



## **An ancient satellite repeat controls gene expression and embryonic development in *Aedes aegypti* through a highly conserved piRNA**

Rebecca Halbach, Pascal Miesen, Joep Joosten, Ezgi Taşköprü, Inge Rondeel, Bas Pennings, Chantal B Vogels, Sarah Merklings, Constantianus J Koenraadt, Louis Lambrechts, et al.

### **► To cite this version:**

Rebecca Halbach, Pascal Miesen, Joep Joosten, Ezgi Taşköprü, Inge Rondeel, et al.. An ancient satellite repeat controls gene expression and embryonic development in *Aedes aegypti* through a highly conserved piRNA. *Nature*, 2020, 580 (7802), pp.274-277. 10.1038/s41586-020-2159-2 . hal-02862278

**HAL Id: hal-02862278**

**<https://hal.science/hal-02862278v1>**

Submitted on 17 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

**An ancient satellite repeat controls gene expression and embryonic development in *Aedes aegypti* through a highly conserved piRNA**

**Rebecca Halbach<sup>1</sup>, Pascal Miesen<sup>1</sup>, Joep Joosten<sup>1</sup>, Ezgi Taşköprü<sup>1</sup>, Inge Rondeel<sup>1</sup>, Bas Pennings<sup>1</sup>, Chantal B.F. Vogels<sup>2,3</sup>, Sarah H. Merklings<sup>4,5</sup>, Constantianus J. Koenraadt<sup>2</sup>, Louis Lambrechts<sup>4,5</sup>, and Ronald P. van Rij<sup>1</sup>**

<sup>1</sup> Department of Medical Microbiology, Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands

<sup>2</sup> Laboratory of Entomology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

<sup>3</sup> Current affiliation: Department of Epidemiology of Microbial Diseases, Yale School of Public Health, 60 College Street, New Haven, CT 06510, USA

<sup>4</sup> Insect-Virus Interactions Group, Department of Genomes and Genetics, Institut Pasteur, 75015 Paris, France

<sup>5</sup> Evolutionary Genomics, Modeling and Health, Unité Mixte de Recherche 2000, Centre National de la Recherche Scientifique, 75015 Paris, France

**Correspondence:** ronald.vanrij@radboudumc.nl

## 28 **Abstract**

29 Tandem repeat elements such as the highly diverse class of satellite repeats occupy large parts of  
30 eukaryotic chromosomes. Most occur at (peri)centromeric and (sub)telomeric regions and have  
31 been implicated in chromosome organization, stabilization, and segregation<sup>1</sup>. Others are located  
32 more dispersed throughout the genome, but their functions remained largely unknown. Satellite  
33 repeats in euchromatic regions were hypothesized to regulate gene expression *in cis* by modulation  
34 of the local heterochromatin, or *in trans* via repeat-derived transcripts<sup>2,3</sup>. Yet, due to a lack of  
35 experimental models, gene regulatory potential of satellite repeats remains largely unexplored. Here  
36 we show that, in the vector mosquito *Aedes aegypti*, a satellite repeat promotes sequence-specific  
37 gene silencing via the expression of two abundant PIWI-interacting RNAs (piRNAs). Strikingly,  
38 whereas satellite repeats and piRNA sequences generally evolve extremely fast<sup>4-6</sup>, this locus was  
39 conserved for approximately 200 million years, suggesting a central function in mosquito biology.  
40 Tandem repeat-derived piRNA production commenced shortly after egg-laying and inactivation of  
41 the most abundant of the two piRNAs in early embryos resulted in an arrest of embryonic  
42 development. Transcriptional profiling in these embryos revealed the failure to degrade maternally  
43 provided transcripts that are normally cleared during maternal-to-zygotic transition. Our results  
44 reveal a novel mechanism by which satellite repeats regulate global gene expression *in trans* via  
45 piRNA-mediated gene silencing, which is fundamental to embryonic development. These findings  
46 highlight the regulatory potential of this enigmatic class of repeats.

## 63 Main

64 Even though satellite repeats have been discovered nearly 60 years ago<sup>7,8</sup>, and comprise a  
65 substantial portion of eukaryotic genomes, little is known about the functions of this class of  
66 repetitive DNA. Many satellite repeats are actively transcribed, and some of them produce small  
67 interfering (si)RNAs required for heterochromatin formation and maintenance<sup>9-16</sup>. Around two-  
68 thirds of the genome of *Aedes aegypti*, the most important vector for arthropod-borne viruses like  
69 dengue, Zika, and yellow fever virus, consists of repetitive elements<sup>17</sup> (Extended Data Fig 1A),  
70 making this mosquito an interesting model to study these sequences. We analysed small RNAs in  
71 somatic (female carcass) and germline (female ovaries) tissues as well as the somatic *Ae. aegypti*  
72 cell line Aag2. Even though satellite repeats constitute less than 10 % of the genome, they were not  
73 only highly covered by siRNAs, but especially by PIWI-interacting (pi)RNAs (Extended Data Fig  
74 1A). piRNAs are a class of small RNAs that protect animal genomes from harmful parasitic  
75 elements like transposons<sup>18</sup>. In the fruit fly *Drosophila melanogaster*, piRNAs are mostly derived  
76 from transposon-rich genomic regions termed piRNA clusters<sup>19</sup>, and predominately restricted to  
77 germline tissues<sup>19</sup>. In contrast, in *Ae. aegypti*, piRNAs are also highly abundant in non-germline  
78 tissues, and piRNAs from transposable elements (TEs) are underrepresented compared to the  
79 abundance of these elements in the genome<sup>20</sup>, especially in the soma (Extended Data Fig 1A).  
80 Instead, we found satellite-repeat derived piRNAs to be highly overrepresented in somatic tissues.  
81 Intriguingly, approximately three-quarters of satellite-repeat derived reads in the soma, and half of  
82 the reads in the germline and Aag2 cells stem from only two individual sequences that map to a  
83 repeat locus on chromosome 3. This locus is about 3.5 kb in size and consists of 20 full repeats and  
84 one disrupted repeat unit organized in a head-to-tail array (Fig 1A). These two highly abundant  
85 satellite-derived small RNAs were 30 and 29 nucleotides in size, respectively (Extended Data Fig  
86 1B), and resistant to  $\beta$ -elimination, suggesting that they are 2'-O-methylated at their 3' end, a  
87 common feature of mature PIWI-bound piRNAs<sup>21-23</sup> (Fig 1B). We named these two sequences  
88 tapiR1 and 2 (tandem repeat-associated piRNA1/2).

89 Expression of tapiR1 and 2 was ubiquitous in both somatic and germline tissues of adult mosquitoes  
90 (Extended Data Fig 2A). In *Ae. aegypti*, the PIWI gene family has expanded to seven PIWI genes  
91 (Piwi2-7 and Ago3) compared to three in flies<sup>24</sup>. Immunoprecipitation (IP) in Aag2 cells of the  
92 aedine PIWI proteins that are expressed both in the soma and gonads (Piwi4-6 and Ago3)<sup>25-27</sup>  
93 followed by northern blotting or deep sequencing indicates that both tapiR1 and tapiR2 exclusively  
94 associate with Piwi4 (Fig 1C, Extended Data Fig 2B, C, Supplementary Fig S1A,B). Indeed, only  
95 knockdown of Piwi4, but not of other PIWI or Argonaute (AGO)-clade genes reduced tapiR1 and 2  
96 levels (Extended Figure 2D,E, Supplementary Fig S1C,D). Thus far, the piRNA repertoire and  
97 function of Piwi4 remained unclear. Piwi4 neither associates with TE nor virus-derived piRNAs<sup>26</sup>,

100 yet was linked to piRNA biogenesis from transposons<sup>26</sup> and to antiviral defense<sup>28,29</sup>. As nearly 90 %  
101 of Piwi4-associated small RNAs only comprise tapiR1, and, to a much lower extent, tapiR2 (Fig  
102 1C), we hypothesize that tapiR1 not only dominates the piRNA repertoire, but also shapes  
103 downstream functions of Piwi4.

104 Satellite repeats are one of the fastest evolving parts of eukaryotic genomes. Except for a few  
105 examples<sup>30-32</sup>, most satellite repeats display high sequence divergence between species, and can  
106 even be species-specific<sup>6,33,34</sup>, akin to piRNAs<sup>4,5</sup>. Hence, we were surprised to find that the  
107 identified tandem repeat locus in *Ae. aegypti* is conserved in the closely related Asian tiger  
108 mosquito *Ae. albopictus*, and even in the more distantly related southern house mosquito *Culex*  
109 *quinquefasciatus* (Extended Data Fig 3A). This locus is, however, not present in the genome  
110 assembly of the malaria vector *Anopheles gambiae*. The tandem repeat locus differed in the number  
111 of monomers across species, and the monomers exhibited substantial length and sequence  
112 divergence, both between species and between monomers within one species. However, the  
113 sequences in the monomers that give rise to tapiR1 and 2, as well as a downstream T were by far  
114 more conserved than the overall monomer, suggesting that these sequences are under extensive  
115 selective constraints (Fig 1D, Extended Data Fig 3B-C). We further analysed whether expression of  
116 the two repeat locus-derived piRNAs is conserved among different mosquitoes, including species  
117 for which no genome assembly is available. We analysed 18 different mosquito species from 6  
118 different genera (*Aedes*, *Culex*, *Culiseta*, *Coquillettidia*, *Toxorhynchites*, and *Anopheles*), as well as  
119 *Culicoides nubeculosus*, a biting midge that also transmits arboviruses but is only distantly related  
120 to mosquitoes, and the fruit fly *Drosophila melanogaster*. Strikingly, even though piRNAs are  
121 usually not conserved even between closely related species<sup>4,5</sup>, we detected both tapiR1 and 2 in all  
122 five genera of the Culicinae subfamily of mosquitoes (Fig 1E, Extended Data Fig 3E). In line with  
123 the absence of the repeat locus in the *Anopheles gambiae* genome, we did not observe tapiR1 or 2  
124 expression in this subfamily of mosquitoes, nor in the two non-mosquito species. These  
125 observations suggest that the locus evolved in the late Triassic after divergence of the Anophelinae  
126 from the Culicinae subfamily of mosquitoes (229-192 mya<sup>35</sup>), but before further divergence of the  
127 culicine genera (226-172 mya<sup>35</sup>). This establishes this repeat locus as one of the very few ancient  
128 and deeply conserved satellite repeats that have hitherto been described<sup>30-32,36</sup>. Conservation of the  
129 locus over million years of mosquito evolution strongly suggests important and conserved functions  
130 for the locus and its associated piRNAs.

131 The satellite repeat locus overlaps with the 3'UTR of the gene AAEL017385 (LOC23687805) of  
132 unknown function in the current genome annotation (Fig 1A). This organization suggests that one  
133 or more splice variants of AAEL017385 are the source of the piRNAs, and that expression of this  
134 gene and piRNA biogenesis might be closely linked. However, knockdown of the different splice

isoforms did not reduce expression of tapiR1 (Extended Data Fig 3F, Supplementary Fig S1E), arguing against this possibility. Rapid amplification of 3' ends (RACE) of AAEL017385 transcripts revealed transcription termination sites directly upstream of the first tapiR1 and tapiR2 repeat, respectively (Extended Data Fig 3G). In addition, we detected a putative tapiR precursor only of ~500 nt which is much smaller than full-length AAEL017385 transcripts (~3.6-5.4 kb) (Extended Data Fig 3H), despite the whole repeat locus being covered by RNA sequencing reads (Extended Data Fig 3I). Even though we cannot exclude that this is a processing intermediate rather than the full-length precursor transcript, and that some AAEL017385 transcripts overlap with the satellite repeat, our data strongly suggest that the two piRNAs and AAEL017385 are expressed from the different transcriptional units. In support, the tapiR repeat locus is not associated with AAEL017385 orthologues in *Ae. albopictus* (AALF011179) and *Cx. quinquefasciatus* (CPIJ011773). We propose that the satellite repeat locus is transcribed from an unknown upstream or internal promoter.

We next characterized the sequence requirements for target silencing by tapiR1, as its expression is approximately one log higher compared to tapiR2 in Aag2 cells. Using a luciferase reporter harbouring a fully complementary target site in the 3' UTR, we validated that this piRNA is able to target RNAs *in trans*. The reporter was silenced more than 35-fold compared to a control reporter with a partially inverted target site (Fig 2A). Addition of an antisense oligonucleotide (AO) complementary to tapiR1, but not tapiR2, relieved silencing in a concentration-dependent manner (Extended Data Fig 4A), confirming that the observed effect is mediated by the piRNA in a sequence-specific fashion. Treatment with the tapiR1 AO did not relieve silencing of a tapiR2 reporter (Extended Data Fig 4A), and reduced only expression of tapiR1, but not tapiR2 (and *vice versa*) (Extended Data Fig 4B), indicating that the antisense oligonucleotides are specific to the mature piRNAs, and do not, or only to a minor extent, target the tapiR1/2 precursor. Additionally, AO treatment did not influence microRNA (miRNA) stability (Extended Data Fig 4B). Unlike most miRNAs<sup>37</sup>, silencing was not dependent on the position of the target site in the mature mRNA, and was efficient in the open reading frame and both the 5' and 3' UTR (Fig 2B, Extended Data Fig 5A). In contrast, silencing was strongly reduced when the target site was located in an intron (Fig 2B), or in a non-translated transcript expressed from a polymerase III promoter (Extended Data Fig 5B). During the course of our study we noticed that the firefly and *Renilla* luciferase genes, which we used as reporter and normalization control, respectively, contain potential target sites for tapiR1. Using reporter assays, we observed that *Renilla* luciferase is indeed potently suppressed by tapiR1 and firefly luciferase slightly (Extended Data Fig 5C-E), yet, mutating these target sites did not affect our conclusions (Extended Data Fig 5F,G).

167 To assess targeting requirements for tapiR1, we introduced mismatches in the piRNA target sites.  
168 Three consecutive mismatches were tolerated unless they were located in the t1 to t9 region of the  
169 target (the nucleotides expected to base-pair to piRNA positions 1 to 9) (Fig 2C, Extended Data  
170 Fig 6A), and single mismatches only impaired silencing at positions t3 to t7 (Extended data Fig  
171 6A,B), reminiscent of a microRNA seed<sup>37</sup> and comparable to the piRNA seed in *C. elegans*<sup>38,39</sup>. In  
172 contrast to observations with slicer (cleavage)-competent Argonautes including *Drosophila* PIWI  
173 proteins<sup>40,41</sup>, basepairing at the potential Argonaute cleavage site between nucleotides t10 and t11  
174 was dispensable for silencing.

175 A mismatch at position t1 did not alter silencing, suggesting that the first nucleotide is anchored in a  
176 binding pocket of Piwi4, similar to other Argonaute proteins<sup>42-44</sup>. Unexpectedly, a mismatch at  
177 position t2 was tolerated as well. This was, however, only the case when the rest of the target site  
178 was perfectly complementary, but not when the target contained mismatches outside of the seed  
179 (Extended Data Fig 6C). We further noticed that, in contrast to *C. elegans*<sup>38</sup>, G:U wobble pairs were  
180 not tolerated inside the seed and had the same effect as a mismatch at the same position (Extended  
181 Data Fig 6D). Whereas the 5' seed region is normally absolutely required for targeting<sup>38,39,41,42,45,46</sup>,  
182 the 3' part of the piRNA might increase specificity and efficiency of the targeting. For this reason  
183 we further assessed the extent of complementarity needed for tapiR1-mediated silencing.

184 Introduction of increasing numbers of mismatches at the 3' end did not interfere with silencing  
185 when at least half of the piRNA could basepair with the target site (Fig 2D), indicating that the  
186 3' part of the piRNA is not necessarily required, yet the seed region alone not sufficient for  
187 silencing.

188 Taken together, these results suggest that tapiR1 needs relatively low sequence complementarity to  
189 efficiently silence targets, a pattern that is more similar to target requirements of miRNAs than  
190 those of piRNAs studied thus far<sup>38,41,45,46</sup>. There are no constraints regarding the position of the  
191 target site on the mature mRNA, and as a consequence, tapiR1 may target a wide range of different  
192 cellular RNAs that perfectly basepair to the seed with additional matches outside the seed. Even  
193 though tapiR1 targeting requirements resemble those of miRNAs, the results are unlikely to be due  
194 to the piRNA being funnelled into the miRNA pathway. First, target silencing of different tapiR1  
195 reporters, or tapiR1 target genes *in vitro* and *in vivo*, is dependent on Piwi4, but not on Ago1  
196 (Extended Data Fig 8B,C, Extended Data Fig. 10H). Second, tapiR1 biogenesis itself is not  
197 dependent on Ago1 (Extended Data Fig 2E), and, third, tapiR1 is 2'-O-methylated (Fig 1C), a  
198 feature of siRNAs and mature PIWI-bound piRNAs, but not Ago1-associated miRNAs.

199 Given the extreme conservation of the tapiR repeat locus and the efficiency of target reporter  
200 silencing, we hypothesized that these piRNAs regulate cellular gene expression. Some satellite  
201 repeats can influence gene expression by modulation of the local chromatin environment *in cis*<sup>47,48</sup>,



202 or have been hypothesized to induce silencing of genes with homologous repeat insertions<sup>31</sup>. In  
203 contrast, the tapiR locus has the potential to silence expression of a broad range of remote genes *in*  
204 *trans*, independent of repeat insertions in target genes, and thus, to regulate diverse and highly  
205 complex cellular processes.

206 To test this idea, we used RNAHybrid<sup>49</sup> to predict tapiR1 target sites and verified these sequences  
207 in luciferase reporter assays. Twelve out of 23 sites in protein-coding genes, lncRNAs, or  
208 transposable elements were indeed sufficient to support suppression of the reporter (Extended Data  
209 Fig 7). Encouraged by these results, we then blocked tapiR1-mediated silencing with the tapiR1 AO  
210 in Aag2 cells and assessed global gene expression by RNAseq two days after treatment.

211 Intriguingly, expression of 134 genes, amongst which many long non-coding (lnc)RNAs, was  
212 significantly increased up to ~450 fold compared to the treatment with a control oligonucleotide  
213 (Fig 2E, Supplementary Table 1). Transposons were not globally affected, although some elements  
214 were up-regulated as well, up to around 850-fold (Extended Data Fig 8A, Supplementary Table 2).

215 Expression of differentially expressed genes and a transposable element was increased in a  
216 concentration-dependent manner upon tapiR1 AO treatment as measured by RT-qPCR (Fig 2F),  
217 validating our RNAseq results. As expected, knockdown of Piwi4, but not Ago1, led to  
218 derepression of target gene expression and reporters carrying cellular target sites (Extended Data  
219 Fig 8B,C). Yet, tapiR2 AO treatment did not relieve silencing of these target genes (Extended Data  
220 Fig 8D), confirming again that the tapiR1 AO is specific for the mature piRNA.

221 Our results indicate that tapiR1 directly and strongly regulates expression of a subset of cellular  
222 genes and transposons in a sequence-dependent manner. Moreover, these data indicate that tapiR1  
223 only needs limited basepairing to mediate silencing. Yet, computationally predicted tapiR1 target  
224 genes were not upregulated as a group upon tapiR1 AO treatment (Extended Data Fig 8E) and  
225 tapiR1 AO modulated genes were not enriched for predicted target sites. Moreover, the minimum  
226 free energy of piRNA/target duplexes was not predictive for effect sizes of tapiR1-mediated  
227 silencing, although we note that some of the strongest predicted targets were functional in a reporter  
228 context and among the genes with the strongest response upon tapiR1 AO treatment (Extended Data  
229 Fig 8F). Thus, similar to miRNAs<sup>50</sup>, factors beyond Watson-Crick base-pairing seem to be essential  
230 to define *bona fide* target genes for tapiR1.

231 Argonaute proteins in complex with their bound small RNAs repress target RNAs in a sequence-  
232 specific manner, however, the mechanism of silencing depends on the nature of Argonaute protein  
233 (Extended Data Fig 9A). Whereas metazoan miRNA-Argonaute complexes can interfere with  
234 translation, or destabilize their targets by inducing deadenylation and decapping, siRNA-Argonaute  
235 complexes, as well as PIWI proteins cleave (slice) their target RNAs. The cleavage products are  
236 then either quickly degraded by cellular RNA degradation pathways, or, in the case of the ping-



pong amplification loop in the piRNA pathway, processed into secondary piRNAs<sup>51</sup>. Furthermore, the *Drosophila* Piwi protein induces transcriptional silencing at target loci in the nucleus. The remarkably efficient silencing by Piwi4-tapiR1 raises questions about its mechanism of action. The ability of tapiR1 to efficiently silence mRNAs with target sites anywhere in the transcript but in introns (Fig 2B) suggest a post-transcriptional rather than a transcriptional mode of action. Most confirmed target sites harbor a bulge at the t10 or t11 sites when basepaired to tapiR1 (Extended Data Fig 7A), which is thought to sterically prohibit cleavage of the target<sup>43</sup>. Additionally, basepairing of these nucleotides is dispensable for silencing (Fig 2C, Extended Data Fig E6B), arguing against cleavage-induced silencing. In agreement, we were unable to detect a slice product that would be the result of a Piwi4 cleavage event for a representative tapiR1 target genes by 5' RACE, despite the fact that we readily detected slice product when targeting the same sites with siRNAs (Extended data Fig 9B). Additionally, we did not observe secondary piRNA generation for confirmed target genes (Extended Data Fig 9C), supporting the notion that Piwi4-tapiR1 complexes neither slice targets, nor are involved in the ping-pong amplification loop. We then tested the ability of tapiR1 to interfere with cap-dependent and cap-independent translation. Both a firefly luciferase reporter translated in a cap-dependent manner, as well as an Renilla Luc reporter translated from the cricket paralysis virus 5' internal ribosomal entry site (IRES) encoded on the same mRNA<sup>52</sup> were efficiently silenced to a similar extent (Extended Data Fig 9D). As miRNA-induced silencing complexes inhibit translation of cap-dependent, but not IRES-driven cap-independent reporters, these results distinguish the mechanism of Piwi4 from that of miRNA-mediated translational inhibition<sup>53-55</sup>. In line with that, tapiR1 AO treatment and Piwi4 knockdown resulted in substantial changes of target mRNA levels, as assessed by RNAseq and qPCR in Aag2 cells and *in vivo* in embryos (Figure 2E,F, Extended data Fig 7B-D, Figure 3E), arguing that Piwi4-tapiR1 complexes regulate RNA stability rather than interfere with protein translation. We thus performed an RNAi knockdown screen of genes encoding components of major cellular RNA decay pathways. However, only few factors had a minor effect on expression of a tapiR1 reporter (Extended Data Fig 9E), and none of these were reproducibly involved in silencing of established tapiR1 target genes, as measured by RT-qPCR (Extended Data Fig 9F, Supplementary Figure S1F,G). It is of note that tapiR1-mediated silencing was also independent of the CCR4/NOT deadenylase complex (Extended Data Fig 9E,F) that is important both for miRNA-mediated target silencing<sup>56-58</sup>, and piRNA-mediated RNA decay in *Drosophila melanogaster* embryos<sup>59</sup>. The CCR4/NOT complex catalyses removal of poly(A) from mRNAs which initiates subsequent degradation<sup>60</sup>. In line with that, we did not consistently observe elongation of poly(A) tails of target RNAs upon AO treatment (Extended Data Fig 9G-I), while being able to detect even the small elongation of Sindbis virus (SINV) RNA upon infection in cells (from 37nt to a mean

length of 70nt)<sup>61</sup> (Extended Data Fig. 9J). Together, these results indicate that Piwi4-tapiR1 efficiently destabilize target mRNA at the post-transcriptional level, but the exact mechanism of silencing remains elusive.

Expression of satellite repeats is often regulated in a developmental or stage-specific manner<sup>31,62</sup>, thus we analysed the expression pattern of tapiR1 and 2 throughout the mosquito life cycle. tapiR1 and 2 were not expressed during the first three hours of embryonic development of *Ae. aegypti* (Fig 3A), but could be detected in all subsequent life stages (Extended Data Fig 10A,B). Likewise, Piwi4 is expressed in all analysed tissues, but is already present in early embryos (Extended Data Fig 10C,D). The very beginning of embryonic development is characterized by a transcriptionally quiescent zygotic genome, and the first mitotic divisions are exclusively driven by maternally deposited transcripts and proteins. Maternal-to-zygotic transition (MZT) is marked by the degradation of these maternal transcripts and concomitant zygotic genome activation<sup>63</sup>.

Destabilization of maternal transcripts initially occurs through maternal decay activities, and later, after onset of zygotic transcription also by zygotic components<sup>63,64</sup>. One of the decay mechanisms involves zygotically expressed miRNAs, for example miR-430, miR-427, and the miR-309 cluster in zebrafish<sup>65</sup>, *Xenopus*<sup>66</sup>, and flies<sup>67</sup>, respectively. Since tapiR1 is expressed early during zygotic genome activation in *Ae. aegypti* embryos (Extended Data Fig 10E, Supplementary Fig S1H) and has strong gene silencing potential, we hypothesized that tapiR1 is involved in the zygotic degradation pathway and necessary for embryonic development. To test this idea we injected either tapiR1-specific AO or control AO into early pre-blastoderm *Ae. aegypti* embryos (before zygotic genome activation and expression of tapiR1) (Fig 3B), and assessed their development using a discrete scoring scheme (Supplementary Fig 1H). Strikingly, more than 90 % of all tapiR1 AO-injected embryos were arrested early in development, whereas about half of all control embryos showed obvious signs of developmental progression (Fig 3C). In accordance, only a small fraction of tapiR1 AO-injected embryos hatched (Fig 3D) and continued to develop as larvae, indicating that tapiR1-deficiency impedes development. RNA sequencing from tapiR1 AO or control AO-injected embryos 20.5 h after injection revealed massive deregulation of cellular transcripts (Fig 3E, Extended Data Fig 10F, Supplementary Table 3, 4). Expression of 205 genes, among which 44 lncRNAs, as well as few transposable elements was increased up to around 1000 and 500 fold in tapiR1 AO-treated embryos, respectively. We confirmed these results, and the dependency of these genes on Piwi4 for a selection of target genes by RT-qPCR upon tapiR1 AO and Piwi4 dsRNA injection (Extended Data Fig 10G,H). Target genes with predicted tapiR1 target sites were more strongly up-regulated than genes without predicted site, as indicated by a shift of the cumulative distribution of RNA fold changes (Fig 3F). Additionally, a subset of these genes was also found to be deregulated in tapiR1 AO-treated Aag2 cells (Extended Data Fig 10I). These findings show that

307 tapiR1 controls regulatory circuits by direct gene targeting *in vivo*, and that this function is essential  
308 for embryonic development, likely by promoting mRNA turnover of a subset of maternal  
309 transcripts. In line with this conclusion, confirmed target genes are down-regulated after the onset  
310 of tapiR1 expression (Fig 3G), and tapiR1 targets are overrepresented in transcripts that are  
311 maternally provided and degraded during MZT (Fig 3H).

312 piRNAs have been shown to promote degradation of *nanos*<sup>59</sup> and other transcripts involved in germ  
313 cell development<sup>68</sup>. Yet, this was dependent on transposon-derived piRNAs, and rather depicts a re-  
314 purposing of the existing piRNA pool, which is, due to its large targeting potential, ideal to be used  
315 to degrade a large number of transcripts. In contrast, we propose that, analogous to abundant  
316 miRNAs in other animals<sup>65-67</sup>, Culicinae mosquitoes have evolved a specific piRNA to destabilize a  
317 defined set of maternally deposited transcripts in early embryonic development. To our knowledge,  
318 this is the first example of sequence-specific gene silencing by transcriptional products from a  
319 satellite repeat *in trans*, and underlines the regulatory potential of tandemly repeated DNA.

320

## 321 **Methods**

### 322 **Cell culture**

323 *Aedes aegypti* Aag2 cells were cultured in Leibovitz's L-15 medium (Invitrogen) supplemented with  
324 10 % heat-inactivated Fetal Bovine Serum (PAA Laboratories), 2 % Tryptose Phosphate Broth  
325 Solution (Sigma Aldrich), 1x MEM Non-Essential Amino Acids (Invitrogen) and 50 U/ml  
326 penicillin/ streptomycin (Invitrogen) at 25 °C. Aag2 cells are a widely-used non-clonal cell line of  
327 probably embryonic origin<sup>69</sup> that expresses all somatic aedine PIWI proteins and produces both  
328 primary and secondary piRNAs via ping-pong dependent amplification<sup>25,26</sup>. For all experiments,  
329 cells were seeded the day before and used at 70-80 % confluency.

330

### 331 **Mosquito rearing**

332 Injections of embryos for RNA sequencing and northern blotting presented in Fig 3A,G were  
333 performed using a Cell fusion agent virus-free, isofemale *Aedes aegypti* strain called Jane. This  
334 strain was initiated from a field population originally sampled in the Muang District of Kamphaeng  
335 Phet Province, Thailand<sup>70</sup>, and reared for 26 generations at 28±1°C, 75±5 % relative humidity,  
336 12:12 hour light-dark cycle. Embryos were hatched under low pressure for 30-60 min. Larvae were  
337 grown in dechlorinated tap water and fed fish food powder (Tetramin) every two days. Adults were  
338 maintained in cages with constant access to a 10 % sucrose solution. Female mosquitoes were fed  
339 on commercial rabbit blood (BCL) through a membrane feeding system (Hemotek Ltd.) using pig  
340 intestine as membrane. For AO injections, female mosquitoes were transferred to 25 °C and 70 %  
341 humidity for at least two days before forced to lay eggs, and embryos were then placed back to  
342 28 °C immediately after the injection. For the time-course experiment in Fig. 3A,G, embryos were  
343 kept at 25 °C during the course of the experiment. All other injection experiments were conducted  
344 with *Ae. aegypti* Liverpool strain, reared at 28 ±1 °C, 70 % humidity as described above, however,  
345 fed on human blood (Sanquin Blood Supply Foundation, Nijmegen, The Netherlands), and  
346 maintained at 28 °C throughout experiments.

347 Experiments presented in Extended data Fig 2A, and Extended data Fig 10A,B were performed with  
348 the *Ae. aegypti* Rockefeller strain, obtained from Bayer AG, Monheim, Germany. The mosquitoes  
349 were maintained at 27±1 °C with 12:12 hour light:dark cycle and 70 % relative humidity, as  
350 described before<sup>71</sup>. For the bloodfeeding experiment mosquitoes were offered human blood  
351 (Sanquin Blood Supply Foundation, Nijmegen, The Netherlands), and five engorged females were  
352 selected and sacrificed at each of the indicated time points.

353 Mosquitoes used in Fig 1F were either different laboratory-reared, or wild-caught species: *Aedes*  
354 *aegypti* Liverpool strain, *Culex pipiens*, *Anopheles coluzzii*, *An. quadriannulatus*, *An. stephensi*  
355 mosquitoes, *Culicoides nubeculosus* biting midges, and *D. melanogaster* w<sup>1118</sup> flies were laboratory

356 strains. The mosquitoes were deep-frozen and stored at -80 °C until use. *Ae. albopictus*, *Ae.*  
357 *cantans*, *Ae. intrudens*, *Ae. pullatus*, *Ae. cinereus*, *Ae. vexans*, *Cx. pusillus*, *Culiseta morsitans*,  
358 *Coquillettidia richiardii*, *An. maculipennis*, *An. claviger*, and *An. coluzzii* were wild-caught  
359 individuals collected in different regions in Italy, Sweden, or the Netherlands between July 2014  
360 and June 2015<sup>72</sup>. Species were identified at the species level, and stored at -20 °C for a maximum of  
361 two years.

362

### 363 **Gene knockdown**

364 Double-stranded RNA was generated by *in vitro* transcription of T7 promoter-flanked PCR  
365 products with T7 RNA polymerase. Primer sequences are given in Supplementary Table S5. The  
366 reaction was carried out at 37°C for 3 to 4 h, then heated to 80 °C for 10 min and gradually cooled  
367 down to room temperature to facilitate dsRNA formation. The dsRNA was purified with the  
368 GeneElute Total RNA Miniprep Kit (Sigma Aldrich).

369 Aag2 cells were seeded in 24-well plates the day before the experiment, and transfected with  
370 X- tremeGENE HP Transfection reagent (Roche) according to the manufacturer's instructions,  
371 using a ratio of 4 µL reagent per µg of dsRNA. The transfection medium was replaced after 3 h with  
372 fully supplemented Leibovitz-15 medium. For Extended Data Fig. 2D,E, Extended Data Fig 3F, and  
373 Extended Data Fig 8B, cells were harvested after 48h; in other experiments, the knockdown was  
374 repeated 48 h after the first transfection and cells were then harvested after 24 h. Knockdown was  
375 confirmed by RT-qPCR. For the RNAi screen shown in Extended Data Fig 9F, a luciferase reporter  
376 harbouring a target site of AAEL001555 in the 3' UTR, and a *Renilla* luciferase reporter were co-  
377 transfected with the dsRNA during the second transfection.

378 For injection of embryos with dsRNA, engorged female *Ae. aegypti* Liverpool mosquitoes were  
379 allowed to lay eggs for 45 min. Embryos were desiccated for 1.5 min, covered with Halocarbon oil  
380 (Sigma Aldrich) and injected with 500ng/µL dsRNA with the Pneumatic PicoPump PV820 (World  
381 Precision Instruments) with 30 psi inject pressure. Injected embryos were then transferred to a wet  
382 Whatman paper and kept at 28 °C and 80 % humidity for 21 h. Per experiment, 30-60 embryos  
383 were injected per condition.

384

### 385 **RNA isolation**

386 RNA from cells and mosquitoes was isolated with Isol-RNA lysis buffer (5PRIME) according to  
387 the manufacturer's instructions. Briefly, 200 µL chloroform was added to 1 mL lysis buffer, and  
388 centrifuged at 16,060 x g for 20 min at 4 °C. Isopropanol was added to the aqueous phase, followed  
389 by incubation on ice for at least one hour, and centrifugation at 16,060 x g for 10 min at 4 °C. The

390 pellet was washed three to five times with 85 % ethanol and dissolved in RNase free water. RNA  
391 was quantified on a Nanodrop photospectrometer.

392

### 393 **Periodate treatment and $\beta$ -elimination**

394 Total RNA was treated with 25 mM NaIO<sub>4</sub> in a final concentration of 60 mM borax and 60 mM  
395 boric acid (pH 8.6) for 30 min at room temperature. In the control, NaIO<sub>4</sub> was replaced by an equal  
396 volume of water. The reaction was quenched with glycerol and  $\beta$ -elimination was induced with a  
397 final concentration of 40 mM NaCl for 90 min at 45 °C. RNA was ethanol precipitated and  
398 analysed by northern blot.

399

### 400 **Generation of antibodies**

401 Custom-made antibodies (Eurogentec) against endogenous PIWI proteins were generated by  
402 immunization of two rabbits per antibody with a mix of two unique peptides (Ago3:  
403 TSGADSSSEDDKQSS, IYKRKQRMSENIQF; Piwi4: HEGRGSPSSRPAYSS,  
404 HHRESSAGGRERSGN; Piwi5: DIVRSRPLDSKVVKQ, CANQGGNWRDNYKRAI; Piwi6:  
405 MADNPQEGSSGGRIR, RGDHRQKPYDRPEQS). After 87 days and a total of four  
406 immunizations (t=0, 14, 28, 56 days), sera of both rabbits were collected, pooled, and purified  
407 against each peptide separately. Specificity of the antibody was confirmed by Western blotting of  
408 Aag2 cells stably expressing PTH (Protein A, TEV cleavage site, 6x His-tag)-tagged PIWI<sup>27</sup> upon  
409 knockdown of the respective PIWI protein, or a control knockdown (dsRLuc).

410

### 411 **Immunoprecipitation and western blotting**

412 Aag2 cells were lysed with RIPA buffer (10 mM Tris-HCl, 150 mM NaCl, 0.5 mM EDTA, 0.1 %  
413 SDS, 1 % Triton-X-100, 10 % DOC, 1x protease inhibitor cocktail), supplemented with 10 %  
414 glycerol and stored at -80 °C until use. The IP was performed with custom-made antibodies against  
415 Piwi4-6 and Ago3 (1:10 dilution) at 4 °C for 4 h on rotation. Protein A/G Plus beads (Santa Cruz)  
416 were added at a dilution of 1:10 and then incubated overnight at 4 °C on rotation. Beads were  
417 washed 3 times with RIPA buffer, and half was used for RNA isolation and protein analysis each.  
418 For RNA extraction, beads were treated with proteinase K for 2 h at 55 °C and isolated with phenol-  
419 chloroform extraction. Equal amounts of RNA for input and IPs were then analysed by northern  
420 blotting. For western blotting, the IP samples were boiled in 2x Laemmli buffer for 10 min at 95 °C,  
421 separated on 7.5 % SDS-polyacrylamide gels, and blotted on 0.2  $\mu$ m nitrocellulose membranes  
422 (Bio-Rad) in a wet blot chamber on ice. Membranes were blocked for 1 h with 5 % milk in PBS-T  
423 (137 mM NaCl, 12 mM phosphate, 2.7 mM KCl, pH 7.4, 0.1 % (v/v) Tween 20) and incubated with  
424 PIWI-specific (dilution 1:1000) and Tubulin primary antibodies (rat anti-Tubulin alpha, MCA78G,



425 1:1000, Sanbio) overnight at 4 °C. The next day, membranes were washed three times with PBS-T  
426 and incubated with secondary antibodies conjugated to a fluorescence dye (IRDye 800CW  
427 conjugated goat anti rabbit, 1:10,000, Li-Cor, and IRDye 680LT conjugated goat anti rat, 1:10,000,  
428 Li-Cor) for 1 h at room temperature in the dark. After washing three times in PBS-T, signal was  
429 detected with the Odyssey-CLx Imaging system (Li-Cor). Alternatively, *Ae.aegypti* Liverpool  
430 embryos were bead-beaten in RIPA buffer supplemented with 1mM PMSF (Sigma Aldrich), and  
431 boiled in 1x Laemmli buffer. Western blotting was performed as described above, however, the  
432 membrane was blocked in 5 % BSA in PBS-T, and incubated with pSer2 PolII (rabbit anti-RNA  
433 polymerase II CTD YSPTSPS (phosphor S2), ab5095, 1:1000, abcam), and Tubulin primary  
434 antibodies.

435

#### 436 **Northern blot**

437 piRNAs were detected by northern blot analyses, as published in ref.<sup>73</sup>. Briefly, RNA was denatured  
438 at 80 °C for 2 min in Gel Loading Buffer II (Ambion) and size-separated on 0.5 x TBE (45 mM  
439 Tris-borate, 1 mM EDTA), 7 M Urea, 15 % denaturing polyacrylamide gels. RNA was then blotted  
440 on Hybond-NX nylon membranes (GE Healthcare) in a semi-dry blotting chamber for 45 min at  
441 20 V and 4 °C and crosslinked to the membrane with EDC crosslinking solution (127 mM 1-  
442 methylimidazole (Sigma-Aldrich), 163 mM N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide  
443 hydrochloride (Sigma-Aldrich), pH 8.0) at 60 °C for two hours. Crosslinked membranes were pre-  
444 hybridized in ULTRAHyb-Oligo hybridization buffer (Thermo Scientific) for one hour at 42 °C and  
445 probed with indicated <sup>32</sup>P 5' end-labelled DNA oligonucleotide probes overnight at 42 °C.

446 Membranes were then washed with decreasing concentrations of SCC (300 mM NaCl, 30 mM  
447 sodium citrate pH 7.0; 150 mM NaCl, 15 mM sodium citrate; 15 mM NaCl, 1.5 mM sodium citrate)  
448 and 0.1 % SDS, and exposed to Carestream BioMax XAR X-Ray films (Kodak). Probe sequences  
449 can be found in the Supplementary Table S5.

450 To detect a putative tapiR precursor transcript, RNA was size-separated on a 0.5 x TBE, 7 M Urea,  
451 6 % denaturing polyacrylamide gel, transferred to Hybond-N+ nylon membranes (GE Healthcare)  
452 in a semi-dry blotting chamber for 2 h at 200 mA at 4 °C, UV-crosslinked at 150 mJ, pre-hybridized  
453 in ULTRAHyb Ultrasensitive hybridization buffer (Thermo Scientific), and probed with <sup>32</sup>P-labeled  
454 DNA probes overnight at 42 °C, washed as described above, exposed to a Storage Phosphor Screen  
455 GP (Kodak), and developed with the Amersham Typhoon 5 Biomolecular Imager (GE Healthcare).  
456 The probe was produced with the Amersham Rediprime DNA Labelling System II (GE Healthcare)  
457 from a PCR-amplified sequence of the tapiR locus spanning 1.5 repeat units (~450 nt), that was  
458 cloned with flanking T7 sites separated by *Eco*RI and *Zra*I restriction sites into the pUC19 vector.

459



## 460 **Reporter cloning and luciferase assay**

461 Reporters were constructed by cloning annealed and phosphorylated oligonucleotides with the  
462 indicated tapiR1 or control target sites in the pMT-GL3 vector<sup>74</sup>. This vector encodes the *Photinus*  
463 *pyralis* firefly luciferase (GL3) under a copper-inducible metallothionein promoter. Sense and  
464 antisense oligonucleotides (Sigma Aldrich) were annealed by heating to 80 °C, and gradually  
465 cooling down to room temperature, phosphorylated with T4 polynucleotide kinase (Roche) at 37 °C  
466 for 30 min, purified and then ligated into the pMT-GL3 vector. For cloning of 3'UTR and 5' UTR  
467 reporters, the target site or the target site and an upstream *Bam*HI site were cloned into the *Pme*I  
468 and *Sac*II, or *Not*I and *Xho*I restriction sites, respectively. ORF reporters were constructed by  
469 cloning a Kozak sequence followed by the first 45 nucleotides of luciferase and the target site into  
470 *Xho*I and *Nco*I sites. For designing the intron reporters, the first intron of RpS7 (AAEL009496) was  
471 cloned behind the duplicated first 45 nt of luciferase in the pMT-GL3 vector, and the original ATG  
472 of firefly luciferase was mutated, and 3 stop codons were introduced at the 3' end of the intron, and  
473 a *Bsa*BI restriction site was inserted in the first third of the intron of by site directed mutagenesis.  
474 Afterwards, oligonucleotides encoding tapiR1 or control target sequences were inserted into *Bsa*BI  
475 as described above. IRES-containing reporters were designed by cloning the 5' UTR of cricket  
476 paralysis virus amplified from infected S2 cells into *Pme*I and *Sac*II restriction sites of the pMT-  
477 GL3 vector with a mutated tapiR1 target site ( $\Delta$ tapiR1) using the HD In-Fusion cloning kit  
478 (Takara). Afterwards oligonucleotide containing tapiR1 or control target sites were cloned into the  
479 *Sac*II site in the 3' UTR as described above. RNAPIII reporters were constructed by cloning the *Ae.*  
480 *aegypti* U6 promoter and the GL3 3' UTR including different tapiR1target sites and a series of 6  
481 T's as termination signal into pUC19 with the HD In-Fusion cloning kit (Takara). As normalization  
482 control, a part of GL3 ORF was cloned downstream of the U6 promoter. Sequences of the  
483 oligonucleotides are provided in Supplementary Table S5. Correct insertion of target sites was  
484 confirmed by Sanger sequencing for all clones.

485 Where indicated, mutated firefly or *Renilla* luciferase versions were used that harbour synonymous  
486 mutations destroying the predicted target sites for tapiR1 (firefly luciferase, nt 782: 5'-  
487 gagtcgtcttaatgtatagattgaagaa-3' mutated to 5'-gtgtcgtgcttatgtaccggttcgaggag-3', and *Renilla*  
488 luciferase, nt 462 5'-tgaatggcctgatattgaagaa-3' mutated to 5'-tgagtggccagatatcgaggag-3'; modified  
489 nucleotides in bold). Aag2 cells were seeded in 96-well plates the day before the experiment and  
490 transfected with 100 ng of the indicated plasmids and 100 ng pMT-*Renilla*<sup>74</sup> per well, using 2  $\mu$ L X-  
491 tremeGENE HP DNA transfection reagent per 1  $\mu$ g plasmid DNA according to the manufacturer's  
492 instructions. Alternatively, 100 ng reporter plasmid and 100 ng pMT-*Renilla* were co-transfected  
493 with the indicated amounts of unlabelled, fully 2'-O-methylated antisense RNA oligonucleotide  
494 using an additional amount of 4  $\mu$ L X-tremeGENE HP DNA transfection reagent (Roche) per 1  $\mu$ g

oligonucleotide. Medium was replaced 3 h after reporter plasmid transfection with 0.5 mM CuSO<sub>4</sub> in fully supplemented Leibovitz's L-15 medium to induce the metallothionein promoter. 24 h later, cells were lysed in 30 µL Passive lysis buffer (Promega) and activity of both luciferases was measured in 10 µL of the sample with the Dual Luciferase Reporter Assay system (Promega) on a Modulus Single Tube Reader (Turner Biosystems). Firefly luciferase was normalized to *Renilla* luciferase activity. For each construct, at least two to three independent clones were measured in triplicate in order to exclude clonal effects.

## **RT-qPCR**

1 µg of total RNA was treated with DNaseI (Ambion) for 45 min at 37 °C and reverse transcribed using the Taqman reverse transcription kit (Applied Biosystems) according to the manufacturer's protocol. Real-time PCR was performed with the GoTag qPCR Master Mix (Promega) and measured on a LightCycler480 instrument (Roche) with 5 min initial denaturation and 45 cycles of 5 s denaturation at 95 °C, 10 s annealing at 60 °C and 20 s amplification at 72 °C. Starting fluorescence values of specific mRNAs were calculated with linear regression method of log fluorescence per cycle number and LinRegPCR program, version 2015.3, as described in ref. <sup>60</sup>.

## **Stem-loop RT-qPCR**

Quantification of tapiR1 piRNA levels were performed similar to miRNA quantification as published ref. <sup>75</sup>. Briefly, 100ng of total RNA was reverse transcribed with 10 pmol stem-loop RT primer using 25u Superscript II reverse transcriptase (Invitrogen) in 1x First Strand buffer, 0.33 mM dNTPs, and 2u RNase inhibitors. cDNA was then measured by Real-time PCR as described above.

## **3' RACE, and 5' RACE of slicer products**

3' Rapid Amplification of cDNA Ends (3' RACE) was performed using the FirstChoice RLM-RACE Kit (Thermo Fischer Scientific) according to the manufacturer's instructions. Amplification products were separated on agarose gel, purified and Sanger sequenced. Slicer products were detected from 1 µg total RNA by 5' RACE following to the instructions, but without prior Calf Intestine Alkaline Phosphatase and Tobacco Acid Pyrophosphatase treatment. After amplification by PCR RACE products in the size range of 150-250 nt were purified from agarose gel, and cloned into pUC19 with HD In-Fusion cloning kit (Takara Bio) according to the manufacturer's instructions, and sequenced by Sanger sequencing.

Primer sequences can be found in Supplementary Table S1.

529 **tapiR1/tapiR2 antisense oligonucleotide treatment and injection**

530 Aag2 cells were seeded in 24-well plates the day before the experiment. Cells were treated with  
531 500 nM 5'Cy5-labelled, fully 2'O-methylated antisense RNA oligonucleotide in 530 µL medium  
532 with 4 µL X-tremeGENE HP DNA transfection reagent (Roche) per 1 µg oligonucleotide. Medium  
533 was replaced after 3 h and cells from three independent experiments were harvested 48 h after  
534 transfection and prepared for RNA sequencing (see below). For 5' RACE, cells were treated with  
535 200 nM antisense oligonucleotides together with 50 nM of siRNA duplexes, and RNA was  
536 harvested 24 h later.

537 *Ae. aegypti* Jane embryos were injected with antisense oligonucleotides similar to as described  
538 above. However, the engorged female mosquitoes were kept at 25 °C and 70 % humidity and  
539 allowed to lay eggs for 45 min. Injection was performed with a FemtoJet 4x (Eppendorf) with 1200  
540 hPa pressure. Injected embryos were then transferred to a wet Whatman paper and kept at 27 °C  
541 and 80 % humidity for the indicated times. Per experiment, 50 to 150 embryos were injected per  
542 condition.

543

544 **Scoring of embryo development and hatching**

545 Injected embryos were allowed to develop for 2.5 days after injection on a moist Whatman paper  
546 and then fixed in 4 % paraformaldehyde for 8 h to overnight. Afterwards, the pigment of the  
547 endochorion was bleached with Trpis solution<sup>76</sup> (0.037 M sodium chlorite, 1.45 M acetic acid) for  
548 24 to 48 h. Embryos were washed five times in PBS and images were taken with a EVOS FL  
549 imaging system (Thermo Fisher Scientific). Embryos with evident larval segmentation (head, fused  
550 thoracical elements and abdomen) were scored as fully developed and embryos without any evident  
551 structure of the ooplasm as undeveloped. Individuals that showed first signs of structural  
552 rearrangements of the ooplasm, but did not complete larval segmentation were scored as  
553 intermediate (see Supplementary Fig S1F). To avoid biases, the scoring was performed blindly.  
554 Hatching rate was counted from injected embryos 4 days post injection. Embryos were kept moist  
555 for two days and then allowed to slowly dry for the rest of the period. The embryos were transferred  
556 to water and then forced to hatch by applying negative pressure for a period of 30 min. The number  
557 of hatched L1 larvae was counted immediately afterwards.

558

559 **Poly(A) tail length (PAT) assay**

560 Ligation-mediated (LM) PAT assay was performed as published in ref.<sup>77</sup>, and RACE PAT as  
561 published in ref.<sup>78</sup>. Superscript IV Reverse transcriptase (Invitrogen) was used for cDNA synthesis.  
562 RACE-PAT products were amplified by PCR with 3 min initial denaturation at 95 °C, 35 cycles of  
563 30s denaturation, 30s annealing at 60 °C and 1 min extension at 72 °C, and final extension for 5

min, while PCR was performed likewise, but with 2 min extension each cycle. PCR products were separated on agarose gel, and stained with ethidium bromide. As control, Sindbis virus poly(A) tails were compared between *in vitro* transcribed RN, produced as described previously ref.<sup>79,80</sup>, and total RNA isolated from infected Aag2 cells.

### Sequence logo

Repeat monomers from the satellite repeat loci in *Ae. aegypti*, *Ae. albopictus*, and *Cx. quinquefasciatus* were extracted manually from the current genome annotations obtained from VectorBase (*Aedes aegypti* Liverpool AaegL5, *Aedes albopictus* Foshan AaloF1, *Culex quinquefasciatus* Johannesburg CpipJ2). A repeat unit was defined as the sequence starting from the first tapiR1 nucleotide until one nucleotide upstream of the next tapiR1 sequence. Sequences were aligned using MAFFT (v7.397)<sup>81</sup> (with options `--genafpair --leavegappyregion --kimura 1 --maxiterate 1000 --retree 1`) and the sequence logo was constructed with the R package ggseqlogo<sup>82</sup>.

### Small RNA sequencing

Small RNAs from Aag2 cells (input) or PIWI immunoprecipitations were cloned with the TruSeq small RNA sample preparation kit (Illumina) according to the manufacturer's instructions. For the input sample, size selected 19-33 nt small RNAs purified from polyacrylamide gel were used to construct the library as described previously<sup>83</sup>, whereas IP samples were not extracted from gel but isolated as described before<sup>27</sup>. Libraries were sequenced on an Illumina HiSeq 4000 instrument by Plateforme GenomEast (Strasbourg, France).

### mRNA sequencing

RNA was isolated from Aag2 cells 48 h after AO transfection (three independent experiments), or embryos 20.5 h after AO injection (50 embryos pooled per experiment from five independent experiments) with RNAsolv reagent following standard phenol-chloroform extraction. Polyadenylated RNAs were extracted and sequencing libraries were prepared using the TruSeq stranded mRNA Library Prep kit (Illumina) following the manufacturer's instructions, and sequenced on an Illumina Hi-Seq 4000 instrument (2x50 bases).

### Analysis of mRNA sequencing

Reads were mapped to the *Ae. aegypti* genome AaegL5 as provided by VectorBase (<https://www.vectorbase.org>) with STAR (version 2.5.2b)<sup>84</sup> in 2-pass mode: first mapping was done for all samples (options: `--readFilesCommand zcat --outSAMtype None --outSAMattrIHstart 0 --outSAMstrandField intronMotif`), identified splice junctions were combined (junctions located

on the mitochondrial genome were filtered out, as these are likely false positives), and this list of junctions was used in a second round of mapping (with `-sjdbFileChrStartEnd`) and default parameters as above. Reads were quantified with the additional option `-quantMode GeneCounts` to quantified reads per gene. Alternatively, reads were quantified on TEfam transposon consensus sequences ([https://tefam.biochem.vt.edu/tefam/get\\_fasta.php](https://tefam.biochem.vt.edu/tefam/get_fasta.php)) with Salmon (v.0.8.2)<sup>85</sup>, default settings and libType set to "ISR". Statistical and further downstream analyses were performed with DESeq2<sup>86</sup> from Bioconductor. Significance was tested at an FDR of 0.01 and a log<sub>2</sub> fold change of 0.5. tapiR1 target sites were predicted with the online tool from RNAHybrid<sup>49</sup> with helix constraints from nucleotide two to seven, and no G:U wobble allowed in the seed. Predictions were made on the AaegL5.1 geneset as provided by VectorBase, and on TEfam transposon consensus sequences. For Fig4H, publicly available sequencing datasets<sup>87</sup> (accession numbers: SRR923702, SRR923826, SRR923837, SRR923853, SRR923704) were mapped and quantified as described above. Genes were categorized on the basis of their expression in embryos at 0-2 h vs. 12-16 h post egg-laying. Genes not detected in the 0-2 h sample were defined as purely zygotic, and genes that did not increase or decrease by more than log<sub>2</sub>(0.5) as maternal stable. Genes that changed in expression by more than log<sub>2</sub>(0.5), log<sub>2</sub>(2), and log<sub>2</sub>(5) from 0-2 h to 12-16 h were categorized as maternal unstable fraction (decreased expression), or as genes that are maternally provided but are transcribed by the zygote in addition to the preloaded maternal pool (increased expression). tapiR1 targets were defined as genes that were significantly upregulated at least two fold in tapiR1 AO injected embryos and harbour a predicted tapiR1 target site (mfe <= -24). Sequenced libraries are available under BioProject number PRJNA482553. A list of publicly available datasets that were used in this study is provided in Supplementary Table S6. The code will be made available on GitHub upon publication.

### Analysis of small RNA sequencing

3' sequencing adapters (TGGAATTCTCGGGTGCCAAGG) were trimmed from the sequence reads with Cutadapt (version 1.14)<sup>88</sup> and trimmed reads were mapped with Bowtie (version 0.12.7)<sup>89</sup> to the *Aedes aegypti* LVP\_AGWG genome sequence AaegL5.1 obtained from VectorBase with at most 1 mismatch. Reads that mapped to rRNAs or tRNAs were excluded from the analyses. Alternatively, 3' sequencing adapters ((NNN)TGGAATTCTCGGGTGCCAAGGC) and three random bases were trimmed from publicly available datasets from *Ae. aegypti* somatic and germline tissues<sup>90</sup> (SRR5961503, SRR5961504, SRR5961505, SRR5961506) and then processed as described above. Oxidized libraries, IPs and input sample were normalized to the total number of mapped reads, all other libraries to the total number of miRNAs (in millions). piRNAs that were at least two fold enriched in a PIWI-IP compared to the corresponding input sample and were present

with at least 10 rpm in the IP sample were considered PIWI-bound. Mapping positions were overlapped with basefeatures and repeatfeatures retrieved from VectorBase and counted with bedtools<sup>91</sup>. Reads that mapped to two or more features were assigned to only one feature with the following hierarchy: open reading frames > non-coding RNAs (incl. lncRNAs, pseudogenes, snoRNAs, snRNAs, miRNAs) > LTR retrotransposons > Non-LTR retrotransposons (SINEs, LINEs, Penelope) > “Cut and paste” DNA transposons > other DNA transposons (Helitrons, MITEs) > satellite and tandem repeat features > DUST > other /unknown repeats. Accordingly, reads that mapped to a repeat feature and an intron or UTR were classified as repeat-derived, whereas all other reads mapping to introns or UTRs were considered as gene-derived. Positions not overlapping with any annotation were summarized as “other”. Results were then visualized with ggplot2<sup>92</sup>, or Gviz<sup>93</sup> in R.

The code will be made available on GitHub upon publication.

## Data availability

Raw sequence data is deposited in the NCBI Sequence Read Archive under the BioProject number PRJNA482553.

## References

- 1 Garrido-Ramos, M. A. Satellite DNA: An Evolving Topic. *Genes (Basel)* **8**, doi:10.3390/genes8090230 (2017).
- 2 Brajkovic, J., Feliciello, I., Bruvo-Madaric, B. & Ugarkovic, D. Satellite DNA-like elements associated with genes within euchromatin of the beetle *Tribolium castaneum*. *G3 (Bethesda)* **2**, 931-941, doi:10.1534/g3.112.003467 (2012).
- 3 Kuhn, G. C., Kuttler, H., Moreira-Filho, O. & Heslop-Harrison, J. S. The 1.688 repetitive DNA of *Drosophila*: concerted evolution at different genomic scales and association with genes. *Mol Biol Evol* **29**, 7-11, doi:10.1093/molbev/msr173 (2012).
- 4 Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199-202, doi:10.1038/nature04917 (2006).
- 5 Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313**, 363-367, doi:10.1126/science.1130164 (2006).
- 6 Melters, D. P. *et al.* Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14**, R10, doi:10.1186/gb-2013-14-1-r10 (2013).
- 7 Kit, S. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *J Mol Biol* **3**, 711-716, doi:10.1016/s0022-2836(61)80075-2 (1961).
- 8 Sueoka, N. Variation and heterogeneity of base composition of deoxyribonucleic acids: A compilation of old and new data. *Journal of Molecular Biology* **3**, 31-IN15 (1961).
- 9 Hall, I. M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232-2237, doi:10.1126/science.1076466 (2002).
- 10 Menon, D. U., Coarfa, C., Xiao, W., Gunaratne, P. H. & Meller, V. H. siRNAs from an X-linked satellite repeat promote X-chromosome recognition in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **111**, 16460-16465, doi:10.1073/pnas.1410534111 (2014).



677 11 Zakrzewski, F. *et al.* Epigenetic profiling of heterochromatic satellite DNA. *Chromosoma*  
678 **120**, 409-422, doi:10.1007/s00412-011-0325-x (2011).

679 12 Fukagawa, T. *et al.* Dicer is essential for formation of the heterochromatin structure in  
680 vertebrate cells. *Nat Cell Biol* **6**, 784-791, doi:10.1038/ncb1155 (2004).

681 13 Kanellopoulou, C. *et al.* Dicer-deficient mouse embryonic stem cells are defective in  
682 differentiation and centromeric silencing. *Genes Dev* **19**, 489-501, doi:10.1101/gad.1248505  
683 (2005).

684 14 Pezer, Z., Brajkovic, J., Feliciello, I. & Ugarkovc, D. Satellite DNA-mediated effects on  
685 genome regulation. *Genome Dyn* **7**, 153-169, doi:10.1159/000337116 (2012).

686 15 May, B. P., Lippman, Z. B., Fang, Y., Spector, D. L. & Martienssen, R. A. Differential  
687 regulation of strand-specific transcripts from Arabidopsis centromeric satellite repeats. *PLoS*  
688 *Genet* **1**, e79, doi:10.1371/journal.pgen.0010079 (2005).

689 16 Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9  
690 methylation by RNAi. *Science* **297**, 1833-1837, doi:10.1126/science.1074973 (2002).

691 17 Matthews, B. J. *et al.* Improved reference genome of *Aedes aegypti* informs arbovirus  
692 vector control. *Nature* **563**, 501-507, doi:10.1038/s41586-018-0692-z (2018).

693 18 Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the  
694 vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**, 246-258, doi:10.1038/nrm3089  
695 (2011).

696 19 Czech, B. & Hannon, G. J. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-  
697 Guided Silencing. *Trends Biochem Sci* **41**, 324-337, doi:10.1016/j.tibs.2015.12.008 (2016).

698 20 Arensburger, P., Hice, R. H., Wright, J. A., Craig, N. L. & Atkinson, P. W. The mosquito  
699 *Aedes aegypti* has a large genome size and high transposable element load but contains a  
700 low proportion of transposon-specific piRNAs. *BMC Genomics* **12**, 606, doi:10.1186/1471-  
701 2164-12-606 (2011).

702 21 Kawaoka, S., Izumi, N., Katsuma, S. & Tomari, Y. 3' end formation of PIWI-interacting  
703 RNAs in vitro. *Mol Cell* **43**, 1015-1022, doi:10.1016/j.molcel.2011.07.029 (2011).

704 22 Saito, K. *et al.* Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-  
705 interacting RNAs at their 3' ends. *Genes Dev* **21**, 1603-1608, doi:10.1101/gad.1563607  
706 (2007).

707 23 Horwich, M. D. *et al.* The *Drosophila* RNA methyltransferase, DmHen1, modifies germline  
708 piRNAs and single-stranded siRNAs in RISC. *Curr Biol* **17**, 1265-1272,  
709 doi:10.1016/j.cub.2007.06.030 (2007).

710 24 Miesen, P., Joosten, J. & van Rij, R. P. PIWIs Go Viral: Arbovirus-Derived piRNAs in  
711 Vector Mosquitoes. *PLoS Pathog* **12**, e1006017, doi:10.1371/journal.ppat.1006017 (2016).

712 25 Vodovar, N. *et al.* Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito  
713 cells. *PLoS One* **7**, e30861, doi:10.1371/journal.pone.0030861 (2012).

714 26 Miesen, P., Girardi, E. & van Rij, R. P. Distinct sets of PIWI proteins produce arbovirus and  
715 transposon-derived piRNAs in *Aedes aegypti* mosquito cells. *Nucleic Acids Res* **43**, 6545-  
716 6556, doi:10.1093/nar/gkv590 (2015).

717 27 Joosten, J. *et al.* The Tudor protein Veneno assembles the ping-pong amplification complex  
718 that produces viral piRNAs in *Aedes* mosquitoes. *Nucleic Acids Res* **47**, 2546-2559,  
719 doi:10.1093/nar/gky1266 (2019).

720 28 Schnettler, E. *et al.* Knockdown of piRNA pathway proteins results in enhanced Semliki  
721 Forest virus production in mosquito cells. *J Gen Virol* **94**, 1680-1689,  
722 doi:10.1099/vir.0.053850-0 (2013).

723 29 Varjak, M. *et al.* *Aedes aegypti* Piwi4 Is a Noncanonical PIWI Protein Involved in Antiviral  
724 Responses. *mSphere* **2**, doi:10.1128/mSphere.00144-17 (2017).

725 30 Plohl, M. *et al.* Long-term conservation vs high sequence divergence: the case of an  
726 extraordinarily old satellite DNA in bivalve mollusks. *Heredity (Edinb)* **104**, 543-551,  
727 doi:10.1038/hdy.2009.141 (2010).



728 31 Li, Y. X. & Kirby, M. L. Coordinated and conserved expression of alphoid repeat and  
729 alphoid repeat-tagged coding sequences. *Dev Dyn* **228**, 72-81, doi:10.1002/dvdy.10355  
730 (2003).

731 32 Martinez-Lage, A., Rodriguez-Farina, F., Gonzalez-Tizon, A. & Mendez, J. Origin and  
732 evolution of Mytilus mussel satellite DNAs. *Genome* **48**, 247-256, doi:10.1139/g04-115  
733 (2005).

734 33 Henikoff, S., Ahmad, K. & Malik, H. S. The centromere paradox: stable inheritance with  
735 rapidly evolving DNA. *Science* **293**, 1098-1102, doi:10.1126/science.1062939 (2001).

736 34 Plohl, M., Mestrovic, N. & Mravinac, B. Satellite DNA evolution. *Genome Dyn* **7**, 126-152,  
737 doi:10.1159/000337122 (2012).

738 35 Reidenbach, K. R. *et al.* Phylogenetic analysis and temporal diversification of mosquitoes  
739 (Diptera: Culicidae) based on nuclear genes and morphology. *BMC Evol Biol* **9**, 298,  
740 doi:10.1186/1471-2148-9-298 (2009).

741 36 Chaves, R., Ferreira, D., Mendes-da-Silva, A., Meles, S. & Adegas, F. FA-SAT Is an Old  
742 Satellite DNA Frozen in Several Bilateria Genomes. *Genome Biol Evol* **9**, 3073-3087,  
743 doi:10.1093/gbe/evx212 (2017).

744 37 Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233,  
745 doi:10.1016/j.cell.2009.01.002 (2009).

746 38 Zhang, D. *et al.* The piRNA targeting rules and the resistance to piRNA silencing in  
747 endogenous genes. *Science* **359**, 587-592, doi:10.1126/science.aao2840 (2018).

748 39 Shen, E. Z. *et al.* Identification of piRNA Binding Sites Reveals the Argonaute Regulatory  
749 Landscape of the *C. elegans* Germline. *Cell* **172**, 937-951 e918,  
750 doi:10.1016/j.cell.2018.02.002 (2018).

751 40 Wang, W. *et al.* The initial uridine of primary piRNAs does not create the tenth adenine that  
752 is the hallmark of secondary piRNAs. *Mol Cell* **56**, 708-716,  
753 doi:10.1016/j.molcel.2014.10.016 (2014).

754 41 Mohn, F., Handler, D. & Brennecke, J. Noncoding RNA. piRNA-guided slicing specifies  
755 transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* **348**, 812-817,  
756 doi:10.1126/science.aaa1039 (2015).

757 42 Matsumoto, N. *et al.* Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to  
758 piRNA. *Cell* **167**, 484-497 e489, doi:10.1016/j.cell.2016.09.002 (2016).

759 43 Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide  
760 DNA and target RNA duplex. *Nature* **456**, 921-926, doi:10.1038/nature07666 (2008).

761 44 Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-  
762 containing argonaute silencing complex. *Nature* **456**, 209-213, doi:10.1038/nature07315  
763 (2008).

764 45 Reuter, M. *et al.* Miwi catalysis is required for piRNA amplification-independent LINE1  
765 transposon silencing. *Nature* **480**, 264-267, doi:10.1038/nature10672 (2011).

766 46 Goh, W. S. *et al.* piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis.  
767 *Genes Dev* **29**, 1032-1044, doi:10.1101/gad.260455.115 (2015).

768 47 Pezer, Z. & Ugarkovic, D. Satellite DNA-associated siRNAs as mediators of heat shock  
769 response in insects. *RNA Biol* **9**, 587-595, doi:10.4161/rna.20019 (2012).

770 48 Feliciello, I., Akrap, I. & Ugarkovic, D. Satellite DNA Modulates Gene Expression in the  
771 Beetle *Tribolium castaneum* after Heat Stress. *PLoS Genet* **11**, e1005466,  
772 doi:10.1371/journal.pgen.1005466 (2015).

773 49 Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction  
774 of microRNA/target duplexes. *RNA* **10**, 1507-1517, doi:10.1261/rna.5248604 (2004).

775 50 Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed  
776 pairing. *Mol Cell* **27**, 91-105, doi:10.1016/j.molcel.2007.06.017 (2007).

777 51 Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon  
778 activity in *Drosophila*. *Cell* **128**, 1089-1103, doi:10.1016/j.cell.2007.01.043 (2007).

779 52 Gross, L. *et al.* The IRES5'UTR of the dicistrovirus cricket paralysis virus is a type III IRES  
780 containing an essential pseudoknot structure. *Nucleic Acids Res* **45**, 8993-9004,  
781 doi:10.1093/nar/gkx622 (2017).

782 53 Mathonnet, G. *et al.* MicroRNA inhibition of translation initiation in vitro by targeting the  
783 cap-binding complex eIF4F. *Science* **317**, 1764-1767, doi:10.1126/science.1146067 (2007).

784 54 Pillai, R. S. *et al.* Inhibition of translational initiation by Let-7 MicroRNA in human cells.  
785 *Science* **309**, 1573-1576, doi:10.1126/science.1115079 (2005).

786 55 Humphreys, D. T., Westman, B. J., Martin, D. I. & Preiss, T. MicroRNAs control translation  
787 initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc*  
788 *Natl Acad Sci U S A* **102**, 16961-16966, doi:10.1073/pnas.0506482102 (2005).

789 56 Behm-Ansmant, I. *et al.* mRNA degradation by miRNAs and GW182 requires both  
790 CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* **20**, 1885-  
791 1898, doi:10.1101/gad.1424106 (2006).

792 57 Braun, J. E., Huntzinger, E., Fauser, M. & Izaurralde, E. GW182 proteins directly recruit  
793 cytoplasmic deadenylase complexes to miRNA targets. *Mol Cell* **44**, 120-133,  
794 doi:10.1016/j.molcel.2011.09.007 (2011).

795 58 Chekulaeva, M. *et al.* miRNA repression involves GW182-mediated recruitment of CCR4-  
796 NOT through conserved W-containing motifs. *Nat Struct Mol Biol* **18**, 1218-1226,  
797 doi:10.1038/nsmb.2166 (2011).

798 59 Rouget, C. *et al.* Maternal mRNA deadenylation and decay by the piRNA pathway in the  
799 early *Drosophila* embryo. *Nature* **467**, 1128-1132, doi:10.1038/nature09465 (2010).

800 60 Temme, C., Simonelig, M. & Wahle, E. Deadenylation of mRNA by the CCR4-NOT  
801 complex in *Drosophila*: molecular and developmental aspects. *Front Genet* **5**, 143,  
802 doi:10.3389/fgene.2014.00143 (2014).

803 61 Frey, T. K. & Strauss, J. H. Replication of Sindbis virus. VI. Poly(A) and poly(U) in virus-  
804 specific RNA species. *Virology* **86**, 494-506, doi:10.1016/0042-6822(78)90088-0 (1978).

805 62 Ugarkovic, D. Functional elements residing within satellite DNAs. *EMBO Rep* **6**, 1035-  
806 1039, doi:10.1038/sj.embor.7400558 (2005).

807 63 Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts.  
808 *Development* **136**, 3033-3042, doi:10.1242/dev.033183 (2009).

809 64 Thomsen, S., Anders, S., Janga, S. C., Huber, W. & Alonso, C. R. Genome-wide analysis of  
810 mRNA decay patterns during early *Drosophila* development. *Genome Biol* **11**, R93,  
811 doi:10.1186/gb-2010-11-9-r93 (2010).

812 65 Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal  
813 mRNAs. *Science* **312**, 75-79, doi:10.1126/science.1122689 (2006).

814 66 Lund, E., Liu, M., Hartley, R. S., Sheets, M. D. & Dahlberg, J. E. Deadenylation of maternal  
815 mRNAs mediated by miR-427 in *Xenopus laevis* embryos. *RNA* **15**, 2351-2363,  
816 doi:10.1261/rna.1882009 (2009).

817 67 Bushati, N., Stark, A., Brennecke, J. & Cohen, S. M. Temporal reciprocity of miRNAs and  
818 their targets during the maternal-to-zygotic transition in *Drosophila*. *Curr Biol* **18**, 501-506,  
819 doi:10.1016/j.cub.2008.02.081 (2008).

820 68 Barckmann, B. *et al.* Aubergine iCLIP Reveals piRNA-Dependent Decay of mRNAs  
821 Involved in Germ Cell Development in the Early Embryo. *Cell Rep* **12**, 1205-1216,  
822 doi:10.1016/j.celrep.2015.07.030 (2015).

823 69 Lan, Q. & Fallon, A. M. Small heat shock proteins distinguish between two mosquito  
824 species and confirm identity of their cell lines. *Am J Trop Med Hyg* **43**, 669-676,  
825 doi:10.4269/ajtmh.1990.43.669 (1990).

826 70 Fansiri, T. *et al.* Genetic mapping of specific interactions between *Aedes aegypti*  
827 mosquitoes and dengue viruses. *PLoS Genet* **9**, e1003621,  
828 doi:10.1371/journal.pgen.1003621 (2013).

829 71 Goertz, G. P., Vogels, C. B. F., Geertsema, C., Koenraadt, C. J. M. & Pijlman, G. P.  
830 Mosquito co-infection with Zika and chikungunya virus allows simultaneous transmission  
831 without affecting vector competence of *Aedes aegypti*. *PLoS Negl Trop Dis* **11**, e0005654,  
832 doi:10.1371/journal.pntd.0005654 (2017).

833 72 Mohlmann, T. W. R. *et al.* Community analysis of the abundance and diversity of mosquito  
834 species (Diptera: Culicidae) in three European countries at different latitudes. *Parasit*  
835 *Vectors* **10**, 510, doi:10.1186/s13071-017-2481-1 (2017).

836 73 Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of  
837 small RNA. *Nat Protoc* **3**, 1077-1084, doi:10.1038/nprot.2008.67 (2008).

838 74 van Rij, R. P. *et al.* The RNA silencing endonuclease Argonaute 2 mediates specific  
839 antiviral immunity in *Drosophila melanogaster*. *Genes Dev* **20**, 2985-2995,  
840 doi:10.1101/gad.1482006 (2006).

841 75 Chen, C. *et al.* Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic*  
842 *Acids Res* **33**, e179, doi:10.1093/nar/gni178 (2005).

843 76 Trpiš, M. A new bleaching and decalcifying method for general use in zoology. *Canadian*  
844 *Journal of Zoology* **48**, 892-893, doi:10.1139/z70-158 (1970).

845 77 Murray, E. L. & Schoenberg, D. R. Assays for determining poly(A) tail length and the  
846 polarity of mRNA decay in mammalian cells. *Methods Enzymol* **448**, 483-504,  
847 doi:10.1016/S0076-6879(08)02624-4 (2008).

848 78 Salles, F. J. & Strickland, S. Analysis of poly(A) tail lengths by PCR: the PAT assay.  
849 *Methods Mol Biol* **118**, 441-448, doi:10.1385/1-59259-676-2:441 (1999).

850 79 Hahn, C. S., Hahn, Y. S., Braciale, T. J. & Rice, C. M. Infectious Sindbis virus transient  
851 expression vectors for studying antigen processing and presentation. *Proc Natl Acad Sci U S*  
852 *A* **89**, 2679-2683, doi:10.1073/pnas.89.7.2679 (1992).

853 80 van Mierlo, J. T. *et al.* Novel *Drosophila* viruses encode host-specific suppressors of RNAi.  
854 *PLoS Pathog* **10**, e1004256, doi:10.1371/journal.ppat.1004256 (2014).

855 81 Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple  
856 sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066,  
857 doi:10.1093/nar/gkf436 (2002).

858 82 Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**,  
859 3645-3647, doi:10.1093/bioinformatics/btx469 (2017).

860 83 van Cleef, K. W. *et al.* Mosquito and *Drosophila* entomobirnaviruses suppress dsRNA- and  
861 siRNA-induced RNAi. *Nucleic Acids Res* **42**, 8732-8744, doi:10.1093/nar/gku528 (2014).

862 84 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,  
863 doi:10.1093/bioinformatics/bts635 (2013).

864 85 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and  
865 bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419,  
866 doi:10.1038/nmeth.4197 (2017).

867 86 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion  
868 for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8  
869 (2014).

870 87 Akbari, O. S. *et al.* The developmental transcriptome of the mosquito *Aedes aegypti*, an  
871 invasive species and major arbovirus vector. *G3 (Bethesda)* **3**, 1493-1509,  
872 doi:10.1534/g3.113.006742 (2013).

873 88 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.  
874 *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).

875 89 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient  
876 alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25,  
877 doi:10.1186/gb-2009-10-3-r25 (2009).

878 90 Lewis, S. H. *et al.* Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence  
879 against transposable elements. *Nat Ecol Evol* **2**, 174-181, doi:10.1038/s41559-017-0403-4  
880 (2018).

- 881 91 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr*  
882 *Protoc Bioinformatics* **47**, 11 12 11-34, doi:10.1002/0471250953.bi1112s47 (2014).  
883 92 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York,  
884 2016).  
885 93 Hahne, F. & Ivanek, R. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods*  
886 *Mol Biol* **1418**, 335-351, doi:10.1007/978-1-4939-3578-9\_16 (2016).  
887

## 888 **Supplementary Information**

- 889 Supplementary Table S1: Differentially expressed genes upon tapiR1 AO treatment in Aag2 cells.  
890 Supplementary Table S2: Differentially expressed transposable elements upon tapiR1 AO treatment  
891 in Aag2 cells.  
892 Supplementary Table S3: Differentially expressed genes upon tapiR1 AO treatment in embryos.  
893 Supplementary Table S4: Differentially expressed transposable elements upon tapiR1 AO treatment  
894 in embryos.  
895 Supplementary Table S5: Oligonucleotide sequences used in this study.  
896 Supplementary Table S6: RNA sequencing datasets used in this study.  
897

## 898 **Acknowledgments**

899 We thank past and current members of the Van Rij laboratory for discussions. We are grateful to  
900 Anna Beth Crist and Artem Baidaliuk (Institut Pasteur, Paris, France) for their help with mosquito  
901 rearing and embryo injections, and to Catherine Bourgouin and Nicolas Puchot (Institut Pasteur,  
902 Paris, France) for assistance with the microinjection apparatus. We thank Bas Dutilh (Radboud  
903 University Medical Center, Nijmegen, The Netherlands) for his support with analysing target site  
904 enrichments, and Geert-Jan van Gemert (Radboud University Medical Center, Nijmegen, The  
905 Netherlands), and Mojca Kristan (London School of Hygiene and Tropical Medicine, London,  
906 United Kingdom) for kindly providing *An. stephensi* mosquitoes. Tim Möhlmann (Wageningen  
907 University, Wageningen, The Netherlands) is acknowledged for collecting wild-caught mosquito  
908 samples. The following reagent was provided by the NIH/NIAID Filariasis Research Reagent  
909 Resource Center for distribution by BEI Resources, NIAID, NIH: *Aedes aegypti*, Strain Black Eye  
910 Liverpool, Eggs, NR-48921. Sequencing was performed by the GenomEast platform, a member of  
911 the ‘France Génomique’ consortium (ANR-10-INBS-0009)  
912 This work is financially supported by a Consolidator Grant from the European Research Council  
913 under the European Union’s Seventh Framework Programme (grant number ERC CoG 615680) and  
914 a VICI grant from the Netherlands Organization for Scientific Research (grant number  
915 016.VICI.170.090). A stay of R.H. at Pasteur Institute, Paris, France was supported by  
916 ERASMUS+.

917

918 **Author contributions**

919 R.H., P.M., and R.P.v.R designed the experiments and analysed the data. R.H. performed the  
920 computational analyses and most of the experiments, except for PIWI-IPs for small RNA  
921 sequencing (J.J. and E.T.), design and validation of PIWI antibodies (B.P.), and tissue isolations  
922 and blood feeding experiment (C.B.F.V. and C.J.K.). C.B.F.V. and C.J.K. provided wild-caught  
923 mosquito samples. I.R. assisted with the experiments, and S.H.M. and L.L. helped with optimizing  
924 embryo injections. R.H. and R.P.v.R. wrote the paper. All authors read and contributed to the  
925 manuscript.

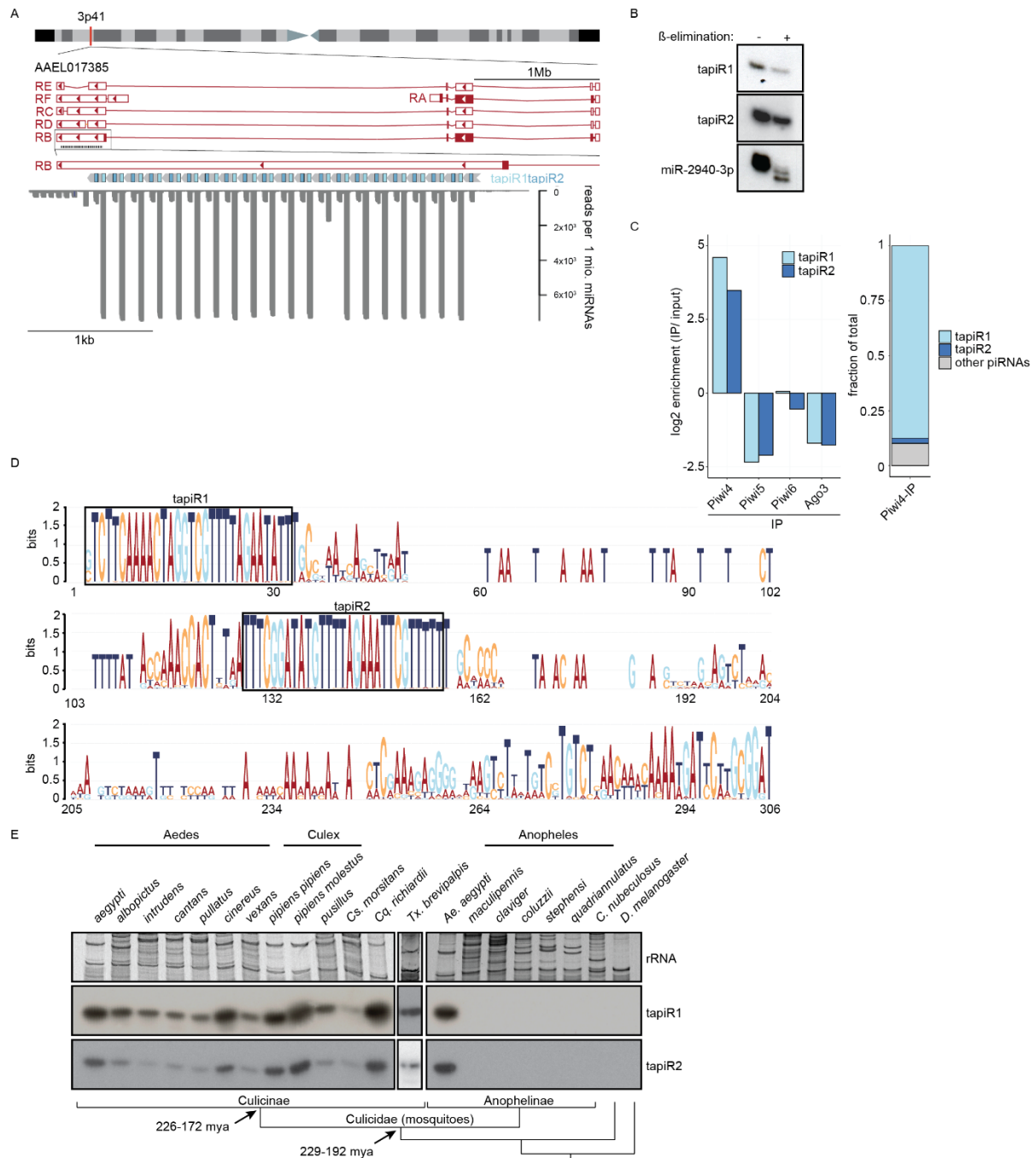
926

927 **Author information**

928 The authors declare no competing financial interests. Correspondence and requests for materials  
929 should be addressed to R.P.v.R. ([ronald.vanrij@radboudumc.nl](mailto:ronald.vanrij@radboudumc.nl)).

930





931

932 **Figure 1:** An evolutionary conserved satellite repeat produces piRNAs that associate with Piwi4.

933 (A) Top panel: Current annotation of the gene AAEL017385 and its splice variants (RA-RF) on  
934 chromosome 3, with the position of the satellite repeat locus and tapiR1 and 2 sequences indicated.  
935 Fill boxes represent open reading frames. Bottom panel: Read coverage of the satellite repeats locus  
936 in Aag2 cells (small RNAs per million mapped miRNAs).

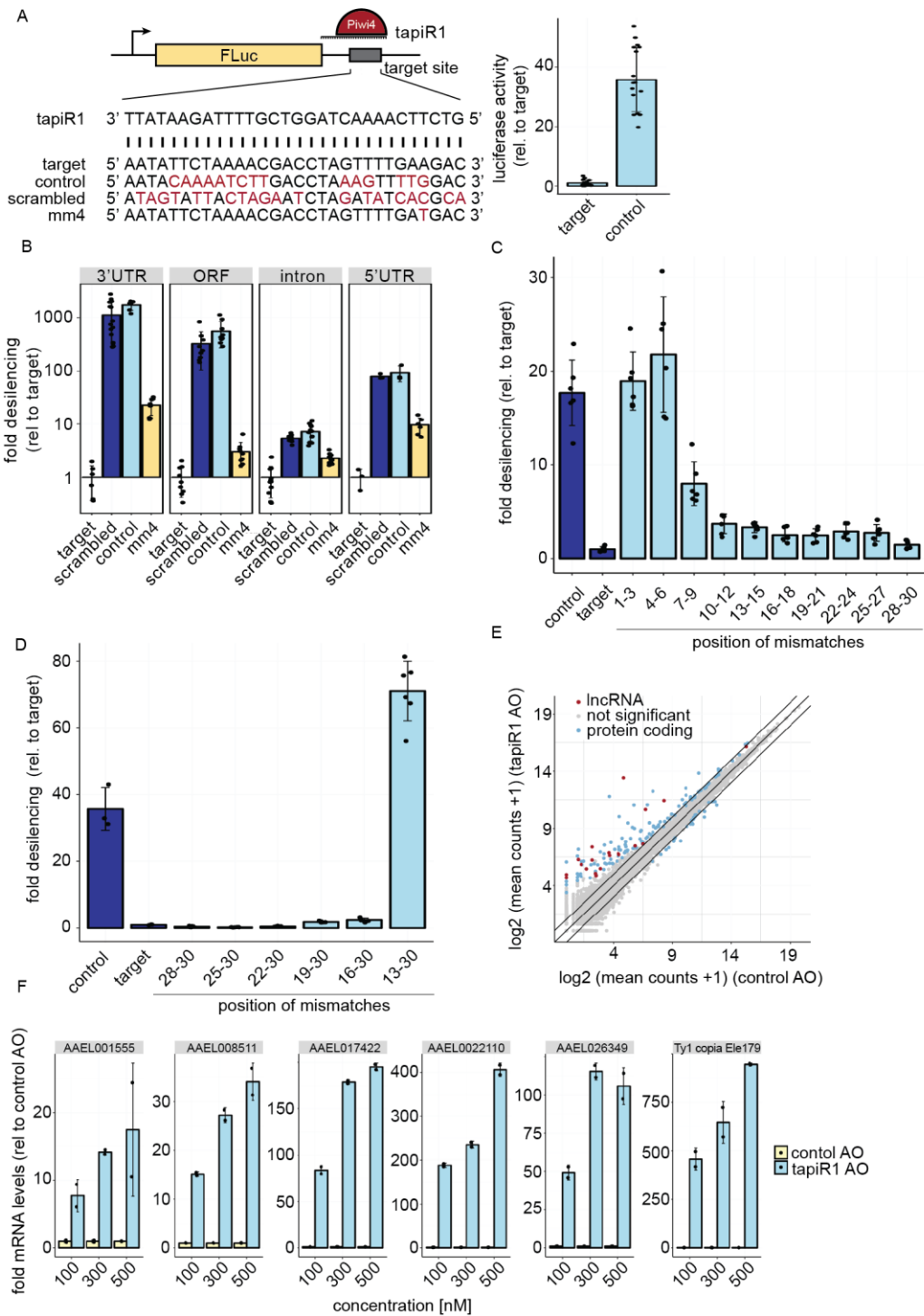
937 (B) Small RNA northern blot of tapiR1 and 2 upon  $\beta$ -elimination in Aag2 cells. miR-2940-3p serves  
938 as positive control for the treatment.

939 (C) Enrichment or depletion of tapiR1 and 2 in small RNA sequencing libraries of the indicated  
940 PIWI-IP compared to input samples (left panel), and fraction of tapiR1 and 2 of total Piwi4-  
941 enriched reads (2-fold, right panel).

942 (D) Sequence conservation of the satellite repeat monomers. All individual repeat monomers from  
943 *Ae. aegypti*, *Ae. albopictus* and *Cx. quinquefasciatus* were used to generate the sequence logo.  
944 Boxes highlight the position of tapiR1 and 2 in the monomer.

945 (E) Northern blot analysis of tapiR1/ 2 in the indicated mosquito species (genera *Aedes*, *Culex*,  
946 *Culiseta*, *Coquilleltidia*, and *Anopheles*) and other insects (*Culicoides* and *Drosophila*). Ethidium  
947 bromide-stained rRNA serves as loading control. For comparison, *Ae. aegypti* was included twice.  
948 Schematic representation of the phylogenetic relationships based on ref.<sup>35</sup> are indicated in the  
949 bottom panel. Bar lengths are arbitrary and do not reflect evolutionary distances.





**Figure 2:** tapiR1 silences target RNAs *in trans* through seed-mediated base-pairing.

(A) Schematic representation of the firefly luciferase (FLuc) reporter constructs (left panel) and luciferase assay in Aag2 cells transfected with reporters containing a fully complementary target site to tapiR1, or a control target site.

(B) Luciferase assay of reporters with a fully complementary tapiR1 target site, a target site with a mismatch at position 4 (mm4), or control sequences introduced at different positions in the reporter mRNA.

(C, D) Luciferase assay of tapiR1 reporters harbouring three consecutive mismatches (C), or

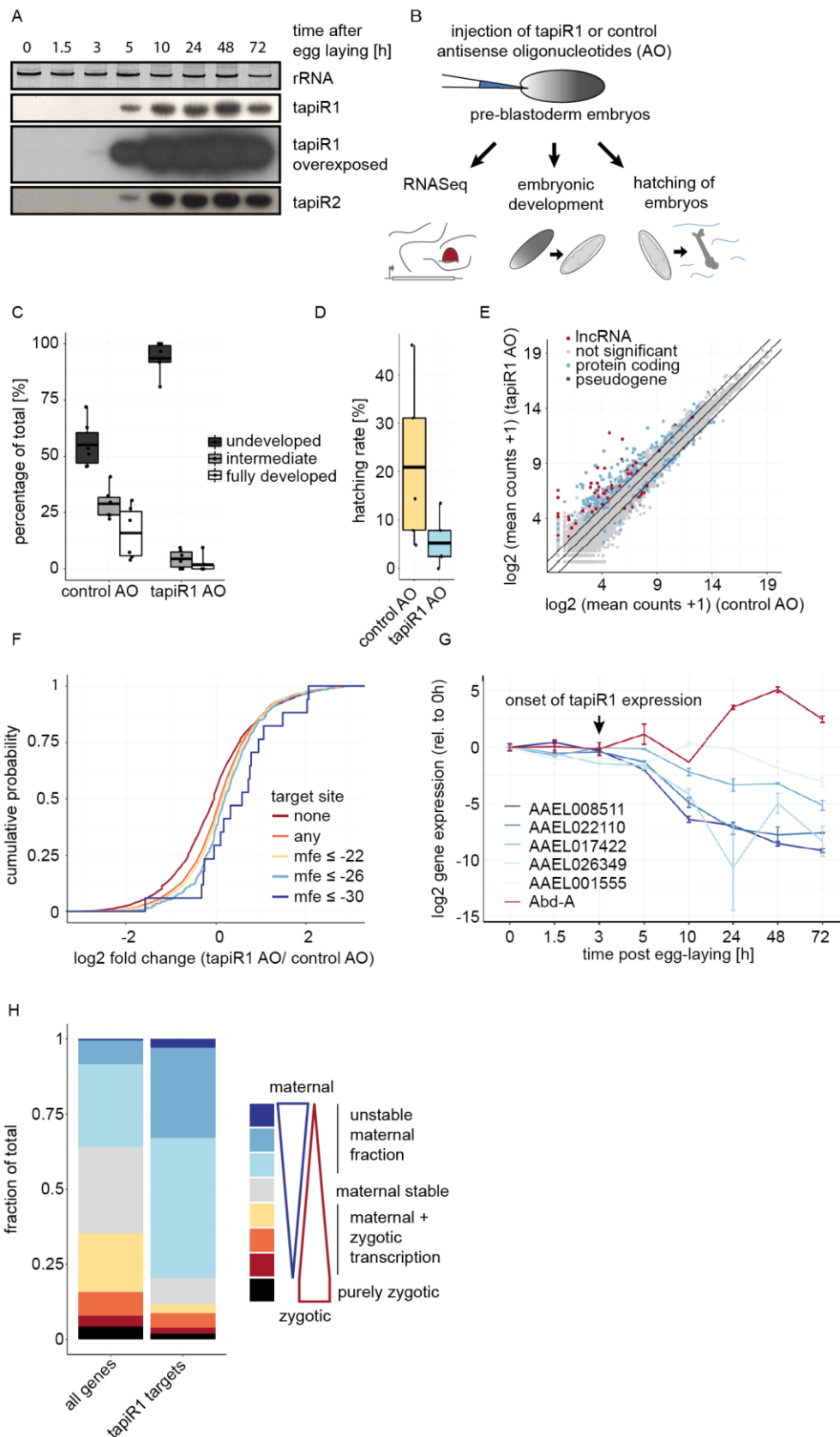
increasing number of mismatches (D) at the indicated positions of the piRNA target site in the

960 3' UTR of firefly luciferase. Firefly luciferase activity was normalized to the activity of a co-  
961 transfected *Renilla* luciferase reporter. Indicated are mean, standard deviation, and individual  
962 measurements of a representative experiment performed with two to three independent clones per  
963 construct and measured in triplicate wells.

964 (E) log<sub>2</sub> expression of mRNAs and lncRNAs in Aag2 cells upon treatment with a tapiR1 specific or  
965 control antisense oligonucleotide (AO). Depicted are average read counts in three biological  
966 replicates. A pseudo-count of one was added to all values in order to plot values of zero. Diagonal  
967 lines indicate a fold change of two. Significance was tested at a false discovery rate (FDR) of 0.01  
968 and a log<sub>2</sub> fold change of 0.5 as indicated by coloured dots.

969 (F) RT-qPCR of tapiR1 target genes upon transfection of Aag2 cells with tapiR1 specific or control  
970 AO. Depicted are mean, standard deviation, and individual measurements of one experiment  
971 measured in technical duplicates.

972



973  
974 **Figure 3:** tapiR1 is essential for embryonic development *in vivo* by promoting turnover of  
975 maternally deposited transcripts.

976 (A) Expression of tapiR1 and 2 in *Ae. aegypti* embryos as analysed by northern blot. Time indicates  
977 the age of the embryos after a 30 min egg laying period. For each time point, around 50 to 150 eggs  
978 were pooled. Ethidium bromide-stained rRNA serves as loading control.

979 (B) Outline of the experimental procedure for panels C-F.

980 (C) Percent of embryos injected with either tapiR1 or control AO that reached the indicated  
981 developmental stages 2.5 days post injection. Individual embryos were scored as either  
982 undeveloped, intermediate or fully developed, as shown in Supplementary Figure S1I. The effect on  
983 tapiR1 AO treatment on development was analysed using a  $X^2$  test of independence:  $X^2(2, N =$   
984  $521) = 105.05, p < 2.2e-16$ .

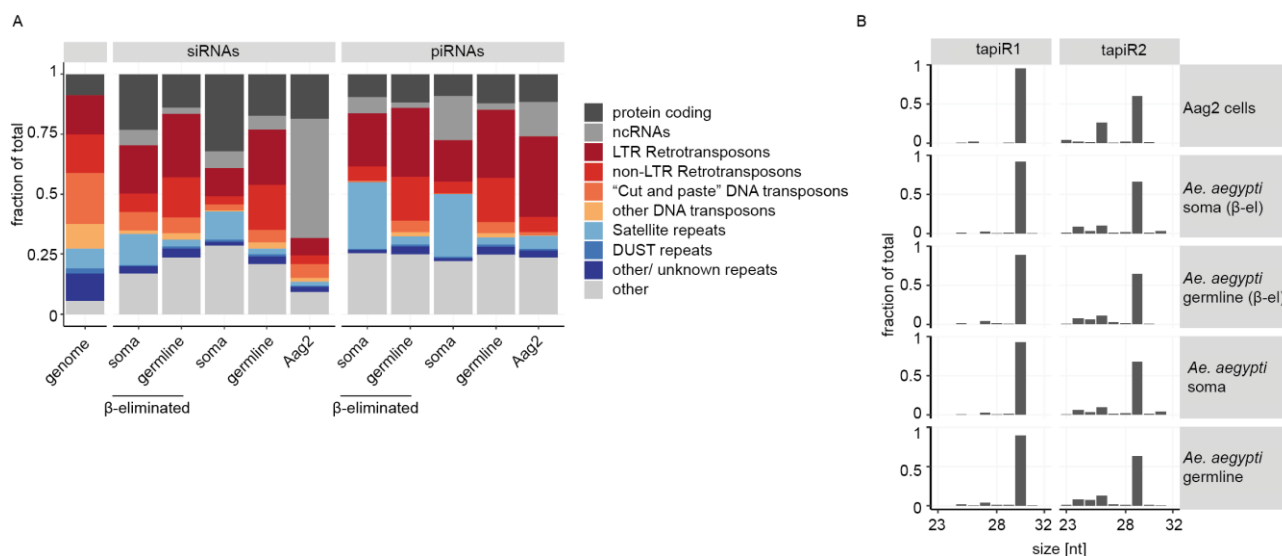
985 (D) Percent of embryos injected with tapiR1 or control AO that hatched four days post injection.  
986 Box-whisker plot represents mean, first and third quartile and maximum and minimum of the data.  
987 Points show the individual experiments with 20 to 60 (C), or 50 to 150 (D) embryos per group.

988 (E) log<sub>2</sub> expression of genes in embryos injected with tapiR1 specific AO or control AO at 20.5 h  
989 post injection. Mean counts from five biological replicates plus a pseudo-count of one are plotted.  
990 Per replicate, 50 embryos per group were pooled. Significance was tested at an FDR of 0.01 and a  
991 log<sub>2</sub> fold change of 0.5. Diagonal lines highlight a fold change of two.

992 (F) Experimental cumulative distribution of log<sub>2</sub> fold changes of genes without or with predicted  
993 target sites for tapiR1. Target sites were grouped based on the predicted minimum free energy (mfe)  
994 of the piRNA/target duplex.

995 (G) Expression of tapiR1 target genes in *Ae. aegypti* embryos. RT-qPCR was performed on samples  
996 shown in (A). Abd-A is a gene not targeted by tapiR1 and serves as negative control.

997 (H) Fraction of genes in different classes expressed in embryos between 0 and 16 h post egg-laying.  
998 Dark blue to dark red represent changes in abundance between 0-2 h and 12-16 h that are larger  
999 than a log<sub>2</sub> fold change of 5, 2, 0.5, between 0.5 and -0.5, or smaller than -0.5, -2, and -5,  
1000 respectively. Genes were considered purely zygotic when they were covered by less than 1 read at  
1001 0-2 h.  
1002



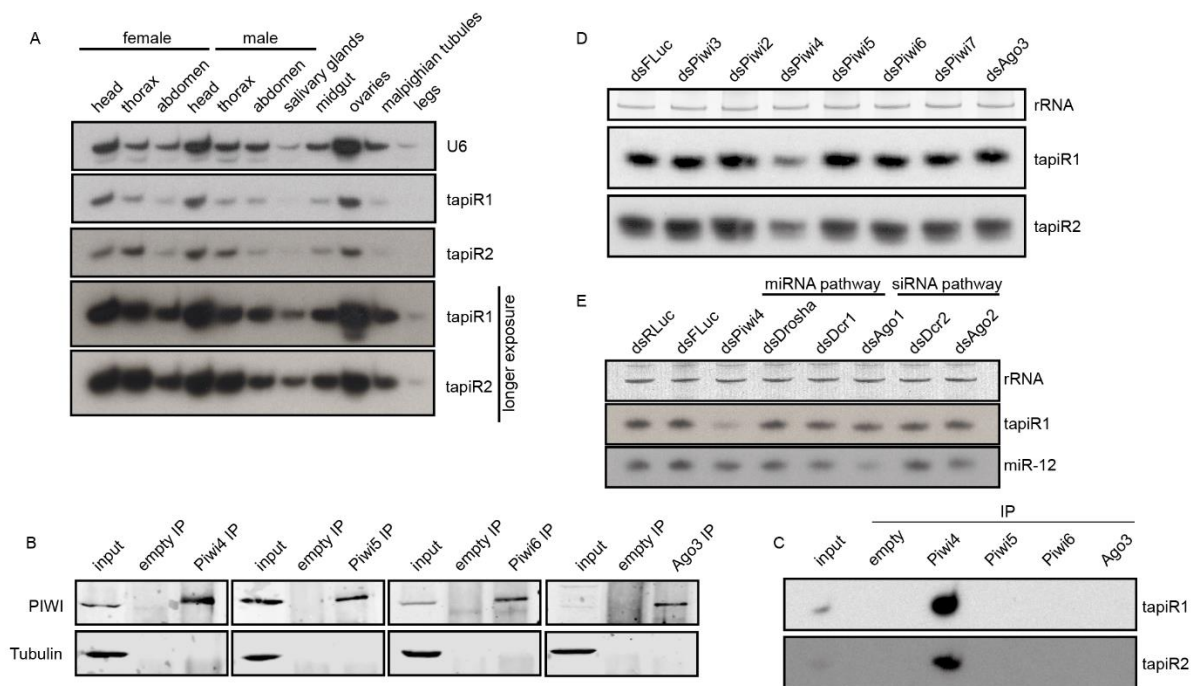
1003

1004 **Extended Data Figure 1:** Expression of piRNAs from a satellite repeat locus.

1005 (A) Fraction of siRNAs and piRNAs mapping on genomic features in adult *Ae. aegypti* female  
 1006 ovaries (germline) or carcasses (soma). Small RNAs that overlapped multiple features were  
 1007 assigned to only one category (see Methods). The leftmost bar depicts the fraction of each feature  
 1008 category in the genome.

1009 (B) Read length distribution of tapiR1 and 2 in Aag2 cells, and adult germline and somatic tissues  
 1010 (β-eliminated, or untreated).

1011



1013

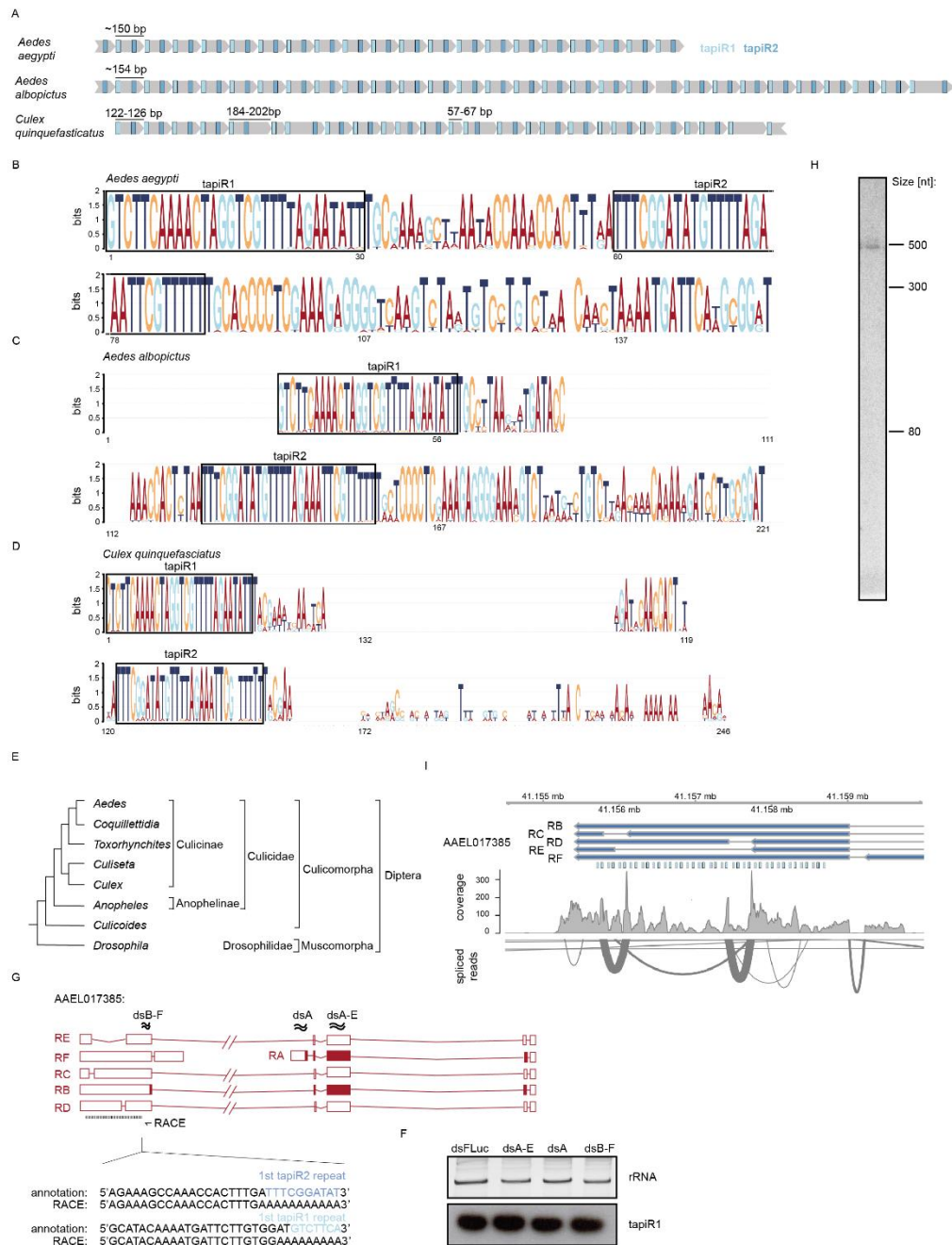
1014 **Extended Data Figure 2:** tapiR1 and 2 are expressed in *Ae. aegypti* mosquitoes and associate with  
1015 Piwi4.

1016 (A, D, E) Northern blot of tapiR1 and 2 in different tissues of adult mosquitoes (A), upon dsRNA-  
1017 mediated knockdown of individual PIWI genes (D), and upon knockdown of miRNA and siRNA  
1018 pathway genes (E), or a control dsRNA treatment (dsFLuc and dsRLuc) in Aag2 cells. U6 snRNA  
1019 or ethidium bromide-stained rRNA serves as loading controls.

1020 (B) Western blot analysis of the indicated PIWI proteins before (input) and after  
1021 immunoprecipitation (IP) used for the small RNA northern blot of panel C. An IP with empty beads  
1022 serves as negative control. Tubulin was used to control for non-specific binding.

1023 (C) Immunoprecipitation of the indicated PIWI proteins from Aag2 cells followed by northern blot  
1024 analyses for tapiR1 and 2.

1025



1026

1027 **Extended Data Figure 3:** Expression of tapiR1 is conserved in culicine mosquitoes, and is  
 1028 independent of AAEL017385.

1029 (A) Schematic representation of the tapiR satellite repeat locus in *Ae. aegypti*, *Ae. albopictus* and  
 1030 *Cx. quinquefasciatus*. Numbers indicate repeat lengths, and, for *Cx. quinquefasciatus*, lengths of  
 1031 deviating repeat monomers.

1032 (B-D) Sequence logos constructed from all individual tapiR repeat units in *Ae. aegypti* (B), *Ae.*  
 1033 *albopictus* (C), or *Cx. quinquefasciatus* (D). Gaps in the sequence logos mainly arise due to size  
 1034 heterogeneity in few repeat monomers.

1035 (E) Evolutionary relationships of dipterous genera based on ref.<sup>35</sup>. Bar lengths are arbitrary and do  
 1036 not reflect evolutionary distances.



1037 (F) Northern blot of tapiR1 in Aag2 cells treated with dsRNA targeting different transcripts of  
1038 AAEL017385 (indicated in panel G), or, as control, firefly luciferase (FLuc). Ethidium bromide  
1039 stained rRNA serves as loading control.

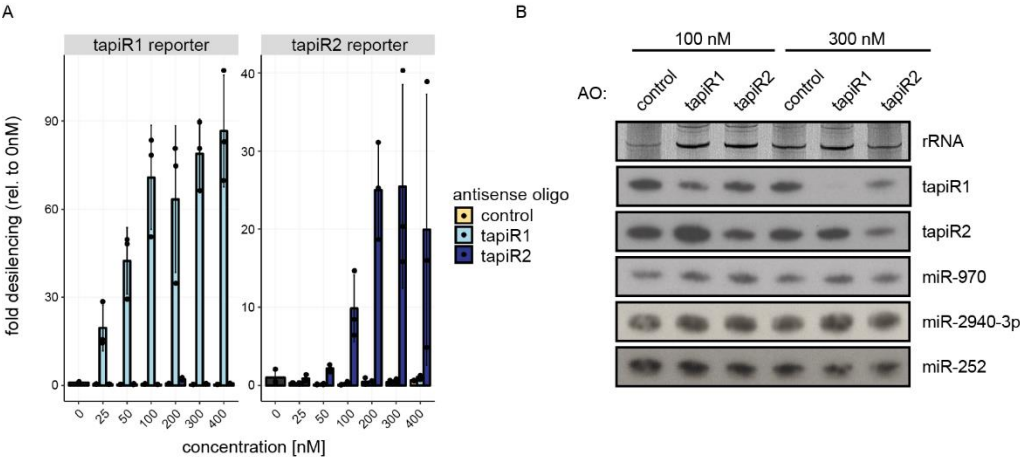
1040 (G) Top panel: Schematic representation of the AAEL017385 locus and satellite repeat. The primer  
1041 used for 3' RACE, and positions targeted by dsRNA in panel F are indicated with an arrow and  
1042 wavy lines, respectively.

1043 Bottom panel: 3' RACE analysis of AAEL017385. Indicated are sequences from the current  
1044 AaegL5 genome annotation and RACE PCR products. The sequences of the 5' terminal part tapiR1  
1045 and 2 repeats are highlighted with colours.

1046 (H) Northern blot of a potential tapiR1/2 precursor transcript.

1047 (I) RNAseq read coverage of the tapiR repeat locus and AAEL017385 (top panel), and sashimi plot  
1048 indicating spliced reads (bottom panel).

1049



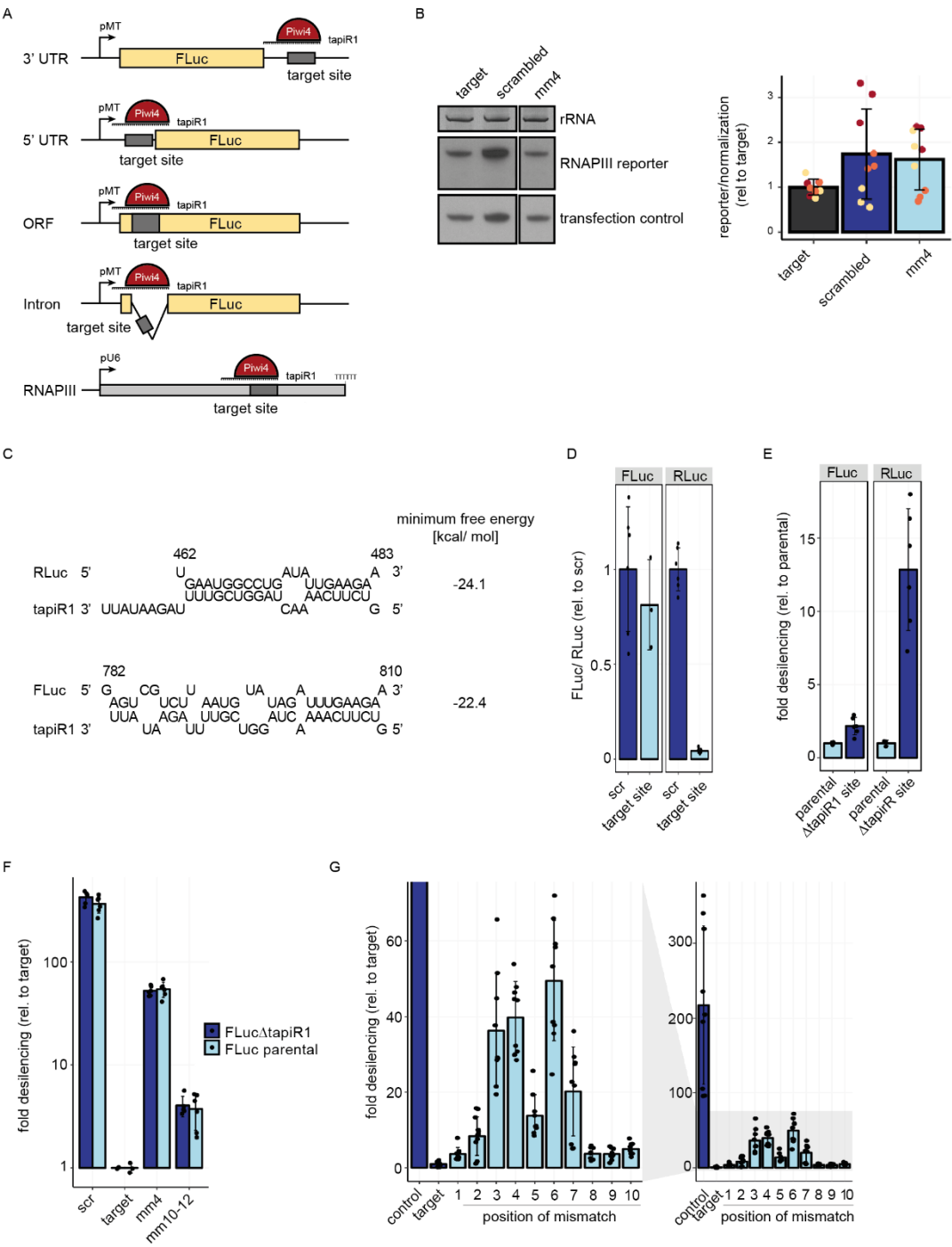
1051

1052 **Extended Data Figure 4:** Antisense oligonucleotides relieve tapiR1-mediated silencing.

1053 (A) Luciferase assay of a reporter with a fully complementary target site for tapiR1 (left panel), or  
1054 tapiR2 (right panel) in the 3' UTR. Cells were co-transfected with the reporter and increasing  
1055 amounts of a fully 2'O-methylated antisense RNA oligonucleotide (AO), or a control AO. Firefly  
1056 luciferase activity was normalized to the activity of a co-transfected *Renilla* luciferase reporter.  
1057 Indicated are mean, standard deviation and individual measurements from a representative  
1058 experiment measured in triplicate wells.

1059 (B) Northern blot detection of tapiR1, tapiR2, and three different miRNAs in Aag2 cells upon  
1060 treatment with a the indicated concentrations of tapiR1, tapiR2, or control AO. Ethidium bromide-  
1061 stained rRNA serves as loading control.

1062



1065 **Extended Data Figure 5: *Renilla* luciferase contains a functional tapiR1 target site.**

1066 (A) Schematic representation of the different reporter constructs used in this study. pMT:

1067 metallothionein promoter; RNAPIII: RNA polymerase III reporter.

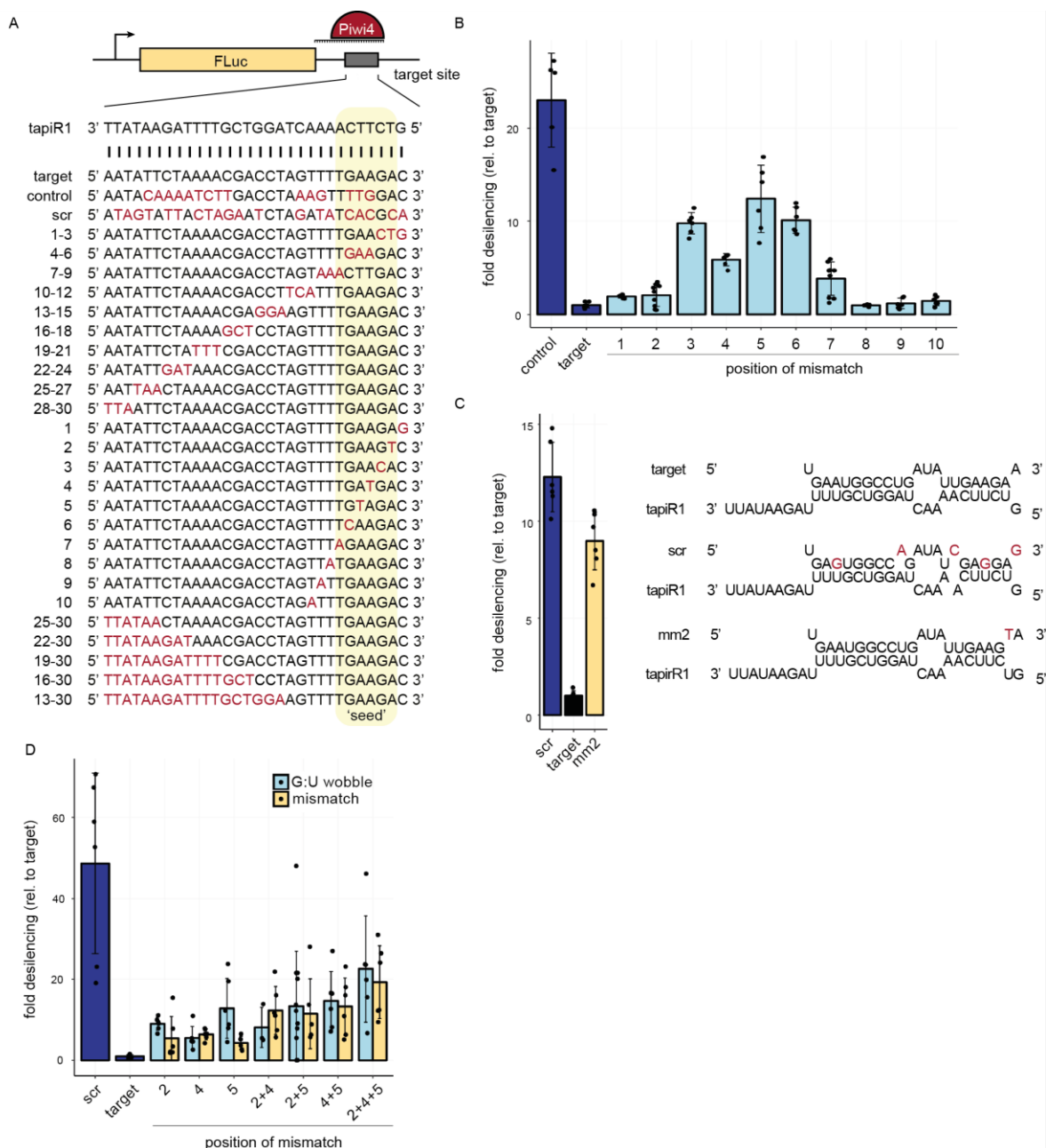
1068 (B) Representative northern blot (right panel), and quantification of an RNAPIII reporter. Values

1069 are normalized to a non-targeted transfection control. Mean, standard deviation, and individual

1070 measures of three independent experiments (indicated with colours) quantified in triplicates are

1071 shown.

1072 (C) Schematic representation of predicted tapiR1 target sites and minimum free energy of the  
1073 indicated structures in the coding sequences of *Renilla* luciferase (RLuc) or firefly luciferase  
1074 (FLuc). Numbers indicate the position of the targets relative to the first nucleotide in the ORFs.  
1075 (D) Luciferase assay of Aag2 cells transfected with reporters carrying either a scrambled (scr) site  
1076 or the predicted target site from firefly luciferase (left panel), or from *Renilla* luciferase (right  
1077 panel) from panel (C) in the 3'UTR of FLuc.  
1078 (E) Luciferase activity of FLuc or RLuc constructs with synonymous mutations introduced into the  
1079 predicted tapiR1 target site ( $\Delta$ tapiR1 site) and the parental clones.  
1080 (F) Luciferase reporter assay of reporters carrying target sites for tapiR1 as indicated in panel C in  
1081 the 3'UTR of either the parental firefly luciferase, or the  $\Delta$ tapiR1 firefly luciferase version.  
1082 (G) Reporter assay with luciferase carrying tapiR1 target sites with single mismatches in the  
1083 3' UTR as used in Extended Data Fig 6B, using RLuc with a mutated tapiR1 target site ( $\Delta$ tapiR1  
1084 site) for normalization. Left panel is a zoom to the x-axis of the right panel. Shown are mean,  
1085 standard deviation and individual measurements from representative experiments performed with at  
1086 least two different clones per construct, and each measured in triplicate wells.



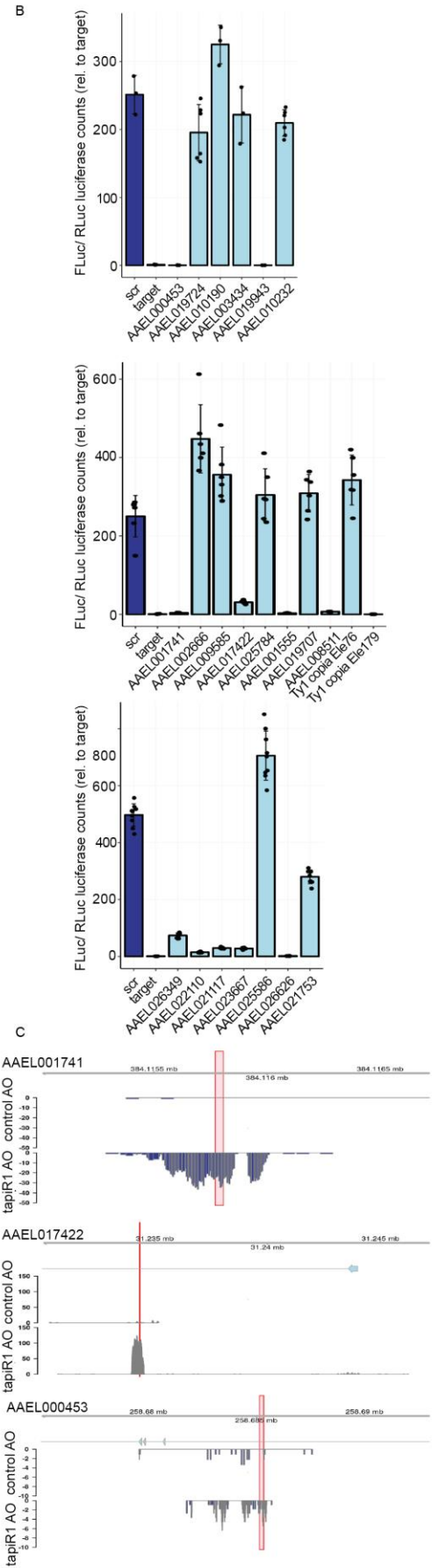
1099 *Renilla* luciferase reporter to control for differences in transfection efficiencies. Data represent  
1100 mean, standard deviation and individual measurements of representative experiments with two  
1101 independent clones per construct and measured in triplicates wells.

1102



**A**

target sites functional:		minimum free energy [kcal/mol]:
AAEL026349	5' C GGUGUU A AAGAC CU UCUGG A UUUGAAGA A 3'	-24.5
tapiR1	3' UUAUA GA UUUUG CU A AACUUCU G 5'	
AAEL022110	5' G AUG UCUGA AGUAA AACGAC G UGG CCG UUGAAGA A 3'	-28.5
tapiR1	3' UUAU A AGAUU AACGUG AUC AA AACUUCU G 5'	
AAEL021117	5' A AUUC GACGACCUA CCAAC UGAAGAC U 3'	-28.6
tapiR1	3' UUA UAAG AUU UUGCUGGAU CAAA ACUUCUG 5'	
AAEL023667	5' U UUC GCU GCGGACCU CC UUUUGAAGA U 3'	-28.3
tapiR1	3' UUAU AUU UUGCUGGA UCA AACUUCUG 5'	
AAEL026626	5' C UUUAAAGCGACUUAAG AA UUGAAGA A 5'	-31.8
tapiR1	3' UUAUA AGAUUUUGCUGGAUC AA AACUUCU G 3'	
AAEL001555/ AAEL022223/ AAEL023448	5' A AAAAUGG CUAGUUUUGAAGA G 3'	-28.0
tapiR1	3' UUAUAAGA UUUUGCU GAUCAAACUUCU G 5'	
AAEL008511	5' C UAUUU GGCGGACUUGGUU UGAAGAU U 3'	-31.6
tapiR1	3' UU UAAG AU UUGCUGGAUCA A ACUUCUG 5'	
AAEL019943	5' U UUCU CCAU CGAUCUAGUUUUGAAGA A 5'	-32.9
tapiR1	3' UUAU AAGA UUUU GCUGGAUCAAAACUUCU G 3'	
Ty1 copia Ele179	5' U GAUGUUUUA UU UU CCUGGU C UUGAAGA A 3'	-29.7
tapiR1	3' UUAUAAGAU UUU GC GAUCA AACUUCU G 5'	
AAEL000453	5' C CUAACACGA AA AGUUUUGAAGA A 5'	-30.0
tapiR1	3' UUAUA GAUUUUGCU UCAAAACUUCU G 3'	
AAEL001741	5' U UAUUUU GGUGGA GAAACGGCCUA U GC UGAAGA A 5'	-27.5
tapiR1	3' UU AUAGA UUUUGCUGGAU C AA ACUUCU G 3'	
AAEL017422	5' G AUUC ACGACCUA CCAAC C UGAAGAU C 5'	-25.1
tapiR1	3' UUA UAAG AUUUU UGCUUGAU CAAA ACUUCUG 3'	
target site not functional:		
AAEL019724	5' C GACGGCCUGG ACGGUA UGAAGA A 3'	-25.7
tapiR1	3' UUAUAAGAUU UUGCUGGAUC AAA ACUUCU G 5'	
AAEL0003434	5' A AUUUUGGGAUGAUC AAU A UUGAAGA U 3'	-27.0
tapiR1	3' UUA UAAGAUUUUGCUGG AU C AACUUCUG 5'	
AAEL010190	5' U GAUGACCUUGG A UGAAGA G 3'	-27.2
tapiR1	3' UUAUAAGAUU UUGCUGGAU AAA ACUUCU G 5'	
AAEL010232	5' U AGCGGCUU UCCA UUGAAGA G 3'	-22.4
tapiR1	3' UUAUAAGAUU UUGCUGGA UCAA AACUUCUG 5'	
AAEL002666	5' GAUAU GAGA CU GCG CU G AGU UGAAGA A 3'	-21.1
tapiR1	3' G UUAUA AG AUUU UGC U A AA ACUUCU G 5'	
AAEL009585	5' G UC UC AAGAAUCGUCACUUC GA AUUC UGAAGA G 3'	-17.7
tapiR1	3' UUAUA AGA AUU AACG UG UC AG UC AAA ACUUCUG 5'	
AAEL025784	5' A CU GAGCGA AG G A UUUUGAAGAC A 3'	-26.0
tapiR1	3' UUAUA GA UUUUGCU G AUC AAAACUUCUG 5'	
AAEL019707	5' G AGA UG GAUCUG A UUGAAGA A 3'	-23.0
tapiR1	3' UUAUAAGA UUG CUGGAU CAA AACUUCU G 5'	
AAEL025586	5' A GAUGU UGG GGAAGAC AGUC GA UGAAGAC G 3'	-23.9
tapiR1	3' UUAUA A GAU GAU UUGCUG GAU CA AA ACUUCUG 5'	
AAEL021753	5' G AAUGUUU CUCU C AUCUG A UUUUGAAGA G 3'	-23.3
tapiR1	3' UUAUAAG AUUU C UGGAU C AAAACUUCU G 5'	
Ty1 copia Ele76	5' U UUCUG U AGACG UACGAUUUGAGAAUU UCUCUAUC GGAAGCUUCA C 3'	-20.7
tapiR1	3' UUAU AAGAU UUUUGC UGGAU GUUU CAAA UGAAGAU ACUUCUG 5'	



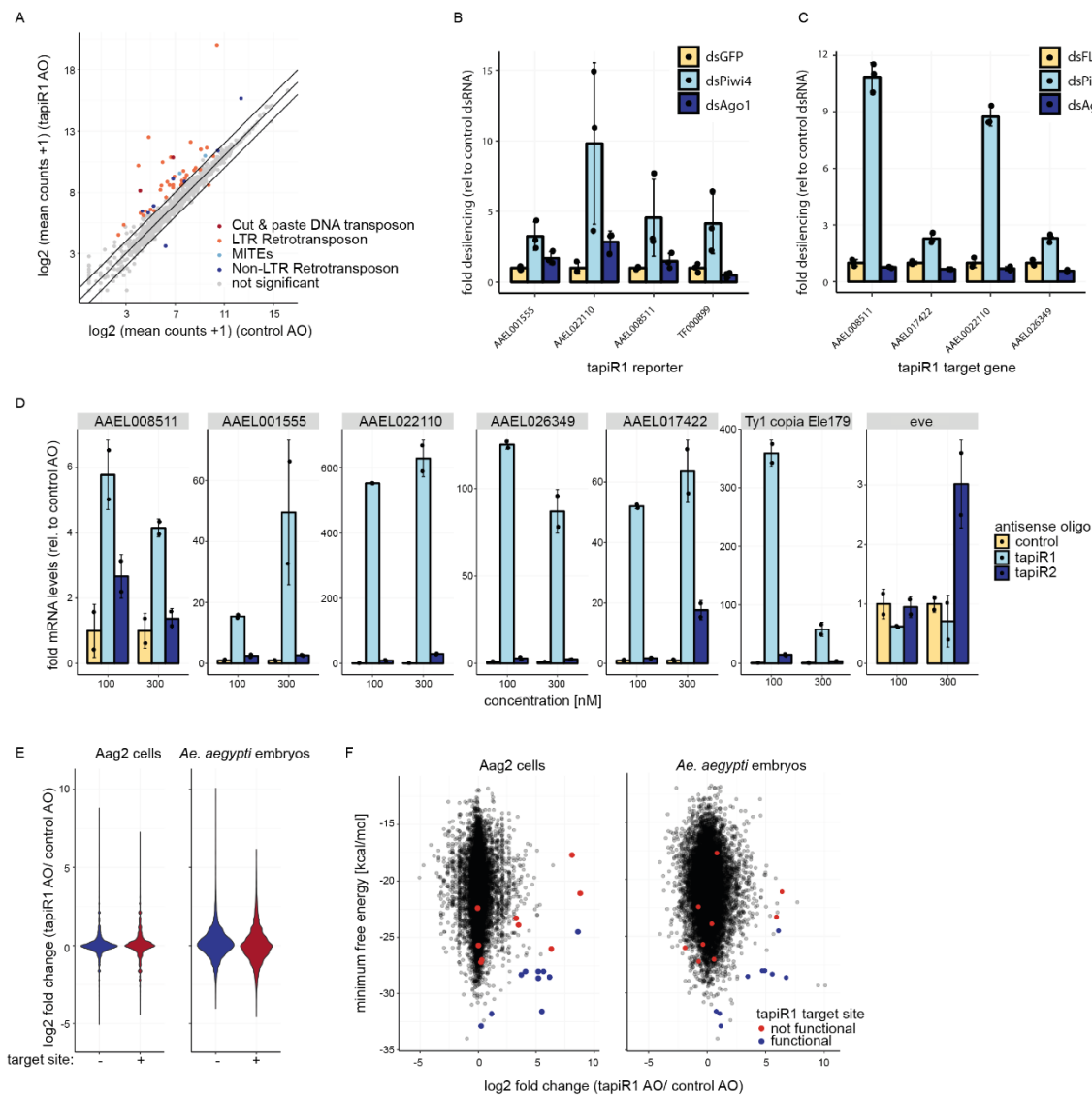
1105 **Extended Data Figure 7:** Validation of tapiR1 target genes.

1106 (A) Predicted structures and minimum free energy of tapiR1/target duplexes analysed in panel B.

1107 (B) Luciferase assay of reporters carrying the predicted target sites from panel A in the 3' UTR of  
1108 firefly luciferase. Firefly luciferase activity was normalized to the activity of a co-transfected  
1109 *Renilla* luciferase reporter to control for differences in transfection efficiencies. Indicated are mean,  
1110 standard deviation and individual measurements from representative experiments performed with  
1111 two to three independent clones per construct and measured in triplicate wells.

1112 (C) AAEL017422, AAEL001741, and AAEL000453 were annotated in the previous AaegL3 gene  
1113 set, but not in the current AaegL5 gene set. Read coverage in tapiR1 AO and control AO treated  
1114 Aag2 cells at these genomic regions suggests that these regions are actively transcribed, but  
1115 repressed by tapiR1. Red boxes indicate the positions of tapiR1 target sites.

1116



1119 **Extended Data Figure 8: *tapiR*/Piwi4 silence gene expression in Aag2 cells.**

1120 (A) log2 mRNA expression of transposable elements in Aag2 cells treated with a *tapiR1* specific  
1121 antisense oligonucleotide (AO) or control AO. Depicted are the means of three biological replicates.  
1122 A pseudo-count of one was added to all values in order to plot values of zero. Diagonal lines  
1123 represent a fold change of two. Significance was tested at an FDR of 0.01 and a log2 fold change of  
1124 0.5.

1125 (B) Luciferase assay of reporters harbouring different *tapiR1* target sites (from Extended Data Fig  
1126 7) in the 3' UTR of firefly luciferase. Firefly luciferase activity was normalized to the activity of a  
1127 co-transfected *Renilla* luciferase reporter to control for differences in transfection efficiencies. Data  
1128 represent mean, standard deviation and individual measurements of representative experiments  
1129 measured in triplicate wells.

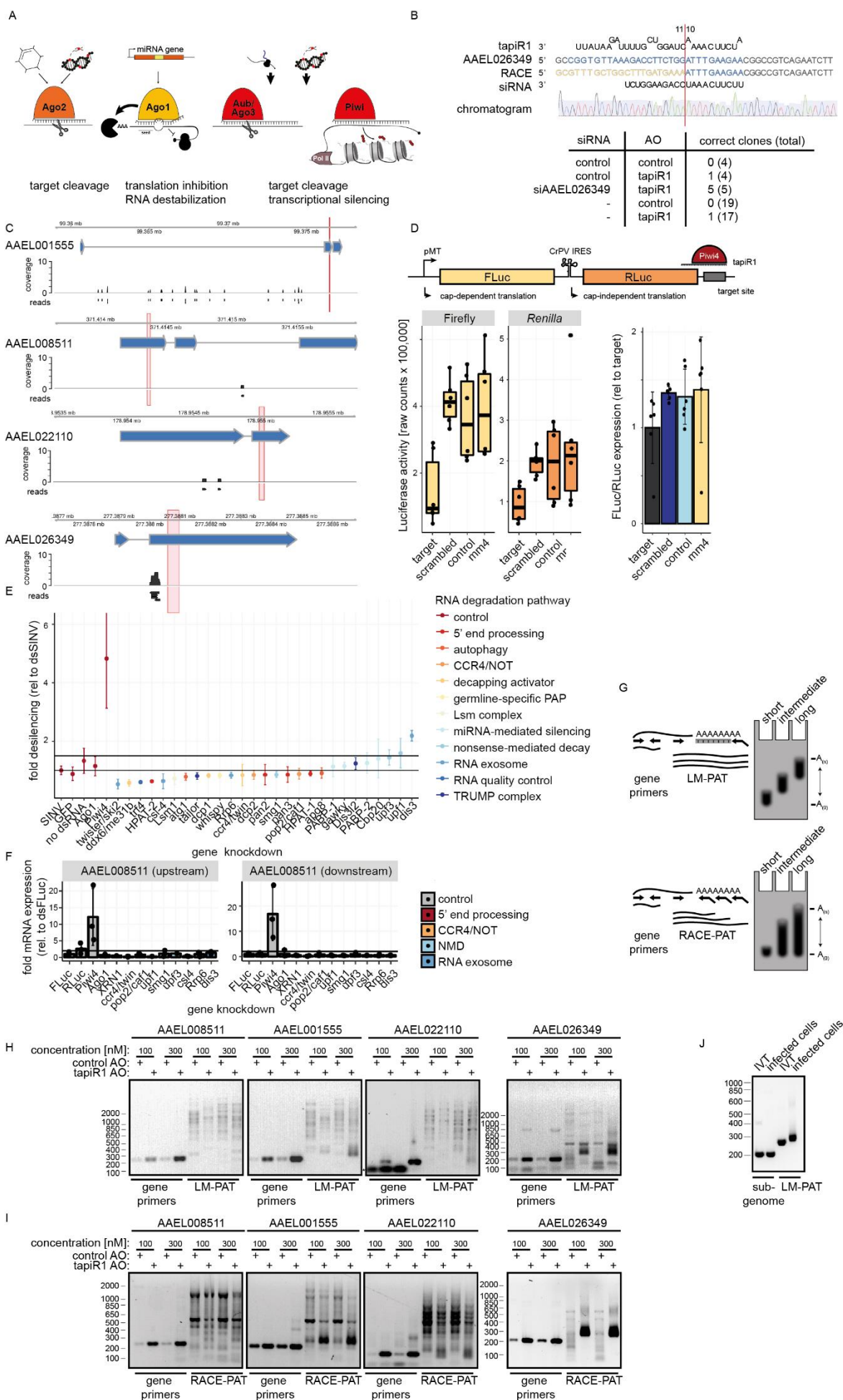
1130 (C, D) RT-qPCR of *tapiR1* target genes upon dsRNA-mediated knockdown of FLuc (control),  
1131 Piwi4, or Ago1 in Aag2 cells (C), or after treatment with different concentrations of control, *tapiR1*,

1132 or tapiR2 AO (D). Depicted are mean, standard deviation, and individual measurements of a  
1133 representative experiment as measured from triplicate wells (C), or from duplicate wells (D). Even  
1134 skipped (eve) does not harbor a tapiR1 target site, and serves as control.

1135 (E) Experimental cumulative distribution of log<sub>2</sub> fold changes of genes without or with predicted  
1136 target sites for tapiR1. Target sites were grouped based on the predicted minimum free energy of  
1137 the piRNA/target duplex.

1138 (F) log<sub>2</sub> fold changes of genes upon treatment with tapiR1 or control AO in Aag2 cells (left) and  
1139 mosquito embryos (right) plotted against the minimum free energy of predicted tapiR1-target  
1140 duplexes. Blue dots indicate target sites that were confirmed to be functional, and red dots indicate  
1141 target sites that were not functional in luciferase reporter assays (see Extended Data Fig 7).

1142 (G) Violin plot of log<sub>2</sub> fold changes of all genes in Aag2 cells (left) and mosquito embryos (right),  
1143 either with or without predicted tapiR1 target site.



1145 **Extended Data Figure 9:** tapiR1 regulates gene expression on a post-transcriptional level.

1146 (A) Schematic representation of different modes of silencing of all three small RNA silencing  
1147 pathways in *Drosophila*.

1148 (B) Top panel: Sequences of tapiR1, target gene AAEL026349, an siRNA targeting the gene at the  
1149 same position, and Sanger sequencing results of the 5' RACE of Aag2 cells treated with tapiR1 AO  
1150 and siRNAs. The tapiR1 target site is indicated in blue, RACE sequencing adapter in yellow, gene  
1151 sequence in dark grey. The predicted splice site between nucleotides 10 and 11 is marked with a red  
1152 vertical line. Bottom panel: Summary of the 5' RACE in the indicated conditions. Numbers refer to  
1153 the number of sequenced clones with the 5'RACE adapter ligated to the predicted splice site, and the  
1154 total number of sequenced clones between brackets.

1155 (C) Small RNA coverage in Aag2 cells (unnormalized) and individual reads (direction of the arrow  
1156 indicates the strand) on tapiR1 target genes. Red boxes indicate positions of the tapiR1 target site on  
1157 the mRNA.

1158 (D) Schematic representation (top panel) and luciferase expression (bottom panels) of IRES-  
1159 containing reporter constructs. Depicted are mean, standard deviation, and individual measurements  
1160 of a representative experiments performed in triplicate wells with two different reporter clones. The  
1161 bottom left panel illustrates firefly luciferase expression normalized to *Renilla* luciferase, while  
1162 bottom right panels show raw luciferase counts from the same experiment.

1163 (E) Luciferase activity of a reporter harbouring the tapiR1 target site of AAEL001555 in the 3'  
1164 UTR of FLuc upon dsRNA-mediated knockdown of the indicated genes. FLuc expression was  
1165 normalized to RLuc expression to control for differences in transfection efficiencies. Depicted are  
1166 mean and standard deviation of one experiment performed in triplicates.

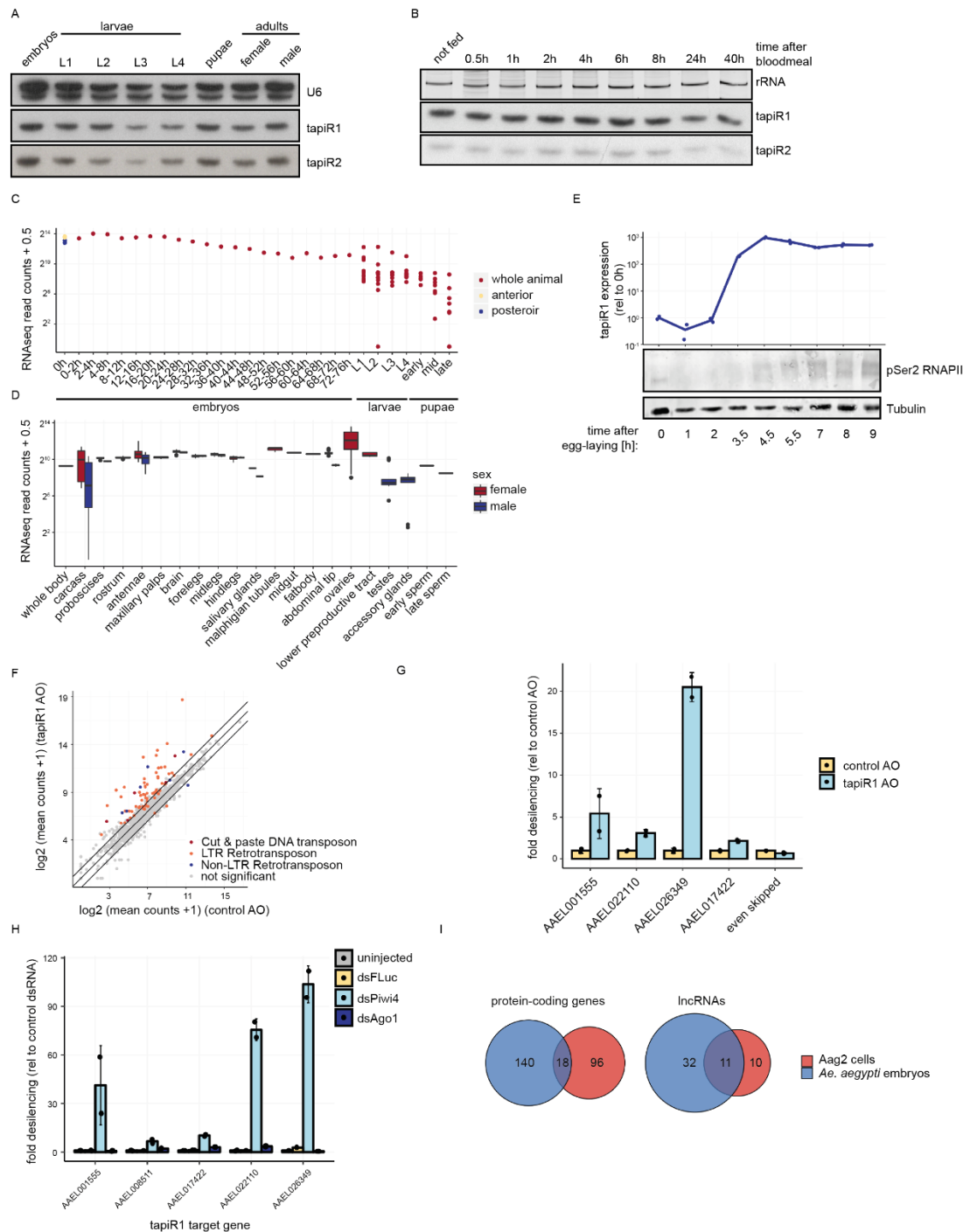
1167 (F) AAEL008511 target gene expression as measured by RT-qPCR upon knockdown of the  
1168 indicated genes. Mean, standard deviation, and individual measurements of one out of three  
1169 experiments performed in triplicate are shown. The other two experiments are presented in  
1170 Supplementary Fig. S1F,G. Bars are color-coded according to the indicated RNA decay pathways.

1171 (G) Schematic illustration of the two poly(A) tail length (PAT) assays and expected results of genes  
1172 with increasing poly(A) tail lengths. LM-PAT, ligation mediated-PAT.

1173 (H-J) Electrophoretic analysis with ethidium bromide stained agarose gels of a LM-PAT assay (H),  
1174 or RACE-PAT assay (I) of different tapiR1 target genes upon treatment with two concentrations of  
1175 tapiR1 or control AO. As positive control, poly(A) tail length was measured from Sindbis virus  
1176 (SINV) RNA *in vitro* transcribed from a plasmid, or in infected Aag2 cells where the poly(A) tail is  
1177 elongated during viral replication<sup>61</sup> (J).

1178





1181 **Extended Data Figure 10:** *tapiR1* regulates gene expression in mosquito embryos.

1182 (A, B) Northern blot analysis of *tapiR1* and 2 in developmental stages of *Ae. aegypti*  
1183 mosquitoes (A), or at different time points after blood feeding (B). U6 snRNA (A) or ethidium  
1184 bromide-stained rRNA (B) were analysed to verify equal loading.  
1185 (C, D) RNAseq read coverage of *Piw4* in the indicated adult *Ae. aegypti* tissues (C), or  
1186 developmental stages (D). Libraries used are listed in Supplementary Table S6.

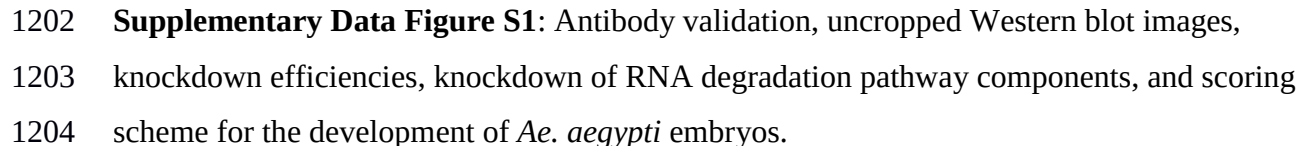
1187 (E) Western blot analysis of phosphorylated polymerase II (Ser2) in embryos at the indicated  
1188 timepoints after egg-laying (bottom panel), and corresponding stem-loop RT-qPCR of tapiR1  
1189 measured in technical duplicates.

1190 (F) log<sub>2</sub> mRNA expression of transposable elements in embryos injected with tapiR1-specific or  
1191 control AO. Mean counts of five biological replicates are shown. Significance was tested at an FDR  
1192 of 0.01 and a log<sub>2</sub> fold change of 0.5. Diagonal lines indicate a fold change of two.

1193 (G, H) RT-qPCR of the indicated tapiR1 target genes 9h after injection of control or tapiR1-specific  
1194 AOs (G), or dsRNA-mediated knockdown of FLuc (control), Piwi4, or Ago1 (H) in embryos.  
1195 Mean, standard deviation and individual measurements of a representative experiment are  
1196 presented. Even skipped (*eve*) is not a tapiR1 target gene and serves as negative control.

1197 (H) Overlap of upregulated tapiR1 target genes (log<sub>2</sub> fold change  $\geq 1$ , with a predicted target site  
1198 with mfe  $\leq 24$ ) in Aag2 cells and *Ae. aegypti* embryos.

1199



1207

1208

1209 (B) Uncropped Western blot images corresponding to Extended Data Fig 2B.  
1210 (C-E) Knockdown efficiencies of PIWI genes shown in Extended Data Fig 2D (C), siRNA and  
1211 miRNA pathway genes shown in Extended Data Fig 2E (D), and AAEL017385 isoforms in the  
1212 experiment shown in Extended Data Fig 3F (E).  
1213 (F,G) RT-qPCR of the tapiR1 target gene AAEL008511 upon knockdown of the indicated genes.  
1214 Presented are two independent repeats of the experiment shown in Extended Data Fig 9G.  
1215 (H) Uncropped Western blot images corresponding to Extended Data Fig 10E.  
1216 (I) Representative images of embryos scored as either undeveloped, intermediate or fully developed  
1217 at 2.5 days post injection with antisense RNA oligonucleotides.  
1218