Sequence dependent but not sequence specific piRNA adhesion traps mRNAs to the germ plasm

Anastassios Vourekas[1], Panagiotis Alexiou[1], Nicholas Vrettos, Manolis Maragkakis and Zissimos Mourelatos[*]

SUPPLEMENTARY INFORMATION

INTRODUCTION

Primary piRNA processing involves phased 5′ to 3′ cleavage of piRNA precursors by the Zucchini (Zuc) endonuclease[18,44–52]. In mice the piRNA Cleavage and Loading complex (piCL), whose core components are mouse Zuc and the MOV10L1 helicase (*Drosophila* homolog is Armitage)[51,53], binds, unwinds and translocates piRNA precursors in a 5′ to 3′ direction to be cleaved by Zuc[35]. The released 5′-fragments that contain 5′-U are preferentially loaded to empty Piwi proteins and further trimmed by an unknown nuclease(s) to mature piRNAs[17,35] that are 2′-O-methylated at their 3′-termini by the Hen1 methyltransferase[54–56]. In *Drosophila*, primary piRNAs are loaded in Piwi and Aub proteins. Active transposable elements (TEs) are targeted and sliced by piRNPs. The cleaved, 3′- fragments of sliced TE RNAs are processed into secondary piRNAs by a mechanism called ping-pong[57,58], and they are loaded mainly to Ago3. Secondary piRNAs can feed back and increase the piRNA output of primary clusters that contain sequences complementary to active TEs[59]

Aub contains arginines that are symmetrically dimethylated (sDMA) by the methyltransferase dPRMT5 (*csul*) and its cofactor *valois (vls);* sDMA-modified Aub binds to extended tudor domains of the Tudor protein[13,14,60,61]. Oocytes from *Drosophila* mutants that lack Tudor (*tud*) or that disrupt the Aub-Tudor interaction (*csul,* or *vls*) fail to localize Aub to the posterior resulting in defective germ plasm; upon fertilization they give rise to agametic progeny because primordial germ cells (PGCs) are never specified[13,15,20,38,62,63]. Truncated forms of Tudor that maintain interaction with Aub rescue PGC formation[13–15,20]. piRNA abundance is not reduced in flies lacking Tudor and most TEs are not upregulated[13], suggesting that the absence of PGC specification in Tudor mutant flies is not caused by piRNA pathway dysregulation or defective TE silencing.

Transport of *oskar* (*osk*) mRNA to the posterior of the oocyte, where it is locally translated to produce Oskar protein, initiates germ plasm assembly[2]. Aub, whose germ plasm enrichment occurs by midoogenesis[64], has been implicated in preventing diffusion of *osk* away from the germ plasm and also for *nos* localization to the posterior[32,65]. Extensive studies of the cytoskeleton-dependent RNP transport during late oogenesis provide the basis for understanding the selective enrichment of germ cell mRNAs to the posterior[8,9,27,28,66,67]. While anterior localization of *bcd* mRNPs is a dynein-dependent, purposeful translocation on microtubules, the

posterior enrichment of many mRNAs *gcl, pgc, osk, nos* first appears, or is enhanced, during the bulk mixing of the ooplasm (ooplasmic streaming) that promotes the diffusion of these mRNAs and their entrapment at the posterior.

Upon PGC formation Tudor is apparently cleaved/degraded rapidly[68], while Aub is incorporated in the cytoplasm of PGCs and persists as they migrate to form the future gonads[69] (**Extended Data Fig. 1a**).

RESULTS

Aubergine CLIP

CLIP was performed using stringent buffers, without adding exogenous ribonuclease to the crosslinked lysates[36]. As performed previously for mammalian Piwi proteins[17,36], we extracted RNAs from the main radioactive signal (enriched in piRNAs) and from a higher position on the nitrocellulose membrane (enriched in larger RNAs -lgClips-) as shown in **Fig. 1a, b** and **Extended Data Fig. 1d**. The segregation of piRNA and longer RNA (lgClip) populations is also supported by the 5′-nucleotide preference and the genomic origin of CLIP tags belonging in these two groups (**Fig. 1b, c, Extended Data Fig. 1e-g**). piRNAs show an overwhelming bias for 5′ U, and this bias is switched and maintained to 5′ A for RNAs longer than 35 nucleotides. Aub loaded piRNAs are predominantly transposon derived with anti-sense orientation (>80%, in agreement with previous observations[31,57], while lgClips are mapped mainly within mRNAs, transposons and repeat elements (**Fig. 1c** and **Extended Data Fig. 1g**). Based on the above we designated the piRNA population at 23-29 nt and the lgClip population at 36 nt and up (30-35 nt tags were not considered to ensure 'purity' of lgClip population).

We aligned CLIP and IP libraries to the consensus (canonical) sequences of *D. melanogaster* transposons, acquired from FlyBase (http://flybase.org/) (**Supplementary Table 1**), and verified the extensive similarity of Aub CLIP piRNA with Aub IP piRNA libraries by pairwise comparisons of piRNA abundances within retrotransposons (germline, intermediate and somatic dominant), which constitute the bulk of piRNAs (**Extended Data Fig. 2a-c**). As shown before, Aub piRNA load is biased towards germline dominant elements. piRNAs and lgClips mapping within elements with strong germline expression show the

highest abundances demonstrating that they are the primary targets for Aub silencing, followed by intermediate and somatic elements. Complementary lgClips for lowly expressed piRNAs are scarce, suggesting that efficient targeting requires relatively large number of complementary piRNAs. We also verified that piRNAs found in Low and High CLIP libraries are essentially identical (**Extended Data Fig. 2d-f**). Finally, on average ~42% of unique piRNAs sequences from every CLIP library can be matched (5′ end match) to piRNAs from the respective standard IP library (**Extended Data Table 1**). The above analyses testify to the high specificity of our CLIP data.

Analysis of Aub RNPs supports association of germ plasm mRNAs by Aub

We performed isopycnic density gradient ultracentrifugation of embryo lysates followed by fractionation, thus acquiring RNPs of different buoyant densities. As seen in **Extended Data Figure 5a, b**, complexes rich in RNA are found in the denser ("heavier") fractions of the Nycodenz gradient (fractions 1-3), and complexes richer in proteins are found in the opposite end of the gradient (fractions 7-12). Aubergine is detected in all fractions, forming two separate peaks in fractions 2 and 8 (**Extended Data Figure 5b**). This distribution resembles that of poly(A)-binding protein (PABP), which is also one of the most abundant co-immunoprecipitating proteins in standard Aub IPs using early embryo lysates (our own unpublished data).

Using qRT-PCR on total RNA extracted from gradient fractions, we also detected the presence of several germ plasm mRNAs that were highly bound by Aub in CLIPs from *yw* embryos (*cycB*, *gcl*, *osk*, *Hsp83*, and *dhd*). Interestingly, the first three formed concentration peaks in fraction 4, *cycB* in fraction 2, and *dhd* in fraction 6 (**Extended Data Figure 5c**). We performed Aub RNA immunoprecipitations (RIP), using dilutions of fractions 2-7 that contained equal amounts of Aub, as detected by Western Blot (**Extended Data Figure 5d**). After extensive washes of the IP beads, we extracted RNA and used it for 5′ end ^{32}P labeling and analysis of small RNAs on Urea/PAGE (**Extended Data Figure 5e**), and preparation of cDNA for qRT-PCR detection of germ plasm mRNAs (**Extended Data Figure 5f**). We observed a clear enrichment of piRNAs in fractions 6 and 7, which are protein rich fractions. Aub RNPs immunoprecipitated from fraction 5 are the most enriched (relative to total RNA from this fraction) in all the germ plasm mRNAs that we tested, even though the abundance

of these mRNAs in total RNA preparations reaches a maximum at gradient fractions other than 5 (**Extended Data Figure 5c**). This differential enrichment attests to the specificity of the mRNA binding by Aub identified by the Aub RIP experiment. We note the partial overlap between the distributions of Aub-piRNA and Aub-mRNA complexes, but the two populations are largely separated. This suggests that the bulk of piRNAs are found in RNPs with low mRNA content.

Embedded transposon sequences within mRNAs is not the reason for Aub binding to mRNAs

We examined whether Aub loaded piRNAs with extensive complementarity could explain the observed binding of mRNAs, in a manner similar to what was shown for retrotransposons (**Extended Data Fig. 3a-c**). On average, 0.1-0.3% of the mRNA derived lgClips overlap with a complementary piRNA (**Extended Data Fig. 6a**). Moreover, we plotted sense lgClips versus antisense piRNAs mapping within mRNAs, and there is no correlation between the two populations (**Extended Data Fig. 6b**), contrary to what was shown for the retrotransposon targets (**Extended Data Fig. 3b, c**).

Another possibility that we examined was the potential targeting of transposons embedded within the mRNA sequences, by repeat derived piRNAs. The origin of repeat derived sequences cannot be traced to a single location on the genome or the transcriptome[70], therefore lgClips derived from transposon sequences embedded within mRNAs are indistinguishable from authentic transposon transcripts that are the primary targets of the Piwi-piRNA mechanism. In our pipeline, for every read that can map multiple times on the genome, one random mapping location is recorded, to avoid artificial inflation of genetic elements that contain repetitive sequences. It is conceivable that if embedded repeats within mRNAs are targeted by piRNAs, adjacent, non-repeat mRNA sequences have increased likelihood of getting crosslinked with piRNPs. With this in mind, we plotted mRNA CLIP tag abundance with the number of annotated embedded repeats per mRNA (retrieved from Repeat Masker). Notably, most mRNAs bound by Aub do not contain embedded transposons, and for the ones that do, there is no positive correlation between the number of CLIP tags and the number of embedded repeats within the mRNA (**Extended Data Figure 6c**). These findings suggest that mRNAs bound by Aub piRNPs are not selected for the presence of

transposon-derived sequences within their length. We note that this analysis is limited by the transposon annotation sensitivity, which might not detect small sequences or sequences with small degree of similarity with the consensus transposon sequence.

We also analyzed CLIP tag density within areas flanking embedded repeats. We reasoned that if embedded repeat sequences can be targeted by complementary piRNAs, flanking areas should be enriched in lgClips. We noticed no such enrichment (**Extended Data Figure 6d**). Finally, we compared the transcriptome of *aub* and *tud* embryos with *yw* embryos (0-2h) to investigate whether the presence of embedded repeats in mRNAs could be correlated with abundance changes in mutant embryos. No such correlation is observed (**Extended Data Figure 6e**). Collectively, the above suggest that the binding of mRNAs by Aub does not depend on mRNA sequences with extensive complementarity towards Aub loaded piRNAs.

DISCUSSION

The question whether piRNAs can target heterotypic, endogenous RNAs via partial complementarity in a regulatory context has beset the field since the discovery of these small RNAs. Their immense sequence repertoire (all 16,384 possible 7mers are represented in *Drosophila* piRNA sequences) suggests that piRNAs can bind to virtually any cellular RNA with partial complementarity. A model that proposes that piRNAs can target mRNAs should be able to address the issue of specificity (how target sequences are recognized), and also clarify whether evolutionary conservation plays a role in the shaping of piRNAs and target RNAs. In this work we capture chimeric reads that contain the *in vivo* complementary sequences and do not rely only on *ad hoc* prediction for piRNA complementarity. Thus we were able to provide information on what type of nucleotide sequences are “targeted” by piRNAs, to explore whether sequence conservation plays a role in the mechanism and to address whether piRNAs utilize a partial, microRNA-like mechanism for binding apart from extensive complementarity that we know they utilize to recognize retrotransposons. Furthermore, we uncover mRNA length as a novel feature of Aub mRNA targeting. We were able to form a cohesive model that provides a framework in which the vast potential

sequence complementarity between the piRNAs and their targets, and the role of mRNA length, can be understood; and to propose testable hypothesis with regard to the formation of the germ granules and Piwi RNPs in other animals.

Our analysis identified retrotransposon fragments generated during secondary piRNA biogenesis, captured by CLIP while still within Aub RNPs; i*n vivo,* after slicing they are released from Aub and are subsequently loaded into Ago3 to be further processed by 3´-5´ exonucleolytic trimming yielding mature Ago3 piRNAs. This suggests the need for a mechanism that performs a "handover" of these fragments to Ago3. Vasa, which is required for secondary piRNA biogenesis, could play this role. Fittingly, retrotransposon fragments of piRNA mediated slicing bearing the hallmarks of secondary processing were identified in Vasa RNP complexes[71]. Overall, our findings concerning retrotransposon targeting, demonstrate that our CLIP approach can capture the piRNA-RNA target duplex formed within Aubergine, and supports recent discoveries on the phased processing of primary piRNA transcripts [18,45].

Although our data demonstrate piRNA mediated mRNA binding by Aub, the degree to which piRNA independent, direct contact of Aub with mRNAs contributes to the overall mRNA binding captured by CLIP cannot be currently estimated. In general we believe that Piwi proteins, just like Agos[72], have the ability to contact mRNA directly without a mediating small RNA *in vivo*, and there is direct evidence for that from Piwi CLIP experiments in genetic backgrounds where no piRNAs are made or loaded on the Piwi protein: for example mouse Mili CLIP in MOV10L1 conditional KO mice shows that Mili devoid of piRNAs can still bind mRNAs and piRNA precursors[35]. Furthermore, as shown in **Extended Data Figure 5**, biochemical fractionation of RNPs from early embryo reveals relative depletion of piRNAs from Aub in fractions containing large amount of mRNAs, suggesting that some of the Aub mRNA binding may be direct and not involve piRNAs. Furthermore, we observe varying amounts of chimeric reads for certain mRNAs that are efficiently bound (high number of non-chimeric reads). The observed variability in the numbers of chimeric reads suggests that in some cases mRNAs may be bound directly, independently of piRNAs. We are now investigating this observation further.

Our findings are in perfect agreement with studies that had identified Aub as part of the RNP complexes that contain germ plasm RNAs *nos, pgc* and *gcl* along with germ cell marker Vasa in the pole plasm and pole cells of the early embryo,[10,11,28,64,65]. Our model predicts that different mRNA species can be found in the same Tud-Aub RNP complexes, as the piRNA mediated adhesion is non-discriminatory; this is in agreement with the observation that at least to a certain extent multiple copies of homotypic and heterotypic mRNAs are assembled and packaged in the same RNP particles[12,28].

Among Aub bound mRNAs we find *osk,* which is transported to the oocyte by an active, microtubule dependent mechanism by midoogenesis. In the absence of Aub, *osk* mRNA diffuses away from the posterior cortex of oocytes[32]; and *osk* mRNA further accumulates at the posterior of late stage oocytes, subsequent to nurse cell dumping and ooplasm streaming [9,73,74], suggesting that Aub may be responsible for this late-phase *osk* mRNA anchoring, hence the presence of *osk* CLIP tags in Aub libraries. The above provide an appropriate explanation for the observation that *osk* mRNA localization appears normal in stage 7-8 *aub* mutant oocytes, but is diffuse in later stages, and almost always absent from the posterior of *aub* mutant early embryos[32]. Furthermore, anterior localized *osk* mRNA bearing a *bcd* 3′ UTR is properly translated and Oskar protein initiates pole plasm formation ectopically[32]. However, *nos* mRNA is not recruited to the ectopic pole plasm in *aub* mutant embryos[32]. Additionally, reduced *osk* mRNA in the germ plasm is also observed in mutants of *capsuleen* (which catalyzes the symmetric dimethylation of Aubergine arginine residues that mediate interaction with Tudor) and *valois* (cofactor of *capsuleen*)[62,63]. Finally, in *capsuleen* and *valois* mutants, the bicaudal phenotype caused by *osk-bcd* 3′UTR is suppressed, suggesting that the posterior determinant *nos* is not recruited efficiently at the ectopic site[62,63]. These observations separate the temporal sequence of *osk* and *aub* functions during pole plasm formation and lend support to our model that Aub piRNPs are required for mRNA anchoring in the germ plasm.

Recently, Aub binding to mRNAs in the soma was found to be associated with destabilization of the bound mRNAs in later stage embryos (after 2h), coincidental with the onset of zygotic transcription and the widespread soma clearance of maternal transcripts[21]. Comparisons of mRNA abundance using microarrays between wild-type and *aub* mutant

embryos showed similar levels at 0-2h embryos but persistence of mRNAs in *aub* mutant embryos at later stages[21]. The molecular mechanism of Aub function in such destabilization is not known, but very likely involves Smaug[21,22]. Our RNA-seq experiments show no change of mRNA abundances between wild-type and *aub* mutant 0-2h embryos (**Extended Data Figure 9d, e**) and for all our CLIP and RNA-Seq experiments we have used 0-2h wild-type and *tud* mutant embryos. We note that *aub* mutant embryos show 100% lethality (**Extended Data Figure 9f, g**) and thus an element of developmental delay in the absence of Aub may be at play. However, in the absence of Tudor (or Csul), embryos do hatch and develop into agametic adults (**Extended Data Figure 9f, g, h.)**. The two functions of Aub (soma versus germ plasm) are not mutually exclusive but must be distinct, as also noted by Barckmann et al[21], because if Aub's role in the germ plasm is the same as in the embryo soma (to degrade mRNAs), then *Drosophila* embryos would never form a germline. In our study we uncover the role of Aub in the formation of germ plasm mRNPs, which is dependent on Aub interacting with Tudor. Furthermore, by performing CLIP in ovaries, we find that Aub binds preferentially to germ plasm/ posterior mRNAs. This enrichment supports the mRNA trapping mechanism that we propose for germ plasm mRNP formation because there is no degradation of maternal mRNAs prior to egg activation and fertilization (Smaug, the critical factor in maternal mRNA degradation is not expressed in the ovaries[75]).

Our analysis indicates that the size differential of germ plasm mRNAs[26] is a previously underappreciated element that facilitates mRNA segregation and specific localization and provides a pertinent explanation to the presence of Aub-bound, redundant localization elements in the 3′ UTR of mRNAs such as *nos* [65]. mRNA size can be a substrate of evolutionary pressure. It remains to be seen if this mechanistic aspect of piRNP formation is encountered in other phyla, for example in the formation of the mammalian chromatoid body.

Cis-acting localization elements are required for regulatory events that occur before and after the entrapment during ooplasmic streaming, such as mRNA stability, regulation of translation and further subcellular partitioning during the formation of primordial germ cells. For example sequence elements that control translation but not localization were found within the 3′UTR of *cycB* mRNA [76], and the 3′ UTR of *nanos* mRNA[30]. Furthermore, it is known that the translation of different germ plasm mRNAs is activated at varying time points

during germ cell formation and migration, and this is controlled entirely by each 3′ UTR[10]. Consequently, there must be different regulatory elements on these mRNAs that control their expression programs after their entrapment in the germ plasm.

We believe that piRNA adhesion for germ plasm entrapment is not mutually exclusive with the existence of such cis-acting sequences embedded within the mRNA (primarily 3′ UTR) for the regulation of translation and/or localization events, but superimposed with this mechanism. The fact that only a small percentage of each mRNA is actually localized in the pole plasm[12], is something that we believe lends support to our model: since posterior localization of these mRNAs is a process that relies mainly on random low complementarity events with posterior localized piRNAs, it is understandable that it is inefficient. It is interesting to note that germ plasm mRNAs are also in general more abundant than mRNAs with other localizations[26]. This may be an auxiliary strategy that ensures localization of adequate amounts to the posterior. Furthermore, taking *bcd* mRNA as an example of anterior localized mRNAs, long mRNAs that need to be localized elsewhere than the pole plasm, may be actively transported on microtubules by dynein dependent mechanism towards the opposite direction at the time of ooplasmic streaming. We envision that active transport away from the posterior could contribute to mRNA partitioning within the oocyte, thus reinforcing the diffusion and entrapment of germ plasm mRNAs to the posterior. Finally, cis-acting elements for posterior localization have been described for *osk* mRNA, which appears at the posterior of the developing oocyte at mid-oogenesis. It is conceivable that other mRNAs have such elements. For example residual *nanos* mRNA is observed at the posterior of *tud*-null embryos. Nevertheless, in wt oocytes, both *osk* and *nanos* mRNAs are reaching their maximum level of posterior presence during ooplasmic streaming stage in late oogenesis, suggesting that cis-elements that direct active transport co-exist with piRNA adhesion mechanism that we propose.

It is interesting to note that efforts to identify consensus primary sequence or secondary structural elements that mediate the localization of all posterior mRNAs have so far met with limited success. We think that our model will be important for the design of future studies of cis-sequence elements. The traditional approach for such studies has been to delete portions of the 3′ UTR; our data suggest that a more appropriate manipulation will be to substitute the

investigated sequence with “neutral” sequences, therefore maintaining the native mRNA length. In this way, the true contribution of potential cis-elements to mRNA localization events will be revealed.

The utilization of piRNAs to trap mRNAs may also have a biochemical basis, in addition to a biological basis in ensuring that the newly formed germline is endowed with piRNA immunity against transposons. In principle, any “non-specific” RNA binding protein (RBP) might have been selected to trap germ plasm mRNAs, irrespective of their sequence. However, it is becoming apparent that there is hidden specificity even in “non-specific” RBPs[77]. A defined protein interface is used by RBPs to interact with RNA and the nucleotide sequence influences binding, even if it involves backbone interactions, since the underlying sequence influences backbone geometry[78]. In contrast, the extreme sequence diversity of piRNAs and partial complementarity with mRNAs may achieve true “random” binding to trap mRNAs. Future biochemical investigations are needed to test this concept.

SUPPLEMENTARY TABLE LEGENDS

Supplementary Table 1. Normalized (reads per million, rpm) abundances of piRNA (23-29 nt) and lgClip (>35 nt) subpopulations from the indicated libraries mapping on various transposon consensus sequences (acquired from Repeat Masker).

Supplementary Table 2. Statistically significant depletion and enrichment in mRNA localization categories[19] in Aub CLIP libraries (non-chimeric lgClips) from *yw* an*d tud* embryos and *yw* ovaries (average normalized values from three replicate libraries for each genotype were used). The specificity of Aub for binding posterior localized mRNAs is significantly reduced compared to *yw*, but not completely lost in *tud* embryos.

Supplementary Table 3. Statistically significant depletion and enrichment in mRNA localization categories[19] for pairwise comparison between *yw* and *tud* embryo (0-2 h) Aub CLIP libraries (non-chimeric lgClips) (two-sided t-test).

Supplementary Table 4. Ranked list of highly bound by Aub (non-chimeric lgClips) posterior localized mRNAs (twelve posterior localization categories marked with an asterisk in **Supplementary Table 3**). "Localized" (enriched) and "protected" (appearing localized in the pole cells only after the degradation of the maternally deposited mRNAs from the somatic part of the early embryo) mRNAs in the posterior and in the germ cells according to Rangan et al[10] are also noted.

Supplementary Table 5. Number of unmapped and piRNA:mRNA chimeric reads identified in each CLIP library. The number of chimeric mRNA fragments overlapping with non-chimeric reads for every library is also shown.

Supplementary Table 6. Enriched and depleted mRNA localization categories for mRNA fragments extracted from chimeric reads from each set of CLIP libraries (*yw* and *tud* embryos, *yw* ovaries), and also for the comparison between *yw* and *tud* embryo libraries.

SUPPLEMENTARY REFERENCES

44. Pane, A., Wehr, K. & Schüpbach, T. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev. Cell 12, 851–62 (2007).

45. Han, B. W., Wang, W., Li, C., Weng, Z. & Zamore, P. D. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science 348, 817–21 (2015).

46. Nishimasu, H. et al. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature 491, 284–7 (2012).

47. Voigt, F. et al. Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc. RNA 18, 2128–34 (2012).

48. Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L. & Hannon, G. J. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. Nature 491, 279–83 (2012).

49. Huang, H. et al. piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. Dev. Cell 20, 376–87 (2011).

50. Watanabe, T. et al. MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. Dev. Cell 20, 364–75 (2011).

51. Haase, A. D. et al. Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. Genes Dev 24, 2499–2504 (2010).

52. Homolka, D. et al. PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. Cell Rep. 12, 418–28 (2015).

53. Saito, K. et al. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev. 24, 2493–8 (2010).

54. Saito, K. et al. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev. 21, 1603–8 (2007).

55. Kirino, Y. & Mourelatos, Z. The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. RNA 13, 1397–401 (2007).

56. Horwich, M. D. et al. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr. Biol. 17, 1265–72 (2007).

57. Gunawardane, L. S. et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science (80-.). 315, 1587–1590 (2007).

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

59. Li, C. et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell 137, 509–521 (2009).

60. Vourekas, A., Kirino, Y. & Mourelatos, Z. Elective affinities: a Tudor-Aubergine tale of germline partnership. Genes Dev. 24, 1963–6 (2010).

61. Nishida, K. M. et al. Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. EMBO J 28, 3820–3831 (2009).

62. Anne, J., Ollo, R., Ephrussi, A. & Mechler, B. M. Arginine methyltransferase Capsuleen is essential for methylation of spliceosomal Sm proteins and germ cell formation in Drosophila. Development 134, 137–46 (2007).

63. Anne, J. & Mechler, B. M. Valois, a component of the nuage and pole plasm, is involved in assembly of these structures, and binds to Tudor and the methyltransferase Capsuléen. Development 132, 2167–77 (2005).

64. Harris, A. N. & Macdonald, P. M. Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. Development 128, 2823–2832 (2001).

65. Becalska, A. N. et al. Aubergine is a component of a nanos mRNA localization complex. Dev. Biol. 349, 46–52 (2011).

66. Weil, T. T., Forrest, K. M. & Gavis, E. R. Localization of bicoid mRNA in Late Oocytes Is Maintained by Continual Active Transport. Dev. Cell 11, 251–262 (2006).

67. Parton, R. M. et al. A PAR-1-dependent orientation gradient of dynamic microtubules directs posterior cargo transport in the Drosophila oocyte. J. Cell Biol. 194, 121–135 (2011).

68. Bardsley, A., McDonald, K. & Boswell, R. E. Distribution of tudor protein in the Drosophila embryo suggests separation of functions based on site of localization. Development 119, 207–219 (1993).

69. Mani, S. R., Megosh, H. & Lin, H. PIWI proteins are essential for early drosophila embryogenesis. Dev. Biol. 385, 340–349 (2014).

70. Marinov, G. K. et al. Pitfalls of Mapping High-Throughput Sequencing Data to Repetitive Sequences: Piwi's Genomic Targets Still Not Identified. Dev. Cell 32, 765–771 (2015).

71. Xiol, J. et al. RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell 157, 1698–711 (2014).

72. Leung, A. K. L. et al. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. Nat. Struct. Mol. Biol. 18, 237–44 (2011).

73. Sinsimer, K. S., Jain, R. a, Chatterjee, S. & Gavis, E. R. A late phase of germ

plasm accumulation during Drosophila oogenesis requires lost and rumpelstiltskin. Development 138, 3431–40 (2011).

74. Glotzer, J. B., Saffrich, R., Glotzer, M. & Ephrussi, A. Cytoplasmic flows localize injected oskar RNA in Drosophila oocytes. Curr. Biol. 7, 326–337 (1997).

75. Dahanukar, A., Walker, J. a. & Wharton, R. P. Smaug, a novel RNA-binding protein that operates a translational switch in Drosophila. Mol. Cell 4, 209–218 (1999).

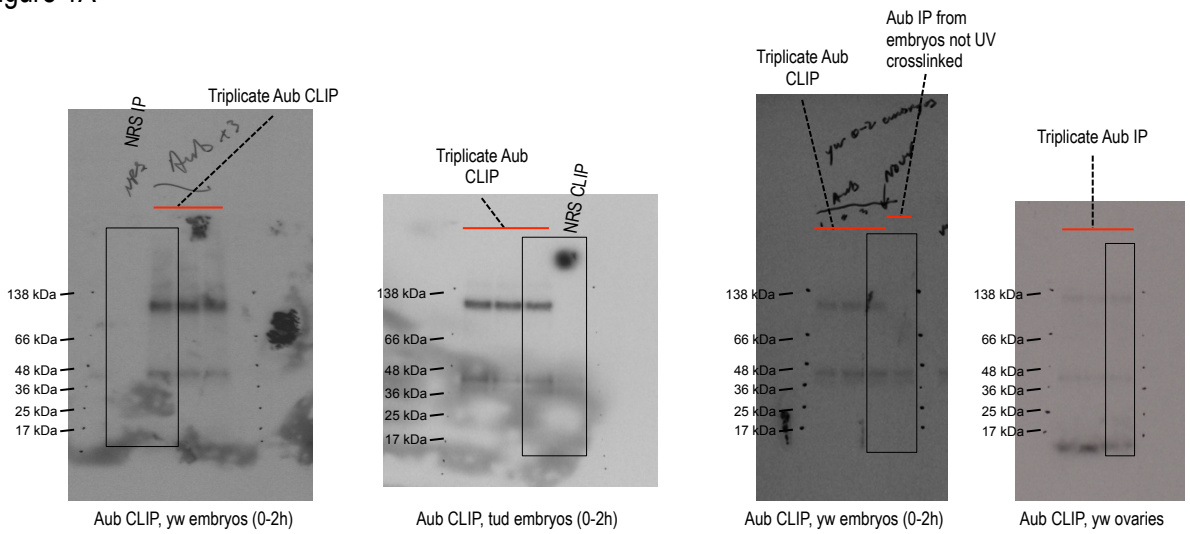
76. Dalby, B. & Glover, D. M. Discrete sequence elements control posterior pole accumulation and translational repression of maternal cyclin B RNA in Drosophila. EMBO J. 12, 1219–27 (1993).

77. Guenther U.P., *et al.* Hidden specificity in an apparently nonspecific RNA-binding protein. *Nature* **502**, 385-58 (2013).

78. Leontis N.B., Lescoute A. & Westhof E. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol.* **16**, 279-87 (2006).

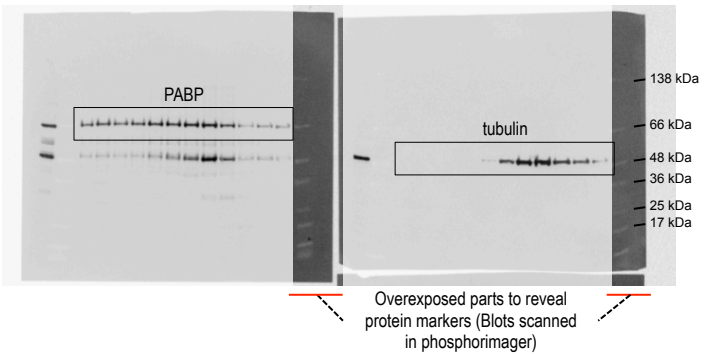
Supplementary Figure - Uncropped scans with size marker indications

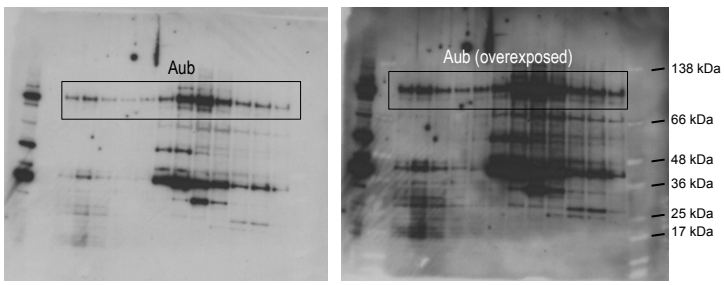
Figure 1A



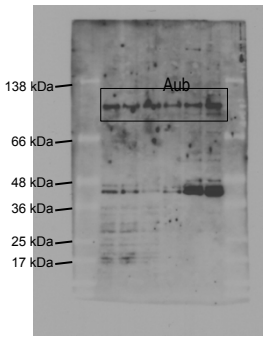
Extended Data Figure 6

Extended Data Figure 6b





Extended Data Figure 6d



Extended Data Figure 6e