IVT RNA was used as
903 control. (**c**) Volcano plot of miRNAs enriched by small RNA immunoprecipitation using a Ψ-
904 specific antibody (Ψ-IP) in mouse fibroblast NIH/3T3 cells compared to input fractions.
905 Significantly enriched ($p<0.01$) miRNAs are highlighted in blue. (**d**) Validation of enriched
906 miRNAs using qRT-PCR. (**e**) Ψ-enrichment of miRNAs affected by PUS1 knockdown (KD) 3
907 days and 6 days after PUS1 shRNA transfection. (**f**) AGO proteins and Dyskerin subunit NHP2
908 were fused at the N-terminus to N- and C- terminally split YFP, respectively. Positive interaction
909 is shown by YFP signal, with 35S::RFP as an expression control. (**g**) Ψ-enrichment of precursor
910 pri-miRNAs in Arabidopsis flower buds plotted against pri-miRNAs bound to NHP2 (blue) and
911 detected by RIP-seq in 2 or more replicates (orange). (**h**) Violin plot showing significant
912 difference between NHP2-enriched and non-enriched pri-miRNA and Ψ enrichment of the
913 mature miRNA by Ψ -IP (Log_2 IP/IgG; (one-way ANOVA; $^{**}p<0.01$).



Extended Data Figure 2

914 **Extended Data Figure 2. Strategies for robust detection of Ψ in small RNAs.**

915 (**a**) Three strategies to detect Ψ in small RNAs and their predicted effect on pseudouridylated
916 sequences. Details are in Methods (**b**) Known sites of pseudouridylation in tRNAs mediated by
917 PUS7 (targets UNUAR) and PUS10 (targets Ψ-55). Ψ-55 is 22 nt "upstream" of the 3′ end of -
918 CCA tailed tRNA fragments. (**c**) Heatmap showing individual tRNA fragments (tRFs) with and
919 without the UNUAR motif in wild-type and *pus7*-/- libraries treated with CMC. (**d)** 3′CCA tRFs
920 ≥22 nt contain canonical Ψ-55 and are enriched by anti-Ψ antibody and depleted by CMC
921 treatment, whereas shorter fragments do not contain this nucleoside and are not enriched/depleted.
922 Ψ-55 is ablated in mutants of PUS10. (**e**) Deletions in 3′ CCA tRFs based on position in
923 CMC/Mn^{2+}- and mock-treated libraries. (**f, g)** Metaplots showing coverage depletion/enrichment
924 in CMC (orange)/IP (blue) libraries at rRNA pseudouridylation sites (f) compared to mismatches
925 from Mn2+/CMC+ libraries (g). Average across 42 predicted pseudouridylated sites in 18S and
926 25S rRNA and from 17 predicted sites in 18S rRNA (CMC mismatch/Mock mismatch). (**h, i**)
927 Enrichment/depletion/mismatches across 18S rRNA in CMC/IP (h) and CMC/Mn^{2+} (i) libraries.
928 Predicted Ψs are marked with dotted lines.



Extended Data Figure 3

929 **Extended Data Figure 3. Arabidopsis mutants in Dyskerin/ NAP57.**

930 (**a**) Generation of *dkc1-2* transgenic plants. *nap57* (SALK_031065) was complemented with the

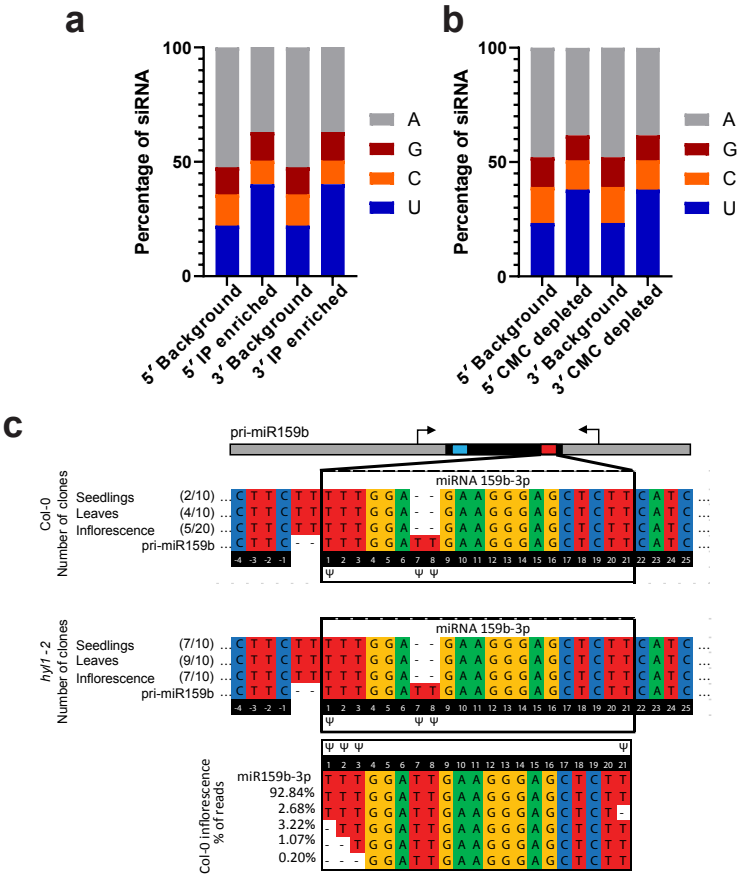
931 *pABI3:AtNAP57* transgene driven by embryo specific *ABI3* promoter (**b**) *AtNAP57* transcript

932 level in *dkc1-2* transgenic lines analyzed by qRT-PCR (n=3) (**c**) *dkc1-2* transgenic plants show

933 severe developmental defects including dwarfism, abnormal leaf shape, deformed and infertile

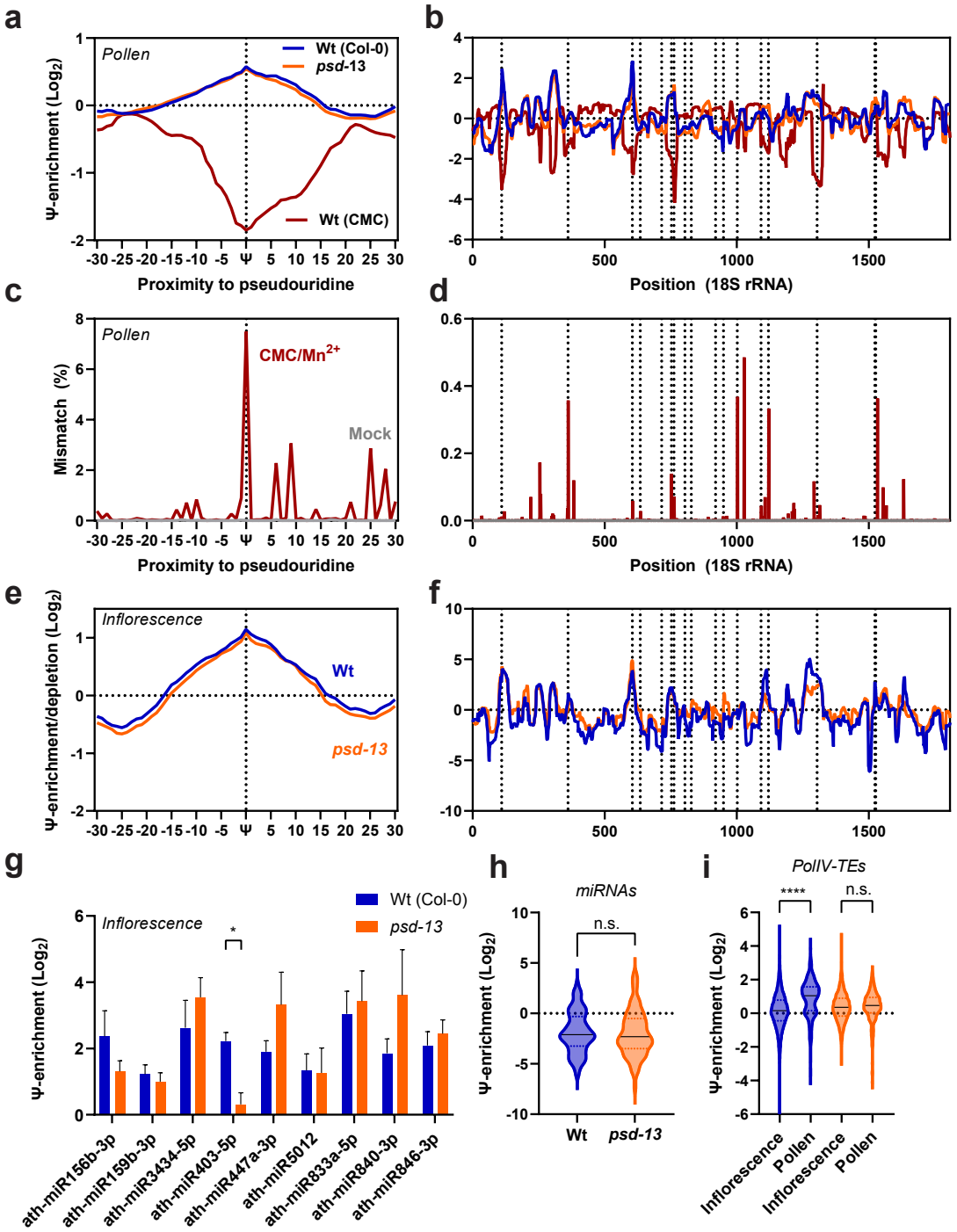
934 flowers. (**d**) Ψ-enrichment of pseudouridylated miRNAs in WT and *dkc1-2* mutants ($P<0.01$ one

935 way ANOVA).



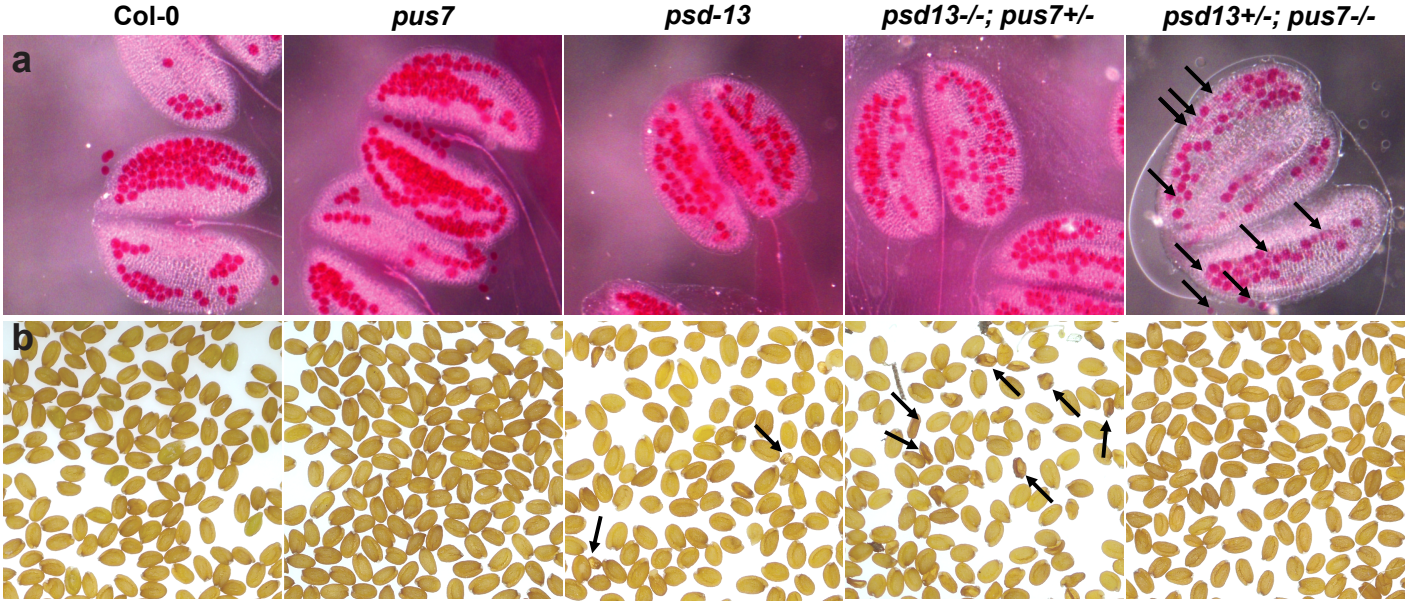
Extended Data Figure 4

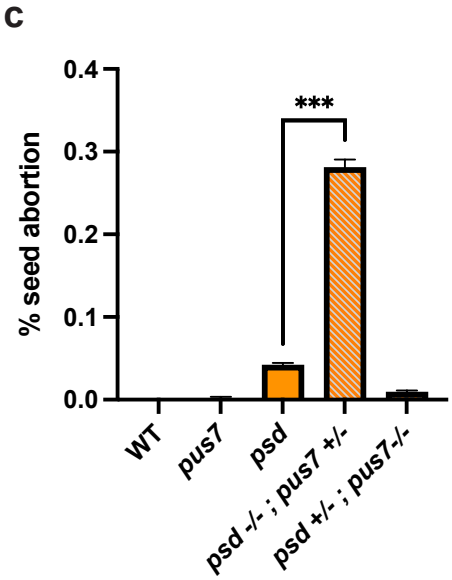
936 **Extended Data Figure 4. Mapping of miRNA pseudouridylation sites in Arabidopsis by**
937 **CMC labeling and reverse transcription in the presence of Mn^{2+} (CMC/Mn^{2+}).**
938 (**a**, **b)** Bar chart showing proportions of 5′ and 3′ nucleotides detected by CMC/Mn2+ mismatch
939 analysis in (**a**) IP-enriched and (**b**) CMC-depleted siRNAs from inflorescence compared to
940 background. IP-enrichment and CMC-depletion show biases towards 5′ and 3′ U. (**c**) Sites of
941 predicted Ψ in precursor pri-miR159b amplified from seedlings, leaves and inflorescence of wild-
942 type (Col-0) and *hyl1-2* microprocessor mutants. had insertions and deletions at varying
943 frequencies (number of clones after PCR sequencing) corresponding to 25% (Col-0) and 75%
944 (*hyl1-2)* overall. Terminal sites were also found in mature miR159b in inflorescence (lower panel).



Extended Data Figure 5

945 **Extended Data Figure 5. Pseudouridine in miRNAs and TE-derived siRNAs in pollen from**
946 **WT and *paused (psd-13)/*exportin-t mutants.**
947 (**a)** Metaplots of read coverage at 42 known pseudouridylated sites in 18S and 25S rRNA, relative
948 to untreated controls, in CMC-treated WT (magenta) and Ψ-IP WT (blue) and *psd-13* (orange)
949 small RNA sequencing libraries from pollen. (**b**) browser of read coverage from the same libraries
950 at 17 rRNA pseudouridylation sites in 18S rRNA. (**c, d**) metaplot and browser analysis of
951 predicted sites in rRNA from CMC/Mn^{2+} analysis relative to mock (**e, f**) Metaplots of read
952 coverage relative to untreated controls in CMC-treated WT (magenta) and WT (blue) and *psd-13*
953 (orange) Ψ-IP small RNA sequencing libraries from inflorescence. Known positions of Ψ in **b, c**
954 and **f** are marked with dotted lines. (**g**) Ψ-enrichment of individual miRNA detected by Ψ-IP and
955 small RNA sequencing in WT and *psd* mutant inflorescence (n=2 biological replicates) (**h**) Ψ-
956 enrichment of miRNA overall in WT and *psd* mutant inflorescence (one-way ANOVA, $p>0.05$)
957 (**i**) Violin plot of Ψ-enrichment of siRNA from individual TEs based on IP-enrichment (Log_2
958 IP/Unbound) in flower buds (inflorescence) and pollen from wt and *psd* mutants. Significant Ψ
959 enrichment was observed in pollen, but not in inflorescence, and depended on *PSD* ($p<0.001$, one-
960 way ANOVA).

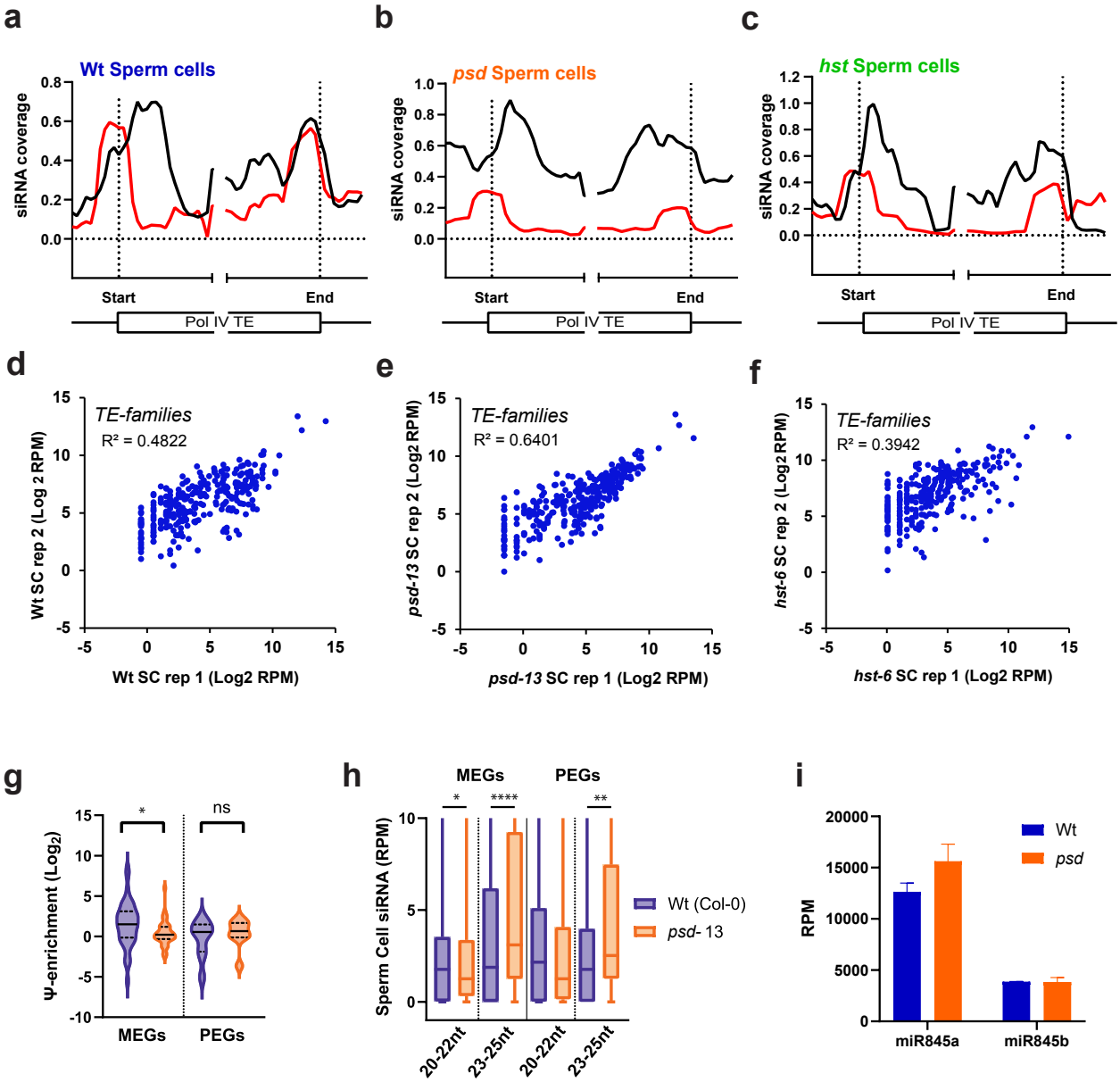




Extended Data Figure 6

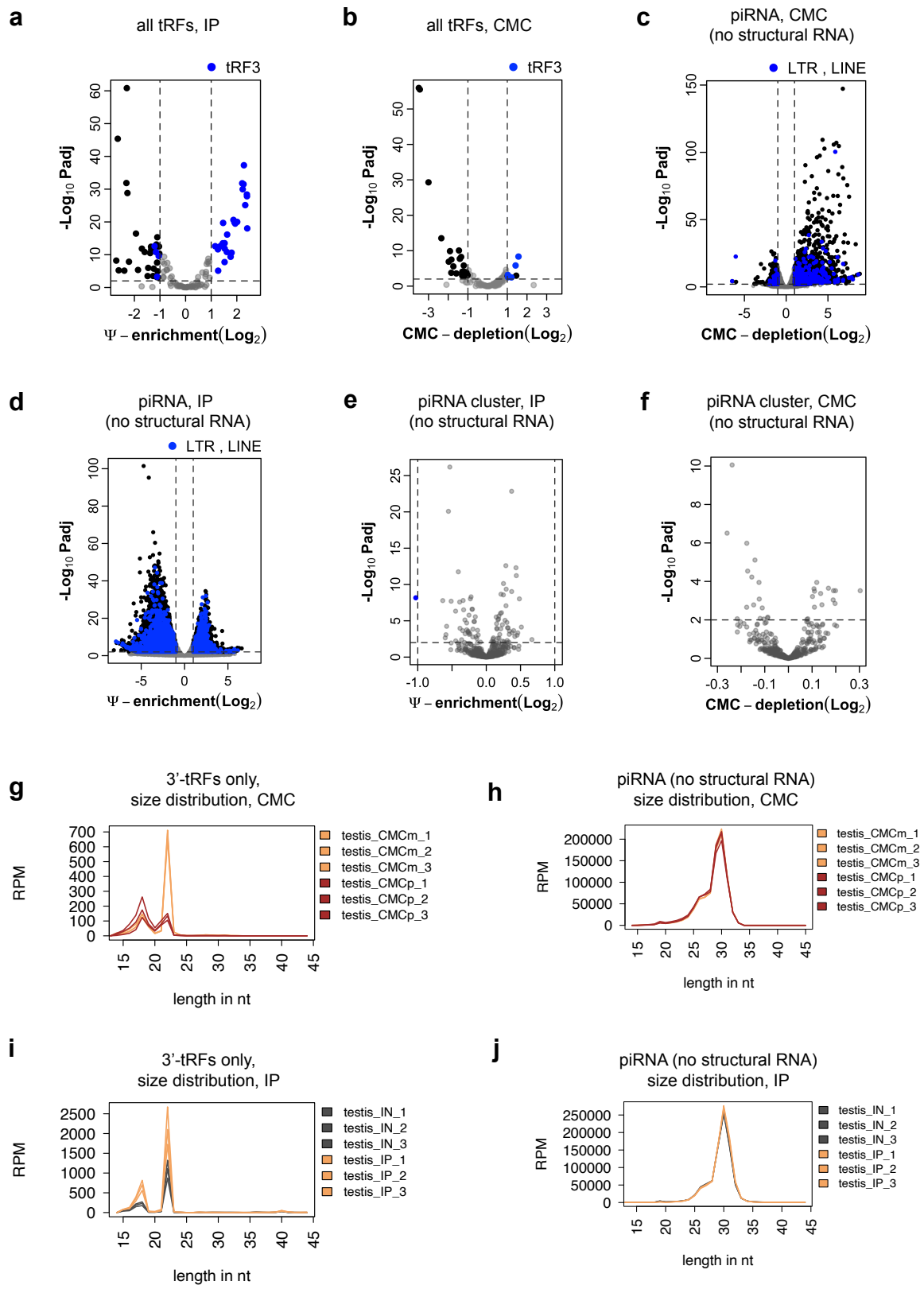
961 **Extended data Figure 6. Mutation of *psd* and *pus7* is synthetically lethal in Arabidopsis.**

962 (**a**) Alexander staining of pollen from Col-0, *pus7*, *psd-13*, *psd-13*-/-; *pus7*+/-, and *psd-13*+/-;
963 *pus7*-/- parents. *psd*-13+/; *pus7*-/- had semisterile pollen abortion (loss of Alexander staining,
964 arrows). (**b**) Mature seeds from each parent. Aborted seeds in *psd-13* are highlighted with arrows.
965 (**c**) Quantification of approximately 500 seeds from each genotype revealed *psd*-13-/-; *pus7*+/-
966 have 25% seed abortion, significantly more than *psd-13* alone (<1%; P<0.001). Double
967 homozygous mutants could not be obtained (n>150 plants). These results suggest that PSD acts in
968 haploid pollen, while PUS7-mediated pseudouridylation of tRNA acts in diploid pre-meiotic
969 tissue. Double homozygous embryos are lethal.



Extended Data Figure 7

970 **Extended Data Figure 7. Characterization of pseudouridylation of siRNA in *psd-13* mutants.**
971 **(a-c)** Metaplots of 20-22 nt (red) and 23-25 nt (black) siRNA abundance across transposable
972 elements transcribed by Pol IV ("Pol IV-TEs") from biological replicates of (**a**) wild-type, (**b**) *psd-*
973 *13* and (**c**) *hst-6* sperm cells sorted by FACS. 250 bp up- and down-stream regions as well as 500
974 bp from each end of the TE are shown. (**d-f**) Scatterplots of small RNA sequencing data from two
975 biological replicates mapped to TE families in (**d**) wild-type, (**e**) *psd-13* and (**f**) *hst-6* sperm cells
976 isolated by FACS. **(g)** Ψ-enrichment of siRNAs from TEs surrounding maternally expressed
977 imprinted genes (MEGs) and paternally expressed imprinted genes (PEGs) in wild type (blue) and
978 *psd-13* (orange) pollen. * = $p<0.05$, ANOVA. **(h)** abundance of 20-22 nt (red) and 23-25 nt (black)
979 siRNA abundance surrounding imprinted genes in wild-type (blue) and *psd-13* (orange) sperm
980 cells**.** siRNA in sperm cells matching MEGs and PEGs was lower in *psd-13* mutants for 21-22nt
981 siRNA ($P<0.05$) but relatively higher for 23-25nt siRNA ($P<0.01$). (**i**) levels of miR845a and
982 miR845b were unchanged or higher in *psd-13* mutant pollen.



Extended Data Figure 8

983 **Extended Data Fig. 8. Pseudouridine in piRNAs and tRNA fragments (tRFs) from mouse**
984 **testis**.
985 **(a, b)** Ψ enrichment by (a) Ψ-IP and (b) CMC-depletion of small RNA from testis from 8 weeks-
986 old male mice. Each dot corresponds to tRFs from a single tRNA anticodon isotype. Statistically
987 significant (p-adj. <0.01, log2 fold change > |1|) enriched/depleted tRFs are in black, 3'-tRFs are
988 colored blue; grey dots mark tRFs without significant enrichment/depletion. **(c)** Volcano plot of
989 individual piRNA sequences in testis (8 weeks old males) by CMC-depletion and sequencing.
990 Black: significant fold changes (p-adj. <0.01, log2 fold change > |1|); grey: not significant; blue:
991 significantly enriched/depleted piRNAs overlapping LTR and LINE transposon sequences. (**d**)
992 Volcano plot of individual piRNA sequences. Black: significant fold changes (p-adj. <0.01, log2
993 fold change > |1|); grey: not significant; blue: significantly enriched/depleted piRNAs overlapping
994 LTR and LINE transposon sequences. (**e**, **f**) Volcano plot of read counts per piRNA cluster using
995 (**e**) Ψ-IP and (**f**) CMC-detection (I). **(g, h)** Size distribution of (g) 3'-tRFs and (h) piRNA sequences
996 (without structural RNAs) using CMC-depletion and sequencing. **(i, j)** Size distribution of (i) 3'-
997 tRFs and (j) piRNA sequences (G, without structural RNAs) using Ψ-IP and sequencing.