The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation

Julia Batki[1,4], Jakob Schnabl[1,4], Juncheng Wang[2,4], Dominik Handler[1], Veselin I. Andreev[1], Christian E. Stieger[1,3], Maria Novatchkova[1,3], Lisa Lampersberger[1], Kotryna Kauneckaite[1], Karl Mechtler[1,3], Dinshaw J. Patel[2] & Julius Brennecke[1,*]

[1] Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA)
Vienna BioCenter (VBC), Dr. Bohrgasse 3, 1030 Vienna, Austria
[2] Structural Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[3] Institute of Molecular Pathology (IMP)
Vienna BioCenter (VBC), 1030 Vienna, Austria
[4] these authors contributed equally to this work
[*] correspondence: julius.brennecke@imba.oeaw.ac.at

ABSTRACT

The PIWI-interacting RNA (piRNA) pathway protects animal genome integrity in part through establishing repressive heterochromatin at transposon loci. Silencing requires piRNA-guided targeting of nuclear PIWI proteins to nascent transposon transcripts, yet the subsequent molecular events are not understood. Here, we identify SFiNX (Silencing Factor interacting Nuclear eXport variant), an interdependent protein complex required for Piwi-mediated co-transcriptional silencing in *Drosophila*. SFiNX consists of Nxf2-Nxt1, a gonad-specific variant of the heterodimeric mRNA export receptor Nxf1-Nxt1, and the Piwi-associated protein Panoramix. SFiNX mutant flies are sterile and exhibit transposon de-repression because piRNA-loaded Piwi is unable to establish heterochromatin. Within SFiNX, Panoramix recruits the cellular heterochromatin machinery, while Nxf2 binds the nascent target RNA and licenses co-transcriptional silencing. Our results reveal an unexpected, RNA-export independent role for Nxf2 in nuclear small RNA biology and suggest that NXF-variants, which have largely unknown function in animals, are functionally linked to the genome-transposon conflict.

The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation

Julia Batki[1,4], Jakob Schnabl[1,4], Juncheng Wang[2,4], Dominik Handler[1], Veselin I. Andreev[1], Christian E. Stieger[1,3], Maria Novatchkova[1,3], Lisa Lampersberger[1], Kotryna Kauneckaite[1], Karl Mechtler[1,3], Dinshaw J. Patel[2] & Julius Brennecke[1,✉]

[1] Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA)
Vienna BioCenter (VBC), Dr. Bohrgasse 3, 1030 Vienna, Austria
[2] Structural Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA
[3] Institute of Molecular Pathology (IMP)
Vienna BioCenter (VBC), 1030 Vienna, Austria
[4] these authors contributed equally to this work
✉ correspondence: julius.brennecke@imba.oeaw.ac.at

The PIWI-interacting RNA (piRNA) pathway protects animal genome integrity in part through establishing repressive heterochromatin at transposon loci. Silencing requires piRNA-guided targeting of nuclear PIWI proteins to nascent transposon transcripts, yet the subsequent molecular events are not understood. Here, we identify SFiNX (Silencing Factor interacting Nuclear eXport variant), an interdependent protein complex required for Piwi-mediated co-transcriptional silencing in *Drosophila*. SFiNX consists of Nxf2-Nxt1, a gonad-specific variant of the heterodimeric mRNA export receptor Nxf1-Nxt1, and the Piwi-associated protein Panoramix. SFiNX mutant flies are sterile and exhibit transposon de-repression because piRNA-loaded Piwi is unable to establish heterochromatin. Within SFiNX, Panoramix recruits the cellular heterochromatin machinery, while Nxf2 binds the nascent target RNA and licenses co-transcriptional silencing. Our results reveal an unexpected, RNA-export independent role for Nxf2 in nuclear small RNA biology and suggest that NXF-variants, which have largely unknown function in animals, are functionally linked to the genome-transposon conflict.

INTRODUCTION

Eukaryotic cells establish heterochromatin at genomic repeats and transposon insertions to suppress transcription and ectopic recombination (Fedoroff, 2012; Slotkin and Martienssen, 2007). One strategy to confer sequence specificity to this process is via repressor proteins (e.g. KRAB-type zinc finger repressors in tetrapods) that bind defined DNA motifs and recruit heterochromatin inducing factors (Yang et al., 2017). A second strategy for sequence-specific heterochromatin formation builds on nuclear small RNAs (Castel and Martienssen, 2013; Grewal, 2010; Holoch and Moazed, 2015). These ~20-30nt long regulatory RNAs guide Argonaute proteins to complementary sequences in nascent target transcripts, which are still attached to chromatin via transcribing RNA polymerases (Shimada et al., 2016). Binding of nuclear Argonautes to nascent transcripts leads to recruitment of chromatin modifying enzymes, ultimately resulting in heterochromatin formation and transcriptional repression. As the nascent target RNA is required for Argonaute recruitment, this process is defined as 'co-transcriptional silencing'. Besides impacting chromatin and transcription, nuclear small RNA pathways have also been linked to the co-transcriptional processes of splicing, RNA quality control and turnover (Dumesic et al., 2013; Reyes-Turcu et al., 2011; Teixeira et al., 2017). Together, this hints at complex molecular connections between nuclear Argonautes, the nascent target RNA, and chromatin. Most of our knowledge on nuclear small RNA-guided silencing is based on pioneering work in fission yeast and plants. Much less is known about how nuclear Argonautes orchestrate heterochromatin formation in metazoans.

The principal nuclear Argonaute pathway in animals is the PIWI-interacting small RNA (piRNA) pathway (Czech et al., 2018; Ozata et al., 2018). It acts preferentially in gonads to safeguard the integrity of the germline genome. In *Drosophila melanogaster*, a single nuclear Argonaute protein (Piwi) orchestrates heterochromatin formation and co-transcriptional silencing at hundreds of transposon insertions throughout the genome (Le Thomas et al., 2013; Rozhkov et al., 2013; Sienski et al., 2012; Wang and Elgin, 2011). Although transposons contain strong promoters and enhancers that respond to cellular transcription factors, binding of Piwi to nascent transposon transcripts effectively suppresses their transcription, resulting in up to several hundred-fold reductions in steady state RNA levels. How binding of the Piwi-piRNA complex to a nascent target RNA leads to silencing at the molecular level is poorly understood. Genetic studies identified three piRNA pathway-specific proteins, Maelstrom, Asterix/Gtsf1, and Panoramix/Silencio, as being required for Piwi-mediated silencing (Donertas et al., 2013; Muerdter et al., 2013; Ohtani et al., 2013; Sienski et al., 2015; Sienski et al., 2012; Yu et al., 2015). In the absence of any of these proteins, Piwi is abundantly expressed,

localizes to the nucleus, is loaded with transposon-targeting piRNAs, but is incapable of target silencing. Among these three factors, only Panoramix is capable of inducing co-transcriptional silencing through heterochromatin formation if targeted to a nascent RNA through aptamer-based tethering. Silencing via tethered Panoramix is independent of Piwi but requires the H3K9 methyltransferase Eggless/SetDB1, the H3K4 demethylase Su(var)3-3/Lsd1, and the H3K9me2/3 reader protein HP1 (Sienski et al., 2015; Yu et al., 2015). This places Panoramix downstream of Piwi and upstream of the cellular heterochromatin machinery. How Panoramix, which does not resemble any known protein, is connected to the nascent RNA, or to chromatin effectors and chromatin itself, is unknown.

Here, we show that Panoramix forms an obligatory protein complex with Nxf2-Nxt1/p15, a variant of the highly conserved Nxf1/Tap-Nxt1/p15 heterodimer. From budding yeast to humans, Nxf1-Nxt1 is the principal nuclear mRNA export receptor that mediates the translocation of export-competent mRNAs through nuclear pore complexes into the cytoplasm (Cullen, 2000; Kohler and Hurt, 2007). Nxf2 is one of three nuclear RNA export factor (NXF) variants in *Drosophila melanogaster*, but no mRNA export function could be attributed to it in Schneider cells (Herold et al., 2001; Herold et al., 2003). We show that *Drosophila* Nxf2 is an essential piRNA pathway factor that binds nascent piRNA-targeted transcripts at chromatin and licenses Panoramix for co-transcriptional silencing and heterochromatin formation. Our data demonstrate how an RNA transport receptor evolved into a co-transcriptional silencing factor, and uncover the molecular diversity of NXF-variants, which are abundant in vertebrate and invertebrate genomes but have largely unknown functions.

RESULTS

The Nxf2-Nxt1/p15 heterodimer interacts with Panoramix

To elucidate the molecular function of Panoramix, we determined its protein interactors in cultured ovarian somatic cells (OSCs), which express a functional nuclear piRNA pathway. We isolated Panoramix via immunoprecipitation (IP) from nuclear lysate of a clonal OSC line expressing FLAG-tagged Panoramix, and identified co-eluting proteins by quantitative mass spectrometry. The most prominent interactors were nuclear RNA export factor 2 (Nxf2), the mRNA export co-factor Nxt1/p15, and eIF-4B (Figure 1A left; Table S1). Among those, Nxf2 and Nxt1/p15 were also identified in genetic transposon de-repression screens (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). We confirmed the interaction between Panoramix, Nxf2 and Nxt1/p15 using reciprocal co-IP mass-spectrometry with FLAG-tagged Nxf2 as bait (Figure 1A right; Table S1). In both experiments, peptide levels for bait and interactors were in a similar range, suggesting that Panoramix, Nxf2, and Nxt1/p15 form a stable protein complex (see below; Figure S1A). In comparison, the previously identified Panoramix-interactor Piwi (Sienski et al., 2015; Yu et al., 2015) was only ~2-fold enriched and Piwi peptide levels in the IP eluates were ~20-fold lower than the other interactors (Figure S1A, Table S1), indicating a transient or sub-stoichiometric association between Piwi and Panoramix or Nxf2. A co-IP experiment using Panoramix monoclonal antibody confirmed these findings for the endogenous Panoramix, Nxf2 and Piwi proteins (Figure 1B).

Nxf2 belongs to the NXF protein family, which in *Drosophila* is composed of Nxf1, the principal mRNA export receptor, and the three NXF variants Nxf2, Nxf3, and Nxf4 (Herold et al., 2001; Herold et al., 2000). Like *piwi* and *panoramix*, *nxf2* is expressed predominantly in ovaries (Figure 1C) (Brown et al., 2014). To follow up on the connection between Panoramix and Nxf2, we generated *nxf2* null mutant flies (Figure 1D; Figure S1B). In contrast to *nxf1* and *nxt1/p15* mutants, which are lethal (Caporilli et al., 2013; Wilkie et al., 2001), *nxf2* mutants were viable and developed gonads (Figure S1C). However, as *panoramix* mutants, *nxf2* mutant females were sterile; Compared to control flies, they laid fewer eggs (Figure 1E) and none of these developed into a larva. To investigate whether the sterility of *nxf2* mutants is linked to defects in transposon silencing, we sequenced total ovarian RNA from *nxf2* mutants and control flies. Similar to *panoramix* or *piwi* mutants, *nxf2* mutants expressed strongly elevated levels of several transposon families (30%; TPM>5) while only very few endogenous mRNAs exhibited changes in their levels (Figure 1F, G; Figure S1D, E; Table S2). Among the de-silenced transposons were germline-specific (e.g. *HeT-A*, *burdock*) and soma-specific (e.g. *gypsy*, *mdg1*) elements, indicating that Nxf2 is required for transposon silencing in both ovarian tissues. In support of this, Nxf2 was expressed, like Panoramix, in ovarian somatic and germline cells (Figure 1H; Figure S1F) and RNAi mediated depletion of Nxf2 specifically in the ovarian soma or germline resulted in de-repression of cell-type specific transposon reporters (Figure 1I). Taken together, Nxf2 interacts with Panoramix, and is required for fertility and transposon silencing during oogenesis.

Nxf2 is required for Piwi-mediated heterochromatin formation

Given its interaction with Panoramix, we hypothesized that Nxf2 is required for Piwi to induce co-transcriptional transposon silencing. If so, loss of Nxf2 should not affect piRNA biogenesis, but would result in piRNA-loaded Piwi being unable to specify heterochromatin at target loci (Sienski et al., 2015; Yu et al., 2015). Indeed, Piwi levels and localization were unchanged in *nxf2* mutants (Figure S2A), indicative of intact piRNA biogenesis and loading. Furthermore, sequencing of small RNAs from Nxf2-depleted OSCs, revealed that Nxf2, similar to Panoramix, is not required for piRNA production (Figure 2A, B). We obtained similar results when we depleted Nxf2 specifically in the ovarian germline via transgenic RNAi (Figure S2B, C). Thus, irrespective of tissue and genomic origin, Nxf2 does not impact piRNA biogenesis but is required for the silencing of piRNA-targeted transposons. To ask whether Nxf2 is required for piRNA-guided co-transcriptional silencing, we turned to OSCs. Here, loss of Piwi results in up to hundred-fold elevated RNA levels of a defined subset of LTR-retrotransposons due to increased transcription accompanied by

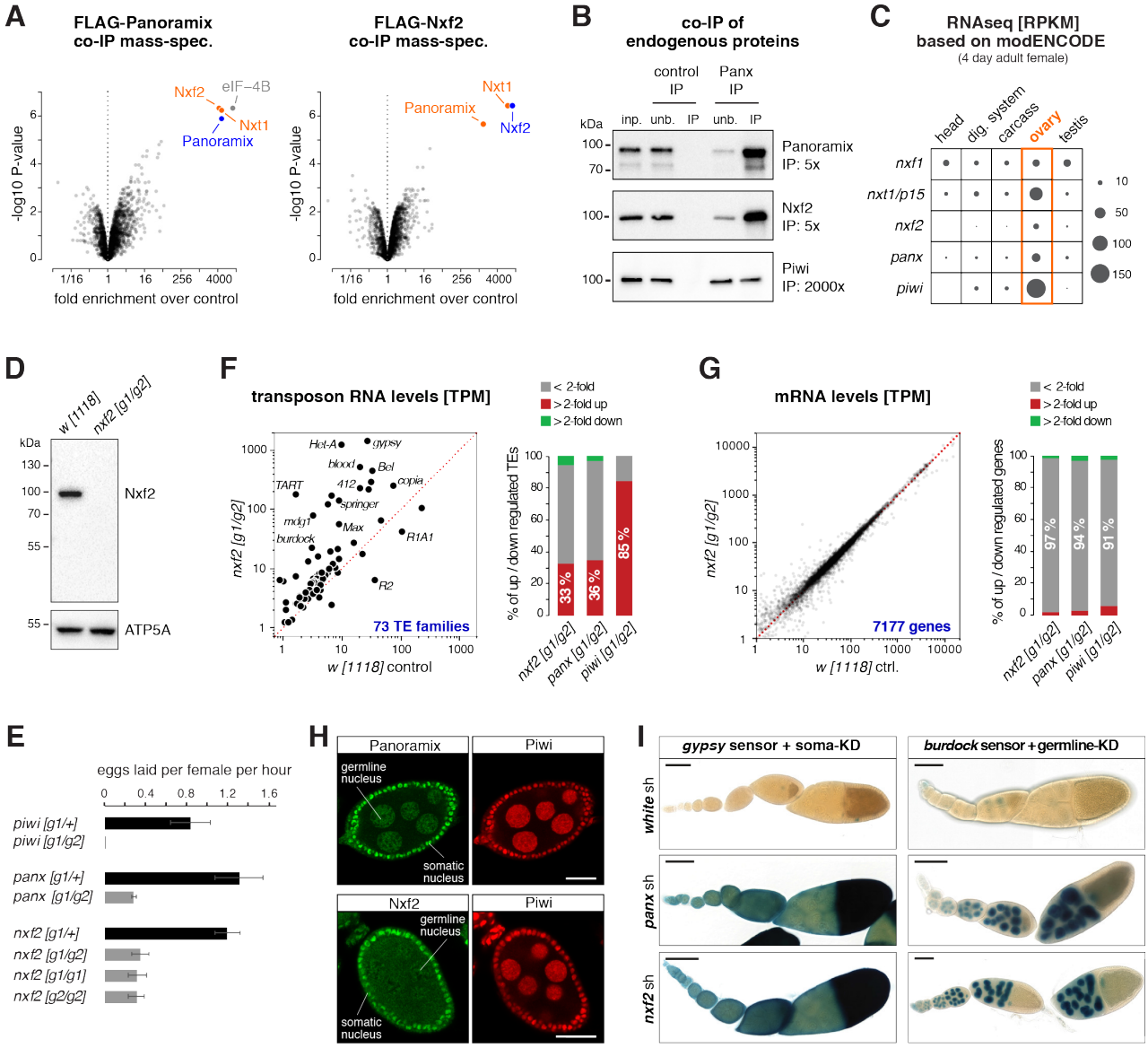

Figure 1. The Nxf2-Nxt1/p15 heterodimer interacts with Panoramix

(A) Volcano plots showing fold enrichments of proteins determined by quantitative mass spectrometry in FLAG-Panoramix (left) or FLAG-Nxf2 (right) co-immunoprecipitates versus control (n = 3). Experiments were performed with OSC cell lines stably expressing cDNA constructs encoding tagged Panoramix or Nxf2 (wildtype OSCs served as control). **(B)** Western blot analysis of co-immunoprecipitation of endogenous Panoramix from nuclear OSC lysate. Relative amount loaded in immunoprecipitation (IP) compared to input (inp.) is indicated for each protein. Mouse IgGs were used as negative control. **(C)** Shown are expression levels of indicated genes based on modENCODE mRNA-seq data in different adult tissues. **(D)** Western blot analysis of ovarian lysates (genotypes indicated) with monoclonal anti-Nxf2 antibody. ATP5A served as loading control. **(E)** Bar plots showing average egg-laying rates of females with indicated genotype (n=3; error bars indicate SD). None of the eggs laid by *panoramix* or *nxf2* mutant females developed into larvae, indicating complete female sterility. **(F, G)** Left: Scatter plots showing steady state ovarian RNA levels (transcripts per million) of genes (F) and transposons (G) in *nxf2* mutants compared to control. Right: Bar diagrams indicate differentially expressed genes or transposons in indicated genotypes. **(H)** Confocal images of egg chambers (scale bar: 20 µm) from wild-type flies co-stained for Panoramix (top) or Nxf2 (bottom) with Piwi. For specificity of antibodies see Figure S1F. **(I)** Shown are X-Gal stained ovarioles (scale bar: 0.1mm) from flies expressing *gypsy*-lacZ (soma sensor; left) or *burdock*-lacZ (germline sensor; right) plus indicated knockdown (KD) constructs (sh-lines).

loss of the heterochromatic H3K9me3 mark (Figure S2D) (Sienski et al., 2012). In Nxf2-depleted OSCs, all piRNA pathway-repressed transposons were strongly de-silenced despite normal piRNA levels (Figure 2C top, 2D; Table S3). The extent of de-repression was virtually indistinguishable to that seen in Panoramix-depleted cells (Figure 2C bottom, 2D). Consistent with a role of Nxf2 in co-transcriptional silencing, ChIP-seq experiments revealed that loss of Nxf2 resulted in increased RNA Polymerase II occupancy and reduced H3K9me3 levels for piRNA-targeted transposons like *gypsy*, or *mdg1* (Figure 2E, F; Figure S2E). In contrast, transposons not under piRNA control in OSCs (e.g. *burdock, F*-element) showed no such changes (Figure S2F, G). To assess these effects at

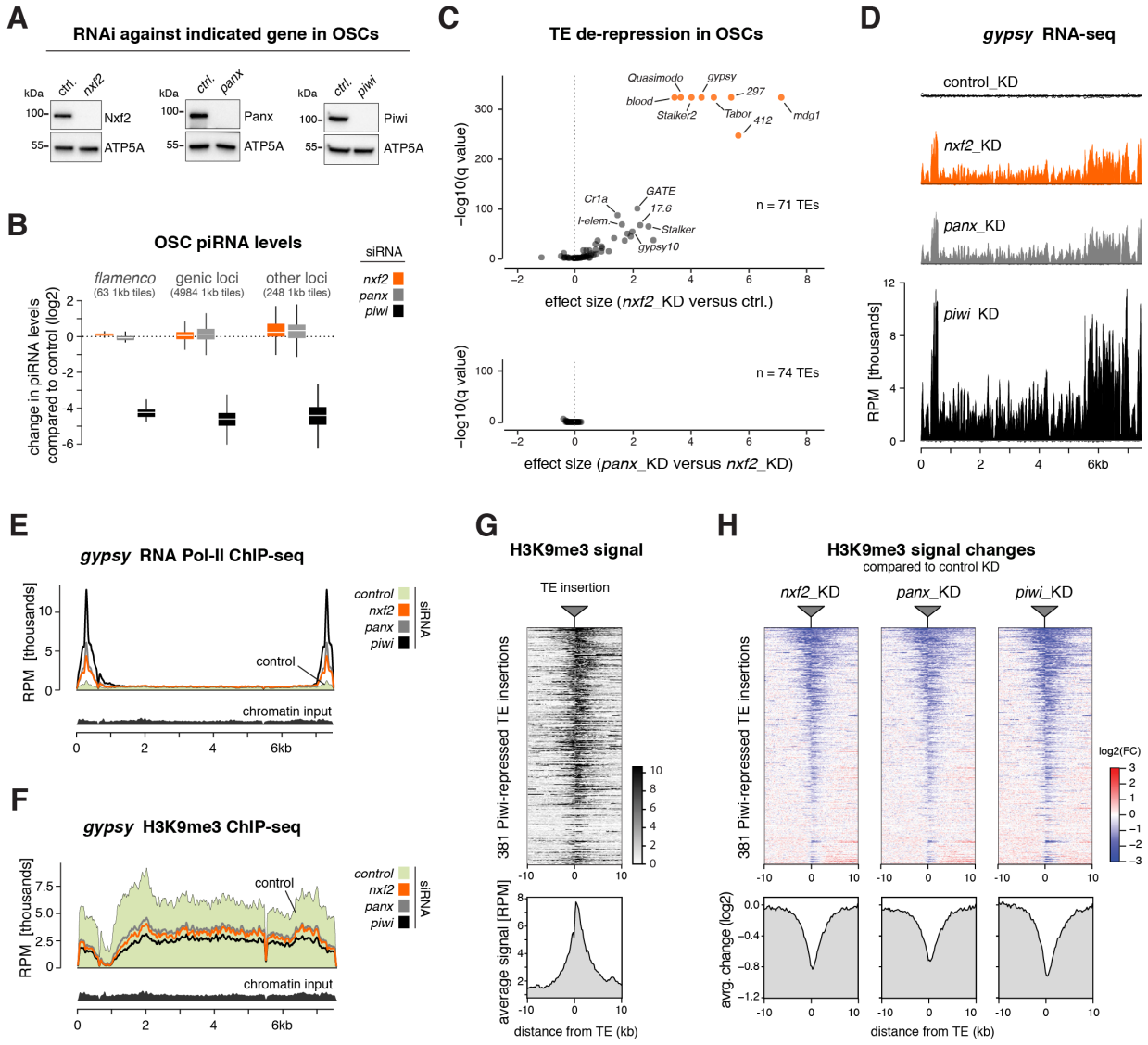

Figure 2. Nxf2 is required for Piwi-mediated heterochromatin formation

(A) Western blots showing knockdown efficiencies of indicated proteins in OSCs. ATP5A served as loading control. **(B)** Box plot showing fold change (compared to control knockdown) of piRNA levels in OSCs depleted of indicated genes (box plots indicate median, first and third quartiles (box), whiskers show 1.5× interquartile range; outliers were omitted). **(C)** Volcano plots showing changes (as effect size) in steady state transposon levels from RNA-seq experiments (n=3; upper plot: *nxf2* knockdown versus control, lower plot: *panoramix* versus *nxf2* knockdowns). **(D)** Density profiles showing normalized reads from OSC RNA-seq experiments mapping to the *gypsy* transposon consensus sequence (knockdowns indicated). **(E, F)** Density profiles showing normalized reads from OSC RNA Pol II ChIP-seq (E), or H3K9me3 ChIP-seq (F) experiments mapping to the *gypsy* transposon consensus sequence (knockdowns indicated). **(G)** Heatmap (top) and metaplot (bottom) showing H3K9me3 levels within the 20kb region flanking all 381 Piwi-repressed euchromatic transposon insertions in the OSC genome (insertions sorted for average H3K9me3 signal per window). **(H)** Heatmaps (top) and metaplots (bottom) showing control-normalized log2-fold changes in H3K9me3 levels at genomic regions surrounding transposon insertions defined in Fig. 2G (sorting as in G; gene knockdowns indicated).

the level of individual genomic loci, we examined the euchromatic insertion sites of Piwi-repressed transposons in the OSC genome. At these stand-alone transposon insertions, H3K9me3 marks spread into flanking genomic regions, to which sequencing reads could be mapped unambiguously (Figure 2G) (Sienski et al., 2012). Focusing on the ~380 Piwi-silenced transposon insertions revealed that Nxf2 loss phenocopies the decrease in H3K9me3 levels seen in cells depleted of Panoramix or Piwi (Figure 2H). Piwi-independent H3K9me3 domains instead were unaffected in Nxf2-depleted cells (Figure S2H).

Several dozen piRNA-repressed transposons are inserted in the vicinity of endogenous gene loci. In the absence of Piwi or Panoramix, loss of repressive heterochromatin at these transposon insertions results in elevated transcription of these genes (Figure S2I; Table S4) (Sienski et al., 2012). A highly similar set of genes was differentially expressed in Nxf2- or Panoramix-depleted OSCs (Figure S2I; Table S4). The rare changes in gene expression caused by Nxf2 loss can, therefore, be attributed directly to impaired Piwi-mediated heterochromatin formation at transposon loci. We confirmed all findings based on RNAi-mediated depletion of Nxf2

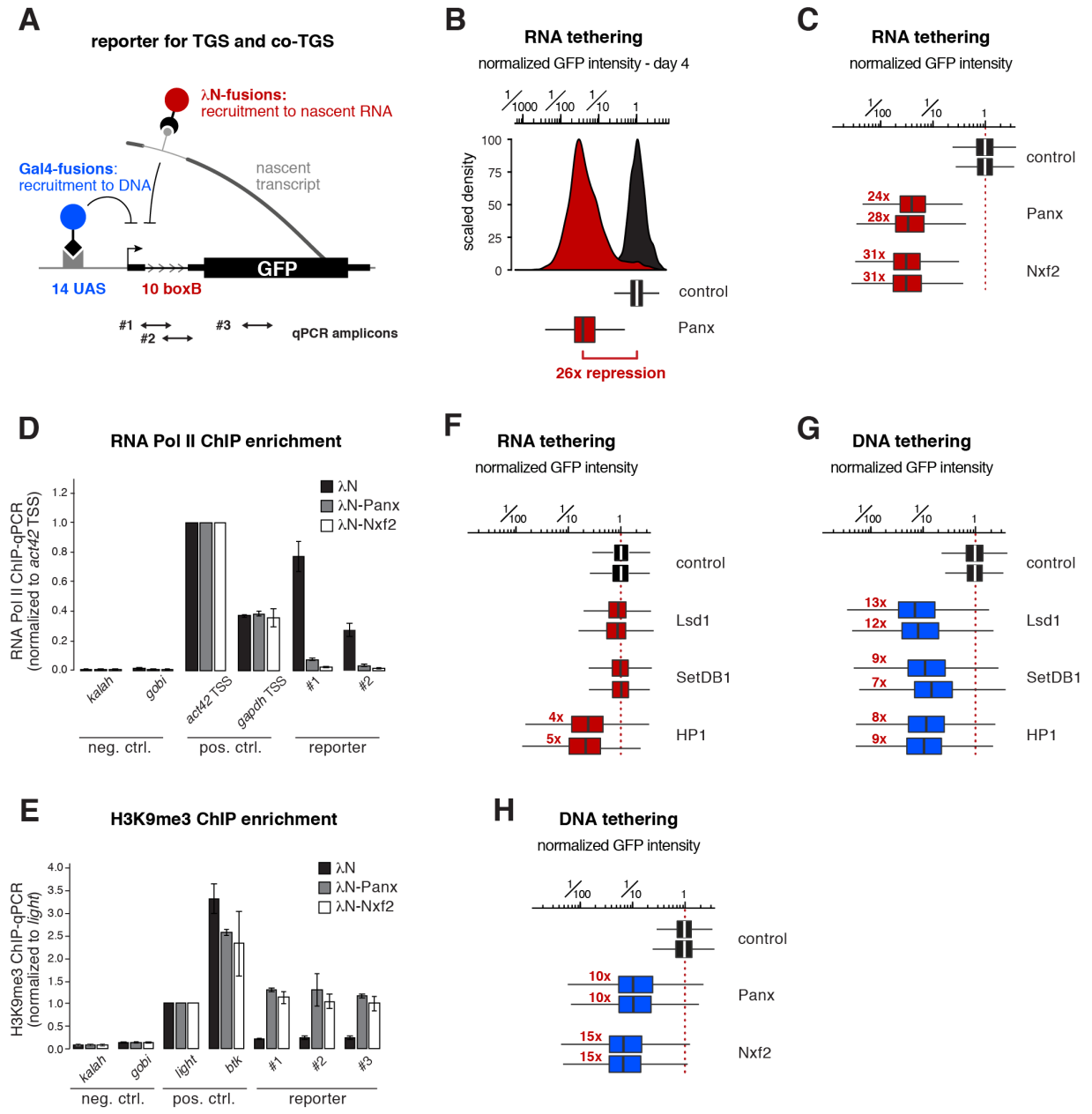

Figure 3. Targeting Nxf2 to nascent RNA induces co-transcriptional silencing
(A) Cartoon of the GFP silencing reporter, which allows direct recruitment of proteins to the nascent RNA via intronic boxB sites and to DNA via UAS sites (positions of qPCR primer amplicons used in D, E are indicated). **(B)** Histogram (top) and corresponding box plot (bottom) showing the λN-control normalized GFP intensity in transfected OSCs (4 days after transfection of λN-tagged fusion proteins; n=2500 cells; box plot definition as in Fig. 2B; red number indicates fold-repression value based on median, normalized to control). **(C)** Box plots showing GFP intensity in OSCs 4 days after transfection with plasmids expressing indicated λN-fusion proteins (red numbers indicate fold-repression value based on median, normalized to control for two replicates; box plot definition as in Fig. 2B). **(D, E)** Bar graphs showing normalized RNA Pol II occupancy (D) and H3K9me3 levels (E) at indicated loci determined by chromatin immunoprecipitation (ChIP) followed by quantitative PCR using OSCs expressing the indicated λN-fusions (n = 2; error bars: SD). **(F)** As in C. **(G, H)** As in C, but for indicated Gal4 DBD fusions.

in OSCs with a second independent siRNA targeting *nxf2* (Figure S2J, K; Table S4). Our combined results support a model where Nxf2—rather than acting as a cellular RNA transport receptor—is required for co-transcriptional silencing and heterochromatin formation downstream of Piwi.

Targeting Nxf2 to nascent RNA induces co-transcriptional silencing

To test Nxf2's involvement in heterochromatin-mediated silencing more directly, we converted a reporter system that assays co-transcriptional silencing independently of piRNAs in ovaries

(Sienski et al., 2015; Yu et al., 2015) into a quantitative cell culture assay. We generated a clonal OSC line harboring a single-copy transgene that expresses GFP under control of a strong enhancer (Figure S3A). To mimic piRNA-guided target silencing, any factor of interest can be recruited to the nascent reporter RNA as a λN-fusion protein, through boxB sites located in the intron of the reporter construct (Figure 3A). The same reporter cell line also allows recruitment of factors of interest to the reporter DNA (via the Gal4-UAS system) upstream of the enhancer in order to assay transcriptional silencing independent of targeting the nascent RNA (Figure 3A).

While the expression of λN alone had no impact on reporter expression, transient expression of λN-Panoramix resulted in more than 25-fold reporter repression for four to five days (Figure 3B; Figure S3B). Expression of λN-Nxf2 led to similarly potent repression (Figure 3C; Figure S3C). In both cases, silencing correlated with reduced RNA Pol-II occupancy and establishment of an H3K9me3 domain at the reporter locus (Figure 3D, E). Experimental tethering of either Panoramix or Nxf2 to a nascent RNA, therefore, induces potent co-transcriptional silencing accompanied by heterochromatin formation. The efficiency of this silencing process is remarkable considering that the boxB sites reside in an intron of the reporter construct, and thus one would expect that the target RNA is only transiently present at the encoding DNA locus. Indeed, the heterochromatin promoting factors Eggless/SetDB1, Su(var)3-3/Lsd1, or Su(var)205/HP1a, which act downstream of Piwi, were not capable of inducing comparable co-transcriptional silencing when targeted to the nascent reporter RNA (Figure 3F). The same factors silenced the reporter as efficiently as Panoramix or Nxf2 when targeted directly to the reporter DNA (Figure 3G, H). Therefore, we propose that co-transcriptional silencing, i.e. silencing via targeting nascent RNA, by Panoramix and Nxf2 requires more than merely recruiting the so-far identified heterochromatin effectors to the nascent RNA.

Panoramix and Nxf2-Nxt1 form the interdependent SFiNX complex

Nxf2 and Panoramix interact, are both capable of inducing co-transcriptional silencing, and loss of either protein results in highly similar phenotypes (Figures 1-3). These findings suggested a close molecular connection between Nxf2 and Panoramix. In support of this, we found a strong reciprocal dependency between both proteins: Depletion of Nxf2 or Panoramix in OSCs resulted in substantial reductions of the respective other protein, while the corresponding mRNA levels were unchanged (Figure 4A, B). Similarly, in *nxf2* mutant ovaries, Panoramix protein was hardly detectable by western blotting or immunofluorescence analysis (Figure 4C, D), despite unchanged *panoramix* mRNA levels (Figure S4A). Conversely, Nxf2 protein levels were reduced in *panoramix* mutant ovaries (Figure 4C). In these ovaries, the remaining Nxf2 protein was excluded from the nucleus in germline and soma (Figure 4E), suggesting that Nxf2's nuclear localization depends on Panoramix. Consistent with this, Panoramix harbors a Lysine-rich sequence stretch (residues 196-262), which is required for its nuclear localization (Figure S4B). Based on these results, we hypothesized that Panoramix and Nxf2 stabilize each other via forming a nuclear protein complex.

To test for a putative Panoramix-Nxf2 complex, we used insect cells (High Five) and co-expressed both proteins together with Nxt1/p15, which we identified as Nxf2 and Panoramix interactor (Figure 1), and which functions as a general NXF cofactor (Herold et al., 2000). While all three proteins were abundantly produced, Panoramix was degraded. Therefore, we expressed instead of full-length Panoramix a 25 kDa fragment that is necessary and sufficient for binding Nxf2 (see below). A single affinity purification step of the Strep-tagged Panoramix fragment, followed by size exclusion chromatography, resulted in a defined protein peak containing all three factors (Figure 4F, G; Figure S4C). Panoramix, Nxf2, and Nxt1/p15 therefore form a stable protein complex, which we named SFiNX (Silencing Factor interacting Nuclear eXport factor variant).

Panoramix-mediated silencing via nascent RNA requires Nxf2

Due to the interdependency between Panoramix and Nxf2, our functional experiments so far interrogated primarily the function of the SFiNX complex rather than that of the individual proteins. To disentangle the molecular roles of Panoramix and Nxf2 within SFiNX, we set out to generate interaction-deficient point-mutant variants. Panoramix consists of two parts, an N-terminal disordered half, and a C-terminal half with predicted secondary structure elements (Figure 5A). Using GFP-tagged full-length Nxf2 as bait, we mapped the interaction site within Panoramix to the first part of the structured domain (Figure 5A; Figure S5A, B), and within there, identified two regions that upon deletion impacted the Nxf2-Panoramix interaction (Figure S5C). Panoramix Δ308-386 failed to bind Nxf2, while Panoramix Δ387-446 interacted less efficiently with Nxf2. The 308-386 peptide harbors a predicted amphipathic alpha-helix (Figure S5D, E). Mutating four hydrophobic residues predicted to line one side of this helix abrogated the Nxf2 interaction (Figure 5B; Figure S5E). Notably, the two Nxf2 binding-deficient Panoramix variants (Panoramix[Δ308-386] and Panoramix[helix mutant]) accumulated to lower levels compared to the wildtype protein (Figure S5F). In contrast, Panoramix Δ387-446 accumulated to higher levels than wildtype Panoramix (Figure S5F). In light of the co-dependency between Panoramix and Nxf2 in vivo, we hypothesized that the 387-446 peptide within Panoramix harbors a destabilizing element that induces protein degradation if not protected by Nxf2. Consistent with this, fusing the 387-446 peptide to GFP led to a ~100-fold reduction in GFP levels (Figure S5G). Mutation of four hydrophobic residues within this "degron" led to increased Panoramix levels (Figure S5D, F). When combining both sets of point mutations, the resulting Panoramix[helix+degron mutant] accumulated to high levels and was unable to interact with Nxf2 (Figure 5B; Figure S5F).

We then turned to Nxf2, which has a similar domain organization like the mRNA export receptor Nxf1/Tap, except that it has a

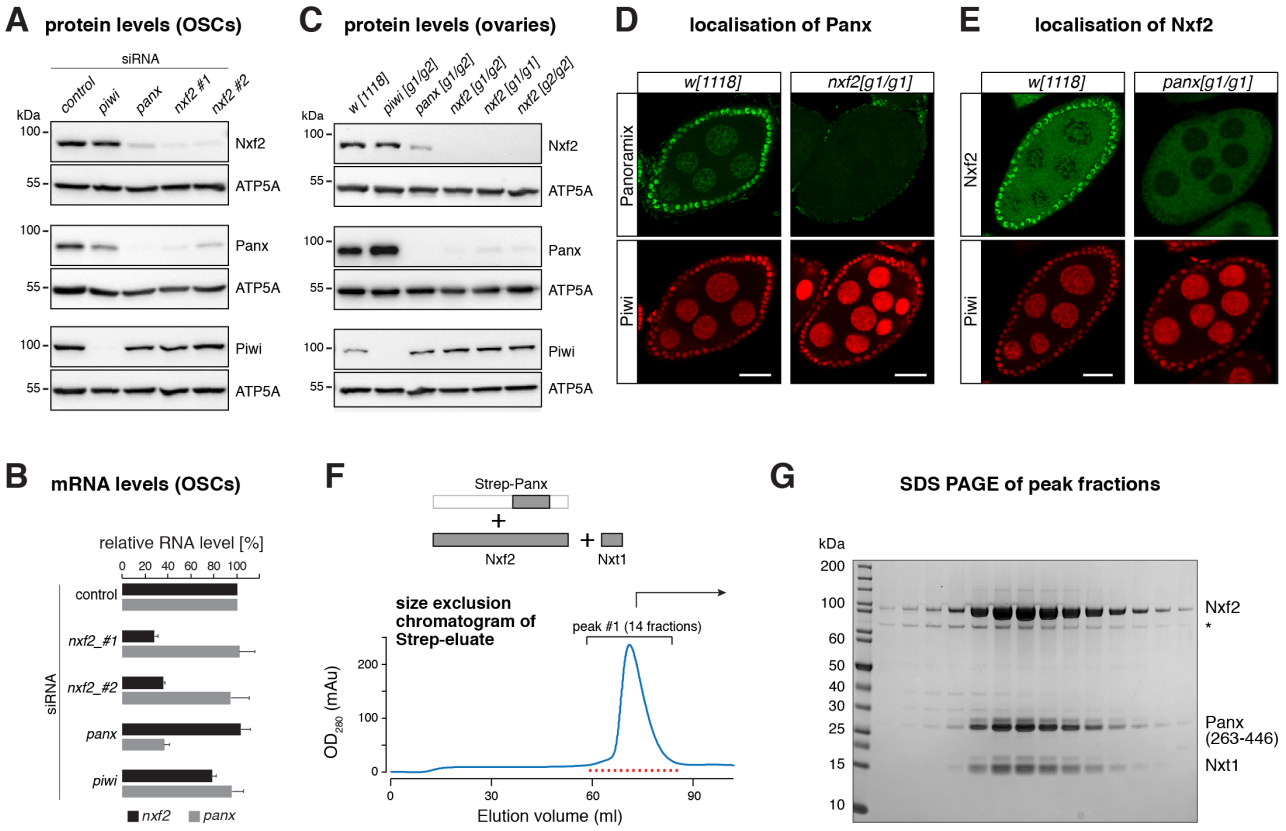

Figure 4. Panoramix and Nxf2-Nxt1 form the interdependent SFiNX complex
(A) Western blots showing levels of indicated proteins in lysates from OSCs with indicated gene knockdowns (ATP5A served as loading control). **(B)** Bar graph showing fold change of RNA-seq levels of indicated genes compared to control knockdown in OSCs (n = 3, error bars: SD). **(C)** Western blots showing levels of indicated proteins in ovary lysates from flies with indicated genotypes (ATP5A served as loading control). **(D)** Confocal images of egg chambers (scale bar: 20 µm) from *w[1118]* control and *nxf2* mutant flies stained for Panoramix (upper panels) and Piwi (lower panels). **(E)** As in D, but for *panoramix* mutants and Nxf2 protein. **(F)** Top: schematics of the trimeric complex expressed in insect cells (grey boxes indicate co-expressed proteins/fragments). Bottom: size exclusion chromatogram of the affinity-purified Strep-Panoramix eluate (mAU = milli-absorbance unit). **(G)** SDS-PAGE of the 14 peak fractions indicated in F (Coomassie blue staining; * indicates HSP70 as determined by mass spectrometry).

duplicated N-terminal putative RNA binding unit, consisting of RNA recognition motif (RRM) and Leucine-rich repeat (LRR) (Figure 5C). Using stabilized Panoramix[degron-mutant] as bait, we determined that Nxf2's two C-terminal domains (NTF2-like plus UBA domains) are sufficient to bind Panoramix, and that the UBA domain is required for the Panoramix interaction (Figure S5H). Thus, we reasoned that the UBA domain harbors a critical binding site for Panoramix. Indeed, when we purified the 30 amino-acid amphipathic Panoramix helix, that is required for Nxf2 binding, from bacterial lysate, the co-expressed and untagged Nxf2 UBA domain co-eluted (Figure S5I). Moreover, we were able to determine the atomic structure of Nxf2's UBA domain in complex with the Panoramix helix (connected via a flexible linker: UBA-linker-Panoramix helix) at 1.9 Å resolution (Table S5). This revealed that the UBA domain of *Drosophila* Nxf2 consists of a three-helix bundle (α1-α3) with a fourth helix at the C terminus (Figure 5D), that is highly similar to that of human NXF1. α1, α3 and α4 form a hydrophobic core that interacts with the hydrophobic face of the Panoramix amphipathic helix involving six Panoramix residues (L323, A324, V325, A328, V331 and L332). In addition, the N terminus of the Panoramix helix is stabilized by two flanking hydrogen bonds and salt bridges (Figure 5E). Our experimentally determined Panoramix[helix mutant] variant that cannot bind Nxf2 (Figure 5B) fully supported this structure: three out of the six hydrophobic residues contributing to the interaction were mutated in this Panoramix variant. Within Nxf2, ten hydrophobic residues contribute to the Panoramix interaction (Figure 5F). Out of these, only two (V800 and I827) were changed to hydrophilic/charged residues in *Drosophila* Nxf1 (Q632 and E657), which does not interact with Panoramix (Figure S6C, D). When we mutated V800 and I827 in Nxf2, together with two flanking residues, into the corresponding Nxf1 amino acids, the resulting Nxf2[UBA mutant] was unable to bind Panoramix (Figure 5G; Figure S6C).

With the interaction-deficient Nxf2 and Panoramix variants in hand, we determined whether the individual proteins are capable of supporting Piwi-mediated silencing. We performed genetic rescue experiments in OSCs and asked whether expression of

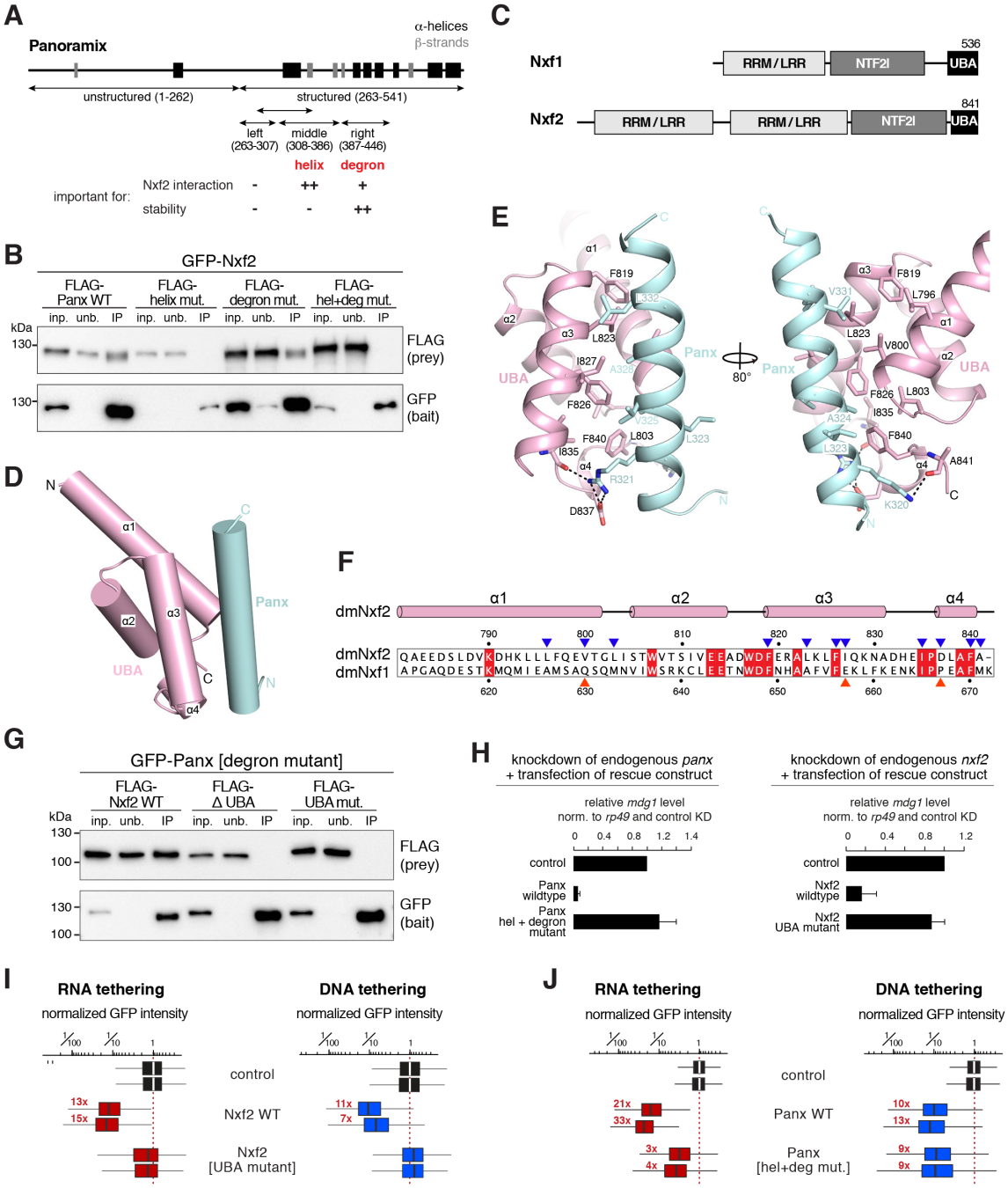

Figure 5. Panoramix-mediated silencing via nascent RNA requires Nxf2

(A) Cartoon of the Panoramix protein showing predicted secondary structure elements and region boundaries used in the interaction mapping experiments. The summary below states the involvement of the [helix] and [degron] sites in the Nxf2 interaction and in Panoramix' protein stability. **(B)** Western blot analysis of GFP-Nxf2 immuno-precipitation experiments using lysate from S2 cells transiently co-transfected with indicated FLAG-Panoramix expressing plasmids (relative amount loaded in immunoprecipitation lanes: 3×). **(C)** Schematics of the domain architecture of Nxf1/Tap and Nxf2. RRM: RNA recognition motif; LRR: leucine rich repeat; NTF2l: nuclear transport factor 2-like domain; UBA: ubiquitin associated domain. **(D)** Cartoon view of the UBA domain of Nxf2 (pink) in complex with the Panoramix helix (cyan). **(E)** Detailed views of interfaces between the UBA domain of Nxf2 (pink) and the Panoramix helix (cyan). **(F)** Protein sequence alignment of the UBA domain of Nxf2 and Nxf1 from *Drosophila melanogaster* with the secondary structure elements of Nxf2 indicated above the alignment. Identical residues are highlighted by red squares, key Nxf2 residues involved in the Panoramix interaction are indicated by blue triangles, while red triangles mark interaction residues that are different in Nxf1. **(G)** Western blot analysis of immuno-precipitation experiments (bait: stabilized degron mutant GFP-Panoramix) using lysate from S2 cells transiently co-transfected with indicated FLAG-Nxf2 expressing plasmids (relative amount loaded in immunoprecipitation lanes: 3×). **(H)** Bar graphs showing the transposon repression rescue potential of indicated constructs transfected into OSCs depleted of endogenous Panoramix (left) or Nxf2 (right). *mdg1* transposon levels determined via qRT-PCR (n = 2, error bars: SD). **(I, J)** Box plots showing GFP intensity in OSCs 4 days after transfection with plasmids expressing indicated λN-tagged (red) or Gal4-tagged (blue) fusion proteins (red numbers indicate fold-repression values based on median, normalized to control for two replicates; box plot definition as in Fig. 2B).

siRNA resistant Panoramix[helix+degron-mutant] or Nxf2[UBA-mutant] variants could restore silencing of the *mdg1* transposon in OSCs depleted for endogenous Panoramix or Nxf2. While the respective wild-type proteins supported *mdg1* silencing, neither of the interaction-deficient mutants (expressed with NLS sequence to assure nuclear localization; Figure S6E, F) displayed rescue activity (Figure 5H). Therefore, both Nxf2 and Panoramix contribute essential activities to the silencing process beyond reciprocal protein stabilization. To investigate the function of Panoramix and Nxf2 as silencing factors more directly, we turned to the OSC tethering assay for co-transcriptional silencing (Figure 3A). This revealed that Nxf2 with point mutated UBA domain was entirely inert in inducing reporter silencing, irrespective of whether it was targeted to the nascent RNA or to the DNA directly (Figure 5I; Figure S6G). Instead, the Panoramix[helix+degron mutant] variant, which is defective in Nxf2 binding, showed clear, though in comparison to the wildtype protein moderate co-transcriptional silencing activity (Figure 5J; Figure S6H). Remarkably, when recruited directly to the reporter DNA, Panoramix[helix+degron mutant] was as potent in inducing silencing as wildtype Panoramix (Figure 5J; Figure S6H). Taken together, our data indicate that Panoramix, and not Nxf2, is the silencing co-factor within SFiNX that links to the heterochromatin machinery. Nxf2 instead is required for Panoramix to achieve potent co-transcriptional silencing, consistent with Nxf2 being a predicted RNA-binding protein.

Nxf2 lost nucleoporin binding but retained RNA binding activity

Nxf2 evolved from the principal mRNA export receptor Nxf1/Tap. To understand how Nxf2 confers co-transcriptional silencing activity to SFiNX, we reasoned that a specific molecular feature intrinsic to Nxf1/Tap was exploited by the rather surprising evolutionary exaptation of an RNA transporter into co-transcriptional silencing. At the same time, Nxf2 must have lost other key features of Nxf1 in order to not get channeled into mRNA export biology. We therefore set out to systematically compare Nxf2 to the well-studied Nxf1/Tap protein.

One central molecular feature of Nxf1/Tap is its ability to shuttle through the selective phenylalanine-glycine (FG) repeat meshwork of the inner nuclear pore complex (NPC). Two nucleoporin FG-binding pockets, one residing in the UBA domain and one in the NTF2-like domain, confer NPC shuttling ability to Nxf1/Tap (Braun et al., 2002). We examined both sites in Nxf2 at the structural level. The putative FG-binding pocket within Nxf2's UBA domain lies on the opposite side of the Panoramix binding surface, making it per se accessible (Figure S7A). However, a salt bridge between E814 and K829 restricts access to the hydrophobic core of the pocket, rendering it probably non-functional (Figure 6A). To inspect the second putative FG-binding pocket, we determined the 2.8 Å resolution crystal structure of Nxf2's NTF2-like domain bound to Nxt1/p15 (Figure S7B; Table S5). Based on this, Nxf2 interacts with Nxt1/p15 in a manner very similar to human NXF1/TAP (Figure S7C) (Fribourg et al., 2001). Nxf2's second putative FG-binding pocket within the NTF2-like domain is also probably non-functional: It is occupied by the bulky side chains of its own Phe735 and Tyr690 and in addition, Arg747 closes access to the hydrophobic pocket by hydrogen bonding with Tyr690 (Figure 6B). In the case of Nxf1/Tap, mutations in either of the two FG-binding pockets abrogate nucleoporin binding (Braun et al., 2002). It is therefore unlikely that Nxf2 retained NPC binding ability. In support of this, in vitro experiments demonstrated lack of affinity between FG-repeat peptides and Nxf2 (in complex with Nxt1 and Panoramix), while human NXF1/TAP (in complex with Nxt1) readily bound FG-peptides (Figure 6C). Furthermore, GFP-tagged Nxf2 did not accumulate at nuclear pores where GFP-tagged Nxf1/Tap is highly enriched (Figure S7D). We conclude that Nxf2 lost the ability to interact with nucleoporins due to specific changes in both of its FG-binding pockets.

Nxf1/Tap's second key feature is its ability to bind mRNA cargo via the N-terminal RRM-LRR domains (Liker et al., 2000). This RRM-LRR fold is duplicated in Nxf2 (Figure 5C). Electrophoretic mobility shift assays revealed that Nxf2's N-terminal RRM-LRR domain (1st unit) is able to bind single-stranded RNA in vitro (Figure 6D, E; we did not succeed in obtaining soluble recombinant 2nd RRM-LRR unit). The RNA binding activity of the 1st RRM-LRR unit was abrogated upon mutating three positively charged amino acids, whose equivalent residues in Nxf1/Tap contact the constitutive transport element of simian type D retroviral transcripts (Figure S7E, F, G) (Aibara et al., 2015; Teplova et al., 2011). To ask whether Nxf2's ability to bind RNA in vitro is relevant in vivo, we performed genetic rescue experiments. In flies and OSCs, expression of Nxf2[$\Delta 1^{st}$ RRM-LRR] instead of the wildtype protein was not able to support transposon silencing (Figure 6F; Figure S7H, I), although this variant localized to the nucleus (Figure S7I) and interacted with Panoramix (Figure 6G). These findings suggested a model where Nxf2, via its N-terminal RNA-binding domain, anchors SFiNX, presumably after its initial recruitment via the Piwi-piRNA complex, to nascent RNAs. If this were true, SFiNX lacking Nxf2's 1st RNA binding unit should remain silencing competent if recruited to the target RNA via the λN-boxB tethering system. Indeed, expression of λN-tagged Nxf2[$\Delta 1^{st}$ RRM-LRR] induced co-transcriptional silencing as efficiently as λN-tagged wildtype Nxf2 (Figure 6H; Figure S7J).

Nxf2 binds nascent Piwi/piRNA targets

In light of Nxf2's RNA binding activity in vitro, we set out to identify its cellular RNA targets via cross-linking immuno-precipitation (CLIP), a challenging experiment considering that piRNA targets are transcriptionally repressed and therefore only present at low steady state levels (Sienski et al., 2012). We UV-cross-linked OSCs carrying a GFP-tag at both endogenous *nxf2* loci, treated lysate from these cells with partial RNase digestion, purified Nxf2-GFP under denaturing conditions, and cloned co-purified RNA fragments for sequencing. The identical procedure

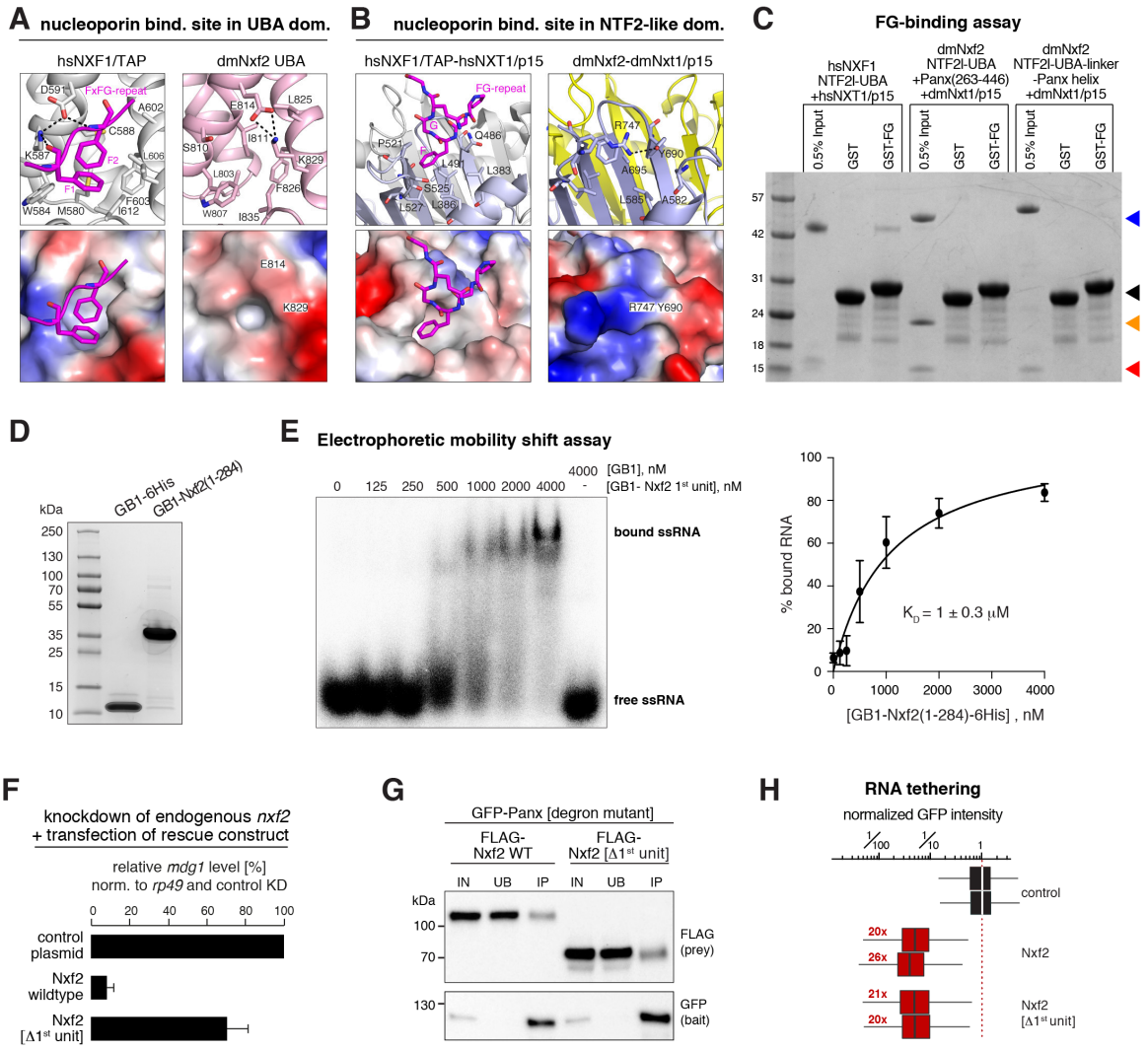

Figure 6. Nxf2 lost nucleoporin binding but retained RNA binding activity

(A) Left: Ribbon (top) and electrostatic surface (bottom) representation showing the FG-repeat binding pocket in the crystal structure of hsNXF1/TAP UBA domain (gray) in complex with a nucleoporin FxFG-repeat peptide (magenta; PDB ID: 1OAI) (Grant et al., 2003). Right: Ribbon (top) and electrostatic surface (bottom) representation showing the closed putative FG-repeat binding pocket in the crystal structure of dmNxf2's UBA domain (pink). **(B)** Left: Ribbon (top) and electrostatic surface (bottom) representation showing the specific recognition of nucleoporin FG-repeat (magenta) in the hsNXF1/TAP (light blue)-NXT1/p15 (gray) structure (PDB ID: 1JN5) by a hydrophobic FG-repeat binding pocket (Fribourg et al., 2001). Right: Ribbon (top) and electrostatic surface (bottom) representation showing the blocked putative FG-repeat binding pocket in the crystal structure of Nxf2's NTF2-like domain (purple) complexed with Nxt1/p15 (yellow). The hydrogen bond is shown as black dashed line. **(C)** Coomassie-stained SDS PAGE testing the interaction between FG-peptides and indicated protein complexes (arrowheads at the right refer to: black: GST or GST-FG repeat; blue: NXF1/TAP, Nxf2(NTF2l + UBA) or Nxf2(NTF2l + UBA)-linker-Panoramix-helix); red: Nxt1; orange: Panoramix (263-446). **(D)** Coomassie stained SDS-PAGE showing purified RNA-binding unit (1st RRM-LRR domain) of Nxf2 and the GB1 control peptide used for the EMSA experiment (E). **(E)** Left: Phosphorimage showing Electrophoretic Mobility Shift Assay (EMSA) with labelled single stranded RNA (ssRNA) and increasing amount of indicated recombinant protein (D). Right: Binding curve for the EMSA calculated from the phosphorimage (n = 3; error bars: SD). **(F)** Bar graph showing the transposon repression rescue potential of indicated constructs transfected into OSCs depleted of endogenous Nxf2. *mdg1* transposon levels determined via qRT-PCR (n = 2, error bars: SD). **(G)** Western blot analysis of immuno-precipitation experiments (bait: stabilized degron-mutant GFP-Panoramix) using lysate from S2 cells transiently co-transfected with indicated FLAG-Nxf2 expressing plasmids (relative amount loaded in immunoprecipitation lanes: 3x). **(H)** Box plots showing GFP intensity in OSCs 4 days after transfection with plasmids expressing indicated λN-tagged fusion proteins (red numbers: fold-repression values based on median, normalized to control for two replicates are indicated; box plot definition as in Fig. 2B).

with wildtype OSCs served as negative control. Most reads in the experimental and control libraries mapped to expressed gene loci (Figure S8A depicts *actin5C* as an example). As these reads were not differentially enriched in either library, we considered them background. When we mapped reads to transposon consensus sequences, three out of 54 transposon families with analyzable read coverage showed >3-fold enrichment in the Nxf2 CLIP library for reads mapping to the sense, but not to the antisense strand (Figure 7A, Figure S8B). These transposons were *gypsy*, *17.6*, and *blood*, which are among the top piRNA-targeted transposons in OSCs.

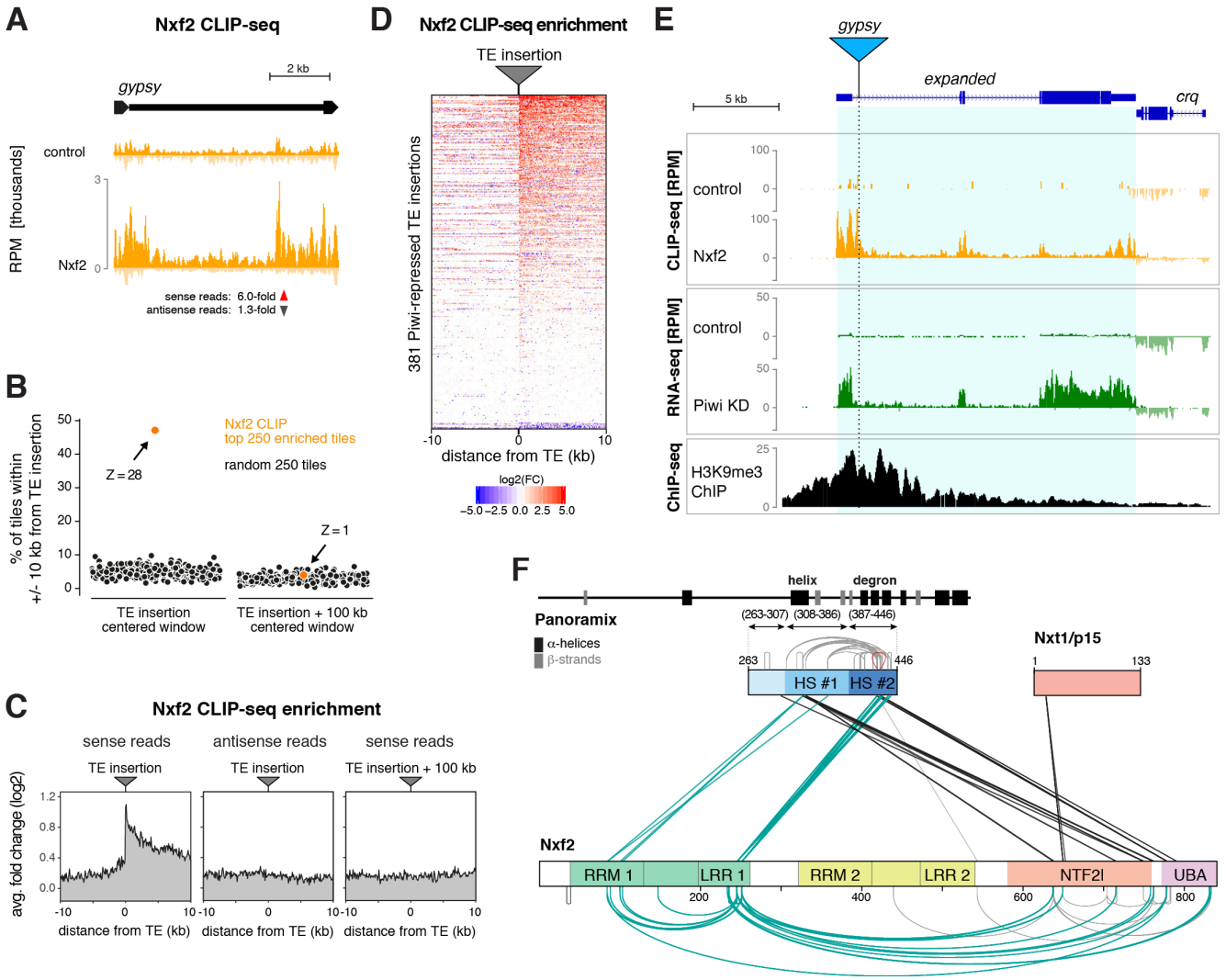

Figure 7. Nxf2 binds nascent Piwi/piRNA targets

(A) Density profiles showing normalized reads from Nxf2 CLIP-seq experiments versus control mapping to the *gypsy* transposon sequence (numbers below indicate the average fold change for sense and antisense reads). **(B)** Left: Jitter plot showing the percentage of 250 randomly chosen genomic 1kb tiles (repeated 100 times: black dots) with measurable CLIP-seq signal that map within the 20kb region flanking the 381 Piwi-repressed euchromatic transposon insertions in OSCs (left). The orange dot corresponds to the 250 1kb tiles exhibiting the highest Nxf2 CLIP-seq enrichment genome-wide. Right: Exactly as in the left plot, but the genomic position of the 381 transposon insertions was shifted by 100kb. **(C)** Metaplots showing the average, control-normalized changes (log2-fold) in Nxf2 CLIP-seq signal at the 381 piRNA repressed transposon insertions in OSCs. Left: sense CLIP-seq reads; middle: antisense CLIP-seq reads; right: sense CLIP-seq reads after shifting the transposon insertion sites by 100 kb; all transposon insertions were oriented on the genomic sense strand. **(D)** Heatmap showing control-normalized changes (log2-fold) in Nxf2 CLIP-seq signal at the 381 Piwi-repressed euchromatic transposon insertions in the OSC genome with all insertions oriented on the genomic plus strand showing sense CLIP reads (insertions are sorted for decreasing average log2-fold change in the region). **(E)** UCSC genome browser panels showing CLIP-seq, RNA-seq and H3K9me3 ChIP-seq reads at the *expanded* locus, which harbors an OSC-specific insertion of *gypsy* (dotted line; insertion not present in reference genome). **(F)** Schematic summary of all crosslinks identified between Panoramix, Nxf2, and Nxt1 in the recombinant SFiNX complex. Protein domains in Nxf2 are indicated (RRM: RNA recognition motif; LRR: leucine rich repeat; NTF2l: nuclear transport factor 2-like domain; UBA: ubiquitin associated domain). Within Panoramix, hotspot (HS) #1 includes the amphipathic helix, and HS #2 includes the degron site. Numbers indicate amino acid positions within the proteins. Straight lines indicate intermolecular crosslinks, half-circles indicate intramolecular crosslinks and the red loop indicates an intermolecular self-link.

Encouraged by this, we set out to identify Nxf2-bound transcripts in an unbiased manner. Focusing on genome-unique mappers, we determined the Nxf2 CLIP-seq enrichment for all genomic 1kb tiles. Strikingly, most tiles with CLIP-seq enrichment mapped in the vicinity of transposon insertions. For example, nearly 50% of the 250 most enriched 1kb tiles (15-200-fold enriched) reside within 20kb of the 381 euchromatic TE insertions that are repressed by piRNAs in OSCs, 9.3-fold more than expected by chance (Z-score: 28; Figure 7B). When we plotted the Nxf2 CLIP-seq enrichment over control at the 381 piRNA repressed TE insertions, most of

these loci exhibited a strong positive signal, predominantly downstream of the insertion site (all transposons were oriented on the genomic plus strand and no enrichment was seen on the opposite strand) (Figure 7C, D). To our surprise, the Nxf2 CLIP-seq enrichment extended for many kilobases, far away from the piRNA targeted transposon sequence (Figure 7D). A typical example is the *expanded* locus shown in Figure 7E. Here, a piRNA targeted *gypsy* transposon is inserted in the first *expanded* intron. Although no piRNAs target *expanded* besides the intronic *gypsy* sequence, Nxf2 CLIP-seq reads were strongly enriched along the entire *expanded* RNA. These data indicate that (1) Nxf2, once recruited to a target RNA (presumably via Piwi), spreads along the entire nascent transcript and that (2) most transposon insertions initiate transcription that extends into the downstream flanking genomic regions, and even under piRNA repressed conditions these nascent transcripts are detectable when experimentally enriched via Nxf2 immunoprecipitation.

Our data suggest that Nxf2, similar to the mRNA export receptor Nxf1/Tap, binds RNA in a sequence-independent manner. This raised the important question of how Nxf2 avoids binding to random nuclear RNAs, which would bear the danger of ectopic silencing and heterochromatin formation. Inspired by the Nxf1 literature, we hypothesized that Nxf2's RNA binding activity is controlled. Prior to mRNA cargo binding, Nxf1/Tap is in a closed conformation with its RNA binding unit folding back onto the NTF2-like domain (Viphakone et al., 2012). Upon adaptor-mediated recruitment of Nxf1/Tap to an export-competent mRNA, this intramolecular interaction is released, RNA cargo is bound, and the complex shuttles through the NPC. To probe for a putative intramolecular regulatory interaction within SFiNX, we took advantage of the recombinant Panoramix-Nxf2-Nxt1 complex (Figure 4). We used chemical crosslinking coupled to mass spectrometry in order to establish an interaction map of residues that are in close physical proximity within the complex (Figure 7F; Figure S8C; Table S6). One set of identified crosslinks (black in Figure 7F) was in full agreement with our structural and biochemical data. (1) The two crosslinks involving Nxt1 map to the NTF2-like domain. (2) Within Panoramix, two crosslinking hotspots are apparent. The first hotspot corresponds to the amphipathic helix, the second one to the degron site. Both hotspots link to the C-terminus of Nxf2, indicating that the Panoramix-Nxf2 interaction is more complex than the helix-UBA interaction that we characterized (Figure 5). Strikingly, nearly all other identified protein crosslinks involve Nxf2's 1st RRM-LRR unit, while the second RRM-LRR unit was virtually devoid of interactions. The 1st RRM-LRR unit exhibited multiple intramolecular crosslinks within Nxf2 (towards the NTF2-like and UBA domains), as well as intermolecular crosslinks to Panoramix, mostly to the degron site. As the 1st RRM-LRR unit is fully dispensable for the Panoramix interaction (Figure 6), our findings are consistent with a model where in the non-target-engaged state, Nxf2's N-terminal RNA binding unit folds back onto the SFiNX complex. Upon recruitment to a target transcript, most likely via Piwi, Nxf2's N-terminal RRM-LRR domain would interact with RNA, thereby anchoring SFiNX via the nascent target transcript to chromatin (Figure S8D). Structural insight into the full SFiNX complex promises to shed light onto the molecular logic of this intriguing silencing complex.

DISCUSSION

The discovery of SFiNX, a nuclear protein complex consisting of Panoramix, Nxf2, and Nxt1/p15, provides the key molecular connection between Piwi, the nascent target RNA, and the cellular heterochromatin machinery. In the absence of SFiNX, piRNA-loaded Piwi is incapable of inducing co-transcriptional silencing (Figure 2). Conversely, experimental recruitment of SFiNX to a nascent RNA independently of Piwi results in potent silencing and local heterochromatin formation (Figure 3). Our data indicate that within SFiNX, Panoramix, but not Nxf2, provides the molecular link to the downstream cellular heterochromatin effectors (Figure 5). Based on genetic experiments, the histone methyl-transferase SetDB1/Eggless, the histone-demethylase Lsd1/Su(var)3-3, and the heterochromatin binding protein HP1/Su(var)205 are required for piRNA-guided co-transcriptional silencing (Sienski et al., 2015; Yu et al., 2015). We did not find any of these factors enriched in our SFiNX co-IP mass-spectrometry experiments. This suggests a transient or regulated molecular interaction between Panoramix and the cellular heterochromatin machinery. In this respect, the recent finding that the SUMOylation machinery is critically involved in piRNA-guided co-transcriptional silencing is of considerable interest (Ninova et al., 2019).

The involvement of Nxf2, a nuclear RNA export variant, in co-transcriptional silencing came as a surprise to us. Based on our biochemical and structural data, we propose that two molecular features of the ancestral NXF protein, the principal mRNA export receptor Nxf1/Tap, facilitated the evolutionary exaptation of Nxf2 into piRNA-guided silencing. First, Nxf2 retained its ability to bind RNA, thereby providing SFiNX a molecular link to the nascent target RNA (Figure 6). Second, our crosslink-mass spectrometry data indicate that, similar to Nxf1/Tap, the RNA binding activity of Nxf2 might be gated (Figure 7). We propose that this could provide a critical regulatory switch to ensure that SFiNX only associates with transcripts that are specified as targets via the Piwi-piRNA complex. In the case of Nxf1/Tap, various proteins (e.g. SR-proteins, THO-complex, UAP56, Aly/Ref) that are recruited during co-transcriptional mRNA maturation, are required to restrict Nxf1/Tap deposition onto export competent mRNAs only (Cullen, 2000; Heath et al., 2016; Kohler and Hurt, 2007). Whether any of these factors is also required for loading Nxf2 onto RNA is currently unclear. It is, however, likely that target-engaged Piwi contributes a critical role in the deposition of SFiNX onto the target RNA, potentially by licensing Nxf2's RNA binding activity. We note, that despite considerable experimental efforts we were not able to establish a direct molecular link between Piwi and either Nxf2 or Panoramix. Considering that Piwi represses transcription of its targets, the steady state level of a Piwi-SFiNX complex is expected to be small. In support of this, piRNA-independent

recruitment of Piwi to a nascent RNA is incapable of target silencing, indicating that the Piwi-SFiNX interaction occurs only once Piwi is bound to a target RNA via a complementary piRNA (Post et al., 2014; Sienski et al., 2015; Yu et al., 2015). An intriguing result from our Nxf2 CLIP experiments is that Nxf2 binds nascent, Piwi-repressed RNAs also far away from piRNA complementary sequences (Figure 7). One possible explanation for this is that Nxf2 or SFiNX, after initial Piwi-mediated recruitment, spread without Piwi along the target RNA. Alternatively, SFiNX might form oligomers in vivo, resulting in multiple RNA binding domains per complex that could tether SFiNX to distant parts of the target transcript. Such an 'entangling' of the target RNA at the transcription locus might serve an additional critical role, namely to keep the target RNA at chromatin, thereby providing sufficient time for the recruited effectors to modify chromatin at the locus.

Although no direct ortholog of Nxf2 is identifiable in vertebrates, our finding that an NXF variant is involved in transposon silencing in *Drosophila* could point to a more general scheme. The Nxf1/Tap ancestor diversified through several independent evolutionary radiations into numerous NXF variants in different animal lineages (Figure S8E; Table S7). In flies, the three NXF variants exhibit gonad-specific expression with Nxf2 and Nxf3 being expressed predominantly in ovaries, and Nxf4 being testis-specific (Gramates et al., 2017). Besides Nxf2 (this study), *Drosophila* Nxf3 is also an essential piRNA pathway component as it is required for the nuclear export of un-processed piRNA cluster transcripts in the germline (ElMaghraby et al., 2019). In mice and humans, several NXF variants are preferentially expressed in testes. Intriguingly, Nxf2 mutant mice are male sterile, a phenotype shared with many piRNA pathway mutants (Pan et al., 2009). Considering this, we speculate that also in vertebrates the host-transposon conflict has been a key driver of NXF protein family evolution through frequent duplication and exaptation events. Our study highlights that some of these variants might have evolved novel functions, not directly related to RNA export biology.

ACKNOWLEDGMENTS

We thank K. Meixner for experimental support, P. Duchek and J. Gokcezade for generating CRISPR edited and transgenic flies, the VBCF NGS unit for deep sequencing, VBCF Protein Technologies Facility for recombinant protein expression, the MFPL monoclonal facility for Nxf2 ans Panoramix antibodies, and the VDRC, TRiP, and Bloomington stock centers for flies. We thank A. Koehler and G. Riddihough (Life Science Editors; http://lifescienceeditors.com) for comments on the manuscript. We thank the Brennecke lab, particularly P. Andersen, for support and feedback. The Brennecke lab is supported by the Austrian Academy of Sciences, the European Community (ERC-2015-CoG - 682181), and the Austrian Science Fund (F 4303). J. Batki was supported by the Boehringer Ingelheim Fonds. X-ray diffraction studies were conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by NIGMS grant P30 GM124165 and U.S. Department of Energy grant DE-AC02-06CH11357. The Pilatus 6M detector on 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). MSKCC core facilities are supported by P30 CA008748. This work was supported by funds from the Maloris Foundation (DJP) and MSKCC core grant (P30 CA008748).

AUTHOR CONTRIBUTIONS

J.Batki, J.S., A.I.V, L.L., and K.K. performed all molecular biology and fly experiments. D.H., J.S., and J.Batki performed the computational analyses, C.S. and K.M. performed the X-link mass spectrometry analysis, M.N. generated the phylogenetic comparisons of NXF proteins. J.W. generated, purified and grew crystals of the UBA-linker-helix and the Nxf2 NTF2l-Nxt1 complex, performed the X-ray crystallographic analyses, and performed the GST pull-down experiments under the supervision of D.J.P. The paper was written by J.Batki, J.S. and J.B. with input from J.W. and D.J.P.

Competing financial interests

The authors declare no competing financial interests.

MATERIALS & METHODS

Fly strains

All fly strains used in this study are listed in Table S8. Flies were kept at 25 °C. For each experiment, flies were aged for 5-6 days and kept on apple juice agar plates with yeast paste to ensure consistent ovarian morphology. Two independent *nxf2* frameshift mutant alleles were generated by injecting the pDCC6 plasmid with *nxf2* targeting gRNAs (Table S9) into *w[1118]* flies (Gokcezade et al., 2014). Sequences for the two frameshift alleles are indicated in Extended Data Fig. 1c. N-terminal 3xFLAG_V5_GFP-tagging of endogenous *nxf2* and *panoramix* loci was done by co-injecting a repair template and the gRNA containing plasmid (Addgene 45956) into act-Cas9 flies (BL-58492). Oligonucleotides used for gRNA cloning are listed in Table S9.

Germline and soma specific gene knockdowns were performed by crossing short hairpin (shRNA) transgene strains with the maternal triple driver (MTD)-GAL4 line or the *traffic jam*-GAL4 driver line, respectively. Oligonucleotides used for shRNA cloning are listed in Table S9. *gypsy* and *burdock*-LacZ sensor strains are described in (Handler et al., 2013).

Rescue strains with different Panoramix and Nxf2 variants were generated by injecting respective rescue transgenes into *panoramix* or *nxf2* mutant flies containing attp landing sites on the same chromosome (attP40 for *panoramix*, attP154 for *nxf2*). Rescue transgenes contained the *panoramix* regulatory control regions (chr2R: 21,308,437-21,313,490) and the *panoramix* coding sequence was replaced by the various *panoramix* and *nxf2* variants.

X-gal staining of ovaries

Ovaries were dissected into ice cold PBS, fixed in 0.5% glutaraldehyde (in PBS) for 15 min at room temperature, and then

washed twice with PBS. Next, samples were incubated in staining solution (10 mM PBS, 1 mM MgCl2, 150 mM NaCl, 3 mM potassium ferricyanide, 3 mM potassium ferrocyanide, 0.1% Triton X-100, 0.1% X-gal (5-bromo-4-chloro-3-indolyl-β -d-galactoside)) overnight (*gypsy*-sensor) or for 2h (*burdock*-sensor) at room temperature.

Generation of Nxf2 and Panoramix antibodies

Purified His-tagged Nxf2 (1-326) protein was used to generate the mouse anti-Nxf2 antibody used for western blot. The mouse anti-Nxf2 and anti-Panoramix antibodies used for immunofluorescence were raised against the SFiNX complex consisting of his-tagged Nxf2 (541-841), Strep-tagged-Panoramix(263-446) and Flag-tagged Nxt1 (full length). All antibodies were generated at the MFPL Monoclonal Antibody Facility.

OSC cell culture

OSCs were cultured as described (Niki et al., 2006; Saito et al., 2009). Plasmid and siRNA transfections were performed using Cell Line Nucleofector kit V (Amaxa Biosystems) with the program T-029, using 8 million cells per transfection. siRNAs used in this study are listed in Table S10.

Stable OSC reporter line generation

The reporter construct (*traffic jam* enhancer driven GFP_P2A-Blasticidin-resistance harboring 10 intronic boxB sites and 14 upstream UAS sites; plasmid submitted to Addgene) was integrated into chromosomal location chr2L:9,094,918, which is devoid of genes and major chromatin marks, using CRISPR-Cas9. In brief, 600bp long homology arms flanking the integration site were amplified from OSC genomic DNA. Oligonucleotides targeting the locus (Table S9) were cloned into the guide RNA expression plasmid (Addgene 49330). Two independent gRNA containing plasmids were mixed 1:1 and 200 ng of this mix were co-transfected with 1200 ng of the integration plasmid into OSCs. After two days, the cells were plated with different dilutions and on the following day Blasticidin containing media was added (1:1000) for a 4-day long selection. Afterwards, the cells were grown in normal medium for about 2 weeks until individual clones could be isolated.

Droplet PCR

To assess the copy number of the reporter construct integrated in the OSC genome, the QX200™ Droplet Digital™ PCR System (BIORAD) was used according to the manufacturer's instructions. In brief, genomic OSC DNA was digested with EcoRI and HindIII restriction enzymes. The PCR reaction was set up with 10ng digested genomic DNA (primer sequences in Table S9) and the QX200™ ddPCR™ EvaGreen Supermix. The PCR mix and the QX200 Droplet Generation Oil for EvaGreen were added into a DG8™ cartridge and droplets were generated with QX200 Droplet generator. Thermal cycling and droplet reading were performed with the instructor's standard protocol which gave the concentration of the amplicon in copies/reaction volume. Based on 2 house-keeping control genes, the copy number in the genome was determined for the integrated reporter.

Stable OSC line generation with extra genomic copy

The pAcm vector (Saito et al., 2009) was modified to create the integration constructs. Downstream of the *act5C* promoter, a 3xFLAG-HA tag was added followed by the open reading frames encoding full length Panoramix or Nxf2. The selection cassette consisted of an independent transcription unit driving mCherry_P2A_Puromycin-resistance via the *traffic jam* enhancer from *Drosophila yakuba* in combination with the *Drosophila* synthetic core promoter (DSCP) (Pfeiffer et al., 2008). The two transgenes were integrated into the chromosomal location chr2L:9,103,945, which is devoid of genes. 1200 bp long homology arms flanking the integration sequence were used for the integration. Stable integration was generated as described above for the reporter cell line, except that Puromycin was used for the selection (1:2000).

Clonal OSC line generation expressing endogenously tagged Nxf2

To generate the endogenously tagged Nxf2 OSC cell line CRIPSR-Cas9 was used to insert a GFP-Prescission-V5-3xFlag tag followed by P2A Puromycin resistance gene at the C-terminus of the Nxf2 gene. As a repair template a PCR product was generated containing the tag sequence flanked by ~500 bp long homology arms surrounding the C-terminus of Nxf2 using biotinylated primers. Oligonucleotides targeting the Nxf2 locus were cloned into a guide RNA expression plasmid containing Cas9 fused to monomeric streptavidin as described (Gu et al., 2018). 1500 ng purified PCR product together with 500 ng of guide RNA containing plasmid were co-transfected into OSCs and after 3 days puromycin containing media was added (1:1000) for a 4-day long selection. Afterwards, the cells were grown in normal medium for about 2 weeks until individual clones could be isolated.

Tethering reporter assay

All tethering constructs are based on λN-entry or Gal4-entry vectors (submitted to Addgene). Various full-length CDS or CDS variants were inserted into the entry vectors to generate N-terminally tagged fusion proteins. Unless having full length genes, the SV40 NLS sequence (PKKKRKV) was included to ensure nuclear localization of the variants. A separate expression cassette driving mCherry via the *traffic jam* enhancer was used to select positively transfected cells.

OSCs harboring stably integrated GFP reporter were transfected with 4μg of λN/Gal4 fusion construct. As a negative control, λN/Gal4 empty vector was used. Two days after transfection, cells were harvested for WB analysis and four days after transfection cells were harvested for flow cytometry analysis using a FACS BD LSR Fortessa (BD Biosciences). Transfected cells were gated based on mCherry expression and the GFP intensity was determined in that population (per experiment 2500 cells). Data analysis was performed using FACS Diva and FlowJo.

OSC rescue assay

OSCs were co-transfected with siRNAs targeting *panoramix* or *nxf2* and a plasmid containing the *act5c* driven siRNA-resistant rescue construct. A second transfection was performed after two days and cells were collected after four days for WB analysis and RNA isolation for RT-qPCR. siRNAs used for the rescue experiments are listed in Table S10.

RT-qPCR

OSCs or 5-10 pairs of ovaries were collected into TRIzol reagent and RNA was isolated according to the manufacturer's instructions. Total RNA was digested with RQ1 RNase-Free DNase (Promega) and cDNA was prepared using random hexamer oligonucleotides and Superscript II (Invitrogen). Primers used for qPCR analysis are listed in Table S9.

Immunofluorescence staining of OSCs

2 days following transfection, cells were plated on concavalin A coated coverslips. After 4 hours, cells were fixed with formaldehyde solution (4% formaldehyde in PBS) for 15 min at room temperature. Fixed cells were washed twice with PBS for 5 min, permeabilized with PBX (0.1 % Triton X-100 in PBS) for 10 min and washed again with PBS for 5 min. Blocking was done in BBS (1 % BSA in PBS) for 30 min and the primary antibody was diluted in BBS and incubated ON at 4°C. Following three washing steps with PBS, the fluorophore-conjugated secondary antibody was diluted in BBS and cells were incubated with it for 1 hour at room temperature in the dark. The stained cells were washed three times with PBS, the second wash containing DAPI. The mounted samples were imaged with a Zeiss LSM-780 confocal microscope and the images were processed using FIJI/ImageJ. Antibodies are listed in Table S11.

Immunofluorescence staining of ovaries

After dissecting ovaries into ice cold PBS (max 30 min), ovaries were fixed with 4% formaldehyde and 0.3 % Triton X-100 in PBS for 20 min at room temperature. Fixed ovaries were washed 3x with PBX (0.3 % Triton X-100 in PBS) for 10 min and blocked in BBX (1 % BSA and 0.3 % Triton X-100 in PBS) for 30 min. Primary antibody was diluted in BBX and ovaries were incubated with it 24 hours at 4 °C. Following three washing steps with PBX, the fluorophore-conjugated secondary antibody was diluted in BBX and ovaries were incubated with it for ON at 4°C in the dark. The stained ovaries were washed three times with PBX, the second wash containing DAPI. The mounted samples were imaged with a Zeiss LSM-780 confocal microscope and the images were processed using FIJI/ImageJ. Antibodies are listed in Table S11.

RNA Fluorescence In Situ Hybridization (FISH)

mdg1 RNA FISH on ovaries was performed as described (Mohn et al., 2014) using CAL Fluor Red 590-labeled Stellaris oligo probes (Table S12). After the RNA FISH protocol, egg chambers were blocked with SBX (1% BSA; 0.1% Triton X-100; 2xSSC) for 30 min and then incubated with primary anti-GFP antibody (Abcam) for 24h at 4°C. After 3x washing (10 min with SBX), samples were incubated with fluorescent secondary antibody for 12h at 4°C. Stacks of soma nuclei were imaged on a Zeiss LSM780 confocal microscope and a maximum intensity projection of 3 slices was generated.

Small RNA-seq

Total RNA was isolated with TRIzol reagent according to the manufacturer's instructions, and 2S rRNA was depleted as described (Hayashi et al., 2016). Small RNA libraries were generated as described (Jayaprakash et al., 2011). In brief, using radio-labelled oligonucleotides as size-markers, 18 to 29nt long RNAs were purified by PAGE. The 3′ linker (containing four random nucleotides) was ligated with T4 RNA ligase 2, truncated K227Q (NEB) overnight at 16°C. Following PAGE purification, the 5′ linker (containing four random nucleotides) was ligated to the small RNAs using T4 RNA ligase (NEB) overnight at 16°C. After PAGE purification, the linker-ligated RNAs were reverse transcribed and PCR amplified. Sequencing was performed with HiSeq2500 (Illumina) in single-read 50 mode.

Small RNA-seq analysis

Sequencing reads were trimmed by removal of the adaptor sequences and the four random nucleotides flanking the small RNA. These reads were pre-mapped to the *Drosophila melanogaster* rRNA precursor, the mitochondrial genome and unmapped reads were mapped to the *Drosophila melanogaster* genome (dm6), all using Bowtie (Langmead et al., 2009) (release 1.2.2) with 0 mismatch allowed. Genome mapping reads were intersected with Flybase genome annotations (r6.18) using Bedtools (Quinlan and Hall, 2010) (2.27.1). Reads mapping to rRNA, tRNA, snRNA, snoRNA loci and the mitochondrial genome were removed from the analysis. The quantification of small RNAs was carried out as described in (Andersen et al., 2017). with the following modifications: as a minimal count per 1kb tile cutoff, a value which includes 80% of all reads in the control libraries was used (98 for OSC-KD, 29 for GLKD). Tiles with a mappability below 20% were excluded from the analysis. Annotation groups were based on RefSeq assembly release 6. Tiles overlapping with genes and piRNA clusters were annotated as genic and respective cluster, tiles without annotation were grouped as 'other'. All sequenced libraries with their GEO Accession number are listed in Table S13.

RNA-seq with rRNA depletion

We modified the protocol published in (Morlan et al., 2012). Total RNA was isolated with TRIzol reagent, which was further purified by RNAeasy columns with on-column DNase I digest (Qiagen), all according to the manufacturer's instructions. Depletion of rRNA from the purified total RNA was done by using a mix of antisense oligonucleotides matching *Drosophila melanogaster* rRNAs (listed in Table S14) and the Hybridase Thermostable RNase H (Epicentre) which specifically degrades RNA in RNA-DNA hybrids. The oligonucleotides were added to the RNA in RNase H Buffer (20 mM Tris-HCl pH=8, 100 mM NaCl) and annealed with

a temperature gradient from 95 °C to 45 °C. The hybrids were digested at 45 °C for 1 hour. Next, DNA was digested with TURBO DNase (Invitrogen) and RNA was purified using RNA Clean & Concentrator-5 (Zymo) according to the manufacturer's instructions. Libraries were prepared using a NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) according to the protocol and sequenced on a HiSeq2500 (Illumina) in single-read 50 mode.

RNA-seq with polyA selection

Total RNA was isolated with TRIzol reagent. Poly(A)+ RNA enrichment was performed with Dynabeads Oligo(dT)$_{25}$ (Thermo Fisher), with two consecutive purifications according to the manufacturer's instructions. Next, cDNA was prepared using NEBNext Ultra II RNA First and Second Strand Synthesis Module. The cDNA was purified with AmpureXP beads and library was prepared with NEBNext Ultra II DNA Library Prep Kit Illumina (NEB) according to the protocol and sequenced on a HiSeq2500 (Illumina) in single-read 50 mode.

RNA-seq analysis

Sequencing reads were trimmed by removal of the adaptor sequences. Reads were mapped to the *Drosophila melanogaster* rRNA precursor and the mitochondrial genome using Bowtie (Langmead et al., 2009) (release 1.2.2) with 0 mismatches allowed. Remaining reads were mapped to the *Drosophila melanogaster* genome (dm6) using STAR (Dobin et al., 2013) (v.2.5.2b; settings: --outSAMmode NoQS --readFilesCommand cat --alignEndsType Local --twopassMode Basic --outReadsUnmapped Fastx --outMultimapperOrder Random --outSAMtype SAM --outFilterMultimapNmax 1000 --winAnchorMultimapNmax 2000 --outFilterMismatchNmax 0 --seedSearchStartLmax 30 --alignSoftClipAtReferenceEnds No --outFilterType BySJout --alignSJoverhangMin 15 --alignSJDBoverhangMin 1). Genome mapping reads were intersected with Flybase genome annotations (r6.18) using Bedtools (Quinlan and Hall, 2010) (2.27.1). Reads mapping to rRNA, tRNA, mitoRNA were excluded from further analysis.

Differential gene expression analysis

Genome matching reads were randomized in order and quantified using Salmon (Patro et al., 2017) (v.0.10.2; settings: --dumpEqWeights --seqBias --gcBias --useVBOpt --numBootstraps 100 -l SF --incompatPrior 0.0 --validateMappings). Salmon results were further processed using wasabi (https://github.com/COMBINE-lab/wasabi commitID=478c133). DGE analysis was performed pairwise between libraries using sleuth (Pimentel et al., 2017) (v0.30.0; settings: extra_bootstrap_summary = TRUE transform_fun_tpm = function(x) log2(x + 0.5), read_bootstrap_tpm = TRUE, gene_mode = TRUE) and running the wald-test function. The sleuth model is a measurement error in the response model. It attempts to segregate the variation due to the inference procedure by Salmon from the variation due to the covariates -- the biological and technical factors of the experiment. For the Wald test, the effect-size represents the estimate of the selected coefficient. It is analogous to, but not equivalent to, the fold-change. The transformed values are on the log2 scale, thus the estimated coefficient is also on the log2 scale. This value takes into account the estimated 'inferential variance' estimated from the Salmon bootstraps. For TEs and mRNAs, we required a minimum of TPM >5 in any of the analyzed libraries.

ChIP-seq

Chromatin immunoprecipitation (ChIP) was carried out according to (Lee et al., 2006), with minor modifications. In brief, OSCs were crosslinked with 1% formaldehyde, quenched with glycine, washed with PBS, collected by centrifugation and pellets were flash-frozen in liquid nitrogen. Chromatin was prepared using Lysis Buffer 1, 2 and 3 from (Lee et al., 2006) and sonication was performed with a Covaris E220 Ultrasonicator for 20 min. For immunoprecipitation, anti H3K9me3 and RNA Pol II antibodies (Table S11), were coupled to Protein G and Protein A Dynabeads, respectively. Sheared chromatin was incubated with the bead-coupled antibodies for 4 hours at 4 °C, beads were washed, and elution plus de-crosslinking was performed at 65 °C overnight. Following RNase A and proteinase K treatment, DNA was purified with ChIP DNA Clean & Concentrator Kit (Zymo). ChIP-qPCR was performed to test the efficiency of the ChIP and libraries were prepared with NEBNext Ultra DNA Library Prep Kit Illumina (NEB) according to the protocol, and sequenced on a HiSeq2500 (Illumina) in single-read 50 mode.

ChIP-seq analysis

Sequencing reads were trimmed by removal of the adaptor sequences and filtered for a minimal length of 18 nucleotides. Reads were mapped to the *Drosophila melanogaster* rRNA precursor, the mitochondrial genome and the genome (dm6) using Bowtie (Langmead et al., 2009) (release 1.2.2), all_with 0 (genome wide analysis) or 3 (TE-consensus analysis) mismatches allowed. BigWig files were generated using Homer (Heinz et al., 2010) and UCSC BigWig tools (Kent et al., 2010). Heatmaps and meta profiles were generated with Deeptools within Galaxy using BigWig files. The genomic coordinates of euchromatic TE insertions were determined in (Sienski et al., 2012) and the same Piwi-regulated TEs were used as in (Sienski et al., 2015). To calculate log2 fold change values relative to control knockdown, bigwigCompare was used with a pseudo-count of 1. To determine Piwi-dependent H3K9me3 regions, the quantification of ChIP-seq reads was carried out as described (Andersen et al., 2017) with the following modifications: As a minimal count per tile cutoff a value of 150 reads was used and tiles with a mappability below 20% were excluded from the analysis. Piwi-dependent regions were classified by a log2 fold change > 2 when comparing control knockdown with Piwi knockdown.

For TE consensus analysis, genome mapping reads longer than 23 nucleotides were mapped to TE consensus sequences using STAR (Dobin et al., 2013) (v.2.5.2b; settings: --outSAMmode NoQS --

readFilesCommand cat --alignEndsType Local --twopassMode Basic --outReadsUnmapped Fastx --outMultimapperOrder Random --outSAMtype SAM --outFilterMultimapNmax 1000 --winAnchorMultimapNmax 2000 --outFilterMismatchNmax 3 --seedSearchStartLmax 30 --outFilterType BySJout --alignSJoverhangMin 15 --alignSJDBoverhangMin 1). Multiple mappings were only allowed within one transposon and read-counts were divided equally to the mapping positions. For plotting, read-counts were normalized to 10 million sequenced reads, converted to bedgraph tracks using Bedtools (2.27.1) (Quinlan and Hall, 2010) and plotted in RStudio. All sequenced libraries with their GEO Accession numbers are listed in Table S13.

CLIP-seq

Cross-linking immunoprecipitation (CLIP) was performed using wild type OSCs or OSCs with endogenously GFP-tagged Nxf2. Cells were irradiated twice with 200 mJ/cm2 UV light (254 nm) on ice in Stratalinker 2400 (Stratagene), scraped into ice cold PBS, collected by centrifugation, then washed with PBS, and centrifuged again. The cell pellets were resuspended in lysis buffer (LB: 30 mM Tris-HCl pH=7.5, 150 mM NaCl, 2 mM $MgCl_2$, 0,5 % Triton X-100, 5 % glycerol, freshly supplemented with Complete Protease Inhibitor Cocktail (Roche)), incubated at 4°C for 20 min, then sonicated with Diagenode Bioruptor (using the medium mode for 10 mins with 15 s ON and 15 s OFF) followed by a centrifugation step. Total cell lysates were treated with RNAse A (Thermo Fisher Scientific) in 1 : 500 000 dilution for 5 min at 37 °C, then placed on ice and stopped with SUPERase In RNase inhibitor (Thermo Fisher Scientific). Immunoprecipitation was performed with GFP-Trap magnetic beads (ChromoTek) for 2h at 4°C. The following washing steps (WASH) were performed: 3x 10 min with LB, 3x 10 min with denaturing wash buffer (DWB: 5M urea, 0,5 % SDS), 3x 10 min with low salt buffer (LSB: 150 mM NaCl, 0,5 % Triton X-100). Next, the beads were washed 3x 5 min with CutSmart Buffer (1X, NEB) and dephosphorylation was performed with calf intestinal phosphatase (NEB) for 30 min at 37 °C. The beads were washed again (WASH), followed by 3x 5 min wash with T4 PNK Reaction Buffer (1X, NEB) and phosphorylation was performed with T4 PNK (NEB) for 30 min at 37 °C in the presence of 1mM ATP. The beads were washed (WASH), followed by 3x 5 min wash with T4 RNA ligase buffer (1X, NEB) and the 3′ linker (containing four random nucleotides, same as for small RNA-seq) was ligated with T4 RNA ligase 2, truncated K227Q (NEB) for 2 h at 25°C. Next, the beads were washed (WASH), followed by 3x 5 min wash with T4 RNA ligase buffer (1X, NEB) and the 5′ linker (containing four random nucleotides, same as for small RNA-seq) was ligated with T4 RNA ligase 1 (NEB) for 2 h at 25°C, in the presence of 1mM ATP. Finally, the beads were washed (WASH), followed by 3x 5 min wash with proteinase K buffer (100 mM Tris-HCl pH=7.5, 50 mM NaCl, 10 mM EDTA). The RNA was eluted from the beads by proteinase K digest (1 mg/ml) for 1 h at 37°C and ethanol precipitated. The linker-ligated RNAs were reverse transcribed using SuperScript II (Thermo Fisher Scientific) and PCR amplified. Sequencing was performed with NextSeq550 (Illumina) in PE75 medium mode.

CLIP-seq analysis

Sequencing reads were trimmed by removal of the adaptor sequences and filtered for a minimal length of 22 nucleotides. Reads were mapped to the *Drosophila melanogaster* rRNA precursor, the mitochondrial genome using Bowtie(Langmead et al., 2009) (release 1.2.2) with 3 mismatches allowed and to the genome (dm6) using STAR (Dobin et al., 2013) (v.2.5.2b; settings: --outFilterMatchNmin 20 --outSAMmode NoQS --readFilesCommand cat --alignEndsType Local --twopassMode Basic --outReadsUnmapped Fastx --outMultimapperOrder Random --outSAMtype SAM --outFilterMultimapNmax 1000 --winAnchorMultimapNmax 2000 --outFilterMismatchNoverLmax 0.05 --seedSearchStartLmax 30 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --alignIntronMax 1 --outFilterType BySJout --alignSJoverhangMin 15 --alignSJDBoverhangMin 1).

For TE consensus histograms, genome mapping reads longer than 22 nucleotides were mapped to TE consensus sequences using STAR (Dobin et al., 2013) (v.2.5.2b; settings: --outSAMmode NoQS --readFilesCommand cat --alignEndsType Local --twopassMode Basic --outReadsUnmapped Fastx --outMultimapperOrder Random --outSAMtype SAM --outFilterMultimapNmax 1000 --winAnchorMultimapNmax 2000 --outFilterMismatchNmax 3 --seedSearchStartLmax 30 --outFilterType BySJout --alignSJoverhangMin 15 --alignSJDBoverhangMin 1). Multiple mappings were only allowed within one transposon and read-counts were divided equally to the mapping positions.

For plotting, read-counts were normalized to 10 million uniquely mapping reads, converted to bedgraph tracks using Bedtools (2.27.1) (Quinlan and Hall, 2010) and plotted in RStudio.

TE expression quantification was performed using Salmon (Patro et al., 2017) (v.0.10.2; settings: --seqBias --gcBias --useVBOpt -l SF --incompatPrior 0.0 –validateMappings) using a transcript set containing all FlyBase transcripts and the TE consensus sequences in sense and antisense.

To determine regions enriched for Nxf2, the quantification of CLIP-seq reads was carried out as described (Andersen et al., 2017) with the following modifications: As a minimal count per tile cutoff a value of 200 reads was used and tiles with a mapability below 20% were excluded from the analysis. To calculate log2 fold change values relative to control CLIP, a pseudo-count of 5 was added to all values.

The genomic coordinates of euchromatic TE insertions were determined in (Sienski et al., 2012) and the same Piwi-regulated TEs were used as in (Sienski et al., 2015). The determine the number of Nxf2 enriched tiles close to TE insertions the top 250 tiles got intersected with the TE coordinates using bedtools (2.27.1) (Quinlan and Hall, 2010). As a control, the tiles got sorted randomly 250 times (sort -R --random-source=/dev/urandom) and the first 250 tiles of this order were intersected with the TE annotations.

BigWig files were generated using Homer (Heinz et al., 2010) and UCSC BigWig tools (Kent et al., 2010). Heatmaps and meta profiles were generated by quantifying CLIP enrichment in the vicinity of TEs using Homer (makeTagDirectory TagDirectory IN.bed -precision 3 -force5th -fragLength given -format bed; annotatePeaks.pl TEpositions.bed dm6.fa -strand + -ghist -hist 100 -size -10000,10000 -normLength 0 -len 0 -noadj -d TagDirectory) and plotted in R (3.5.2) using ggplot2 (3.1.1) (Wickham, 2016). All sequenced libraries with their GEO Accession numbers are listed in Table S13.

Protein co-immunoprecipitation from nuclear OSC cell lysates

OSCs were collected after trypsinization by centrifugation, washed with PBS and centrifuged again. The cell pellet was resuspended in LB1 (10 mM Tris-HCl pH=7.5, 2 mM $MgCl_2$, 3 mM $CaCl_2$, freshly supplemented with Complete Protease Inhibitor Cocktail (Roche)), incubated at 4°C for 10 min followed by a centrifugation step. The pellet was resuspended in LB2 (10 mM Tris-HCl pH=7.5, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 0,5 % IGEPAL CA-630, 10 % glycerol, freshly supplemented with Complete Protease Inhibitor Cocktail (Roche)), incubated at 4°C for 10 min followed by a centrifugation step. The isolated nuclei were lysed in LB3 (50 mM Tris-HCl pH=8, 150 mM NaCl, 2 mM $MgCl_2$, 0,5 % Triton X-100, 0,25 % IGEPAL CA-630, 10 % glycerol, freshly supplemented with Complete Protease Inhibitor Cocktail (Roche)), incubated at 4°C for 20 min followed by a centrifugation step. Nuclear lysate was used for immunoprecipitation with Flag M2 Magnetic Beads (Sigma) for 2h at 4°C. The beads were washed 3x 10 min with LB3 and were either used for mass spectrometry analysis or the proteins were eluted in 1× SDS buffer with 5 min incubation at 95°C for western blotting.

Protein co-immunoprecipitation from S2 cell lysates

S2 cells were transfected using Cell Line Nucleofector kit V (Amaxa Biosystems) with the program G-030, using 8 million cells per transfection. S2 cells were co-transfected with FLAG-tagged and GFP-tagged protein encoding plasmids. After two days, cells were collected by centrifugation, washed with PBS and collected again. The cell pellet was resuspended in LB (30 mM Tris-HCl pH=7.5, 150 mM NaCl, 2 mM $MgCl_2$, 0,5 % Triton X-100, 10 % glycerol, freshly supplemented with Complete Protease Inhibitor Cocktail (Roche)), incubated at 4°C for 20 min followed by a centrifugation step. The total cell lysate was used for immunoprecipitation with GFP-Trap magnetic beads (ChromoTek) for 2h at 4°C. The beads were washed 3x 10 min with LB and the proteins were eluted in 1× SDS buffer with 5 min incubation at 95°C.

Western blot

Proteins were separated by SDS–polyacrylamide gel electrophoresis (PAGE) and transferred to a 0.2 µm nitrocellulose membrane (Bio-Rad). The membrane was blocked with 5% milk in PBX (0.05 % Triton X-100 in PBS) and were incubated with primary antibody ON at 4°C. After three washes with PBX, the membrane was incubated with HRP-conjugated secondary antibody for 1h, followed by three PBX washes. The membrane was incubated with Clarity Western ECL Blotting Substrate (Bio-Rad) and imaged with a ChemiDoc MP imaging system (Bio-Rad). Antibodies are listed in Table S11.

Mass spectrometry analysis

Co-immunoprecipitated proteins coupled to magnetic beads were digested with LysC on the beads, eluted with glycine followed by trypsin digestion. Peptides were analyzed using an UltiMate 3000 RSLCnano System (Thermo Fisher Scientific) coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific), equipped with a Proxeon nanospray source (Thermo Fisher Scientific). Peptides were loaded onto a trap column (Thermo Fisher Scientific, PepMap C18, 5 mm × 300 µm ID, 5 µm particles, 100 Å pore size) at a flow rate of 25 µL/min using 0.1% TFA as mobile phase. After 10 min, the trap column was switched in line with the analytical column (Thermo Fisher Scientific, PepMap C18, 500 mm × 75 µm ID, 2 µm, 100 Å). Peptides were eluted using a flow rate of 230 nl/min and a binary 3h gradient. The gradient starts with the mobile phases: 98% A (water/formic acid, 99.9/0.1, v/v) and 2% B (water/acetonitrile/formic acid, 19.92/80/0.08, v/v/v), increases to 35%B over the next 180 min, followed by a gradient in 5 min to 90%B, stays there for 5 min and decreases in 2 min back to the gradient 98%A and 2%B for equilibration at 30°C.

The Q Exactive HF mass spectrometer was operated in data-dependent mode, using a full scan (m/z range 380-1500, nominal resolution of 60,000, target value 1E6) followed by MS/MS scans of the 10 most abundant ions. MS/MS spectra were acquired using normalized collision energy of 27, isolation width of 1.4 m/z, resolution of 30.000 and the target value was set to 1E5. Precursor ions selected for fragmentation (exclude charge state 1, 7, 8, >8) were put on a dynamic exclusion list for 60 s. Additionally, the minimum AGC target was set to 5E3 and intensity threshold was calculated to be 4.8E4. The peptide match feature was set to preferred and the exclude isotopes feature was enabled.

For peptide identification, the RAW-files were loaded into Proteome Discoverer (version 2.1.0.81, Thermo Scientific). All hereby created MS/MS spectra were searched using MSAmanda v2.1.5.9849, Engine version v2.0.0.9849 (Dorfer et al., 2014). For the first step search the RAW-files were searched against *Drosophila melanogaster* reference translations retrieved from Flybase (dmel_all-translation-r6.13; 21,983 sequences; 20,112,742 residues), using the following search parameters: The peptide mass tolerance was set to ±5 ppm and the fragment mass tolerance to 15ppm. The maximal number of missed cleavages was set to 2. The result was filtered to 1 % FDR on protein level using Percolator algorithm integrated in Thermo Proteome Discoverer. A sub-database was generated for further processing. Peptide areas were quantified using an in-house developed tool APQuant: http://ms.imp.ac.at/index.php?action=peakjuggler (Doblmann et al., 2018).

Protein expression and purification

The UBA domain of *Drosophila melanogaster* Nxf2 (residues

781-841) and Panoramix helix (residues 311-340) were covalently linked through a KLGSHM linker in one expression cassette. The NTF2-like domain of *Drosophila melanogaster* Nxf2 (residues 573-777, NTF2l) and full-length Nxt1 (residues 1-133) were cloned into a modified RSFduet-1 vector (Novagen) with an N-terminal His6-SUMO tag on the NTF2-like domain and no tag on Nxt1. Proteins were expressed in *E. coli* strain BL21(DE3) RIL (Stratagene). The cells were grown at 37°C until OD600 reached 0.8, then the media was cooled to 16°C and IPTG (isopropyl β-D-1-thiogalactopyranoside) was added to a final concentration of 0.35 mM to induce protein expression overnight at 16°C. The cells were harvested by centrifugation at 4°C and disrupted by sonication in Binding buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole) supplemented with 1 mM PMSF (phenylmethylsulfonyl fluoride) and 3 mM β-mercaptoethanol. After centrifugation, the supernatants were loaded onto 5 ml HisTrap Fastflow column (GE Healthcare). After extensive washing with Binding buffer, the complex was eluted with Binding buffer supplemented with 500 mM imidazole. The His6-SUMO tag was removed by Ulp1 protease digestion during dialysis against Binding buffer and separated by reloading onto HisTrap column. The flow-through fraction was further purified by HiTrap Q FF column and Superdex 75 16/60 column (GE Healthcare). The pooled fractions were concentrated to 35 mg/ml (UBA-linker-helix) and 10 mg/ml in crystallization buffer (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM DTT).

Crystallization, data collection and structure determination

As the complex of the UBA domain with the Panoramix helix was not very stable and the yield was very low by co-expression, we covalently linked the UBA domain and the Panoramix helix in one cassette by linkers of different length ($KL(GS)_nHM$, n=1, 2, 4, 6). Only the construct with the six-residue KLGSHM linker produced crystals. Crystals of the UBA-linker-helix were grown in 0.095 M sodium citrate pH 5.6, 19% (v/v) Isopropanol, 19% (w/v) PEG 4000, 5% (v/v) glycerol. Crystals of the NTF2-like domain-Nxt1 complex were grown from a solution containing 0.1 M MES pH 6.5, 1.6 M magnesium sulfate using the hanging drop vapor diffusion method at 20°C. For data collection, the crystals were flash frozen (100 K) and collected on NE-CAT beam lines 24ID-C and 24ID-E at the Advanced Photo Source (APS), Argonne National Laboratory. The diffraction data of both, the UBA-linker-Panoramix helix and the NTF2l-Nxt1 complex were processed with the NECAT RAPD online server and iMosfilm (Battye et al., 2011), respectively. The structures of the UBA-linker-Panoramix helix and the NTF2l-Nxt1 complex were solved by molecular replacement (MR) in PHENIX (Adams et al., 2002) using the structure of the UBA domain of human NXF1/TAP in complex with FxFG peptide (PDB ID: 1OAI) (Grant et al., 2003) and the structure of the human NXF1/TAP(NTF2l)-NXT1 complex (PDB ID: 1JKG) (Fribourg et al., 2001) as search templates. The automatic model building was carried out using the program PHENIX AutoBuild (Adams et al., 2002). The resulting model was refined by PHENIX refinement (Adams et al., 2002) and Refmac5 (Murshudov et al., 1997), and completed manually using COOT (Emsley et al., 2010). The statistics of the diffraction and refinement data are summarized in Table S6. All the molecular graphics were generated with the PyMOL program (https://pymol.org/2/).

GST pull-down assay

The C-terminal fragment of *Drosophila melanogaster* Nup358 (residues 2395-2426), which contains three FG-repeats, was fused to GST. 40 µg of GST-FG-repeat were incubated with 50 µL Glutathione Sepharose 4B beads (GE heathcare) at 4°C for 3 h. The beads were washed twice with pull-down buffer (20 mM Tris pH 7.5, 100 mM NaCl, 2 mM DTT and 2 mg/ml BSA) and then incubated with 100 µg protein complexes for 4 h at 4°C. An additional two washes were applied using pull-down buffer without BSA. Each sample was analyzed with SDS-PAGE followed by Coomassie staining.

Recombinant protein expression in insect cells

To generate a Panoramix (263-446) / Nxf2 (full length) / Nxt1 (full length) co-expression plasmid, the individual open reading frames were cloned into a modified version of the pACEBac1 vector (Geneva Biotech), in which the expression cassette is flanked by BsaI restriction enzyme sites. Panoramix was cloned with an N-terminal Twin-Strep-tag, Nxf2 with an N-terminal His6-tag and Nxt1 with an N-terminal FLAG-tag. All constructs contained the intact polyhedrin leader sequence which harbors a mutated ATG (ATT) upstream of the actual start codon. A low-level of non-canonical initiation from an upstream ATT site results in low level of slightly larger versions of the proteins. The three expression cassettes were then combined into a single destination vector via Golden Gate cloning (Engler et al., 2008). The resulting plasmid was transposed into the EmBacY bacmid backbone (Trowitzsch et al., 2010) and transfected into *Spodoptera frugiperda* Sf9 cells to generate a single baculovirus expressing all three genes. The resulting virus was used to infect *Trichoplusia ni* High5 cells at a density of 1×10^6 cells/ml and expression was performed at 21°C. The cells were harvested 4 days after growth arrest, approximately 96-120 hours after infection, collected by centrifugation and the cell pellets were flash-frozen in liquid nitrogen.

Affinity purification with Twin-Strep-tag

High5 cells were lysed in lysis buffer (LB) (50 mM Tris-HCl pH=8, 150 mM NaCl, 0,05 % TX100, 1mM DDT) freshly supplemented with Complete Protease Inhibitor Cocktail (Roche) and with Benzonase (~10U/ml) for 30 min at 4°C and the lysate was cleared by centrifugation. For purification, a StrepTactin Superflow HC resin (IBA GmbH) was used with the AKTA Purifier FPLC system and the column was equilibrated with 2 column volumes of LB before sample loading. The bound protein complex was eluted with LB supplemented with 5 mM desthiobiotin and analyzed by SDS-PAGE and InstantBlue (Expedeon) staining.

Size exclusion chromatography (SEC)

After affinity purification, the complex containing fractions were

pooled and further purified by SEC using a Superdex 200 10/300 (GE Healthcare) with the AKTA Purifier FPLC system in SEC buffer (SB) (50 mM Tris-HCl pH=8, 150 mM NaCl, 1mM DDT). The column was equilibrated with 2 column volumes of SB before sample loading. The purified complex was analyzed by SDS-PAGE and Coomassie staining.

Crosslinking-mass spectrometry (XL-MS)

Protein crosslinking and digestion: The purified complex was crosslinked with 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methyl-morpholinium chloride (DMTMM; final concentration 4mM). To identify the optimal crosslinker concentration, the complex was titrated with different concentrations and crosslinking yield was checked with SDS-PAGE. The complex was reacted with the crosslinker for 30 minutes at room temperature. To stop the reaction, Tris-HCl pH=8 was added to a final concentration of 100 mM. Sodium deoxycholate (SDC) was added to a final concentration of 1.5%. Samples were reduced with dithiothreitol (DTT, 10 mM, 30 min, 60°C), alkylated with iodoacetamide (IAA, 15 mM, 30 min at room temperature in the dark) and diluted to 1% SDC. Proteins were digested for 3h using trypsin (protein/enzyme 30:1, 37°C). With the addition of 1% trifluoro acetic acid (TFA), SDC was precipitated and the digest was stopped. The supernatant was decanted and stored until measurement.

Size-exclusion-chromatography (SEC) enrichment: The digested samples were enriched for crosslinks (XLs) prior to LC-MS/MS analysis using SEC. Therefore, approx. 15 µg of the digest was separated on a TSKgel SuperSW2000 column (300 mm × 4.5 mm × 4 µm, Tosoh Bioscience). The three high mass fractions were collected and measured on the mass spectrometer.

LC-MS/MS: Digested peptides were separated using a Dionex UltiMate 3000 HPLC RSLC nanosystem prior to MS analysis. The HPLC was interfaced with the mass spectrometer via a Nanospray Flex™ ion source. For sample concentrating, washing and desalting, the peptides were trapped on an Acclaim PepMap C-18 precolumn (0.3x5mm, Thermo Fisher Scientific), using a flowrate of 25 µl/min and 100% buffer A (99.9% H2O, 0.1% TFA). The separation was performed on an Acclaim PepMap C-18 column (50 cm x 75 µm, 2 µm particles, 100 Ä pore size, Thermo Fisher Scientific) applying a flowrate of 230 nl/min. For separation, a solvent gradient ranging from 2-35% buffer B (80% ACN, 19.92% H2O, 0.08% TFA) was applied. The applied gradient varied from 60-90 min, depending on the sample complexity.

Mass Spectrum Acquisition: MS1 spectra were recorded at a resolution of 120000 ranging from 350-1600 m/z (AGC 1e6, 60 ms max. injection time). The top 10 most intense ions from MS1 were selected for fragmentation. MS2 spectra were recorded at 30,000 resolution (AGC 5e4, max. injection time 150 ms, isolation width 1.0 m/z). DMTMM crosslinks were fragmented with higher energy C-trap dissociation (HCD) using a stepped collision energy of 30-33-35%. Once a precursor was selected for MS2, it was excluded from fragmentation for 30sec.

Data Analysis: Raw files were analyzed with pLink (Fan et al., 2015) (Version 2.3.3) using the settings as described above. Used crosslinker: DMTMM (-18.0116 Da, reactivity towards lysine, protein N-terminus, serine, threonine and tyrosine or aspartate, glutamate and the protein C-terminus, respectively); MS1 accuracy: 10 ppm; MS2 accuracy: 20 ppm; used enzyme: trypsin; max. missed cleavages: 4; minimum peptide length: 5; max. modifications: 4; static modifications: carbamidomethylation (cysteine, +57.021 Da); dynamic modifications: oxidation (methionine, +15.995 Da). For the database search a database containing the three crosslinked proteins was used and the false discovery rate (FDR) was set to 1%. To reduce the number of false positives, XLs were manually validated. For XL visualization, xiNET was used (Combe et al., 2015).

Recombinant protein expression in bacterial cells

The open reading frame encoding the Nxf2(1-284) fragment or its point mutant variant was cloned into pET21a with an N-terminal GB1 solubility enhancing tag. Protein expression was in BL21DE3 cells, which were grown at 37°C until OD600=0.6-0.8 and then induced with 0.1 mM IPTG for 18 hours at 18°C. The cell pellet was resuspended in lysis buffer (50mM $NaPO_4$ pH=8, 150 mM NaCl, 0.1 % Triton X-100, 10 mM imidazole, 10% glycerol, 5 mM 2-mercaptoethanol, freshly supplemented with 1 mM PMSF and Complete Protease Inhibitor Cocktail (Roche). Lysozyme (10 mg/ml) was added and the cell suspension was incubated for 30 min at 4°C and then sonicated for 15 min at 40% power output with 30% duty cycle (3 sec ON / 7 sec OFF). The sonicated suspension was centrifuged for 20 min at 19,000 g at 4°C. The supernatant was loaded onto pre-equilibrated TALON Metal Affinity Resin (Takara) and incubated for 2 hours. The resin was washed with 20 x bed volume wash buffer (50 mM Tris-HCl pH=7.25, 500 mM NaCl, 20 mM imidazole, 0.1 % Triton X-100, 10 % glycerol, 5 mM 2-mercaptoethanol). The proteins were eluted with 5 x bed volume elution buffer (50 mM Tris-HCl pH=8 300 mM NaCl, 300 mM imidazole, 0.1 % Triton X-100, 10 % Glycerol) and dialyzed against PBS supplemented with 10 % glycerol and 5 mM 2-mercaptoethanol. Post dialysis the protein was aliquoted, flash-frozen in liquid nitrogen and stored at – 80°C for further analysis.

Electrophoretic mobility shift assay (EMSA)

10 nM [32P] 5'-labelled single-stranded 35 nt RNA (CUCAUCUUGGUCGUACGCGGAAUAGUUUAAACUGU) was incubated with various concentrations of recombinant protein in 10 µl total volume with EMSA binding buffer (10 mM Tris-HCl pH=7.9, 2 mM $MgCl_2$, 0.1 mM EDTA, 4% glycerol, 50 mM KCl, 1 mM DTT, 10 µg/ml BSA) for 20 min at 4°C. 2 µl of EMSA loading buffer (50% glycerol, 0.075% bromophenol blue) was added to the samples which were analyzed by 4.8 % PAGE gel in 0.5 x TBE. The radioactive bands were visualized with a Phosphorimager.

Alpha helix characterization

To determine the physicochemical properties of the predicted alpha-helix within Panoramix, the HeliQuest web server (Gautier et al., 2008) was used. Using an 18-amino acid sliding window

(corresponding to a complete helical wheel), the tool predicts hydrophobic surfaces and shows the sequence in a helical wheel representation.

Phylogenetic tree, orthologue identification and multiple sequence alignment

For phylogenetic reconstruction, Nxf family members from a set of species was extracted, aligned using mafft (v7.407) and the obtained protein sequence alignment was converted to a codon alignment using pal2nal (v14). From this alignment, a maximum-likelihood tree was inferred with iqtree (v1.6.7) using best-fitting codon model selection by Modelfinder and 1000 ultrafast bootstrap replicates; visualization using iTol (v4.2.3).

For Nxf protein family collection, proteins showing significant sequence similarity to the NTF2-like domain of *Drosophila melanogaster* Sbr (Nxf1) were collected from the NCBI non-redundant protein database (NCBI nr) using blastp (query NP_524660.1:372-531; species filter: arthropoda and selected other organisms). Hits were retained if they also showed reciprocal best blast hits to one of the 4 known *Drosophila melanogaster* Nxf family members in blastp searches against the *Drosophila melanogaster* proteome (PTHR10662 members: Nxf3/ FBgn0263232, nxf2/ FBgn0036640, nxf4/ FBgn0051501, sbr/ FBgn0003321). The obtained Nxf protein family set was supplemented with *Drosophila melanogaster* nxf4 protein and its orthologs identified in reciprocal-best-blast-procedure.

For UBA protein domain alignments, a subset of representative Nxf2 protein sequences was selected using an 80% identity cutoff over the C-terminal region (corresponding to NP_524111.3: 726-841). The species selection was also used in alignments of the Nxf1/Tap ortholog groups.

Panoramix ortholog identification was performed using reciprocal psi-blast searches against NCBI nr, using the region of highest conservation among Drosophilid Panoramix proteins as a query (NP_611576.1:292-510). Multiple sequence alignments were visualized in Jalview (v2.10.4), and the secondary structure was predicted with JPRED. Sequence accessions from the alignments are listed in Table S7.

Data availability statement

All sequencing data used for this study (Table S13) have been deposited with NCBI GEO (accession code GSE120617). The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaino et al., 2016) (partner repository under data set identifier PXD011201). All fly lines used in this study are available from the VDRC (http://stockcenter.vdrc.at/control/main). Source data for all gel images are provided as Supplementary Information. Coordinate and structure factors of the UBA-linker-helix and the dmNxf2-Nxt1 complex have been deposited in the Protein Data Bank (PDB) under the accession number 6OKL and 6MRK, respectively.

Code availability statement

All custom code is based on the publicly available code used in (Andersen et al., 2017) with modifications indicated in the Methods section.

REFERENCES

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta crystallographica. Section D, Biological crystallography *58*, 1948-1954.

Aibara, S., Katahira, J., Valkov, E., and Stewart, M. (2015). The principal mRNA nuclear export factor NXF1:NXT1 forms a symmetric binding platform that facilitates export of retroviral CTE-RNA. Nucleic acids research *43*, 1883-1893.

Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54-59.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta crystallographica. Section D, Biological crystallography *67*, 271-281.

Braun, I.C., Herold, A., Rode, M., and Izaurralde, E. (2002). Nuclear export of mRNA by TAP/NXF1 requires two nucleoporin-binding sites but not p15. Molecular and cellular biology *22*, 5405-5418.

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., *et al.* (2014). Diversity and dynamics of the Drosophila transcriptome. Nature *512*, 393-399.

Caporilli, S., Yu, Y., Jiang, J., and White-Cooper, H. (2013). The RNA export factor, Nxt1, is required for tissue specific transcriptional regulation. PLoS genetics *9*, e1003526.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nature reviews. Genetics *14*, 100-112.

Combe, C.W., Fischer, L., and Rappsilber, J. (2015). xiNET: cross-link network maps with residue resolution. Mol Cell Proteomics *14*, 1137-1147.

Cullen, B.R. (2000). Nuclear RNA export pathways. Molecular and cellular biology *20*, 4181-4187.

Czech, B., Munafo, M., Ciabrelli, F., Eastwood, E.L., Fabry, M.H., Kneuss, E., and Hannon, G.J. (2018). piRNA-Guided Genome Defense: From Biogenesis to Silencing. Annual review of genetics *52*, 131-157.

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. Molecular cell *50*, 749-761.

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.

Doblmann, J., Dusberger, F., Imre, R., Hudecz, O., Stanek, F., Mechtler, K., and Durnberger, G. (2018). apQuant: Accurate Label-Free Quantification by Quality Filtering. J Proteome Res.

Donertas, D., Sienski, G., and Brennecke, J. (2013). Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. Genes & development *27*, 1693-1705.

Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S., and Mechtler, K. (2014). MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. J Proteome Res *13*, 3679-3684.

Dumesic, P.A., Natarajan, P., Chen, C., Drinnenberg, I.A., Schiller, B.J., Thompson, J., Moresco, J.J., Yates, J.R., 3rd, Bartel, D.P., and Madhani, H.D. (2013). Stalled spliceosomes are a signal for RNAi-mediated genome defense. Cell *152*, 957-968.

ElMaghraby, M.F., Andersen P.R., Pühringer F., Meixner K., Lendl T., Tirian L., Brennecke J. A heterochromatin-specific RNA export pathway facilitates piRNA production. BioRxiv [Preprint]. (2019) Available from: https://doi.org/10.1101/596171

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta crystallographica. Section D, Biological crystallography *66*, 486-501.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PloS one *3*, e3647.

Fan, S.B., Meng, J.M., Lu, S., Zhang, K., Yang, H., Chi, H., Sun, R.X., Dong, M.Q., and He, S.M. (2015). Using pLink to Analyze Cross-Linked Peptides. Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.] *49*, 8 21 21-19.

Fedoroff, N.V. (2012). Presidential address. Transposable elements, epigenetics, and genome evolution. Science *338*, 758-767.

Fribourg, S., Braun, I.C., Izaurralde, E., and Conti, E. (2001). Structural basis for the recognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor. Molecular cell *8*, 645-656.

Gautier, R., Douguet, D., Antonny, B., and Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific alpha-helical properties. Bioinformatics *24*, 2101-2102.

Gokcezade, J., Sienski, G., and Duchek, P. (2014). Efficient CRISPR/Cas9 Plasmids for Rapid and Versatile Genome Editing in Drosophila. G3.

Gramates, L.S., Marygold, S.J., Santos, G.D., Urbano, J.M., Antonazzo, G., Matthews, B.B., Rey, A.J., Tabone, C.J., Crosby, M.A., Emmert, D.B., *et al.* (2017). FlyBase at 25: looking to the future. Nucleic acids research *45*, D663-D671.

Grant, R.P., Neuhaus, D., and Stewart, M. (2003). Structural basis for the interaction between the Tap/NXF1 UBA domain and FG nucleoporins at 1A resolution. J Mol Biol *326*, 849-858.

Grewal, S.I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. Current opinion in genetics & development *20*, 134-141.

Gu, B., Posfai, E., and Rossant, J. (2018). Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. Nature biotechnology *36*, 632-637.

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. Molecular cell *50*, 762-777.

Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S.L., and Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3'-end formation. Nature *539*, 588-592.

Heath, C.G., Viphakone, N., and Wilson, S.A. (2016). The role of TREX in gene expression and disease. The Biochemical journal *473*, 2911-2935.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.

Herold, A., Klymenko, T., and Izaurralde, E. (2001). NXF1/p15 heterodimers are essential for mRNA nuclear export in Drosophila. Rna *7*, 1768-1780.

Herold, A., Suyama, M., Rodrigues, J.P., Braun, I.C., Kutay, U., Carmo-Fonseca, M., Bork, P., and Izaurralde, E. (2000). TAP (NXF1) belongs to a multigene family of putative RNA export factors with a conserved modular architecture. Molecular and cellular biology *20*, 8996-9008.

Herold, A., Teixeira, L., and Izaurralde, E. (2003). Genome-wide analysis of nuclear mRNA export pathways in Drosophila. The EMBO journal *22*, 2472-2483.

Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. Nature reviews. Genetics *16*, 71-84.

Jayaprakash, A.D., Jabado, O., Brown, B.D., and Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic acids research *39*, e141.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204-2207.

Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nature reviews. Molecular cell biology *8*, 761-773.

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

Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., and Toth, K.F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes & development *27*, 390-399.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nature protocols *1*, 729-748.

Liker, E., Fernandez, E., Izaurralde, E., and Conti, E. (2000). The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. The EMBO journal *19*, 5587-5598.

Mohn, F., Sienski, G., Handler, D., and Brennecke, J. (2014). The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell *157*, 1364-1379.

Morlan, J.D., Qu, K., and Sinicropi, D.V. (2012). Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. PloS one *7*, e42882.

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. Molecular cell *50*, 736-748.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta crystallographica. Section D, Biological crystallography *53*, 240-255.

Niki, Y., Yamaguchi, T., and Mahowald, A.P. (2006). Establishment of stable cell lines of Drosophila germ-line stem cells. Proceedings of the National Academy of Sciences of the United States of America *103*, 16325-16330.

Ninova, M., Chen, Y.A., Godneeva, B. Rogers, A.K., Luo, Y., Aravin, A.A., Fejes Tóth, K. The SUMO ligase Su(var)2-10 links piRNA-guided target recognition to chromatin silencing. BioRxiv [Preprint]. (2019) Available from: https://doi.org/10.1101/533091

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 & development *27*, 1656-1661.

Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P.D. (2018). PIWI-interacting RNAs: small RNAs with big functions. Nature reviews. Genetics.

Pan, J., Eckardt, S., Leu, N.A., Buffone, M.G., Zhou, J., Gerton, G.L., McLaughlin, K.J., and Wang, P.J. (2009). Inactivation of Nxf2 causes defects in male meiosis and age-dependent depletion of spermatogonia. Developmental biology *330*, 167-174.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature methods *14*, 417-419.

Pfeiffer, B.D., Jenett, A., Hammonds, A.S., Ngo, T.T., Misra, S., Murphy, C., Scully, A., Carlson, J.W., Wan, K.H., Laverty, T.R.*, et al.* (2008). Tools for neuroanatomy and neurogenetics in Drosophila. Proceedings of the National Academy of Sciences of the United States of America *105*, 9715-9720.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nature methods *14*, 687-690.

Post, C., Clark, J.P., Sytnikova, Y.A., Chirn, G.W., and Lau, N.C. (2014). The capacity of target silencing by Drosophila PIWI and piRNAs. Rna *20*, 1977-1986.

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

Reyes-Turcu, F.E., Zhang, K., Zofall, M., Chen, E., and Grewal, S.I. (2011). Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. Nature structural & molecular biology *18*, 1132-1138.

Rozhkov, N.V., Hammell, M., and Hannon, G.J. (2013). Multiple roles for Piwi in silencing Drosophila transposons. Genes & development *27*, 400-412.

Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., and Siomi, M.C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature *461*, 1296-1299.

Shimada, Y., Mohn, F., and Buhler, M. (2016). The RNA-induced transcriptional silencing complex targets chromatin exclusively via interacting with nascent transcripts. Genes & development *30*, 2571-2580.

Sienski, G., Batki, J., Senti, K.A., Donertas, D., Tirian, L., Meixner, K., and Brennecke, J. (2015). Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. Genes & development *29*, 2258-2271.

Sienski, G., Donertas, 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.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nature reviews. Genetics *8*, 272-285.

Teixeira, F.K., Okuniewska, M., Malone, C.D., Coux, R.X., Rio, D.C., and Lehmann, R. (2017). piRNA-mediated regulation of

transposon alternative splicing in the soma and germ line. Nature *552*, 268-272.

Teplova, M., Wohlbold, L., Khin, N.W., Izaurralde, E., and Patel, D.J. (2011). Structure-function studies of nucleocytoplasmic transport of retroviral genomic RNA by mRNA export factor TAP. Nature structural & molecular biology *18*, 990-998.

Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., and Berger, I. (2010). New baculovirus expression tools for recombinant protein complex production. J Struct Biol *172*, 45-54.

Viphakone, N., Hautbergue, G.M., Walsh, M., Chang, C.T., Holland, A., Folco, E.G., Reed, R., and Wilson, S.A. (2012). TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nature communications *3*, 1006.

Vizcaino, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., *et al.* (2016). 2016 update of the PRIDE database and its related tools. Nucleic acids research *44*, D447-456.

Wang, S.H., and Elgin, S.C. (2011). Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proceedings of the National Academy of Sciences of the United States of America *108*, 21164-21169.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).

Wilkie, G.S., Zimyanin, V., Kirby, R., Korey, C., Francis-Lang, H., Van Vactor, D., and Davis, I. (2001). Small bristles, the Drosophila ortholog of NXF-1, is essential for mRNA export throughout development. Rna *7*, 1781-1792.

Yang, P., Wang, Y., and Macfarlan, T.S. (2017). The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. Trends Genet *33*, 871-881.

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.