HAL
open science

Non-coding RNA Transcription in Tetrahymena Meiotic Nuclei Requires Dedicated Mediator Complex-Associated Proteins

Miao Tian, Kazufumi Mochizuki, Josef Loidl

▶ **To cite this version:**

Miao Tian, Kazufumi Mochizuki, Josef Loidl. Non-coding RNA Transcription in Tetrahymena Meiotic Nuclei Requires Dedicated Mediator Complex-Associated Proteins. Current Biology - CB, 2019, 29 (14), pp.2359-2370.e5. 10.1016/j.cub.2019.05.038 . hal-02387365

HAL Id: hal-02387365

https://hal.science/hal-02387365v1

Submitted on 7 Dec 2020

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

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

Non-coding RNA transcription in Tetrahymena meiotic nuclei requires dedicated Mediator complex -associated proteins

Miao Tian,[1,3,*] Kazufumi Mochizuki,[2] and Josef Loidl[1]

[1]Department of Chromosome Biology, Max F. Perutz Laboratories, University of Vienna, Vienna, A-1030, Austria
[2]Institute of Human Genetics (IGH), CNRS, University of Montpellier, Montpellier, 34090, France
[3]Lead Contact
*Correspondence: tian.miao@univie.ac.at

Summary
To preserve genome integrity, eukaryotic cells use small RNA-directed mechanisms to repress transposable elements (TEs). Paradoxically, in order to silence TEs, precursors of the small RNAs must be transcribed from TEs. However, it is still poorly understood how these precursors are transcribed from TEs under silenced conditions. In the otherwise transcriptionally silent germline micronucleus (MIC) of *Tetrahymena*, a burst of non-coding RNA (ncRNA) transcription occurs during meiosis. The transcripts are processed into small RNAs that serve to identify TE-related sequences for elimination. The Mediator complex (Med) has an evolutionarily conserved role for transcription by bridging gene-specific transcription factors and RNA polymerase II. Here, we report that three Med-associated factors, Emit1, Emit2, and Rib1, are required for the biogenesis of small ncRNAs. Med localizes to the MIC only during meiosis and both Med localization and MIC ncRNA transcription require Emit1 and Emit2. In the MIC, Med occupies TE-rich pericentromeric and telomeric regions in a Rib1-dependent manner. Rib1 is dispensable for ncRNA transcription but is required for the accumulation of double-stranded ncRNAs. Nuclear and sub-nuclear Localization of the three Med-associated proteins is interdependent. Hence, Emit1 and Emit2 act coordinately to import Med into the MIC, and Rib1 recruits Med to specific chromosomal locations to quantitatively or qualitatively promote the biogenesis of functional ncRNA. Our results underscore that the transcription machinery can be regulated by a set of specialized Med-associated proteins to temporally transcribe TE-related sequences from a silent genome for small RNA biogenesis and genome defense.

Introduction
Transposable elements (TEs) are mobile DNA species that make up substantial fractions of almost all eukaryotic genomes. The transpositions of TEs can induce insertions, deletions, inversions, and translocations, thereby disrupting gene and regulatory sequences [1-3]. Thus, the transposition of TEs in the germline can lead to hereditary diseases, germ cell death, and sterility [4, 5].

In order to preserve genome integrity, eukaryotic cells use small ncRNA-directed mechanisms to repress TE mobility [3]: transcripts produced from active or degenerate TEs are processed into small ncRNAs, which associate with Argonaute/Piwi family proteins and repress TE activity through transcriptional and/or post-transcriptional mechanism(s) [6]. Therefore, transcription at TE-containing loci is the first step in all small ncRNA-directed TE repression pathways. Studies from protists to humans indicate that RNA polymerase II (Pol II), or its specialized form in plants (Pol IV), is required for the transcription of TE-derived small ncRNA precursors [7, 8]. The precursor RNA can be transcribed from a single DNA strand under the control of specific transcription factors via the canonical mechanism [9, 10] or bidirectionally via an as-yet not fully understood mechanism [11, 12].

Tetrahymena thermophila and other ciliated protozoa have evolved an

ultimate method of combating TEs, that is, by eliminating them from the somatic genome [13]. This unique feature makes *Tetrahymena* a good model organism for investigating the production of TE-targeting small ncRNAs. *Tetrahymena* carries two morphologically and functionally distinct nuclei within a single cell. The large polyploid somatic macronucleus (MAC) is transcriptionally active and determines the phenotype, whereas the diploid germline micronucleus (MIC) is transcriptionally inactive and serves as a vault for the genome that is passed to progeny MACs and MICs during conjugation, the sexual reproduction process (Figure 1). The MIC genome contains approximately 12,000 internal eliminated sequences (IESs) that are removed from the MAC during conjugation via programmed DNA elimination. IESs occupy about one third of the germline genome and the vast majority are TE-related sequences. Some IESs are suggested to be structural and/or regulatory elements, such as centromeres, that are needed for the canonical condensation and division cycle to maintain MIC genome integrity [13-15]. Thus, DNA elimination removes sequences that are potentially harmful and/or dispensable for the function of MAC.

IES elimination in *Tetrahymena* is guided by small ncRNAs [16, 17]. The transcription of small RNA precursors occurs concomitantly with meiotic recombination and is the only detectable transcriptional activity of the MIC [18]. Transcription takes place preferentially from so-called Type-A IESs, which are enriched in pericentromeric and telomeric regions. The clustered arrangement of these IESs and the use of the resultant small ncRNAs in targeting TEs resemble the properties of piRNA clusters found in animals [15, 19]. Transcription of both DNA strands by Pol II produces complementary transcripts, which form double-stranded RNA (dsRNA) molecules [20-23]. The dsRNAs are cleaved into 28–30 nt scan RNAs (scnRNAs) by a dicer-like protein, Dcl1 [24-26]. scnRNAs are exported from the MIC and form a complex with Twi1, a *Tetrahymena* Piwi protein, which eventually transports the scnRNAs to the new MAC, where they guide heterochromatinization and IES elimination [27, 28].

Although the scnRNA-guided IES elimination pathway has been elucidated, the molecular mechanism regulating scnRNA precursor transcription remains unknown. Major outstanding questions related to MIC transcription are (1) how does it take place in the normally repressive heterochromatic environment of the MIC? (2) how is it limited to the pericentromeric and telomeric regions? Here, we report three Med-associated proteins that localize to the transcribed MIC during conjugation and are required for scnRNA biogenesis.

Results

Emit1, Emit2, and Rib1 are required for DNA elimination

Previous studies have shown that most of the genes involved in scnRNA biogenesis are exclusively expressed during conjugation [16, 24, 29, 30]. Hence, we systematically knocked out conjugation-specific genes and found that deletion of three, *TTHERM_00039000*, *TTHERM_01055420*, and *TTHERM_00474920*, caused the arrest of exconjugant cells at stages where

cells have two MICs and two MACs (Figure 2A), the phenotype characteristic of mutants defective in scnRNA biogenesis. We named these genes *EMIT1, EMIT2* (Enables Micronucleus Transcription), and *RIB1* (Ripe Banana, owing to the inhomogeneous spotty localization of Rib1 protein in the elongated MIC).

No conserved domain was detected in Emit1, the protein encoded by *EMIT1*. Emit2 has low sequence similarity to the alpha catalytic subunit of DNA polymerase III, but the significance of this is unclear. Rib1 contains nine tandem repeats of the "NQ[M/I]NQN[P/Q]" motif (Figure 2B). Additionally, this N/Q-rich low complexity region was identified as an intrinsically disordered region and contains a prion-like domain (Figure S2). Similar N/Q-rich tandem repeats are also present in the Med15-related proteins of several lepidopterans. Since Rib1 interacts with the Med complex (see below), this motif may be involved in regulating transcription.

In *EMIT1*, *EMIT2*, and *RIB1* knockout strains (*emit1*Δ, *emit2*Δ, and *rib1*Δ), vegetative propagation and conjugation were normal, except that the pachytene-like stage IV was extended in *emit2*Δ (Figure S1). We next analyzed DNA elimination by fluorescence in situ hybridization (FISH) using a probe complementary to the repetitive *REP2* IESs [31]. In the progeny of wild-type (WT) cells, *REP2* IESs were retained only in the new MIC at 32 h after the induction of conjugation. In contrast, in the mutants, *REP2* IESs were present in both of the new MIC and new MAC at the same time point (Figure 2C). This result indicates that *EMIT1, EMIT2*, and *RIB1* are required to complete DNA elimination. Consequently, none of the knockout mating pairs produced viable sexual progeny (Table S1).

Emit1, Emit2, and Rib1 are required for scnRNA biogenesis

IES elimination requires scnRNAs that are processed from long dsRNAs by Dcl1 in the MIC [24, 25]. In contrast to the WT, we could not detect scnRNAs in the *emit1*Δ, *emit2*Δ, and *rib1*Δ strains (Figure 2D). Therefore, we tested whether dsRNA was formed in the first place. dsRNA was probed with an antibody that recognizes dsRNAs longer than 40 bp. Although we detected dsRNAs in the meiotic prophase MIC of WT cells, they were undetectable in all three mutants (Figure 2E). In contrast, dsRNA hyperaccumulation occurred in *dcl1*Δ control cells, as previously reported [24, 25], probably because dsRNAs are not processed to scnRNAs in the absence of Dcl1. These results indicate that the production of dsRNA precursors of scnRNAs requires *EMIT1, EMIT2*, and *RIB1*.

dsRNAs are believed to be formed by the convergent transcription of both DNA strainds of IESs [20, 22]. Thus, transcription of the MIC genome was examined by northern blotting using RNA probes complementary to the sense and antisense strand of a well-characterized IES, the M-element (Figure 2F, i). As previously reported [25], both sense and antisense RNA probes detected RNAs of heterogeneous lengths in total RNA from WT cells (Figure 2F, ii and iii). In contrast, such RNAs were not detected in total RNA from *emit1*Δ and *emit2*Δ cells, suggesting that Emit1 and Emit2 are required for ncRNA

transcription (at least for the M-element) in the MIC. In the WT cells, dsRNAs are quickly processed by Dcl1 to scnRNAs [23], therefore ncRNAs detected by northern blot (which detects both ssRNA and dsRNA) are dramatically increased in *dcl1*Δ [32]. In contrast, although ncRNAs in *rib1*Δ were not processed to scnRNAs, the level of ncRNAs was comparable to the WT, but much lower than in *dcl1*Δ cells (Figure 2F and Figure S3B). Because scnRNA was not detected in *rib1*Δ cells, these results suggest that *RIB1* deletion does not completely prevent ncRNA transcription in the MIC but instead inhibits the formation of functional scnRNA precursors by either attenuating ncRNA transcription, destabilizing the transcripts, and/or disrupting the annealing of complementary transcripts.

Emit1, Emit2, and Rib1 interact with Med subunits

In order to understand how Emit1, Emit2, and Rib1 regulate MIC transcription, we first investigated their interactors by immunoprecipitation and mass spectrometry analysis. For this, strains expressing C-terminally hemagglutinin (HA)-tagged Emit1, Emit2, or Rib1 from the endogenous promoter were generated. Consistent with *EMIT1, EMIT2*, and *RIB1* mRNA expression patterns, HA-tagged Emit1, Emit2, and Rib1 were detected only during conjugation (Figure 3A). Hence, the HA-tagged proteins were immunoprecipitated during meiotic prophase (Figure S3C) and co-purified proteins were identified by mass spectrometry. A total of 108, 29, and 101 proteins were identified as interacting partners of Emit1, Emit2, and Rib1, respectively (Table S2). Furthermore, Emit1, Emit2, and Rib1 co-immunoprecipitated and shared 22 interactors, suggesting that they probably act in a complex *in vivo* (Figure 3B).

The 22 interactors in common included eight of the 14 putative *Tetrahymena* Med components (Figure 3B) [33]. In addition, another common interactor (encoded by *TTHERM_00918460*) had amino acid sequence similarity to *Schizosaccharomyces pombe* Med10 (26% identity), and was therefore named Med10L (Med10-like). Since there is only one copy of each Med subunit coding gene in the *Tetrahymena* genome [33], the nine Med subunits that interacting with Emit1, Emit2, and Rib1 are most likely genuine Med components. These results suggest that Emit1, Emit2, and Rib1 regulate transcription through Med. Consistent with this hypothesis, three Pol II subunits also co-immunoprecipitated with Emit1 (Figure 3B).

Given the essential role of Med in regulating Pol II transcription in other eukaryotes, we next asked whether mRNA transcription in the MAC is affected by deletion of *EMIT1, EMIT2*, or *RIB1*. Reverse transcription (RT)-PCR analysis showed that knocking out of any of these three genes does not inhibit expression of the two others or of any other conjugation-specific gene tested (Figure S3D). Therefore, Emit1, Emit2, and Rib1 are unlikely required for basal mRNA transcription of the somatic genome. Based on above results, we conclude that Emit1, Emit2, and Rib1 are conjugation-specific Med-associated

proteins that mainly, if not exclusively, regulate MIC transcription.

Emit1, Emit2, and Rib1 colocalize with Pol II and Med in the meiotic MIC

To investigate the spatial relationships between Med-associated proteins and the transcriptional machinery, we first determined Pol II and Med localization in WT cells. Pol II localization was determined by immunostaining with an antibody against Rpb3, a *Tetrahymena* Pol II subunit [34], and Med localization was determined with immunostaining for HA-tagged Med31, a conserved Med subunit [33]. Rpb3 and Med31 were constitutively localized in the MAC (Figure 4A). They were also localized in the elongated meiotic prophase MIC where centromeres and telomeres are clustered at opposite ends [35]. Rpb3 and Med31 were particularly abundant in the centromere-proximal half and the telomeric tip (Figures 4B and 4C), where Type-A IESs are also concentrated [15] (Figure 4B, yellow). By late prophase (stage IV), Rpb3 and Med31 had disappeared from the pericentromeric region but remained at the telomeric tip. At all stages examined, the Rpb3 and Med31 localization patterns were indistinguishable (Figure 4C and Figure S4). A similar localization pattern was observed for dsRNAs, suggesting that they are formed concomitantly with transcription (Figures 4B and 4C).

We next compared the localization of HA-tagged Emit1, Emit2, and Rib1 with that of Rpb3. All three colocalized with Rpb3 in the meiotic MIC (Figure 4C and Figure S5). Emit1 was also detected in the MAC, although the biological significance of this finding is unclear. Importantly, the localization of Rpb3, Med31, Emit1, Emit2, and Rib1 was maintained after pre-fixation detergent treatment (Figure 5 and Figure S5), which removes nucleoplasmic proteins from the nucleus [36]. This result indicates that these factors are tightly bound to chromatin.

Emit1, Emit2, and Rib1 are required for the proper Pol II and Med localization in the MIC

The failure of *emit1*Δ, *emit2*Δ, and *rib1*Δ cells to produce functional ncRNA (Figure 2E), and the colocalization of Emit1, Emit2, and Rib1 with Pol II and Med in the meiotic MIC prompted us to ask whether Pol II and Med localization in the MIC is dependent on Emit1, Emit2, and Rib1. In the absence of Emit1, Emit2 or Rib1 alone, Rpb3 was present in the MIC but became dispersed along the elongated nucleus throughout meiotic prophase (Figure 5A, upper panel). The MIC localization of Rpb3 was resistant to the pre-fixation detergent treatment in both WT and the mutants (Figure 5A, lower panel). Hence, Rpb3 binds to chromatin independently of Emit1, Emit2, and Rib1. Taken together, these results indicate that all three proteins are required for the subnuclear localization of Pol II in the MIC, but not for its chromatin association.

Next, to investigate whether Med localization is dependent on the three proteins, the Med31-HA construct was introduced into *emit1*Δ, *emit2*Δ, and *rib1*Δ cells. In *emit1*Δ and *emit2*Δ cells, Med31 was undetectable in the MIC,

although normal levels of Med31 were detected in the MAC (Figure 5B, upper panel). In contrast to its preferential centromere-proximal/telomeric location in the WT MIC, Med31 was present throughout the *rib1*Δ MIC. Pre-fixation detergent treatment had no effect on Med31's association with chromatin in *rib1*Δ cells (Figure 5B, lower panel). Thus, Emit1 and Emit2 are needed for Med localization to the MIC, whereas Rib1 mediates Med localization to specific regions of MIC chromosomes. The presence of Med in *rib1*Δ and its absence in *emit1*Δ and *emit2*Δ cells, may explain the presence or absence, respectively, of MIC transcription in these mutants (Figure 2F).

Since Med and Pol II are homogeneously distributed along the MIC in the absence of *RIB1*, we investigated whether the redistribution of these proteins causes ectopic transcription. We investigated the expression of Type-B IESs (*IES737* and *IES4092*) and the *TPB6* gene, which are not usually transcribed in the meiotic MIC in WT cells [19, 37]. RT-PCR analysis did not detect transcripts from any of the tested loci during meiotic prophase in *rib1*Δ cells (Figure S3E). Therefore, the ubiquitous distribution of Med and Pol II throughout the MIC in the absence of *RIB1* does not result in a global ectopic transcription in the MIC. Thus, other factor(s) must function to restrict scnRNA production to specific chromosomal regions even when the transcriptional machinery is mislocalized in the absence of Rib1.

Localization of Emit1, Emit2, and Rib1 is interdependent

As Emit1, Emit2, and Rib1 interact with one another (Figure 3B), we next analyzed whether their localization is interdependent. For this, HA-tagged Emit1, Emit2, and Rib1 were expressed individually in *emit1*Δ, *emit2*Δ, and *rib1*Δ strains and then localized by immunostaining with an anti-HA antibody (Figure 6). In the *emit1*Δ mutant, Emit2 was distributed uniformly throughout the MIC, but Rib1 localization was only moderately affected (Figure 6A). In the *emit2*Δ mutant, Emit1 was not detected in the MIC (Figures 6B). Curiously, in the absence of Emit2, Rib1 accumulated at the MIC periphery, especially near to the centromeric and telomeric termini (Figures 6B, 6C and Figure S6), suggesting that the centromeric/telomeric localization of Rib1 is independent of interaction with chromosomes. Interestingly, the localization pattern of Rib1 in *emit2*Δ resembles that of α-tubulin in the MIC [38]. Therefore, Rib1 may be transported to MIC termini through microtubules or other components of the cytoskeleton and then loaded onto chromatin in an Emit2-dependent manner. In the *rib1*Δ mutant, both Emit1 and Emit2 were distributed uniformly throughout the MIC (Figure 6D).

The interdependency of Emit1, Emit2, and Rib1 localization and the requirement for these proteins in the correct localization of Med and Pol II are summarized in Table 1. Because Emit2 is essential for transferring Emit1 and Med to the MIC and Rib1 to the MIC chromatin, the *emit2*Δ phenotype can be explained by the absence of Emit1, Rib1 and/or Med from MIC chromosomes. Once Emit2 and Rib1 are in the MIC, Rib1 can localize to pericentromeric and

subtelomeric regions without Emit1, whereas Emit1, Emit2, Med, and Pol II require Rib1 for proper subnuclear localization. Therefore, Rib1 has the ability to localize to the "scnRNA producing regions," where it recruits the other factors.

Discussion

Specialized factors promote unconventional ncRNA transcription by the Med complex in the MIC

This study identified three conjugation-specific Med-associated proteins that are essential for the biogenesis of functional ncRNAs in the meiotic prophase MIC. Two of these, Emit1 and Emit2, are essential for transporting Med into the MIC: Emit1 requires Emit2 to localize to the MIC, but not vice versa (Figures 5B, 6A, and 6B). Interestingly, a MIC-specific importin, Ima5 [39], was identified as an Emit2 interaction partner (Figure 3B and Table S2). Thus, Emit2 may serve as an adaptor protein between Ima5 and Emit1 to import Med components into the MIC. Moreover, Dcl1 was identified as the most prevalent protein that co-immunoprecipitated with Emit2, suggesting an additional function for Emit2 in either importing Dcl1 into the MIC or linking the transcription and processing of transcripts into scnRNAs.

In contrast to the MIC-limited Emit2 localization, Emit1 is localized in both the MIC and the MAC during meiotic prophase. However, Emit1 is unlikely needed for mRNA production from the MAC (Figure S3D), suggesting that, if Emit1 has any role in the MAC, then it would likely be in ncRNA production. This possibility is supported by the fact that an ectopically introduced IES can be transcribed in the MAC [20]. Based on Emit1′s role in mediating the subnuclear localization of Pol II and Emit2 in the MIC (Figures 5A and 6A), its role in maintaining the MIC localization of Med (Figure 5B), and its interaction with Pol II subunits (Figure 3B), we speculate that Emit1 couples Med to the other components for initiating and/or regulating ncRNA transcription.

Rib1 directs the ncRNA transcription machinery to the pericentromeric and telomeric regions of MIC chromosomes

The homogeneous distribution of Med and Pol II in the *rib1*Δ MIC suggests that Rib1 is likely to be a MIC transcriptional regulator that functions by recruiting the transcriptional machinery to the TE-rich pericentromeric and telomeric regions. The Rib1-dependent subnuclear localization of Med and Pol II may also be important to avoid the steric hindrance of meiotic recombination by the transcriptional machinery. Indeed, the distribution of γ-H2AX (a marker for DNA double-strand breaks) suggests that meiotic recombination occurs in the centromere-distal part of the nucleus [40].

All three Med-associated proteins, together with Med and Pol II, are quickly lost from the pericentromeric regions of the meiotic MIC as meiosis progresses. The observation of several Rib1 fragments in the immunoprecipitation product (Figure S3C) suggests that this process may be regulated by the degradation of Rib1. Notably, some peptidase homologs were highly enriched in Rib1

immunoprecipitation product (Figure 3B and Table S2). Thus, these proteins may regulate Rib1 turnover, and hence spatiotemporal transcriptional activity within the MIC.

Growing evidence has shown that protein phase separation drives the formation of the membraneless organelle (e.g., nucleolus and processing bodies), which concentrates a specific group of macromolecules for specific biological functions [41]. Interestingly, the phase-separation driven in part by intrinsically disordered regions (IDRs) of Med components at some clustered enhancers (called superenhancer) can concentrate transcription apparatus at those sites to regulate gene expressions [42]. Because prion-like IDRs also function as drivers of phase separation [43], the compartment of the transcriptional machinery at the TE-rich regions and the present of N and/or Q rich prion-like IDRs in Rib1, as well as in Med22 and Med31 (Figure S2), implicating that the biased accumulation of the transcriptional machinery in the MIC may be mediated by phase separation.

Unlike Emit1 and Emit2, Rib1 is not required for activating MIC transcription. Because the ncRNAs in the WT cells form dsRNAs and quickly processed to scnRNAs by dcl1 [23]. The comparable steady-state levels of ncRNAs in WT and *rib1*Δ cells (Figure 2F) suggests that ncRNA transcription is reduced in *rib1*Δ and/or, ncRNA is destabilized by some degradation pathways in the absence of *RIB1*. Nonetheless, ncRNAs accumulated in *rib1*Δ are not processed to scnRNAs. It is possible that the level of ncRNA in *rib1*Δ might be too low to form enough amount of dsRNAs for Dcl1 processing. Alternatively, Rib1 may be essential for the formation of dsRNA. The formation of dsRNA may rely on an RNA-dependent RNA polymerase (RdRP). There is one identifiable RdRP encoding gene (*RDR1*) in the *Tetrahymena* genome [44]. However, localization analysis of Rdr1 shows that it is present only in the cytoplasm of both vegetatively growing cells and conjugating cells (http://www.suprdb.org, ID: SUPR000375). Therefore, RdRP is unlikely to be involved in the biogenesis of the double-stranded scnRNA precursors, although we cannot exclude the possibility that some unknown RdRP, which we could not identify in the fully sequenced genome, is involved in this process. Instead, Rib1 may be required for efficiently annealing complementary transcripts. Alternatively, it may be essential for bi-directional transcription of the MIC genome or necessary for efficiently annealing complementary transcripts. Although our analysis of ncRNA transcripts suggested that they are derived from both strands of an IES (Figure 2F), investigation of unique IES sequences at the single-cell level is necessary to explicitly determine the strand specificity of ncRNA transcripts in *rib1*Δ cells and thus the role of Rib1.

The Med complex has a conserved role in genome defense

The usage of developmental stage-specific Med-associated proteins to regulate transcription in this early branching eukaryote is reminiscent of the requirement for specialized Med and Med-associated proteins in cell differentiation in flies

and mice [45-47]. Thus, transcriptional regulation via altering Med composition or its association partners might be an ancient mechanism.

An intriguing observation was that ncRNA transcription occurs in the normally transcriptionally inactive MIC, which is wrapped by heterochromatic histone marks (e.g. H3K23 me3 and H3K27 me3) and lacks euchromatic marks (e.g. H3K4 me3) [40, 48]. Studies in *Drosophila* suggested that the Rhino heterochromatin-binding protein and Deadlock, its partner protein, recruit Moonshiner, a basal transcription factor IIA paralog, for transcription initiation at heterochromatic piRNA clusters [11, 49, 50]. Moreover, Med is reported to be not only involved in producing mRNAs but also required for transcribing ncRNAs that are needed for heterochromatin formation and TE repression in both *Arabidopsis* and fission yeast [51, 52]. In this context, it would be interesting to test whether Emit1, Emit2, or Rib1 associates with heterochromatin marks and recruits Med and other factors to promote transcription in these regions. We believe that further investigation into Emit1, Emit2, Rib1, and their interacting proteins will provide insight into the mechanism of Med regulation in non-canonical ncRNA transcription and genome defense.

Acknowledgements

We thank Douglas Chalker (Washington University in St. Louis) for sharing *dcl1*Δ strains. We also thank Rachel Howard-Till, Takahiko Akematsu (University of Vienna) and Tai-Ting Woo (Institute of Molecular Biology, YAO lab) for their technical supports in FISH experiment, in generating the MicNup98A mCherry tagging construct and in dsRNA staining. We are grateful to the members of the Loidl, Mochizuki, Campbell (University of Vienna) laboratories and Gang Dong (Medical University of Vienna) for fruitful discussions and for sharing reagents. We also thank Markus Hartl and his team for their supports in proteomics analysis. Mass Spectrometry analyses were performed by the MFPL Mass Spectrometry Facility using the VBCF instrument pool. This research was funded by grants P27313-B20 and P31606-B20 from the Austrian Science Fund (FWF) to J.L. and ANR-10-LABX-12–01 and ANR-16-ACHN-0017 from the French National Research Agency to K.M.

Author contributions

J.L. conducted the knockout screening and did initial characterization of *EMIT2*, K.M. identified, named and did initial characterization of *RIB1* and M.T. identified *EMIT1* and conducted most of the experiments. M.T., K.M., and J.L. designed the experiments and wrote the manuscript.

Declaration of Interests

The authors declare no competing interests.

Figure legends

Figure 1. *Tetrahymena* reproductive cycles.

Tetrahymena cells possess two functionally distinct nuclei: the transcriptionally silent micronucleus (MIC), which functions as the germline; and the transcriptionally active macronucleus (MAC) representing the soma. The MAC is polyploid but each single genome is only about two thirds of the size of the MIC genome owing to the elimination of adverse sequences (mostly transposons) from the soma.

(A) During vegetative propagation, the two nuclei divide asynchronously.

(B) Starving cells may enter a sexual reproductive cycle, in which two cells of opposite mating types conjugate.

(C) The MICs of conjugating cells undergo simultaneous meioses. Meiotic prophase is characterized by the extreme elongation of MICs. Prophase substages I–IV are classified according to the degree of nuclear elongation. During prophase, the otherwise genetically silent MIC transcribes ncRNA (green underline), which carries information about which parts of the genome will be eliminated from the progeny MAC precursor.

(D) After meiosis, one of the four haploid nuclei survives and then doubles by mitosis. One of the resulting nuclei, the "sperm" nucleus enters the partner cell and fertilizes the stationary "egg" nucleus. Such reciprocal fertilization provides both cells with a diploid zygotic nucleus, which divides into the precursors of progeny MICs and MACs.

(E) IESs within progeny MAC precursors (outlined in red) are eliminated, while the old MAC is degraded.

(F) Upon feeding, four progeny cells are formed, each containing one MAC and one MIC.

Figure 2. Emit1, Emit2, and Rib1 are required for scnRNA production and IES elimination.

(A) The percentage of sexual progeny cells in the different development stages at 32 h after induction of conjugation. Stage 1, 2 MACs/2 MICs, conjugating partners are still connected, the old MAC remains; stage 2, 2 MACs/2 MICs, conjugating partners have separated, the old MAC remains; Stage 3, 2 MACs/2 MICs, the old MAC is degraded; stage 4: 2 MACs/1 MIC. Only the wild type (WT) proceeds to stage 4 prior to feeding. Developmental profiles of WT and knockout mutants during conjugation can be found in Figure S1. To generate the histogram, at least 100 cells were counted for each genotype. Error bars represent the standard deviation from three independent experiments.

(B) Rib1 has an N/Q-rich region (84–142 aa) with nine imperfect tandem repeats of the “NQ[M/I]NQN[P/Q]” motif. The same region was also identified as an intrinsically disordered region and contains a prion-like domain. (see also Figure S2). The *y*-axis is information content in bits. Hydrophilic, neutral, or hydrophobic residues were colored in blue, green,

or black, respectively.

(C) DNA FISH analysis of *REP2* IESs in *emit1*Δ (Panels i–iii) and WT cells (Panel iv). The old parental MAC and new MICs are indicated by red and white arrowheads, respectively. The developing MAC is indicated by white circle. Stages shown in panels i–iv correspond to those shown in (A) (DNA-FISH results for *emit2*Δ and *rib1*Δ are identical to that of *emit1*Δ, and are thus not shown). Knockout mutants with DNA elimination defects were unable to produce viable sexual progeny (Table S1).

(D) scnRNA production in WT, *emit1*Δ, *emit2*Δ, and *rib1*Δ cells. Numbers indicate the number of hours after induction of conjugation.

(E) dsRNA immunostaining in WT, *emit1*Δ, *emit2*Δ, *rib1*Δ, and *dcl1*Δ cells.

(F) Northern blotting analysis of MIC genome transcription. Panel i: Schematic diagram of the M-element locus. Flanking MAC-destined sequences are shown by the black horizontal line. IES regions that are eliminated in the progeny MAC are shown by the black cylinder. The positions and directions of T7 and T3 promoters used for in vitro transcription of the RNA probes for the sense (+) and antisense (−) strands are indicated by arrows. Panels ii, iii: Northern blotting analysis of ncRNAs produced from the MIC using RNA probes complementary to the (+) and (−) strand of the M-element. The time after induction of conjugation is indicated. Separated lanes in panels ii and iii are cropped from the same exposure of the same blot (uncropped images are shown in Figure S3A). Panel iv: The membrane from panel iii was reprobed with a DNA probe complementary to the exon region of a conjugation-specific gene, *PDD1*, which served as the conjugation and loading control.

Figure 3. Emit1, Emit2, and Rib1 interact with Pol II and Med subunits.

(A) Panel i: *EMIT1*, *EMIT2*, and *RIB1* are specifically expressed during conjugation. The *x*-axis indicates different stages of *Tetrahymena* life cycle: Vegetative growth (G), starvation, and conjugation. Numbers indicate the time after induction of conjugation. The *y*-axis indicates mRNA expression in arbitrary units (AU) (values lower than 100 AU indicate no expression). Panels ii–iv: Western blot analyses of HA-tagged Emit1, Emit2, and Rib1 indicate that they are conjugation-specific proteins. Twi1 is a marker for conjugating cells. Alpha-tubulin (Tub) is the loading control.

(B) Panel i: Venn diagram of Emit1, Emit2, and Rib1 interacting proteins, as identified by immunoprecipitation-coupled mass spectrometry (IP-MS) analyses. Panels ii, iii: Network diagram of IP-MS data. High-confidence protein interactions are indicated by arrows between bait and prey proteins. For more details, see Table S2.

Figure 4. Emit1, Emit2, and Rib1 colocalize with Pol II and Med in the meiotic MIC.

(A) Co-immunostaining of HA-tagged Med31 and Rpb3 at different stages of

Tetrahymena life cycle (see also Figure S4).

(B) Left: Schematic view of chromosomal organization in the elongated MIC at meiotic prophase, showing the centromeres (C) and telomeres (T) clustered at opposite ends. Right: Localization of *REP2* IESs (yellow), Pol II subunit Rpb3 (red), Cna1 centromeric protein (green, marks the centromeric tip), and dsRNA (magenta) in the WT MIC at early (stage III) to mid (stage IV) prophase. The presumable centromeric and telomeric tip of MICs in the right panels are also labeled C and T, respectively.

(C) Colocalization of HA-tagged Med31, Emit1, Emit2, and Rib1 with Rpb3 (Pol II) in meiotic prophase MICs (see also Figures S4 and S5).

Figure 5. Deletion of *EMIT1*, *EMIT2*, and *RIB1* disrupts the Pol II and Med localization in the MIC.

(A) The Pol II subunit Rpb3 is abundant in the centromere-proximal half of the stage III MIC and in the telomere tip of the stage IV MIC in the WT, but is homogenously distributed in the mutants.

(B) Deletion of *EMIT1* and *EMIT2* prevents the Med31 incorporation into the MIC, and *RIB1* deletion disrupts the specific localization of Med31 in the MIC.

Figure 6. Interdependent localization of Emit1, Emit2, and Rib1 in the meiotic MIC.

(A) Localization of Emit2 and Rib1 in *emit1*Δ.

(B) Localization of Emit1 and Rib1 in *emit2*Δ.

(C) Co-immunostaining of Rib1 and a MIC-specific nucleoporin, MicNup98A, in WT and *emit2*Δ cells (see also Figure S6). A line profile of fluorescence intensity along the white dash line shows Rib1 localization to the MIC periphery in *emit2*Δ.

(D) Emit1 and Emit2 localization in *rib1*Δ cells.

Table

Table 1. MIC localization of the investigated proteins in *emit1*Δ, *emit2*Δ, and *rib1*Δ cells

Protein	WT	*emit1*Δ	*emit2*Δ	*rib1*Δ
Emit1	Enriched in pericentromeric and telomeric regions	–	Not present	Uniform
Emit2	Same as above	Uniform	–	Uniform
Rib1	Same as above	Same as WT	At the periphery	–
Mediator complex (Med31)	Same as above	Not present	Not present	Uniform
RNA polymerase II (Rpb3)	Same as above	Uniform	Uniform	Uniform

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Miao Tian (tian.miao@univie.ac.at).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

***Tetrahymena* strains and culture conditions**

WT *Tetrahymena thermophila* B2086 (mating type II) and CU428 (mating type VII) were obtained from the *Tetrahymena* Stock Center at Cornell University. Other *Tetrahymena* strains used in this work are listed in the Key Resources Table. *Tetrahymena* cells were grown in modified Neff medium (0.25% proteose peptone, 0.25% yeast extract, 0.5% glucose, 33.3 μM $FeCl_3$) at 30°C without shaking [53]. To make cells competent for mating, they were starved at a concentration of ~3×10^5 cells/mL in 10 mM Tris-HCl (pH7.5) at 30°C for at least 16 h. Conjugation was induced by mixing equal amounts of cells of different mating types. To induce the expression of mCherry-tagged MicNup98A, 0.5 μg/mL of $CdCl_2$ was added to the cells during starvation. $CdCl_2$ was removed by washing cells with 10 mM Tris-HCl (pH7.5) before induction of conjugation.

METHOD DETAILS

Generation of somatic gene knockout strains

Plasmid constructs used for creating the somatic gene knockout strains were generated as previously described [54, 55]. Sequences of all primers used are listed in Table S3. Briefly, to generate the plasmid construct for deleting *EMIT1* from the somatic genome, the 5' and 3' flanking regions of *EMIT1* were amplified from genomic DNA by PCR using primer pairs EMIT1_5UTRf2904_Not1/EMIT1_5UTRr3731_N4 and EMIT1_3UTR_f6283_N4/EMIT1_3UTR_r7101_Not1, respectively. Q5 High-Fidelity DNA Polymerase (New England Biolabs, Beverly, MA, USA) was used for PCR. Due to the presence of overlapping sequences, the *EMIT1* flanking sequences and the *neo4* cassette [56] released from pNeo4_SmaI [55] by SmaI digestion were cloned into the NotI site of pBluescript SK(-) via Gibson assembly method [57], using a NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, Beverly, MA, USA). The *EMIT2* knockout construct containing the *neo4* cassette was also generated in this way. The *RIB1* and *EMIT2* knockout constructs containing the *chx* cassette [55] were generated in a similar way. The only difference was that the gene-specific flanking sequences were fused with the *chx* cassette released by SmaI digestion of pChx_SmaI. To generate the DNA fragment for deleting *RIB1* (using the *neo4* cassette), the 5' and 3' flanking regions of *RIB1* were amplified by PCR using primer pairs RIB1KO 5´FW/RIB1KO 5´RV and RIB1KO 3´FW/RIB1KO 3´RV, respectively. Also, the *neo4* cassette was released from pNeo4 by SmaI digestion. Due to the presence of overlapping sequences, these three fragments were assembled via overlapping PCR, using the primer pair RIB1KO 5´FW/ RIB1KO

3´RV.

Before the transformation, plasmid constructs for knocking out *EMIT1* (using the *neo4* cassette), *EMIT2* (using the *neo4* or *chx* cassette), and *RIB1* (using the *chx* cassette) were linearized by NotI digestion. The linearized plasmid constructs or the DNA fragment assembled via overlapping PCR were introduced into starved B2086 and CU428 cells via biolistic transformation, and transformants were selected with increasing paromomycin or cycloheximide (Sigma-Aldrich, St Louis, MO, USA) concentration [58] until all MAC loci were replaced via phenotypic assortment [59]. Somatic gene knockout was confirmed by RT-PCR (Figure S3D). Viability testing of sexual progeny was done as previously described [16].

Bioinformatics analyses and visualization

Tandem amino acid repeats in Rib1 and other proteins were detected by using XSTREAM algorithm [60]. Motif logos were generated by analyzing the aligned tandem repeats using WebLogo [61]. Rib1 sequence features were illustrated by using IBS [62]. Prion-like domains and intrinsically disordered regions in Rib1, Med22, and Med31 were predicted by using PLAAC [63] and PONDER®VLS2 [64], respectively. *EMIT1*, *EMIT2*, and *RIB1* expression data were retrieved from TetraFGD database (http://tfgd.ihb.ac.cn/) [65].

Generation of epitope-tagged strains

Plasmid constructs used for creating strains that express C-terminally HA-tagged Emit2, Rib1, or Med31 from the endogenous promoter were generated as previously described [55]. In brief, to add a codon-optimized HA tag-coding sequence to the C terminus of the *EMIT2* somatic ORF, and a *neo4* or a *chx* cassette into its 3' flanking sequence, a DNA fragment was amplified from the C-terminus of its ORF using the primer pair EMIT2_CDSf2617_Not1/EMIT2_CDSr3434_HA. Also, two adjacent DNA fragments were amplified from the 3' flanking region using the primer pairs EMIT2_3UTRf3435_HA/EMIT2_3UTRr3874_N4, and EMIT2_3UTRf3869_N4/EMIT2_3UTRr4659_Not1, respectively. Due to the presence of overlapping sequences, the above three DNA fragments from the *EMIT2* locus, together with a *neo4* cassette (prepared as described above) were fused and cloned into the NotI site of pBluescript SK(-) using Gibson assembly. Plasmid constructs used for making strains expressing C-terminally HA-tagged Rib1 and Med31 using the *chx* cassette were also generated in this way. For transformation, these constructs were linearized by NotI digestion and introduced into starved WT cells or knockout mutants via biolistic transformation as described above.

Plasmid constructs used for creating strains expressing C-terminally HA or GFP-tagged Emit1 and Rib1 from the endogenous promoter were generated as previously described, with slight modifications [54]. In brief, for *EMIT1*, DNA fragments were amplified from the ORF and the 3' flanking region using the

primer pairs EMIT1_CDSf5599_Sac1Gib/EMIT1_CDSr6234_BamH1Gib and EMIT1_3UTRf6283_Xho1Gib/EMIT1_3UTRr7101_Kpn1Gib, respectively. Due to the presence of overlapping sequences, they were fused with an HA-tag coding sequence containing DNA fragment released from pHA-neo4 using BamHI and XhoI double digestion and cloned into the SacI and KpnI sites of pBluescript SK(+) using Gibson assembly. The construct used for C-terminal GFP tagging of the *RIB1* gene was generated in a similar way. The major difference was that the *RIB1* ORF sequence and 3' flanking sequence were fused with an EGFP coding sequence containing a DNA fragment released from pEGFP-neo4 using BamHI and XhoI double digestion. To generate the plasmid construct for creating strain expressing N-terminally mCherry-tagged MicNup98A, the GFP coding sequence of pBNMB1-EGFP [66] was first replaced with the mCherry coding sequence cloned from pmCherry-neo4. Meanwhile, DNA fragments were amplified from the 5' flanking sequence and the N-terminus ORF of *MICNUP98A* (*TTHERM_01080600*) using primer pairs 1MicNup98A/2MicNup98A and 3MicNup98A/4MicNup98A, respectively. Subsequently, the 5' flanking sequence and the N-terminus ORF sequence were cloned into the SacI-SalI sites and BamHI-KpnI sites of the modified pBNMB1-EGFP plasmid, respectively, using Gibson assembly. For transformation, all three constructs were linearized by SacI and KpnI double digestion and introduced into starved WT cells or knockout mutants via biolistic transformation as described above.

RNA analyses

Total RNA was extracted from 5 mL samples of conjugating WT or mutant cells (~3×10^5 cells/mL) using an RNeasy Plus Mini Kit (Qiagen, Valencia, CA, USA). To generate cDNA for RT-PCR, 5 μg RNA was first treated with RNase-free DNase I (Thermo Fisher Scientific, Waltham, MA, USA) to remove genomic DNA. This preparation was used for cDNA synthesis: the first-strand cDNA was synthesized using random hexamers (Integrated DNA Technologies, Leuven, Belgium) and a RevertAid H Minus Reverse Transcriptase kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions, but with RiboLock RNase Inhibitor (Thermo Fisher Scientific, Waltham, MA, USA) added. OneTaq DNA Polymerase (New England Biolabs, Beverly, MA, USA) was used to amplify DNA fragments from cDNA. Sequences of primers used for amplifying cDNA fragments of *EMIT1*, *EMIT2*, *RIB1*, *DCL1*, *TWI1*, *SPO11* [67], *PARS11* [58], *TPB6*, *IES737* and *IES4092* were listed in Table S3. Some of them are the same as previously described [37, 58]. To examine scnRNA production, 10 μg total RNA was separated by 12% polyacrylamide-urea gel electrophoresis, as previously described [25]. RNA was stained with ethidium bromide.

To generate RNA probes complementary to the sense and antisense strands of the MIC M-element for northern blotting, an 1154 bp DNA fragment was amplified from this region by overlapping PCR using primers T7-NB-

Melement_F, M_alt06IES_R121066, M_alt06IES_F121041, and T3-NB-Melement_R. During PCR, T7 and T3 promoter sequences were fused to the 5′ and 3′ termini of the M-element, respectively. The DNA fragment was then used as the template for *in vitro* transcription using either T7 or T3 RNA polymerase (New England Biolabs, Beverly, MA, USA) according to the manufacturer's instructions. During *in vitro* transcription, ^{32}P-ATP (6000 Ci/mM; Hartmann Analytic, Braunschweig, Germany) was incorporated; unincorporated ^{32}P-ATP was removed by passing the reaction mixture through an RNase-free Sephadex G-50 Quick Spin Column (Roche, Indianapolis, IN, USA). The template DNA was then removed using RNase-Free DNase I (Thermo Fisher Scientific, Waltham, MA, USA).

Before northern blotting, 15 μg total RNA isolated from *Tetrahymena* cells was denatured and separated by electrophoresis on a 1% agarose gel containing 6.66% formaldehyde. RNA electrophoresis was performed in 1× MOPS buffer (containing 20 mM 4-morpholinepropanesulfonic acid, 2 mM sodium chloride, and 1 mM ethylenediaminetetraacetic acid, pH7.0). After hydrolysis with 50 mM sodium hydroxide, the gel was neutralized with 200 mM sodium acetate solution (pH4.0) and RNA was blotted onto a Hybond-N+ nylon membrane (GE Healthcare, Piscataway, NJ, USA) in 20× SSC buffer (3 M sodium chloride and 0.3 M sodium citrate). After UV-crosslinking the RNA to the membrane, the radiolabeled RNA probe was denatured and hybridized to the RNA at 68°C in ULTRAhyb Ultrasensitive Hybridization Buffer (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions. The signal was detected using a storage phosphor screen (Imaging Screen K, Bio-Rad, Hercules, CA, USA) and scanned with a Typhoon 9200 image analyzer (GE Healthcare, Piscataway, NJ, USA).

To generate the *PDD1* DNA probe, a 302 bp DNA fragment (991–1292 bp of the *PDD1* ORF) was PCR amplified from the last exon of the *PDD1* gene using the primer pair PDD1_NB_Exon_F2514/PDD1_NB_Exon_R2815 and radiolabeled by random priming with ^{32}P-dATP, random hexamers, and a Klenow Fragment (exo-) kit (Thermo Fisher Scientific, Waltham, MA, USA). After denaturing, the DNA probe was hybridized to RNA on the membrane at 42°C in ULTRAhyb. The signal was detected as described above.

Western blotting and cytological staining

For western blotting, crude proteins were extracted from 1.5 mL cells by trichloroacetic acid precipitation, separated by SDS-PAGE and analyzed by western blotting using the indicated antibodies: HA-tagged proteins, mouse anti-HA monoclonal antibody (1:1000 dilution; clone HA-7; Sigma-Aldrich, St Louis, MO, USA); alpha-tubulin, mouse anti-tubulin-α Ab-2 monoclonal antibody (1:5000 dilution; Clone DM1A; Lab Vision, Fremont, CA, USA); Twi1, rabbit anti-Twi1p polyclonal antibody (1: 5000 dilution) [68]. HRP-conjugated anti-mouse IgG secondary antibody (1:5000 dilution; Bio-Rad, Hercules, CA, USA) or HRP-conjugated anti-rabbit IgG secondary antibody (1:5000 dilution;

Cell Signaling Technology, Danvers, MA, USA) was used for detection.

For immunostaining, conventional cell fixation was performed as previously described [69]. For pre-fixation detergent treatment, cells were first treated with 1% Triton X-100 containing 0.37% formaldehyde on ice for 25 minutes, then formaldehyde was added to a final concentration of 3.7%. The cell suspension was applied to a slide and air-dried. Slides were washed with 1× PBS and 1× PBS containing 0.05% Triton X-100. HA-tagged proteins were detected with mouse anti-HA monoclonal antibody (1:500 dilution) or rabbit anti-HA polyclonal antibody (1:200 dilution; Sigma-Aldrich, St Louis, MO, USA). mCherry-tagged MicNup98A was detected with rabbit anti-dsRed polyclonal antibody (1:100 dilution; Clontech Laboratories, Mountain View, CA, USA). Rpb3 was detected with a custom rabbit anti-*Tetrahymena* Rpb3 polyclonal antibody (1:100 dilution) [34]. Dmc1 and Rad51 were stained with mouse anti-Rad51/Dmc1 monoclonal antibody (1:50 dilution; Lab Vision, Fremont, CA, USA). For immunostaining of dsRNA, 3 mL of conjugating cells were harvested and resuspended in an equal volume of 1× PBS. 900 μL 1.5× PBS containing 2.5% Triton X-100 was then added to the resuspended cells and gently mixed for exactly three seconds, then 480 μL of 37% formaldehyde was added to the mixture to fix cells at room temperature for 30 minutes. Fixed cells were pelleted and resuspended in 300 μL of a fixative solution containing 4% paraformaldehyde and 3.4% sucrose. Fixed cells were spread onto a poly-L-lysine coated slide and dried. Immunostaining of dsRNA was carried out using mouse anti-dsRNA monoclonal antibody (1:1000 dilution; J2; SCICONS, Hungary), as previously described [23]. Primary antibodies were detected with Cy3-conjugated goat anti-rabbit IgG secondary antibody (1:1000 dilution; GE Healthcare, Piscataway, NJ, USA) or FITC-conjugated goat anti-rabbit IgG secondary antibody (1:200 dilution; Sigma-Aldrich, St Louis, MO, USA) or Alexa Fluor 488-conjugated goat anti-mouse IgG secondary antibody (1:1000 dilution; Thermo Fisher Scientific, Waltham, MA, USA). After immunostaining, slides were mounted under a coverslip in Vectashield anti-fading agent (Vector Laboratories, Burlingame, CA, USA) supplemented with 0.5 mg/mL DAPI for inspection by fluorescence microscopy.

For Cna1 and Rpb3 co-immunostaining, 5 mL of cells were harvested at meiotic prophase and resuspended in 500 μL Carnoy's fixative (6:3:2 mixture of methanol, chloroform, acetic acid), and then spread onto a slide as previously described [69]. Rpb3 was detected with the rabbit anti-Rpb3 polyclonal antibody (1:100 dilution) and Cy3-conjugated goat anti-rabbit IgG secondary antibody (1:1000 dilution). Images of Rpb3 stained cells were taken and the coordinates were recorded. Slides were washed with 1× PBS and 1× PBS containing 0.05% Triton X-100. After drying, a custom rabbit anti-*Tetrahymena* Cna1 polyclonal antibody (1:200 dilution; kindly provided by Harmit Malik) [70] was applied on the slide and detected with FITC-conjugated goat anti-rabbit IgG secondary antibody (1:200 dilution). Cells were located by their coordinates and their Cna1 signals were recorded. To detect *REP2* IESs by FISH, cells in meiotic prophase

were fixed with Carnoy's fixative and cells at a late stage of conjugation (32 h after induction of meiosis) were fixed with the conventional fixation method as described above. Fixed cells were spread onto slides and dried for at least three days. *REP2* FISH was performed as previously described [71]. For live-cell imaging, Hoechst 33342 (Thermo Fisher Scientific, Waltham, MA, USA) was added to the mating cells at a final concentration of 0.5 μg/mL to stain the nuclei. Z-stack images taken from DAPI- and immuno-stained nuclei (except for cells fixed using Carnoy's fixative and live cells) were deconvolved, projected and colored as previously described [72]. Fluorescence line profile analyses of Rib1, DAPI, and MicNup98A were carried out using the Line Profile tool in AutoQuant X3 software.

Immunoprecipitation and mass spectrometry

Immunoprecipitation (IP) of HA-tagged proteins from conjugating cells was based on a published protocol [73]. Briefly, conjugating cells with both partners expressing C-terminally HA-tagged Emit1, Emit2, or Rib1 were grown to $OD_{540\ nm} \approx 0.7$ in 200 mL growth medium, washed with 10 mM Tris-HCl (pH7.5), and starved in the same buffer. At 3 h after mixing, the pairing rate and progression of conjugation were examined. IP was performed only if >80% of cells were undergoing conjugation and their MICs were elongating. The equal amount of WT strains were used for the control IP. Emit2 and control IPs were performed with two and three biological replicates, respectively.

After IP, the protein samples were run into an SDS-PAGE gel for 2 cm and then the Coomassie-Blue-Stained gel pieces were excised for tryptic digestion. After reduction and alkylation of thiols using dithiothreitol and iodoacetamide (Sigma-Aldrich, St Louis, MO, USA), the proteins were digested with trypsin (Promega, Madison, WI, USA) and then the solutions with tryptic peptides were desalted on custom-made C18 stagetips, as previously described [74, 75]. Tryptic peptides were separated on an Ultimate 3000 RSLC nano-flow chromatography system and analyzed on a Q Exactive HF Orbitrap mass spectrometer, equipped with a Proxeon nanospray source (all from Thermo Fisher Scientific, Waltham, MA, USA). Raw data were processed with MaxQuant software package [76] searching against a merged *Tetrahymena* protein sequence database, which contains the latest *Tetrahymena* protein database (Version 2014, http://ciliate.org) and an older version (Version 2008, https://www.jcvi.org/) [77]. Tryptic digestion specificity was set to allowing two missed cleavages. Results were filtered at a protein and peptide false discovery rate of 1%. Search results were further processed with the Perseus software package [78]. The detailed instrument settings for the mass spectrometry analysis and proteomics raw data have been deposited to the ProteomeXchange Consortium via the PRIDE [79] partner repository with the dataset identifier PXD012372. A network diagram of the IP-MS data was generated based on the high-confidence interactions using Cytoscape (version, 3.6.0) [80].

QUANTIFICATION AND STATISTICAL ANALYSIS

The SAINTexpress algorithm (v3.6; parameter set to SAINTexpress-spc.exe –L4) was used to identify high-confidence protein–protein interactions from mass spectrometry data [81, 82]. In this analysis, protein interactions with a Bayesian False Discovery Rate of ≤ 0.05 were considered highly confident. The number of cells used for testing the viability of sexual progeny and for evaluating the progression of conjugation can be found in the table or figure legends.

Supplemental Information

Table S1. Sexual progeny viability in WT and *emit1*Δ, *emit2*Δ, and *rib1*Δ cells. Related to Figure 2.

Table S2. SAINT analysis of mass spectrometry data of immunoprecipitation samples. Related to Figure 3.

Table S3. Oligonucleotides used in this study. Related to STAR Methods, Key Resources Table.

(Table S2 and S3 are supplied as separated Excel files)

References

1. Hedges, D.J., and Deininger, P.L. (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutation research *616*, 46-59.
2. Huang, C.R., Burns, K.H., and Boeke, J.D. (2012). Active transposition in genomes. Annual review of genetics *46*, 651-675.
3. Klein, S.J., and O'Neill, R.J. (2018). Transposable elements: genome innovation, chromosome diversity, and centromere conflict. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *26*, 5-23.
4. Vorechovsky, I. (2010). Transposable elements in disease-associated cryptic exons. Human genetics *127*, 135-154.
5. Toth, K.F., Pezic, D., Stuwe, E., and Webster, A. (2016). The piRNA Pathway Guards the Germline Genome Against Transposable Elements. Advances in experimental medicine and biology *886*, 51-77.
6. Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet *8*, 272-285.
7. Chen, X.M. (2009). Small RNAs and Their Roles in Plant Development. Annu Rev Cell Dev Bi *25*, 21-44.
8. Iwasaki, Y.W., Siomi, M.C., and Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions. Annual review of biochemistry *84*, 405-433.
9. Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z., et al. (2013). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol Cell *50*, 67-81.
10. Goriaux, C., Desset, S., Renaud, Y., Vaury, C., and Brasset, E. (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO reports *15*, 411-418.
11. Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54-59.
12. Chen, Y.C.A., Stuwe, E., Luo, Y.C., Ninova, M., Le Thomas, A., Rozhavskaya, E., Li, S.S., Vempati, S., Laver, J.D., Patel, D.J., et al. (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. Mol Cell *63*, 97-109.
13. Chalker, D.L., Meyer, E., and Mochizuki, K. (2013). Epigenetics of ciliates. Cold Spring Harbor perspectives in biology *5*, a017764.
14. Eisen, J.A., Coyne, R.S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J.R., Badger, J.H., Ren, Q., Amedeo, P., Jones, K.M., et al. (2006). Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS biology *4*, e286.
15. Hamilton, E.P., Kapusta, A., Huvos, P.E., Bidwell, S.L., Zafar, N., Tang, H., Hadjithomas, M., Krishnakumar, V., Badger, J.H., Caler, E.V., et al. (2016). Structure of the germline genome of Tetrahymena thermophila and relationship to the massively rearranged somatic genome. Elife *5*.
16. Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689-699.

17. Yao, M.C., Fuller, P., and Xi, X. (2003). Programmed DNA deletion as an RNA-guided system of genome defense. Science *300*, 1581-1584.
18. Sugai, T., and Hiwatashi, K. (1974). Cytologic and autoradiographic studies of the micronucleus at meiotic prophase in Tetrahymena pyriformis. J Protozool *21*, 542-548.
19. Noto, T., Kataoka, K., Suhren, J.H., Hayashi, A., Woolcock, K.J., Gorovsky, M.A., and Mochizuki, K. (2015). Small-RNA-Mediated Genome-wide trans-Recognition Network in Tetrahymena DNA Elimination. Mol Cell *59*, 229-242.
20. Chalker, D.L., and Yao, M.C. (2001). Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila. Genes & development *15*, 1287-1298.
21. Mochizuki, K., and Gorovsky, M.A. (2004). RNA polymerase II localizes in Tetrahymena thermophila meiotic micronuclei when micronuclear transcription associated with genome rearrangement occurs. Eukaryotic cell *3*, 1233-1240.
22. Schoeberl, U.E., Kurth, H.M., Noto, T., and Mochizuki, K. (2012). Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in Tetrahymena. Genes & development *26*, 1729-1742.
23. Woo, T.T., Chao, J.L., and Yao, M.C. (2016). Dynamic distributions of long double-stranded RNA in Tetrahymena during nuclear development and genome rearrangements. Journal of cell science *129*, 1046-1058.
24. Malone, C.D., Anderson, A.M., Motl, J.A., Rexer, C.H., and Chalker, D.L. (2005). Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila. Molecular and cellular biology *25*, 9151-9164.
25. Mochizuki, K., and Gorovsky, M.A. (2005). A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. Genes & development *19*, 77-89.
26. Mochizuki, K., and Kurth, H.M. (2013). Loading and pre-loading processes generate a distinct siRNA population in Tetrahymena. Biochemical and biophysical research communications *436*, 497-502.
27. Schoeberl, U.E., and Mochizuki, K. (2011). Keeping the soma free of transposons: programmed DNA elimination in ciliates. J Biol Chem *286*, 37045-37052.
28. Noto, T., and Mochizuki, K. (2017). Whats, hows and whys of programmed DNA elimination in Tetrahymena. Open biology *7*.
29. Kurth, H.M., and Mochizuki, K. (2009). 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. Rna *15*, 675-685.
30. Woehrer, S.L., Aronica, L., Suhren, J.H., Busch, C.J., Noto, T., and Mochizuki, K. (2015). A Tetrahymena Hsp90 co-chaperone promotes siRNA loading by ATP-dependent and ATP-independent mechanisms. The EMBO journal *34*, 559-577.
31. Fillingham, J.S., Thing, T.A., Vythilingum, N., Keuroghlian, A., Bruno, D., Golding, G.B., and Pearlman, R.E. (2004). A non-long terminal repeat retrotransposon family is restricted to the germ line micronucleus of the ciliated protozoan Tetrahymena thermophila. Eukaryotic cell *3*, 157-169.
32. Mochizuki, K., and Gorovsky, M.A. (2005). A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic

prophase. Genes & development *19*, 77-89.

33. Bourbon, H.M. (2008). Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. Nucleic Acids Res *36*, 3993-4008.
34. Kataoka, K., and Mochizuki, K. (2017). Heterochromatin aggregation during DNA elimination in Tetrahymena is facilitated by a prion-like protein. Journal of cell science *130*, 480-489.
35. Loidl, J., Lukaszewicz, A., Howard-Till, R.A., and Koestler, T. (2012). The Tetrahymena meiotic chromosome bouquet is organized by centromeres and promotes interhomolog recombination. Journal of cell science *125*, 5873-5880.
36. Lukaszewicz, A., Howard-Till, R.A., Novatchkova, M., Mochizuki, K., and Loidl, J. (2010). MRE11 and COM1/SAE2 are required for double-strand break repair and efficient chromosome pairing during meiosis of the protist Tetrahymena. Chromosoma *119*, 505-518.
37. Feng, L., Wang, G., Hamilton, E.P., Xiong, J., Yan, G., Chen, K., Chen, X., Dui, W., Plemens, A., Khadr, L., et al. (2017). A germline-limited piggyBac transposase gene is required for precise excision in Tetrahymena genome rearrangement. Nucleic Acids Res *45*, 9481-9502.
38. Kushida, Y., Takaine, M., Nakano, K., Sugai, T., and Numata, O. (2015). Dynamic Change of Cellular Localization of Microtubule-Organizing Center During Conjugation of Ciliate Tetrahymena thermophila. Zoological science *32*, 25-32.
39. Malone, C.D., Falkowska, K.A., Li, A.Y., Galanti, S.E., Kanuru, R.C., LaMont, E.G., Mazzarella, K.C., Micev, A.J., Osman, M.M., Piotrowski, N.K., et al. (2008). Nucleus-specific importin alpha proteins and nucleoporins regulate protein import and nuclear division in the binucleate Tetrahymena thermophila. Eukaryotic cell *7*, 1487-1499.
40. Papazyan, R., Voronina, E., Chapman, J.R., Luperchio, T.R., Gilbert, T.M., Meier, E., Mackintosh, S.G., Shabanowitz, J., Tackett, A.J., Reddy, K.L., et al. (2014). Methylation of histone H3K23 blocks DNA damage in pericentric heterochromatin during meiosis. Elife *3*.
41. Boeynaems, S., Alberti, S., Fawzi, N.L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B., Van Den Bosch, L., et al. (2018). Protein Phase Separation: A New Phase in Cell Biology. Trends in cell biology *28*, 420-435.
42. Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*.
43. Franzmann, T., and Alberti, S. (2018). Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. J Biol Chem.
44. Lee, S.R., Talsky, K.B., and Collins, K. (2009). A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis. Rna *15*, 1363-1374.
45. D'Alessio, J.A., Ng, R., Willenbring, H., and Tjian, R. (2011). Core promoter recognition complex changes accompany liver development. P Natl Acad Sci USA *108*, 3906-3911.
46. Yin, J.W., Liang, Y., Park, J.Y., Chen, D., Yao, X., Xiao, Q., Liu, Z., Jiang, B., Fu, Y.,

Bao, M., et al. (2012). Mediator MED23 plays opposing roles in directing smooth muscle cell and adipocyte differentiation. Genes & development *26*, 2192-2205.

47. Lu, C., and Fuller, M.T. (2015). Recruitment of Mediator Complex by Cell Type and Stage-Specific Factors Required for Tissue-Specific TAF Dependent Gene Activation in an Adult Stem Cell Lineage. Plos Genet *11*, e1005701.
48. Strahl, B.D., Ohba, R., Cook, R.G., and Allis, C.D. (1999). Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. Proc Natl Acad Sci U S A *96*, 14967-14972.
49. Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., et al. (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.
50. Pane, A., Jiang, P., Zhao, D.Y., Singh, M., and Schupbach, T. (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. The EMBO journal *30*, 4601-4615.
51. Kim, Y.J., Zheng, B., Yu, Y., Won, S.Y., Mo, B., and Chen, X. (2011). The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. The EMBO journal *30*, 814-822.
52. Oya, E., Kato, H., Chikashige, Y., Tsutsumi, C., Hiraoka, Y., and Murakami, Y. (2013). Mediator directs co-transcriptional heterochromatin assembly by RNA interference-dependent and -independent pathways. Plos Genet *9*, e1003677.
53. Cassidy-Hanley, D., Bowen, J., Lee, J.H., Cole, E., VerPlank, L.A., Gaertig, J., Gorovsky, M.A., and Bruns, P.J. (1997). Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment. Genetics *146*, 135-147.
54. Kataoka, K., Schoeberl, U.E., and Mochizuki, K. (2010). Modules for C-terminal epitope tagging of Tetrahymena genes. Journal of microbiological methods *82*, 342-346.
55. Gao, S., Xiong, J., Zhang, C.C., Berquist, B.R., Yang, R.D., Zhao, M., Molascon, A.J., Kwiatkowski, S.Y., Yuan, D.X., Qin, Z.H., et al. (2013). Impaired replication elongation in Tetrahymena mutants deficient in histone H3 Lys 27 monomethylation. Genes & development *27*, 1662-1679.
56. Mochizuki, K. (2008). High efficiency transformation of Tetrahymena using a codon-optimized neomycin resistance gene. Gene *425*, 79-83.
57. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.
58. Tian, M., and Loidl, J. (2018). A chromatin-associated protein required for inducing and limiting meiotic DNA double-strand break formation. Nucleic Acids Res *46*, 11822-11834.
59. Sonneborn, T.M. (1974). Tetrahymena pyriformis. In Handbook of genetics, R.C. King, ed. (New York: Springer), pp. 433-467.
60. Newman, A.M., and Cooper, J.B. (2007). XSTREAM: a practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. Bmc Bioinformatics *8*, 382.
61. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a

sequence logo generator. Genome Res *14*, 1188-1190.

62. Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., et al. (2015). IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics *31*, 3359-3361.
63. Lancaster, A.K., Nutter-Upham, A., Lindquist, S., and King, O.D. (2014). PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. Bioinformatics *30*, 2501-2502.
64. Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., and Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. Bmc Bioinformatics *7*, 208.
65. Xiong, J., Lu, Y., Feng, J., Yuan, D., Tian, M., Chang, Y., Fu, C., Wang, G., Zeng, H., and Miao, W. (2013). Tetrahymena functional genomics database (TetraFGD): an integrated resource for Tetrahymena functional genomics. Database : the journal of biological databases and curation *2013*, bat008.
66. Akematsu, T., Fukuda, Y., Garg, J., Fillingham, J.S., Pearlman, R.E., and Loidl, J. (2017). Post-meiotic DNA double-strand breaks occur in Tetrahymena, and require Topoisomerase II and Spo11. Elife *6*.
67. Mochizuki, K., Novatchkova, M., and Loidl, J. (2008). DNA double-strand breaks, but not crossovers, are required for the reorganization of meiotic nuclei in Tetrahymena. Journal of cell science *121*, 2148-2158.
68. Aronica, L., Bednenko, J., Noto, T., DeSouza, L.V., Siu, K.W., Loidl, J., Pearlman, R.E., Gorovsky, M.A., and Mochizuki, K. (2008). Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in Tetrahymena. Genes & development *22*, 2228-2241.
69. Loidl, J., and Scherthan, H. (2004). Organization and pairing of meiotic chromosomes in the ciliate Tetrahymena thermophila. Journal of cell science *117*, 5791-5801.
70. Cervantes, M.D., Xi, X., Vermaak, D., Yao, M.C., and Malik, H.S. (2006). The CNA1 histone of the ciliate Tetrahymena thermophila is essential for chromosome segregation in the germline micronucleus. Molecular biology of the cell *17*, 485-497.
71. Noto, T., Kurth, H.M., Kataoka, K., Aronica, L., DeSouza, L.V., Siu, K.W., Pearlman, R.E., Gorovsky, M.A., and Mochizuki, K. (2010). The Tetrahymena argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus. Cell *140*, 692-703.
72. Howard-Till, R.A., Lukaszewicz, A., and Loidl, J. (2011). The recombinases Rad51 and Dmc1 play distinct roles in DNA break repair and recombination partner choice in the meiosis of Tetrahymena. Plos Genet *7*, e1001359.
73. Tian, M., Yang, W., Zhang, J., Dang, H., Lu, X., Fu, C., and Miao, W. (2017). Nonsense-mediated mRNA decay in Tetrahymena is EJC independent and requires a protozoa-specific nuclease. Nucleic Acids Res *45*, 6848-6863.
74. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature protocols *1*, 2856-2860.
75. Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature protocols *2*, 1896-1906.

76. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

77. Coyne, R.S., Thiagarajan, M., Jones, K.M., Wortman, J.R., Tallon, L.J., Haas, B.J., Cassidy-Hanley, D.M., Wiley, E.A., Smith, J.J., Collins, K., et al. (2008). Refined annotation and assembly of the Tetrahymena thermophila genome sequence through EST analysis, comparative genomic hybridization, and targeted gap closure. Bmc Genomics *9*, 562.

78. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods *13*, 731-740.

79. 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 Res *44*, D447-D456.

80. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

81. Choi, H., Larsen, B., Lin, Z.Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.C., and Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat Methods *8*, 70-U100.

82. Teo, G., Liu, G., Zhang, J., Nesvizhskii, A.I., Gingras, A.C., and Choi, H. (2014). SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. J Proteomics *100*, 37-43.