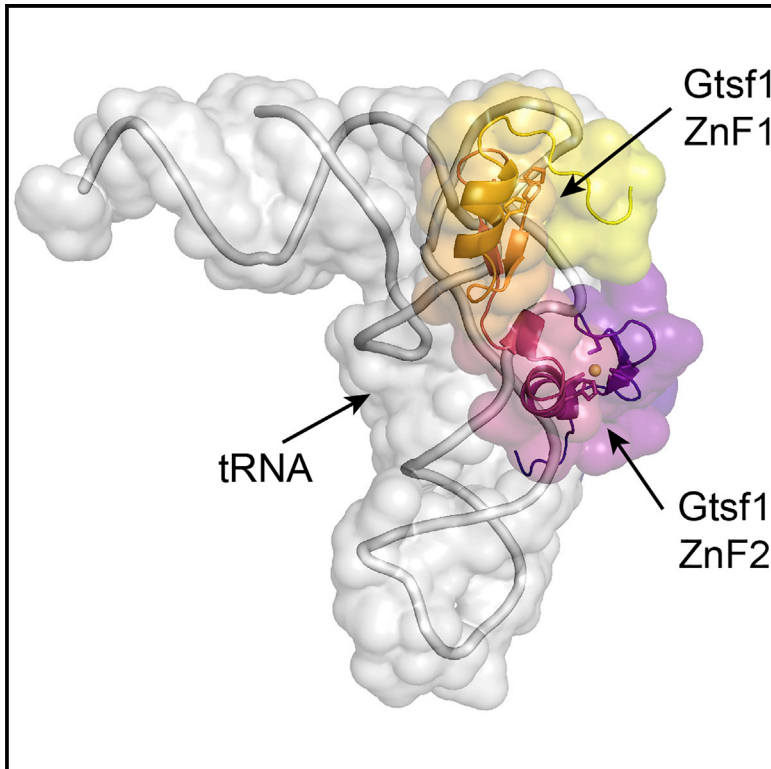


## Asterix/Gtsf1 links tRNAs and piRNA silencing of retrotransposons

### Graphical abstract



### Authors

Jonathan J. Ipsaro, Paul A. O'Brien, Shibani Bhattacharya, Arthur G. Palmer III, Leemor Joshua-Tor

### Correspondence

leemor@cshl.edu

### In brief

Retrotransposon silencing by the piRNA pathway is critical for genomic stability in metazoan germlines. Ipsaro et al. characterize the structure and function of Asterix/Gtsf1, a piRNA effector step protein, showing that it directly binds tRNAs. Ultimately, Asterix may exploit the tRNA dependence of LTR retrotransposons to enforce their piRNA-mediated silencing.

### Highlights

- Asterix/Gtsf1 directly binds RNA via its first CHHC zinc finger
- tRNAs are highly enriched Gtsf1 binding partners
- Cryo-EM structure of Gtsf1 bound to tRNA reveals the complex's architecture
- Asterix knockdown disrupts silencing of tRNA-dependent (LTR) retrotransposons



## Report

# Asterix/Gtsf1 links tRNAs and piRNA silencing of retrotransposons

Jonathan J. Ipsaro,<sup>1</sup> Paul A. O'Brien,<sup>2</sup> Shibani Bhattacharya,<sup>3</sup> Arthur G. Palmer III,<sup>2</sup> and Leemor Joshua-Tor<sup>1,4,\*</sup><sup>1</sup>Howard Hughes Medical Institute, W.M. Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA<sup>2</sup>Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, 650 West 168th Street, New York, NY 10032, USA<sup>3</sup>New York Structural Biology Center, 89 Convent Avenue, New York, NY 10027, USA<sup>4</sup>Lead contact\*Correspondence: [leemor@cshl.edu](mailto:leemor@cshl.edu)<https://doi.org/10.1016/j.celrep.2021.108914>

## SUMMARY

The Piwi-interacting RNA (piRNA) pathway safeguards genomic integrity by silencing transposable elements (transposons) in the germline. While Piwi is the central piRNA factor, others including Asterix/Gtsf1 have also been demonstrated to be critical for effective silencing. Here, using enhanced crosslinking and immunoprecipitation (eCLIP) with a custom informatic pipeline, we show that Asterix/Gtsf1 specifically binds tRNAs in cellular contexts. We determined the structure of mouse Gtsf1 by NMR spectroscopy and identified the RNA-binding interface on the protein's first zinc finger, which was corroborated by biochemical analysis as well as cryo-EM structures of Gtsf1 in complex with co-purifying tRNA. Consistent with the known dependence of long terminal repeat (LTR) retrotransposons on tRNA primers, we demonstrate that LTR retrotransposons are, in fact, preferentially de-repressed in Asterix mutants. Together, these findings link Asterix/Gtsf1, tRNAs, and LTR retrotransposon silencing and suggest that Asterix exploits tRNA dependence to identify transposon transcripts and promote piRNA silencing.

## INTRODUCTION

To maintain genomic integrity, the activity of mobile genetic elements (transposons) must be repressed. This is particularly important in the germline, where transposon silencing, enforced by the Piwi-interacting RNA (piRNA) pathway (Czech and Hannon, 2016; Siomi et al., 2011), affords genetic stability between generations. piRNA silencing is accomplished through interrelated mechanisms that function in distinct cellular compartments. In the cytoplasm, piRNA-directed cleavage leads to post-transcriptional target degradation (Brennecke et al., 2007; Gunawardane et al., 2007). In the nucleus, however, Piwi-piRNA complexes are believed to recognize nascent transposon transcripts, recruit additional factors, and ultimately enforce the deposition of histone H3 lysine 9 trimethylation (H3K9me3) repressive marks (Klenov et al., 2011; Le Thomas et al., 2013; Rozhkov et al., 2013; Sienski et al., 2012).

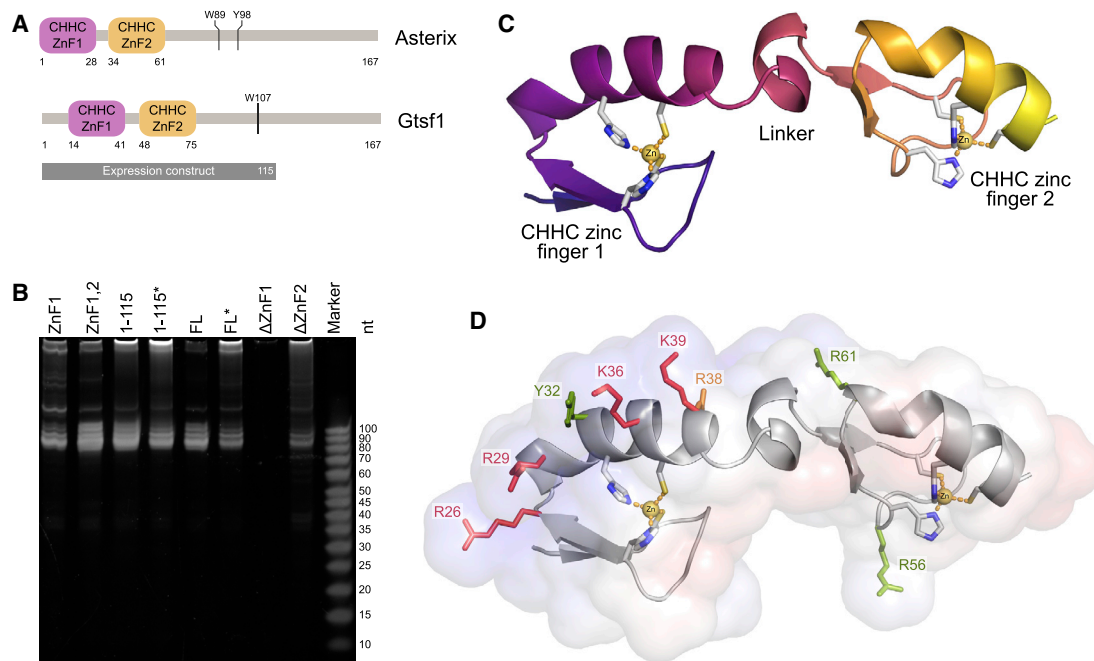
The results of three independent, genome-wide screens revealed a number of candidate proteins that are essential for piRNA silencing (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). While many of these factors affected piRNA biogenesis, others had no obvious effect on piRNA levels or composition. This suggests that this second class of factors principally acts downstream of piRNA biogenesis, most likely in transcriptional gene silencing (TGS) or in the ping-pong cycle (reviewed in Czech and Hannon, 2016).

Recent work on several of these proteins (including, but not limited to, Panoramix [Yu et al., 2015], Nxf2, and Nxt1 [Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019]) has provided a framework for linking the piRNA pathway to deposition of heterochromatic silencing marks. However, many of the molecular and mechanistic underpinnings that govern these connections remain obscure. With this in mind, we endeavored to detail the role of one of the strongest hits in the aforementioned screens, the protein CG3893/Cue110/Asterix/Gtsf1, in piRNA transposon silencing.

In *Drosophila*, expression of Asterix is largely restricted to the female germline, where it is critical not only to transposon silencing, but also more broadly for ovarian development. There, Asterix localizes to the nucleus and has been shown to interact with Piwi (Dönertas et al., 2013; Muerdter et al., 2013; Ohtani et al., 2013). Similarly, gametocyte-specific factor 1 (Gtsf1), the mammalian homolog of Asterix, is involved in retrotransposon suppression and is also important in both oogenesis and spermatogenesis (Krotz et al., 2009; Yoshimura et al., 2018). Reports on the sub-cellular localization of Gtsf1 are mixed, with the most recent findings revealing focal localization in both nuclei and cytoplasmic processing bodies (piP bodies) (Yoshimura et al., 2018).

Asterix and Gtsf1 are small proteins—167 amino acids in length—predicted to consist of two N-terminal CHHC-type zinc fingers and a disordered C-terminal domain (Figure 1A).





**Figure 1. Structure and RNA-binding activity of Asterix/Gtsf1**

(A) Domain architecture. Asterix/Gtsf1 comprises two N-terminal CHHC zinc fingers and a C-terminal tail predicted to be intrinsically disordered. Aromatic residues that interact with Piwi proteins are indicated.

(B) Urea-PAGE analysis of RNAs that co-purify with Gtsf1 truncation constructs. Domains or amino acid ranges of each recombinantly expressed mouse Gtsf1 construct are indicated above the corresponding lane. FL, full-length protein. Asterisks indicate constructs containing four cysteine-to-serine point mutations that were included in the NMR construct to limit aggregation.

(C) Solution structure of mouse Gtsf1. The lowest-energy structure for the protein's core (residues 13–72) is depicted as a ribbon diagram. Zinc-coordinating residues are shown as sticks, with zinc atoms displayed as yellow spheres.

(D) Mapping of the RNA-binding interface. The calculated electrostatic surface of mouse Gtsf1 (scaled from  $-5 k_B T$  in red to  $+5 k_B T$  in blue) displays a positively charged ridge on ZnF1. Zinc-coordinating residues and point mutations tested for effects on RNA binding are shown as sticks (red, abolishes binding; orange, hinders binding; green, no effect).

See also [Figures S1–S4](#) and [Table S1](#).

CHHC zinc fingers have only been identified in eukaryotes and are found in just three protein groups: spliceosomal U11-48K proteins, tRNA methyltransferases, and gametocyte-specific factors (such as Asterix/Gtsf1) ([Andreeva and Tidow, 2008](#)). In the former two cases, these motifs have been demonstrated to bind RNA ([Tidow et al., 2009](#); [Wilkinson et al., 2007](#)).

To detail the role of Asterix/Gtsf1 in retrotransposon silencing, we implemented a combination of biochemical, structural, cell-based, and informatic analyses. Here, we present biochemical evidence that Asterix/Gtsf1 directly binds RNA. We determined the structure of mouse Gtsf1 using nuclear magnetic resonance (NMR) spectroscopy and mapped the RNA-binding site through mutational analysis. Using eCLIP and a customized informatic workflow, we demonstrate that Asterix/Gtsf1 preferentially binds to tRNAs in cells. Using cryo-electron microscopy (cryoEM), we solved a low-resolution structure of Gtsf1 in complex with tRNA. Together, these findings led us to propose a model of how Asterix uses tRNA biology to effect transposon silencing. Informatic analysis of existing datasets implicated Asterix as particularly relevant in silencing long terminal repeat (LTR) transposons, a transposon class that shares an evolutionary history with tRNA.

## RESULTS

### Asterix/Gtsf1 is an RNA-binding protein

We initiated structural studies with recombinantly produced mouse Gtsf1 to systematically characterize its molecular role in retrotransposon silencing. During purification from *Sf9* cells, we observed that Gtsf1 co-purified with endogenous nucleic acids ([Figure S1](#)). These species could be separated by ion exchange chromatography ([Figure S1A](#)), resulting in monodispersed and highly purified protein ([Figures S1B](#) and [S1C](#)) in addition to isolated nucleic acids.

When assessed by urea-PAGE ([Figure S1D](#)), the nucleic acids displayed a narrow size distribution ( $\sim 70$ – $90$  nucleotides), suggesting that specific ligands were being bound. We hypothesized that this Gtsf1-bound material was RNA by analogy to the ligands of other CHHC zinc fingers' proteins' ligands. Treatment with RNase A or sodium hydroxide degraded this material, whereas treatment with DNase I did not, verifying that this was indeed the case ([Figure S1E](#)).

To further pinpoint the RNA-binding activity of Gtsf1, we created a panel of truncation constructs, similarly expressed each in *Sf9*, and assessed which of these co-purified with RNA

(Figure 1B). The first CHHC zinc finger (ZnF1) was found to be both necessary and sufficient for the majority of RNA binding (Figure 1B). Additional inclusion of the second CHHC zinc finger fully recapitulated the RNA size profile as compared of the full-length protein's pull-down.

Purified RNAs are usually unstable, and RNAs of this size are unlikely to be fully protected by a single, 45-amino-acid (~5 kilodalton) protein-binding partner such as ZnF1. Thus, this result suggests that the isolated RNAs were structured, affording them some protection from degradation.

### Overall structure of Gtsf1

As the zinc finger RNA-binding modules were now of primary interest, we examined a construct of mouse Gtsf1, spanning residues 1–115, using NMR spectroscopy (Figures 1C, S2, and S3; Table S1). In agreement with folding and domain predictions, initial heteronuclear single quantum coherence (HSQC) experiments suggested the protein contained both ordered and disordered segments (Figure S2A). Subsequent backbone assignment more specifically indicated the structured core of the protein spanned residues 13–73, and residues outside this range tended to be disordered. Analysis of secondary chemical shifts (Figures S2D and S2E), as well as backbone conformation predictions using torsion angle likelihood obtained from shift and sequence similarity (TALOS) and chemical shift index (CSI) methods (Figures S2F and S2G), indicated strand-strand-helix architectures for both ZnF1 and ZnF2, similar to that observed for the only other reported CHHC zinc finger structure (Tidow et al., 2009).

Structure determination of residues 1–80 revealed two tandem, CHHC zinc finger domains (ZnF1, ZnF2) connected by an  $\alpha$ -helix-containing linker (Figures 1C and S3) with the N and C termini being intrinsically disordered. In preliminary structure calculations, which did not include restraints for the CHHC residues with zinc, each zinc finger already displayed a strand-strand-helix fold with the appropriate zinc-coordinating residues in proximity to one another.

In the NMR-derived structural ensemble of the 20 final, lowest-energy structures (Figures S3A–S3C), the relative positions of the zinc finger domains varied somewhat owing to flexibility in the intervening linker. Nonetheless, structure calculations for the individual domains were highly superimposable (Figures S3F and S3G) with root-mean-square deviation (RMSD) values of 0.2 Å for backbone atoms for each zinc finger, further allowing for confident interpretation of each domain's individual structure. Moreover, these domains were highly superimposable with each other and the only other CHHC zinc finger structure available (from the U11-48K spliceosomal protein; Tidow et al., 2009) (Figure S3H). Co-evolution analysis (Ovchinnikov et al., 2014) additionally corroborated the overall protein fold, with several intra-ZnF residues displaying evidence of co-evolution (Figure S3E). Final validation of the structural ensemble with Molprobit (Davis et al., 2007) indicated reasonable geometry overall, with the core (residues 13–73) possessing very few violations (Table S1).

### ZnF1 presents a conserved RNA-binding interface

Guided by the protein structure, we next mapped the RNA-binding interface. Calculation of the electrostatic surface of Gtsf1

revealed a pronounced, positively charged ridge running the length of ZnF1 (Figure 1D). Mutagenesis of single basic residues along this patch abrogated or reduced RNA-binding activity with no apparent effects on expression or solubility (Figures S4A and S4C), indicating that they indeed form part of the RNA-binding interface. Correspondingly, mutations of basic residues on ZnF2 (Figures 1D, S4B, and S4D) did not affect RNA binding.

Evolutionary analysis corroborated the importance of ZnF1, with residues important for RNA binding among the most highly conserved in the structure (Figure S3D). Although some key residues—notably, in the CHHC metal-coordination site—of ZnF2 were also highly conserved, ZnF2 was more variable overall. Together, these findings bolstered our initial characterization that ZnF1 mediates RNA interactions (Figure 1B) and precisely identified basic residues in this region as forming a conserved interface for RNA binding.

### Recombinantly produced Gtsf1 co-purifies with tRNAs

To complement the biochemical characterization of Gtsf1 protein and gain insight into the possible identities of biologically relevant ligands, we next analyzed the RNAs that were being retained during recombinant expression in *Sf9*. RNAs that co-purified with mouse Gtsf1 were isolated by phenol:chloroform extraction, ethanol precipitated, then subjected to size selection and next-generation sequencing.

Consistent with the previous observation that the bound RNAs were approximately 70–90 nucleotides in size, we found considerable enrichment of tRNAs in the Gtsf1 pulldown compared to size-matched controls (Table S2). This enrichment was readily apparent, even though the *Sf9* genome is not fully annotated, as approximately 15% of the sequencing library comprised a single tRNA species. Moreover, each of the 20 most abundant sequences was determined to be tRNA derived, with 50% of all library reads corresponding to these 20 sequences.

Enrichment of tRNA sequences contrasted with size-matched controls from extracted *Sf9* total RNAs where the top sequence was derived from the highly abundant large ribosomal subunit, yet nonetheless made up only ~4% of the library. The top tRNA read in the size-matched control contributed only approximately 0.3% of the total reads.

### Gtsf1 directly binds tRNAs in cellular contexts

To catalog RNAs interacting directly with Gtsf1 in a mammalian cellular context, we employed eCLIP (Van Nostrand et al., 2016). Strep-tagged Gtsf1 was transfected into a mouse embryonal teratoma cell line (P19), bound RNAs were covalently linked using UV crosslinking, the complexes were isolated by affinity purification, and the RNA was subjected to next-generation sequencing.

Many classes of RNA—such as rRNAs, tRNAs, and highly repetitive genetic elements like transposons—are typically excluded from downstream analysis due to ambiguity in read mapping and/or their high abundance. Given the relevance of these gene classes in the context of Asterix/Gtsf1, we therefore developed a custom bioinformatic workflow to ensure their inclusion. Read mapping was performed, allowing for multimapping with up to 50 genomic sites per read (Dobin et al., 2013). Various sources of well-curated gene annotations (including gencode

[Frankish et al., 2019], miRBase [Kozomara et al., 2019], tRNADB [Chan and Lowe, 2016], piRNAclusterdb [Rosenkranz, 2016], and Tetrascripts [Jin et al., 2015]) were integrated for customized annotation calling. We sequentially filtered each read into a single annotation class based on the known biological abundance of that class. Subsequently, the enrichment for a given gene or annotation class was determined, first by subtracting levels in a background (non-crosslinked) pulldown and then by comparing to input read counts, with multimapping reads being 1/n-normalized across genes within the assigned annotation class.

Analysis of these data by annotation category displayed a substantial enrichment of tRNAs (Figure 2A), and when analyzed as a distribution of fold enrichments for individual genes in each annotation class (i.e., per locus), we again found a preponderance of tRNA enrichment (Figure 2B). Inspection of the tRNA reads revealed widespread coverage across the tRNA body, suggesting that full-length (or nearly full-length) tRNAs were being bound. Although some variability was observed in the degree of enrichment across different tRNAs, in contrast with the *Sf9* pulldowns, no particular tRNA or set of tRNAs was selected preferentially based on the anticodon (Figure 2C).

We repeated the informatic analysis to test the robustness of these results (Figure S5). We disfavored tRNA enrichment by setting tRNAs as the lowest priority in our annotation ranking (Figure S5B). Additionally, we developed a complementary informatic pipeline that built an alternative gene model to accommodate multimapping reads and then used tools such as DESeq2 (Love et al., 2014) to quantify enrichment (Figure S5C). While the absolute strength of the enrichment signals did vary among these analyses, each of these workflows supported the conclusion that tRNAs were the most enriched annotation class.

### Asterix directly binds tRNAs in *Drosophila* OSS cells

To date, the most productive model organism for dissecting piRNA biology has been *Drosophila melanogaster*, especially given the availability of an ovary-derived cell culture line (ovarian somatic sheath [OSS] cells) with an intact primary piRNA silencing pathway. Indeed, the requirement of Asterix for effective transposon silencing in *Drosophila* was discovered in OSS cells (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013).

Therefore, to compare our observations from mouse Gtsf1 to *Drosophila* Asterix and establish a framework for better cross-referencing observations between mammals and flies, we similarly performed eCLIP experiments in OSS using transfected, strep-tagged Asterix. Once more, tRNAs were found to be highly enriched both as a class and individually (Figures 2C–2E).

Finally, to verify that these findings were not due to overexpression artifacts, we performed eCLIP experiments using FLAG-tagged Asterix under the control of its endogenous promoter. Again, tRNAs were enriched both individually and as a class (Figure S6). Interestingly, some piRNA enrichment was also observed in this experiment; however, unlike tRNAs, this was not found as universally across piRNA annotations. This observation may be explained by Asterix's known association with piRNA-induced silencing complexes (piRISC) (Muerdter et al., 2013; Ohtani et al., 2013; Yoshimura et al., 2009) coupled

with a preponderance of basic residues in the protein's C-terminal, Piwi-interacting tail that are absent in the mammalian ortholog.

### Gtsf1 binds tRNAs in the D-arm

To gain insight into the interaction between tRNA and Gtsf1, we further scrutinized the eCLIP data. In eCLIP, a pileup of read ends is expected at the cross-linking site, presumably due to interference from the cross-link with reverse transcription during preparation of the library. Analysis of library 5' ends can thus be used to inform potential sequence motifs that are specifically engaged with the cross-linked protein.

An initial analysis of genomic sequences in the vicinity of library 5' ends did not reveal obvious binding motifs. With the apparent preference for tRNAs as a Gtsf1 ligand, and recognizing that tRNAs are highly structured, we hypothesized that RNA binding by Gtsf1 could be driven by structural determinants, perhaps more so than by RNA sequence.

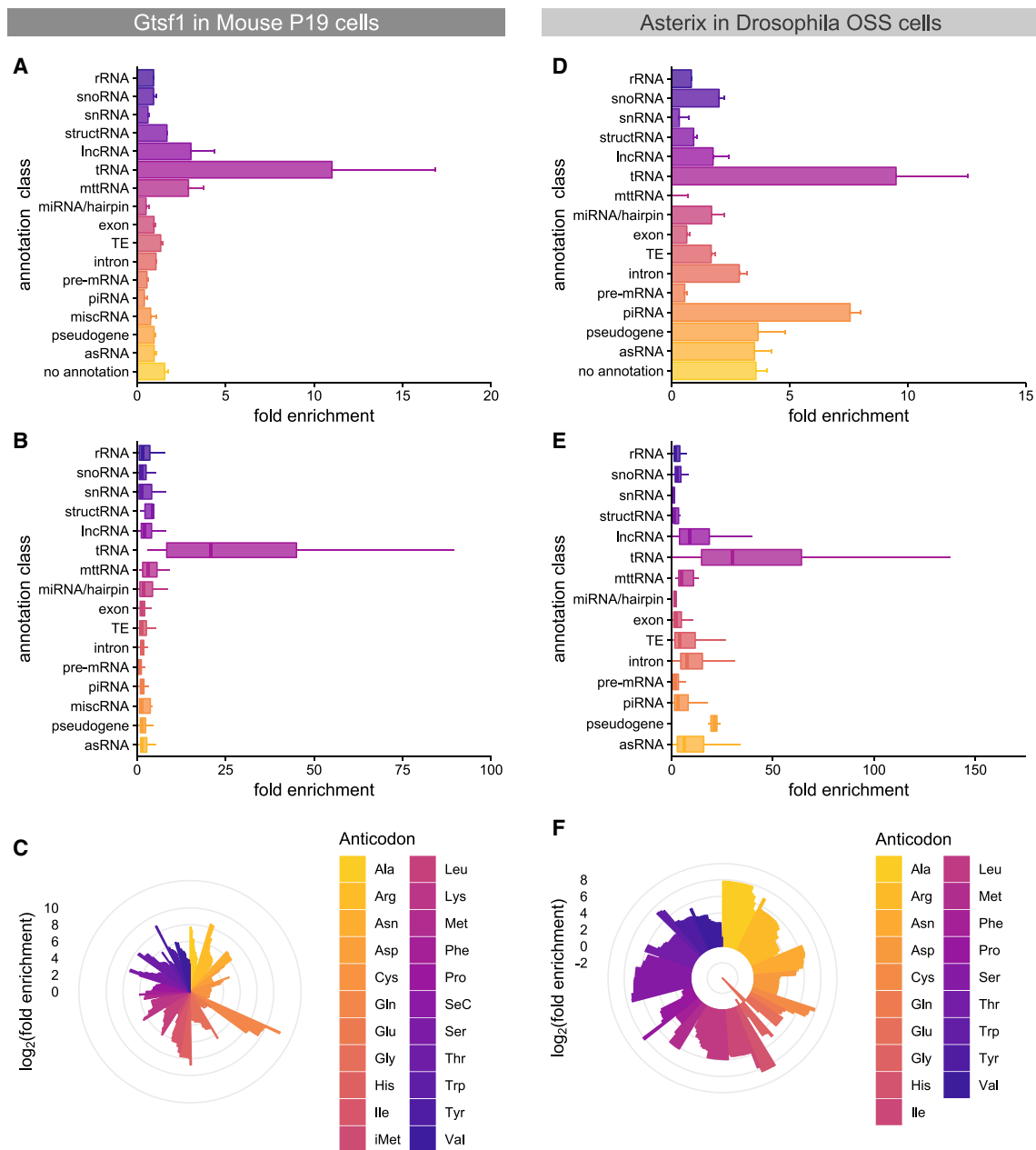
To test this, the 5'-end positions of mapped tRNA reads were plotted as a histogram on a model tRNA 73 nucleotides in length (not including the CCA tail) and scored according to fold enrichment weighted by Z score (Figure 3A). Using the analysis that retained the most tRNA reads, we were able to identify two high-scoring sites at nucleotides 18 and 22 in the D-arm (Figure 3B).

### Structure of Gtsf1 in complex with tRNA

Having characterized both Gtsf1 and its RNA ligands in several contexts, we next aimed to determine a structure of the protein-RNA complex. Initial NMR experiments on Gtsf1 reconstituted with RNA ligands showed evidence of binding but were hampered by poor-quality spectra, which likely resulted from slow tumbling of the complex. Attempts at crystallization were met with similar difficulties, presumably due to the inherent flexibility present in the protein structure. Although the molecular weight for the complex is a mere ~45 kDa (19 kDa for Gtsf1 and 25 kDa for a typical tRNA), we speculated that the density of the bound RNA nevertheless could allow for structure determination using cryo-EM.

We subjected recombinantly expressed mouse Gtsf1 from *Sf9* cells—which, as mentioned, co-purifies with endogenous RNA—to cryo-EM. Given the relatively small size of this complex, the presence of disordered regions in the protein, and the fact that the sample included a heterogeneous population of RNAs, we opted to image this material at 200 keV (rather than the customary 300 keV) to increase contrast and aid in particle picking. Nearly 5,000 micrographs were collected and resulted in almost 500,000 particles (see Method details).

From these data, we were able to obtain a low-resolution reconstruction (Figures 3C, S7A, and S7B) that had the dimensions and shape of a tRNA with two additional domains. To more accurately orient the Gtsf1 structure into the reconstruction, we applied a similar workflow to an even smaller complex, comprising only ZnF1 in complex with co-purifying RNAs (estimated total molecular weight of 31 kDa) (Figures 3D and S7C). By comparing the two reconstructions, we were able to generate a difference map to unambiguously deduce the location of ZnF2 (Figure 3E, fuchsia surface).



**Figure 2. Asterix/Gtsf1 specifically and directly binds tRNAs in cellular contexts**

(A and D) Gene class enrichment of Asterix/Gtsf1-bound RNAs. The fold enrichment of each annotation class in eCLIP experiments for (A) mouse Gtsf1 in P19 cells and (D) *Drosophila* Asterix in OSS cells is shown as a bar chart. Values indicate the average fold enrichment for two replicate libraries. Error bars indicate the standard error.

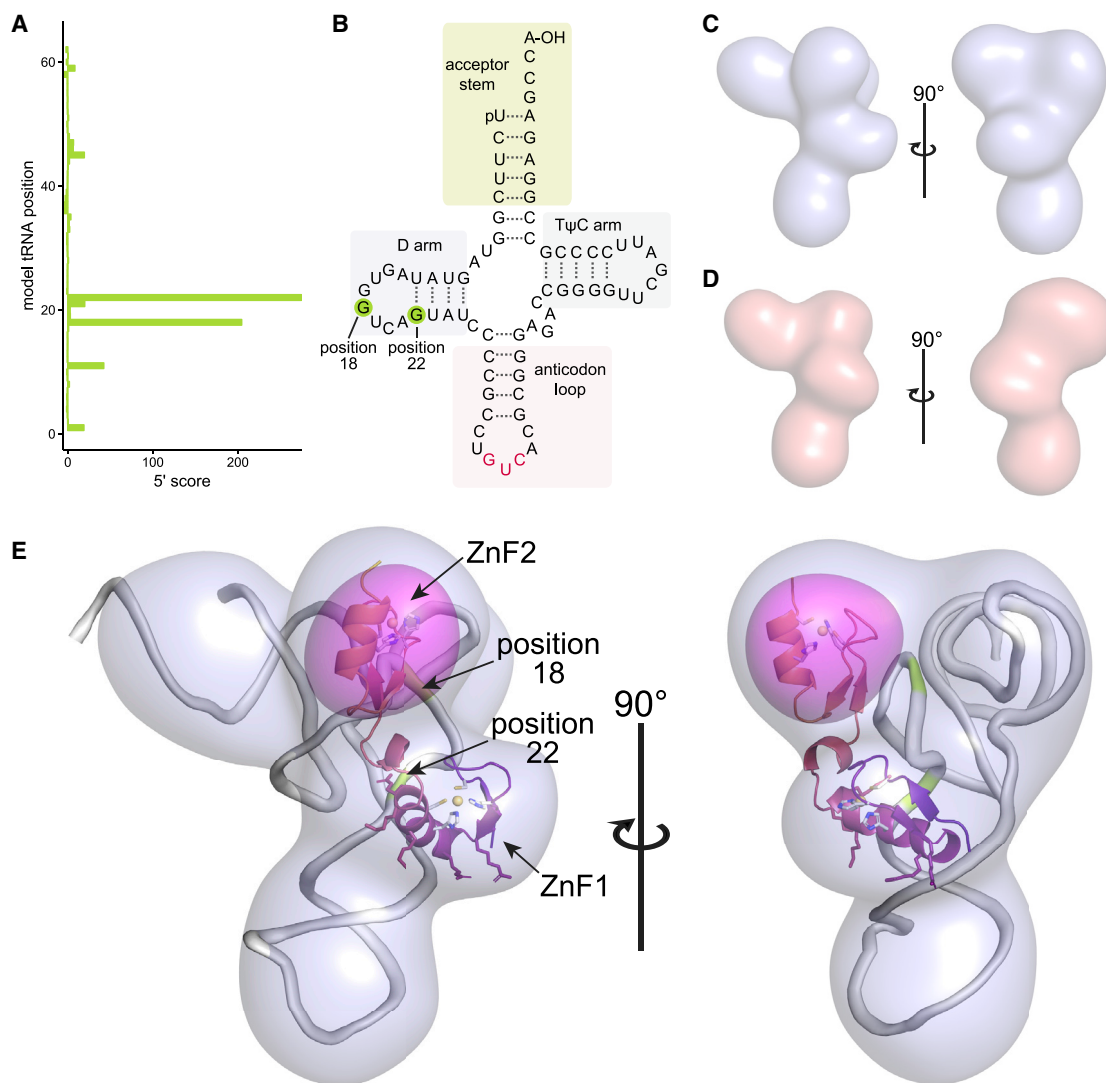
(B and E) Annotation enrichment distribution plots. Fold enrichment distributions among gene annotations within each class are displayed as boxplots for (B) P19 mouse and (E) *Drosophila* OSS eCLIP experiments.

(C and F) Fold enrichment scores per tRNA, sorted by anticodon. tRNA enrichment for (C) P19 mouse and (F) *Drosophila* OSS eCLIP experiments are plotted as  $\log_2(\text{fold enrichment})$  on the radial bar graph. Multiple bars of the same colors indicate distinct gene annotations for that anti-codon.

See also [Figures S5](#) and [S6](#).

Ultimately, we were able to place tRNA and the zinc finger domains of mouse Gtsf1 into the cryo-EM map with little residual density, resulting in a low-resolution structure of the complex. Consistent with the biochemical analysis, ZnF1, the

domain primarily responsible for binding RNA, formed an interface with the most probable cross-linking tRNA nucleotides. The second zinc finger extended toward the tRNA acceptor stem.



### Figure 3. Structure of mouse Gtsf1 in complex with tRNA

(A and B) Mapping of favored RNA crosslinking sites. (A) Analysis of tRNA reads indicates two preferred cross-linking sites that correspond to nucleotides 18 and 22 of a model tRNA. (B) These sites are in the tRNA D-arm and are highlighted (green) on a tRNA secondary structure diagram.

(C and D) Cryo-EM maps of mouse Gtsf1 bound to co-purifying tRNA. (C) EM density map of full-length MmGtsf1 in the presence of co-purifying RNA. The map is filtered to 10 Å and displayed with a 3.2- $\sigma$  cutoff. (D) cryo-EM map of a truncation construct of MmGtsf1 (only the first zinc finger; residues 1–45) in the presence of co-purifying RNA. Filtered to 10 Å and displayed with a 1.4- $\sigma$  cutoff, this map shows a very similar shape to that observed in (C), but lacking one lobe of the structure.

(E) Difference map and structure. The full-length reconstruction (light blue) is shown with the modeled tRNA (gray), the Gtsf1 NMR structure (core residues 13–72; colored as in Figure 1C), and the difference map calculated from reconstructions (C) and (D) (pink surface; contoured at 5.8  $\sigma$ ). Positions of favored tRNA cross-linking are highlighted in green and labeled. Residues of ZnF1 that are important for binding are shown as sticks, colored according to their position in the ribbon diagram. This alignment accommodates both molecules within the full-length reconstruction, positions the RNA-binding residues of MmGtsf1 ZnF1 near to the tRNA, places the most highly cross-linked RNA residues at the protein interaction surface, and identifies the placement of ZnF2.

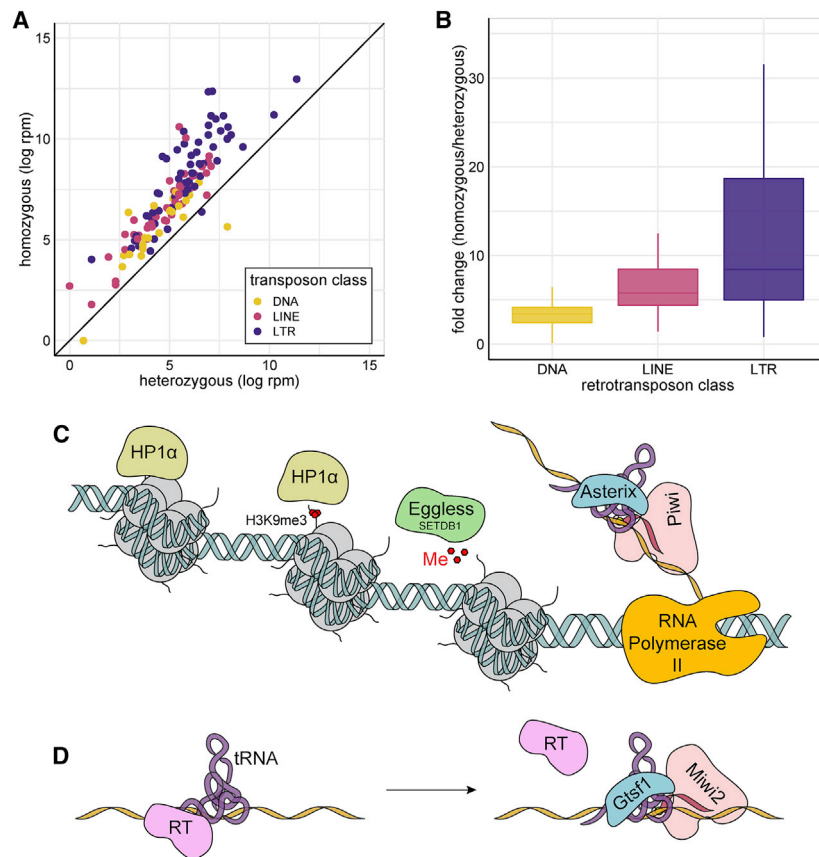
See also Figure S7.

### Asterix knockout predominantly affects LTR-class transposons

To understand how binding of Asterix/Gtsf1 to tRNA might be involved in piRNA silencing and repression of transposon expression, we noted that certain groups of retroviruses and retrotransposons require host tRNAs as primers for their replication

by reverse transcription (Martinez, 2017; Schorn et al., 2017). Such retrotransposons belong to the LTR family and are characterized by the presence of repeated DNA sequences that flank the transposon body.

In order to transpose, LTR transposon transcripts must be reverse transcribed. The reverse transcriptase enzyme



**Figure 4. LTR transposons (a class that relies on tRNAs for retrotransposition) are preferentially de-repressed upon Asterix knockout**

(A) Comparison of transposon levels in Asterix knockout versus Asterix heterozygous flies. Transposon expression as determined by RNA-seq in mutant versus heterozygous flies is plotted and color-coded by transposon class.

(B) Transposon fold change distribution plots. Fold changes among gene annotations within each transposon class are displayed as boxplots.

(C and D) Models for the role of Asterix/Gtsf1 in silencing.

(C) In the nucleus, Asterix may utilize tRNAs to recognize primer binding sites in nascent transposon transcripts. Enhanced recruitment of piRISCs (pink) could then be achieved through interactions with Asterix's C-terminal tail in addition to the protein-protein interactions between Asterix and Piwi and between the piRNA and the transcript. This leads to further recruitment of histone modification enzymes (such as Egless/SETDB1; green) and eventual binding of HP1α (beige) to establish heterochromatin.

(D) In the cytoplasm, similar interactions among Gtsf1, tRNAs, transposons, and Piwi proteins may also occur to displace reverse transcriptase and/or enhance the slicing activity of cytoplasmic piRISCs.

See also Figure S8.

requires priming, which is most often accomplished using host tRNAs recognizing a primer binding site (PBS) immediately downstream of the 5' LTR. This dependence on host tRNA recognition thus makes the PBS a conspicuous feature of LTR transposons, which can indeed be exploited for LTR recognition, as has been shown with tRNA fragments (Schorn et al., 2017; Schorn and Martienssen, 2018).

We reasoned that if Asterix/Gtsf1 is indeed using tRNAs to recognize LTR transposon transcripts, then this class of transposons should be highly affected by Asterix/Gtsf1 knockdown. Re-analysis of RNA-seq data from Asterix knockout flies (Muerdter et al., 2013) supported this finding and indicated that both in absolute read counts and in distributions of fold changes among loci, LTR retrotransposons were indeed the most affected transposon class (Figures 4A and 4B).

## DISCUSSION

Several lines of evidence now establish Asterix/Gtsf1 as a bona fide tRNA-binding protein: the presence of CHHC zinc fingers that, in other proteins, bind structured RNAs; co-purification of tRNAs from the recombinant expression of Gtsf1; the ability to abolish these interactions with individual Gtsf1 point mutants; and direct binding of Gtsf1 to tRNAs in multiple relevant cell culture systems.

Taken together with the marked effects of Asterix knockout on LTR retrotransposons and the evolutionary history that LTR retrotransposons share with tRNAs, tRNA binding by Asterix/Gtsf1 suggests that these proteins

are co-opting molecular epitopes of tRNAs to facilitate transposon silencing.

In the nucleus, where Asterix/Gtsf1 localizes in both mice and flies, LTR tRNA primer binding could be used to augment the specificity of Piwi/MIWI2 (both of which have been identified as binding partners [Dönertas et al., 2013; Yoshimura et al., 2018]) (Figure 4C). The interactions between Asterix/Gtsf1 and Piwi/MIWI2, between Asterix/Gtsf1 and tRNA, between tRNAs and transposon transcripts, and between piRISCs and nascent transcripts could reinforce one another, thereby enhancing target recognition. In the cytoplasm, Gtsf1 could likewise assist in the recruitment of Piwi partners—in this case, for ping-pong processing—while potentially acting simultaneously to interfere with tRNA-primer/reverse transcriptase engagement and limit retrotransposon replication (Figure 4D).

Given that the precise ordering of complex formation is presently unknown, an attractive possibility is that tRNAs engaged with PBSs could recruit Asterix/Gtsf1 more effectively than free tRNAs and, in doing so, assist in Piwi/MIWI2 target recognition. Such an assembly mechanism effectively narrows the pool of tRNAs recognized by Asterix/Gtsf1, which is likely important, given the high concentration of cellular tRNAs and the known observation that certain tRNAs are favored in reverse transcription of particular retroelements. It remains to be understood, however—both in typical retroelement replication and in its inhibition through the mechanisms proposed—how tRNA



unwinding is accomplished, whether it be by additional co-factors or simply by part of the dynamic nature of the acceptor stem (Chan et al., 2020). One noteworthy observation from the cryo-EM reconstruction is the placement of the second zinc finger and, by extension, the intrinsically disordered C terminus of the protein: projecting toward the tRNA acceptor stem (and thus the tRNA primer for reverse transcription). While our biochemical data support that only the first zinc finger is necessary and sufficient for binding tRNAs *in vitro*, it is possible that more elaborate interactions between the tRNA and an engaged transposon target could be recognized by the second zinc finger. This sort of interaction would be reminiscent of those observed in the related CHHC zinc finger protein U11-K48 from the minor spliceosome (Tidow et al., 2009).

As RNA interference pathways are studied across many species and cell types, variations on several themes continue to emerge. In addition to the most obvious presence of a small-RNA-loaded RISC as the central component of the pathway, complexes that establish multivalent interactions with silencing targets are also prevalent—GW182-mediated recruitment of Ago2 in humans and the RITS complex in *S. pombe* are prime examples (Debeauchamp et al., 2008; Elkayam et al., 2017; Mota-medi et al., 2004) (Figure S8). Moreover, GTSF-1 in *C. elegans* has been demonstrated as critical in the formation of a functional RNA-dependent RNA polymerase complex (RDRP), where it is believed to aid in the assembly of RNA silencing complexes (Almeida et al., 2018). These multipartite binding platforms confer enhanced molecular specificity while also allowing flexibility in the repertoire of silencing targets. In this case, our findings suggest that Gtsf1/Asterix exploits a key vulnerability in many retroelements: their dependence on host tRNAs for their replication.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Sf9 cell culture
  - P19 cell culture
  - OSS cell culture
- METHOD DETAILS
  - Cloning
  - Expression and purification
  - Characterization of co-purifying Sf9 RNAs
- ECLIP LIBRARY GENERATION
  - Cell culture
  - P19 cell transfection of MmGtsf1-TEV-Strep
  - OSS cell transfection of Asterix-TEV-Strep
  - Library preparation
  - Next-generation sequencing
- ECLIP PROCESSING
  - Rationale

- Summary
- Gene annotations
- tRNA analysis
- NMR spectroscopy
- Cryo-electron microscopy
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2021.108914>.

## ACKNOWLEDGMENTS

We thank Sonam Bhatia, Sara Ballouz, Sarah Diermeier, Paloma Guzzardo, Matt Jaremko, Justin Kinney, Molly Gale Hammell, Greg Hannon, Katie Meze, Felix Muerdter, Kathryn O'Neill, Nikolay Rozhkov, David Spector, and Dennis Thomas for advice. Astrid Haase provided custom piRNA cluster gene annotations for *D. melanogaster*. CRISPR-tagged Asterix OSS cells were gifted by Julius Brennecke. P19 cells and advice were provided by Andrea Schorn and Rob Martienssen. We thank Amanda Epstein and visiting students Michael Jacobs, Edward Twomey, and Dexter Adams for technical assistance. The 600- and 800-MHz NMR spectrometers were supported by NIH grants S10RR026540 and S10OD016432, respectively, and The Center on Macromolecular Dynamics by NMR Spectroscopy at the New York Structural Biology Center is supported by NIH grant P41 GM118302. This work was performed with assistance from the CSHL Next-generation Sequencing, Sequence Technologies and Analysis, and Mass Spectrometry Shared Resources, which are supported by the Cancer Center Support Grant 5P30CA045508. Cryo-EM was performed at the CSHL cryo-EM facility. OSS cells, medium components, and selected plasmids were acquired from the Drosophila Genomics Resource Center, which is supported by NIH grant 2P40OD010949. We thank Life Science Editors for editorial assistance. This work was supported by National Institutes of Health (NIH) grants F32GM097888 (J.J.I.), R01GM050291 (A.G.P.), R35GM130398 (A.G.P.), and T32 GM008281-28 (P.A.O.). L.J. is an Investigator of the Howard Hughes Medical Institute.

## AUTHOR CONTRIBUTIONS

J.J.I. conceptualized and performed biochemical, next-generation sequencing, informatic analysis, NMR sample preparation, and cryo-EM. P.A.O. and S.B. collected and processed NMR data. A.G.P. conceptualized and supervised NMR data collection and assisted with analysis. L.J. conceptualized and supervised biochemical, sequencing and informatic, and cryo-EM experiments. All authors contributed to manuscript preparation.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 29, 2020

Revised: January 15, 2021

Accepted: March 5, 2021

Published: March 30, 2021

## REFERENCES

Almeida, M.V., Dietz, S., Redl, S., Karaulanov, E., Hildebrandt, A., Renz, C., Ulrich, H.D., König, J., Butter, F., and Ketting, R.F. (2018). GTSF-1 is required for formation of a functional RNA-dependent RNA Polymerase complex in *Caenorhabditis elegans*. *EMBO J.* 37, e99325.

- Andreeva, A., and Tidow, H. (2008). A novel CHHC Zn-finger domain found in spliceosomal proteins and tRNA modifying enzymes. *Bioinformatics* 24, 2277–2280.
- Antczak, M., Popena, M., Zok, T., Sarzynska, J., Ratajczak, T., Tomczyk, K., Adamiak, R.W., and Szachniuk, M. (2016). New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta Biochim. Pol.* 63, 737–744.
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–50.
- Barraud, P., Schubert, M., and Allain, F.H. (2012). A strong <sup>13</sup>C chemical shift signature provides the coordination mode of histidines in zinc-binding proteins. *J. Biomol. NMR* 53, 93–101.
- Batki, J., Schnabl, J., Wang, J., Handler, D., Andreev, V.I., Stieger, C.E., Novatchkova, M., Lampersberger, L., Kauneckaite, K., Xie, W., et al. (2019). The nascent RNA binding complex SF1NX licenses piRNA-guided heterochromatin formation. *Nat. Struct. Mol. Biol.* 26, 720–731.
- Bersch, B., Bougault, C., Roux, L., Favier, A., Vernet, T., and Durmort, C. (2013). New insights into histidine triad proteins: solution structure of a *Streptococcus pneumoniae* PhtD domain and zinc transfer to AdcAll. *PLoS ONE* 8, e81168.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089–1103.
- Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54, 905–921.
- Cavanagh, J., Fairbrother, W.J., Palmer, A.G., Rance, M., and Skelton, N.J. (2007). *Protein NMR Spectroscopy: Principles and Practice*, Second Edition (Academic Press).
- Chan, P.P., and Lowe, T.M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* 44, D184–D189.
- Chan, C.W., Badong, D., Rajan, R., and Mondragon, A. (2020). Crystal structures of an unmodified bacterial tRNA reveal intrinsic structural flexibility and plasticity as general properties of unbound tRNAs. *RNA* 26, 278–289.
- Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem. Sci.* 41, 324–337.
- Czech, B., Preall, J.B., McGinn, J., and Hannon, G.J. (2013). A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol. Cell* 50, 749–761.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375–83.
- Debeauchamp, J.L., Moses, A., Noffsinger, V.J., Ulrich, D.L., Job, G., Kosinski, A.M., and Partridge, J.F. (2008). Chp1-Tas3 interaction is required to recruit RITS to fission yeast centromeres and for maintenance of centromeric heterochromatin. *Mol. Cell Biol.* 28, 2154–2166.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 6, 277–293.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Dönertas, D., Sienski, G., and Brennecke, J. (2013). *Drosophila* Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev.* 27, 1693–1705.
- Elkayam, E., Faehnle, C.R., Morales, M., Sun, J., Li, H., and Joshua-Tor, L. (2017). Multivalent Recruitment of Human Argonaute by GW182. *Mol. Cell* 67, 646–658.e3.
- Fabry, M.H., Ciabrelli, F., Munafò, M., Eastwood, E.L., Kneuss, E., Falciatori, I., Falconio, F.A., Hannon, G.J., and Czech, B. (2019). piRNA-guided co-transcriptional silencing coopts nuclear export factors. *eLife* 8, e47999.
- Findeisen, M., Brand, T., and Berger, S. (2007). A <sup>1</sup>H-NMR thermometer suitable for cryoprobes. *Magn. Reson. Chem.* 45, 175–178.
- Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* 47, D766–D773.
- Grant, T., Rohou, A., and Grigorieff, N. (2018). cistEM, user-friendly software for single-particle image processing. *eLife* 7, e35383.
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315, 1587–1590.
- Hadley, W. (2016). *ggplot2: Elegant Graphics for Data Analysis* (Springer).
- Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F.S., and Brennecke, J. (2013). The genetic makeup of the *Drosophila* piRNA pathway. *Mol. Cell* 50, 762–777.
- Holm, L. (2019). Benchmarking Fold Detection by DaliLite v.5. *Bioinformatics* 35, 5326–5327.
- Jin, Y., Tam, O.H., Paniagua, E., and Hammell, M. (2015). TETranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* 31, 3593–3599.
- Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J., Liles, K., et al. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* 27, 112–128.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006.
- Klenov, M.S., Sokolova, O.A., Yakushev, E.Y., Stolyarenko, A.D., Mikhaleva, E.A., Lavrov, S.A., and Gvozdev, V.A. (2011). Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc. Natl. Acad. Sci. USA* 108, 18760–18765.
- Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162.
- Krotz, S.P., Ballow, D.J., Choi, Y., and Rajkovic, A. (2009). Expression and localization of the novel and highly conserved gametocyte-specific factor 1 during oogenesis and spermatogenesis. *Fertil. Steril.* 91, 2020–2024.
- Lakomek, N.A., Ying, J., and Bax, A. (2012). Measurement of <sup>15</sup>N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR* 53, 209–221.
- Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., and Tóth, K.F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* 27, 390–399.
- Lee, W., Tonelli, M., and Markley, J.L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31, 1325–1327.
- Linge, J.P., Habeck, M., Rieping, W., and Nilges, M. (2003). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* 19, 315–316.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963.
- Martinez, G. (2017). tRNAs as primers and inhibitors of retrotransposons. *Mob. Genet. Elements* 7, 1–6.
- Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* 119, 789–802.
- Muerdter, F., Guzzardo, P.M., Gillis, J., Luo, Y., Yu, Y., Chen, C., Fekete, R., and Hannon, G.J. (2013). A genome-wide RNAi screen draws a genetic

- framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol. Cell* **50**, 736–748.
- Murano, K., Iwasaki, Y.W., Ishizu, H., Mashiko, A., Shibuya, A., Kondo, S., Adachi, S., Suzuki, S., Saito, K., Natsume, T., et al. (2019). Nuclear RNA export factor variant initiates piRNA-guided co-transcriptional silencing. *EMBO J.* **38**, e102870.
- Ohtani, H., Iwasaki, Y.W., Shibuya, A., Siomi, H., Siomi, M.C., and Saito, K. (2013). DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the *Drosophila* ovary. *Genes Dev.* **27**, 1656–1661.
- Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030.
- Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K.J., Lukasiak, P., Bartol, N., Blazewicz, J., and Adamiak, R.W. (2012). Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **40**, e112.
- Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinformatics* **47**, 11.12–11.34.
- R Team (2019). R: A Language and Environment for Statistical Computing. <https://www.r-project.org/>.
- Rosenkranz, D. (2016). piRNA cluster database: a web resource for piRNA producing loci. *Nucleic Acids Res.* **44**, D223–D230.
- Rozhkov, N.V., Hammell, M., and Hannon, G.J. (2013). Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes Dev.* **27**, 400–412.
- Schorn, A.J., and Martienssen, R. (2018). Tie-Break: Host and Retrotransposons Play tRNA. *Trends Cell Biol.* **28**, 793–806.
- Schorn, A.J., Gutbrod, M.J., LeBlanc, C., and Martienssen, R. (2017). LTR-Retrotransposon Control by tRNA-Derived Small RNAs. *Cell* **170**, 61–71.e11.
- Schrödinger, LLC (2019). The PyMOL Molecular Graphics System, Version 2.0. <https://pymol.org/2/#page-top>.
- Shaikh, T.R., Gao, H., Baxter, W.T., Asturias, F.J., Boisset, N., Leith, A., and Frank, J. (2008). SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protoc.* **3**, 1941–1974.
- Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223.
- Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964–980.
- Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12**, 246–258.
- Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* **16**, 1146–1152.
- Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., et al.; FlyBase Consortium (2019). FlyBase 2.0: the next generation. *Nucleic Acids Res.* **47**, D759–D765.
- Tidow, H., Andreeva, A., Rutherford, T.J., and Fersht, A.R. (2009). Solution structure of the U11-48K CHHC zinc-finger domain that specifically binds the 5' splice site of U12-type introns. *Structure* **17**, 294–302.
- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514.
- Wilkinson, M.L., Crary, S.M., Jackman, J.E., Grayhack, E.J., and Phizicky, E.M. (2007). The 2'-O-methyltransferase responsible for modification of yeast tRNA at position 4. *RNA* **13**, 404–413.
- Yoshimura, T., Toyoda, S., Kuramochi-Miyagawa, S., Miyazaki, T., Miyazaki, S., Tashiro, F., Yamato, E., Nakano, T., and Miyazaki, J. (2009). Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. *Dev. Biol.* **335**, 216–227.
- Yoshimura, T., Watanabe, T., Kuramochi-Miyagawa, S., Takemoto, N., Shimoto, Y., Kudo, A., Kanai-Azuma, M., Tashiro, F., Miyazaki, S., Katanaya, A., et al. (2018). Mouse GTSF1 is an essential factor for secondary piRNA biogenesis. *EMBO Rep.* **19**, e42054.
- Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J.B., Ma, J., Czech, B., and Hannon, G.J. (2015). Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science* **350**, 339–342.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE                                            | SOURCE                                             | IDENTIFIER                    |
|----------------------------------------------------------------|----------------------------------------------------|-------------------------------|
| <b>Antibodies</b>                                              |                                                    |                               |
| ANTI-FLAG® M2 antibody (produced in mouse)                     | Sigma-Aldrich                                      | Cat#F1804                     |
| IRDye® 800CW Goat anti-Mouse IgG secondary antibody            | LI-COR                                             | Cat#926-32210                 |
| StrepMAB-Classic (murine Strep-tag® II specific monoclonal)    | IBA                                                | Cat#2-1507-001                |
| <b>Bacterial and virus strains</b>                             |                                                    |                               |
| BL21-CodonPlus (DE3)-RIPL ( <i>E. coli</i> )                   | Agilent                                            | Cat#230280                    |
| DH10MultiBac ( <i>E. coli</i> )                                | Geneva Biotech                                     | Cat#DH10MultiBac              |
| MAX Efficiency DH5 $\alpha$ competent cells ( <i>E. coli</i> ) | Invitrogen                                         | Cat#18258012                  |
| <b>Chemicals, peptides, and recombinant proteins</b>           |                                                    |                               |
| Ammonium chloride (15N, 99%)                                   | Cambridge Isotope Laboratories, Inc.               | Cat#NLM-467-10                |
| Antibiotic-antimycotic                                         | GIBCO / ThermoFisher Scientific                    | Cat#15-240-062                |
| D-desthiobiotin                                                | Sigma-Aldrich / MilliporeSigma                     | Cat#D1411                     |
| D-glucose, (U-13C6, 99%)                                       | Cambridge Isotope Laboratories, Inc.               | Cat#CLM-1396-1                |
| Deuterium oxide (D2O, 99%)                                     | Cambridge Isotope Laboratories, Inc.               | Cat#DLM-4-100                 |
| DNase I                                                        | Zymo Research                                      | Cat#E1010                     |
| ExoSAP-IT PCR product cleanup reagent                          | Applied Biosystems / ThermoFisher Scientific       | Cat#78200.200.UL              |
| FastAP thermosensitive alkaline phosphatase                    | Thermo Scientific / ThermoFisher Scientific        | Cat#EF0654                    |
| Fetal Bovine Serum (FBS), Heat Inactivated                     | Seradigm / VWR or GIBCO / ThermoFischer Scientific | Cat#97068-091 or Cat#16140063 |
| Fly Extract                                                    | Drosophila Genomics Resource Center                | Stock#1645670                 |
| GlutaMAX supplement                                            | GIBCO / ThermoFisher Scientific                    | Cat#35050061                  |
| HyClone Insect Cell Culture Media (CCM3)                       | Cytiva / VWR                                       | Cat#16777-272                 |
| Insulin solution (human)                                       | Sigma-Aldrich / MilliporeSigma                     | Cat#I9278                     |
| IPTG                                                           | Gold Biotechnology                                 | Cat#I2481C50                  |
| Iron Supplemented Bovine Calf Serum (BCS)                      | Seradigm / VWR                                     | Cat#10158-358                 |
| L-glutathione (reduced)                                        | Sigma-Aldrich / MilliporeSigma                     | Cat#G6013                     |
| Minimum Essential Medium $\alpha$ , nucleosides                | GIBCO / ThermoFisher Scientific                    | Cat#12-571-063                |
| Proteinase K                                                   | NEB                                                | Cat#P8107S                    |
| RNase A                                                        | Ambion / Invitrogen / ThermoFisher Scientific      | Cat#AM2271                    |
| RNase I, cloned                                                | Ambion / Invitrogen / ThermoFisher Scientific      | Cat#AM2294                    |
| RNase inhibitor, murine                                        | NEB                                                | Cat#M0314L                    |
| Shields and Sang M3 Insect Medium                              | Sigma-Aldrich / MilliporeSigma                     | Cat#S8398                     |
| T4 polynucleotide kinase (PNK)                                 | NEB                                                | Cat#M0201S                    |
| T4 RNA ligase 1 (ssRNA Ligase)                                 | NEB                                                | Cat#M0437M                    |
| TEV protease                                                   | Produced in-house                                  | N/A                           |
| TURBO DNase                                                    | Invitrogen / ThermoFisher Scientific               | Cat#AM2238                    |
| X-tremeGENE HP DNA transfection reagent                        | Roche / MilliporeSigma                             | Cat#6366236001                |
| Xfect transfection reagent                                     | Takara                                             | Cat#631317                    |

(Continued on next page)

**Continued**

| REAGENT or RESOURCE                                                                                                  | SOURCE                                       | IDENTIFIER                                                                                                                            |
|----------------------------------------------------------------------------------------------------------------------|----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| <b>Critical commercial assays</b>                                                                                    |                                              |                                                                                                                                       |
| MycAlert Mycoplasma Detection kit                                                                                    | Lonza                                        | Cat#LT07-118                                                                                                                          |
| Nucleospin gel and PCR cleanup                                                                                       | Takara                                       | Cat#740971.50                                                                                                                         |
| PowerUp SYBR Green master mix                                                                                        | Applied Biosystems / ThermoFisher Scientific | Cat#A25776                                                                                                                            |
| PrimeScript RT-PCR kit                                                                                               | Takara                                       | Cat#RR014B                                                                                                                            |
| RNA Clean & Concentrator-5                                                                                           | Zymo Research                                | Cat#R1013                                                                                                                             |
| SMARTer® smRNA-Seq kit for Illumina®                                                                                 | Takara                                       | Cat#635030                                                                                                                            |
| <b>Deposited data</b>                                                                                                |                                              |                                                                                                                                       |
| Annotations: Custom composite annotations                                                                            | This paper                                   | <a href="https://github.com/jonipsaro/asterix_gtsf1">https://github.com/jonipsaro/asterix_gtsf1</a>                                   |
| Annotations: miRNAs                                                                                                  | miRBase release 22.1                         | <a href="http://www.mirbase.org/">http://www.mirbase.org/</a>                                                                         |
| Annotations: piRNA clusters                                                                                          | piRNA Cluster Database                       | <a href="https://www.smallnagroup.uni-mainz.de/piCdb/">https://www.smallnagroup.uni-mainz.de/piCdb/</a>                               |
| Annotations: rRNAs                                                                                                   | UCSC Genome Browser                          | <a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a>                                                                         |
| Annotations: transposable elements                                                                                   | TEtranscripts                                | <a href="http://hammelllab.labsites.cshl.edu/software/#TEtranscripts">http://hammelllab.labsites.cshl.edu/software/#TEtranscripts</a> |
| Annotations: tRNAs                                                                                                   | Genomic tRNA database release 17             | <a href="http://gtmadb.ucsc.edu/GtRNADB_archives/release17/">http://gtmadb.ucsc.edu/GtRNADB_archives/release17/</a>                   |
| Genome: Drosophila reference genome and annotations, dm6                                                             | Flybase, release 6.27                        | <a href="https://flybase.org/">https://flybase.org/</a>                                                                               |
| Genome: Mouse reference genome and annotations version M21, GRCm38                                                   | GENCODE                                      | <a href="https://www.gencodegenes.org/">https://www.gencodegenes.org/</a>                                                             |
| Scripts: Custom processing scripts                                                                                   | This paper                                   | <a href="https://github.com/jonipsaro/asterix_gtsf1">https://github.com/jonipsaro/asterix_gtsf1</a>                                   |
| Sequencing: eCLIP of Drosophila Asterix/Gtsf1 (endogenous promoter) in Drosophila ovarian somatic sheath (OSS) cells | This paper                                   | GEO: GSE151109                                                                                                                        |
| Sequencing: eCLIP of Drosophila Asterix/Gtsf1 transfected in Drosophila ovarian somatic sheath (OSS) cells           | This paper                                   | GEO: GSE151107                                                                                                                        |
| Sequencing: eCLIP of mouse Asterix/Gtsf1 transfected in mouse P19 embryonal teratoma cells                           | This paper                                   | GEO: GSE151108                                                                                                                        |
| Sequencing: RNA-sequencing data of homozygous and heterozygous Asterix knock-out D. melanogaster                     | <a href="#">Muerdter et al. (2013)</a>       | GEO: GSE46009                                                                                                                         |
| Sequencing: Small RNA sequencing of Gtsf1-bound RNAs from Sf9                                                        | This paper                                   | GEO: GSE151110                                                                                                                        |
| Structure: Asterix/Gtsf1 from mouse (full-length protein) bound to co-purifying tRNA                                 | This paper                                   | EMDB: EMD-22040                                                                                                                       |
| Structure: Asterix/Gtsf1 from mouse (residues 1-45; zinc finger 1) bound to co-purifying tRNA                        | This paper                                   | EMDB: EMD-22041                                                                                                                       |
| Structure: NMR solution structure of Asterix/Gtsf1 from mouse (CHHC zinc finger domains)                             | This paper                                   | PDB: 6X46; BMRB: 30754                                                                                                                |
| Structure: U11-48K CHHC Zn-Finger Domain                                                                             | <a href="#">Tidow et al. (2009)</a>          | PDB: 2VY4                                                                                                                             |
| <b>Experimental models: Cell lines</b>                                                                               |                                              |                                                                                                                                       |
| <i>Drosophila melanogaster</i> : ovarian somatic sheath cells (OSS)                                                  | Drosophila Genomics Resource Center          | RRID: CVCL_1B46                                                                                                                       |
| <i>Mus musculus</i> : embryonal teratocarcinoma (P19)                                                                | ATCC                                         | ATCC: CRL-1825; RRID: CVCL_2153                                                                                                       |
| <i>Spodoptera frugiperda</i> : pupal ovarian cells (Sf9)                                                             | GIBCO / ThermoFisher Scientific              | Cat#11496015; RRID: CVCL_0549                                                                                                         |

(Continued on next page)

| REAGENT or RESOURCE                                                                                         | SOURCE                              | IDENTIFIER                                                                                                                                                                                                                      |
|-------------------------------------------------------------------------------------------------------------|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Continued</b>                                                                                            |                                     |                                                                                                                                                                                                                                 |
| <b>Oligonucleotides</b>                                                                                     |                                     |                                                                                                                                                                                                                                 |
| eCLIP oligos 1a: RNA linker ligation RNA_X1A:<br>AUUAGGNNNNNAGAUCCGGAAGAGCGUCGUGUAG                         | Van Nostrand et al. (2016)          | N/A                                                                                                                                                                                                                             |
| eCLIP oligos 1b: RNA linker ligation RNA_X1B:<br>AAUAGCANNNNNAGAUCCGGAAGAGCGUCGUGUAG                        | Van Nostrand et al. (2016)          | N/A                                                                                                                                                                                                                             |
| eCLIP oligos 1c: RNA linker ligation RNA_RiL19:<br>AGAUCGGAAGAGCGUCGUG                                      | Van Nostrand et al. (2016)          | N/A                                                                                                                                                                                                                             |
| eCLIP oligo 2: reverse transcription primer DNA_AR17:<br>ACACGACGCTCTTCCGA                                  | Van Nostrand et al. (2016)          | N/A                                                                                                                                                                                                                             |
| eCLIP oligo 3: DNA linker ligation DNA_rand103Tr3:<br>NNNNNNNNNAGATCGGAAGAGCACACGTCTG                       | Van Nostrand et al. (2016)          | N/A                                                                                                                                                                                                                             |
| <b>Recombinant DNA</b>                                                                                      |                                     |                                                                                                                                                                                                                                 |
| DmGtsf1, codon-optimized for expression in <i>Sf9</i> ;<br>pAWG backbone                                    | This paper                          | Synthesized in-house                                                                                                                                                                                                            |
| DmGtsf1, codon-optimized for expression in <i>Sf9</i> ;<br>pFL backbone                                     | This paper                          | Synthesized in-house                                                                                                                                                                                                            |
| MmGtsf1 for expression in P19 cells (wildtype<br>codon usage); pEF backbone                                 | This paper                          | Subcloned from GenScript Accession#<br>NM_028797.1; Clone ID: OMu06141D                                                                                                                                                         |
| MmGtsf1, codon-optimized for expression in <i>Sf9</i> ,<br>various constructs and truncations; pFL backbone | This paper                          | Synthesized in-house                                                                                                                                                                                                            |
| pAGW transfection control                                                                                   | Drosophila Genomics Resource Center | Stock Number: 1071                                                                                                                                                                                                              |
| pmaxGFP transfection control                                                                                | Lonza                               | Based on Cat#VDC-1040                                                                                                                                                                                                           |
| <b>Software and algorithms</b>                                                                              |                                     |                                                                                                                                                                                                                                 |
| ARIA version 2.3                                                                                            | Linge et al. (2003)                 | <a href="http://aria.pasteur.fr/">http://aria.pasteur.fr/</a>                                                                                                                                                                   |
| bedtools version 2.29.0                                                                                     | Quinlan (2014)                      | <a href="https://bedtools.readthedocs.io/en/latest">https://bedtools.readthedocs.io/en/latest</a>                                                                                                                               |
| cisTEM version 1.0.0                                                                                        | Grant et al. (2018)                 | <a href="https://cistem.org/">https://cistem.org/</a>                                                                                                                                                                           |
| Crystallography and NMR System version 1.2                                                                  | Brünger et al. (1998)               | <a href="https://www.mrc-lmb.cam.ac.uk/public/xtal/doc/cns/cns_1.3/installation/frame.html">https://www.mrc-lmb.cam.ac.uk/public/xtal/doc/cns/cns_1.3/installation/frame.html</a>                                               |
| DALI                                                                                                        | Holm (2019)                         | <a href="http://ekhidna2.biocenter.helsinki.fi/dali/">http://ekhidna2.biocenter.helsinki.fi/dali/</a>                                                                                                                           |
| DESeq2 version 1.24.0                                                                                       | Love et al. (2014)                  | <a href="http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html">http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html</a>                                                                             |
| FLASH version 1.2.11                                                                                        | Magoč and Salzberg (2011)           | <a href="http://www.cbcb.umd.edu/software/flash">http://www.cbcb.umd.edu/software/flash</a>                                                                                                                                     |
| Gremlin                                                                                                     | Ovchinnikov et al. (2014)           | <a href="http://gremlin.bakerlab.org/">http://gremlin.bakerlab.org/</a>                                                                                                                                                         |
| Illumina bcl2fastq2 version 2.19                                                                            | Illumina                            | <a href="https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software/downloads.html">https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software/downloads.html</a> |
| MolProbity                                                                                                  | Davis et al. (2007)                 | <a href="http://molprobity.biochem.duke.edu/">http://molprobity.biochem.duke.edu/</a>                                                                                                                                           |
| NMRFAM-SPARKY                                                                                               | Lee et al. (2015)                   | <a href="http://pine.nmrfam.wisc.edu/download_packages.html">http://pine.nmrfam.wisc.edu/download_packages.html</a>                                                                                                             |
| NMRPipe                                                                                                     | Delaglio et al. (1995)              | <a href="https://www.ibbr.umd.edu/nmrpipe/install.html">https://www.ibbr.umd.edu/nmrpipe/install.html</a>                                                                                                                       |
| PyMOL version 2.0                                                                                           | SchrödingerLLC, 2019                | <a href="https://www.schrodinger.com/products/pymol">https://www.schrodinger.com/products/pymol</a>                                                                                                                             |
| RNA Composer version 1.0                                                                                    | Antczak et al. (2016)               | <a href="http://rnacomposer.cs.put.poznan.pl/">http://rnacomposer.cs.put.poznan.pl/</a>                                                                                                                                         |
| SPIDER version UNIX 24.01                                                                                   | Shaikh et al. (2008)                | <a href="https://spider.wadsworth.org/spider_doc/spider/docs/spider.html">https://spider.wadsworth.org/spider_doc/spider/docs/spider.html</a>                                                                                   |

(Continued on next page)

**Continued**

| REAGENT or RESOURCE                          | SOURCE                                    | IDENTIFIER                                                                                        |
|----------------------------------------------|-------------------------------------------|---------------------------------------------------------------------------------------------------|
| STAR version 2.5.2b                          | Dobin et al. (2013)                       | <a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>                 |
| TALOS                                        | Shen et al. (2009)                        | <a href="https://spin.niddk.nih.gov/NMRPipe/talos/">https://spin.niddk.nih.gov/NMRPipe/talos/</a> |
| WARP version 1.0.6                           | Tegunov and Cramer (2019)                 | <a href="http://www.warpem.com/warp/">http://www.warpem.com/warp/</a>                             |
| <b>Other</b>                                 |                                           |                                                                                                   |
| Agencourt AMPure XP beads                    | Beckman Coulter / ThermoFisher Scientific | Cat#A63880 / Cat#NC9959336                                                                        |
| Anti-FLAG M2 Magnetic Beads                  | Sigma-Aldrich                             | Cat#M8823-1ML                                                                                     |
| Blue Pippin 2% agarose gel cassette          | Sage Science                              | Cat#BDF2010                                                                                       |
| Dynabeads MyOne Silane                       | Invitrogen / ThermoFisher Scientific      | Cat#37002D                                                                                        |
| Lacey carbon grids                           | Electron Microscopy Sciences              | Cat#LC325-Cu                                                                                      |
| MagStrep "type3" XT beads                    | IBA                                       | Cat#2-4090-002                                                                                    |
| Mono Q 5/50 GL                               | Cytiva (formerly GE Healthcare)           | Cat#17516601                                                                                      |
| Ni-NTA Agarose                               | QIAGEN                                    | Cat#30250                                                                                         |
| Phenol:Chloroform:IAA, 25:24:1, pH 6.6       | Invitrogen / ThermoFisher Scientific      | Cat#AM9732                                                                                        |
| Strep-Tactin® Superflow® high capacity resin | IBA                                       | Cat#2-1208-025                                                                                    |
| Superdex 75 Increase 10/300 GL               | Cytiva (formerly GE Healthcare)           | Cat#29148721                                                                                      |

**RESOURCE AVAILABILITY**

**Lead contact**

Requests for resources, reagents, or further information should be directed to and will be fulfilled by Leemor Joshua-Tor ([leemor@cshl.edu](mailto:leemor@cshl.edu)).

**Materials availability**

Plasmids generated in this study are available upon request.

**Data and code availability**

Coordinates and NMR data have been deposited in the Protein Data Bank (PDB: 6X46) and the Biological Magnetic Resonance Bank (BMRB: 30754). Cryo-electron microscopy maps for complexes isolated from full-length MmGtsf1 protein and ZnF1 domain pull-downs have been deposited in the EMD (EMD-22040 and EMD-22041, respectively). Sequencing data have been deposited in Gene Expression Omnibus (GEO) repository with accession numbers GSE151110 (Sf9 RNA pull-down), GSE151108 (eCLIP data from P19 cells), GSE151107 (eCLIP data from OSS cells), GSE151109 (eCLIP data from OSS cells using CRISPR-tagged Asterix). Custom gene annotation files and data processing scripts are available on GitHub ([https://github.com/jonipsaro/asterix\\_gtsf1](https://github.com/jonipsaro/asterix_gtsf1)). Intermediate files used for generating gene annotations or processing of the data are available upon request.

**EXPERIMENTAL MODELS AND SUBJECT DETAILS**

**Sf9 cell culture**

Sf9 (*Spodoptera frugiperda* pupal ovarian; RRID: CVCL\_0549; female) cells were maintained in CCM3 medium (Cytiva). Cells were cultured at 27°C ambient atmosphere with orbital at 115 rpm. Cultures were monitored for *Mycoplasma* contamination using the MycoAlert Mycoplasma Detection Kit (Lonza). *Mycoplasma* contamination was not detected.

**P19 cell culture**

P19 (mouse embryonal carcinoma; ATCC: CRL-1825; RRID: CVCL\_2153; male) cells were maintained in minimum essential medium with ribonucleosides and deoxyribonucleosides (GIBCO), supplemented with bovine calf serum and fetal bovine serum (7.5% and 2.5% final concentration, respectively) (Seradigm). Cells were cultured at 37°C in a 5% CO<sub>2</sub> atmosphere. Cultures were monitored for *Mycoplasma* contamination using the MycoAlert Mycoplasma Detection Kit (Lonza). *Mycoplasma* contamination was not detected. The identity of the cultured cells was confirmed by short tandem repeat (STR) profiling, serviced by ATCC.

**OSS cell culture**

*Drosophila* OSS (ovarian somatic sheath; RRID: CVCL\_1B46; female) cells were maintained in OSS medium (Shields and Sang M3 Insect Medium [Sigma-Aldrich] supplemented with approximately 5 mM potassium glutamate, 5 mM potassium bicarbonate, 10% heat-inactivated fetal bovine serum [Seradigm], 10% fly extract [*Drosophila* Genomics Resource Center], 2 mM reduced glutathione

[Sigma-Aldrich], 1x GlutaMAX [GIBCO], 0.01 mg/mL human insulin [Sigma-Aldrich], and an antibiotic-antimycotic [GIBCO] consisting of penicillin, streptomycin, and Amphotericin B). Cells were cultured at ~23°C. Cultures were monitored for *Mycoplasma* contamination using the MycoAlert Mycoplasma Detection Kit (Lonza). *Mycoplasma* contamination was not detected. OSS cells with Asterix C-terminally FLAG-tagged at its native locus (Asterix-GFP-FRT-Precision-V5-FLAG<sub>3</sub>-P2A) were provided by the lab of J. Brennecke and cultured in the same way as unmodified OSS cells.

## METHOD DETAILS

### Cloning

#### Overview

In order to screen for well-behaved targets for recombinant protein expression, a panel of constructs was generated from *H. sapiens*, *M. musculus*, and *D. melanogaster* Gtsf1 cDNAs (codon-optimized for expression in Sf9) by SLIC (sequence- and ligation-independent cloning) in DH5 $\alpha$  cells (Invitrogen). These constructs presented various N- or C-terminal tags for enhanced expression and purification using either *E. coli* or insect cell culture systems. In addition, natural sequences of the *D. melanogaster* and *M. musculus* proteins were used for transfection in eCLIP experiments. The sequence of each construct was verified by GenScript. Constructs presented in this work are described in further detail below and summarized at the end of this section.

#### Constructs for structure determination by NMR

To obtain sufficient quantities of isotopically-labeled, purified protein, numerous MmGtsf1 constructs were screened for high expression in *E. coli*. A fragment corresponding to the first 115 residues of MmGtsf1 with a C-terminal TEV-His<sub>6</sub>-tag showed highest expression and produced sufficiently soluble material for structure determination by NMR. To prevent aggregation over the duration of NMR data collection, four of the cysteines (those not involved in zinc chelation) were mutated to serine. These constructs were cloned into the vector pET-22 and also included TEV-cleavable linker for His<sub>6</sub>-tag removal (MmGtsf1-115-TEV-His and MmGtsf1-115-TEV-His C28S, C76S, C100S, C103S).

#### Constructs for RNA binding studies

Constructs were similarly screened for expression in Sf9 cells. Data presented for RNA interaction studies include the full-length protein (167 residues), point mutants, and truncations as indicated in each figure. All Sf9-derived material included a C-terminal Strep<sub>2</sub>-tag and TEV-cleavable linker and was cloned into the vector pFL for baculoviral-induced insect cell culture.

#### Constructs for eCLIP

MmGtsf1 cDNA (not codon-optimized) was obtained from Genscript (Accession Number NM\_028797.1; Clone ID: OMu06141D) and subcloned by SLIC into the vector pEF $\alpha$  under the control of the EF1 $\alpha$  promoter. A C-terminal TEV-Strep<sub>2</sub> tag was included to allow for affinity pull-down during eCLIP processing. pmaxGFP (Lonza) was used as a transfection control. pEF $\alpha$  plasmid was a kind gift from A. Schorn and R. Martienssen. *Drosophila* Asterix cDNA (cDNA codon-optimized for expression in Sf9) was synthesized in-house and cloned into the vector pAWG (*Drosophila* Genome Resource Center) under the control of the actin promoter. A C-terminal TEV-Strep<sub>2</sub> tag was included to allow for affinity pull-down during eCLIP processing. pAGW (*Drosophila* Genome Resource Center) was used as a transfection control.

### Expression and purification

#### Recombinant expression in *E. coli*

To generate isotopically-labeled, purified protein, target constructs were transformed into BL21-CodonPlus (DE3)-RIPL (Agilent). Cultures were then grown in M9 media supplemented with <sup>15</sup>NH<sub>4</sub>Cl and/or <sup>13</sup>C-labeled glucose (Cambridge Isotopes) at 37°C to a culture density of approximately 0.7. Protein expression was induced with 1 mM IPTG (final concentration) and proceeded for 3.5 hours.

#### Purification

Cells were harvested by centrifugation at 4000g, resuspended in lysis buffer (50 mM sodium phosphate, pH 8.0, 50 mM NaCl, 10 mM imidazole; ~20 mL per liter culture), and lysed by sonication. The cell lysate was clarified by ultracentrifugation at 125,000g for 1 h after which the supernatant applied to a Ni-NTA column equilibrated with lysis buffer. The column was washed (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 40 mM imidazole) and the protein then eluted (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 200 mM imidazole). To prevent precipitation and proteolysis, DTT was added to the elution at a final concentration of 10 mM and EDTA at a final concentration of 1 mM. The C-terminal His<sub>6</sub>-tag was then removed by overnight treatment with TEV protease (1:25 mass ratio of protease:target) at 4°C. The cleaved protein was further purified by ion-exchange chromatography (MonoQ column) in a buffer of 25 mM Tris, pH 8.0, and 2 mM DTT with a NaCl gradient from 0 to 1 M. MmGtsf1-115 eluted approximately between 17 and 24 mS. Peak fractions were pooled, concentrated, and used for further purification by gel filtration chromatography (Superdex75 increase) in 50 mM MES, pH 6.5, 200 mM NaCl, and 5 mM TCEP. Peak fractions were pooled, concentrated and mixed with ZnCl<sub>2</sub> (2:1 molar ratio Zn<sup>2+</sup>:protein) and MgCl<sub>2</sub> (4:1 molar ratio Mg<sup>2+</sup>:protein). Upon addition of ZnCl<sub>2</sub>, the protein solution became temporarily turbid, but clarified upon gentle mixing. For NMR structure determination, sodium azide was added at a final concentration of 0.02% as a preservative. Typical yields were 2-3 mg of purified protein (> 98% pure as assessed by SDS-PAGE) per liter culture.



### Expression in Sf9 and RNA pull-down

Constructs (each with a C-terminal TEV-Strep<sub>2</sub> tag) were cloned into the vector pFL then integrated into bacmids using DH10MultiBac cells (Geneva Biotech). Isolated bacmids were then transfected into Sf9 cells for baculoviral-driven expression. For details regarding growth and maintenance of Sf9, refer to the Experimental Models and Subject Details. After expression, cells were harvested by centrifugation at 1000g, resuspended in lysis buffer (50 mM Tris, pH 8.0, 100 mM KCl, 1 mM DTT) (~20 mL per liter culture), and lysed by sonication. The cell lysate was then clarified by ultracentrifugation at 125,000g for 1 h and the supernatant applied to a Strep-Tactin (IBA) column equilibrated with lysis buffer. The bound MmGtsf1 proteins were subsequently washed with lysis buffer, further washed with lysis buffer containing 2 mM ATP, and finally eluted in lysis buffer containing 5 mM D-desthiobiotin. Protein purity was assessed by SDS-PAGE. Co-purifying nucleic acids were isolated by phenol:chloroform extraction, precipitated with ethanol, then assessed by Urea-PAGE.

### Characterization of co-purifying Sf9 RNAs

#### Initial nucleic acid characterization

After phenol:chloroform extraction and alcohol precipitation, pulled-down nucleic acids were characterized by treatment with RNase, DNase, or by alkaline hydrolysis. For each treatment, approximately 50 ng of nucleic acid was mixed with either RNase A (Ambion; 1 μg), DNase I (Zymo Research; 0.1 units), or 1 μL of 1 M sodium hydroxide in total volume of 40 μL under suitable buffer conditions (10 mM Tris, pH 8.0, 1 mM EDTA for RNase A treatment; no added buffer for alkaline hydrolysis treatment; 10 mM Tris, pH 7.6, 2.5 mM MgCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub> for DNase I treatment). Murine RNase inhibitor (NEB; 40 units) was included in all conditions with the exception of the RNase A treatment. Samples were incubated for 15 minutes at 37°C for nuclease treatments or 70°C for alkaline hydrolysis. After treatment, the sodium hydroxide was neutralized by the addition of 1 μL of 1 M hydrochloric acid. As a control, a 50 nucleotide DNA duplex was treated under the same set of conditions. All samples were then denatured and assessed by 12% Urea-PAGE.

#### sRNA library preparation

Affinity co-purifying nucleic acids which bound to MmGtsf1 during expression in Sf9 were separated from the protein by ion exchange chromatography (Mono Q column, as described above, eluting between 45 and 55 mS). Peak fractions were pooled, and the RNA isolated by phenol:chloroform extraction and alcohol precipitation. Small RNA libraries were prepared using the SMARTer smRNA-Seq Kit for Illumina sequencing (Takara). Size-selection was performed using Blue Pippin 2% agarose gel cassettes (Sage Science). All libraries were assessed by fluorometric quantification (Qubit 3.0) and by Bioanalyzer chip-based capillary electrophoresis. The average fragment size was 228 bp with most insert sizes ranging from 20–100 bp. Libraries were pooled in equimolar ratios according to their quantification (determined above). Single-end reads with two 8-basepair barcodes were generated on an Illumina NextSeq resulting in approximately 10 million reads per library. Base calling was performed with Illumina bcl2fastq2 v2.19 software.

#### sRNA library data processing

Owing to the incomplete assembly of the Sf9 genome and the lack of annotations, processing for sRNA was straightforward, but limited. Reads were first trimmed to remove sequences appended during library preparation (adapters, polyA sequences at the 3' end, as well as the first three nucleotides after the adaptor at the 5' end). Removal of the polyA sequence was performed using a custom script (*polyA\_trim.py*). Reads were then filtered based on size and quality scores. Reads in the processed libraries were collapsed and the most abundant sequences were manually inspected.

## eCLIP LIBRARY GENERATION

### Cell culture

For details regarding growth and maintenance of P19 (mouse embryonal carcinoma) cells and OSS (*Drosophila* ovarian somatic sheath) cells, refer to the [Experimental models and subject details](#).

### P19 cell transfection of MmGtsf1-TEV-Strep

P19 cells were grown to 75% confluency in 150 mm culture dishes. Four hours prior to transfection, the medium was refreshed. To transfect, 30 μg of DNA (either MmGtsf1-TEV-Strep in pEF or eGFP in pMAX [transfection control]) was premixed with 60 μL of X-tremeGENE HP DNA transfection reagent (Roche) in serum-free medium for 15 minutes. After a 15-minute incubation, this mixture was added to the cultures. Sixteen hours post-transfection, the cells were visibly perturbed and the medium was again refreshed. Expression of eGFP in the transfection control was confirmed by UV microscopy. Forty-eight hours post-transfection, the cells were rinsed with ice-cold phosphate-buffered saline (PBS) and taken for processing.

### OSS cell transfection of Asterix-TEV-Strep

OSS cells were grown to 75% confluency in 150 mm culture dishes. Four hours prior to transfection, the medium was refreshed. To transfect, 50 μg of DNA (either Asterix-TEV-Strep in pAWG or pAGW [transfection control]) was premixed with 15 μL of Xfect Polymer transfection reagent (Takara) in 500 μL Xfect buffer. OSS medium was removed from the cells and replaced with Shields and Sang M3 Insect Medium supplemented only with potassium bicarbonate and potassium glutamate. After a 10-minute incubation of the DNA with the transfection reagent, the transfection mixture was added to the cultures. Two hours post-transfection, the M3 medium was removed and replaced with fully-supplemented OSS medium. Expression of GFP in the transfection control was confirmed by UV

microscopy. Seventy-two hours post-transfection, the cells were rinsed with ice-cold phosphate-buffered saline (PBS) and taken for processing.

### Library preparation

eCLIP Libraries were prepared essentially as in [Van Nostrand et al. \(2016\)](#) with the following parameters and modifications. UV cross-linking was performed at 254 nm for ~45 s (400 mJ) in an HL-2000 Hybrilinker. For MmGtsf1-TEV-Strep in P19 cells and Asterix-TEV-Strep in OSS cells, protein pull-down was accomplished using MagStrep “type 3” XT beads (IBA) with 50  $\mu$ L of bead resuspension used per sample. Asterix-GFP-FRT-Precision-V5-FLAG<sub>3</sub>-P2A, pull-down was similarly accomplished with Anti-FLAG M2 magnetic beads (Sigma-Aldrich). The suppliers of molecular reagents used in the eCLIP procedure (ExoSAP-IT, FastAP, Proteinase K, RNase I, RNase inhibitor, T4 PNK, T4 RNA ligase, TURBO DNase), commercial kits (Nucleospin cleanup kit, PrimeScript RT-PCR kit, RNA Clean & Concentrator-5 kit, and SYBR Green master mix), and antibodies used in western blotting (mouse ANTI-FLAG M2 primary, mouse StrepMAB-Classical primary, and goat anti-Mouse IgG IRDYE 800CW secondary) are detailed in the Key Resources Table. Library adaptor oligonucleotide sequences are also provided.

Final libraries were amplified and barcoded using Illumina compatible primers as described below.

Non-crosslinked input D504, D701  
 Crosslinked input (replicate 1) D504, D702  
 Crosslinked input (replicate 2) D501, D703  
 Non-crosslinked IP (background) D503, D701  
 Crosslinked IP (replicate 1) D502, D703  
 Crosslinked IP (replicate 2) D503, D704

For samples from mouse P19 cells, 8 amplification cycles were used for the inputs and 14 cycles for the IPs. For samples from *Drosophila* OSS cells, 13 amplification cycles were used for the inputs and 18 cycles for the IPs.

For quality control, all libraries were assessed by fluorometric quantification (Qubit 4.0) and by Bioanalyzer chip-based capillary electrophoresis. The average fragment size was typically 240–250 bp with most insert sizes ranging from 15–200 bp. A detailed version of the complete eCLIP library preparation is available upon request.

### Next-generation sequencing

Libraries were pooled in equimolar ratios according to their quantification (determined above). Paired-end reads with two, 8-basepair barcodes were generated on an Illumina NextSeq resulting in approximately 100 million paired-end reads (~15–20 million reads per library). Base calling was performed with Illumina bcl2fastq2 v2.19 software.

## eCLIP PROCESSING

### Rationale

Based on our previous findings when sequencing endogenous *Sf9* RNAs copurifying with recombinantly-expressed MmGtsf1, we surmised that it would be necessary to include multi-mapping reads in our analysis pipeline. This stems from the fact that many of the RNA species of interest arise from known multi-mapping regions (tRNAs, transposable elements, and piRNA clusters).

### Summary

The pipeline begins with demultiplexed paired-end libraries. Given that most all of the paired-end reads were short enough to overlap, they were joined into single sequences using FLASH ([Magoč and Salzberg, 2011](#)). Sequencing adapters were then trimmed, PCR duplicates removed, and the reverse complement of the read (corresponding to the sense strand of the original RNA) was taken for downstream processing. Identical reads were collapsed and counted, then mapped to the genome using STAR ([Dobin et al., 2013](#)). The aligned reads were then annotated and filtered based on feature type using a combination of custom scripts and bedtools ([Quinlan, 2014](#)). Full descriptions of custom scripts accompany the deposited code (see [Resource availability](#)).

### Gene annotations

As many of the gene classes of interest have dedicated communities of their own (tRNAs, miRNAs, piRNAs, and transposons), we incorporated these multiple annotation sources into the workflow. The sources of annotations are listed below for both the mouse and *Drosophila* analyses. In brief, annotations from each source were compared, matched when possible, and if matched the outer bounds of each annotation were taken. The resulting composite annotations have been deposited (see [Resource availability](#)).

Mouse: Gencode version M21 ([Frankish et al., 2019](#)), miRBase release 22.1 ([Kozomara et al., 2019](#)), piRNA cluster DB ([Rosenkranz, 2016](#)), Tetranscripts ([Jin et al., 2015](#)), tRNA DB ([Chan and Lowe, 2016](#)), UCSC rRNA annotations ([Kent et al., 2002](#))  
 Fly: FlyBase ([Thurmond et al., 2019](#)), miRBase release 22.1 ([Kozomara et al., 2019](#)), piRNA cluster DB ([Rosenkranz, 2016](#)), custom annotations provided by A. Haase for piRNAs, Tetranscripts ([Jin et al., 2015](#)), tRNA DB, UCSC ([Kent et al., 2002](#))

### tRNA analysis

Following multi-mapping normalization, reads belonging to the tRNA annotation class were further characterized. To begin, the size of each tRNA annotation was scaled to a “model tRNA” size of 73 nucleotides. Each tRNA read was then re-mapped to its annotation, now scaled to the model tRNA length. By aggregating all tRNA-mapping reads, we were able to generate histograms of read statistics (5' end, 3' end, read length, and nucleotides covered). It is expected that eCLIP reads will have a pileup at their 5' end corresponding to the cross-linking site. We scored this pileup by determining the fold enrichment for each metric (essentially calculated as [IP – background] / input) and weighting it by its Z-score.

### NMR spectroscopy

#### Instrumentation

NMR spectroscopy was performed using Bruker AVANCE500 (New York Structural Biology Center, NYSBC), DRX600 (Columbia University), AVANCE700 (NYSBC), AVANCE800 (NYSBC), and AVANCE900 (NYSBC) NMR spectrometers equipped with 5 mm cryoprobes.

#### Sample preparation

MmGtsf1 samples were prepared in 50 mM MES, pH 6.5, 200 mM NaCl, 5 mM TCEP, and 2:1 stoichiometric ZnCl<sub>2</sub>, 4:1 stoichiometric MgCl<sub>2</sub>, and 0.02% azide. For data acquisition, samples were either supplemented with a final concentration of 10% D<sub>2</sub>O or lyophilized and resuspended in 99% D<sub>2</sub>O. Sample concentrations were 0.5 mM for the [U-<sup>15</sup>N]-labeled protein and 0.8 mM for the [U-<sup>13</sup>C, U-<sup>15</sup>N]-labeled protein. The sample temperature was calibrated to 298 K using 98% <sup>2</sup>H<sub>4</sub>-methanol (Findeisen et al., 2007). 100 μM DSS was included in samples for internal referencing of <sup>1</sup>H chemical shifts, followed by indirect referencing for <sup>13</sup>C and <sup>15</sup>N chemical shifts (Cavanagh et al., 2007).

#### Resonance assignments

Backbone resonance assignments were obtained using <sup>1</sup>H-<sup>15</sup>N HSQC, <sup>1</sup>H-<sup>13</sup>C HSQC, HNCA, HN(CO)CA, HNCO, HN(CA)CO, HNCACB, and HN(COCA)CB experiments (Cavanagh et al., 2007). Side chain resonance assignments were obtained using HCCH-TOCSY, HBHA(CO)NH, H(CCO)NH, and (H)C(CO)NH experiments (Cavanagh et al., 2007). Spectra were processed using NMRPipe (Delaglio et al., 1995) and analyzed using NMRFAM-SPARKY (Lee et al., 2015).

#### Distance restraints

Distance restraints for structure determination were obtained from <sup>1</sup>H-<sup>15</sup>N NOESY-HSQC, <sup>1</sup>H-<sup>13</sup>C NOESY-HSQC, and <sup>1</sup>H-<sup>13</sup>C NOESY-HSQC (with spectral parameters optimized for detection of aromatic spins) (Cavanagh et al., 2007). <sup>1</sup>H-<sup>13</sup>C NOESY experiments were performed for samples prepared in 99% D<sub>2</sub>O.

#### Zinc coordination

Protonation states of histidine residues were determined by long-range HMQC experiments together with the empirical correlation between the chemical shift difference <sup>13</sup>C<sub>ε1</sub> - <sup>13</sup>C<sub>δ2</sub> (Barraud et al., 2012). H23 and H57 are designated with N<sub>δ1</sub> coordination, and H33 and H67 are designated with N<sub>ε2</sub> coordination to the Zn<sup>2+</sup> ion.

#### Relaxation parameter determination

Backbone <sup>15</sup>N R<sub>1</sub> relaxation rate constants, <sup>15</sup>N R<sub>2</sub> relaxation rate constants, and the steady-state {<sup>1</sup>H}-<sup>15</sup>N NOE were measured at 500 MHz (NYSBC) using the pulse sequences of Lakomek et al. (2012). R<sub>1</sub> measurements used relaxation delays of 24 (× 2), 176, 336 (× 2), 496, 656, 816, 976, and 1200 ms. R<sub>2</sub> measurements used relaxation delays of 16.3 (× 2) 32.6, 49.0 (× 2), 65.3, 97.9, 130.6, 163.2, and 195.8 ms. NOE measurements used a recycle delay of 7 s for the control experiment and 2 s of recovery followed by 3 s of saturation for the saturated experiment. Duplicate relaxation delays were used for error estimation for measurement of <sup>15</sup>N R<sub>1</sub> and R<sub>2</sub> relaxation rate constants. Duplicate experiments were used for error estimation for the steady-state {<sup>1</sup>H}-<sup>15</sup>N NOE experiment.

#### Structure determination

Automatic NOESY cross-peak assignments and structure calculations were performed with ARIA 2.3 (Ambiguous Restraints for Iterative Assignment) (Linge et al., 2003) using an eight step iteration scheme supported by partial manual assignments of aliphatic/aromatic <sup>13</sup>C-edited NOESY-HSQC and amide <sup>15</sup>N-edited NOESY-HSQC spectra, respectively. Less than 10% of all assignments were labeled ambiguous after initial and final ARIA structure calculations. The unambiguous distance restraints output from the automation run was recalibrated by increasing all the upper distance limits by ~10% and further elimination of lone and consistent NOE violations by manually inspecting the lower quality peak assignments. Dihedral angle restraints for residues in the structured zinc finger domains were derived from the analysis of the backbone chemical shifts in TALOS (Shen et al., 2009). Structure calculations were performed in two stages by initially excluding Zn<sup>2+</sup> during automated NOESY cross-peak assignments followed by water refinement of the Zn<sup>2+</sup>-bound structures. The tetrahedral Zn<sup>2+</sup> metal ion coordination was implemented in CNS 1.1 by adding a CHHC patch with the experimentally verified tautomeric states for the two histidine side chains in the topallhdg5.3.pro file (Bersch et al., 2013; Tidow et al., 2009). Bond lengths and angles used to define the Zn<sup>2+</sup>-bound CHHC motif in the parallhdg5.3.pro file was obtained from the published structure (PDB 2VY4) of the homologous ribonuclear protein U11-K48 (Tidow et al., 2009). The final ensemble of 20 representative Zn<sup>2+</sup>-bound structures was generated by calculating 500 structures with water refinement in CNS 1.2 (Brünger et al., 1998). Table S1 summarizes the final restraints used in the calculations, NMR ensemble statistics, and the overall quality of the structures determined by MolProbity (Davis et al., 2007).

#### Local variability analysis

A sliding window of ± 3 amino acids was used to align the 20 lowest energy structures to one another in all combinations at each residue. Average RMSDs were calculated for each window's alignment, then mapped onto the central residue in the window. Res-

idues near the termini included as many residues as possible while maintaining up to 3 residues on either side of the queried residue (e.g., the score for residue 1 derived from RMSDs using residues 1-4 for alignment; the score for the final residue, 115 derived from RMSDs using residues 112-115).

### Cryo-electron microscopy

#### Sample preparation

Affinity-purified MmGtsf1-TEV-Strep constructs from *Sf9* (which included co-purified RNAs) at  $\sim 0.25$  mg/mL in elution buffer (50 mM Tris, pH 8.0, 100 mM KCl, 1 mM DTT, 5 mM d-desthiobiotin) were first cross-linked at 254 nm for  $\sim 45$  s (400 mJ) in an HL-2000 Hybrilinker. It should be noted however, that assessment of RNAs by Urea-PAGE following this treatment did not seem to result in significant covalent cross-linking. For cryo-EM grid preparation, 4  $\mu$ L of solution was applied to a glow-discharged Lacey carbon grid, incubated for 10 s at 25°C and 95% humidity, blotted for 2.5 s, then plunged into liquid ethane using an Automatic Plunge Freezer EM GP2 (Leica).

#### Data acquisition

Data were acquired on Titan Krios transmission electron microscope (ThermoFisher) operating at 200 keV. Dose-fractionated movies were collected using a K2 Summit direct electron detector (Gatan) operating in electron counting mode. In total, 32 frames were collected over a 4 s exposure. The exposure rate was 7.6  $e^-/\text{pixel}/\text{second}$  (approximately 19  $e^-/\text{\AA}^2/\text{second}$ ), which resulted in a cumulative exposure of approximately 76  $e^-/\text{\AA}^2$ . EPU data collection software (ThermoFisher) was used to collect micrographs at a nominal magnification of 215,000x (0.6262  $\text{\AA}/\text{pixel}$ ) and defocus range of  $-1.0$  to  $-3.0$   $\mu\text{m}$ . For the full-length protein construct sample (MmGtsf1-TEV-Strep with RNA), 4,849 micrographs were collected. For the construct containing only the first zinc finger (MmGtsf1-[1,45]-TEV-Strep with RNA), 2,461 micrographs were collected.

#### Micrograph processing and 3D reconstruction

Real-time image processing (motion correction, CTF estimation, and particle picking) was performed concurrently with data collection using WARP (Tegunov and Cramer, 2019). Automated particle picking was initiated with the BoxNet pretrained deep convolutional neural network bundle included with WARP that implemented in TensorFlow. Following this first round of particle picking, the particle selections on  $\sim 20$  micrographs were manually inspected and adjusted. This process was iterated one additional time. For the full-length construct, a particle diameter of 100  $\text{\AA}$  and a threshold score of 0.6 yielded 495,299 particle coordinates for the full-length construct. These particles were then subjected to a 2D classification in cisTEM (Grant et al., 2018) after which a subset of 346,643 particles were used for *ab initio* reconstruction and autorefinement in cisTEM. For the truncation construct, a particle diameter of 100  $\text{\AA}$  and a threshold score of 0.5 yielded 159,646 particle coordinates. These were then taken for 2D classification in cisTEM (Grant et al., 2018) after which a subset of 96,036 particles were used for *ab initio* reconstruction and autorefinement. After refinement, structures of tRNA—modeled incorporating the sequence of the most abundantly pulled-down RNA from *Sf9* expression of MmGtsf1-TEV-Strep (Table S2) and generated with RNAComposer (Antczak et al., 2016; Popenda et al., 2012)—and the zinc finger domains of MmGtsf1 were manually placed in the reconstructed volume based on the molecular shapes and the likely interaction surfaces as defined by mutagenesis data and most probable eCLIP cross-linking sites.

#### Difference map calculation

Reconstructed volumes for the full-length and truncated MmGtsf1 constructs (both with co-purifying RNA as described above) were filtered to 10  $\text{\AA}$  with cisTEM (Grant et al., 2018). Using SPIDER (Shaikh et al., 2008), a 90 pixel ( $\sim 56$   $\text{\AA}$ ) radius mask was applied to the filtered volumes after which each was normalized and aligned. This map for the truncation construct was then subtracted from the corresponding full-length map (MmGtsf1-TEV-Strep with RNA).

#### Figures

Figures of molecular models were generated using PyMOL (SchrödingerLLC, 2019). Electrostatic surface calculations were performed with APBS (Jurrus et al., 2018) with a solvent ion concentration of 0.15 M using the AMBER force field. Superpositioning of structural homologs was performed by the DALI server (Holm, 2019). Conservation analysis was performed using the ConSurf server (Ashkenazy et al., 2016). Co-evolution analysis was performed using the Gremlin server (Ovchinnikov et al., 2014). Graphs were produced in R (R Team, 2019) using the ggplot2 package (Hadley, 2016).

## QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters are described in the corresponding figure legends. All data presented for eCLIP experiments are from two replicate library preparations.

## ADDITIONAL RESOURCES

In addition to custom gene annotations and data processing scripts, extended readme documentation is provided for running and modifying the analysis code at [https://github.com/jonipsaro/asterix\\_gtsf1/](https://github.com/jonipsaro/asterix_gtsf1/).