A potential role of inefficient and non-specific piRNA production from the whole transcriptome

Keisuke Shoji[1,2,3*], Yukihide Tomari[1,2,*]

[1]Laboratory of RNA Function, Institute for Quantitative Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan.

[2]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0032, Japan.

[3]Graduate school of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology, Koganei-shi, Tokyo 184-8588, Japan.

*Email: kshoji@go.tuat.ac.jp (K.S.) and tomari@iqb.u-tokyo.ac.jp (Y.T.)

Author Contributions: K.S. designed the study. K.S. performed bioinformatic analysis of sequencing data. Y.T. supervised the study. K.S. and Y.T. wrote the manuscript.

Competing Interest Statement: The authors declare no competing interests.

Keywords: piRNA, RNA virus, transposons.

Abstract

PIWI-interacting RNAs (piRNAs) are small RNAs that play a critical role in repressing transposons in the germline by guiding PIWI proteins to complementary target sequences. It is well known that the 5' ends of piRNAs are typically defined through cleavage by PIWI proteins or by the endonuclease Zucchini, and piRNAs are amplified via reciprocal cleavage of sense and antisense transcripts by PIWI proteins. Here, we show that low levels of sense-stranded piRNAs are produced from the whole transcriptome in silkworms, flies, and mice, in a manner independent from the known biogenesis pathways. Notably, this mechanism is utilized in silkworm cultured cells for the production of piRNAs derived from the extremely abundant sub-genomic RNAs of *Bombyx mori* latent virus, which have been suggested to play a role in suppressing viral replication. We propose that this inefficient and non-specific piRNA-production pathway acts as a primitive sensor for aberrantly abundant transcripts in the germline, potentially seeding piRNA sequence repertoire for more efficient and established biogenesis pathways.

Main Text

Introduction

PIWI-interacting RNAs (piRNAs) are a class of small RNAs of ~24–32 nucleotides (nt) inlength expressed in the germline. They function by associating with PIWI-subfamily proteins, guiding them to repress transposons (1, 2). Many piRNAs are derived from genomic regions called piRNA clusters, which serve as reservoir for transposon fragments and are often transcribed in the antisense orientation (3, 4). Consequently, these piRNAs possess sequences complementary to transposons, allowing the piRNA-PIWI complex to recognize and cleave transposon mRNAs for silencing (5–8). After the cleavage, the 3' fragment of the cleavage product is subsequently loaded into another PIWI protein and acts as a precursor for a new piRNA, which shows 5' 10-nt overlap with the original piRNA. The new piRNA then reciprocally cleaves complementary RNAs, supplying new precursors of piRNAs. This piRNA amplification pathway, which depends on the endoribonuclease (slicer) activity of PIWI proteins, is known as the ping-pong cycle (3, 5). In addition, piRNA-guided RNA slicing also recruits an endonuclease named Zucchini (Zuc) downstream of the PIWI protein at the cleavage site. The 3' fragment of the Zuc-mediated cleavage is then loaded into a new PIWI protein, defining the 5' end of a new piRNA and recruiting another Zuc to the downstream region. This process, known as the phased piRNA pathway, expands the sequence repertoire of piRNAs beyond the ping-pong sites (9–11). The phased piRNA pathway can be also initiated by specific RNA-binding proteins, for example, Yb in *Drosophila*. A number of protein-coding genes in flies contain Yb-binding sequences in their 3' untranslated regions (UTR) and generate specific genic piRNAs (12–16). Accordingly, it is thought that, for transcripts to be processed into piRNAs, they must meet one of the following requirements: integration of the sequence into a piRNA cluster, recognition and cleavage by a piRNA, or possession of a specific sequence recognized by a specific RNA-binding protein, such as Yb.

BmN4 cells, derived from silkworm ovaries, have both the ping-pong and phased piRNA pathways, making them valuable for piRNA research (17, 18). Like silkworm ovaries, BmN4 cells express two PIWI proteins, Siwi and BmAgo3, which participate in the ping-pong amplification of piRNAs (17). Siwi preferentially associates with piRNAs with uracil at the first position (referred to as 1U bias), whereas BmAgo3-bound piRNAs show a bias toward adenine at the tenth position (referred to as 10A bias) (19, 20). The ping-pong process is thought to occur in a non-membranous liquid-like structure called nuage, which is located around the nucleus (3, 21). Interestingly, we have previously observed that a slicer-deficient Siwi mutant is mislocalized from nuage to P-bodies, another non-membrane structure that contains mRNA decay factors, accompanied by the overproduction of piRNAs derived from genic mRNAs (22). In fact, these gene-derived piRNAs are produced even under the normal condition, albeit at low levels (22). However, their biogenesis mechanism remains unclear, because they do not meet the requirements for the known piRNA biogenesis pathways; they are generally uni-stranded in the sense direction, they lack the ping-pong signature, and their sequences do not appear to contain specific motifs for RNA-binding proteins or to be integrated into piRNA clusters (22).

In this study, we found that sense-stranded piRNAs are generated from the whole transcriptome, albeit inefficiently, through a mechanism distinct from any known biogenesis pathways. The production of these piRNAs is strongly correlated with the abundance of their source transcripts, suggesting the involvement of non-specific fragmentation processes. Interestingly, this mechanism is utilized to produce piRNAs from *Bombyx mori* latent virus, which have been suggested to play a role in suppressing viral replication in silkworm cultured cells. We propose that this primitive piRNA production pathway acts as a sensor for unusually highly expressed transcripts in the germline.

Results

Low levels of sense piRNAs are generated from many genes.

To understand the characteristics of gene-derived sense piRNAs we previously noted (22), we first mapped piRNA-sized small RNAs (≥26 nt) from BmN4 cells to the entire genes annotated in silkworms (23), which include "transposon-like" genes that are homologous to 1,811 previously well-annotated transposons (24) at the amino acid level. We then performed a metagene analysis to determine the distribution of piRNAs across different gene regions: 5' UTRs, coding sequences (CDSs), and 3' UTRs. Many piRNAs were mapped to the antisense strand of genes (Figure 1A, top), which can be attributed to the antisense transcription from transposon-like genes or repetitive sequences in the silkworm genome, such as those in piRNA clusters. We therefore decided to focus on the genes that only have piRNAs mapped to the sense strand. We observed that sense piRNAs are generally produced from all the gene regions, with an apparent peak in the 5' UTR (Figure 1A, middle). However, this peak was eliminated when we excluded multi-mappable piRNAs (Figure 1A, bottom), suggesting that the peak contained piRNAs homologous to repetitive sequences. In other words, uniquely mapped sense piRNAs showed an almost uniform distribution across 5' UTRs, CDSs, and 3' UTRs (Figure 1A, bottom), suggesting that any region of mRNAs can produce gene-derived sense piRNAs. We then performed similar analyses for piRNAs in flies and mice using previously reported PIWI-interacting small RNA datasets (9, 10). Again, after excluding genes with antisense piRNAs and focusing on uniquely mapped piRNAs, we observed largely uniform distributions of gene derived sense piRNAs across 5' UTRs, CDSs, and 3' UTRs (Figure S1A–E), supporting the idea that the whole transcriptome can potentially generate low levels of sense piRNAs. Given that the 5' and 3' UTRs generally contain many different regulatory sequences and show substantial length variations among different species, we decided to use CDS-derived piRNAs as a representative dataset for further analyses to investigate the nature of gene-derived sense piRNAs.

For the in-depth analysis, we re-sorted the silkworm genes that produce piRNAs from their CDSs (including antisense piRNAs and multi-mappable piRNAs) into three categories: [1] 3,220 genes homologous to the 1,811 well-annotated transposons at the amino acid level ("transposon-like genes"), [2] 2,033 genes that are not homologous to the transposons but produce antisense piRNAs from the CDSs ("ping-pong piRNA genes"), and [3] 7,422 genes that are not

homologous to the transposons and produce only sense piRNAs from the CDSs ("sense piRNA genes"). We did not detect CDS-derived piRNAs in the remaining 4,205 genes (Figure S1F, Genes without piRNAs). As a reference, we also added piRNAs derived from the CDSs of well-annotated 1,811 transposons (24) in our analysis ("transposons").

First, we mapped small RNAs from naive BmN4 cells to these three gene categories and transposons, and examined their length distribution. We confirmed that the piRNAs derived from the sense piRNA genes have a peak at 27–28 nt, consistent with the typical piRNA length observed with transposon-like genes, ping-pong piRNA genes, and transposons (Figure 1B). The genes that produce sense piRNAs occupied as much as ~44% of the entire gene set (Figure S1F). However, these gene-derived sense piRNAs represented only ~0.7% of the total piRNA pool (Figure S1G), suggesting that their production efficiency is much lower than typical piRNAs generated from transposons, transposon-like genes, and ping-pong piRNA genes.

Next, we comprehensively compared the levels of CDS-derived piRNAs and their corresponding source mRNAs across each gene category. We found that transposons, transposon-like genes, and ping-pong piRNA genes produce piRNAs at disproportionally high levels compared to the abundance of their source mRNAs (Figure S1H–J). In contrast, the amount of piRNAs derived from the sense piRNA genes showed a strong proportional correlation to the abundance of their source mRNAs (Figure 1C). These results suggest that many genes are capable of producing sense piRNAs, likely via random fragmentation, depending on their mRNA abundance. However, this occurs with a much lower efficiency compared to typical piRNAs (Figure 1D).

Gene-derived piRNAs are produced independently of the ping-pong or phased pathway.

To investigate the characteristics of these gene-derived sense piRNAs, we re-analyzed various small RNA libraries previously reported. In general, mature piRNAs are 2'-*O*-methyl modified at the 3' ends by the methyltransferase Hen1, making them resistant to $NaIO_4$-mediated oxidation (25). Indeed, the levels of piRNAs produced from transposons, transposon-like genes, and ping-pong piRNA genes were unchanged by the $NaIO_4$ treatment (Figure 2A, top). Similarly, gene-derived sense piRNAs were also resistant to $NaIO_4$, indicating that they are also 2'-*O*-methylated (Figure 2A, top).

As the name indicates, the definition of piRNAs is that they interact with PIWI proteins. piRNAs deriving from transposons, transposon-like genes, and ping-pong piRNA genes were significantly enriched by both Siwi- and BmAgo3-immunoprecipitation (IP), compared to microRNAs (miRNAs) that are known to interact with AGO-subfamily proteins (Figure 2A, middle). Interestingly, gene-derived sense piRNAs were even more enriched by Siwi- and BmAgo3-IP, presumably because they are too rare to be quantitatively detected without being concentrated by IP. On the other hand Siwi- and BmAgo3-IPed gene-derived piRNAs showed a strong correlation with the abundance of their source mRNAs, suggesting that they are generally associated with PIWI proteins without specific selection (Figure S2A, B). Conversely, gene-derived sense piRNAs were significantly depleted by the transient knockout of Siwi, like typical

153 piRNAs (Figure 2A, bottom). We concluded that small RNAs produced from sense piRNA genes
154 are bona fide piRNAs, bearing 2'-*O*-methyl modification at the 3' end and interacting with PIWI
155 proteins.

156 Given that these sense piRNAs from genes lack antisense ping-pong pairs, a plausible
157 explanation is that they are produced by the Zuc-mediated phased piRNA biogenesis pathway. To
158 examine the requirement of Zuc in the production of gene-derived sense piRNAs, we investigated
159 the change in their expression upon the knockdown (KD) of Zuc (Figure 2B). Transposon-derived
160 piRNAs were significantly reduced by Zuc KD, suggesting that they are at least in part generated
161 by the phased pathway in addition to the ping-pong pathway. piRNAs derived from transposon-
162 like genes and ping-pong piRNA genes were less affected by the Zuc KD, suggesting that they are
163 mainly generated by the ping-pong pathway. In contrast, gene-derived sense piRNAs were rather
164 upregulated by Zuc KD, suggesting that they are not produced by either the ping-pong pathway or
165 the Zuc-mediated phased pathway. Instead, they appear to be generated through a third mechanism
166 that becomes more dominant in the absence of Zuc.

167

168 **Gene-derived sense piRNAs are also produced in flies and mice.**

169 To investigate whether gene-derived sense piRNAs are specific to silkworms or a more general
170 phenomenon, we performed similar analyses using publicly available datasets and CDSs from
171 *Drosophila* and mice. We found that gene-derived sense piRNAs are also generated with low
172 efficiency in *Drosophila*; they are associated with all three fly PIWI proteins (Aubergine [Aub],
173 Ago3, and Piwi) and their abundance is proportional to their source mRNA levels (Figure 3A–C).
174 As observed in silkworms, transposons and some ping-pong piRNA genes in flies also produced
175 disproportionally abundant piRNAs compared to their source mRNAs (Figure S3A–F; note that the
176 gene model of *Drosophila* is so well-annotated that transposon-like genes are virtually excluded).
177 Moreover, gene-derived sense piRNAs associated with Piwi and Aub showed typical length
178 distributions peaking at ~25–26 nt, although the length peak of those associated with Ago3 was
179 less clear (Figure 3D). Similarly in mice, we found that gene-derived sense piRNAs are associated
180 with MILI or MIWI with typical length peaks at ~27 nt and ~30 nt, respectively (Figure 3E). Their
181 abundance is proportional to their source mRNAs (Figure 3F, G) and lower than piRNAs deriving
182 from ping-pong piRNA genes and transposons (Figure S3G–J). These data support the idea that
183 genes generally produce low levels of piRNAs in flies and mice, like in silkworms.

184 Long non-coding RNAs (lncRNAs) are particularly well annotated in the mouse RefSeq
185 genome database (26). Therefore, we investigated if sense piRNAs are also produced from mouse
186 lncRNAs. We extracted 2,056 lncRNAs that produce only sense piRNAs (sense piRNA lncRNAs)
187 and 1,392 lncRNAs that produce antisense piRNAs (ping-pong piRNA lncRNAs), and then
188 mapped piRNAs to them. We found that, as observed with gene-derived piRNAs, lncRNA-derived
189 sense piRNAs are produced as efficient as gene-derived sense piRNAs, and their abundance is
190 proportional to their sources (Figure S3K–N). We concluded that a low levels of sense piRNAs are
191 produced from the whole transcriptome.

192

193 **BmLV-derived piRNAs are generated by the same pathway as gene-derived piRNAs.**

194 It is known that a positive-sense single-stranded RNA virus called Bombyx mori latent virus
195 (BmLV) persistently infects BmN4 cells and that this virus is suppressed by piRNAs in BmN4 cells
196 (27–29). BmLV-derived piRNAs primarily originate from the extremely abundant sense-stranded
197 sub-genomic RNAs, which are thought to recognize and suppress the antisense strand of the virus
198 that is synthesized during viral replication (28). However, it remains unknown how BmLV-
199 associated piRNAs are produced. To answer this question, we first analyzed the levels of BmLV-
200 derived piRNAs and their source mRNAs (RNA-dependent RNA polymerase (Rdrp), Coat protein
201 (Cp), and P15), as we did for gene-derived sense piRNAs (Figure 4A). We found that both piRNAs
202 and mRNAs from BmLV are extremely abundant but the ratio between them is similar to that of
203 general gene-derived sense piRNAs and their source mRNAs in BmN4 cells (Figure S4A). We
204 have previously reported that overexpression of a Siwi mutant (Siwi-D697A) that disrupts its slicer
205 activity causes selective upregulation of gene-derived piRNAs, presumably because of the
206 suppression of the slicer-dependent ping-pong pathway (22). By reanalyzing this dataset, we found
207 that BmLV-derived piRNAs are also significantly increased by Siwi-D697A (Figure 4B). In
208 contrast, overexpression of Siwi-D697A did not cause any changes in the expression levels of
209 BmLV-derived small interfering RNAs (siRNAs), which are mainly produced from the genomic
210 region of BmLV and contribute to virus silencing together with BmLV-derived piRNAs (28, 29).
211 Moreover, BmLV-derived piRNAs were resistant to $NaIO_4$ treatment, bound to PIWI proteins
212 sensitive to Siwi knockout, and unaffected or slightly upregulated by Zuc knockdown (Figure 4C,
213 S4B, C). These results suggest that BmLV-derived sense piRNAs are produced by the same
214 pathway as gene-derived sense piRNAs; BmLV-derived piRNAs are abundant simply because their
215 source mRNAs are abundant.

216

217 **DISCUSSION**

218 Here we show that genes and virtually any transcripts can produce piRNAs, albeit inefficiently,
219 through a mechanism distinct from the known ping-pong and phased piRNA biogenesis pathways
220 (Figure 4D). This mechanism most likely involves non-specific RNA fragmentation, because the
221 production of these piRNAs appears to be solely dependent on the abundance of their source RNAs
222 (Figure 1C, 3A–C, E, F, S3K, L). These sense piRNAs do not have apparent antisense targets to
223 silence, and thus may be regarded as an intrinsic noise of the piRNA system. Importantly, however,
224 this mechanism (we call the "primitive piRNA biogenesis pathway" hereafter) is utilized to produce
225 piRNAs from BmLV (Figure 4A, C, S4A, B), which function to silence this virus in BmN4 cells
226 (28, 29). BmLV-derived piRNAs are almost exclusively sense-stranded, presumably because the
227 transcripts from BmLV are heavily biased toward the sense strands (29); the synthesis of the
228 antisense strand during virus replication is likely too limited and transient to drive efficient ping-
229 pong amplification of piRNAs, even though it can be targeted by sense piRNAs for silencing.
230 However, if both the sense and antisense transcripts were abundantly available, the piRNAs

231 produced by the primitive biogenesis pathway could in theory initiate the ping-pong amplification
232 and the subsequent phased piRNA biogenesis (Figure 4D).

233 In general, existing transposons are stably silenced by abundant piRNAs produced
234 through the ping-pong and phased piRNA biogenesis pathways (Figure 1B, D, S1J) (3, 10, 11).
235 However, when the genome is invaded by novel transposons, the piRNA system needs to adapt
236 itself to silence these new elements from scratch. We envision that the primitive piRNA biogenesis
237 pathway initiates this adaptation process by generating a small pool of piRNAs de novo. These
238 primitive piRNAs can then act as the seed for the ping-pong and phased piRNA pathways as long
239 as both sense and antisense transcripts are abundantly available, perhaps even before the new
240 transposons are captured by piRNA clusters for established silencing (30–34). In this way, the
241 primitive piRNA biogenesis pathway may act as a primitive sensor for aberrantly abundant
242 transcripts in the germline.

243 A recent study reported that, in *Drosophila melanogaster* testes, piRNAs are produced
244 from CDSs, but not from 5' or 3' UTRs, in a manner dependent on Ago2, an AGO protein that loads
245 small interfering RNAs (siRNAs) and some miRNAs, and Dicer-2, the core biogenesis factor for
246 siRNAs (35). It is thought that targeting by Ago2-loaded siRNAs and miRNAs defines the
247 precursors to be processed into piRNA in fly testes. Indeed, our reanalysis of the small RNA
248 sequencing data (35) confirmed that these CDS-piRNAs are markedly down-regulated by the KD
249 of Ago2 (Fig. S4D) or Dicer-2 (Fig. S4E). In contrast, gene-derived primitive sense piRNAs were
250 only slightly decreased by Ago2 KD and remained unchanged by Dicer-2 KD, like piRNAs derived
251 from ping-pong genes or transposons (Fig. S4D and E), suggesting that CDS-piRNAs and primitive
252 piRNAs are distinct species.

253 How are the primitive piRNAs produced? A clue lies in our finding that overexpression
254 of the slicer-deficient Siwi-D697A mutant promotes the production of sense-stranded primitive
255 piRNAs from genes (22) as well as from BmLV (Figure 4B). Siwi-D697A is mislocalized to P-
256 bodies, forming more solid-like structures than normal liquid-like nuage (22). P-bodies are
257 generally thought to function as a storage site for mRNAs (36, 37), but they also accommodate an
258 important subset of piRNA biogenesis factors in germ cells (22, 38, 39). In a normal situation, Siwi
259 shuttles between nuage and P-bodies by changing its interactors, but Siwi-D697A loses such
260 dynamic molecular exchanges and remains immobilized to P-bodies (22). We envision that Siwi-
261 D697A forces P-body-stored mRNAs to join the piRNA biogenesis pathway. In such a scenario,
262 the non-specific RNA fragmentation, which is hypothesized to produce primitive piRNAs, may
263 occur within P-bodies even under normal conditions. This idea may also shed light on the mystery
264 in the field as to why the piRNA system is compartmentalized to multiple different subcellar
265 platforms, such as nuage, P-bodies, and the outer membrane of mitochondria, which accommodates
266 other important subsets of piRNA biogenesis factors (40–42). Future studies are required to
267 understand the molecular mechanism and biological roles of the random piRNA biogenesis
268 pathway.

269
270

Materials and Methods

Cell culture

BmN4 cells were cultured at 27°C in IPL-41 medium (Applichem) supplemented with 10% fetal bovine serum.

Sequence analysis of small RNAs

Small RNA sequences from BmN4 cells (DRR181866) were reported previously (18). For informatics analysis, the 3′-adaptor sequences were identified and removed by cutadapt 2.6 with Python 3.7.0 with "--minimum-length 20" parameter. Reads that could be mapped by bowtie 1.2.3 (43) to the *Bombyx mori* genome (23) or *Drosophila melanogaster* genome r6.27 from FlyBase (44) or mouse genome GRCm38.p5 from NCBI, allowing up to two mismatches, were used to calculate the mapping rate and to normalize each library. Mapping to the BmLV genome (GenBank *AB624361*.1) was performed under a similar condition but the reads were not used for normalization. In the comparison between the silkworm $NaIO_4$-treated library and its control, small RNA reads were normalized by the spike sequences (Supplementary Table 1) introduced into the library. In the comparison of the Siwi KO library with its control, normalization was performed using the median read of miRNAs (Supplementary Table 1) present in the library, as in the previous analyses (45, 46). Of the 96,858 mouse CDSs, the top-listed isoform in the FASTA file for each gene was treated as a representative, and 22,858 CDSs were used for analysis. Reads that could be mapped by bowtie 1.2.3 (43) to the *Bombyx mori* gene model (23) or *Drosophila melanogaster* CDSs dataset r6.27 from FlyBase (44) or mouse CDSs dataset GRCm38.p5 form NCBI allowing up to two mismatches were used to downstream analyses. SAM files were converted to BAM files by SAMtools (47), then to BED files, and the mapped reads of each gene and transposons were calculated by BEDTools (48). Those reads with a length between 26 nt and 32 nt were defined as piRNAs. Silkworm genes with transposon sequences were defined as in the previous paper (22). Specifically, TBLASTX was performed using 1,811 transposon sequences (24) in the database, and silkworm gene model sequences as the queries, and genes having an E-value of 1e-50 or less were defined as transposons. Silkworm ping-pong genes were defined by those with piRNAs mapping to the antisense strand of genes with more than 1 read in the total small RNA library of the GFP knockout BmN4 cells. For the *Drosophila* and mouse data, we used only the top-listed splice variants from the CDSs FASTA files. In addition, we excluded those genes whose piRNAs mapped to the antisense strands with more than 1 read in the three libraries, Piwi-IP, Aub-IP, and Ago3-IP for *Drosophila* and in the two libraries, MILI-IP and MIWI-IP for mice.

Sequence analysis of mRNAs

mRNA reads that could be mapped by Hisat 2.1.0 (43) to the *Bombyx mori* genome (23) or *Drosophila melanogaster* genome r6.27 from FlyBase (44) or mouse genome GRCm38.p5 form

309 NCBI with -k 3 option were used to calculate the mapping rate and to normalize each library.
310 Mapping to the BmLV genome (GenBank *AB624361*.1) was performed under a similar condition
311 but the reads were not used for normalization. SAM files were converted to BAM files by
312 SAMtools (47), then to BED files, and the mapped reads of each gene and transposons were
313 calculated by BEDTools (48). Results were imported into R and graphed.

314

315 **Analysis of small RNA lengths**

316 Reads mapped to genes, transposons, and BmLV were extracted from FASTA files and calculated
317 the length distributions of the mapped piRNAs. The alphabetFrequency function of the biostrings
318 package in R was used for the calculation.

319

320 **Knockdown of Zucchini, RNA extraction, and small RNA-seq**

321 Preparation of a small RNA library from Zuc KD cells was performed as described previously (46).
322 For double-stranded RNA (dsRNA) preparation for KD, template DNAs were prepared by PCR
323 using primers containing T7 promoter listed in Supplementary Table 1. dsRNAs were transcribed
324 using T7 Scribe Standard RNA IVT Kit (Cell Script) and purified with MEGAclear Transcription
325 Clean-Up Kit (Thermo Fisher/Invitrogen). For dsRNA transfection, 5 μg of dsRNAs were
326 transfected into BmN4 cells (6×10^5 cells per 10 cm dish) with X-tremeGENE HP DNA
327 Transfection Reagent (Sigma). dsRNAs were repeatedly transfected every 3 days for four times.
328 Total RNAs were prepared by TRI Reagent (Molecular Research Center, Inc.). Small RNA libraries
329 were prepared from ~20–45 nt total RNAs using Small RNA Library Preparation kit (Illumina) and
330 analyzed by Illumina HiSeq 2500 platform.

331

332
333
334 **Acknowledgments**
335 We thank all the members of the Tomari laboratory for discussion and critical comments on the
336 manuscript. We thank Dr. Natsuko Izumi of The University of Tokyo for preparation of the Zuc
337 KD library and careful and critical comments on this paper. We thank Dr. Masashi Iwanaga of
338 Utsunomiya University and Dr. Susumu Katsuma of The University of Tokyo for their fruitful
339 discussion.

340
341
342 **References**
343 1. C. Klattenhoff, W. Theurkauf, Biogenesis and germline functions of piRNAs. *Development*
344 **135**, 3–9 (2008).

345 2. M. Ghildiyal, P. D. Zamore, Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.*
346 **10**, 94–108 (2009).

347 3. J. Brennecke, *et al.*, Discrete small RNA-generating loci as master regulators of transposon
348 activity in Drosophila. *Cell* **128**, 1089–1103 (2007).

349 4. A. A. Aravin, R. Sachidanandam, A. Girard, K. Fejes-Toth, G. J. Hannon, Developmentally
350 regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744–747 (2007).

351 5. L. S. Gunawardane, *et al.*, A slicer-mediated mechanism for repeat-associated siRNA 5' end
352 formation in Drosophila. *Science* **315**, 1587–1590 (2007).

353 6. M. Reuter, *et al.*, Miwi catalysis is required for piRNA amplification-independent LINE1
354 transposon silencing. *Nature* **480**, 264–267 (2011).

355 7. S. De Fazio, *et al.*, The endonuclease activity of Mili fuels piRNA amplification that silences
356 LINE1 elements. *Nature* **480**, 259–263 (2011).

357 8. K. M. Nishida, *et al.*, Gene silencing mechanisms mediated by Aubergine piRNA complexes
358 in Drosophila male gonad. *RNA* **13**, 1911–1922 (2007).

359 9. I. Gainetdinov, C. Colpan, A. Arif, K. Cecchini, P. D. Zamore, A Single Mechanism of
360 Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most
361 Animals. *Mol. Cell* **71**, 775-790.e5 (2018).

362 10. F. Mohn, D. Handler, J. Brennecke, piRNA-guided slicing specifies transcripts for Zucchini-
363 dependent, phased piRNA biogenesis. *Science* **348**, 812–817 (2015).

364 11. B. W. Han, W. Wang, C. Li, Z. Weng, P. D. Zamore, piRNA-guided transposon cleavage
365 initiates Zucchini-dependent, phased piRNA production. *Science* **348**, 817–821 (2015).

366 12. H. Ishizu, *et al.*, Somatic Primary piRNA Biogenesis Driven by cis-Acting RNA Elements
367 and trans-Acting Yb. *Cell Rep.* **12**, 429–440 (2015).

368 13. Y. Murota, *et al.*, Yb Integrates piRNA Intermediates and Processing Factors into Perinuclear
369 Bodies to Enhance piRISC Assembly. *Cell Rep.* **8**, 103–113 (2014).

370 14. K. Saito, *et al.*, Roles for the Yb body components Armitage and Yb in primary piRNA
371 biogenesis in *Drosophila*. *Genes Dev.* **24**, 2493–2498 (2010).

372 15. D. Olivieri, M. M. Sykora, R. Sachidanandam, K. Mechtler, J. Brennecke, An in vivo RNAi
373 assay identifies major genetic and cellular requirements for primary piRNA biogenesis in
374 Drosophila. *EMBO J.* **29**, 3301–3317 (2010).

375 16. N. Robine, *et al.*, A broadly conserved pathway generates 3'UTR-directed primary piRNAs.
376 *Curr. Biol.* **19**, 2066–2076 (2009).

377 17. S. Kawaoka, *et al.*, The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-
378 interacting RNA complexes. *RNA* **15**, 1258–1564 (2009).

379 18. N. Izumi, K. Shoji, Y. Suzuki, S. Katsuma, Y. Tomari, Zucchini consensus motifs determine
380 the mechanism of pre-piRNA production. *Nature* **578**, 311–316 (2020).

381 19. S. Kawaoka, N. Izumi, S. Katsuma, Y. Tomari, 3' end formation of PIWI-interacting RNAs
382 in vitro. *Mol. Cell* **43**, 1015–1022 (2011).

383 20. W. Wang, *et al.*, The initial uridine of primary piRNAs does not create the tenth adenine that
384 is the hallmark of secondary piRNAs. *Mol. Cell* **56**, 708–716 (2014).

385 21. A. Webster, *et al.*, Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form
386 a Ping-Pong Complex Assembled by Krimper. *Mol. Cell* **59**, 564–575 (2015).

387 22. P. Y. Chung, K. Shoji, N. Izumi, Y. Tomari, Dynamic subcellular compartmentalization
388 ensures fidelity of piRNA biogenesis in silkworms. *EMBO Rep.* **22**, e51342 (2021).

389 23. M. Kawamoto, *et al.*, High-quality genome assembly of the silkworm, Bombyx mori. *Insect*
390 *Biochem. Mol. Biol.* **107**, 53–62 (2019).

391 24. M. Osanai-Futahashi, Y. Suetsugu, K. Mita, H. Fujiwara, Genome-wide screening and
392 characterization of transposable elements and their distribution analysis in the silkworm,
393 Bombyx mori. *Insect Biochem. Mol. Biol.* **38**, 1046–1057 (2008).

394 25. M. D. Horwich, *et al.*, The Drosophila RNA Methyltransferase, DmHen1, Modifies Germline
395 piRNAs and Single-Stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272 (2007).

396 26. N. A. O'Leary, *et al.*, Reference sequence (RefSeq) database at NCBI: current status,
397 taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

398 27. S. Katsuma, *et al.*, Novel Macula-Like Virus Identified in *Bombyx mori* Cultured Cells. *J.*
399 *Virol.* **79**, 5577–5584 (2005).

400 28. S. Katsuma, K. Shoji, Y. Suzuki, T. Kiuchi, CRISPR/Cas9-mediated mutagenesis of Ago2
401 and Siwi in silkworm cultured cells. *Gene* **768**, 145314 (2021).

402 29. S. Katsuma, *et al.*, Transcriptome profiling reveals infection strategy of an insect maculavirus.
403 *DNA Res.* **25**, 277–286 (2018).

404 30. J. S. Khurana, *et al.*, Adaptation to P element transposon invasion in Drosophila melanogaster.
405 *Cell* **147**, 1551–1563 (2011).

406 31. K. Shoji, Y. Umemura, S. Katsuma, Y. Tomari, The piRNA cluster torimochi is an expanding
407 transposon in cultured silkworm cells. *PLoS Genet.* **19**, e1010632 (2023).

408 32. S. Shpiz, S. Ryazansky, I. Olovnikov, Y. Abramov, A. Kalmykova, Euchromatic Transposon
409 Insertions Trigger Production of Novel Pi- and Endo-siRNAs at the Target Sites in the
410 Drosophila Germline. *PLoS Genet.* **10**, e1004138 (2014).

411 33. I. Olovnikov, *et al.*, De novo piRNA cluster formation in the Drosophila germ line triggered
412 by transgenes containing a transcribed transposon fragment. *Nucleic Acids Res.* **41**, 5757–
413 5768 (2013).

414 34. P. George, *et al.*, Increased production of piRNAs from euchromatic clusters and genes in
415 Anopheles gambiae compared with Drosophila melanogaster. *Epigenetics Chromatin* **8**, 50
416 (2015).

417 35. T. Iki, S. Kawaguchi, T. Kai, miRNA/siRNA-directed pathway to produce noncoding piRNAs
418 from endogenous protein-coding regions ensures *Drosophila* spermatogenesis. *Sci Adv* **9**,
419 eadh0397 (2023).

420 36. C. J. Decker, R. Parker, P-Bodies and Stress Granules: Possible Roles in the Control of
421 Translation and mRNA Degradation. *Cold Spring Harb. Perspect. Biol.* **4**, a012286 (2012).

422 37. Y. Luo, Z. Na, S. A. Slavoff, P-Bodies: Composition, Properties, and Functions. *Biochemistry*
423 **57**, 2424–2431 (2018).

424 38. A. A. Aravin, *et al.*, Cytoplasmic Compartmentalization of the Fetal piRNA Pathway in Mice.
425 *PLoS Genet.* **5**, e1000764 (2009).

426 39. A. K. Lim, L. Tao, T. Kai, piRNAs mediate posttranscriptional retroelement silencing and
427 localization to pi-bodies in the *Drosophila* germline. *J. Cell Biol.* **186**, 333–342 (2009).

428 40. Y. Shiromoto, *et al.*, GPAT2, a mitochondrial outer membrane protein, in piRNA biogenesis
429 in germline stem cells. *RNA* **19**, 803–810 (2013).

430 41. D. Handler, *et al.*, The genetic makeup of the Drosophila piRNA pathway. *Mol. Cell* **50**, 762–
431 777 (2013).

432 42. V. V. Vagin, *et al.*, Minotaur is critical for primary piRNA biogenesis. *RNA* **19**, 1064–1077
433 (2013).

434 43. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment
435 of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

436 44. L. S. Gramates, *et al.*, FlyBase: a guided tour of highlighted features. *Genetics* **220**, iyac035
437 (2022).

438 45. S. Liu, *et al.*, MicroRNAs of Bombyx mori identified by Solexa sequencing. *BMC Genomics*
439 **11**, 148 (2010).

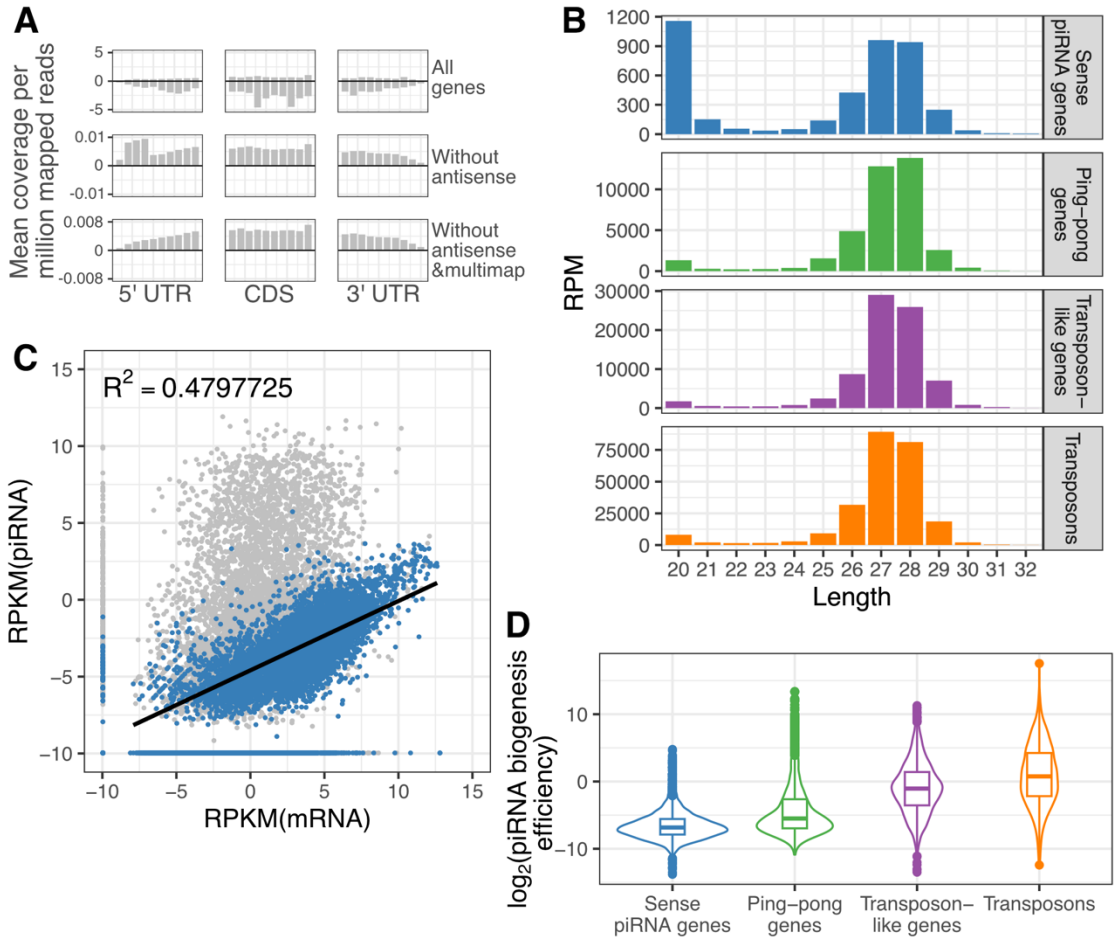
440 46. N. Izumi, *et al.*, Identification and Functional Analysis of the Pre-piRNA 3′ Trimmer in
441 Silkworms. *Cell* **164**, 962‑973 (2016).

442 47. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
443 2079 (2009).

444 48. A. R. Quinlan, I. M. Hall, BEDTools : a flexible suite of utilities for comparing genomic
445 features. *Bioinformatics* **26**, 841–842 (2010).

446
447

448 Figure 1



449

450 **Figure 1. Sense piRNAs are produced from many genes with low efficiency.**

451 (A) The distribution of average piRNA coverage within 5' UTRs, CDSs, and 3' UTRs of silkworm
452 genes is shown. For each gene, the UTRs and CDSs were each divided into 10 segments. The
453 positive Y-axis indicates coverage in the sense direction, while the negative Y-axis indicates
454 coverage in the antisense direction. Out of 16,880 genes analyzed in total, 10,771 genes had no
455 piRNAs mapped to the antisense strand.

456 (B) Length distribution of small RNAs mapped to the sense piRNA genes, ping-pong genes,
457 transposon-like genes, and transposons. Silkworm piRNAs have a peak at 27–28 nt and siRNAs at
458 20 nt.

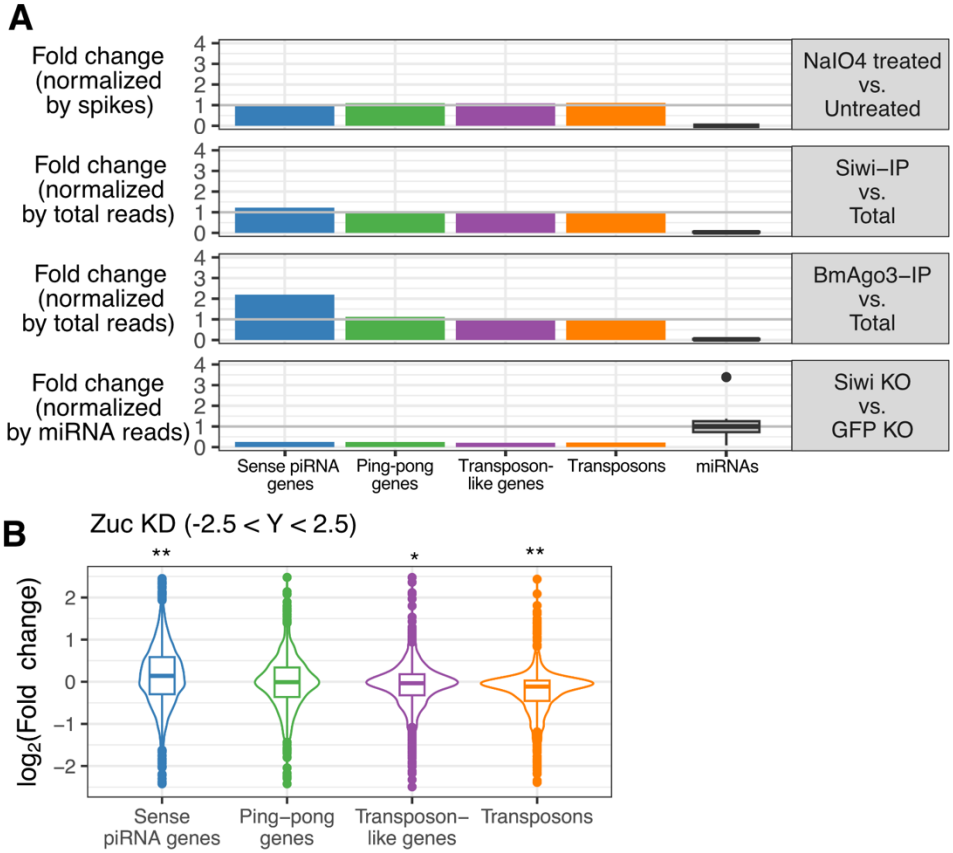
459 (C) Scatter plot illustrating the $\log_2$ scale relationships between mRNAs and piRNAs (26–32 nt)
460 mapped to the CDSs of the genes. A minimum value of $\log_2(0.001)$ is added to both the vertical
461 and horizontal axes. The 7,422 genes defined as sense piRNA genes in this study are shown in blue.
462 See Figure S1H–J for additional information. Other genes and transposons are shown in gray. A

463 regression line was drawn using only those piRNAs and mRNAs whose levels were greater than
464 $\log_2(0)$. The adjusted R^2 value is shown in the upper left of the graph.

465 (D) Violin plots of piRNA production efficiency. RPKM (piRNA) / RPKM (mRNA) was
466 calculated and shown for each category: sense piRNA genes, ping-pong piRNA genes, transposon-
467 like genes, and transposons. Center line, median; box limits, upper and lower quartiles; whiskers,
468 1.5 × interquartile range; points, outliers.

469
470

471 Figure 2



472

473 **Figure 2. Gene-derived piRNAs are bona fide piRNAs.**

474 (A) Changes in the abundance of gene-derived sense piRNAs, other canonical piRNAs, miRNAs
475 by different treatments. From top to bottom: $NaIO_4$ treated vs. untreated (degree of 2'-*O*-methyl
476 modification), Siwi-IP vs. total library (degree of binding to Siwi), BmAgo3-IP vs. total library
477 (degree of binding to BmAgo3), Siwi KO vs. control (dependence on the presence of Siwi). Total
478 values are shown for piRNAs, and boxplots are shown for miRNAs. Center line, median; box limits,
479 upper and lower quartiles; whiskers, 1.5 × interquartile range; points, outliers.

480 (B) Violin plots of the change in piRNA abundance by Zuc KD. RPKM (Zuc KD) / RPKM (control
481 KD) was calculated and shown for each category. Transcripts with an average RPKM greater than

482 $\log_2(-2.5)$ were shown in these plots. Center line, median; box limits, upper and lower quartiles;
483 whiskers, 1.5 × interquartile range; points, outliers. Only those with Y-axis values between -2.5
484 and 2.5 were shown in the inset. * $p = 1.6e-5$, ** $p < 1e-16$.

485

Figure 3

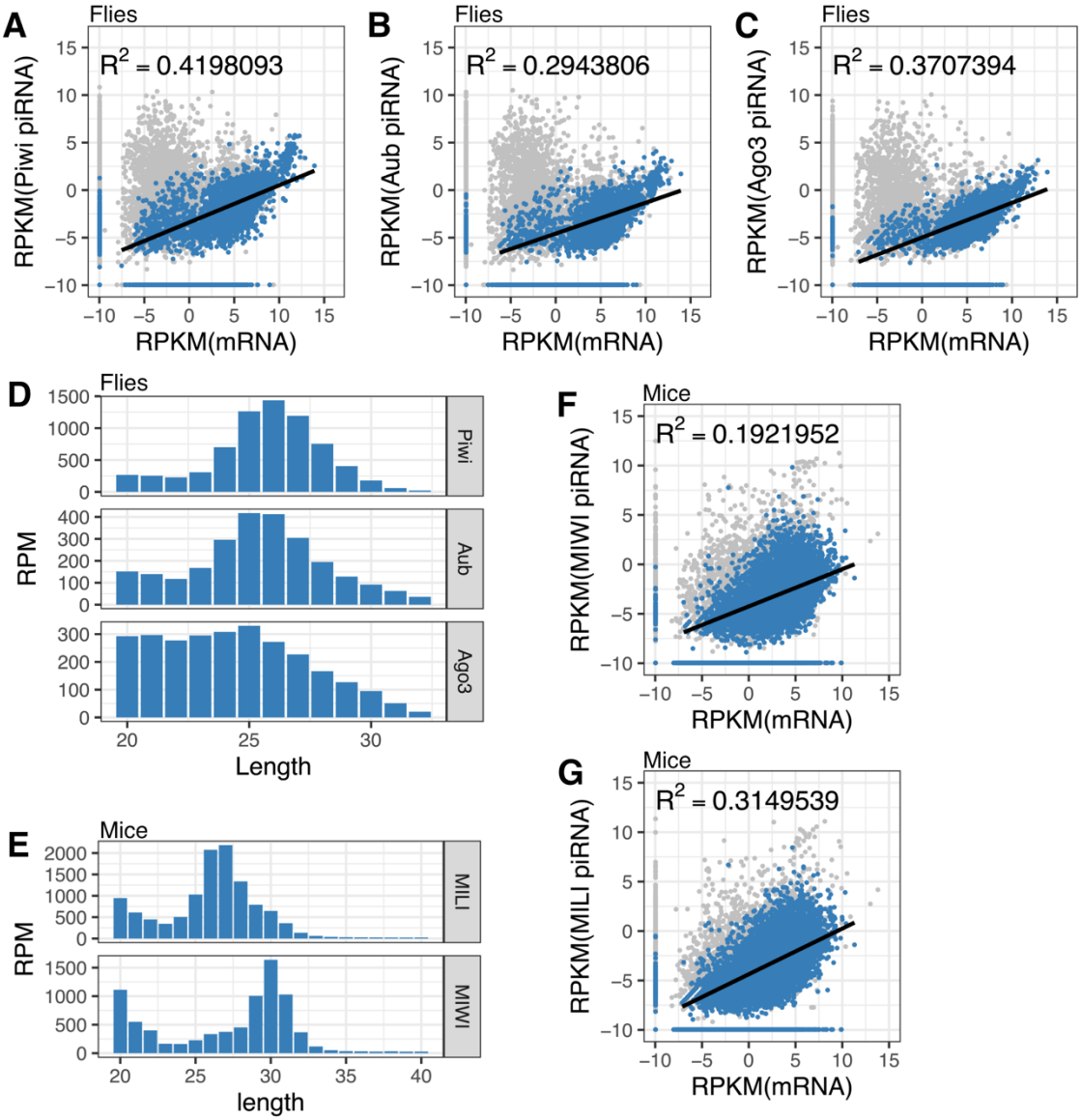


Figure 3. Gene-derived piRNAs in flies and mice.

(A–C) The scatter plots show the relationship between the abundance of ovarian mRNAs and Piwi (A), Aub (B) and Ago3 (C)-bound piRNAs in *Drosophila melanogaster*. 7,106 genes that do not map piRNAs to the antisense strand in all three libraries (Piwi, Aub, Ago3) are plotted. A value of $\log_2(0.001)$ is added to both the vertical and horizontal axes as the minimum value. Regression lines were drawn using only those piRNAs and mRNAs whose levels were greater than $\log_2(0)$. The adjusted R^2 values are shown in the upper left of the graph.

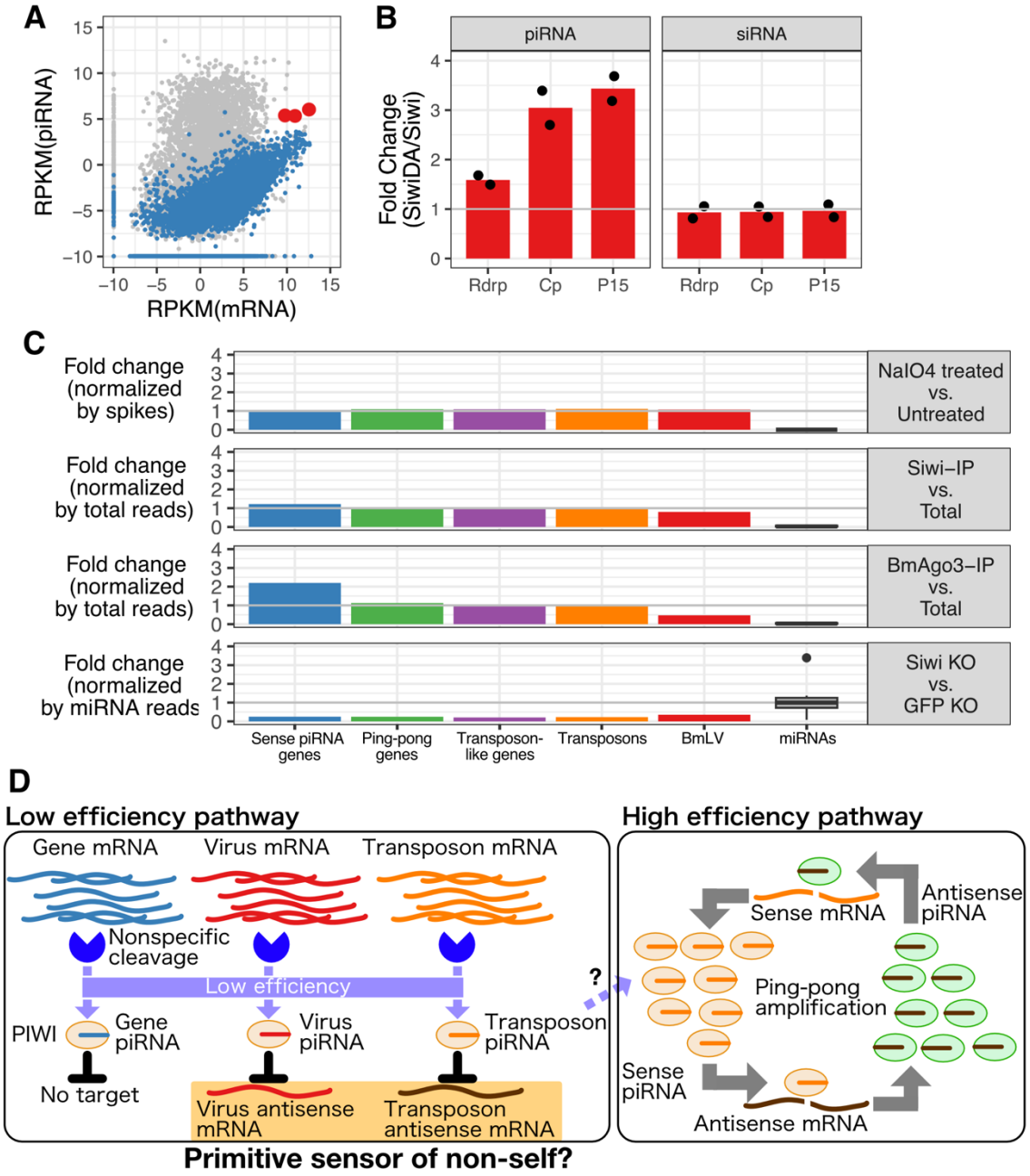
495 (D) The length distribution of gene-derived piRNAs that bind to the three *Drosophila* PIWI proteins
496 (Piwi, Aub, and Ago3). Piwi-IPed piRNAs have a length peak at 26 nt, Aub-IPed piRNAs at 25 nt,
497 and Ago3-IPed piRNAs at 24 nt in these libraries as reported previously (10).

498 (E) The length distribution of gene-derived piRNAs that bind to MILI and MIWI, which are
499 expressed in primary spermatocytes. MILI-IPed piRNAs have a length peak at 27 nt and MIWI-
500 IPed piRNAs at 30 nt as reported previously (9).

501 (F, G) The relationships between the abundance of mRNAs and MIWI- and MILI-bound piRNAs
502 in primary spermatocytes. 20,656 genes for which no piRNAs were mapped to the antisense strand
503 are plotted. A value of $\log_2(0.001)$ is added as the minimum value to both the vertical and horizontal
504 axes. Regression lines were drawn using only those piRNAs and mRNAs whose levels were greater
505 than $\log_2(0)$. The adjusted R^2 values are shown in the upper left of the graph.

506

507 Figure 4



508

509 **Figure 4. BmLV-derived piRNAs are produced by the same pathway as gene-derived piRNAs.**

510 (A) The scatter plot shows the $\log_2$ scale relationships between mRNAs and piRNAs (26–32 nt)
511 mapped to the CDS of genes. A value of $\log_2(0.001)$ is added as the minimum value to both the
512 vertical and horizontal axes. 7,422 sense piRNA genes are shown in blue and three ORFs of BmLV
513 are highlighted in red.

514 (B) Changes in BmLV-derived piRNA levels by Siwi-DA expression. BmLV-derived piRNAs of
515 26–32 nt but not siRNAs of 20 nt are increased by Siwi-DA expression where gene-derived piRNAs
516 are also increased (22). The results for the two replicate libraries are shown, with dots and bars
517 representing individual values and the means, respectively. A straight line is drawn in gray to
518 indicate no change. (C) Changes in the abundance of BmLV piRNAs, other canonical piRNAs, and
519 miRNAs by different treatments. From top to bottom: $NaIO_4$ treated vs. untreated (degree of 2'-*O*-
520 methyl modification), Siwi-IP vs. total library (degree of binding to Siwi), Ago3-IP vs. total library
521 (degree of binding to BmAgo3), Siwi KO vs. control (dependence on the presence of Siwi). Total
522 values are shown for piRNAs, and boxplots are shown for miRNAs. Center line, median; box limits,
523 upper and lower quartiles; whiskers, 1.5 × interquartile range; points, outliers.

524 (D) A model diagram illustrating the results of this study. mRNAs are fragmented by putative non-
525 specific enzymes and incorporated into PIWI protein as piRNAs with low efficiency in proportion
526 to the amount of mRNA amount. Gene-derived piRNAs do not have their complementary targets,
527 but virus-derived piRNAs recognize and repress viral antisense mRNAs during virus replication.
528 This mechanism may serve as a primitive sensor for aberrantly abundant transcripts, seeding
529 piRNA sequence repertoire for the ping-pong amplification, e.g., upon the invasion of new
530 transposons.