universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

Genetic and molecular characterization of the chromatin binding specificity of the HP1 protein variant Rhino

verfasst von / submitted by

Lisa Christine Baumgartner

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 794 685 490
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Molekulare Biologie
Betreut von / Supervisor:	Dr. Julius Brennecke

Acknowledgements

I am lost for words trying to express how deeply grateful I am for the amazing experience that were the last five years, spent doing exciting science with wonderful friends, mentors, and role models. I am truly indebted to everyone who made this possible.

Thank you, Julius, for giving me the opportunity and guidance to grow into the scientist I am today. Thank you for this beautiful, challenging project we conquered together, the many fun discussions and discoveries, and for everything you've taught me. Your true passion and curiosity for biology have been an inspiration, and your smart and thorough way of doing and presenting science will always be something I aspire to maintain in future projects. Thank you for never giving up, for your encouraging words in difficult times, for your trust and patience, and the genuine support I felt from you. You have set an incredibly high bar for future mentors.

My gratitude also goes to the past and present members of the Brennecke lab for the supportive and familial atmosphere that fostered many interesting discussions and fond memories. Special thanks go to Peter Andersen for introducing me to fly work and making sure I had a smooth start in the lab. Many tips and tricks you taught me, both technical and personal, will stay with me for a long time to come. I'm also truly grateful to Dominik for countless hours spent fixing the annotation pipeline or playing around with weird analysis requests and for being a dear friend and constant in the lab. Thank you for your support and for enabling me to acquire skills I will surely make use of many times in the future. Thank you also to Ulrich for introducing me to the world of protein purifications and biochemistry. Despite rather little biological insight gained in this exercise, I learned a lot from you and am grateful for all the time you took to teach me and for every discussion we had about piRNA pathway mysteries. I thank Sebastian, Hannah, and Sarah for joining me for their Master thesis or summer school projects. Working with you and getting the chance to pass on little pieces of lab-wisdom has been a lot of fun, and I hope you will keep equally fond memories of this time. I am also grateful to Mosti, Júlia, Maya, Katherina, Changwei, Laszlo, Arturo, and Maja for fun beer hours, lab outings, coffee breaks, helpful discussions and the friendships that came out of my time in the lab. One of the most constant parts of my PhD time has been the breakfast/coffee club at 8:30. I want to thank all of its members over the years for a good start of the day.

Many members and services of the VBC have supported me in my PhD. My gratitude goes to Peter Ducheck, Balazs Erdi, and Joseph Gokcezade for countless engineered flies and injected constructs, which were crucial to make my project possible. The same holds true for the fantastic NGS, Protein Chemistry, BioOptics, Monoclonal Antibody, and MBS facilities that offered all the support I could hope for. Also the VCDI and EHS teams who ensured our smooth passage through the COVID-19 pandemic were highly appreciated. Many thanks also to Maria Novatchkova for bioinformatics support and to Clemens Plaschka for his generosity in letting me use his equipment and for his input. Great thanks go to all members of the RNA

club for their helpful input and for being a supportive and fun community. I am grateful also to Andi Pauli, Fred Berger, and Dònal O'Carroll for their support and feedback as part of my thesis advisory committee, as well as Joao Matos, Madhav Jagannathan, and Ortrun Mittelsten Scheid for reviewing and examining my thesis. I also acknowledge the generous support of the Böhringer Ingelheim Fonds and Vienna RNA DK in financing my project and providing opportunities to be a part of scientific communities beyond the VBC.

Lastly, I want to thank my friends and family who supported me over the years. I have been lucky to meet many amazing people at the VBC that have enriched my time here in one way or another. I thank all the Hasi-Friends for countless fun memories, for their friendship, and for playing a big part in how much I enjoyed my time in Vienna. I thank Oliver for being my friend and companion for much of this journey. I thank my childhood friends Yasmin, Nadja, and Chantal for their support and friendship, for their advice and perspective, and for celebrating my victories like their own. I thank my parents for supporting me throughout my education and in all my decisions, and for giving me the freedom to become who I want to be. Finally, I thank Jakob for always being there, for sharing frustration and excitement, for giving me something to move towards in this last chapter of my PhD, and for many wonderful moments in the past, present, and future.

Abstract

Transposable elements are abundant constituents of eukaryotic genomes and their activity poses a danger for genome integrity. In animal gonads, transposons are silenced through a small RNA based genome defense system, termed piRNA pathway. In the germline of *Drosophila melanogaster*, the target spectrum of this silencing pathway is defined by the sequence of piRNA precursors originating from heterochromatic piRNA clusters. The *Drosophila* Heterochromatin Protein 1 (HP1) variant Rhino binds to dual-strand piRNA clusters and permits their transcription despite the presence of repressive, heterochromatic histone modifications. With this function Rhino not only holds a central role in the piRNA pathway, but also represents a remarkably specialized member of the HP1 protein family with strongly opposing function compared to canonical HP1a. The specific chromatin binding of Rhino at piRNA clusters is therefore thought to be crucial to maintain transposon silencing and to prevent unwanted piRNA production from ectopic heterochromatic loci. Current models propose a role for maternally inherited piRNAs and histone marks as the basis of Rhino's specific chromatin occupancy at piRNA source loci. However, these factors are insufficient to explain Rhino's binding specificity, which implies the existence of other, undiscovered specificity cues.

My PhD work uncovered Kipferl, a member of the diverse family of zinc finger associated domain (ZAD)-C2H2 zinc finger proteins, as critical Rhino specificity factor in ovaries. In the absence of Kipferl, Rhino is lost from most of its genomic binding sites, resulting in decreased levels of transposon targeting piRNAs and impaired fertility. Kipferl binds to Guanosine-rich DNA motifs and interacts with the Rhino chromo domain, thereby recruiting Rhino to specific loci. Moreover, Kipferl and Rhino stabilize each other's chromatin binding and occupy large chromatin domains together. In *kipferl* mutant flies, Rhino accumulates on pericentromeric Satellite arrays, leading to a characteristic accumulation of Rhino signal at the nuclear envelope of germline nurse cells in immunofluorescence stainings. This redistribution of Rhino to selfish DNA Satellites might indicate a competition for Rhino between different repetitive elements. By guiding Rhino to piRNA clusters and stabilizing it at non-Satellite loci, Kipferl might have evolved to bring balance to the ovarian piRNA pool.

Altogether, my findings reveal that DNA sequence, in addition to the H3K9me3 mark, determines the identity of piRNA source loci. Furthermore, they point to the existence of alternative specificity factors acting in different developmental stages of ovarian development as well as in testes, where Kipferl is not expressed and Rhino's chromatin occupation differs from that observed in ovaries. This work critically advances our understanding of how heterochromatin sub-compartments are specified for HP1 variant protein binding and provides insight into how Rhino might be caught in the crossfire of genetic conflicts.

Zusammenfassung

Transposons sind ein häufiger Bestandteil eukaryotischer DNA und ihre Aktivität stellt eine Gefahr für die Integrität des Genoms dar. In der Keimbahn von Tieren werden Transposons durch ein auf kleinen RNAs basierendes Abwehrsystem, den so genannten piRNA-Signalweg, zum unterdrückt. In *Drosophila melanogaster* wird das Zielspektrum des piRNA-Signalweg durch die Sequenz der piRNA-Vorläufer definiert, die aus heterochromatischen piRNA-Clustern stammen. Die *Drosophila* Heterochromatin Protein 1 (HP1)-Variante Rhino bindet an sogenannte dual-strand piRNA-Cluster und ermöglicht deren Transkription trotz der Anwesenheit repressiver, heterochromatischer Histonmodifikationen. Mit dieser Funktion nimmt Rhino nicht nur eine zentrale Rolle im piRNA-Signalweg ein, sondern stellt auch ein bemerkenswert spezialisiertes Mitglied der HP1-Proteinfamilie dar. Es wird daher angenommen, dass die spezifische Chromatinbindung von Rhino an piRNA-Clustern entscheidend ist für die Aufrechterhaltung der Transposonregulation und die Verhinderung unerwünschter piRNA-Produktion aus ektopischen Regionen. Aktuelle Modelle gehen davon aus, dass maternal vererbte piRNAs und Histonmarkierungen die Grundlage für die spezifische Chromatinbesetzung von Rhino an piRNA-Ursprungsloci bilden. Diese Faktoren reichen jedoch nicht aus, um die Bindungsspezifität von Rhino zu erklären, was die Existenz anderer, unentdeckter Spezifitätsmerkmale voraussetzt.

Im Rahmen meiner Doktorarbeit habe ich Kipferl, ein Mitglied der vielfältigen Familie der Zinkfinger-assoziierten Domäne (ZAD)-C_2H_2-Zinkfingerproteine, als wichtigen Rhino-Spezifitätsfaktor in *Drosophila* Eierstöcken entdeckt. In Abwesenheit von Kipferl geht Rhino an den meisten seiner genomischen Bindungsstellen verloren, was zu einem Verlust von Transposon-bindenden piRNAs und einer Beeinträchtigung der Fruchtbarkeit führt. Kipferl bindet an Guanosin-reiche DNA-Motive und interagiert mit der Rhino-Chromodomäne, wodurch Rhino an spezifische Loci rekrutiert wird. Außerdem stabilisieren sich Kipferl und Rhino gegenseitig und besetzen gemeinsam große Chromatindomänen. In *kipferl*-mutierten Fliegen akkumuliert Rhino auf DNA-Satelliten, was in Immunfluoreszenzfärbungen zu einer charakteristischen Anhäufung von Rhinosignal an der Kernhülle von Keimbahnzellen führt. Diese Umverteilung von Rhino auf egoistische DNA-Satelliten könnte auf einen Kampf um Rhino zwischen verschiedenen repetitiven Elementen hinweisen. Indem Rhino zu piRNA-Clustern gelenkt und an diversen Loci stabilisiert wird, könnte sich Kipferl entwickelt haben, um ein Gleichgewicht im piRNA-Pool der Eierstöcke herzustellen.

Insgesamt zeigen meine Ergebnisse, dass die DNA-Sequenz zusätzlich zur H3K9me3-Markierung die Identität der piRNA-Ursprungsloci bestimmt. Darüber hinaus weisen sie auf die Existenz alternativer Spezifitätsfaktoren hin, die in verschiedenen Entwicklungsstadien der Eierstockentwicklung sowie in den Hoden wirken, wo Kipferl nicht exprimiert wird und die Chromatinbindung von Rhino sich von der in den Eierstöcken unterscheidet. Diese Arbeit trägt entscheidend zu unserem Verständnis darüber bei, wie Heterochromatin-Subkompartimente für die Bindung von HP1-Proteinvarianten spezifiziert werden, und gibt Aufschluss darüber, wie Rhino in das Kreuzfeuer genetischer Konflikte geraten konnte.

Table of Contents

1. Introduction

The genic information required for the proper functioning of any organism in encoded in its DNA (Crick et al., 1961). In addition, genomes also contain vast amounts of non-coding, often repetitive, and seemingly non-functional DNA sequences. These sequences, which make up between 20 and 70% of all genetic material depending on the organism, for a long time referred to as “junk DNA”, were thought to be non-functional remnants of past attempts of gene evolution, inert spacers between genes, or simply unavoidable accumulations of often repetitive elements (Ohno, 1972, Schnable et al., 2009, Wessler, 2006, Orgel and Crick, 1980).

In recent years accumulating evidence has suggested that much of this “junk DNA” is neither inert nor lacks biological function. Non-coding RNAs of different lengths with various functions have been discovered, and the regulatory potential of distal DNA elements influencing gene expression and chromatin organization are now fully recognized (ENCODE, 2012). Furthermore, several classes of repetitive elements were found to contribute strongly to the evolution of gene regulatory networks and genomic innovations (reviewed by Chuong et al., 2017). In the following I will give an overview of the nature and role of repetitive DNA elements and introduce the systems in place to regulate their activity, focusing on the fruit fly *Drosophila melanogaster*.

1.1 Repetitive elements

Repetitive elements, which are found across the eukaryotic kingdom, are distinguished based on their length and encoded features. In the following sections I will introduce two classes of selfish repetitive elements found ubiquitously in eukaryotic genomes.

1.1.1 DNA Satellites

Short and highly repetitive DNA Satellites were initially discovered due to their characteristic base composition, which caused them to separate from bulk genomic DNA fragments in CsCl gradients (Lohe and Brutlag, 1986, Thiery et al., 1976, Corneo et al., 1966, Kit, 1961). Distinct populations of highly AT-rich or GC-rich elements migrated above and below bulk DNA, respectively, inspiring their description as “Satellites”. Different classes of Satellites are distinguished based on the length of an individual repeat (reviewed by Thakur et al., 2021). Microsatellites, or simple sequence repeats, (2-10 bp) form arrays of several kilobases and are prevalent, for example, in human telomeres, making up 3% of the human genome. Minisatellites (10-100 bp) form arrays of up to 30 kb. Satellites and macrosatellites, spanning few hundred to few thousand, or up to several thousand base pairs, respectively, form large arrays in telomeric or centromeric regions and comprise the most studied classes. Despite their high prevalence in virtually all genomes, the direct functions of Satellites are still largely elusive, owing mostly to the difficulty of studying highly repetitive sequences. Satellites generally do not encode proteins, and little is known about the prevalence, function, or

regulation of transcripts originating from these sequences. Potential functions have been ascribed to DNA secondary structures formed by repetitive sequences, although conclusive insight is still largely lacking (Thakur et al., 2021, Jagannathan and Yamashita, 2017). The conserved accumulation of Satellite sequences at specialized genomic sites, however, suggests important structural functions of these elements.

1.1.1.1 Genomic distribution and function of DNA Satellites

The highly repetitive nature of large Satellite arrays long presented a major difficulty for the telomere to telomere assembly of chromosomes, leaving blank spaces in centromeric regions whose size and content could only be estimated through *in situ* hybridization studies (Lohe et al., 1993). Only recent advances in long read DNA sequencing have enabled the full coverage of centromeric repeat regions in selected genomes (Hoyt et al., 2022, Nurk et al., 2022, Altemose et al., 2022). While the sequence of the respective elements differs vastly between species, most eukaryotic centromeres are characterized by arrays of Satellite repeats, often spanning several mega bases (Henikoff et al., 2001, Melters et al., 2013, Round et al., 1997, Rudd and Willard, 2004). Centromeric Satellite repeats in several systems have been found to be required for the recruitment of CENP-A, the centromere-specific variant of histone H3, which is needed for the formation of a kinetochore structure (Palmer et al., 1987, Westermann et al., 2003). Interestingly, while centromeric Satellite repeat units in many mammalian and plant genomes are of roughly nucleosomal length, *Drosophila melanogaster* centromeric Satellites are very short with 5-10 bp per unit (Sun et al., 2003, Thakur et al., 2021). The observation that highly diverse centromeric Satellite sequences all have a common role in the deposition of CENP-A implies that Satellites share a conceptually conserved function despite stark differences in the mechanistic details. A marked exception is the budding yeast *Saccharomyces cerevisiae*, where CENP-A is recruited in a sequence-specific manner to a point centromere (Fitzgerald-Hayes et al., 1982).

Further structural functions have been proposed for the large numbers of pericentromeric microsatellites found in many organisms: By allowing the bundling of different chromosomes into so called chromocenters they are thought to ensure their packaging into a single nucleus and avoid the formation of micronuclei in both *Drosophila* and mouse (Jagannathan et al., 2018, Pardue and Gall, 1970, Jones and Robertson, 1970). This represents another example of conserved functions of Satellites and Satellite binding proteins, despite the lack of direct evolutionary conservation, as both components differ vastly between organisms (Jagannathan et al., 2018, Thakur et al., 2021).

Satellite DNA sequences with longer repeat units also often occur at conserved positions within genomes. In *Drosophila melanogaster*, four groups of such Satellites exist (Figure 1), whose functions have not been fully elucidated.

Telomeric TAS repeats

The telomere-associated sequence (TAS) repeats form the boundary between euchromatic chromosome arms and telomeric repeats of retro-transposons, which carry out telomere maintenance in *Drosophila* (Pryde et al., 1997). The individual telomeres of chromosomes 2, 3 and X are each characterized by distinct TAS elements that differ in the sequence of their basic repeat units (Pryde et al., 1997, Karpen and Spradling, 1992, Walter et al., 1995, Biessmann et al., 2005). Tandem arrays of TAS Satellites are therefore thought to aid the distinction of different chromosomes and the correct pairing of homologs in mitosis (see further below).

Ribosomal IGS

The ribosomal intergenic spacer (IGS) contains promoter sequences for ribosomal genes (Simeone et al., 1985). IGSs are found across eukaryotes in varying sizes and sequences (Long and Dawid, 1980, Fedoroff, 1979, Rae et al., 1981). In *Drosophila*, they are known to form highly polygenic arrays at the rDNA locus of the X chromosome (Wellauer and Dawid, 1977). How IGS elements contribute to the expression and regulation of rDNA is not fully understood.

Pericentromeric *1.688*

Immediately adjacent to the rDNA cluster, *Drosophila melanogaster* harbors a large (11 Mb) array of 359-bp repeats of the 1.688 g/cm^3 buoyant density family of Satellite repeats (Lohe et al., 1993). Pericentromeric regions of the three major chromosomes of *Drosophila* all harbor arrays of 1.688 family repeats, with different repeat sequences occurring on different chromosomes (Lohe and Brutlag, 1986, Losada and Villasante, 1996, Abad et al., 2000). 359-bp elements are found on the X chromosome, while 353-bp and 356-bp elements occur on chromosome 3, and a 260-bp repeat is found on chromosome 2. The most prominent block of 1.688 repeats was prominently described as the Zygotic hybrid rescue (*zhr*) allele, which rescued the female-specific species incompatibility between *Drosophila melanogaster* males and *Drosophila simulans* females (Sawamura et al., 1993, Sawamura and Yamamoto, 1997).

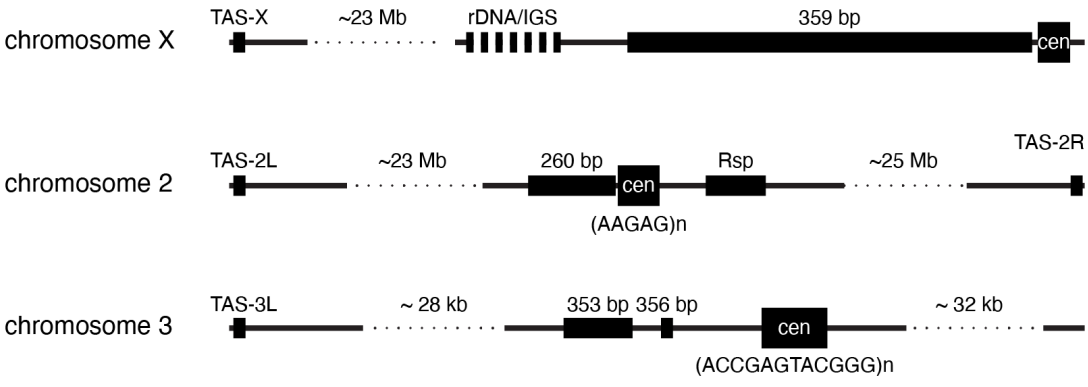


Figure 1: Satellite landscape of the *Drosophila melanogaster* chromosomes 2, 3, and X.

It was later found that the large *melanogaster*-specific accumulation of 359-bp repeats caused mitotic failure due to a lack of proper heterochromatin formation or maintenance at this locus (Ferree and Barbash, 2009). The same incompatibility was observed between *zhr* females and wildtype males, and it was proposed that maternal deposition of Satellite products (potentially small RNAs) was required for proper heterochromatic packaging of *1.688* repeats in the embryo (Usakin et al., 2007, Ferree and Barbash, 2009). Of note, fly stocks harboring the *zhr* allele are viable and fertile, indicating that the large array of 359-bp elements is not required for development (Sawamura et al., 1993). Similar to TAS elements, chromosome-specific variants of *1.688* family Satellites might nevertheless aid the correct pairing of homologous chromosomes (Ferree and Prasad, 2012).

Pericentromeric Responder

Head-to-tail pairs of the 120 bp long left and right Responder (*Rsp*) Satellites form arrays of varying sizes in the region directly proximal to the centromere of chromosome 2 (Pimpinelli and Dimitri, 1989). *Rsp* Satellites participate in a long known, but still incompletely understood case of meiotic drive in spermatogenesis. Extended *Rsp* arrays confer sensitivity to Segregation Distorter (*Sd*), a truncated version of RanGAP that prevents the histone-to-protamine exchange at chromosomes containing sensitive *Rsp* configurations by an unknown mechanism (Larracuente and Presgraves, 2012, Larracuente, 2014). Sensitivity to *Sd* varies with *Rsp* copy number, with highly sensitive *Rsp* alleles containing up to 2500 copies, and insensitive alleles harboring only 100-200 or as little as <20 copies. *Sd* and *Sd*-sensitive *Rsp* configurations are found on different alleles of chromosome 2, which leads to the preferential, non-mendelian passage of Sd-containing chromosomes to the progeny of *Sd* males (Wu et al., 1988, Lyttle, 1991). The continued presence of *Rsp* in *Drosophila* genomes, despite this fitness cost, implies functional benefits of the locus, which are, however, not known to date.

1.1.1.2 Satellite sequences are the most diverse sequences in the genome

The highly polymorphic and often drastically different Satellite sequences within and between species present major drivers of evolution and are thought to contribute to species incompatibility (Ferree and Prasad, 2012). Little is known about how repeat length, expansion, or reduction are regulated, and recent advances in long read genome sequencing hold the potential to shed more light on these questions in the future. Just like their sequences, the functions or organismal fitness impacts of Satellites are most likely diverse.

1.1.2 Transposable elements

Transposable elements (TEs) are defined as mobile DNA sequences that have the ability to propagate and increase their copy number within a host's genome (Fedoroff, 2012). Two

classes of TEs are defined by their mode of propagation via an RNA or DNA intermediate. Further subclasses, orders, and families are distinguished based on an element's structural and mechanistic differences and phylogenetic relationships. Moreover, autonomous and non-autonomous elements are distinguished based on the functions encoded by their highly compact genome: autonomous elements encode all enzymatic components required for their propagation, while non-autonomous elements rely on proteins encoded by other elements (Mérel et al., 2020, McCullers and Steiniger, 2017).

Propagation of class I retrotransposons is characterized by a "copy-and-paste" mechanism. The copy number of class I elements is increased as a direct result of the insertion of a cDNA copy derived from a TE transcript at a new location in the host genome (Boeke et al., 1985). To this end, most retroelements encode a reverse transcriptase (Flavell, 1995). Based on their transcriptional regulation and transposition mechanisms, several orders or retroelements are distinguished. Long terminal repeat (LTR)-containing elements form the largest and most diverse group of transposons in *Drosophila melanogaster*. Their replication cycle shares mechanistic principles with retroviruses, including the formation of virus-like particles. Non-LTR retroelements like long interspersed nuclear elements (LINEs) are also found abundantly in *Drosophila*, whereas the non-autonomous, RNA Polymerase III transcribed short interspersed nuclear elements (SINEs) constitute large parts of mammalian genomes (Kapitonov and Jurka, 2008).

Class II DNA transposons do not rely on an RNA intermediate, but rather propagate via a "cut-and-paste" mechanism. Here, the transposon's DNA sequence is excised and re-inserted with the help of a transposase or recombinase (Mitra et al., 2008, Yuan and Wessler, 2011). This mode of transposition leads to an increase in copy number when it takes place during S phase of the cell cycle, where replicated elements insert into non-replicated target sites. A prominent example of a class II transposon in *Drosophila* are piggyBac and the P-element, which have been extensively studied and employed as early genetic tools for mutagenesis and transgenesis in *Drosophila* as well as other organisms (Sandoval-Villegas et al., 2021, Ryder and Russell, 2003).

1.1.2.1 Genomic distribution and function of transposable elements

Transposition of both RNA and DNA elements results in double strand breaks, insertional mutagenesis, and potential mis-regulation of nearby genes, and thereby presents a direct threat for genome integrity. Moreover, current models propose that elevated copy numbers of the same sequence can increase the frequency of ectopic recombination, further endangering genome stability (Bourgeois and Boissinot, 2019). However, examples of considerable transposon content of many genomes including human (~45%), *Drosophila* (~20%), and maize (~85%) indicate that transposon sequences are not just to be seen as a burden to be tolerated (Wells and Feschotte, 2020). Indeed, both transposon regulatory sequences as well as their encoded proteins have been coopted at multiple occasions in

evolution and also the evolutionary benefits of regulatory sequences provided by mobile genetic elements that can contribute to novel gene expression patterns is now fully recognized (Feschotte, 2008, Cosby et al., 2019). Moreover, transposons have been implicated in the establishment of heterochromatin during development in *Drosophila* as well as mammals, and a structural and functional role of transposons has been described for diverse eukaryotic centromeres, as well as *Drosophila* telomeres (Sentmanat and Elgin, 2012, Moschetti et al., 2020, Pimpinelli et al., 1995). Taken together, these beneficial contributions of transposons present a plausible explanation for the vast contribution of TEs to overall genome size.

Transposon sequences are not uniformly distributed across genomes. They preferentially accumulate in pericentromeric heterochromatin, where only few coding genes are found, and meiotic recombination rates are low (Bourque et al., 2018). Different proposed reasons for the accumulation of transposons in these regions observed at steady state are hard to distinguish. One explanation could lie in insertion preferences of certain TEs for heterochromatic regions, where the particular chromatin environment and the late replication timing might be favorable for copy number maintenance or increase. Alternatively, this biased distribution might passively result from a combination of the low recombination potential and the absence of adverse effects of TE insertions in these gene-poor regions. Both factors lower the probability of elimination from a population. Upon insertion in euchromatic chromosome arms on the other hand, insertional damage might be higher and increased recombination frequency allows fast removal of insertions form the genome.

1.1.3 Transposons and Satellites together

Several instances of an interaction between Satellite repeats and TEs have been reported, and it was even proposed that Satellite repeats can be derived from transposon sequences (Caizzi et al., 1993, Heikkinen et al., 1995, Kapitonov et al., 1998, Kidwell, 2002). Specialized chromosomal compartments such as centromeric and telomeric regions in *Drosophila* are both characterized by a combination of Satellite and transposon accumulations. Islands of *G2/Jockey-3* retrotransposon arrays surrounded by large blocks of microsatellites have recently been found at the center of all *Drosophila melanogaster* centromeres (Chang et al., 2019). The sequence identity of the surrounding Satellites, as well as the length of the *G2/Jockey-3* islands differ between the five *melanogaster* centromeres, but the shared architectural features strongly suggest a functional role of these elements. Interestingly, islands of *G2/Jockey-3* retroelements are also highly enriched in CENP-A ChIPseq from *Drosophila simulans*, where the centromeric Satellite complement is very different from *melanogaster* and harbors longer repeat units. This indicates that a combination of Satellites with *G2/Jockey-3* elements is a conserved feature of *Drosophila* centromeres.

Drosophila telomeres represent another example of co-occurrence of Satellites with TEs. While telomeres in most eukaryotes are composed of short repeat sequences maintained by the specialized telomerase enzyme, *Drosophila* does not encode telomerase and instead relies on a group of specialized non-LTR retroelements for telomere function (Casacuberta, 2017). Head-to-tail arrays of *HetA*, *TART*, and *TAHRE* (referred to as HTT) at telomeres are maintained by continuous insertion of these elements at chromosome ends. HTT repeats are flanked by TAS Satellites, which form a barrier to adjacent euchromatin (Mason and Biessmann, 1995, Biessmann et al., 2005). TAS sequences are found also at telomeres in various other species and differ greatly in sequence from one organism to another and even between telomeres in the same species (Thakur et al., 2021). Similar to pericentromeric regions, telomeres represent a repressive environment (Cacchione et al., 2020). Interestingly, silencing at telomeres seems to be maintained via different mechanisms than at pericentromeric heterochromatin. It was shown to be centered around the TAS repeats, rather than HTT, and to involve the Polycomb system (Doheny et al., 2008). The chromatin state of telomeres likely differs between somatic and germline cells, due to the targeting of HTT by the piRNA pathway in ovaries (see below).

Finally, *R1* and *R2* retrotransposons are known to specifically occur at the highly repetitive rDNA locus, interrupting more than 70% of rDNA units in *Drosophila melanogaster* (Eickbush and Eickbush, 2007, Jakubczak et al., 1992). The expression of these elements is expected to disrupt the functionality of 28S rRNA units, but is kept silent by unknown mechanisms (Pérez-González and Eickbush, 2002, Ye and Eickbush, 2006). A recent study revealed that R2 retroelements are required to maintain rDNA copy number, indicating a beneficial role of this transposon for the host genome (Nelson et al., 2021).

Overall, accumulating evidence strongly indicates that a description of non-genic DNA elements as simple “junk” does not do justice to its various highly complex regulatory, architectural, but also selfish functions. This also implies, that effective genome maintenance mechanisms must be in place to control the activity of selfish DNA elements.

1.2 Constitutive heterochromatin

Despite beneficial roles in gene regulation and genome structure and organization, deleterious effects of repetitive elements prevail and have to be contained. Repetitive elements cause danger to the genome not just by transposition, but also through ectopic recombination (Bourque et al., 2018). Large stretches or high copy number of repetitive sequences endanger genome stability via illicit recombination events that can lead to the formation of large rearrangements or even loss of parts of chromosomes. Therefore, repressive chromatin structures are thought to be required to compact repetitive regions (Kabi and Filion, 2021).

1.2.1 Characteristics and function

Most repeats across eukaryotic genomes lie within densely packed, largely transcriptionally repressed constitutive heterochromatin, which is characterized by the di- and tri-methylation of Lysine 9 on histone H3 (Jenuwein and Allis, 2001). This mark serves as an anchor for the main heterochromatin protein HP1 (first identified as Su(var)2-5 in *Drosophila melanogaster* (Eissenberg et al., 1992, James et al., 1989, James and Elgin, 1986), hereafter referred to as HP1a), which serves as hub for chromatin compaction and silencing (Lachner et al., 2001). Moreover, constitutive heterochromatin is characterized by a lack of histone tail acetylation, as well as low general accessibility as determined by nuclease cleavage assays (Jenuwein and Allis, 2001). An additional characteristic of heterochromatin in several mammalian and plant systems is DNA methylation, which enhances silencing, but is absent in *Drosophila* (reviewed by Iyer et al., 2011). The majority of constitutive heterochromatin is found in pericentromeric regions which harbor most of the genomic transposon content and only few genes. Smaller heterochromatin islands are also found along chromosome arms where they often surround dispersed transposon insertions, as well as at telomeric repeats (Janssen et al., 2018).

The function of constitutive heterochromatin is usually restricted to the repression of repetitive sequences, rather than dynamic regulation of gene expression (Trojer and Reinberg, 2007). However, some reports also demonstrate important roles for H3K9-methyltransferases for mouse development and it is unclear if these are only related to TE silencing or also to the regulation of gene expression (Padeken et al., 2022). The repressive function of constitutive heterochromatin was long thought to stem from its compact nature, aided also by the global lack of acetylation and close nucleosomal spacing. Accumulating evidence now suggests that heterochromatin is more dynamic than previously appreciated, and that phase separation properties of HP1 proteins crucially contribute to the exclusion of transcription and recombination factors that might activate or destabilize heterochromatic sequences (Larson et al., 2017, Strom et al., 2017). With these combined mechanisms, the repressive nature of heterochromatin reduces the recombination capacity of underlying sequences, which in turn is thought to have facilitated the accumulation of TE sequences in pericentromeric regions, further enhancing the need for efficient compaction. Of note, few genes are found in these regions, and their expression has been found to rely on the heterochromatic nature of their surroundings, although the mechanistic basis for this dependency is not clear (Schoelz and Riddle, 2022).

1.2.2 HP1 proteins

Originally identified in *Drosophila* as heterochromatin binding protein and a suppressor of position effect variegation, Su(var)2-5 is the founding member of the large heterochromatin protein 1 (HP1) family, which is found ubiquitously from yeast to humans (Eissenberg and Elgin, 2000). HP1 is widely recognized as the main non-histone heterochromatin protein as it directly reads out the characteristic histone H3 lysine 9 trimethylation mark (H3K9me3) and

serves as a hub for the recruitment of other repressive factors. HP1 proteins hold essential functions around chromatin architecture that go beyond simple repression of repetitive sequences, influencing chromosome condensation and segregation, telomere maintenance, as well as regulation of transcription and DNA damage repair (Eissenberg and Elgin, 2000, Eissenberg and Elgin, 2014).

1.2.2.1 HP1 variants

HP1 proteins often exist as small gene families, and most organisms that harbor HP1 proteins encode multiple variants with only partially overlapping functions (Vermaak and Malik, 2009). In human, mouse, and *Drosophila*, three HP1 family proteins are expressed ubiquitously in all cells. *Drosophila* HP1a and human HP1alpha hold canonical HP1 functions: they broadly localize to constitutive heterochromatin and mediate the repression and compaction of these regions (Saunders et al., 1993). HP1b and HP1c in *Drosophila*, as well as HP1beta and HP1gamma in humans partially or predominantly localize to euchromatic compartments and are functionally distinct from HP1a/alpha (Vakoc et al., 2005, Canzio et al., 2014, Smothers and Henikoff, 2001). Their functions are not fully characterized and seem to be at least partially overlapping and non-essential, with involvements in both activation and repression of transcription. Despite strong functional overlaps in line with similar nomenclature, the evolutionary relationship between mammalian HP1alpha/beta/gamma with *Drosophila* HP1a/b/c in not straight forward (Figure 2). All mammalian HP1 proteins share most similarity with *Drosophila* HP1b, suggesting that HP1b is closest to the ancestral version. These observations indicate that the HP1 family is highly dynamic across evolution. In further support of this, several additional, often germline specific HP1 variants exist in many organisms. A recent study found that some Dipteran species harbor more than 20 HP1

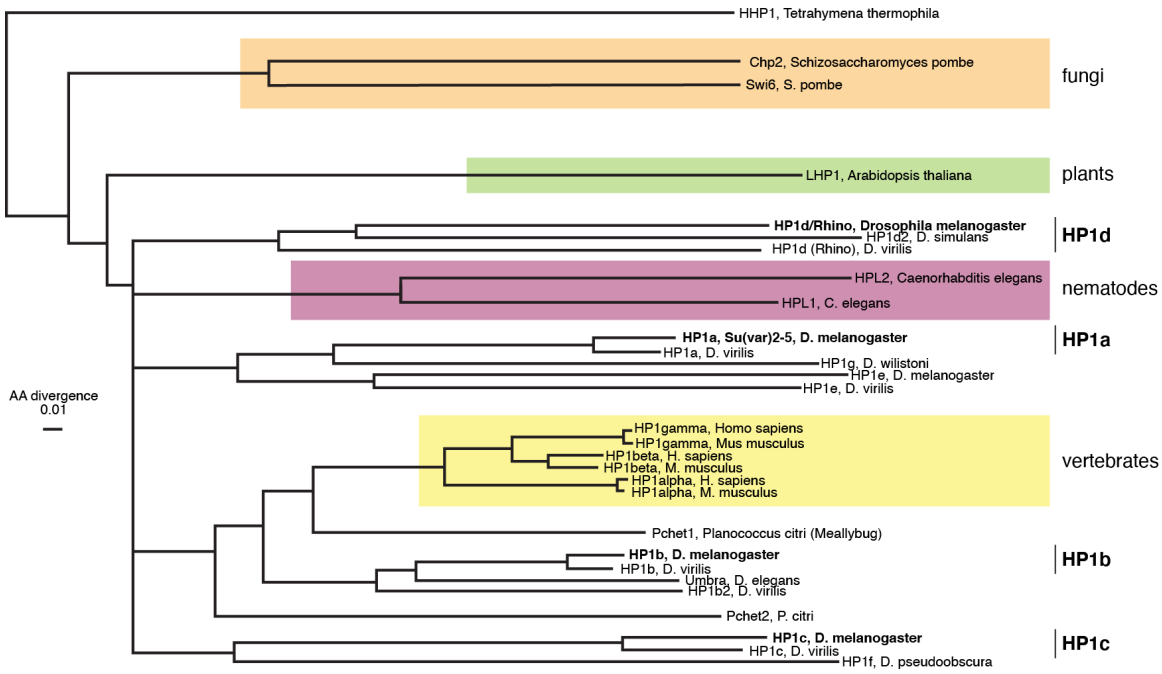


Figure 2: Phylogenetic tree of HP1 protein variants across eukaryotes. (Adapted from Vermaak et al., 2009)

proteins with recurrent duplications and gene deaths leading to a large diversity in HP1 proteins (Helleu and Levine, 2018). *Drosophila melanogaster* harbors two additional full length HP1 proteins, HP1d/Rhino and HP1e, which are expressed in ovaries (Rhino) and testes (both). Rhino represents a remarkable HP1 protein whose function strongly adapted to its role in transcriptional licensing of piRNA clusters, which will be further outlined below. The function of the least studied member of the *Drosophila* HP1 family, HP1e, is largely unknown, with only few studies describing its heterochromatic localization and a requirement for paternal DNA packaging in the male germline (Levine et al., 2015).

Besides full length HP1 proteins harboring all canonical protein domains (see below), several partially truncated HP1 variants also exist (Levine et al., 2012). These proteins likely originated through partial duplications of existing HP1 proteins, but their putative functions remain poorly characterized.

1.2.2.2 Domain architecture

HP1 proteins are characterized by a shared domain architecture, featuring an N-terminal chromo domain connected via a flexible hinge region to a C-terminal chromoshadow domain, supplemented additionally with variable N- and C-terminal extensions (Aasland and Stewart, 1995, Paro and Hogness, 1991, Singh et al., 1991).

Chromo domain

The chromo domain binds to H3K9 methylated histone tails, with its binding affinity increasing with higher methylation state (Lachner et al., 2001, Nielsen et al., 2002, Jacobs and Khorasanizadeh, 2002). The chromo domain fold is also found in several other chromatin reader proteins (e.g. Polycomb, Su(var)3-9) and mediates binding to methylated lysines at different positions within the histone tail with remarkable specificity (Aasland and Stewart, 1995). Despite a highly similar amino acid context surrounding the methylated lysine residue (ARKS), the HP1 and Su(var)3-9 chromo domains strictly bind to methylated H3K9, while Polycomb binding is specific to H3K27me2/3 (Fischle et al., 2003). This specificity is conferred by differences in protein-protein interactions of the chromo domain with surrounding residues of the histone tail peptide. While chromo domains generally exist as monomers, the *S. pombe* Swi6 and *D. melanogaster* Rhino chromo domains were found to exist as dimers (Canzio et al., 2013, Yu et al., 2015a, Le Thomas et al., 2014). In Swi6, chromo domain dimers are formed through binding of one monomer to a histone tail mimicking ARK motif on the partner chromo domain, potentially allowing the formation of higher order multimers or regulating chromatin binding (Canzio et al., 2013). However, no such motif is found in other HP1 proteins. Dimerization of the Rhino chromo domain was observed in X-ray crystallography studies and, according to the author's comment, is not sterically compatible with the binding to two neighboring histone tails (Yu et al., 2015a). The physiological relevance of Rhino chromo domain dimerization is therefore unclear.

Chromoshadow domain

The chromoshadow domain serves as a dimerization surface for all HP1 proteins and forms the contact point to several downstream factors recruited by HP1 (Brasher et al., 2000, Smothers and Henikoff, 2000). Dimerization and partner protein recruitment are tightly coupled, as the dimerization interface forms the binding cleft for a conserved PxVxL amino acid motif in downstream factors that are recruited via the chromoshadow domain (Thiru et al., 2004). Recent reports further show that dimerization of HP1 proteins (as well as other chromatin readers) is essential for their chromatin binding ability, presumably through increased binding affinity via cooperative binding (Villaseñor et al., 2020). The residues involved in the chromoshadow domain dimerization are conserved in most HP1 proteins, supporting the formation of both homo- and heterodimers, which has been confirmed *in vitro* (Brasher et al., 2000, Nielsen et al., 2002, Ye et al., 1997, Lee et al., 2019). The *in vivo* role of HP1 heterodimerization is not clear.

Flexible hinge and terminal extensions

An unstructured and highly variable hinge region connects the two folded domains and contains the nuclear localization signal (Singh and Georgatos, 2002, Wang et al., 2000). The hinge region has been implicated in interactions with selected partner proteins, as well as nucleic acids, and can be post translationally modified (Lomberk et al., 2006, Eissenberg and Elgin, 2000). Its interaction with DNA is thought to strengthen the recruitment to chromatin in light of the relatively low affinity of the chromo domain for H3K9me3. Most prominently, the hinge region has been recently found to be important for liquid-liquid phase separation properties of human HP1alpha and *Drosophila* HP1a (Larson et al., 2017, Strom et al., 2017). HP1alpha phase separation is further regulated by phosphorylated residues in the N-terminal extension of the chromo domain. C-terminal extensions to the chromoshadow domain of variable length are thought to regulate the association with downstream factors, for example for HP1b/c, further highlighting the functional relevance of unstructured protein domains (Canzio et al., 2014).

1.2.2.3 Partner proteins

A plethora of factors has been described as interactors of HP1 proteins. Most interactions are mediated via the chromoshadow domain, with only few exceptions that were mapped to the hinge region, while interactions with the chromo domain are reserved for histone tails (Lomberk et al., 2006, Eissenberg and Elgin, 2000). Most partner proteins contain the consensus PxVxL motif or a derivative thereof, which interacts with the chromoshadow domain dimer of HP1 (Brasher et al., 2000, Smothers and Henikoff, 2000). The most prominent interactors of canonical HP1a/alpha are deacetylases and histone methyl transferases that further amplify the repressive nature of constitutive heterochromatin and

allow the spreading of repressive chromatin through continued methylation of H3K9, providing additional binding sites for HP1.

Besides the binding to the H3K9methylated histone tail, additional interactions with core histones have been reported. HP1alpha – but not beta and gamma – was found to interact preferentially with histone variant H2A.Z, which is found in pericentromeric heterochromatin, and *S. pombe* Swi6 was shown to contact histone H2B via its chromoshadow domain dimer (Fan et al., 2004, Ryan and Tremethick, 2018, Sanulli et al., 2019). Furthermore, interactions between HP1alpha and linker histone H1 have also been reported and are thought to be mediated primarily through the HP1 hinge region (Lu et al., 2009, Hale et al., 2006). H1 has been proposed to hold a repressive function in heterochromatin biology by promoting chromatin compaction through binding and shielding the negatively charged linker DNA. The detailed mechanistic interplay of histone variants, their modifications, and other players in heterochromatin formation and maintenance awaits further investigation. While the self-reinforcing nature of constitutive heterochromatin is well described, how pericentromeric heterochromatin is initially formed is not fully understood. An important role of transposon and repeat sequences in the process cannot be denied. In the following section I will highlight two orthogonal strategies that are known to result in heterochromatin formation on transposon sequences.

1.3 Heterochromatin formation and transposon silencing

With their high copy number, transposons can serve as nucleation sites for heterochromatin formation across pericentromeric regions. Their high sequence diversity, however, poses a significant challenge to the system. Mechanisms for heterochromatin specification have to be able to recognize a diverse and ever-changing set of sequences, while avoiding the targeting of genic sequences. Different transposon silencing mechanisms have been described that work in partially complementary manners.

1.3.1 Transposon silencing via KRAB-Zinc Finger proteins

The diverse family of Krüppel associated box (KRAB) zinc finger proteins serve as an important line of defense against TEs in tetrapods (reviewed by Bruno et al., 2019). Through sequence-specific DNA binding mediated by highly variable C2H2 zinc fingers (ZnFs), these proteins have adapted to target TE sequences for heterochromatic repression. While ZnF folds have been described to mediate the interaction with both proteins and nucleic acids, regularly spaced assays of ZnFs are known to mediate the recognition of DNA sequences with high accuracy (Wolfe et al., 2000, Brayer and Segal, 2008). The KRAB domain recruits the co-repressor KRAB-associated protein 1 (KAP1), which in turn links to several factors inducing the formation of repressive heterochromatin, including the histone methyltransferase SET Domain Bifurcated Histone Lysine Methyltransferase 1 (SETDB1), the nucleosome remodeling, and deacetylation

(NuRD) complex as well as HP1a and the DNA-methylation machinery (Iyengar and Farnham, 2011, Friedman et al., 1996). While the highly flexible sequence-specificity of ZnFs offers a versatile defense system against diverse TE sequences, adaptation of the system is relatively slow, as the evolution of new KRAB-ZnF genes takes time. It is therefore thought, that the KRAB-ZnF system mostly maintains the silent state of long domesticated TE (Castro-Diaz et al., 2014).

1.3.1.1 Other types of zinc fingers with N-terminal domains

KRAB-ZnF proteins are restricted to the tetrapod lineage, and no corresponding system has been described in *Drosophila*. Of note, other protein families have been described that share a common overall domain architecture with KRAB-ZnF proteins. A shared N-terminal domain connected to an array of multiple regularly spaced C_2H_2 ZnF folds is characteristic to the families of ZAD, and SCAN domain proteins (Fedotova et al., 2017). The ZAD-ZnF family stands out as the most diverse group of ZnF-containing transcription factors in *Drosophila*, harboring over 90 members, and has been proposed to act as the insect equivalent of the KRAB-ZnF family (Chung et al., 2007, Chung et al., 2002). However, a link to transposon biology is so far missing, with only one ZAD-ZnF protein having a reported role in transposon regulation (Czech et al., 2013). Importantly, the ZAD and SCAN domains are not known to recruit any downstream factors but are characterized as homodimerization domains. Whether the DNA binding capacity of ZnF proteins has been employed for transposon silencing and heterochromatin specification outside of tetrapods remains to be determined.

1.3.2 Small RNA-mediated heterochromatin formation

Several systems form yeast to plants and animals have been described to rely on small non-coding RNAs (sRNA) for the formation of heterochromatin on repetitive sequences (Gutbrod and Martienssen, 2020). sRNAs of 21 to 32 nucleotides in length are loaded onto Argonaute proteins and serve as sequence-specific guides for heterochromatin formation through recruited downstream factors (Grewal and Elgin, 2007, Martienssen and Moazed, 2015). Arguably the best studied system is the fission yeast *S. pombe*, where seminal insights into sRNA-dependent heterochromatin formation has been gained by several groups over the last decades. In *S. pombe*, short interfering RNAs (siRNAs) derived from centromeric repeats are loaded onto the single Argonaute protein Ago1, which associates with the Chromo domain-containing protein 1 (Chp1) and the adaptor protein called Targeting complex subunit 3 (Tas3) to form the RNA-induced transcriptional silencing (RITS) complex (Volpe et al., 2002, Verdel et al., 2004, Verdel and Moazed, 2005). RITS thereby associates with nascent transcripts at centromeric repeats, where it recruits the histone methyltransferase Clr4 (Su(var)3-9 in *Drosophila melanogaster*) to the underlying chromatin. This results in the deposition of the heterochromatic H3K9me2/3 histone mark, and the subsequent recruitment of the *S. pombe* HP1 protein Swi6. The system is further enhanced by the recruitment of the RNA-dependent

RNA polymerase complex (RDRC) through RITS (Motamedi et al., 2004). RDRC allows the reverse transcription of the targeted nascent RNA into double-stranded (ds) RNA, which then serves as substrate for the production of more targeting siRNAs.

A conceptually similar system, also including intricate feedback loops, has been extensively described in *A. thaliana*, where repressive histone marks as well as DNA methylation are deposited at repetitive regions via several Argonaute family proteins (Pikaard and Mittelsten Scheid, 2014, Gutbrod and Martienssen, 2020).

In several animal models the small RNA mediated transcriptional silencing of transposons is mediated by Piwi-induced wimpy testes (PIWI)-interacting RNAs (piRNAs), which guide heterochromatin formation at TE loci in germline cells (Ozata et al., 2019, Sienski et al., 2012, Le Thomas et al., 2013, Aravin and Bourc'his, 2008, Wang and Elgin, 2011). In the piRNA pathway, 23-32 nt piRNAs complementary to TE sequences are loaded into nuclear PIWI clade Argonaute proteins, which mediate the silencing of transposons through co-transcriptional formation of heterochromatin at target loci (Ozata et al., 2019). Practical experimental advantages have enabled the deepest understanding of the piRNA pathway in *Drosophila melanogaster*, which is extensively described below.

Transposon silencing via small RNAs offers one clear advantage: The recognition of target sequences via base complementarity allows for high adaptivity simply by incorporation of new sequences into Argonaute proteins. How this is achieved with particular efficiency in the piRNA pathway will be further outlined in section 1.4.1.1.

1.4 The piRNA pathway in *Drosophila*

In *Drosophila melanogaster*, the piRNA pathway is best studied in the adult ovary, though it is also active in testes and presumably in developing gonads (Quénerch'du et al., 2016, Marie et al., 2017). Three PIWI proteins are expressed in *Drosophila* ovaries: the nuclear, catalytically inactive Piwi, as well as the cytoplasmic, cleavage-competent Aubergine (Aub) and Ago3. Aub and Ago3 mediate post transcriptional TE silencing via slicing of TE transcripts upon recognition via complementarity with their loaded piRNA. This slicing of TE transcripts at the same time contributes majorly to piRNA production, as a slicing event catalyzed by either protein serves as a trigger for processing of the 5’ cleavage product into piRNAs: The free 3’ end of the cleaved target RNA is incorporated into the next PIWI protein, and a mature piRNA is generated through RNA cleavage, trimming, and O-methylation (Aravin et al., 2008, Aravin et al., 2007b, Brennecke et al., 2007, Gunawardane et al., 2007, Zhang et al., 2011, Saito et al., 2007, Hayashi et al., 2016). Loading of the next piRNA into Aub/Ago3 feeds into the so-called ping-pong amplification cycle, where the next cleavage- and trigger-competent piRNA-PIWI complex is formed and can again induce the processing of complementary transcripts.

piRNAs loaded into Piwi, on the other hand, mediate the targeting of complementary, nascent transcripts in the nucleus, inducing the formation of heterochromatin at the underlying locus via co-transcriptional gene silencing (Wang and Elgin, 2011, Sienski et al., 2012, Le Thomas et al., 2013, Rozhkov et al., 2013, Sienski et al., 2015, Batki et al., 2019, Yu et al., 2015b). Mechanistic similarities between piRNA-mediated heterochromatin formation in *Drosophila* and analogous siRNA-mediated processes in *S. pombe*, and plants extend from the targeting mechanism to the connection to core heterochromatin machinery, as well as recently delineated involvement of multivalent interactions (Schnabl et al., 2021).

Besides the germline cells of the *Drosophila* ovary where the piRNA pathway is active, somatic support cells surrounding the developing oocyte and the accompanying nurse cells also harbor a simplified version of the piRNA pathway (Sato and Siomi, 2020). In these cells, Aub and Ago3, along with many other piRNA pathway proteins are not expressed and ping-pong amplification of piRNAs does not take place. Instead, transposon silencing in these cells depends exclusively on the co-transcriptional silencing via piRNA-loaded Piwi. Much of the knowledge about Piwi-dependent co-transcriptional silencing was obtained through a cell culture system of ovarian somatic cells (OSCs) recapitulating the piRNA pathway of somatic cells of the ovary (Niki et al., 2006, Saito et al., 2009). Due to the absence of ping-pong, as well as other processes in the soma, the origin and processing of piRNA precursor RNAs differs significantly between the germline and somatic compartment of the ovary, as summarized below. While the study of the somatic piRNA pathway was strongly facilitated by the availability of OSCs, no cell culture system exists that replicates the germline pathway in flies.

1.4.1 Single stranded piRNA precursors as a challenge

Marked differences between the piRNA pathway and other RNA interference pathways involve the origin and biogenesis of the targeting small RNA. siRNAs, as well as gene regulatory miRNAs, are produced from double stranded RNA precursors through processing by the endonuclease Dicer (Filipowicz et al., 2005). To achieve dsRNA precursors, both the plant and yeast sRNA production for transcriptional silencing is coupled to RNA-dependent RNA polymerases, converting nascent single stranded transcripts into dsRNA to feed them into siRNA biogenesis. The processing of double stranded RNA as a source for guidance factors that determine targets of repression has the inherent benefit that it facilitates the distinction of substrates from cellular, single stranded mRNA. Furthermore, an important target of post-transcriptional silencing via siRNAs are viral transcripts, and the double stranded nature of certain RNA virus genomes presents an advantage in the fast recognition of these foreign nucleic acids. The piRNA pathway on the other hand does not involve dsRNA precursors (Vagin et al., 2006). Instead, long, single stranded piRNA precursor transcripts are processed into mature piRNAs by in a complex pathway involving a host of piRNA pathway specific factors concentrated in perinuclear germ granules of the germline known as “nuage” or at the outer mitochondrial membrane (Pippadpally and Venkatesh, 2020). The distinction of piRNA precursors from other RNA species therefore presents a certain challenge to the piRNA

pathway, as they do not structurally differ from mRNAs in a way that siRNA or miRNA precursors do. Insights from *Drosophila* indicate that the different cell types of the ovary employ two independent ways for piRNA precursor specification which I outline below.

1.4.1.1 The advantage of piRNA clusters

The discovery of piRNAs as a distinct class of Argonaute-interacting small RNAs went hand in hand with the discovery that these small RNAs are produced in large amounts from specialized genomic source loci in both *Drosophila* and mouse (Brennecke et al., 2007, Houwing et al., 2007, Aravin et al., 2008). These loci harbor large numbers of nested transposon insertions and were coined as "piRNA cluster". Different classes of piRNA clusters were soon described in *Drosophila*, where one gave rise to abundant piRNAs on one genomic strand (termed uni-strand cluster), whereas the other produced piRNAs from both strands (dual-strand cluster) (Brennecke et al., 2007). Both uni- and dual-strand clusters are transcribed by RNA Pol II to give rise to long piRNA precursor molecules, but the initiation of transcription as well as the downstream fate of these cluster transcripts differ.

Uni-strand clusters and piRNA biogenesis in the soma

Uni-strand piRNA clusters are transcribed from canonical promoters at a single transcription start site and are the exclusive source of piRNAs in the somatic cells of the ovary. The most prominent example is *flamenco*, a remarkable locus spanning more than 300 kb made up almost entirely of *gypsy* element insertions oriented in antisense relative to its transcription which is expressed specifically in the soma (Yamanaka et al., 2014, Pelisson et al., 1994). The long, non-coding, single-stranded *flamenco* transcripts are capped and spliced proximal to the promoter. Upon export to the cytoplasm via canonical mRNA export, somatic piRNA precursors are selected for piRNA biogenesis based on their Uridine-rich nucleotide composition which is read out by the soma-specific protein Fs(1)yb (Szakmary et al., 2009, Saito et al., 2010, Qi et al., 2010, Handler et al., in preparation). The somatic piRNA precursor selection does not seem to distinguish between piRNA cluster transcripts and mRNAs per se, as Uridine-rich mRNA 3' UTR sequences are equally efficiently bound by Fs(1)yb and fed into piRNA biogenesis. Being restricted to the sense strand of mRNAs, the resulting piRNAs however, lack a target and are to be regarded as non-functional by products of somatic piRNA biogenesis (Handler et al., in preparation).

Dual-strand clusters and piRNA biogenesis in the germline

In the germline, the predominant source of piRNAs are dual-strand piRNA clusters. These clusters are often located at the border between pericentromeric heterochromatin and euchromatic chromosome arms and are themselves prominently marked by heterochromatic H3K9me3 (Brennecke et al., 2007, Mohn et al., 2014, Klattenhoff et al., 2009, Zhang et al.,

2014). Despite this repressive chromatin environment, dual-strand piRNA clusters are bidirectionally transcribed, which is achieved through the direct recruitment of core transcription machinery to piRNA clusters independent of promoters. To this end, the germline-specific TFIIA paralog Moonshiner is recruited to piRNA clusters to initiate transcription at random sites on both genomic strands within the cluster (Andersen et al., 2017). Along with Moonshiner, the piRNA pathway specific RNA export factor Nxf3 and its partner proteins Bootlegger and UAP56 are enriched at piRNA clusters (ElMaghraby et al., 2019, Kneuss et al., 2019, Zhang et al., 2012b, Zhang et al., 2018). This allows the direct handover of newly transcribed piRNA precursor transcripts to the export machinery, forming a direct and dedicated connection to the cytoplasmic site of precursor processing into mature piRNAs. This likely ensures the enrichment of piRNA precursors at the site of processing. Furthermore, biogenesis of mature piRNAs from piRNA precursors strictly depends on PIWI-mediated slicing events of the ping-pong cycle: cleavage of piRNA precursor RNAs by piRNA-loaded Aub or Ago3 provide an entry point for the subsequent loading of all three Piwi proteins (Han et al., 2015, Mohn et al., 2015). Once initiated at a so-called trigger site (piRNA complementary site), the downstream RNA is processed into Piwi-loaded piRNAs by a process referred to as “phasing”, which is shared with the somatic piRNA pathway. In a seemingly eternal chicken-and-egg scenario, only transcripts that already contain sequence complementarity to pre-existing piRNAs can give rise to new piRNAs. Dual-strand piRNA cluster transcripts are thus the ideal substrates for piRNA production in the germline, as they are tunneled directly to nuage via Nxf3 and contain a multitude of trigger sites where piRNA biogenesis can be initiated.

Similar to the somatic pathway, the piRNA biogenesis machinery in the germline does not strictly distinguish between piRNA precursors and mRNAs, as the introduction of trigger sites into a reporter mRNA induces its processing into mature piRNAs (Mohn et al., 2015). This on one hand presents a danger to the cell, as any RNA can be consumed by the piRNA pathway if complementary piRNAs exist. On the other hand, this ability to process any targeted sequence into piRNAs is thought to increase the efficiency of transposon silencing.

One interesting exception to the dual-strand cluster mode is piRNA cluster 20A, a uni-strand cluster that is transcribed in both the soma and the germline of the ovary. How piRNA biogenesis from 20A is initiated in the germline, and how this cluster avoids the conversion into a dual-strand cluster is not fully understood.

Role of piRNA clusters in new invasions

piRNA clusters are thought to serve as heritable and adaptive memory loci, with transposon fragments in the clusters providing crucial silencing information about previous transposon invasions (Aravin et al., 2007a). Any sequence complementary to transposon fragments contained in piRNA clusters will be silenced by the pathway in trans. Besides maintaining transposon repression at steady state, this is thought to also serve as an important protection

against repeated invasions. Adapted versions of known elements entering via horizontal transfer from other species could thereby still by targeted as long as sufficient complementarity still exists. Current models further propose that piRNA clusters serve as traps for newly invading transposons: the insertion of new sequences into a piRNA cluster leads to the fast incorporation of the new sequence into the piRNA pool (Kofler, 2019, Kofler, 2020, Gebert et al., 2021). This is greatly aided by the phased production of piRNAs from entire piRNA precursor molecules upon recognition at an initial site of high U-content or piRNA complementarity in the soma or germline, respectively, which circumvents the need for the biogenesis machinery to recognize a new sequence. Thereby, unlike the KRAB-ZnF system, the piRNA pathway is thought to be highly adaptable to new invading sequences.

piRNA clusters outside of *Drosophila*

Following their initial discovery in flies and mouse, piRNA clusters have been described in organisms as diverse as silk worm, zebra fish, and oyster (Kawaoka et al., 2008, Houwing et al., 2007, Jehn et al., 2018). Although the core concept of piRNA clusters as transposon sequence memory loci seem to be largely conserved, their regulation differs strongly between organisms. Much of the machinery found at dual-strand piRNA clusters in *Drosophila* is not conserved in other systems. The translation of novel findings to other systems is therefore not always straight forward. However, several factors active at *Drosophila* dual-strand piRNA cluster are homologs of highly conserved protein found from yeast to humans (Klattenhoff et al., 2009, Mohn et al., 2014, Zhang et al., 2014, Andersen et al., 2017, ElMaghraby et al., 2019, Kneuss et al., 2019). Although none of these proteins or their relatives have been connected to the piRNA pathway in other organisms up to date, studying the role of piRNA-specific paralogs of core transcription and chromatin factors presents a unique opportunity to learn more about newly evolved, but also highly conserved functions of the respective protein families.

1.4.2 Rhino as an anchor for dual-strand piRNA cluster factors

The recruitment of specialized machinery for the non-canonical transcription and targeted export of piRNA precursors hinges on the genome-wide recognition of dual-strand piRNA clusters by the HP1 protein variant Rhino. Rhino was initially characterized as a female germline-specific HP1 variant protein important for chromosome structural organization and establishment of egg polarity during oogenesis (Volpe et al., 2001). DAPI staining of nurse cell nuclei indicated impaired chromatin organization beyond the so called “five blob stage” and patterning defects induced by mislocalization of *gurken* and *oskar* mRNAs lead to the fusion of the dorsal egg appendages, resulting in the name-giving *rhino* phenotype. Similar dorsal appendage phenotypes are observed for other piRNA pathway factors such as *zucchini* and *aubergine*, and were later found to be caused by the activation of the ATR/Chk2 DNA damage response (Klattenhoff et al., 2007). The fast-evolving properties of Rhino later implied its

involvement in a genomic conflict, even before the initial discovery of the piRNA pathway (Vermaak et al., 2005).

Since then Rhino was characterized as a central recruitment platform at piRNA clusters (Klattenhoff et al., 2009). It associates constitutively with the adaptor protein Deadlock and the Rai1 homolog Cutoff in the Rhino-Deadlock-Cutoff (RDC) complex (Mohn et al., 2014, Zhang et al., 2014). Genetic removal of either of the three proteins destabilizes the other two complex members at the protein level (Mohn et al., 2014). Cutoff likely is a catalytically dead version of the cap cleaving factor Rai1 and is thought to bind the 5' end of piRNA precursors instead of the canonical cap binding complex. Besides protecting their 5' end, Cutoff was shown to inhibit the termination and splicing of piRNA cluster transcripts (Chen et al., 2016, Pane et al., 2011). Deadlock is a *Drosophila*-specific adaptor protein and the only described direct interactor of Rhino, binding to the Rhino chromoshadow domain via a small folded domain at its N-terminus (Mohn et al., 2014, Zhang et al., 2014). A long unstructured linker connects this N-terminal domain of Deadlock to a C-terminal domain which acts as a hub for downstream factor recruitment. This domain harbors the interaction site for Cutoff, as well as Bootlegger and Moonshiner, coordinating all factors known to be involved in transcription, protection, and export of piRNA precursors (Andersen et al., 2017, ElMaghraby et al., 2019, Kneuss et al., 2019) (unpublished data from the Brennecke lab). The RDC complex and its downstream factors present fascinating examples of specialized protein variants connecting to highly conserved gene expression machinery and thereby hijacking these processes for piRNA biogenesis.

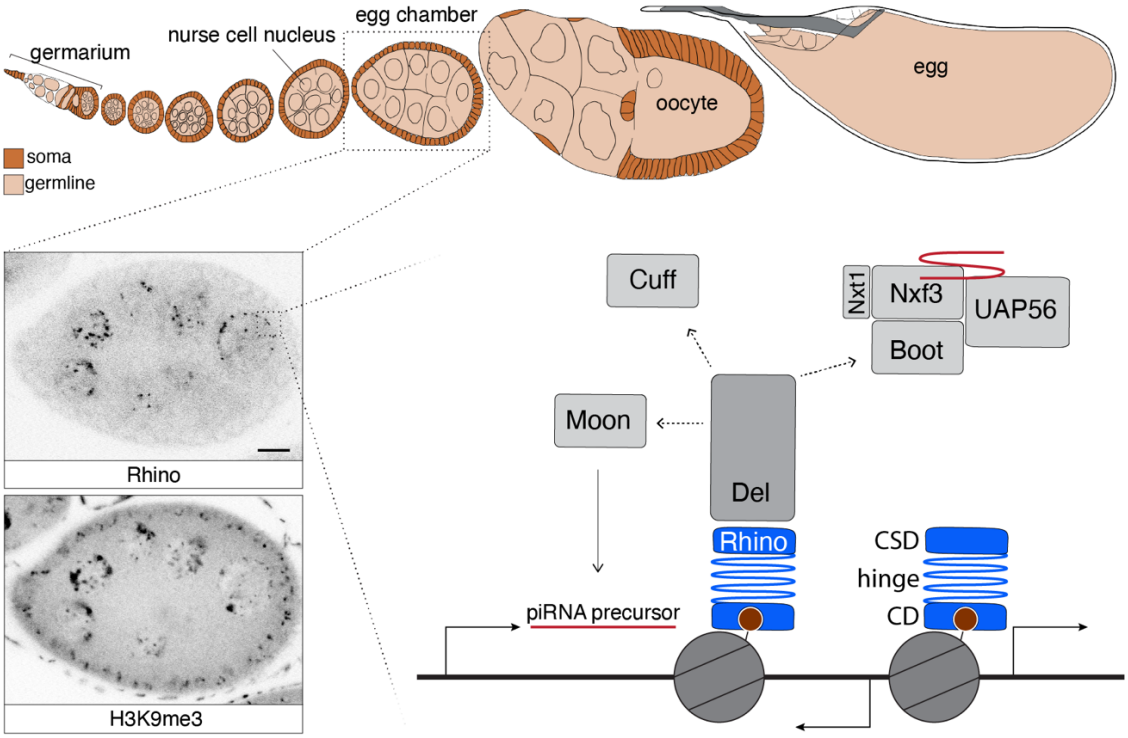


Figure 3: Rhino anchors piRNA cluster factors to heterochromatic dual-strand clusters. *Drosophila* ovarioles (top) represent egg production chains of progressively maturing egg chambers. Rhino forms foci in nurse cell nuclei corresponding to piRNA clusters (left), where Rhino recruits Moonshiner, Cutoff, and Bootlegger through Deadlock.

Rhino acts as a chromatin anchor for all of these factors and is central to their enrichment in discrete dots in nurse cell nuclei of the *Drosophila* ovary visible in immunofluorescence stainings (Figure 3). In the absence of Rhino piRNA production collapses and transposons are upregulated, leading to full female sterility (Klattenhoff et al., 2009). With Rhino binding to any chromatin locus, all downstream factors are recruited, leading to the Moonshiner-dependent transcription and Nxf3-Bootlegger-dependent export of potential piRNA precursor transcripts. Therefore, Rhino's genomic binding profile is of central importance to the definition of the germline piRNA pool. Like canonical HP1a, Rhino binds to the heterochromatic H3K9me3 histone mark present at dual-strand clusters using its chromo domain. Importantly, however, Rhino does not associate with H3K9me3-marked chromatin outside of piRNA clusters.

1.4.3 Comparison of Rhino and canonical HP1a

As a member of the HP1 protein family, Rhino contains the canonical HP1 protein domain architecture with a chromo domain specific for H3K9me3-binding, a flexible hinge domain, and a dimerizing chromoshadow domain (Figure 4). Rhino's involvement in the arms race between the piRNA pathway and transposable elements is thought to underly the positive selection of Rhino, which has led to considerable specialization of this HP1 variant. Signatures of positive selection are found to varying degrees in all protein domains (Vermaak et al., 2005). The chromo domain maintains binding affinity to H3K9me2/3 but contains several Rhino-specific amino acid substitutions with unknown function (Le Thomas et al., 2014, Yu et al., 2015a). The hinge domain massively expanded in Rhino, reaching more than four times the length of the HP1a hinge. Most prominent signs of Rhino's specialization to a new function are found at the chromoshadow domain, which completely changed its interaction surface to downstream protein. Instead of recognizing a PxVxL-like motif present in many heterochromatin proteins, the only known Rhino interactor is Deadlock, which interacts with the Rhino chromoshadow domain in a unique way involving two Deadlock molecules (Parhad et al., 2017, Yu et al., 2018b). The changed set of downstream factors results in a fundamentally different consequence of Rhino binding at a chromatin locus as compared to canonical HP1 function: while HP1a recruits factors that induce the compaction and repression of heterochromatin, Rhino's downstream factors act in the opposite way, inducing and stabilizing active transcription within heterochromatin. With this, Rhino has a very

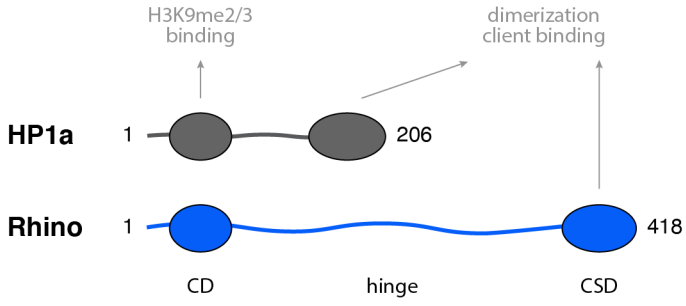


Figure 4: Cartoon comparing the domain architecture of HP1a and Rhino.

important function in the piRNA pathway, but also presents a dangerous protein that can fundamentally reverse heterochromatin identity upon binding. To prevent large scale decompaction of heterochromatin and most importantly aberrant piRNA production it seems crucial to limit Rhino to selected loci relevant for piRNA production. How this is achieved represents one of the most intriguing open questions in the piRNA pathway field.

1.4.4 Known players in Rhino deposition

The association of Rhino with dual-strand piRNA clusters is thought to be mediated by binding of its chromo domain to H3K9me3 present at these sites. However, how binding of Rhino to other H3K9me3-marked regions is avoided is not clear. Parallel studies form the Brennecke and Kalmykova labs revealed that Rhino does not only bind to large piRNA clusters composed of nested transposon fragments, but that it also associates with single transposon insertions on chromosome arms, where it induces the piRNA production from flanking regions (Mohn et al., 2014, Shpiz et al., 2014). Importantly, Rhino's binding at these sites was lost upon depletion of Piwi. This inspired a model for Piwi/piRNA-mediated Rhino deposition, where heterochromatin deposited though Piwi downstream factors was bound by Rhino. This was supported and further expanded by reports from several groups indicating that the specification of Rhino-dependent piRNA source loci depended on maternally deposited piRNAs, ensuring the specification of Rhino binding at the same sites that gave rise to abundant piRNAs in previous generations (De Vanssay et al., 2012, Hermant et al., 2015, Todeschini et al., 2010, Casier et al., 2019, Akulenko et al., 2018, Olovnikov et al., 2013, Komarov et al., 2020, Le Thomas et al., 2014). Interestingly, the efficiency of Rhino recruitment to piRNA targeted reporter constructs with identical sequence differed strongly depending on the genomic location of their insertion, indicating context-dependent regulation of this process.

While a Piwi/piRNA-dependent model of Rhino specification explains many aspects of Rhino's chromatin binding profile, it leaves many open questions. Firstly, Rhino binding at large piRNA clusters is not affected by the depletion of Piwi, arguing that other mechanisms specify these sites for Rhino binding. Secondly, the Piwi-mediated deposition of H3K9me3 upstream of Rhino cannot explain the selective binding of Rhino to only a subset of all H3K9me3-marked genomic locations. A requirement for additional specificity factors to achieve Rhino's specific binding to piRNA clusters is therefore undeniable.

1.5 Aims

In my PhD project, I set out to unravel how Rhino's chromatin binding specificity at a subset of H3K9me3-marked heterochromatin is achieved in *Drosophila* ovaries. I aimed to identify the factor(s) required to guide Rhino to piRNA source loci, and to investigate the contribution of the Rhino protein domains to this localization pattern. We hypothesized that additional specificity cues have to be present at piRNA clusters to be recognized by Rhino, to allow the distinction of these loci from regions of general heterochromatin. The Rhino protein, in turn, must harbor the ability to read out these cues, while canonical HP1a promiscuously binds to H3K9me3-marked nucleosomes genome wide. Furthermore, I aimed to test putative specificity factors for their ability to induce the recruitment of Rhino upon expression in an ectopic system.

The basis of Rhino's specific binding to piRNA cluster chromatin presented a long-standing question in the piRNA pathway field, and I envisioned my work to critically advance our understanding of this central step of piRNA source locus specification. Rhino is one of many HP1 variant proteins, of which most serve unknown and often tissue-specific functions. In light of this, I expected my findings to point to potential new concepts defining HP1 protein specificity that might be applicable outside of piRNA biology.

2. Results

2.1 The genome wide binding pattern of Rhino

As a basis for the investigation of how Rhino specifically associates with piRNA clusters, we first performed an extensive characterization of the genome-wide Rhino binding pattern. To set Rhino into context with general heterochromatin factors, we performed chromatin immunoprecipitation experiments followed by next generation sequencing (ChIP-seq) for Rhino, as well as HP1a, H3K9me2 and me3 in *Drosophila* ovaries.

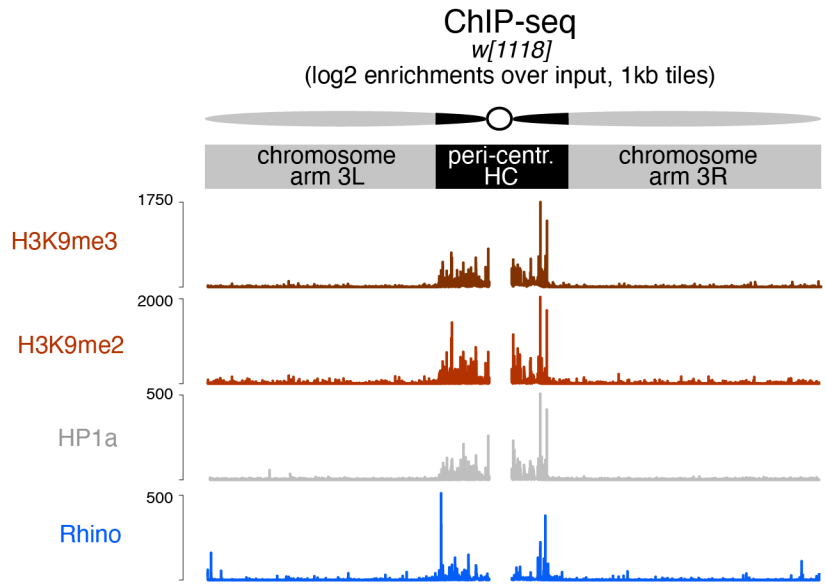


Figure 5: Chromosome wide ChIP-seq pattern of H3K9-methyl marks along with HP1a and Rhino shows distinct chromatin binding pattern of the two related proteins. ChIP-seq enrichment (genome unique reads per 1-kb tiles, one representative replicate each) of H3K9me3, HP1a, H3K9me2, and Rhino along the assembled chromosome 3 sequence in w^{1118} ovaries. Pericentromeric heterochromatin and euchromatic chromosome arms are indicated.

In accordance with previous studies, H3K9me3 and HP1a occupy large continuous pericentromeric regions, while slightly higher levels of H3K9me2 are also found along chromosome arms (Figure 5) (Janssen et al., 2018). For Rhino, we find a clear enrichment in pericentromeric heterochromatin, with the marked difference that unlike HP1a, Rhino binding is much less uniform across these regions. Furthermore, we also detect several Rhino-enriched regions in chromosomal arms and at telomeres where levels of HP1a and H3K9me3 are low. The literature suggests a connection of piRNA mediated heterochromatin formation at stand-alone transposon insertions with Rhino deposition as outlined above (section 1.4.4). We therefore set out to determine the contribution of strain specific transposon insertion profiles on the Rhino landscape. To this end, we characterized the localization of Rhino and general heterochromatin components via ChIP-seq in three *Drosophila* laboratory strains. We analyzed the commonly used w^{1118} wildtype strain, an MTD-Gal4 control knock down cross, as well as the *iso1* strain that was used to sequence the *Drosophila melanogaster* reference genome. Rhino-bound regions as well as H3K9me3-marked domains often lack defined boundaries, which complicates analysis via peak calling. We therefore determined ChIP-seq

enrichment values for nonoverlapping genomic 1-kb tiles reaching a minimum mappability of 25% (84.3% of the assembled genome). Using a stringent binary cutoff of 4-fold Rhino enrichment over input (p<0.036, Z-score >2.1), between 3.3% and 3.7% of the analyzable genome (3,136 1-kb tiles) were bound by Rhino in two independent replicates per genotype (dissected ovaries of the *w*^{1118}, MTD-Gal4, and *iso1* wild-type strains). Assuming that differences in general chromatin environment might influence Rhino biology, we divided the *Drosophila* genome into "pericentromeric heterochromatin" and "euchromatic chromosome arms" for further analyses, based on the sharp decline of continuous signal for HP1a and H3K9me3 found in pericentromeric regions (see Figure 5 and Supplementary Table 1). The prominent dual-strand piRNA clusters *38C*, *42AB*, and *80F* are all found at the border of pericentromeric heterochromatin and euchromatin (Brennecke et al., 2007). They were not assigned to either category and were analyzed separately as reference loci throughout this study. A similar number of tiles were bound by Rhino in pericentromeric heterochromatin and in chromosome arms in all three genotypes, despite the vast difference in size of the two compartments. Thereby, 14.5 - 28.3% of heterochromatic tiles, but only 1 to 2.7% of euchromatic tiles were bound by Rhino. While the well described piRNA clusters range among the most prominently Rhino bound regions, several other tiles, both in heterochromatin and in euchromatin, showed similar enrichment levels for Rhino ChIP-seq signal (Figure 6). This indicated that a reduction of Rhino biology to these three clusters would not provide a full picture of the cellular reality.

The comparison of Rhino binding profiles in different genotypes revealed largely congruent binding patterns in large clusters and pericentromeric regions, in line with the high degree of similarity expected for transposon profiles in these regions displaying low recombination rates (Figure 7). Interestingly, also among euchromatic tiles, where Rhino domains are expected to correspond to potentially strain-specific TE insertions, we observed considerable

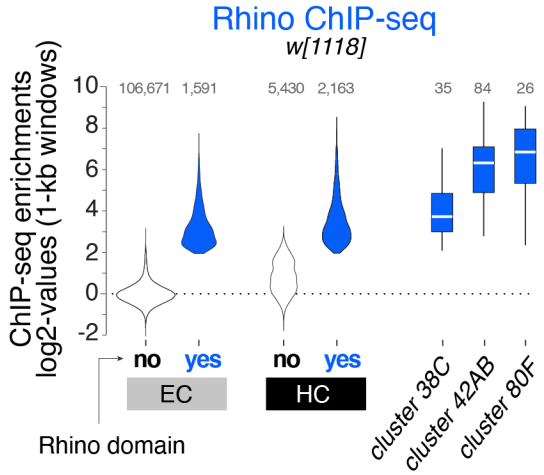


Figure 6: Rhino occupancy at many 1-kb tiles compares to the high enrichment at piRNA clusters. Violin plots showing average log2 fold enrichment of Rhino ChIP-seq over input for 1-kb tiles (n=2) from *w*^{1118} ovaries. Tiles were grouped into Rhino-bound and non-Rhino-bound based on a cutoff of 4-fold enrichment (corresponding to p = 0.036, Z-score = 2.1) of Rhino ChIP-seq signal over input in each replicate experiment. Rhino-dependent piRNA clusters *38C, 42AB*, and *80F* were analyzed separately (shown as box plots due to low number of tiles).

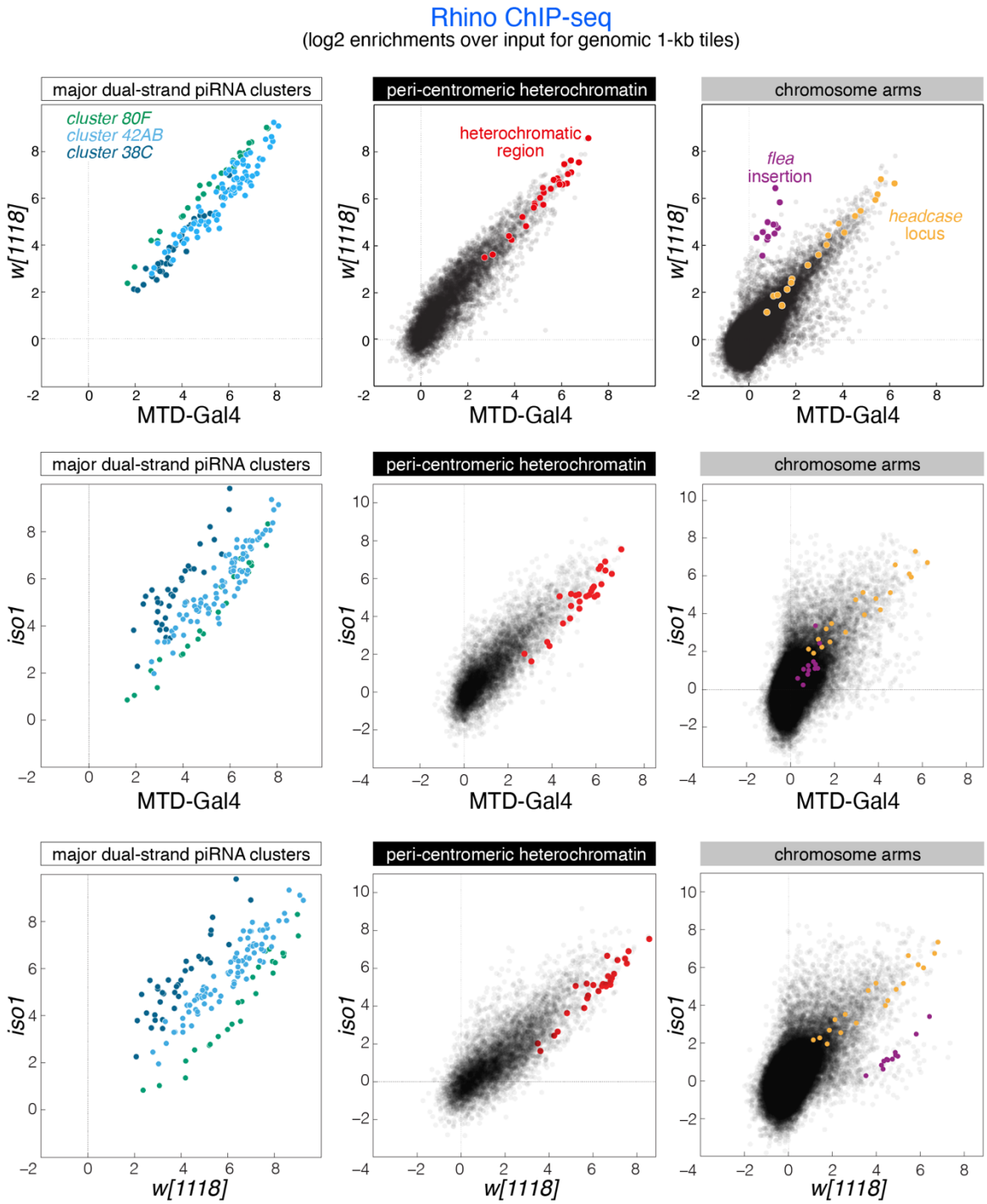


Figure 7: Comparison of Rhino ChIP-seq enrichment in different strains shows largely overlapping binding pattern. Scatter plot comparing average log2 ChIP-seq enrichments for Rhino in ovaries from the indicated genotypes (w^{1118} (n=2), MTD-Gal4 > *w*-sh (n=3), and *iso1* (n=2)). 1-kb tiles are separated into pericentromeric heterochromatin and chromosomal arms; piRNA clusters *38C, 42AB*, and *80F* are shown separately; colored 1-kb tiles correspond to example loci in Figure 8.

overlap between the Rhino binding pattern in the three genotypes. While several tiles did show strain-specific Rhino enrichment, the majority of tiles behaved the same.

We therefore asked, if this high degree of correlation was due to a more similar TE profile than expected, or if Rhino domains formed also independently of TEs at fixed genomic sites.

Manual inspection showed that many Rhino domains that were shared among all genotypes (most importantly also the *iso1* reference genome strain) had no association with TEs in neither the reference genome nor a previously determined TE insertion profile of MTD-Gal4 flies (Figure 8). No ChIP-seq enrichment for Rhino was detectable in experiments using ovaries from *rhino* mutant flies (Figure 8). We conclude that strain-specific TE insertions do impact the Rhino binding pattern, and that the impact of genetic perturbations on the Rhino profile should therefore be compared to genotype-matched controls. However, overall we find that Rhino domains are highly similar in different laboratory strains and that transposon-independent Rhino domains exist.

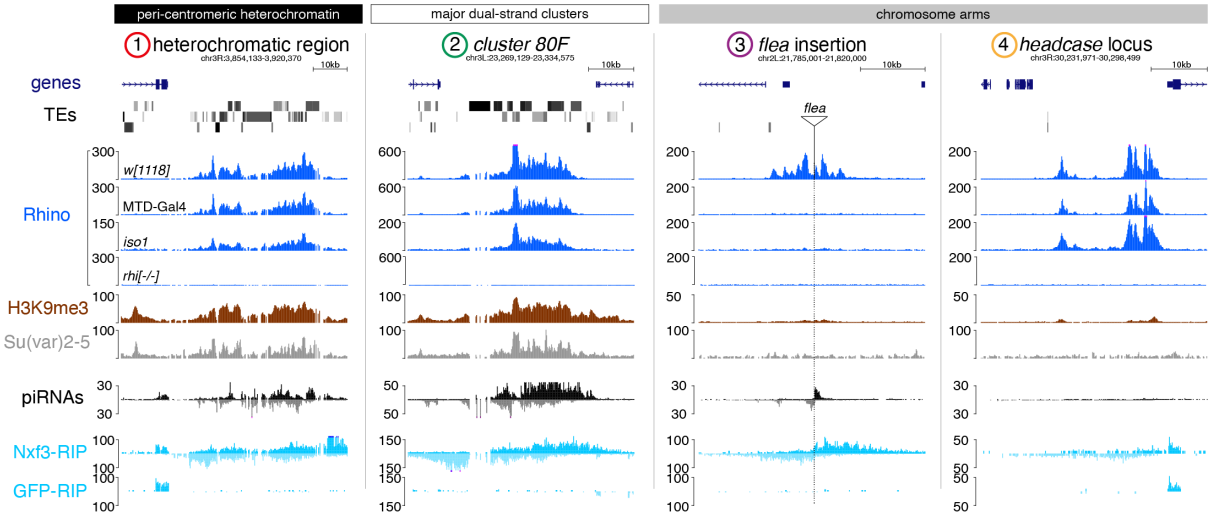


Figure 8: UCSC genome browser tracks depicting the diversity of Rhino domains. 1: heterochromatic, transposon-rich locus; 2: piRNA cluster *80F*; 3: strain-specific flea insertion (ElMaghraby et al., 2019) in w^{1118}; 4: Rhino domain proximal to euchromatic headcase locus. Unless indicated otherwise, data are from w^{1118} ovaries (ChIP and RIP signal are depicted as coverage per million reads, piRNA coverage normalized to miRNA reads, data is displayed for one representative replicate). GFP-RIP-seq serves as control for non-specific mRNA binding.

2.1.1 Rhino domains are diverse in size, content, and piRNA output

To further characterize the different types of Rhino domains in terms of their functionality, we sequenced piRNAs from all three strains, and made use of a previously published Nxf3 RNA immunoprecipitation (RIP)-seq data set from w^{1118} ovaries, which informs about the RDC-Moonshiner-dependent piRNA precursor pool (ElMaghraby et al., 2019). We consistently observed Nxf3 RIP-seq signal from Rhino-bound chromatin, regardless of the underlying sequence or TE association (Figure 8). This indicated that all Rhino-bound regions are functional and able to recruit all downstream factors required for the production and export of piRNA precursors.

The piRNA output, however, varied strongly between domains of similar Rhino enrichments. In line with their initial discovery due to the high levels of piRNAs they produce, piRNA clusters *42AB*, *38C*, and *80F* stood out as piRNA production hotspots. Genome unique flanking regions of euchromatic transposon insertions give rise to trailing phased piRNAs in a butterfly pattern,

as described previously (Figure 8) (Shpiz and Kalmykova, 2014, Mohn et al., 2014). Within TE-rich pericentromeric heterochromatin, large differences in piRNA levels reflect the binding profile of Rhino, ranging from strongly piRNA producing regions to long stretches of low – but detectable – continuous piRNA traces. Rhino domains devoid of TE association do not give rise to piRNAs (Figure 8). We argue that Rhino domains that are devoid of transposon sequences are poor piRNA producers as the emerging transcripts lack trigger sites to stimulate piRNA biogenesis in nuage (Han et al., 2015, Mohn et al., 2015).

Our data demonstrate that Rhino binds to a remarkably diverse set of genomic loci in euchromatin and heterochromatin. The focus on large, strongly piRNA producing piRNA clusters therefore does not adequately depict Rhino biology. We hypothesized that the deposition of Rhino will not be fundamentally different at large clusters versus TE-free regions, and that important insights can be gained from a more diverse set of Rhino binding sites. This work will therefore refer to Rhino-bound regions in general, and not be limited to piRNA producing piRNA clusters.

2.1.2 Rhino domains are marked by varying levels of H3K9me3

The H3K9me3 mark has been implicated in the specification of Rhinos' chromatin occupancy. Immunofluorescence stainings in nurse cell nuclei, as well as comparison of global ChIP-seq profiles of H3K9me3, HP1a, and Rhino support this connection, but highlight a clear discrepancy between the connection of HP1a and Rhino towards the H3K9me3 mark they both associate with. Consistent with the comparable *in vitro* affinity to H3K9me2/3 (~20 micromolar) of their respective chromo domains, both HP1a and Rhino signals were enriched at H3K9me3 loci (Mohn et al., 2014, Le Thomas et al., 2014, Yu et al., 2015a)(Figure 9). However, HP1a was found throughout the nucleus and was enriched in several large regions likely corresponding to constitutive heterochromatin. In contrast, Rhino was concentrated in multiple, discrete foci, distinct from the prominent HP1a domains (Figure 9).

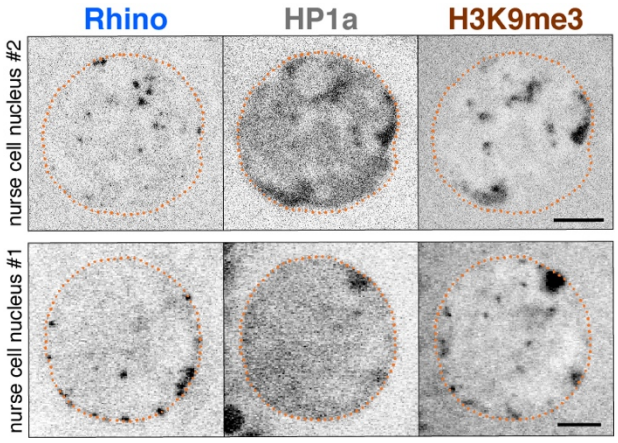


Figure 9: Immunofluorescence stainings for Rhino, HP1a, and H3K9me3 in single representative nurse cell nuclei of *Drosophila melanogaster* ovary egg chambers. Scale bar: 5 µm.

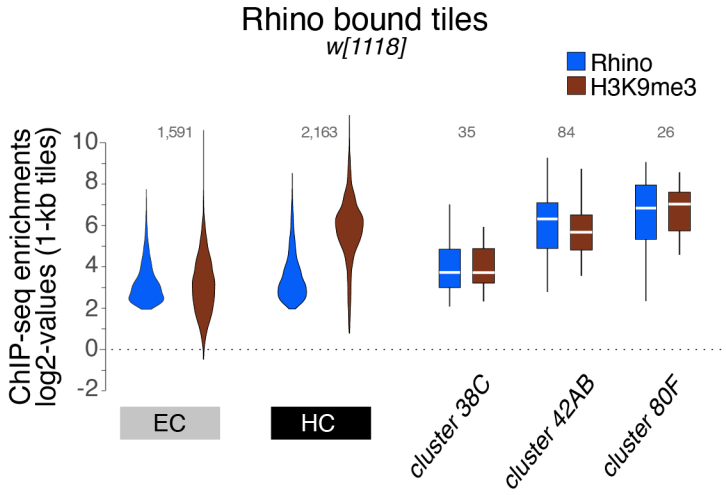


Figure 10: Rhino-bound genomic tiles are consistently marked by H3K9me3. Violin plots showing average log2 fold enrichment of Rhino and H3K9me3 ChIP-seq over input for 1-kb tiles (n=2) from *w*^{1118} ovaries. Tiles were grouped into Rhino-bound and non-Rhino-bound based on a cutoff of 4-fold enrichment (corresponding to p = 0.036, Z-score = 2.1) of Rhino ChIP-seq signal over input in each replicate experiment. Rhino-dependent piRNA clusters *38C, 42AB*, and *80F* were analyzed separately (shown as box plots due to low number of tiles).

In agreement with this, essentially all Rhino domains, in euchromatin and heterochromatin, were enriched in H3K9me3 ChIP-seq signal (Figure 10). However, H3K9me3 levels did not directly correlate with Rhino enrichments: Despite similar Rhino levels at heterochromatic and euchromatic Rhino-bound tiles, H3K9me3 levels were considerably higher for heterochromatic tiles (Figure 10). Importantly, while H3K9me3 levels globally correlated with HP1 enrichments, numerous loci, mostly in pericentromeric regions, displaying high H3K9me3 signal were not or only poorly bound by Rhino (Figure 11).

HP1a bound to H3K9me3 irrespective of the concomitant occupancy by Rhino, as evidenced by the enrichment of a tagged HP1a protein expressed exclusively in germline cells (Figure 12A), indicating that Rhino cannot out-compete HP1a for binding to this histone mark. Besides H3K9me3, both Rhino and HP1a have been shown to bind to H3K9me2 with

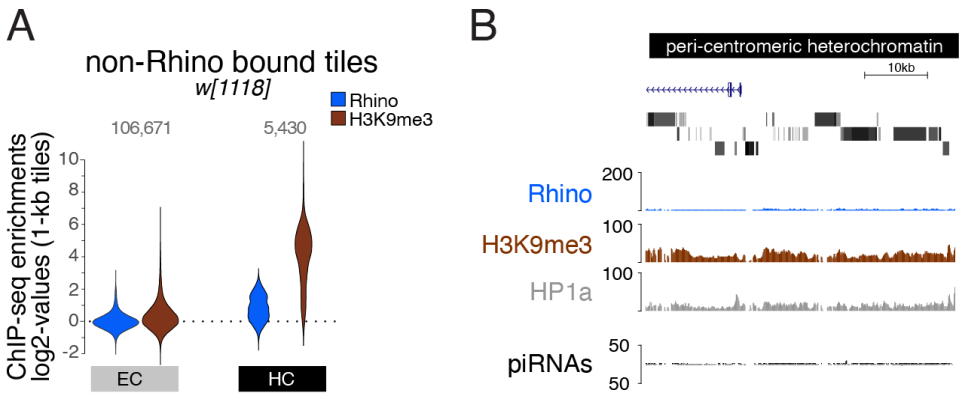


Figure 11: Not all heterochromatin is bound by Rhino. *(A)* Violin plot showing average log2 fold enrichment of Rhino and H3K9me3 ChIP-seq over input for 1-kb tiles (n=3) from *w*^{1118} ovaries at non-Rhino bound tiles. (B) UCSC genome browser tracks depicting indicated ChIP-seq and sRNA-seq data at a heterochromatic, non-Rhino bound locus. ChIP signal is depicted as coverage per million reads, piRNA coverage normalized to miRNA reads, data is displayed for one representative replicate from *w*^{1118} ovaries.

comparable, yet slightly lower affinity. H3K9me2 is abundant on chromosome arms where it often coincides with Rhino (Figure 12B). Similar to H3K9me3, however, a large fraction of H3K9me2 is found outside of Rhino-bound tiles. Thus, methylation of H3K9 alone, while being a hallmark of Rhino domains, cannot explain Rhino's chromatin profile.

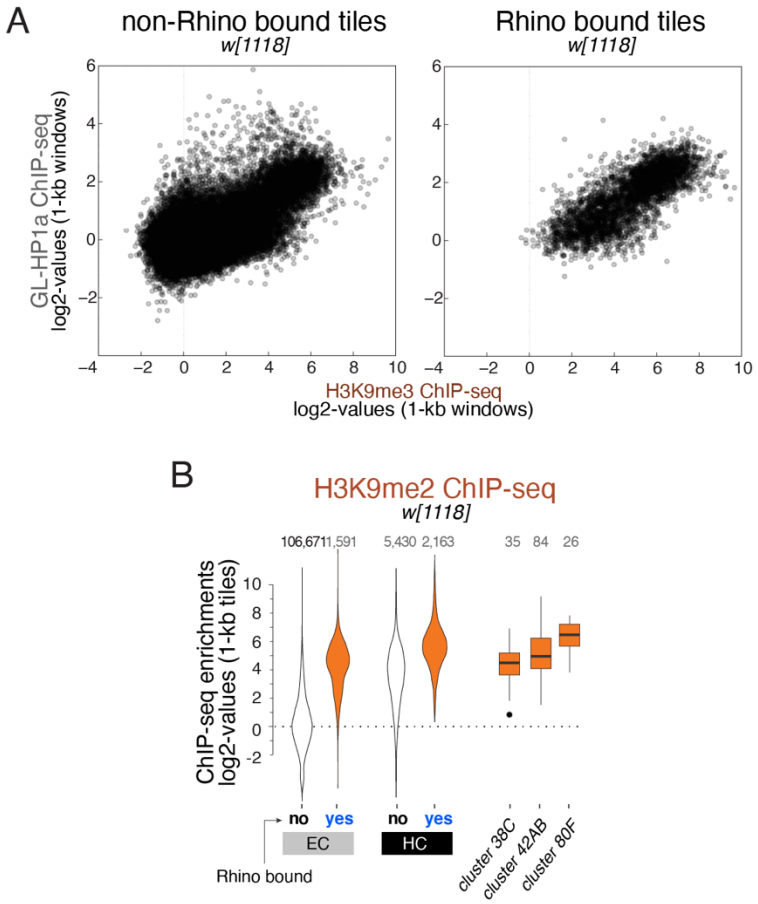


Figure 12: H3K9me3 marked chromatin is consistently bound by HP1a, irrespective of co-occupation by Rhino. (A) Scatter plot comparing the HP1a and H3K9me3 ChIP-seq enrichment at 1-kb tiles in the absence (left) or presence (right) of Rhino. (B) Violin plots showing average log2 fold enrichment of H3K9me2 ChIP-seq over input for 1-kb tiles (n=2) from *w*^{1118} ovaries. Tiles were grouped into Rhino-bound and non-Rhino-bound based on a cutoff of 4-fold enrichment (corresponding to p = 0.036, Z-score = 2.1) of Rhino ChIP-seq signal over input in each replicate experiment. Rhino-dependent piRNA clusters *38C, 42AB*, and *80F* were analyzed separately (shown as box plots due to low number of tiles).

2.1.3 HP1 variant proteins have different requirements for H3K9 methylation

Besides HP1a and Rhino, the ubiquitously expressed HP1b and HP1c variants are also present in ovaries and show broad nuclear localization without any discernable pattern (Figure 13A). As for canonical HP1a, the chromo domains of HP1b and HP1c have been shown to bind to methylated H3K9 peptides *in vitro* (Font-Burgada et al., 2008). However, while a clear connection to H3K9me3 levels is apparent for the binding pattern of Rhino and HP1a in ChIP-seq experiments, binding of HP1b and HP1c does not correlate with H3K9me3 (Figure 13B). Instead, both proteins are enriched at thousands of defined narrow peaks that are mostly located in euchromatic chromosome arms and are devoid of H3K9me2/3, HP1a, or Rhino

(Figure 13C). The chromatin binding of HP1b/c has been described to depend on the zinc-finger transcription factors Woc and Row (Font-Burgada et al., 2008). This indicates that, despite shared affinity for a chromatin mark, dedicated interaction partners can strongly modulate the chromatin binding pattern of HP1 proteins.

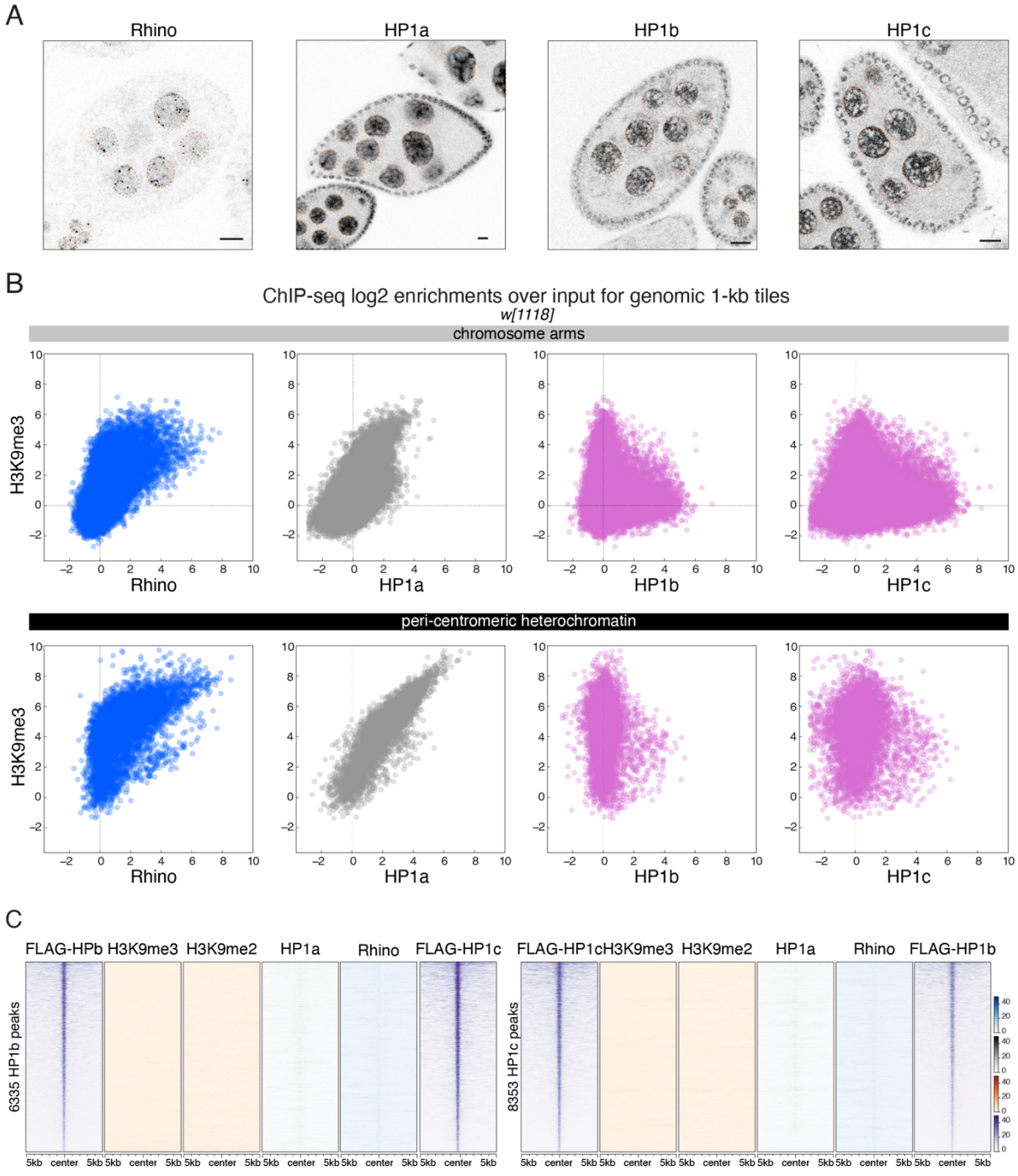


Figure 13: Different HP1 variant proteins expressed in ovaries have different requirements for H3K9 methylation. (A) Confocal microscopy images of *Drosophila* egg chambers expressing GFP-tagged HP1 variant proteins. Scale bar: 10 µm. (B) Scatter plot comparing average log2 ChIP-seq enrichments for indicated HP1 proteins and H3K9me3 in w^{1118} ovaries (1-kb tiles separated into pericentromeric heterochromatin and chromosomal arms). (C) Heat map centered on HP1b (left) or HP1c (right) peaks, respectively, depicting the indicated ChIP-seq signal (n=1) in ovaries from the w^{1118} strain.

2.2 The role of Rhino's protein domains in binding specificity

Despite similar domain architecture and comparable affinity to methylated H3K9, Rhino specifically binds to a subset of H3K9me2/3 marked regions, while HP1a globally associates with this histone mark. We therefore hypothesized that additional molecular determinants must be read out differentially by Rhino, but not HP1a, to achieve Rhino's specific localization pattern. To determine how such a specificity cue might be recognized by Rhino, we aimed to delineate which protein domain of Rhino was most important for its correct localization. Early reports comparing the binding patterns of the three main HP1 variant proteins in *Drosophila* concluded that all protein domains contributed to the correct localization of HP1a, b, and c (Smothers and Henikoff, 2001). While the chromo domain establishes the direct link to chromatin, its specificity did not seem to differ strongly between the different variants, indicating that interacting proteins bound at the chromoshadow domain or biophysical properties mediated via the hinge domain contributed more strongly to the different localization of HP1 variant proteins.

In an analogous approach, we constructed a set of domain swap constructs, harboring different combinations of chromo-, hinge, and chromoshadow domains of HP1a and Rhino. In addition, prompted by recent findings about the importance of the N-terminal extension (NTE) in human HP1alpha (Larson et al., 2017), we added one construct with the HP1a NTE on Rhino, while all remaining constructs harbored the Rhino NTE. An overview of all constructs is shown in Figure 14. Domain boundaries were determined based on amino acid alignments of various HP1 proteins in *Drosophila* species as well as the mouse and human counterparts (Appendix 1). Investigation of the localization of these proteins by immunofluorescence imaging and ChIP-seq revealed interesting results, but also several limitations of this approach.

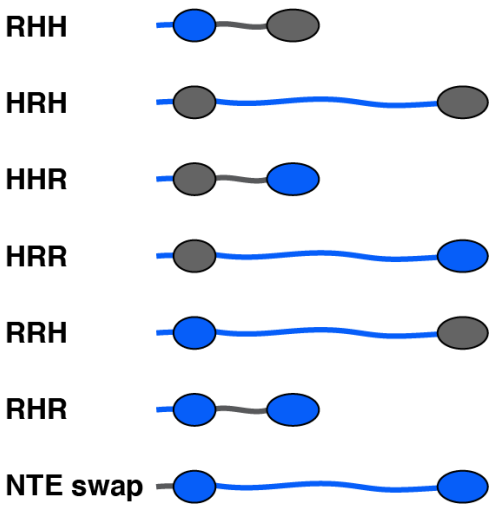


Figure 14: Overview of chimeric domain swap constructs. Blue indicates domains from Rhino, grey indicates HP1a domains.

2.2.1 Chimeric domain swap constructs show different localization patterns

Due to the homodimerization ability of HP1 chromoshadow domains, as well as potential diffuse biophysical interactions via the hinge region, we argued that the presence of wildtype Rhino protein would influence the localization pattern of chimeric constructs, most importantly of those constructs harboring the Rhino chromoshadow domain. To minimize such confounding effects, we stably integrated chimeric Rhino-HP1a domain swap constructs into a landing site recombined with a *rhino* frame shift mutant allele. Expression of wildtype 3xFLAG-GFP-tagged Rhino protein from an ectopic cDNA transgene driven by the *rhino* promoter fully restored fertility of *rhino* knock out females, indicating the functionality of tagged Rhino protein (Figure 15).

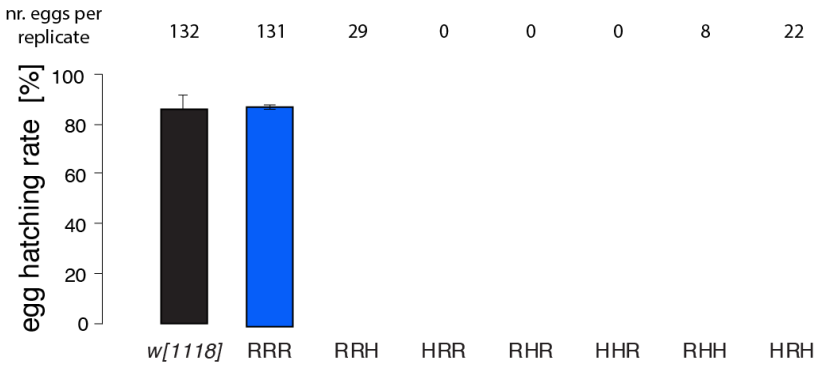


Figure 15: Wildtype Rhino, but not chimeric constructs combining domains from Rhino and HP1a, is able to rescue female fertility. Hatching rates for eggs laid by w^{1118} control females in comparison to females carrying a chimeric domain swap construct or wildtype Rhino (RRR) expressed from the *rhino* promoter. Total number of eggs laid is indicated as n for each genotype.

None of the chimeric proteins rescued the fertility defect of *rhino* KO flies. This demonstrates that all protein domains are important for full functionality of Rhino, which is in agreement with positive selection described for all domains of Rhino (Vermaak and Malik). Moreover, we observed strong differences in the number of eggs laid by females carrying chimeric Rhino-HP1a constructs: while loss of Rhino does not impact the number of eggs laid per female, we observed strongly reduced numbers of eggs laid by females expressing chimeric constructs. In line with this, we observed drastically reduced ovary size mainly for flies expressing constructs combining the HP1a chromo domain with the Rhino chromoshadow domain, irrespective of the hinge (HRR and HHR) in the *rhino* KO background. Importantly, loss of Rhino does not strongly affect ovary morphology, indicating a dominant negative effect of these constructs (Volpe et al., 2001).

Tagged Rhino localized in a pattern reminiscent of endogenous Rhino in immunofluorescence analysis (Figure 16). The localization patterns observed for the chimeric constructs in the absence of wildtype Rhino protein are summarized in Figure 16: Constructs containing the HP1a chromoshadow domain localized in few, large patches, reminiscent of wildtype HP1a protein, regardless of other protein domains. The Rhino chromoshadow domain was not able to support formation of local accumulations in the presence of the HP1a hinge, which

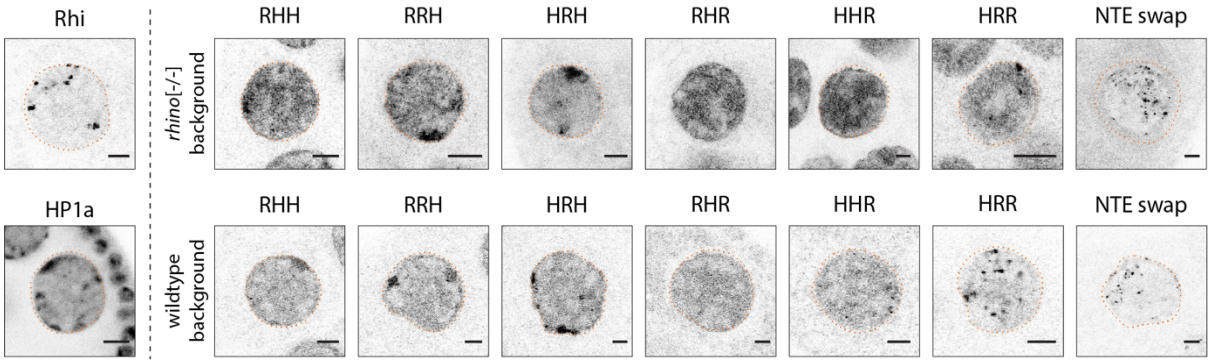


Figure 16: Localization pattern of chimeric domain swap constructs expressed under the control of the *rhino* promoter. Individual nurse cell nuclei of flies harboring GFP-tagged domain swap constructs are depicted. Scale bar: 5 µm. Dotted line: nuclear outline based on DAPI.

conferred diffuse nuclear localization. Finally, the localization pattern of the HRR construct containing the chromo domain of HP1a in combination with the Rhino hinge and chromoshadow domain had a largely diffuse localization pattern with few discrete foci in a subset of nuclei. This seemingly Rhino-like localization pattern of HRR was in disagreement with the ovary morphology defect caused by expression of HRR and HHR constructs, which clearly indicate adverse effects of these proteins. Replacement of the NTE had no effect on Rhino localization, mimicking the small dots observed in flies harboring the wildtype rescue construct.

Dimerization with wildtype Rhino changes the localization of chimeric constructs

To test if presence of wildtype Rhino indeed influenced the immunofluorescence pattern of chimeric constructs, we analyzed ovaries from heterozygous flies carrying one wildtype copy of Rhino (Figure 16 bottom). In the presence of wildtype Rhino, constructs containing the HP1a chromoshadow domain displayed a pattern resembling the localization of HP1a, with several larger patches per nucleus. Constructs harboring the HP1a hinge in combination with the Rhino chromoshadow domain showed diffuse nuclear localization, with only few dim foci observed for the HHR construct. The HRR construct harboring the HP1a chromo domain in the context of the Rhino protein formed small dots in nurse cell nuclei, as did the NTE replacement construct.

Based on these combined data we conclude that the exchange of the chromo domain, despite having little impact on the immunofluorescence localization pattern, has the strongest effect on Rhino functionality. The Rhino hinge region is required for formation of foci, which might be due to its biophysical properties as previously observed for HP1 proteins (Keenen et al., Strom et al., 2017). The role of the chromoshadow domain is hard to delineate, due to its putative dimerization with wildtype proteins. Due to embryonic lethality of an HP1a knock out, we were not able to distinguish if this effect was caused by genuine guidance factors interacting at the chromoshadow domain, or due to homodimerization with wildtype HP1a.

2.2.2 Chimeric domain swap constructs show different chromatin binding patterns

Strongly reduced ovary size with dramatically decreased germline tissue are usually caused by defects in germline stem cell fitness, and the lack of germline tissue complicates large-scale experiments such as ChIP-seq, which we deemed crucial for the precise assessment of protein domain function. To leave the germ line stem cells unaffected by our constructs and allow for better ovary development, we therefore revised our swap construct expression strategy and created flies expressing the same constructs under the control of maternal tubulin *tub67C*, which starts to be expressed after cyst formation (Matthews et al., 1989). Starting from mid stage cysts, wildtype Rhino expressed under the control of the maternal tubulin promoter (*mtub*>GFP-Rhino) formed distinct dots in nurse cell nuclei reminiscent of endogenously expressed wildtype Rhino (Figure 17). Ovary size upon expression of HP1a-CD-containing constructs under control of the maternal tubulin promoter was still reduced in the *rhino* KO background, but intact egg chambers and good overall ovary morphology now allowed full analysis of these flies.

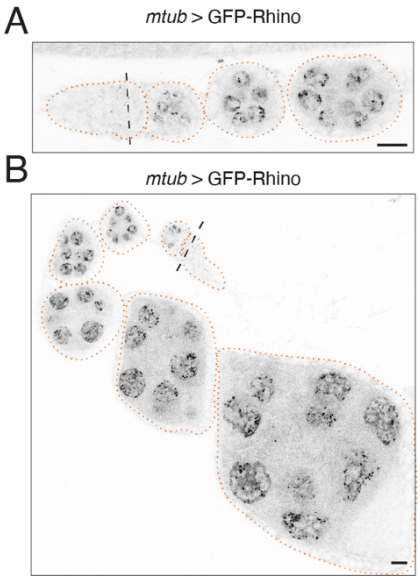


Figure 17: Rhino expressed under the control of maternal tubulin regulatory elements (*mtub*>GFP-Rhino) forms dots reminiscent of the wildtype Rhino localization pattern. The GFP signal for *mtub*>GFP-Rhino in the germarium (A) and ovariole (B) is depicted in green. Scale bar: 5µm, dashed line indicates onset of expression, orange dotted line indicates outline of egg chambers.

To determine if localization patterns observed in stainings are predictive of chromatin occupancy patterns, we conducted a more functional analysis of the tubulin promoter-controlled domain swap constructs via ChIP-seq. Importantly, late expression of Rhino results in a ChIP-seq pattern highly similar to wildtype (Figure 18A). The HP1-like localization pattern of RRH, RHH, and HRH constructs was reflected also in their ChIP-seq pattern, with the HRH construct almost perfectly mimicking the HP1a enrichment (Figure 18A). RRH and RHH were found in the same regions, albeit at slightly lower levels (Figure 18A).

Upon closer inspection, RRH in certain locations more resembled the Rhino pattern, indicating that either the hinge or chromo domain of Rhino do not perfectly share their function with

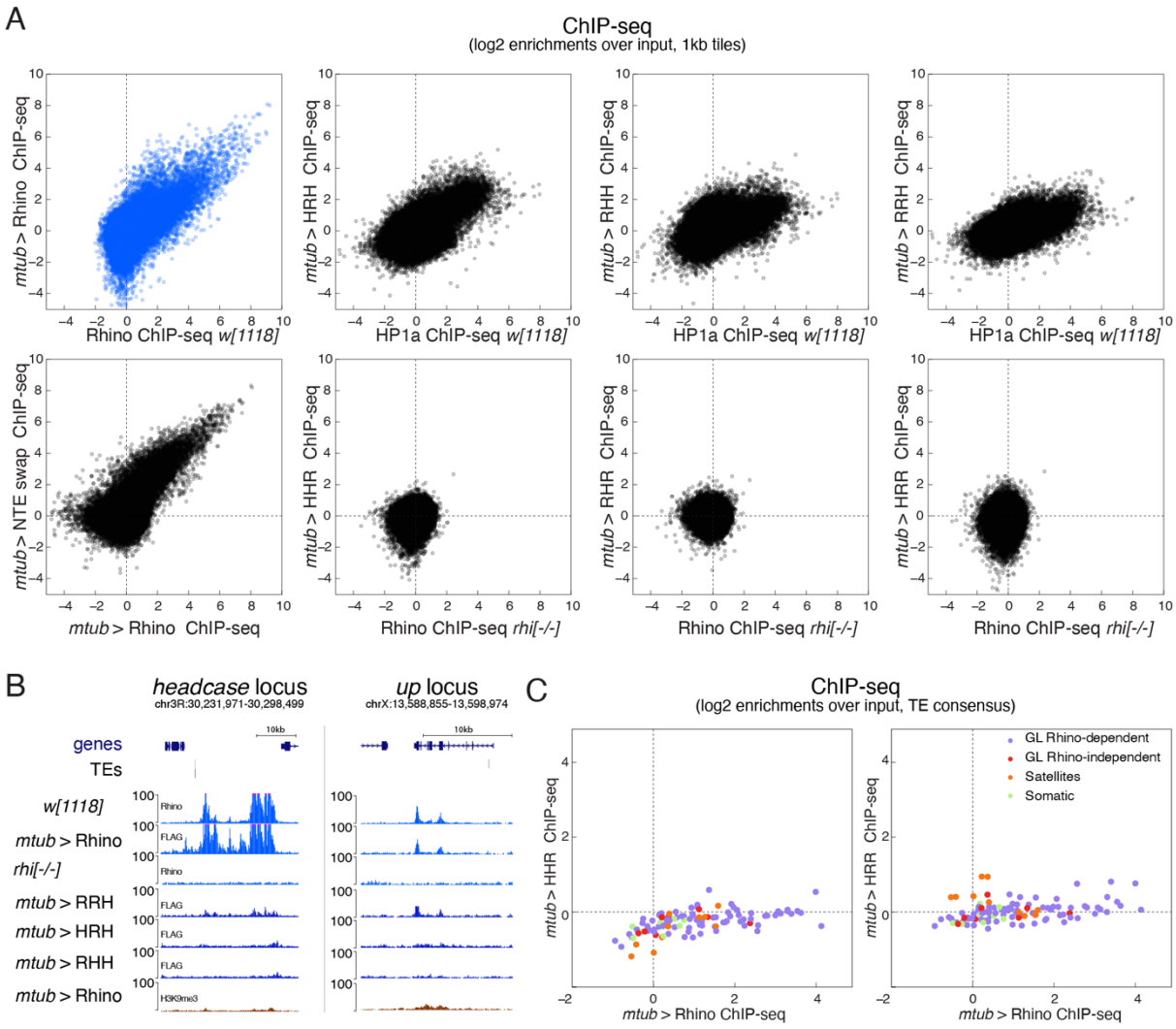


Figure 18: ChIP-seq pattern of chimeric domain swap constructs expressed under the maternal tubulin regulatory elements. (A) Scatter plots comparing the log2-fold ChIP-seq enrichment of various chimeric domain swap constructs to the chromatin binding pattern of wildtype HP1a and Rhino for 1-kb tiles. (B) UCSC genome browser tracks depicting the indicated signal at two example Rhino domains. (ChIP-seq signal: coverage per million reads; data is displayed for one representative replicate). (C) Scatter plots comparing the log2 ChIP-seq enrichment for indicated constructs to the wildtype Rhino enrichment at transposon consensus sequences.

their HP1a counterparts (Figure 18B). Similar to the immunofluorescence experiment, we detect no effect of the NTE exchange on the chromatin binding pattern of Rhino (Figure 18A). Finally, we don't find strong enrichment of the RHR and HHR constructs in any chromosomal regions in the *rhino* mutant background, which is in line with their diffuse nuclear localization (Figure 18A). Curiously, this is also true for the HRR construct, for which we had observed discrete dots in nurse cell nuclei. The lack of enrichment for HRR in our analyses was surprising, as the strong effect on ovary morphology of this construct would indicate association of the HRR chimeric protein to ectopic regions where it could cause deleterious effects. Analysis of ChIP-seq reads mapping to transposable element consensus sequences also showed no enrichment (Figure 18C). To determine if the HHR and HRR constructs impaired general heterochromatin integrity, we analyzed the H3K9me3 levels in the respective ovaries. We found a global reduction of H3K9me3 enrichments in both genotypes,

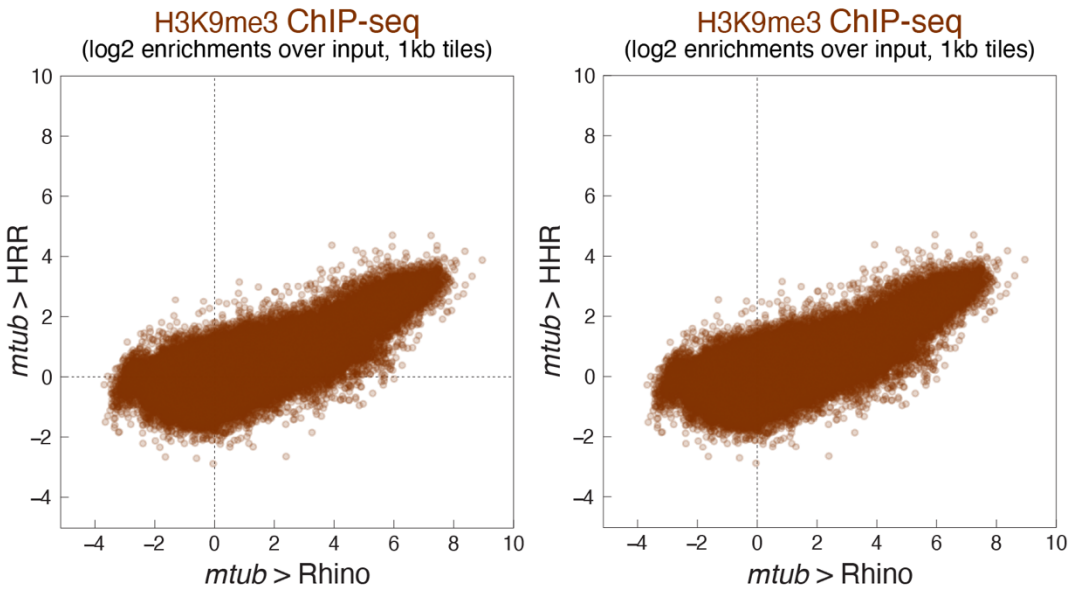


Figure 19: H3K9me3 levels in ovaries expressing chimeric domain swap constructs with negative effects on ovary morphology is reduced. Scatter plots depicting the H3K9me3 ChIP-seq enrichment in log2 scale for the indicated constructs.

but not in ovaries harboring the other chimeric constructs (Figure 19). It is hard to delineate, if this reduction is a direct effect of the presence of these constructs, or a secondary consequence of the distorted ovarian morphology of these samples, which impacts the tissue composition and complicates a direct quantitative comparison.

Analysis of heterozygous flies indicated that, as expected, dimerization with wildtype Rhino can slightly influence the localization pattern of constructs harboring the Rhino chromoshadow domain. Especially at piRNA clusters *42AB* and *80F* the binding pattern of HRR differs strongly between heterozygous and homozygous flies (Figure 20). This indicated that

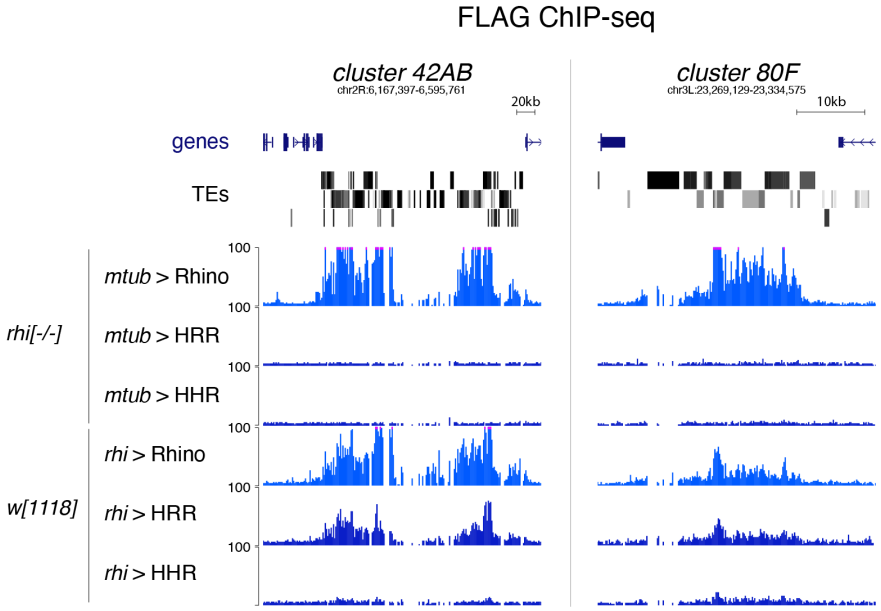


Figure 20: Presence of wildtype Rhino influences the binding pattern of chimeric swap constructs containing the Rhino chromoshadow domain. UCSC genome browser tracks depicting the indicated signal at two example Rhino domains. (ChIP-seq signal: coverage per million reads; data is displayed for one representative replicate).

dimerization with wildtype Rhino, and likely wildtype HP1a can strongly influence the binding pattern of chimeric domain swap constructs.

In conclusion, we find that, similar to previous reports, each domain of HP1 proteins contributes to the exact chromatin binding pattern of the respective variant protein. Our data indicate that dimerization with intact wildtype proteins via the chromoshadow domain is a strong determinant for protein localization, and that the chromo domain is likely more than just a reader of H3K9me3, and might integrate combinations with other histone marks or interact with other factors.

2.2.3 Caveats of chimeric domain swap constructs

The unavoidable dimerization with wildtype HP1a makes it hard to distinguish between true recruitment mechanisms acting on the chromoshadow domain on the one hand and dimerization-mediated influences of fully functional wildtype HP1a on the other. This strongly limits the conclusions drawn from all constructs containing the HP1a chromoshadow domain. Furthermore, reassessment of domain borders for the chromo domain indicated that our constructs were likely constructed too stringently for Rhino, potentially interfering with an alpha helical extension of the Rhino chromo domain that is not present in HP1a (Le Thomas et al., 2014, Yu et al., 2015a). Therefore, constructs harboring the Rhino chromo domain followed by the HP1a hinge have to be analyzed with caution, as the integrity of the Rhino chromo domain might be impaired in these constructs. This could explain the lack of any defined binding pattern of RHR in either homozygous or heterozygous background. Together these limitations confound the analysis of all but the HHR construct and the NTE swap, strongly limiting any firm conclusions (Alignment in Appendix).

2.2.4 Revised design for testing of Rhino protein domain function

It was recently shown that chromo domains of various protein families are able to recognize their cognate histone modifications and reproduce the *in vivo* binding pattern of the full proteins when expressed as head-to-tail dimers (Villaseñor et al., 2020). The advantage of such constructs is the full circumvention of dimerization, allowing the characterization of the localization potential of a tandem chromo domain in isolation. We implemented respective constructs for Rhino, HP1a, and HP1c and overexpressed them using the MTD-Gal4 driver system in otherwise wildtype flies. Overexpression of the HP1a chromo domain construct resulted in highly atrophic ovaries. This was reminiscent of the ovary morphology phenotype observed for chimeric domain swap constructs harboring the HP1a chromo domain, further supporting the notion that an excess of HP1a chromo domain has a dominant negative effect due to unknown mechanisms. Constructs containing the Rhino and HP1c chromo domains showed largely diffuse nuclear localization with the exception of few accumulations in a subset of nurse cell nuclei (Figure 21A). To determine if this largely diffuse pattern was caused by the lack of chromatin binding for these two constructs or merely due to strong

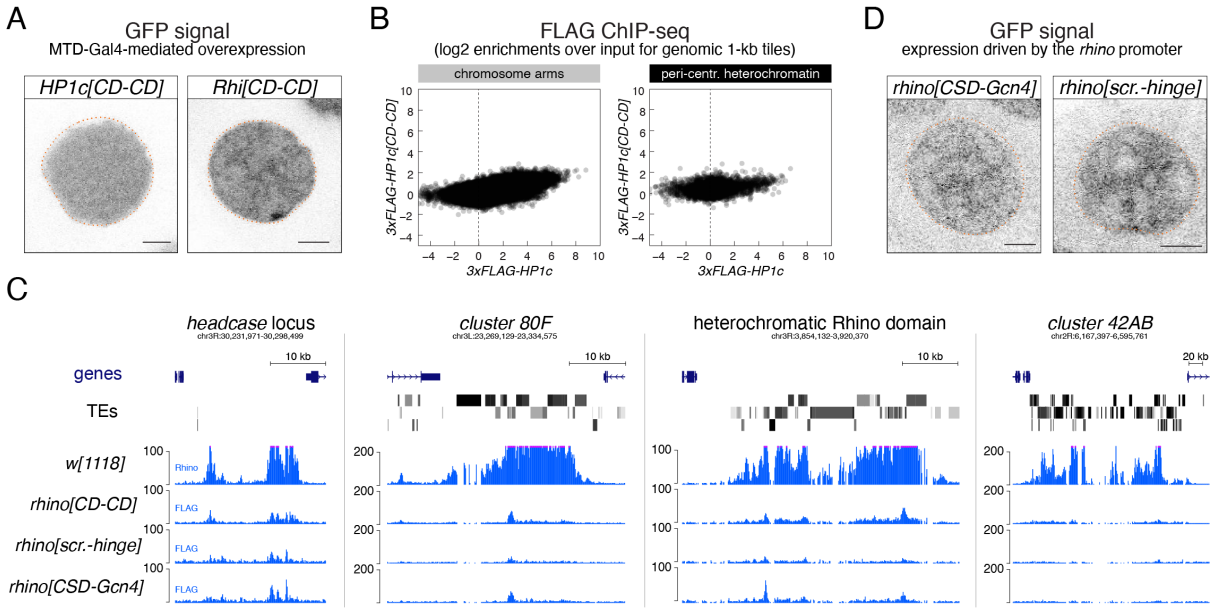


Figure 21: The chromo domain of Rhino allows specific low-level binding at Rhino domain. (A) Confocal microscopy images showing the localization of chromo domain dimer constructs in nurse cell nuclei upon MTD-Gal4-driven expression. Dotted line: nuclear outline based on DAPI. Scale bar: 5 µm. (B) Scatter plot contrasting the chromatin binding profile of the HP1c chromo domain dimer versus endogneously tagged wildtype HP1c for genomic 1-kb tiles. (C) UCSC genome browser tracks showing the indicated ChIP-seq signal (coverage per million reads) at diverse Rhino domains. (D) Confocal microscopy images showing the localization of modified Rhino constructs in nurse cell nuclei. Dotted line: nuclear outline based on DAPI. Scale bar: 5 µm.

overexpression masking more defined patterns, we characterized their chromatin binding profile. The HP1c chromo domain dimer failed to associate with chromatin. We did not detect any enrichment of this construct at HP1c binding sites or in H3K9me3-marked pericentromeric heterochromatin (Figure 21B), indicating that the HP1c chromo domain requires dedicated supporting factors for chromatin binding, mediated by the hinge or chromoshadow domain. The Rhino chromo domain dimer behaved differently. It bound to several Rhino domains of diverse characteristics, although at strongly reduced levels compared to wildtype Rhino (Figure 21C). Among the three largest piRNA clusters, only *cluster 80F* showed clear low-level enrichment of the Rhino chromo domain dimer. Importantly, binding of the Rhino chromo domain construct remained restricted to Rhino domains and did not extend to general heterochromatin (Figure 22). This strongly indicated that the Rhino chromo domain held the potential to read out specificity cues for the targeted binding of Rhino at selected sites.

Despite binding to many Rhino domains, enrichment levels of the Rhino chromo domain dimer often remained low and seemed to be restricted to more narrow regions of large Rhino domains (Figure 21C). This indicated that full length Rhino relies on the hinge or chromoshadow domain for the full extent of chromatin binding. To differentiate which domain was more important for this function, we generated flies expressing artificial Rhino variants in the absence of wildtype Rhino. We substituted the Rhino chromoshadow domain with the Saccharomyces cerevisiae Gcn4 dimerization domain to maintain dimerization capability (Rhi^{GCN4}, (O'Shea et al., 1991)) to assess if the chromo domain and hinge were

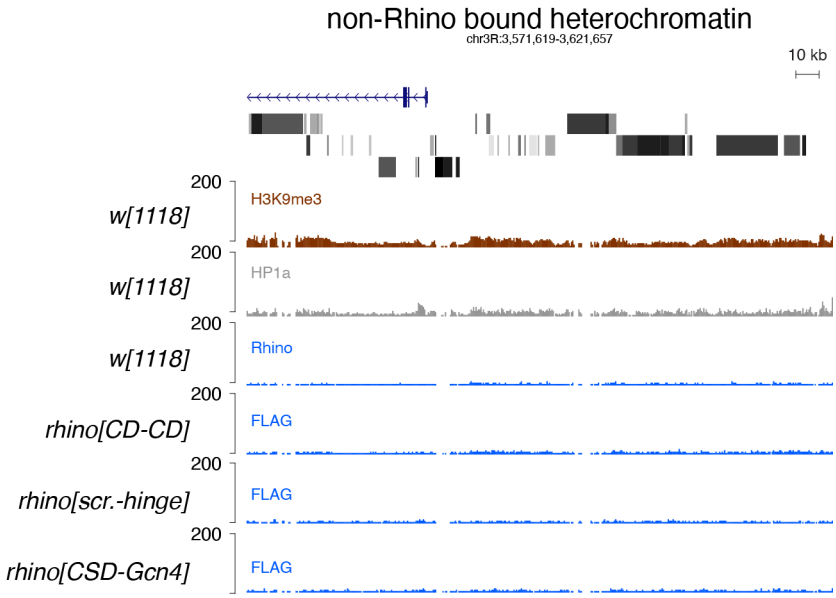


Figure 22: Rhino chromo domain constructs don't associate with general heterochromatin. UCSC genome browser tracks showing the ChIP signal for indicated Rhino constructs at an example region in general heterochromatin (Coverage per million reads).

sufficient to establish the wildtype Rhino binding pattern. To test for the sufficiency of the chromoshadow domain, we scrambled the amino acids of the hinge region to create an unstructured linker to the chromo domain ($Rhi^{scr.hinge}$). Rhi^{GCN4} and $Rhi^{scr.hinge}$, expressed from the *rhino* promoter in the *rhino* mutant background, showed broad nuclear localization (Figure 21D). ChIP-seq analysis revealed that both constructs failed to form extended Rhino domains, but demonstrated clear signal for both constructs at many sites, resembling the ChIP-seq enrichment for the Rhino chromo domain dimer (Figure 21C). This indicated that, while the chromo domain can mediate low level binding at many sites, full length Rhino was required for binding to Rhino domains to the full extent.

2.3 The role of uncharacterized ZAD-ZnF protein CG2678/Kipferl in Rhino specification

To achieve its specific localization to a subset of H3K9me3-marked chromatin, Rhino must rely on additional specificity cues. These can either constitute chromatin components such as additional histone tail modifications, or interacting protein factors binding to chromatin upstream of Rhino. Inspired by the described role of interacting proteins Woc and Row for the chromatin binding pattern of HP1b and HP1c (Font-Burgada et al., 2008), we asked whether DNA-binding proteins impact Rhino's chromatin binding profile. To this end, we performed co-immunoprecipitation experiments with tagged Rhino from ovary lysate, followed by quantitative mass spectrometry. This yielded no significant interactor besides the known Rhino-interactor Deadlock, which connects Rhino to downstream effectors in transcription. Reasoning that protein-protein interactions relevant for Rhino guiding might be restricted to chromatin and disrupted during lysate preparation, we performed a proximity labeling experiment by co-expressing GFP-tagged Rhino together with low levels of the TurboID biotin ligase fused to a GFP nanobody in ovaries (Figure 23A). After isolating biotinylated proteins under denaturing conditions, quantitative mass spectrometry revealed proteins enriched in the Rhino-GFP sample versus a control sample. Besides Deadlock and Cutoff, which together with Rhino make up the RDC complex (Mohn et al., 2014), as well as the piRNA cluster factor Bootlegger (ElMaghraby et al., 2019, Kneuss et al., 2019), the uncharacterized protein CG2678 stood out as a putative direct Rhino interactor as it was also a top hit in a yeast two-hybrid screen with Rhino as bait and a cDNA library from ovaries as prey (Figure 23B, Supplementary Table 2).

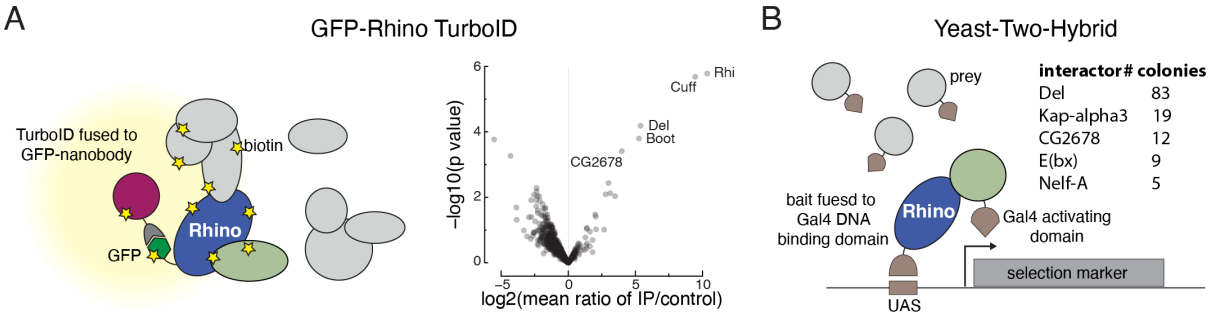


Figure 23: Two-fold strategy for the identification of Rhino interactors finds CG2678 as putative candidate. (A) Cartoon illustrating flexible GFP-nanobody-TurboID setup used for the characterization of proteins in close proximity to GFP-tagged Rhino (left) and volcano plot showing fold enrichment of proteins determined by quantitative mass spectrometry (Doblmann et al., 2018) in GFP-Rhino TurboID samples versus nuclear GFP TurboID control (n = 3 biological replicates; statistical significance based on two-sided t-test; P values corrected for multiple testing (Benjamini–Hochberg)). (B) Cartoon depicting yeast two-hybrid screen set up, with top Rhino interactors listed.

CG2678 further stood out among other candidates because of its domain architecture: two splice isoforms are annotated for CG2678, containing one or two arrays of four and three C2H2 Zinc fingers (ZnFs), respectively (Figure 24). Both isoforms harbor an N-terminal zinc finger associated domain (ZAD), which is a homodimerization domain frequently found in insect zinc finger proteins (Chung et al., 2007, Chung et al., 2002). *D. melanogaster* encodes

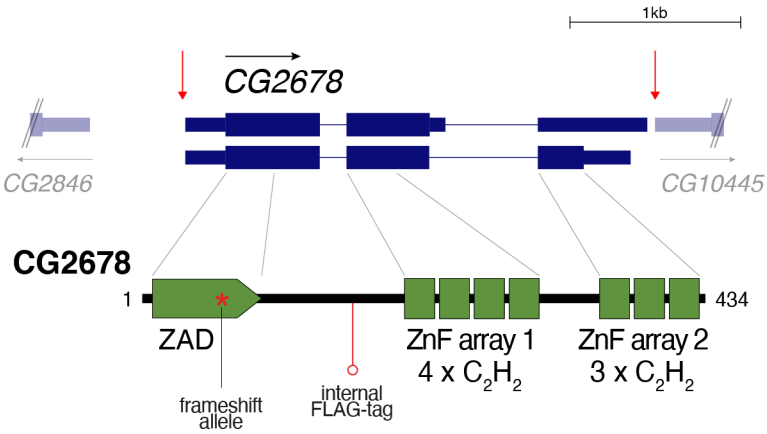


Figure 24: CG2678 encodes a ZAD-ZnF protein harboring two arrays of C_2H_2 zinc fingers. Genomic *CG2678* locus depicting the two annotated transcripts and CG2678 protein domain architecture (location of the frame shift mutation (red asterisk), internal 3xFLAG affinity tag (red circle), and cleavage sites for full locus deletion (red arrows) are indicated).

more than 90 ZAD-ZnF proteins, of which only a handful has been characterized for their functions in transcription (Bag et al., 2021, Harms et al., 2000), genome organization (Maksimenko et al., 2015, Sabirov et al., 2021a), and heterochromatin biology (Kasinathan et al., 2020a). As a predicted DNA-binding protein and a putative Rhino interactor, CG2678 was an intriguing candidate for a factor defining Rhino's chromatin specificity.

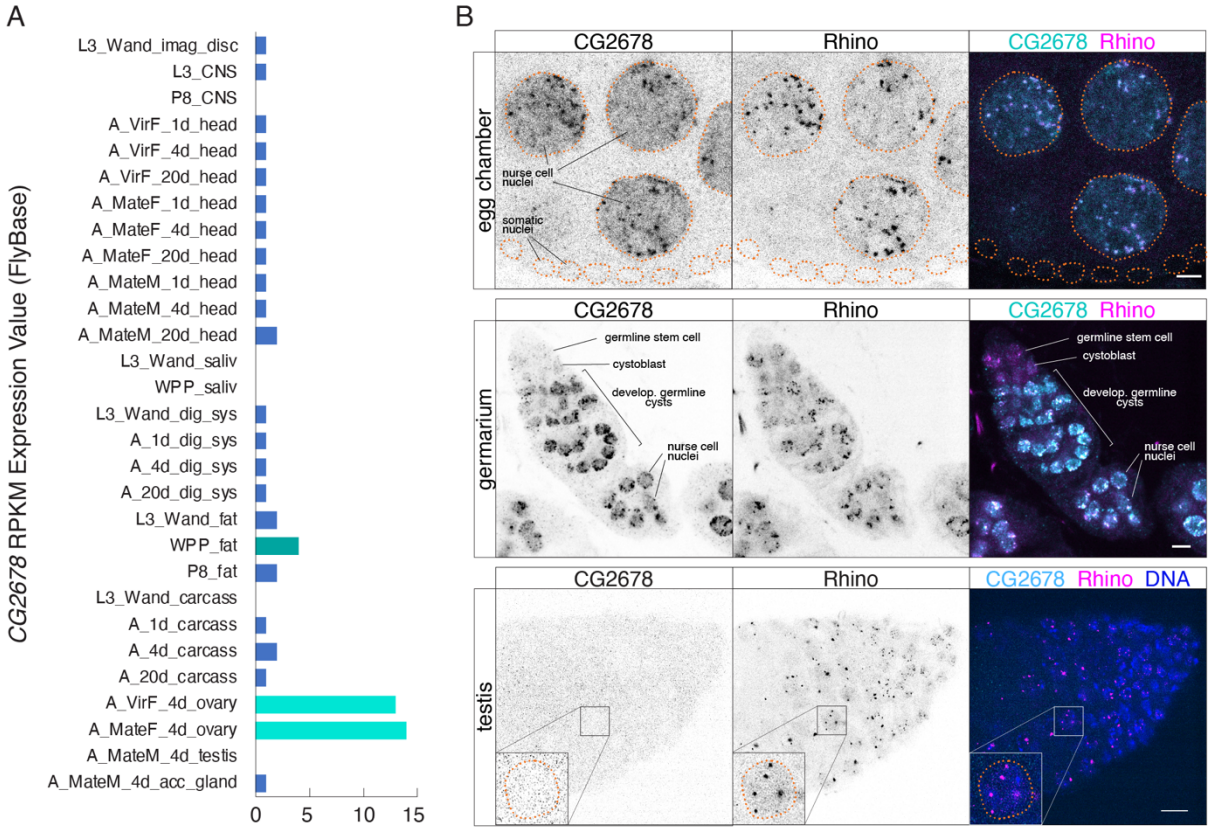


Figure 25: *CG2678* is expressed in ovaries in a non-uniform pattern. (A) Bar graph indicating the expression pattern of *CG2678* (Flybase modENCODE Anatomy RNA-Seq; L: larva; P: pupa; A: adult; M: male; F: female). (B) Confocal images of nurse cell nuclei (top), germarium (middle) and testes (bottom). Single channel and merged color images depict immunofluorescence signal for endogenous CG2678 (left, cyan) and Rhino (middle, magenta). Scale bar: 5 µm, dotted line: nuclear outline based on DAPI.

Similar to *rhino*, *CG2678* mRNA levels were detected primarily in ovaries (Flybase gene expression atlas, Figure 25A, (Larkin et al., 2021)). Immuno-fluorescence experiments showed that CG2678 was expressed specifically in germline cells, and showed no detectable signal above background in somatic cells which lack Rhino-defined piRNA clusters (Figure 25B). In germline nurse cells, CG2678 was enriched in nuclei and accumulated in numerous discrete foci in a pattern indistinguishable from Rhino (Figure 25B top). However, we noticed that the expression of CG2678 was not uniform within ovarioles: CG2678 protein was barely detectable in germline stem cells and up to germarium region 2a despite strong Rhino signal in these cells (Figure 25B middle). This is supported by single cell RNA-seq data indicating low mRNA levels in germline stem cells and strongly increased levels in differentiating nurse cells (Rust et al., 2020). CG2678 was not detectable in testes where Rhino is also expressed (Figure 25A, B bottom). The differential expression of Kipferl would be consistent with a function in determination of Rhino's chromatin binding specificity, which has recently been described to differ between ovaries and testes for yet unknown reasons (Chen et al., 2021).

To determine whether the co-localization of CG2678 and Rhino in nuclear foci is due to the co-occurrence at the same chromatin sites, we performed ChIP-seq experiments for CG2678 in wildtype ovaries. This revealed that Rhino and CG2678 co-occupy the same sites genome-wide with 99,6% of all Rhino-enriched 1-kb tiles also being enriched in CG2678 in w^{1118} ovaries (Figure 26A). We obtained largely identical results for the *iso1* and MTD-Gal4 genotype

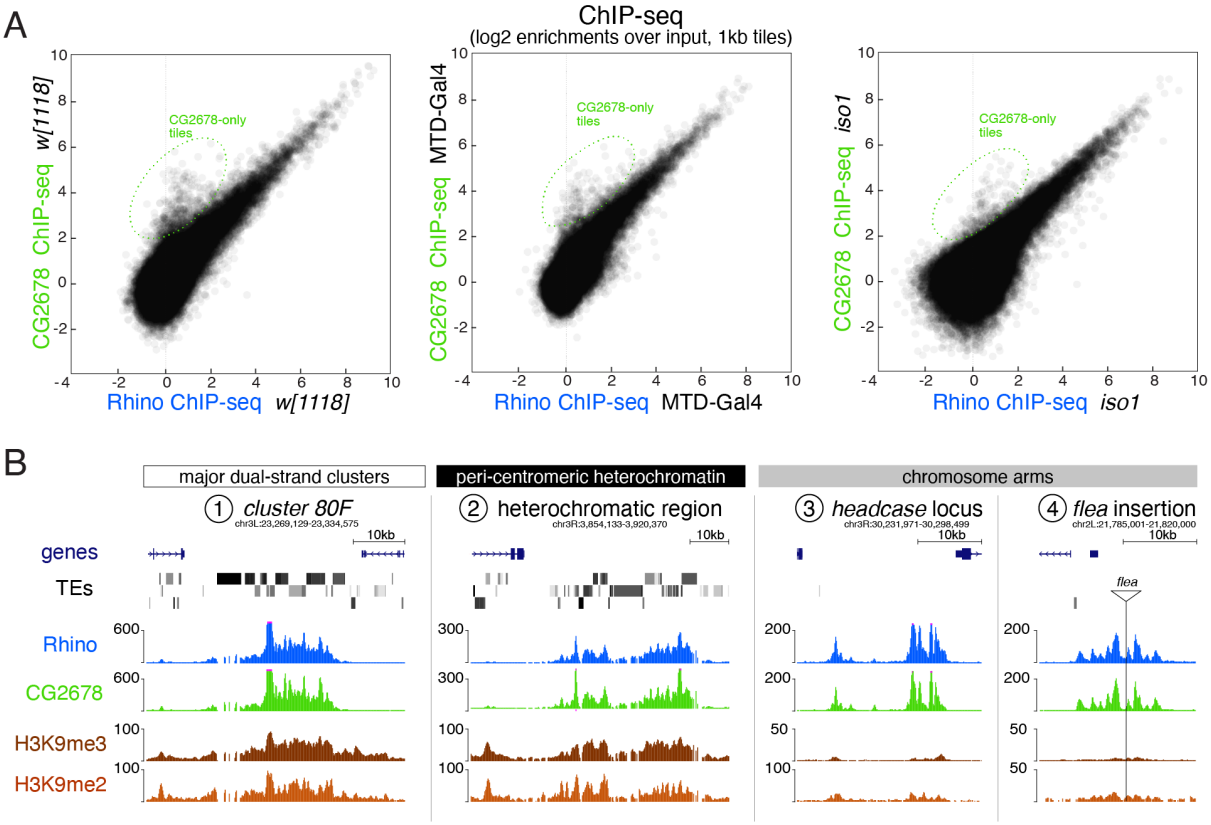


Figure 26: CG2678 mirrors Rhino's chromatin occupancy at all sites. (A) Scatter plots depicting the correlation of log2-fold Rhino- versus CG2678 ChIP-seq enrichments in w^{1118} (n=2), MTD-Gal4 (n=2) or *iso1* ovaries (n=1). (B) USCS genome browser tracks depicting indicated ChIP-seq signal (coverage per million sequenced reads for one representative replicate) at diverse Rhino domains.

backgrounds (Figure 26A). Closer inspection of example regions revealed that CG2678 and Rhino bind to chromatin in a virtually indistinguishable local pattern (Figure 26B). This was true at major piRNA clusters, Rhino domains in pericentromeric heterochromatin, and Rhino domains in euchromatic chromosome arms irrespective of their TE content (Figure 26B).

We find that 366 genomic 1-kb tiles showed enrichment for CG2678 but not for Rhino. At these sites, CG2678 is found in narrow peaks (Figure 27A). The overwhelming majority of these CG2678-only loci were located within the euchromatic chromosome arms and free of H3K9 methylation (Figure 27B). We speculate that a lack of H3K9me2/3 makes these sites unsuitable for Rhino binding. Overall, our experiments identified CG2678 as molecular Rhino cofactor and we suggest that the potentially sequence-specific DNA binding protein represents a promising candidate for a Rhino specificity factor.

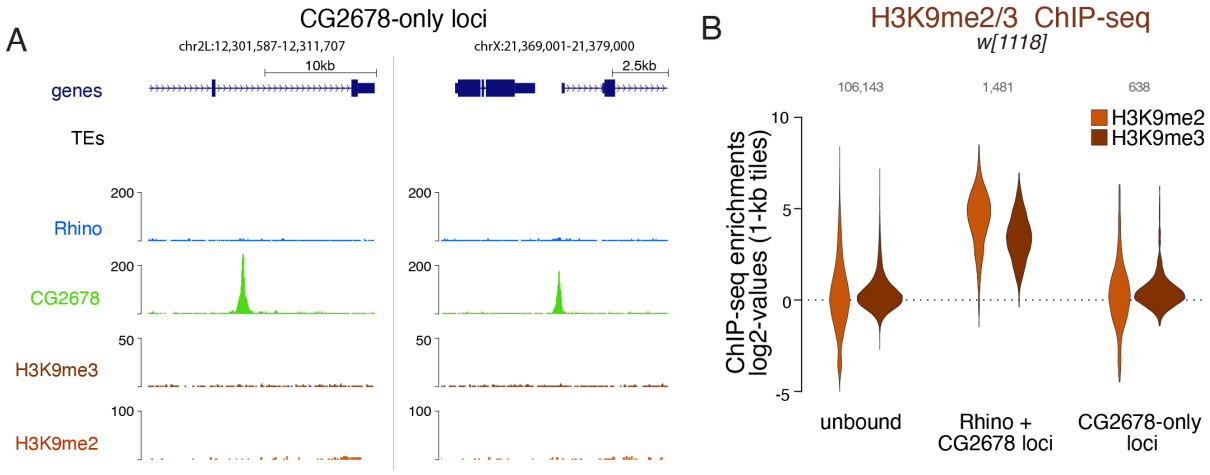


Figure 27: CG2678 forms stand-alone peaks at sites lacking H3K9 methylation. (A) UCSC browser tracks illustrating CG2678 signal at CG2678-only peaks. ChIP-seq signal is shown as coverage per million sequenced reads for w^{1118} ovaries for one representative replicate. (B) Violin plots depicting log2-fold enrichment of H3K9me2 (orange, n=1) and H3K9me3 (brown, n=2) at euchromatic 1-kb tiles bound by neither Rhino nor CG2678, both proteins, or CG2678 only. Classification into groups was performed based on binary cutoffs for Rhino (4-fold) and a linear fit for CG2678 co-occupancy in two independent replicate ChIP-seq experiments from w^{1118} ovaries to extract CG2678-only tiles highlighted in (E).

2.3.1 Kipferl loss of function leads to redistribution of Rhino

To determine the impact of CG2678 on Rhino localization we generated mutant fly lines carrying a frameshift allele or a complete deletion of the *CG2678* locus. No CG2678 protein was detectable by western blot for *CG2678* null mutant flies which were viable and displayed normal ovary morphology (Figure 28A). *CG2678* mutant females showed normal egg-laying rates but reduced fertility, which further dropped in older females (Figure 28B). The deleted locus was equipped with flanking *attP* sites, allowing the insertion of modified rescue constructs into the endogenous context via recombination-mediated cassette exchange (see Figure 24 for location of *attP* sites, (Bateman et al., 2006)). Insertion of the CG2678 genomic

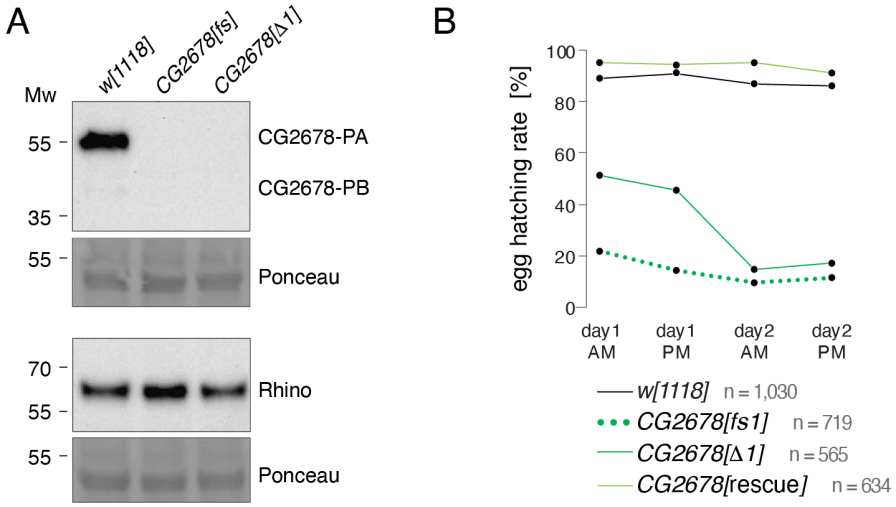


Figure 28: CG2678 is required for full female fertility. (A) Western blot analysis verifying *CG2678* frame shift (fs1) and locus deletion (Δ1) alleles using a monoclonal antibody against CG2678 (top; CG2678-PB is a minor protein isoform) and depicting Rhino levels in the absence of CG2678 (bottom). Ponceau staining: loading control. (B) Time-resolved hatching rates for eggs laid by w^{1118} control females in comparison to females carrying a *CG2678* frame shift (*fs1*), locus deletion (*Δ1*), or tagged rescue construct instead of the CG2678 locus, respectively (AM, PM indicates egg laying time). Total number of eggs laid is indicated for each genotype.

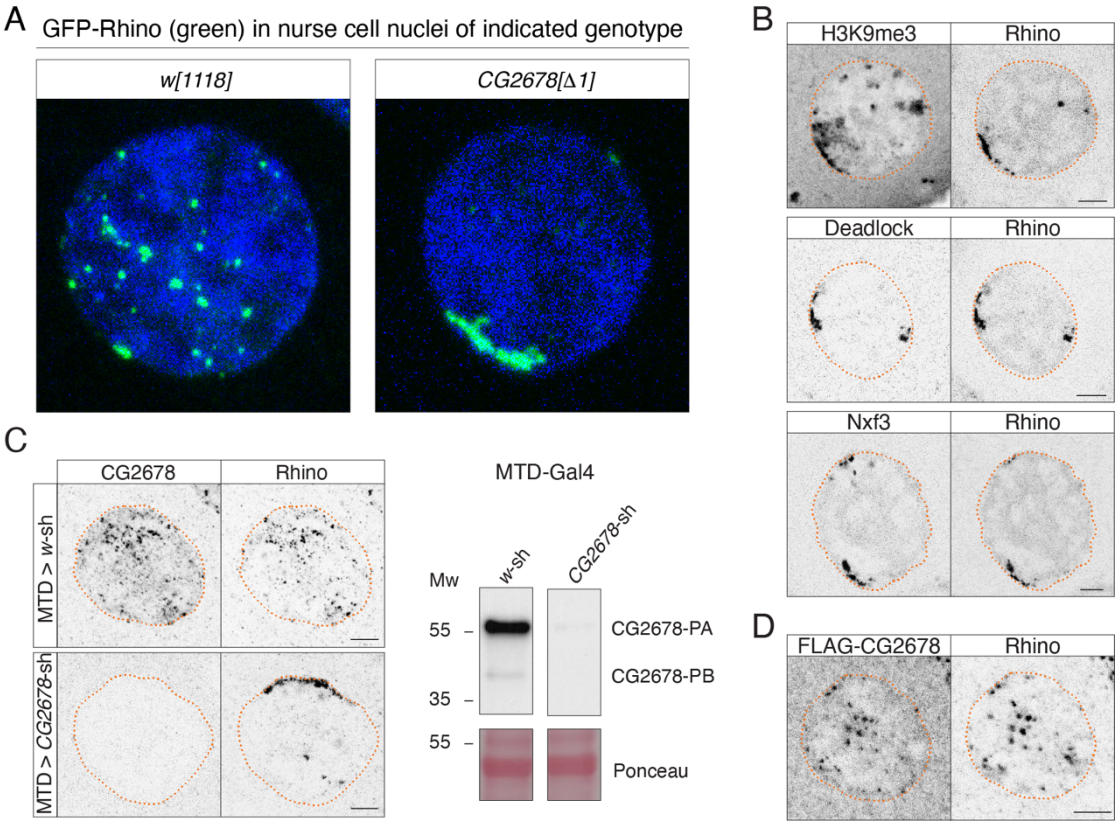


Figure 29: Rhino localizes in crescent shaped accumulations at the nuclear envelope in *CG2678/kipferl* mutants. (A) Confocal images illustrating localization of GFP-Rhino in nurse cell nuclei of w^{1118} and *CG2678* locus deletion (*Δ1*) females (scale bar: 5 µm). (B) Confocal images showing the localization of GFP-Rhino with H3K9me3 (top), Deadlock (middle), or Nxf3 (bottom) in CG2678 null mutant nurse cells (scale bar: 5 µm). (C) Confocal images showing immunofluorescence signal for CG2678 and Rhino in nurse cells of MTD-Gal4 driven control or *CG2678* knock down ovaries (scale bar: 5 µm) and Western blots showing CG2678 or Rhino protein levels in ovarian lysate from indicated genotypes (Ponceau staining served as loading control). Changes in protein levels and isoform ratios between control and rhino depleted ovaries were not reproducible and are likely caused by differences in ovary morphology, which distort the protein composition across samples.

sequence with an internal FLAG-tag (between ZAD and first ZnF array; Figure 24) at the deleted locus restored fertility to wildtype levels (Figure 28B).

Loss of CG2678 strongly affected the Rhino localization pattern: while Rhino accumulated in many distinct foci throughout nurse cell nuclei in wildtype, Rhino signal accumulated in few large, continuous structures at the nuclear periphery in *CG2678* mutants (Figure 29A). These accumulations overlapped with dense DAPI signal and were enriched in several piRNA cluster factors, as well as H3K9me3, indicating that they were functional chromatin-bound Rhino domains (Figure 29B). Germline-specific RNAi-mediated depletion of CG2678 caused a similar phenotype, and expression of tagged wildtype CG2678 protein restored wildtype Rhino foci in *CG2678* deletion flies (Figure 29C, D). In reference to the peculiar crescent-shaped Rhino accumulations in *CG2678* mutant flies resembling a popular Austrian pastry, we named *CG2678 'kipferl*'.

To distinguish if the altered nuclear localization of Rhino was caused by a redistribution of Rhino onto chromatin regions naturally accumulated at nuclear periphery, or by a mere reorganization of nurse cell chromatin in three-dimensional space, we performed ChIP-seq

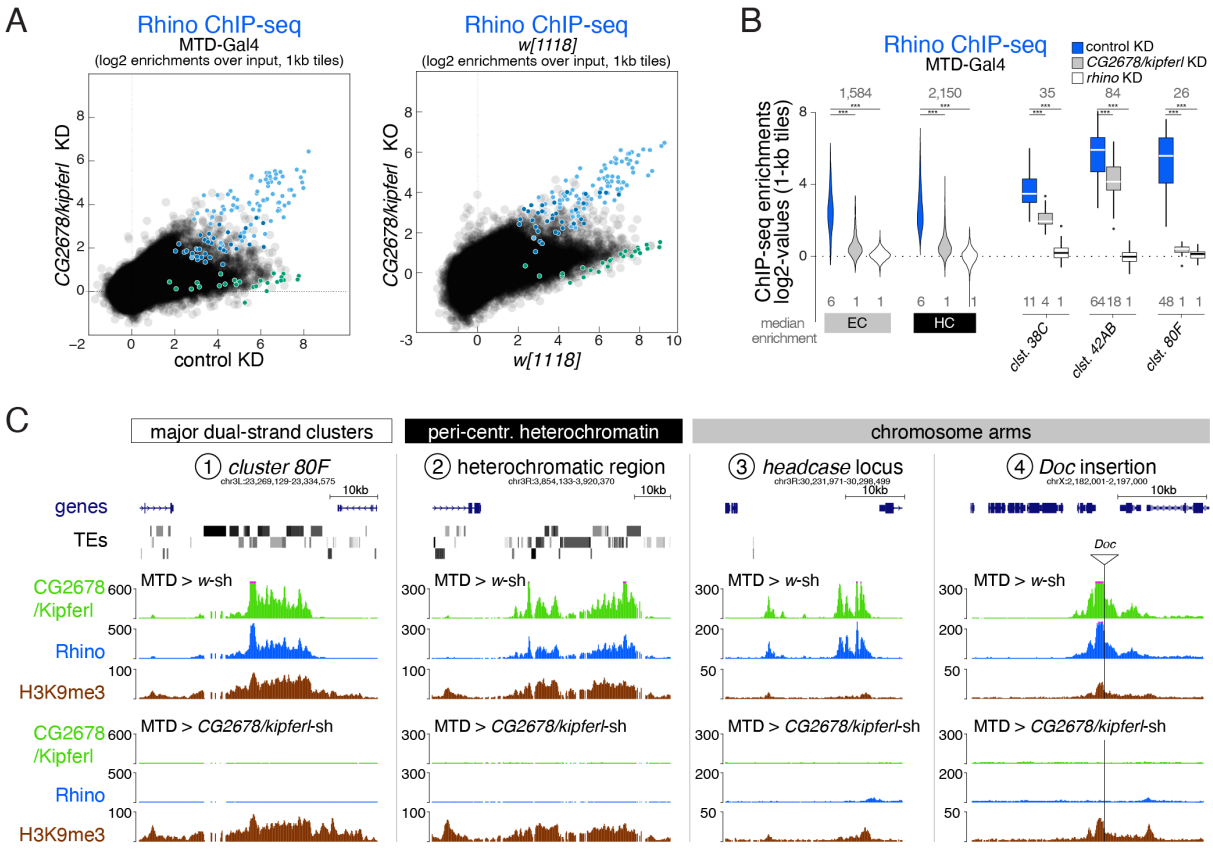


Figure 30: Rhino is lost from most of its genomic binding sites in the absence of Kipferl. (A) Scatter plot of genomic 1-kb tiles contrasting average log2-fold Rhino ChIP-seq enrichment in ovaries with MTD-Gal4 driven *CG2678/kipferl* knock down versus control ovaries (left) or full *CG2678/kipferl* mutants versus w^{1118} (right) (average of two replicate experiments each). (B) Violin plots showing average log2-fold Rhino ChIP-seq enrichment in control (n=3) as well as *CG2678/kipferl* (n=2) or *rhino* (n=1) germline knock down ovaries on Rhino-bound 1-kb tiles (defined in Fig.1D) in heterochromatin (HC) and chromosome arms (EC). piRNA clusters *38C, 42AB,* and *80F* are depicted separately. *** corresponds to p < 0,001 based on student's t-test. (C) UCSC browser tracks (ChIP-seq) depicting diverse Rhino domains in control and *CG2678/kipferl* germline knock down ovaries (signal shown as coverage per million sequenced reads for one representative replicate).

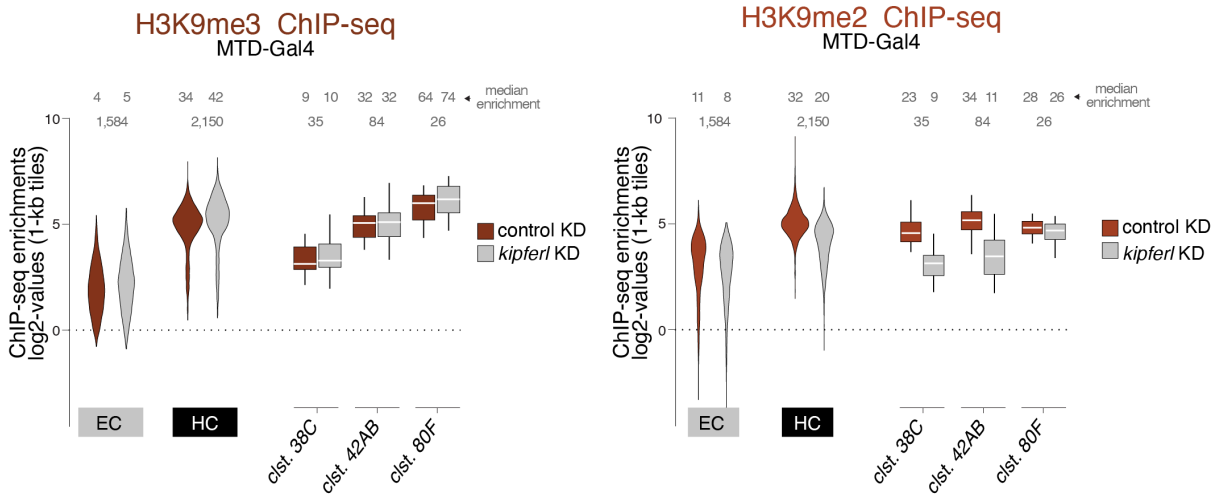


Figure 31: H3K9me2/3 levels are not perturbed in the absence of Kipferl. Violin plots showing average log2-fold H3K9me3 (left) and H3K9me2 (right) ChIP-seq enrichment in control and *CG2678/kipferl* germline knock down for Rhino-bound 1-kb tiles (defined in Fig.1D) in heterochromatin (HC) and along chromosome arms (EC). piRNA clusters *38C, 42AB,* and *80F* are depicted separately as box plots.

for Rhino in CG2678 depleted ovaries. In the absence of Kipferl, Rhino enrichment was reduced to background levels at most genomic sites. The only exceptions were piRNA clusters *42AB* and *38C*, as well as parts of telomere flanking regions (Figure 30A, B). All other Rhino domains were lost, including cluster *80F* as well as virtually all domains in pericentromeric heterochromatin and along chromosome arms, irrespective of TE association or domain size (Figure 30C). We observed largely overlapping effects in *kipferl* mutant ovaries as we did upon RNAi-mediated Kipferl depletion (Figure 30A). Importantly, while Rhino was lost we did not observe reduced H3K9 methylation levels, indicating that, despite continued presence of underlying H3K9me2/3, Rhino is not able to stably associate with chromatin in the absence of Kipferl (Figure 30C, Figure 31).

Remaining Rhino levels at clusters *42AB* and *38C* were, nevertheless, strongly reduced in comparison to wildtype ovaries. To determine if these low remaining levels of Rhino could account for the strong accumulations visible in nurse cell nuclei, we performed RNA

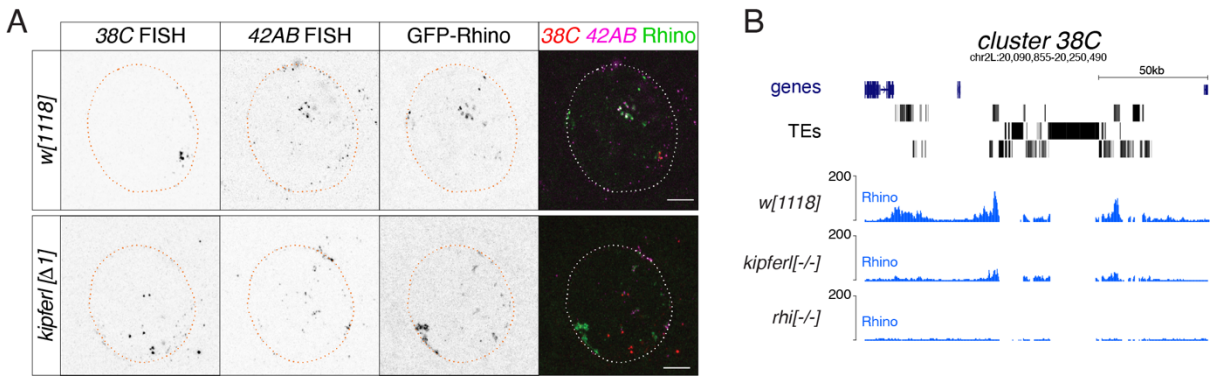


Figure 32: RNA FISH for piRNA clusters *42AB* and *38C* does not overlap with large Rhino accumulations in *kipferl* mutant ovaries. (A) Confocal images showing the localization of piRNA cluster RNA via RNA FISH in respect to GFP-Rhino in nurse cells of w^{1118} or *CG2678/kipferl* locus deletion flies (scale bar: 5 µm). Imaging conditions were chosen to avoid photo bleaching and to optimally display the large Rhino-GFP accumulations to show that these are not corresponding to either of these clusters. (B) UCSC browser tracks depicting Rhino ChIP-seq signal at piRNA cluster 38C in the indicated genotypes (signal shown as coverage per million sequenced reads for one representative replicate).

fluorescence in situ hybridization (FISH) experiments. These confirmed that both clusters are still transcribed in *kipferl* mutants. However, the nuclear RNA-FISH signal of these two clusters did not show increased signal as compared to wildtype, and did not co-localize with the strong, crescent-shaped Rhino accumulations at the nuclear envelope (Figure 32A). In line with low residual Rhino levels especially at cluster *38C*, the overlap with GFP signal was minimal for this piRNA cluster (Figure 32B). We therefore speculated that the prominent Rhino accumulations in *kipferl* mutants must correspond to repetitive loci, not included in our tile-based analysis of genome unique reads.

Analysis of all ChIP-seq reads revealed that Rhino enrichment on transposon consensus sequences was strongly decreased upon Kipferl depletion (Figure 33A). In contrast, Rhino occupancy was maintained on several Satellite sequences, foremost the Responder and *1.688* family Satellites which had been previously described as sources of Rhino-dependent piRNAs (Figure 33B) (Wei et al., 2021, Usakin et al., 2007, Chen et al., 2021). In agreement with increased Rhino occupancy, the nuclear RNA FISH signal for both Satellites was strongly increased in *kipferl* mutant ovaries (Figure 33C). Importantly, the FISH signal detected for *Rsp* and *1.688*, presumably corresponding to nascent transcripts, precisely overlapped with the crescent-shaped Rhino accumulations at the nuclear periphery. It is conceivable that the vast number of Satellite repeats, in particular the Mega-basepair long 359bp array on the X chromosome, explain the elongated, continuous structure of Rhino domains in Kipferl-deficient nurse cell nuclei.

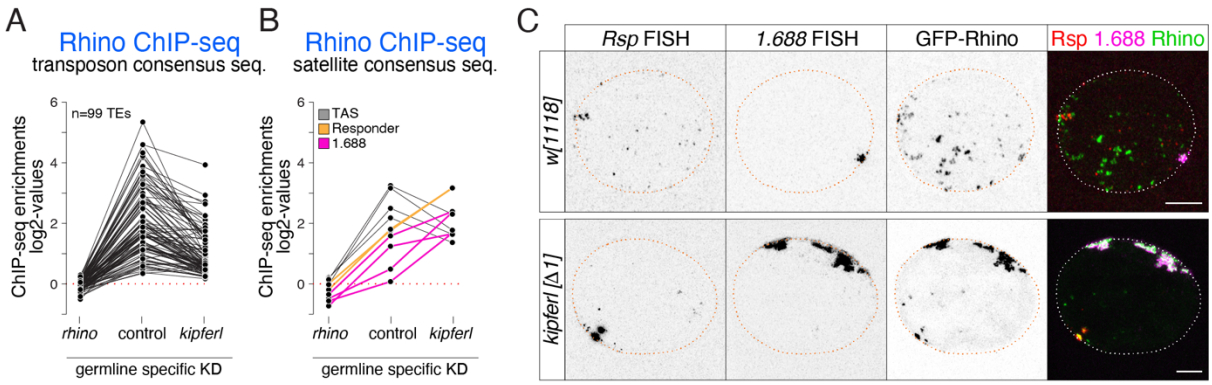


Figure 33: Rhino accumulates at Satellite repeats in the absence of Kipferl. Jitter plots depicting the log2-fold Rhino ChIP-seq enrichments on transposon (A) and Satellite (B) consensus sequences in indicated genetic backgrounds. (C) Confocal images showing *Rsp* and *1.688* Satellite RNA FISH signal and GFP-Rhino in nurse cells of w^{1118} or *CG2678/kipferl* mutant flies (scale bar: 5 µm).

With the chromosomal content of the large Rhino accumulations identified, we were still intrigued by the strong accumulation at the nuclear envelope. Rhino domains have previously been described to preferentially localize to the nuclear envelope also in wildtype ovaries (ElMaghraby et al., 2019, Zhang et al., 2012b). This is thought to result in a direct precursor export route from piRNA clusters to nuage (Zhang et al., 2012a). We hypothesized that the piRNA precursor export via Nxf3 might exert a pulling force on transcripts still associated with chromatin of origin. Indeed, in *kipferl,nxf3* double mutant flies, Rhino still accumulated in

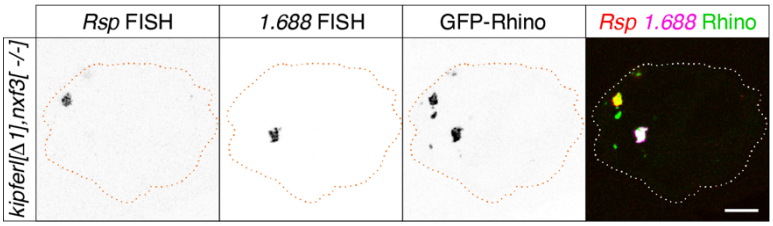


Figure 34: Nxf3 is responsible for the localization of Rhino accumulations to the nuclear envelope in *kipferl* mutant nurse cells. Confocal images showing *Rsp* and *1.688* Satellite RNA FISH signal in respect to GFP-Rhino in nurse cells of *CG2678/kipferl,nxf3* double mutant flies (scale bar: 5 µm).

large foci on Satellite repeats, but these now formed condensed structures that were no longer confined to the nuclear envelope (Figure 34).

Collectively these data indicate that Kipferl is required for Rhino to stably associate with the vast majority of its chromosomal binding sites. The loss of Kipferl leads to a large-scale redistribution of Rhino to Satellite repeats, which accumulate at the nuclear envelope due to piRNA precursor export-mediated forces, explaining the name-giving Rhino phenotype in *kipferl* mutants.

2.3.2 The ovarian piRNA pool changes drastically in the absence of Kipferl

Rhino-dependent piRNA production accounts for close to 80% of ovarian piRNAs in wildtype ovaries (Figure 35A). Other major contributors are the somatic piRNA source loci (eg. *flamenco*), as well as Rhino-independent germline source loci (eg. cluster *20A*). Depletion of Rhino leads to a collapse of piRNA production from Rhino-dependent source loci, while piRNA production at Rhino-independent piRNA source loci in the germline and ovarian soma are unaffected (Figure 35B). The large-scale rearrangements of Rhino on chromatin in the absence of Kipferl were also reflected in the piRNA pool in Kipferl-depleted ovaries: piRNA levels were reduced specifically at nearly all Rhino-dependent piRNA source loci, with only the major Rhino-dependent piRNA clusters *38C* and *42AB* left unaffected (Figure 35C). This was in line with these clusters retaining Rhino binding in the absence of Kipferl. At cluster *80F*, which depended entirely on Kipferl for Rhino binding, piRNA production collapsed. No notable changes were observed for piRNA levels at somatic or Rhino-independent germline source loci. Overall, this resulted in a reduction of miRNA-normalized piRNAs to 65% of wildtype counts in Kipferl-depleted ovaries (Figure 35A).

The specific impact of Kipferl depletion on Rhino-dependent piRNA production was also apparent when all piRNAs mapping in antisense to transposon consensus sequences were analyzed (Figure 36A). Transposons under control of somatic piRNAs (e.g. *Tabor, gypsy5, ZAM*; mostly originating from *flamenco*) were not affected by either Rhino or Kipferl depletion. For transposons controlled in the germline, piRNA levels were largely maintained for elements silenced through Rhino-independent piRNAs (e.g. *297, roo*) also upon Kipferl-depletion. Among the transposons controlled by Rhino-dependent piRNAs, we observed a differential impact of Kipferl-depletion: few transposons showed retained or increased piRNA

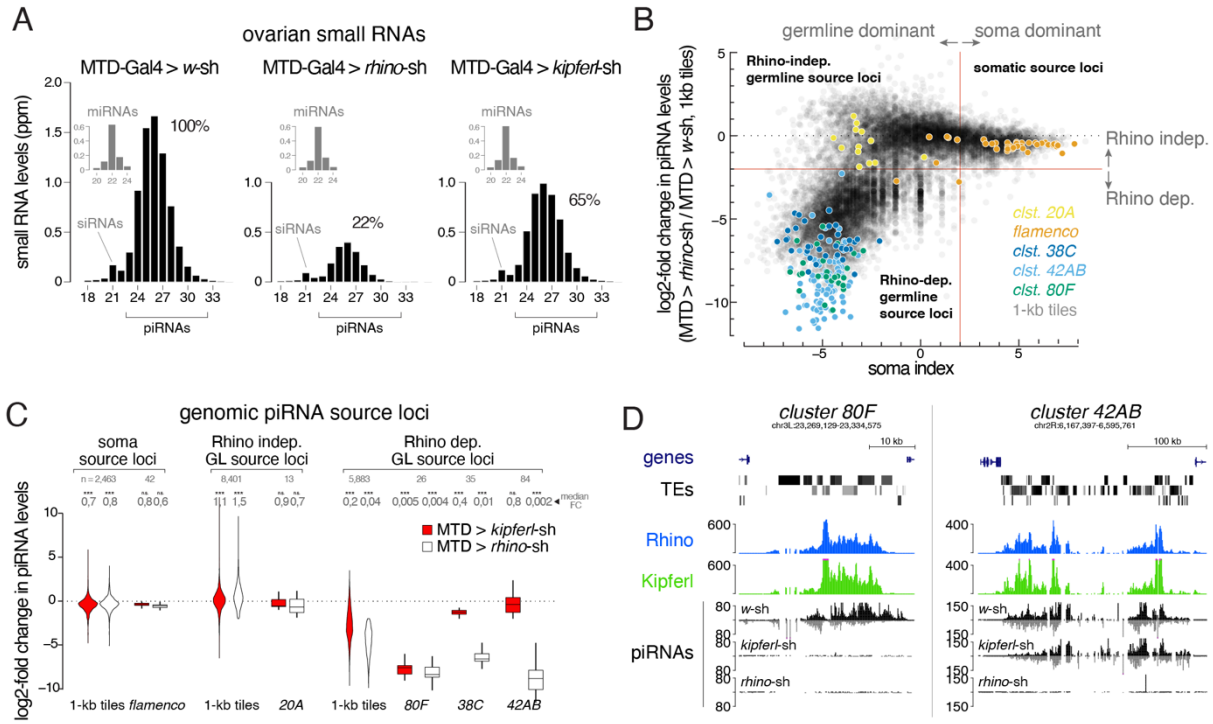


Figure 35: Rhino-dependent piRNA sources show differential dependence on Kipferl. (A) Size profiles of Argonaute-bound small RNAs (miRNA-normalized) isolated from ovaries of indicated genotypes. microRNA reads are shown separately (grey bars). Percentages represent total piRNA levels (23-33nt) relative to control (100%). (B) Classification of genomic 1-kb tiles into somatic (e.g. *flamenco*), Rhino-independent (e.g. cluster *20A*) and Rhino-dependent germline source loci (e.g. clusters *38C, 42AB, 80F*). Binary cutoffs are applied at a soma index of 2 do distinguish somatic from germline source loci and a 4-fold reduction in piRNAs upon rhino depletion to determine dependency on Rhino. (C) Violin plots showing log2-fold changes in levels of uniquely mapping piRNAs on 1-kb tiles relative to control upon MTD-Gal4 mediated knock down of *rhino* or *kipferl* (1-kb tiles were categorized into somatic source loci, Rhino-independent germline source loci and Rhino-dependent germline source loci according to Fig. S4B, n=1). *** and n.s. corresponds to $p < 0{,}001$ or $p > 0{,}05$, respectively, based on Wilcoxon signed-rank test. (D) UCSC genome browser tracks displaying piRNA levels at cluster *80F* and *42AB* in control, *kipferl*, and *rhino* knock down ovaries (ChIP-seq signal from MTD-Gal4 control ovaries is depicted as coverage per million reads, piRNA coverage was normalized to miRNA reads, data is given for one replicate).

levels in the absence of Kipferl (Figure 36B, C). These elements were represented in Kipferl-independent piRNA clusters *42AB* and *38C*, explaining their high remaining antisense piRNA counts.

Despite a strong effect on piRNA levels, unlike Rhino, depletion of Kipferl did not lead to a complete loss of piRNAs mapping in antisense to most transposons (Figure 36B). In line with incomplete loss of Rhino-dependent piRNAs in Kipferl-depleted ovaries, RNA-seq experiments showed that only few transposons were upregulated upon Kipferl germline knock down (Figure 36D). RNA FISH experiments confirmed that the same, relatively small subset of TEs was upregulated also in *kipferl* mutant ovaries (Figure 36E). We note that several elements with similar losses of antisense piRNAs were not derepressed and that transposon de-repression in ovaries lacking Kipferl was less severe compared to ovaries lacking Rhino (Mohn et al., 2014). This might be explained by the milder loss of piRNAs in *kipferl* mutants compared to *rhino* mutants, leaving selected piRNA source loci intact. Alternatively, the onset of Kipferl expression only upon germline cell differentiation might leave a window of Rhino-mediated, but Kipferl-independent transposon silencing in early developmental stages, which might be sufficient to repress certain elements.

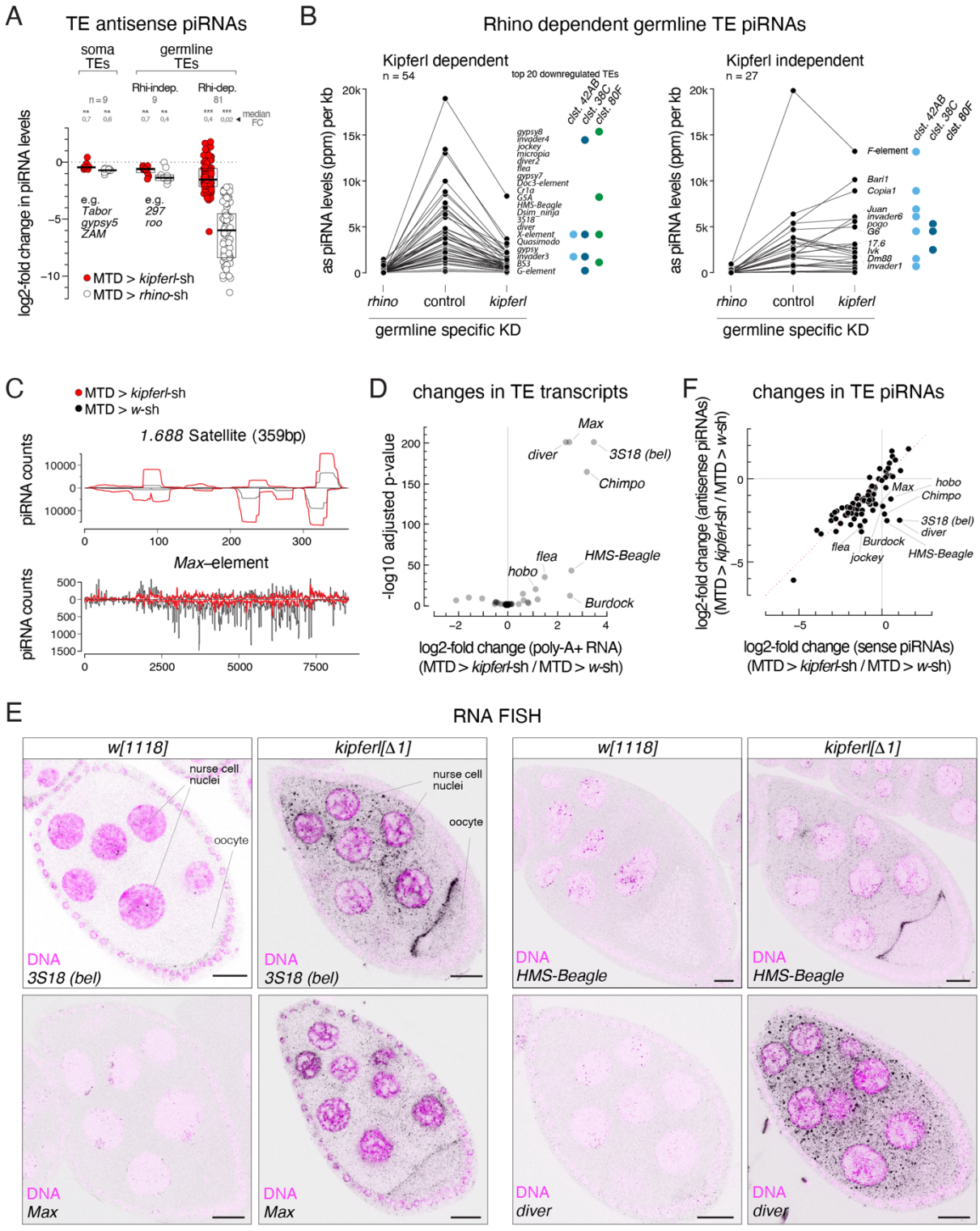


Figure 36: Loss of Kipferl differentially impacts the regulation of Rhino-dependent transposons. (A) Jitter plot depicting log2-fold changes for piRNA levels mapping antisense to transposon consensus sequences in indicated MTD-Gal4 mediated knock downs compared to control (transposons classified analogous to panel A, n=1). *** and n.s. corresponds to $p < 0{,}001$ or $p > 0{,}05$, respectively, based on Wilcoxon signed-rank test. (B) Jitter plots showing piRNA levels (per kb sequence) in indicated genotypes mapping to transposons (antisense only) giving rise to Rhino-dependent piRNAs. Blue and green dots in panel D indicate fragments of the respective transposon in piRNA clusters *38C, 42AB*, or *80F* (n=1). (C) piRNA profiles across the consensus sequences for a representative transposon (*Max*) and the 359bp *1.688* Satellite. piRNA counts (normalized to miRNAs) are displayed for indicated genotypes. (D) Volcano plot depicting the log2-fold changes in poly-adenylated transposon transcripts in *kipferl* depleted versus control ovaries (n=3). (E) Confocal images showing RNA FISH signal for indicated transposon transcripts in w^{1118} and *kipferl* null mutant ovaries (scale bar: 20 µm). (F) Scatter plot contrasting fold change of antisense versus sense piRNAs mapping to transposon consensus sequences.

In the *Drosophila* germline, the ping-pong amplification cycle induces piRNA production from mRNAs targeted by antisense piRNAs. Thereby, transcripts of de-regulated transposons targeted by residual piRNAs give rise to abundant piRNAs mapping to the sense strand of their consensus sequence. Sequencing of Argonaute-bound piRNAs nicely demonstrates that precisely the few elements upregulated at the level of poly-adenylated transcripts gave rise to strongly increased amounts of sense piRNAs in Kipferl-depleted ovaries (Figure 36F). For other Rhino-dependent elements, levels of sense and antisense piRNAs were highly correlated, likely owing to the bidirectional piRNA production at Rhino-dependent piRNA clusters.

While Rhino is lost from most of its canonical binding sites in the absence of Kipferl, it relocalizes strongly to Satellite repeats of the *Rsp* and *1.688* families (Figure 33). As a consequence, the levels of piRNAs mapping to these repeats increased strongly in Kipferl-depleted ovaries compared to control (Figure 37). More than 20% of all piRNAs in *kipferl* deficient ovaries mapped to Satellites, while this was the case for only 4% of all piRNAs in wildtype ovaries. Together, these data demonstrate that Kipferl is required to maintain a balanced and functional piRNA pool in *Drosophila* ovaries that allow the efficient silencing of a diverse set of transposons.

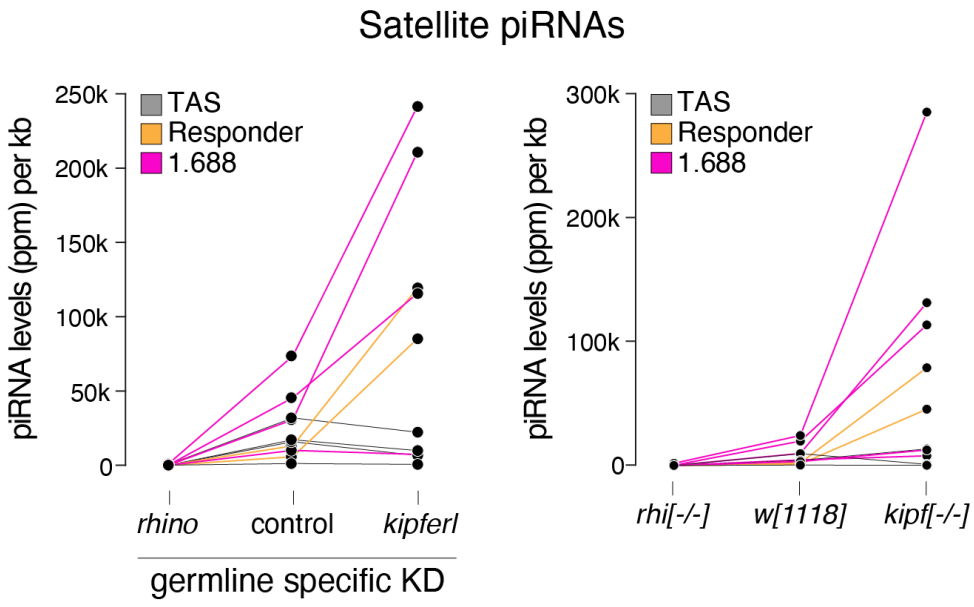


Figure 37: piRNA production from Satellite repeats increases drastically in the absence of Kipferl. Jitter plots showing piRNA levels (per kb sequence) in indicated genotypes mapping to Satellite repeats.

Similar to Rhino, depletion of Kipferl did not lead to expression changes for ovarian mRNAs, establishing Kipferl as a piRNA specific factor defining the genomic distribution of Rhino's chromatin binding profile (Figure 38).

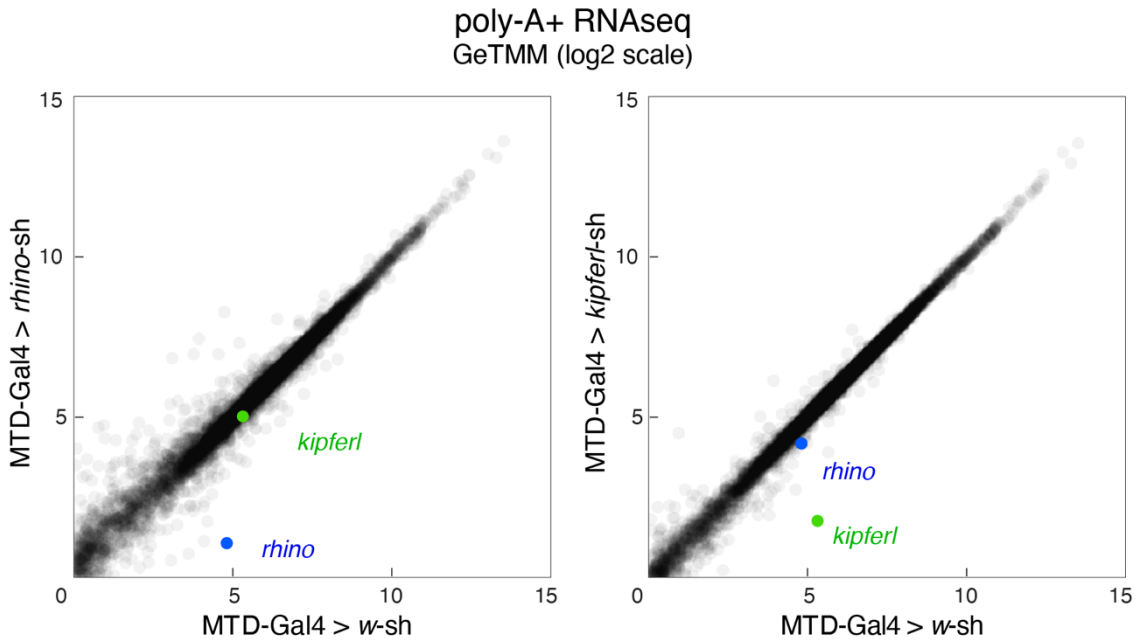


Figure 38: Depletion of Kipferl does not lead to gene expression changes. Scatter plots contrasting RNAseq signal (poly-A plus) for expressed genes in indicated genotypes. Values are displayed as Gene length corrected trimmed mean of M-values (GeTMM; Blue dot: *rhino*. Green dot: *kipferl*. n=3)

2.3.3 Kipferl binds to chromatin upstream of Rhino

Given the genomic redistribution of Rhino in the absence of Kipferl, together with the putative DNA-binding ability of Kipferl via its ZnFs, we hypothesized that Kipferl might act as sequence-specific DNA binder that functions upstream of Rhino. As such, Kipferl would mediate the recruitment and/or stabilization of Rhino on chromatin at its binding sites located within an H3K9me2/3 domain. This would require Kipferl's ability to bind to chromatin in the absence of Rhino.

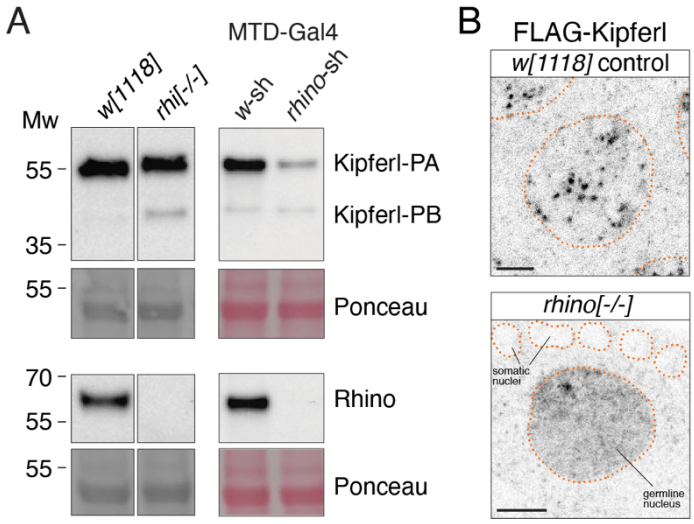


Figure 39: Kipferl localizes diffusely in nurse cell nuclei in the absence of Rhino. (A) Western blot image depicting the levels of Kipferl in the indicated genotypes (top) where Rhino is absent (bottom). Ponceau images serve as loading control. (B) Confocal images depicting FLAG staining for internally tagged Kipferl in the indicated background genotype. Scale bar: 5 µm. Orange lines represent nuclear outline based on DAPI.

In *rhino* mutant ovaries, Kipferl was readily detectable and localized diffusely throughout nurse cell nuclei, with only occasional enrichment in few nuclear foci (Figure 39). We observed a slight change in isoform abundance for Kipferl in *rhino* mutants compared to control, which was not reproducible and can likely be attributed to differences in ovary morphology between different genotypes. Despite the lack of an obvious localization pattern in immunofluorescence stainings, most genomic 1-kb tiles occupied by Kipferl in wildtype ovaries retained Kipferl enrichment in the absence of Rhino (Figure 40). Kipferl enrichment levels, however, were highly variable between *rhino*-depleted and control ovaries, and we observed a strong decrease at the largely Kipferl-independent piRNA clusters *42AB* and *38C*, as well as several other regions (Figure 40).

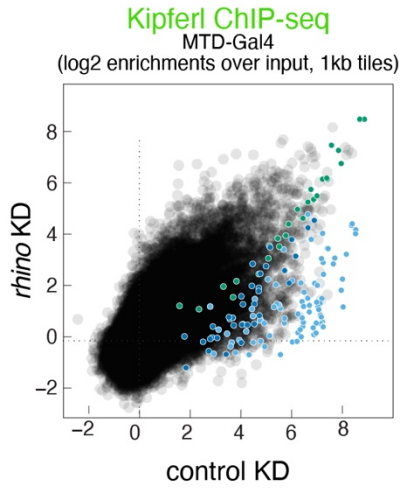


Figure 40: Kipferl remains chromatin-bound in the absence of Rhino. Scatter plot of genomic 1-kb tiles contrasting average log2-fold Kipferl ChIP-seq enrichment in ovaries with MTD-Gal4 driven *rhino* knock down (n=2) versus control ovaries (n=3).

To investigate the relationship of Kipferl and Rhino at chromatin from a different angle, we determined all continuous Kipferl-bound regions in wildtype ovaries to assess the effect of Rhino depletion on Kipferl's chromatin occupancy, and vice versa (Figure 41A). This revealed that the two proteins were highly interdependent in their chromatin association at regions they co-occupied in wildtype ovaries, with both proteins showing strongly reduced signal at these sites in the absence of the other. This included the majority of heterochromatic Kipferl binding sites, as well as roughly 60 % of Kipferl-bound regions in euchromatin. The strong dependence of Rhino's chromatin occupancy on Kipferl confirmed our previous results. Loss of Rhino, on the other hand, had a more complex effect on Kipferl's chromatin association: despite a clear reduction of the Kipferl ChIP-seq signal at sites co-occupied by Rhino, residual signal was visible at many of these loci. Manual inspection of several regions occupied by Kipferl and Rhino in wildtype ovaries confirmed that Kipferl remained bound at these regions in *rhino* mutant ovaries, but that it's binding was reduced to more defined peaks of lower intensity (Figure 41B). At Kipferl binding sites lacking Rhino enrichment in wildtype ovaries, Kipferl binding was fully maintained in the absence of Rhino, further supporting the ability of Kipferl to bind to chromatin independently of Rhino (Figure 41B).

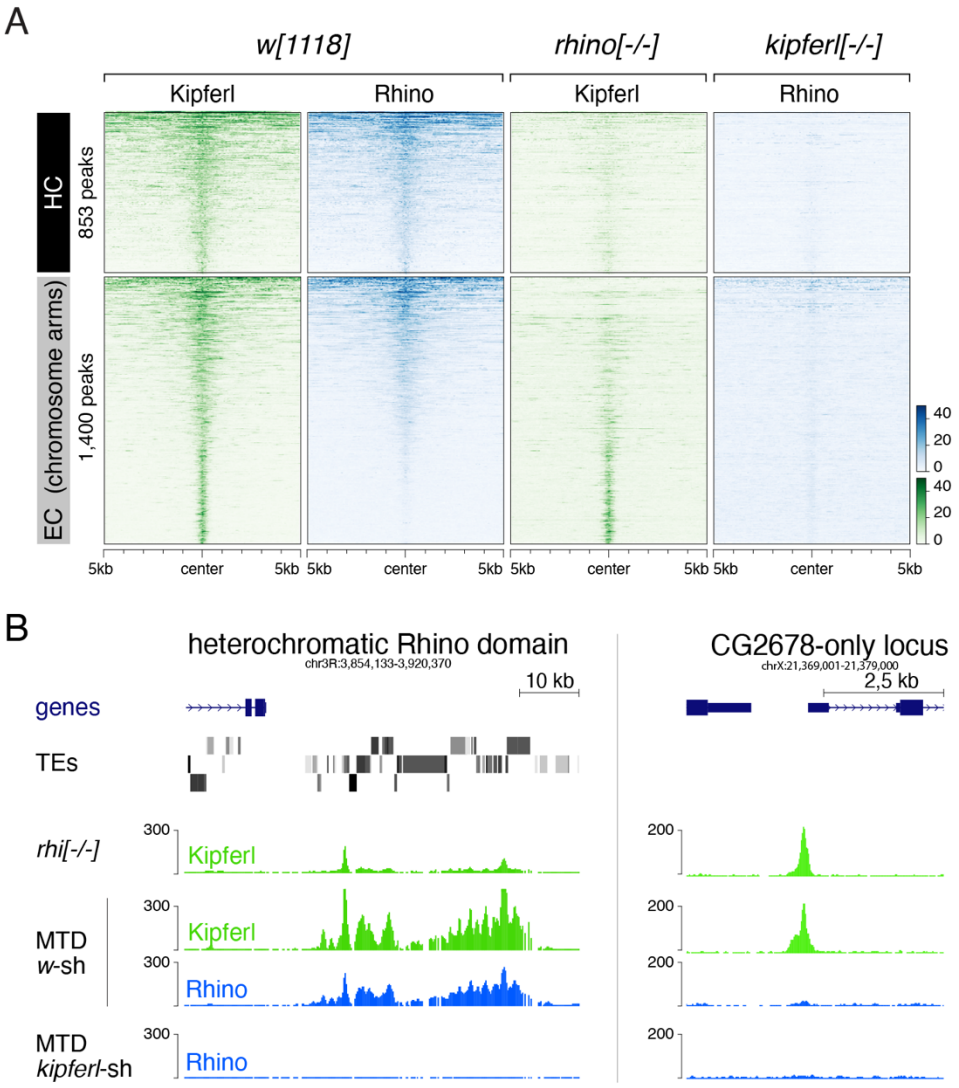


Figure 41: Rhino and Kipferl show a strong interdependence at Rhino domains. (A) Heat maps depicting ChIP-seq signal for Kipferl and Rhino in indicated genotypes from representative replicate experiments, centered on broad Kipferl peaks detected in two independent ChIP-seq experiments of w^{1118} ovaries (data sorted by Kipferl signal in w^{1118}). (B) USCS genome browser tracks depicting Kipferl and Rhino ChIP-seq signal in indicated genotypes at a Rhino domain (left) or Kipferl-only locus (right).

For a better characterization of Kipferl's intrinsic, Rhino-independent chromatin binding profile, we re-focused our analysis on Kipferl peaks determined in *rhino* mutants, where we found Kipferl enriched at thousands of narrow peaks (Figure 42). At sites that were co-occupied by Rhino in wildtype ovaries, we found that Kipferl's chromatin occupancy was reduced in *rhino* mutants compared to wildtype, while it was unaffected at Kipferl-only peaks (Figure 42A). Of note, Kipferl bound to chromatin in narrow peaks at both stand-alone Kipferl peaks and Rhino-occupied regions in *rhino* mutants, which widened into larger domains in the presence of Rhino (Figure 42A). This suggested that Rhino strengthens Kipferl binding at chromatin and supports the spreading of Kipferl to larger domains in wildtype ovaries. Indeed, Rhino-independent Kipferl binding sites were also bound by Kipferl in wildtype ovaries, and we observed strengthening and expansion of Kipferl's chromatin enrichment from the narrow peaks seen in *rhino* mutant ovaries to broad domains in wildtype, specifically at sites where

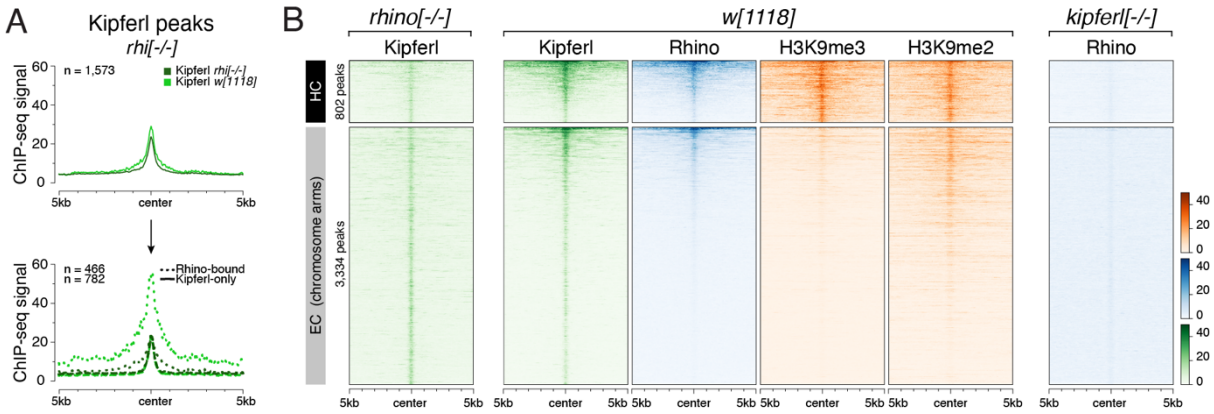


Figure 42: Kipferl binds to chromatin in thousands of narrow peaks independently of Rhino. (A) Meta profile showing Kipferl signal in wildtype and *rhino* mutant ovaries at all narrow Kipferl peaks determined in *rhino* mutant ovaries (top). The same peaks are divided into Rhino-Kipferl co-occupied (dotted line) and Kipferl-only peaks (dashed line) based on intersection with wildtype Rhino domains (bottom). (B) Heat maps depicting indicated ChIP-seq signal in indicated genotypes from representative replicate experiments centered on narrow Kipferl peaks detected in two independent ChIP-seq experiments of *rhino* mutant ovaries (data sorted by Rhino signal in w^{1118}).

Rhino was present (Figure 42B). Kipferl peaks that were not occupied by Rhino, on the other hand, did not extend into larger domains. This further demonstrated that Kipferl binds the same sites in both genotypes, yet in different patterns.

Importantly, local H3K9 methylation was required for Rhino binding at Kipferl peaks: Rhino occupied most Kipferl peaks in pericentromeric heterochromatin where H3K9me2/3 levels were high. Within chromosomal arms, Rhino accumulated selectively at Kipferl sites co-occurring with a local H3K9me2/3 domain (Figure 42B). In line with our previous observations, these findings supported a model where Kipferl binds chromatin independently of Rhino, and Kipferl binding sites act as nucleation sites for larger Rhino domains when located inside local heterochromatin. Moreover, Kipferl and Rhino cooperate and spread from nucleation sites into flanking heterochromatic regions, resulting in extended Rhino/Kipferl domains. Consistent with this, Rhino occupancy at and around Kipferl binding sites was dependent on Kipferl (Figure 42B).

We find that roughly 60% of regions that depend on Kipferl for their association with Rhino were located within less than 5 kb of a Kipferl nucleation site. At these regions, Kipferl's intrinsic chromatin binding profile in *rhino* mutants often mirrored the non-uniform enrichment of Rhino in wildtype ovaries, only at lower levels (Figure 43A). However, several Kipferl-dependent Rhino domains did not harbor prominent Kipferl peaks (e.g. the *eyeless* gene, Figure 43B). Based on the codependency of Kipferl and Rhino at these sites, we conclude that alternative guidance cues for Rhino likely exist, but that nucleation of a stable Rhino domain through these factors nevertheless requires the stabilization through Kipferl.

Finally, despite strongly associating with piRNA clusters *42AB* and *38C* in wildtype ovaries, we find only weak intrinsic Kipferl binding at these largely Kipferl-independent piRNA clusters, which is in line with their maintained Rhino binding in *kipferl* mutant ovaries (Figure 43B). This implied that Kipferl can also be stabilized by Rhino on chromatin, which likely contributes

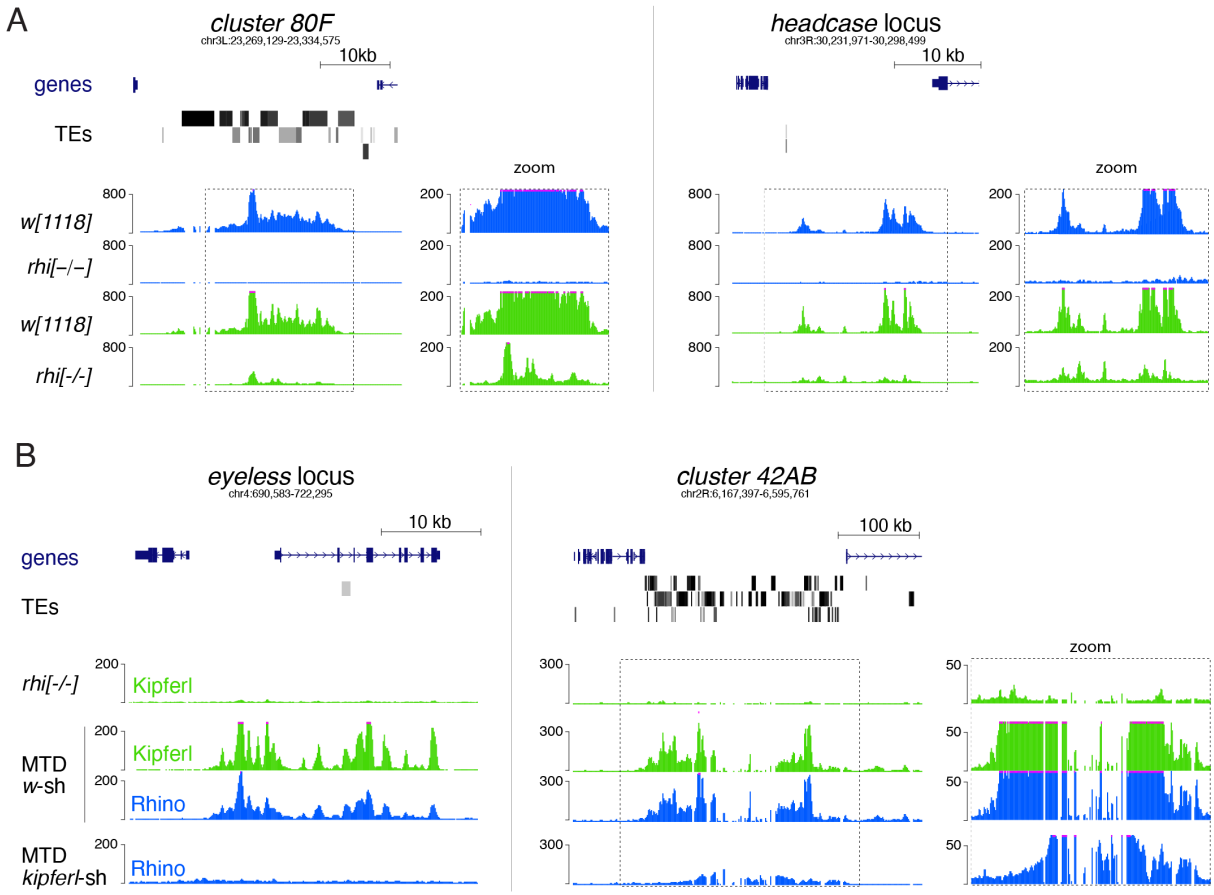


Figure 43: Rhino-independent Kipferl binding explains roughly 60% of Kipferl-dependent Rhino domains. USCS genome browser tracks depicting Rhino and Kipferl signal in the indicated genotypes at Rhino domains with (A) or without (B) distinct Kipferl nucleation sites. ChIP-seq signal is depicted as coverage per million reads.

to the two proteins' ability to spread from narrow nucleation sites into extended domains. Upon closer inspection, Kipferl influenced Rhino's chromatin profile also at the Kipferl-independent piRNA cluster *42AB*: Rhino binding in the absence of Kipferl was restricted to the central regions of the cluster. We observed several sites of small intrinsic Kipferl enrichments towards the periphery of the cluster, which were not picked up by our stringent peak calling approach. These low affinity Kipferl binding sites likely support the formation of a fully extended Rhino domain in wildtype ovaries (Figure 43B, zoom). Based on these data, we conclude that Kipferl and Rhino share a complex relationship, with both proteins supporting each other's chromatin occupancy. We find that Kipferl is a major specificity factor for Rhino in ovaries, and that Kipferl cooperates with Rhino to form extended domains from defined nucleation sites. Our data further indicate that Kipferl is not the only Rhino specificity factor but demonstrate that it is also required for the stabilization of Rhino domains nucleated by alternative means.

2.3.3.1 Kipferl binds to DNA with a preference for GRGGN-motifs

We argued that Kipferl's DNA binding pattern was likely due to the specific association with DNA motifs through its ZnF arrays, which are predicted to bind to GGA-rich motifs (Figure 44A). *De novo* discovery of enriched sequence motifs within the top 3000 Kipferl ChIP-seq peaks in *rhino* mutant ovaries recovered a motif that was present in >80% of peaks and closely matched the *in silico* predicted motif (Figure 44A). Additional motifs were identified at lower frequency and mostly corresponded to variations of the same GRGGN core motif. We find that this GRGGN motif is locally enriched at Kipferl peaks, and that higher motif counts tend to induce higher Kipferl association (Figure 44B and C). The motif was enriched at Kipferl peaks irrespective of Rhino co-occupancy, indicating that Kipferl-only peaks were not qualitatively different from Rhino-recruiting peaks, which further supported our observation that the presence or absence of local H3K9me2/3 determined whether or not Rhino was recruited.

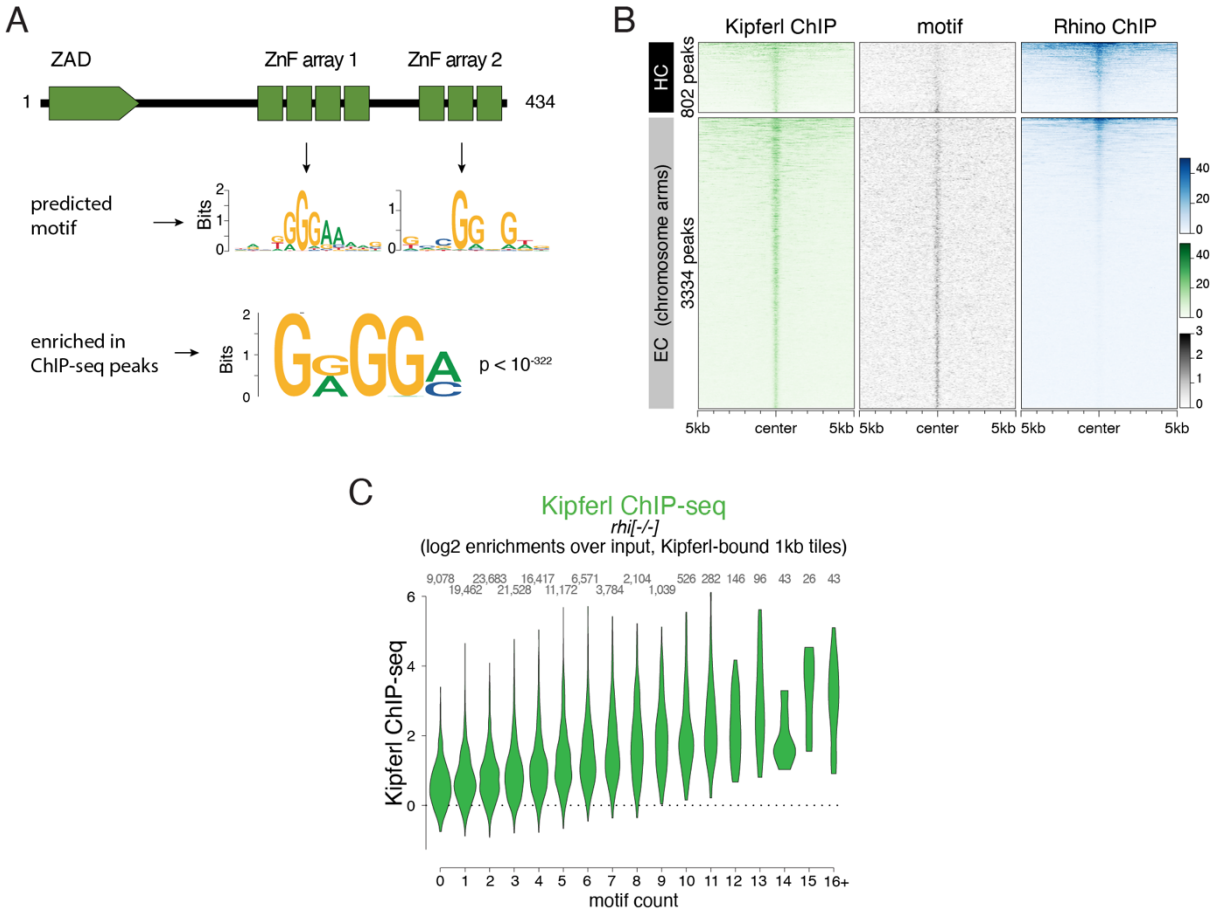


Figure 44: Kipferl binding sites are enriched in a Guanosine-rich sequence motif. (A) Schematic representation of Kipferl protein domain architecture and position weight matrix of Kipferl binding motifs predicted for the two ZnF arrays (Princeton Cys2His2 PWM predictor) or determined experimentally by HOMER from top 3,000 narrow Kipferl peaks found in two independent ChIP-seq replicates from *rhino* mutant ovaries. The GRGGN motif was found at least once in 83.4 % of peaks versus 50.3% in control sequences ($p < 10\text{-}322$). (B) Heat map depicting Rhino-independent Kipferl ChIP-seq signal and GRGGN motif enrichment centered on narrow Kipferl peaks analogous to panel E, sorted by Kipferl signal in *rhi[-/-]* (motif count: # of motifs per non-overlapping genomic 100 bp window). (C) Violin plot showing the relation of Kipferl ChIP-seq enrichment with GRGGN motif count on 1-kb tiles.

We were not able to purify Kipferl or its zinc finger arrays as recombinant proteins to study DNA binding directly. We therefore expressed FLAG-tagged Kipferl from an integrated transgene at high levels in cultured ovarian somatic stem cells (OSCs), which do not express endogenous Rhino. Kipferl ChIP-seq using endogenous antibodies from wildtype OSCs did not show any enrichment, despite low level expression. Instead, the ectopically expressed FLAG-tagged Kipferl bound the OSC genome at thousands of defined sites (Figure 45A). Kipferl ChIP-seq peaks were strongly enriched in the GRGGN sequence motif characteristic for Kipferl binding sites in ovaries, and a *de novo* motif enrichment analysis confirmed the motifs

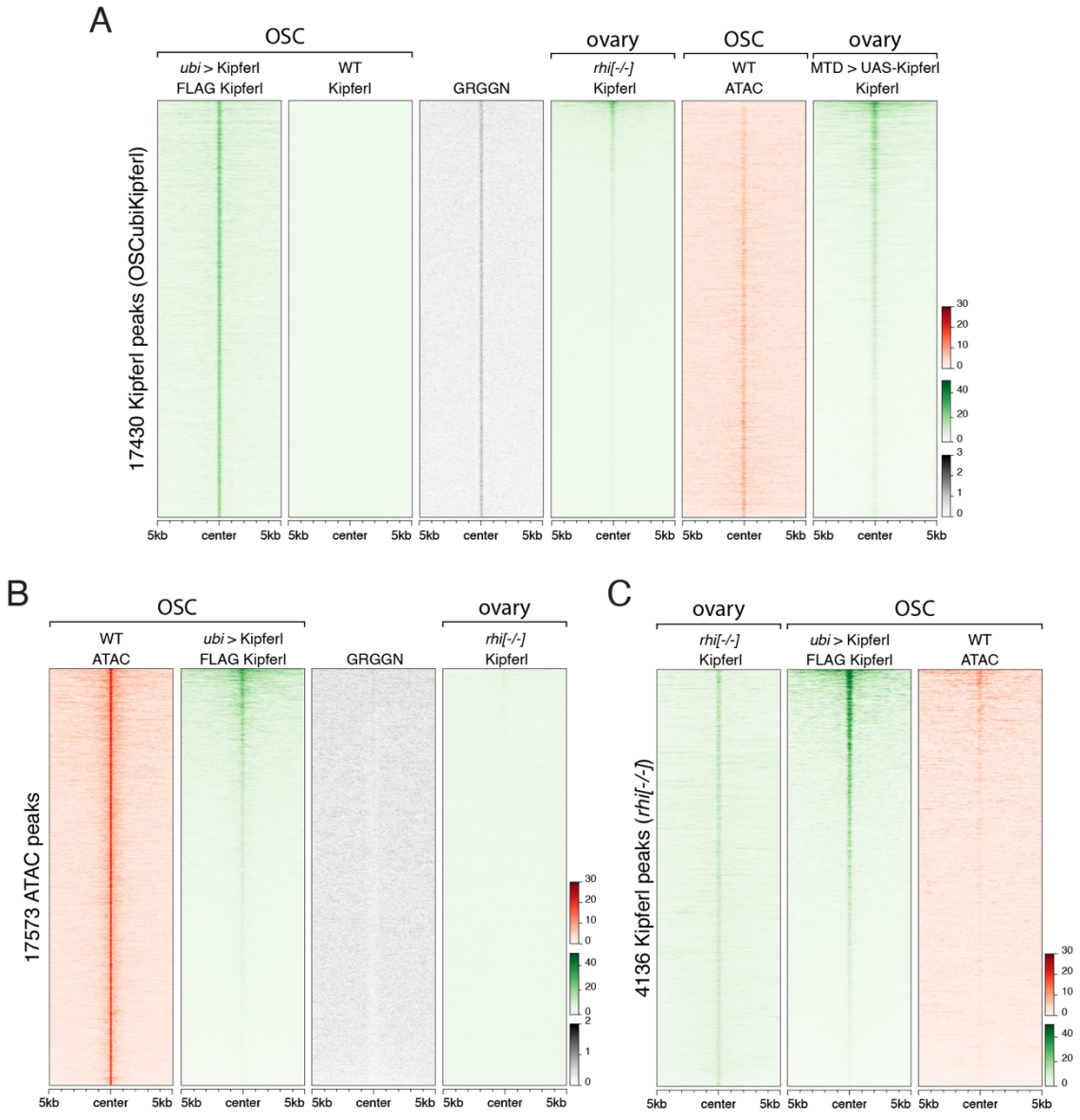


Figure 45: Ectopic overexpression of Kipferl in OSCs leads to the formation of thousands of Kipferl peaks at accessible GRGGN-enriched sites. (A) Heat map centered on Kipferl peaks detected via FLAG ChIP in OSCs stably overexpressing tagged Kipferl under control of the ubi63E promoter. Indicated data are sorted by Kipferl ChIP signal in *rhino* knock out ovaries. Motif count is given as count per non-overlapping genomic 100 bp window. (B) Heat map centered on ATAC seq peaks found in wildtype OSCs, sorted by Kipferl signal detected in OSCs overexpressing tagged Kipferl under control of the ubi63E promoter. Motif count is given as count per non-overlapping genomic 100 bp window. (C) Heat map centered on Kipferl peaks detected in two independent replicates of Kipferl ChIP in *rhino* knock out ovaries, sorted by Kipferl signal detected in OSCs stably overexpressing tagged Kipferl under control of the ubi63E promoter depicting the indicated signal.

identified in ovaries. Many of the OSC Kipferl peaks were also bound by Kipferl in *rhino* mutant ovaries. The majority of Kipferl binding sites in OSCs, however, were not bound in ovaries (Figure 45A). We hypothesized that this might be due to differences in protein levels in the two systems, as well as DNA accessibility in ovaries and OSCs. Indeed, Kipferl binding sites in OSCs generally fell in regions of local chromatin accessibility as measured by ATAC-seq in wildtype OSCs, indicating that Kipferl preferentially binds to its DNA motifs if those are intrinsically accessible (Figure 45A). Furthermore, overexpression of Kipferl in ovaries resulted in Kipferl enrichment at many additional OSC peaks (Figure 45A). Importantly, accessible sites devoid of GGGG/GGAG were not bound by Kipferl, neither in OSCs nor in *rhino* mutant ovaries (Figure 45B). The majority of Kipferl binding sites detected in *rhino* mutant ovaries was also bound by Kipferl in OSCs, and a lack of Kipferl signal in OSCs was often accompanied by absence of ATAC-seq signal, indicating that a lack of accessibility might hinder Kipferl's chromatin association at these sites (Figure 45C). Taken together, the OSC data further supported a direct and specific Kipferl-DNA interaction.

Overexpression of Kipferl in ovaries strengthened the Kipferl signal at preexisting binding sites, and lead to the formation of many new peaks (Figure 46A). New Kipferl peaks were bound by Rhino if they overlapped with local H3K9me2/3. They generally corresponded to sites bound by Kipferl upon overexpression in OSCs and were formed at sites enriched in the GRGGN motif. Many sites showed low levels of Kipferl binding already in wildtype ovaries, which however was not sufficient for detection via peak calling. This indicates that the low Kipferl levels in wildtype ovaries are limiting the number and the selection of nucleation sites bound by Kipferl and Rhino. Female fertility was not negatively impacted by the strong MTD-Gal4-mediated overexpression of Kipferl, leaving the physiological importance of low Kipferl levels up for debate (Figure 46B). Overall, our data are in agreement with a model where Kipferl binds to GRGGN motifs and recruits Rhino to sites concomitantly marked by H3K9me2/3.

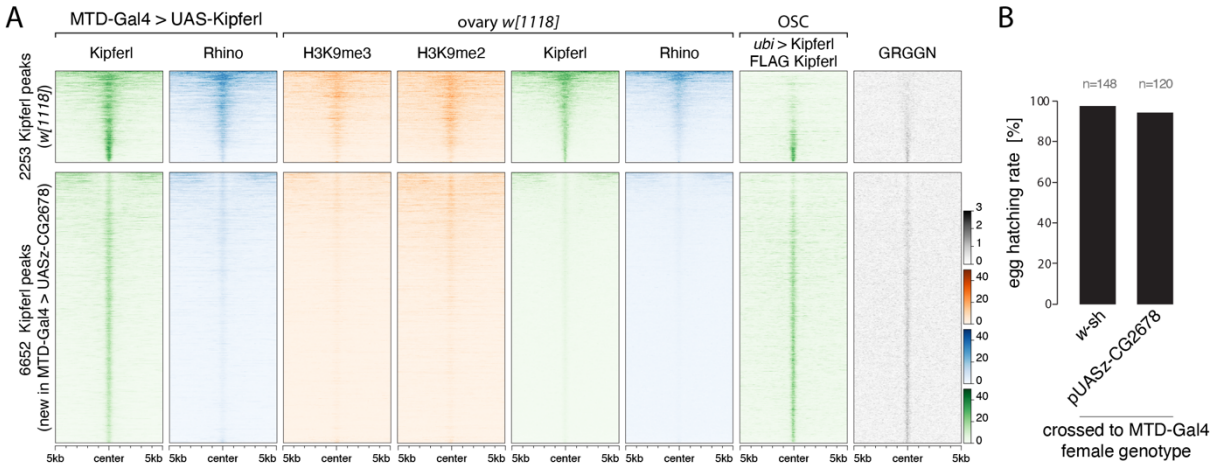


Figure 46: Overexpression of Kipferl in the ovary does not impact fitness. Heat map centered on broad Kipferl peaks detected in two independent w^{1118} ChIP-seq replicates (top) or in ovaries with MTD-Gal4-mediated Kipferl overexpression. Indicated data are sorted by Kipferl ChIP-seq signal in w^{1118} ovaries (motif count: # of motifs per non-overlapping genomic 100 bp window). (B) Bar graph depicting the hatching rate for eggs laid by females of the indicated genotype. Number of eggs is given as n.

2.3.4 Structure function analysis and interaction mapping

2.3.4.1 Lessons from tagged Kipferl

The ChIP-seq analyses and Kipferl's identification through a yeast two-hybrid screen suggest that Kipferl directly interacts with both DNA and Rhino. In order to further study how the Kipferl protein mediates these functions, we made use of recombination mediated cassette exchange (RMCE) to introduce various tagged and truncated Kipferl variants into flies carrying the Kipferl locus deletion (Bateman et al., 2006). In the process, we found that the Kipferl locus harbors an additional 25 amino acid exon, flanked by 109 and 31 nucleotides of intronic sequence up and downstream, respectively, which is not annotated in the *Drosophila* reference genome. PCR analysis showed that the respective DNA sequence was present in several laboratory fly strains, but that it was absent from *iso1* genomic DNA, which served as the basis of the reference genome. We compared the Kipferl protein sequence with its homologs in other *Drosophila* species, which confirmed the presence of the additional in frame coding sequence, inserted between P347 and K348 of the annotated *melanogaster* Kipferl protein (Figure 47). As the reference genome version of Kipferl, lacking the additional

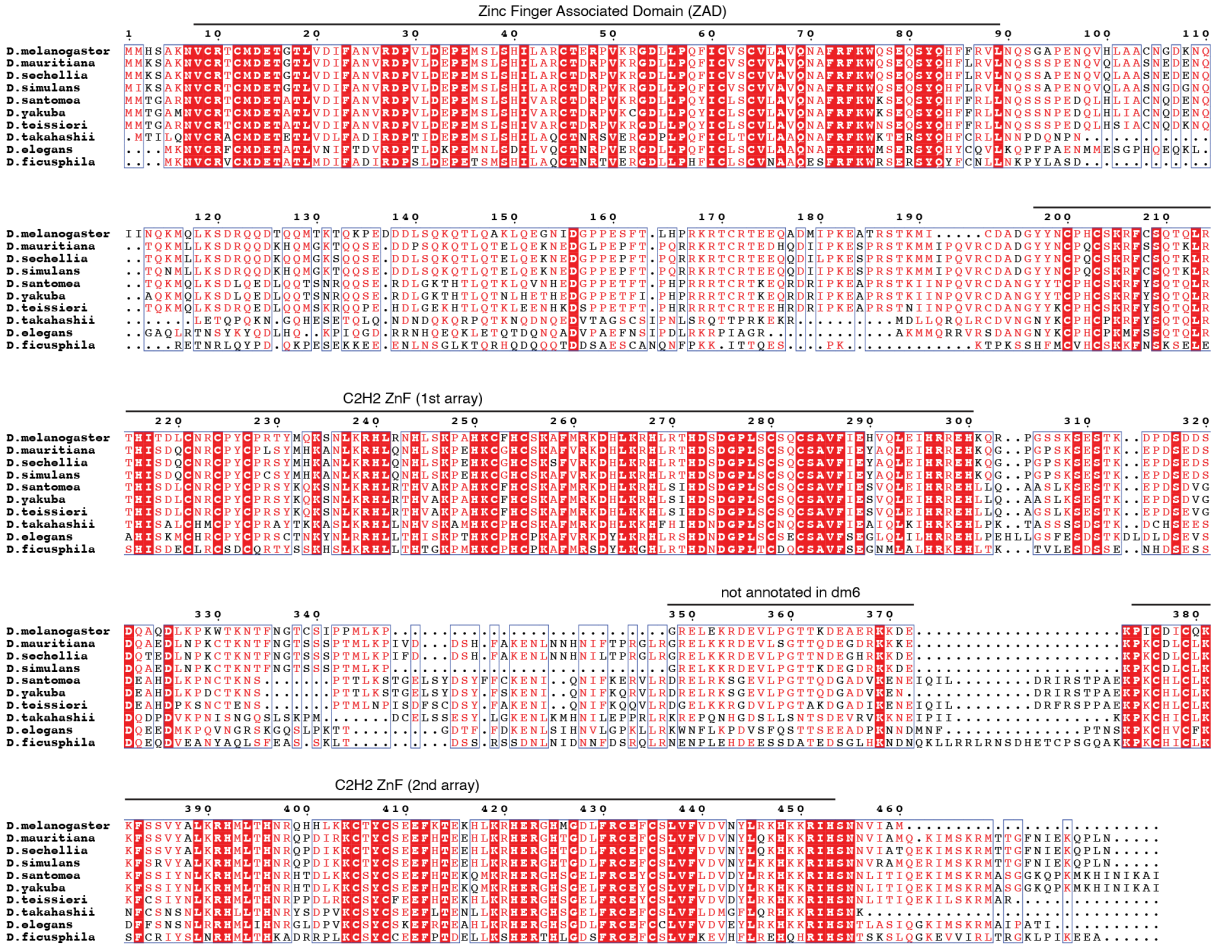


Figure 47: Multiple sequence alignment of Kipferl protein from different *Drosophila* species confirms additional coding sequence. Multiple sequence alignment of selected reciprocal best CG2678 homologs generated with mafft v7.505 (Rozewicki et al., 2019) and visualized with espript v3.0 (Robert and Gouet, 2014). Protein sequences included in the alignment were obtained from the NCBI non-redundant database, their protein accessions being XP_033162862, XP_032579476, EDX12039, XP_039231961, XP_039493628, XP_043654173, XP_044252272, XP_017119505, XP_017057177.

exon, fully restored female fertility to slightly higher levels than the version containing the additional sequence, all RMCE rescue constructs were designed to harbor the reference genome version of Kipferl (Figure 48A).

While expression of untagged Kipferl was able to fully rescue the fertility defect of *kipferl* mutant females, the addition of a GFP tag at either terminus of the Kipferl protein abrogated the ability of a tagged construct to fully rescue the mutant phenotype, with clear fertility defects remaining upon introduction of GFP-tagged Kipferl variants (Figure 48B) . Expression levels for C-terminally GFP-tagged Kipferl were low and the signal only partially overlapped with Rhino, which accumulated in elongated structures at the nuclear envelope (Figure 48A). Interestingly, N-terminal tagging of Kipferl with GFP led to a complete rescue of the Rhino accumulations at the nuclear envelope characteristic of the *kipferl* mutant phenotype and also increased fertility to near-wildtype levels. N-terminally GFP tagged Kipferl colocalized with Rhino in several small nuclear foci (Figure 48A). The ChIP-seq pattern observed for Rhino in these flies, however, corresponded precisely to the pattern observed in *kipferl* mutant ovaries, with Rhino being enriched only at piRNA clusters *42AB* and *38C* (Figure 49A, B, C).

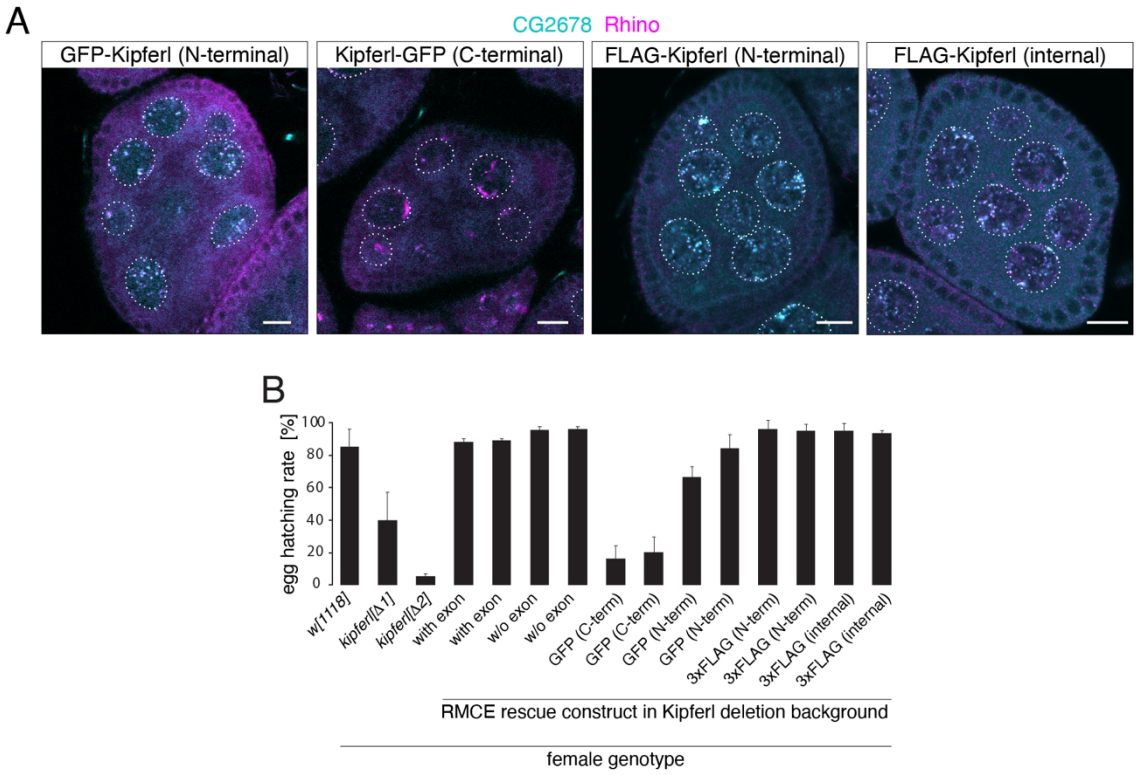


Figure 48: Different tagged versions of Kipferl show varying degree of functionality. (A) Bar graph depicting the hatching rate for eggs laid by females of the indicated genotype. (B) Confocal images depicting immunofluorescence signal for CG2678 constructs with indicated tags (cyan) and Rhino (magenta). Scale bar: 5 µm.

The same combination of phenotypes (rescue of fertility and immunofluorescence pattern, but not ChIP-seq distribution) was also observed for a much smaller N-terminally 3xFLAG tagged Kipferl rescue construct (Figure 48, Figure 49A, C). This indicated that even small disturbances of the very N-terminal protein fold interfered with proper Kipferl function. We argue that this was not due to a lack in dimerization capability, as the colocalization with

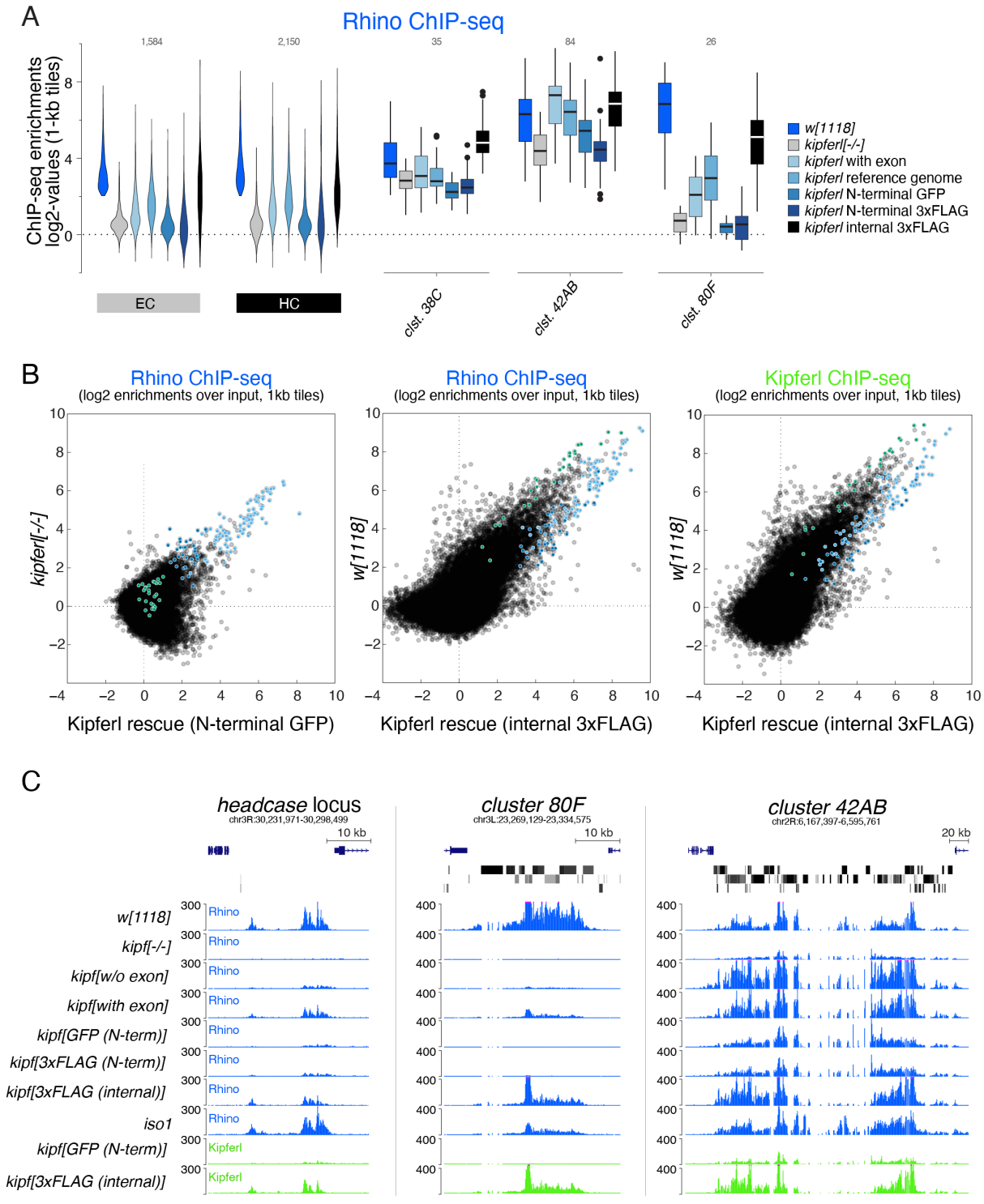


Figure 49: N and C-terminal tagging of Kipferl impairs the functionality of the protein. (A) Violin plot showing average log2 fold enrichment of Rhino ChIP-seq over input for Rhino-bound 1-kb tiles (defined through two independent w^{1118} data sets analogous to previous analyses) from ovaries expressing the indicated Kipferl RMCE rescue constructs. Rhino-dependent piRNA clusters *38C, 42AB,* and *80F* were analyzed separately (shown as box plots due to low number of tiles). (B) Scatter plots contrasting the indicated ChIP-seq enrichments for genomic 1-kb tiles. (C) UCSC genome browser tracks depicting Rhino and Kipferl ChIP-seq signal in the indicated genotypes.

Rhino was maintained and N-terminally GFP-tagged Kipferl was enriched at Kipferl-independent piRNA clusters *42BA* and *38C* in ChIP-seq, presumably due to functional interaction with Rhino, which we found to depend on dimerization (see next section).

To circumvent these complications, we introduced a 3xFLAG tag into the unstructured linker region between the ZAD and the first ZnF array, which led to a full rescue of fertility, colocalization with Rhino in small nuclear foci, and largely restored wildtype Rhino chromatin binding pattern (Figure 48, Figure 49). We note that, despite a strong rescue of the *kipferl* mutant ChIP-seq phenotype, many sites don't fully recover to wildtype Rhino levels (Figure 49C). These sites often correspond to regions where the wildtype Rhino binding profile differs between the w^{1118} and *iso1* genotypes. Both the *iso1* background as well as flies harboring internally FLAG tagged Kipferl lack the additional exon found in Kipferl in w^{1118}. Future experiments might help elucidate if this difference in Kipferl sequence is the cause for the discrepancies we observed for the Rhino profiles between these genotypes.

2.3.4.2 *In vivo* functions of Kipferl's protein domains

To examine how the DNA binding ability, as well as interaction with Rhino are carried out by Kipferl, we expressed truncated FLAG-tagged versions of the Kipferl protein in flies harboring the Kipferl deletion through RMCE. Deletion of either one of the ZnF arrays (Kipferl$^{\Delta\text{-1st array}}$, Kipferl$^{\Delta\text{-2nd array}}$), the N-terminal ZAD (Kipferl$^{\Delta\text{-ZAD}}$), or the unstructured linker between the ZAD and the first ZnF array (Kipferl$^{\Delta\text{-linker}}$) allowed us to assign different functions to each domain (Figure 50A).

Deletion of the second ZnF-array had only mild impacts on the chromatin binding pattern of Kipferl, implying a stabilizing role of this domain (Figure 50B). This indicates that the long and short Kipferl protein isoform likely functions in the same pathway. Kipferl$^{\Delta\text{-1st array}}$ instead showed strongly reduced chromatin binding ability, demonstrating that the first ZnF array is crucial for Kipferl functionality (Figure 50B). Like the first ZnF array, deletion of the ZAD

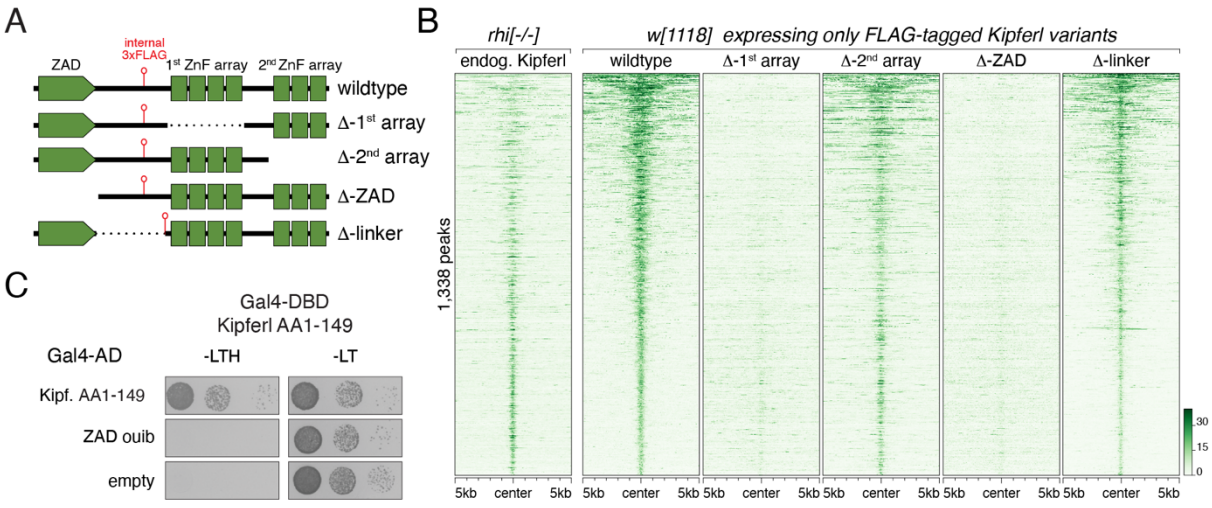


Figure 50: ZnF array 1 is crucial for chromatin binding. (A) Schematic representation of Kipferl rescue constructs harboring the wildtype protein sequence or indicated deletions, as well as an internal 3xFLAG tag. Rescue constructs were introduced into the endogenous kipferl locus via RMCE. (B) Heat map displaying Kipferl variant ChIP-seq signal centered on peaks bound by Kipferl in *rhino* mutants (data sorted by the ChIP signal detected for the wildtype Kipferl rescue construct; shown are only peaks that are Kipferl-bound in ovaries expressing wildtype tagged Kipferl). (C) Yeast two-hybrid assay on selective (left) and nonselective medium (right). AD: activation domain; DBD: DNA binding domain.

resulted in a complete loss of Kipferl ChIP-seq signal (Figure 50B). The ZAD, which is shared among the >90 ZAD-ZnF family members in *D. melanogaster*, was described to act as an anti-parallel homodimerization domain and has not been directly implicated in DNA binding (Jauch et al., 2003, Bonchuk et al., 2021). We confirmed the homodimerization capability of the Kipferl ZAD using yeast two-hybrid experiments (Figure 50C). In line with our results, previous reports demonstrated that dimerization is crucial for the DNA binding capability of ZAD-ZnF proteins (Maksimenko et al., 2020). The deletion of the linker region had only a minor effect on Kipferl's DNA binding patter. Kipferl$^{\Delta\text{-linker}}$ bound to Kipferl binding sites in a pattern resembling that of wildtype protein (Figure 50B). Together, these data suggest that dimeric Kipferl binds to DNA primarily through its first ZnF array.

To gain insight into the molecular interaction between Kipferl and Rhino, we made use of the characteristic localization of the two proteins in nurse cell nuclei. While we observed perfect overlap of FLAG-tagged wildtype Kipferl with Rhino in nuclear foci, both variants with defects in DNA binding, Kipferl$\Delta^{\text{-1st array}}$ and Kipferl$^{\Delta\text{-ZAD}}$, failed to colocalize with Rhino (Figure 51). In these genotypes, Rhino accumulated in elongated domains at the nuclear periphery, reminiscent of the *kipferl* mutant phenotype. Rescue of the dimerization through incorporation of the Gcn4 dimerization domain or the ZAD of Ouija board, a ZAD-ZnF protein not expressed in ovaries, also restored the co-localization of Kipferl with Rhino (Kipferl^{GCN4} and Kipferl^{ouib}, Figure 51). Neither deletion of the unstructured linker between ZAD and first ZnF array, nor the second ZnF array had any impact on the colocalization with Rhino (Figure 51). Collectively, this indicates that the ZAD, the unstructured linker, and the second ZnF array are dispensable for the interaction with Rhino. Thus, the first ZnF array within dimeric Kipferl mediates the interaction with Rhino either via DNA binding or through a direct protein-protein interaction.

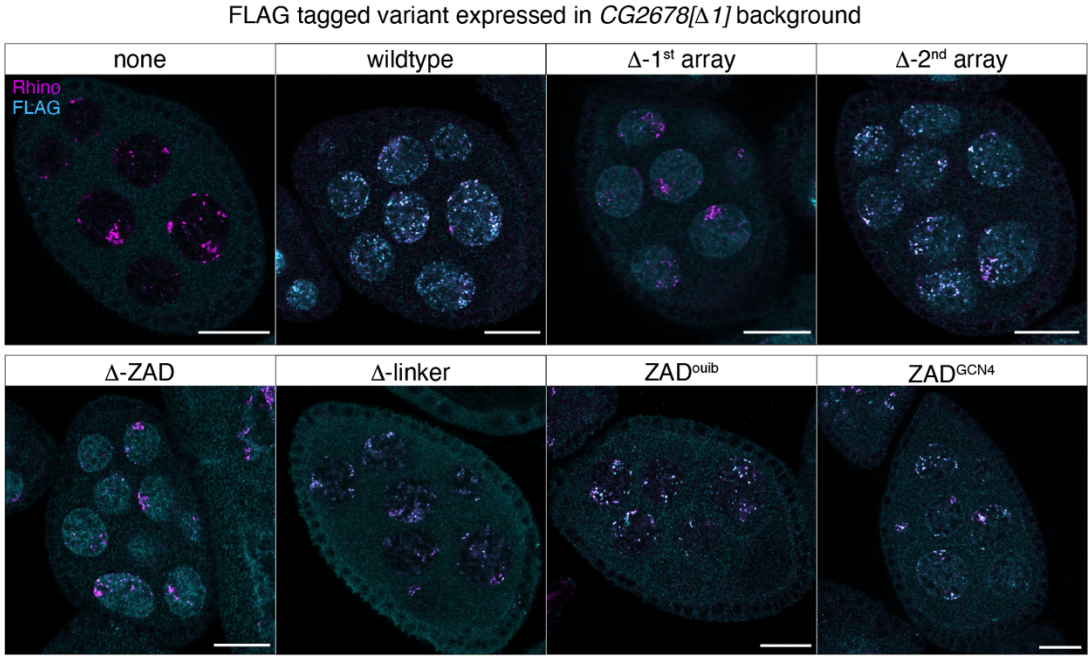


Figure 51: ZnF array 1 is required for the interaction with Rhino. Confocal images showing entire egg chambers with immunofluorescence stainings for Rhino (magenta) and indicated FLAG-tagged Kipferl variants (cyan) (scale bars: 20 µm).

2.3.4.3 Kipferl and Rhino interact at Kipferl's 4th ZnF and Rhino's chromo domain

To determine if Kipferl's first ZnF array interacted directly with Rhino, we replicated the direct interaction of the two full length proteins in a yeast two-hybrid experiment. Probing full length Rhino against partly overlapping fragments of Kipferl confirmed an interaction between Rhino and the first ZnF array within Kipferl, and did not reveal any additional interaction interfaces (Figure 52). To separate the interactions with Rhino and DNA, respectively, we divided the first ZnF array into smaller fragments, identifying the 4th ZnF as the smallest unit sufficient for the interaction with Rhino (Figure 52). We verified this finding *in vivo*, where flies expressing Kipferl$^{\Delta\text{-ZnF\#4}}$ instead of the endogenous protein showed no colocalization of Kipferl and Rhino, and the latter displayed the characteristic *kipferl* mutant phenotype (Figure 54A). This extended also to Rhino's ChIP-seq pattern, which resembled the *kipferl* mutant situation: Rhino was lost from all Kipferl-dependent regions, but remained bound at *42AB* and *38C* (Figure 54B, F). Rhino was bound to Satellite repeats, and female fertility was reduced (Figure 54C, D). Kipferl$^{\Delta\text{-ZnF\#4}}$, which was diffusely localized in nurse cell nuclei similar to wildtype Kipferl in the absence of Rhino, fully retained its DNA binding ability and mirrored the chromatin binding pattern of wildtype Kipferl in *rhino* mutant ovaries (Figure 54A, E, F). This demonstrates that the sequence-specific binding to DNA is mediated by ZnFs 1-3, while ZnF 4 interacts with Rhino.

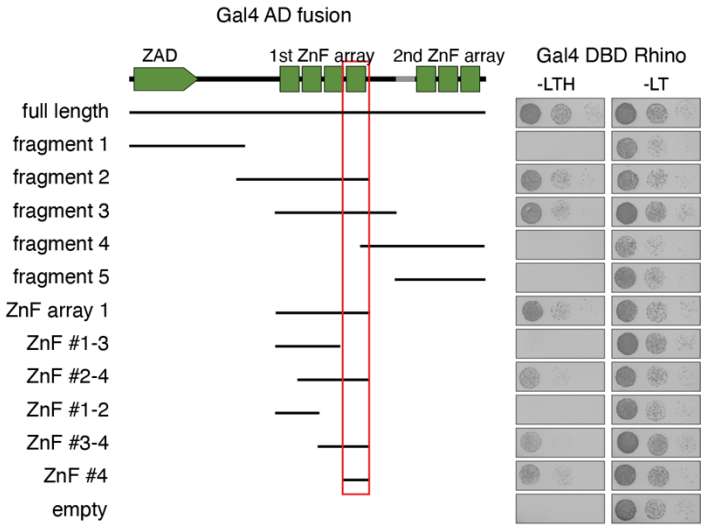


Figure 52: Yeast two hybrid interaction mapping identifies ZnF#4 as interaction point with Rhino. Yeast two-hybrid assay on selective (left) and nonselective medium (right). AD: activation domain; DBD: DNA binding domain.

Similar to other HP1 proteins, the only described interface for protein-protein interactions within Rhino is located at its chromoshadow domain. To determine if Kipferl's 4th zinc finger interacts with the same domain, potentially competing with Rhino's interaction partner Deadlock, we inverted the yeast two-hybrid assay to probe the three domains of Rhino against Kipferl's 4th ZnF. This revealed no interaction with the chromoshadow or the hinge domain, but mapped the binding site for Kipferl to Rhino's chromo domain (Figure 53). Kipferl did not interact with the chromo domain of Su(var)2-5, indicating that the binding was specific for Rhino.

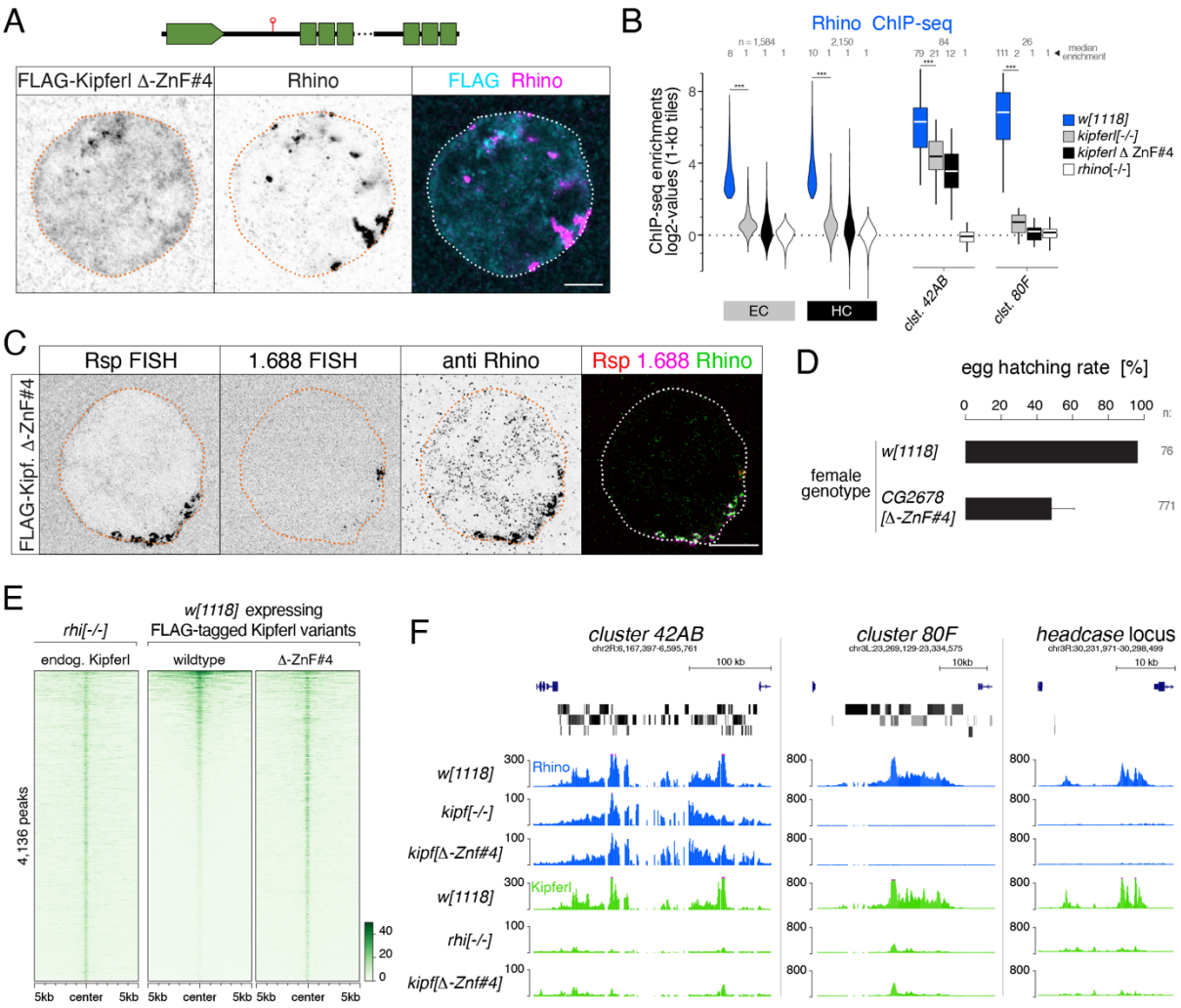


Figure 54: Deletion of ZnF#4 separate Rhino and Kipferl functionally. (A) Confocal images of a representative nurse cell nucleus depicting localization of Kipferl lacking ZnF#4 and Rhino in flies expressing only Kipferl$^{\Delta\text{-ZnF\#4}}$ (scale bar: 5 µm). (B) Violin plots showing the average log2-fold enrichment of Rhino ChIP-seq signal over input for Rhino-bound 1-kb tiles in *w^{1118}* (n=2), *kipferl* mutant (n=3), Kipferl$^{\Delta\text{-ZnF\#4}}$ (n=1) and *rhino* mutant (n=1) ovaries. *** corresponds to p < 0,001 based on student's t-test. (C) Confocal images showing *Rsp* and *1.688* Satellite RNA FISH signal and Rhino in nurse cells of Kipferl$^{\Delta\text{-ZnF\#4}}$ flies (scale bar: 5 µm). (D) Bar graph depicting the hatching rate of eggs laid by females of the indicated genotype. The total number of eggs is given as n. (E) Heat map showing indicated ChIP-seq signal, centered on narrow Kipferl peaks detected in two independent ChIP-seq experiments of *rhino* mutant ovaries (data sorted by ChIP-seq signal detected for wildtype Kipferl rescue). (F) USCS genome browser tracks showing ChIP-seq signal (coverage per million sequenced reads for one representative replicate) for indicated proteins and genotypes at piRNA clusters *42AB* and *80F* and the *headcase* locus.

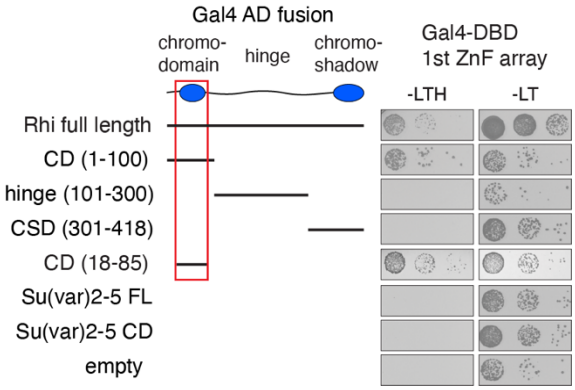


Figure 53: Rhino interacts with Kipferl via its chromo domain. Yeast two-hybrid assay on selective (left) and nonselective medium (right). AD: activation domain; DBD: DNA binding domain.

These findings were in line with our previous observation that a Rhino chromo domain dimer held the capacity to bind to a subset of Rhino domains in a pattern reminiscent of Kipferl nucleation sites (Figure 21). Our combined data thus demonstrate that Kipferl recruits Rhino to chromatin via a direct interaction between its 4th ZnF and the Rhino chromo domain. Furthermore, we conclude that functional Rhino is required for the extension and strengthening of Rhino/Kipferl domains.

2.3.5 The Kipferl consensus motif is enriched in few transposon sequences

The primary role ascribed to the piRNA pathway is the silencing of transposable elements. Considering their sequence diversity, the identification of a sequence specific guidance factor for Rhino, and thereby as a determinant for piRNA production, came as a surprise. Transposon sequences vary strongly in their overall nucleotide content (Figure 55A). Concomitantly, AT-rich elements exist that harbor none or few Kipferl binding motifs. While the GC-rich *gypsy8* and *Rt1a/b/c* elements harbored between 29 and 38 motives (Figure 55B), we do not find an overrepresentation of the GRGGN motif in transposon sequences as compared to random

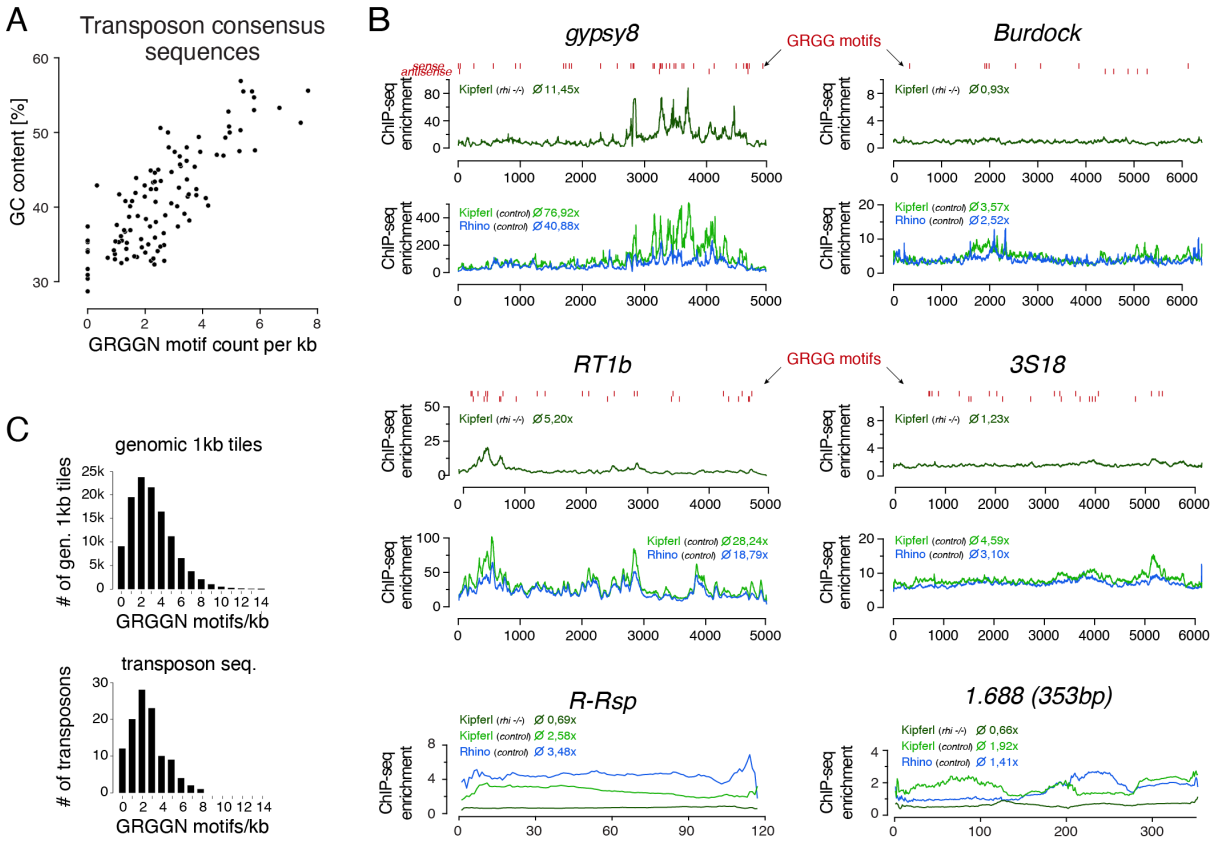


Figure 55: TE sequences have variable GC content. (A) Scatter plot of overall G/C content per transposon sequence versus the number of GRGGN motifs normalized to transposon length. (B) ChIP-seq enrichment profiles on consensus sequences of transposons with high (*gypsy8* and *Rt1b*) or low (*Burdock* and *3S18*) number of Kipferl DNA binding motifs per kb sequence and on Satellite sequences lacking Kipferl motifs (*R-Rsp* and *353bp 1.688* element). Indicated ChIP-seq signals are displayed as average enrichment over input in two (Kipferl) or three replicates (Rhino) of ovaries from *rhino* mutant (top tracks) or MTD-Gal4 > *w*-sh control ovaries (bottom tracks). Red bars indicate motif instances on the sense or antisense strand. Numbers indicate average enrichment across the entire element. (C) Bar graphs showing the distribution of motif frequency in all genomic 1-kb tiles (top) and in transposon consensus sequences (bottom).

genomic tiles (Figure 55B, C). Kipferl's ChIP-seq enrichment on transposon consensus sequences in *rhino* mutant ovaries showed a significant correlation with motif density per kilobase (Figure 56A, R = 0.64, p < 2.2e-16). Moreover, the extent of Rhino-independent Kipferl enrichment directly correlated with Rhino enrichment levels in wildtype ovaries for most elements (Figure 56B, R = 0.83 , p < 2.2e-16). Few exceptions of elements that were strongly occupied by Rhino in wildtype ovaries despite lack of baseline Kipferl binding were represented by telomeric transposons and elements contained in clusters *42AB* or *38C* (Figure 56C). These elements maintained Rhino binding in the absence of Kipferl, supporting the likely existence of alternative chromatin recruitment mechanisms for Rhino.

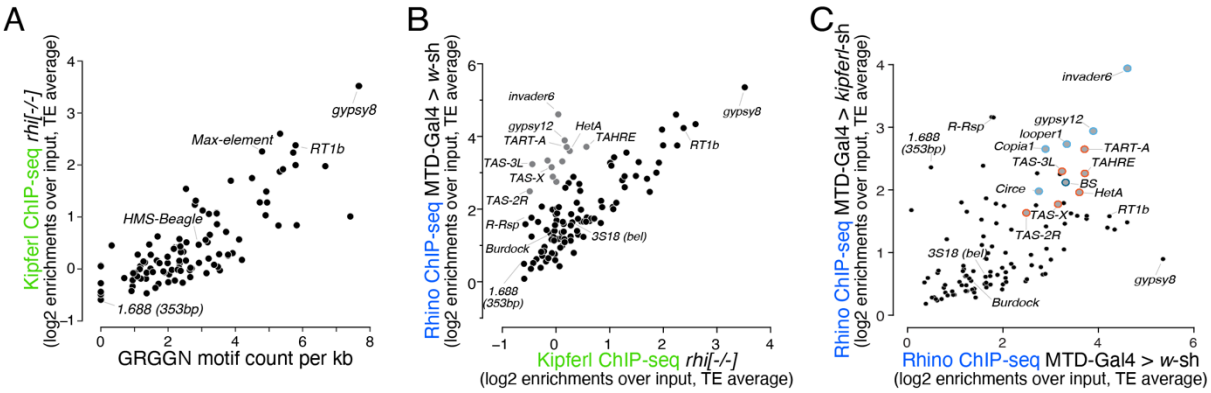


Figure 56: The number of GRGGN motifs dictate the occupancy of Kipferl and Rhino on TEs. (A) Scatter plot correlating the GRGGN motif count (normalized to element length) to the Rhino-independent Kipferl ChIP-seq enrichment for each transposon (ChIP-seq enrichments depict average of two independent experiments). (B) Scatter plot depicting the relation between wildtype Rhino ChIP-seq enrichments and Rhino-independent Kipferl ChIP-seq enrichments per transposon (average of two (Kipferl) or three (Rhino) independent experiments; elements indicated in grey are bound by Rhino in a largely Kipferl-independent manner, see also panel C). (C) Scatter plot contrasting Rhino ChIP-seq enrichment (over input) on transposon sequences in control (n=3) versus kipferl-depleted (n=2) ovaries. Elements bound by Rhino independently of Kipferl (see panel B) are enlarged and highlighted in grey (those elements with insertions in piRNA clusters *42AB* and *38C* are additionally indicated in light and dark blue, respectively, and telomeric elements are indicated in red).

As the number of GRGGN motifs per element, and thereby Kipferl binding, correlated with overall GC-content, enrichments for Rhino on transposons also correlated with GC-content (Figure 57A, B, $R_{motif\text{-}GC}$ = 0.78 , p < 2.2e-16; $R_{GC\text{-}Kipf}$ = 0.71 , p < 2.2e-16; $R_{GC\text{-}Rhi}$ = 0.59 , p < 1.1e-13). This correlation was abolished in ovaries depleted for Kipferl, indicating that Rhino's

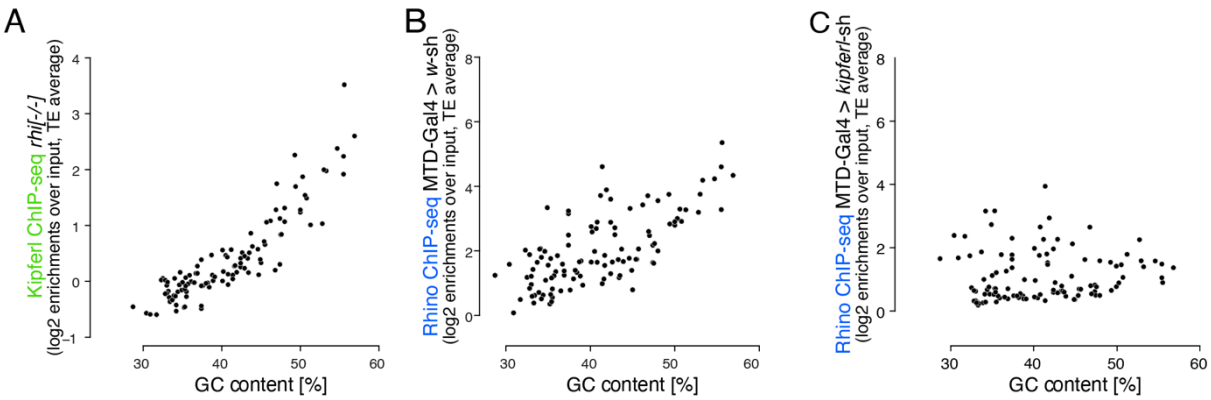


Figure 57: Rhino binding to transposable elements correlates with GC content in a Kipferl-dependent manner. Scatter plots correlating the overall GC nucleotide content per transposon to the ChIP-seq enrichment detected for Kipferl in *rhino* mutant ovaries (A) and for Rhino in wildtype (B) or Kipferl-depleted ovaries (C).

preferential binding to GC-rich transposons is due to Kipferl-mediated recruitment and/or stabilization of Rhino on DNA sequences with Kipferl motifs (Figure 57C, R = 0.08 , p = 0.34).

2.3.6 Kipferl stabilizes Rhino also at transposons lacking Kipferl binding sites

Many transposons did not show intrinsic Kipferl binding, yet showed low-level Rhino enrichment in wildtype ovaries which depended on Kipferl. Moreover, the silencing of several of these elements depended strongly on Rhino (e.g. *Burdock*, *HMS-Beagle*, *3S18*), and they were upregulated also upon loss of Kipferl (Figure 58).

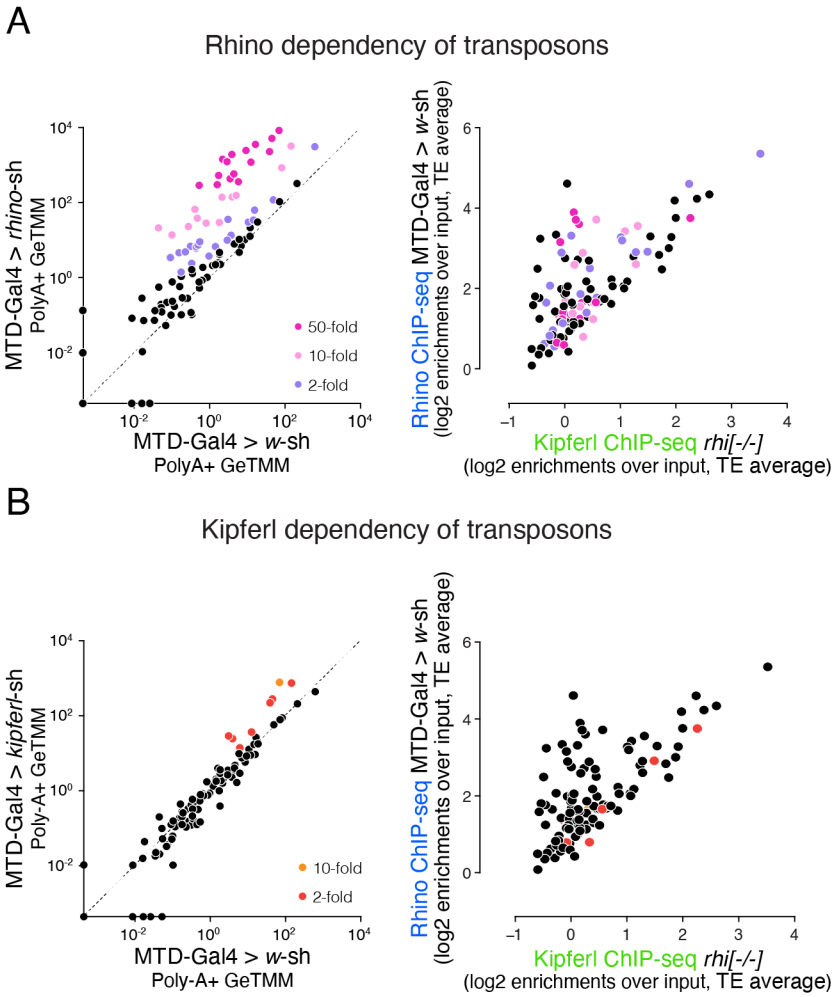


Figure 58: Not all Kipferl-dependent transposons harbor Kipferl binding sites. Left: Classification of transposons into different grades of Rhino (A) and Kipferl (B) dependency based on poly-adenylated sense transcripts detected in wildtype control ovaries versus *rhino* or *kipferl*-depleted ovaries (three replicates each). Values are displayed as Gene length corrected trimmed mean of M-values (GeTMM). Right: Scatter plot depicting the distribution of Rhino (A) or Kipferl (B) dependent elements in respect to their occupancy by Rhino in MTD-Gal4 > *w*-sh ovaries, as well as their Rhino-independent Kipferl affinity.

Stand-alone insertions of certain transposons have been previously proposed to act as independent Rhino domains, providing piRNAs capable of targeting other insertions in trans (Mohn et al., 2014, Shpiz et al., 2014). This model is supported by the recent observation that deletion of three major piRNA clusters (*38C, 42AB*, and *20A*) does not lead to upregulation of

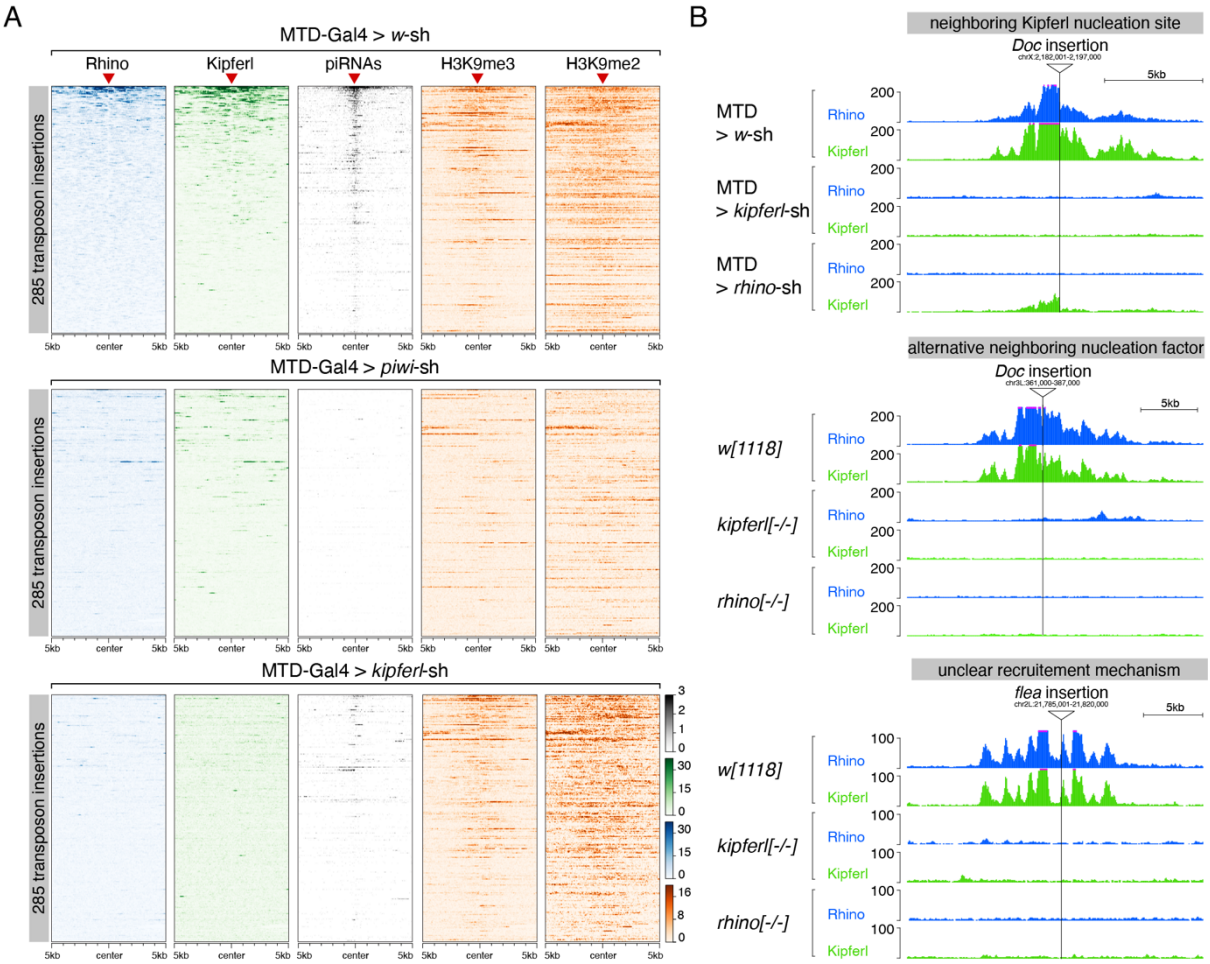


Figure 59: Kipferl has varying roles at stand-alone transposon insertions. (A) Heat maps depicting indicated ChIP-seq signal in the genomic regions flanking 285 euchromatic stand-alone insertions (red triangles) of Rhino-dependent transposons with low Rhino-independent Kipferl binding (data sorted by the ChIP-seq signal detected for Rhino in MTD-Gal4 > *w*-sh ovaries). (B) UCSC genome browser tracks of stand-alone transposon insertions found in MTD-Gal4 or w^{1118} strains, depicting examples of different potential modes of Kipferl dependency. ChIP-seq signal is shown as coverage per million reads for one representative replicate.

these elements and leaves flies fully fertile (Gebert et al., 2021). Considering that only a minority of single insertions are strongly bound by Rhino, the average Rhino enrichment mapped to the consensus sequence would be low, as observed for *HMS-Beagle* and *Burdock* (Akulenko et al., 2018). Based on stand-alone transposon insertions mapped in the MTD-Gal4 strain, we determined all insertions of Rhino-dependent transposons lacking intrinsic Kipferl binding. While nearly all insertions were marked with local H3K9me2/3, Rhino and Kipferl were only enriched at a subset of these insertions (Figure 59A). In agreement with previous reports, RNAi-mediated depletion of Piwi resulted in the loss of Rhino, Kipferl, H3K9me2/3, and piRNAs from all insertions. Depletion of Kipferl, on the other hand, induced the loss of Rhino and piRNAs, but did not impair heterochromatin at transposon insertions (Figure 59B).

We conclude that stable binding of Rhino requires a combination of Piwi-dependent H3K9me2/3 deposition and Kipferl-mediated recruitment or stabilization, even on transposons that do not contain strong Kipferl nucleation sites. Manual inspection of individual transposon insertions led us to propose that Kipferl can support Rhino at stand-

alone transposon insertions in different ways: Some of the strain-specific transposon insertions embedded within strong Rhino domains are located in the proximity of a Kipferl stand-alone peak present also in the other strains examined in this study (Figure 59 top). Neighboring Kipferl binding sites might therefore be a critical factor for the establishment of a strong local Rhino domain. In other instances, we find no nearby Kipferl binding site (Figure 59 middle, bottom). At these sites, Kipferl likely stabilizes Rhino upon recruitment through alternative nucleation factors, similar to other regions (eg. *ey* locus) that lack clear Kipferl binding sites. Indeed, we occasionally observe small residual enrichments of Rhino partly overlapping with Kipferl-dependent Rhino domains (Figure 59 middle). Taken together, our data indicate that Kipferl likely represents the first of several guidance factors determining Rhino's chromatin binding profile at a subset of cellular heterochromatin. We further conclude that Kipferl might hold a unique function, being required not just for Rhino binding at its cognate recruitment sites, but also at domains nucleated by alternative specificity factors.

2.4 The Rhino chromatin binding profile differs between ovaries and cultured GSCs

Our collective data imply Kipferl as an important specificity factor for Rhino in the adult *Drosophila* ovary. However, other than Rhino, Kipferl is not uniformly expressed in the germline lineage (Figure 25). No Kipferl protein was detected in testis and only very low levels were seen in early stages of the germarium in the ovary. In line with this, unpublished data from the Brennecke group indicate that in ovaries, Rhino-dependent piRNA populations differ between early (germline stem cells) and late stages of germ cell development, which make up the majority of ovarian tissue. A detailed characterization of germline stem cell chromatin is difficult, as these cells represent only a very small fraction of the entire ovary and are almost impossible to isolate in sufficient quantities without relying on genetic tricks (Niki et al., 2006, Kai et al., 2005). However, a stable germline stem cell (GSC) culture system has been established in the Brennecke group (Tirian et al., in preparation), allowing us to characterize Rhino's chromatin binding profile in a pure germline stem cell population. The GSC line was obtained from female flies genetically tailored to contain non-differentiating germline stem cells, that continuously divide in culture (Tirian et al., in preparation). Although it is unclear whether the GSC line fully recapitulates the *in vivo* germline stem cell fate, it expresses a fully

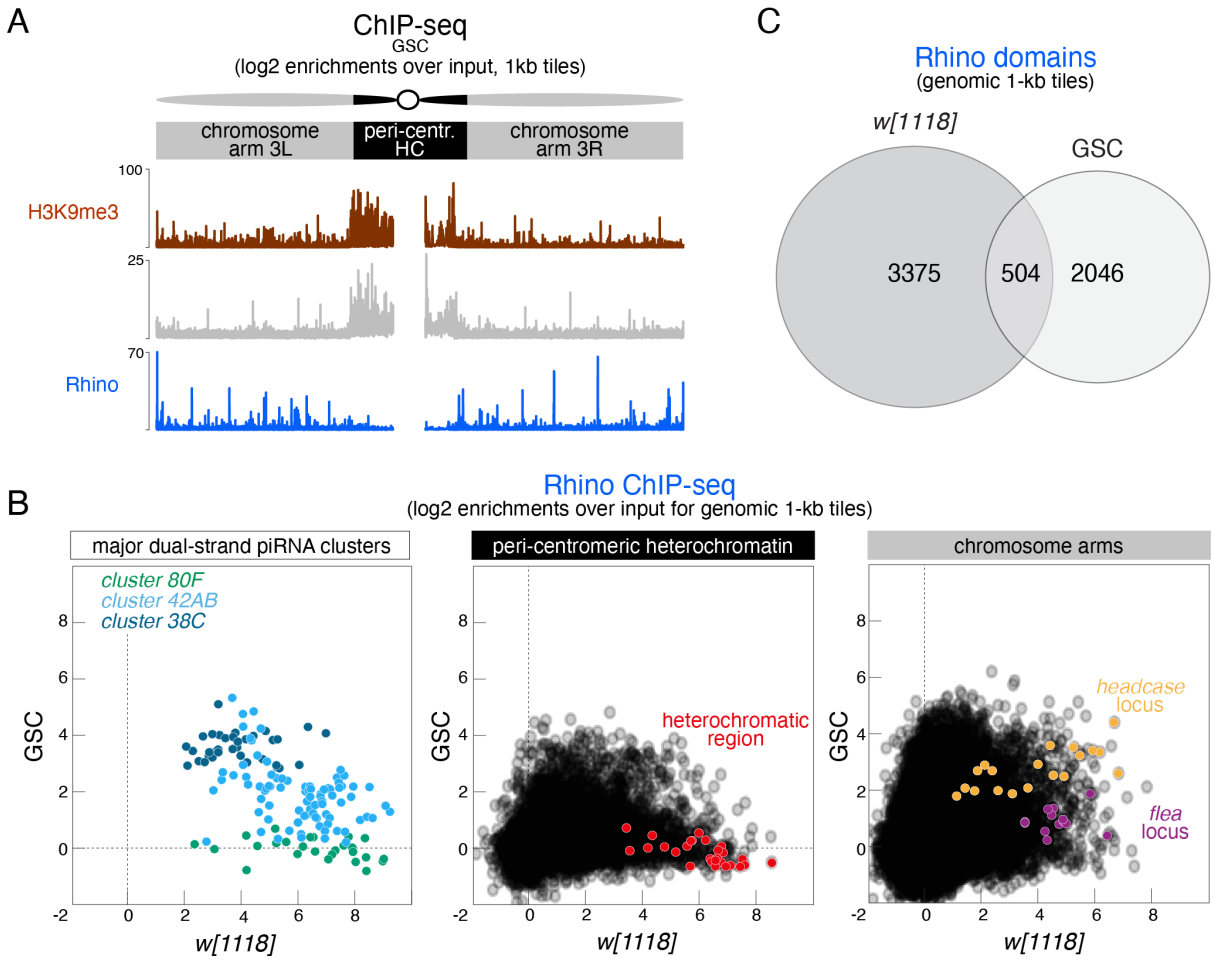


Figure 60: The Rhino chromatin occupancy pattern in cultured GSCs differs strongly from the ovarian Rhino pattern. *Legend on the next page.*

Figure 60: (A) ChIP-seq enrichment (genome unique reads; 1-kb tiles, one representative replicate each) of H3K9me3, HP1a, and Rhino along the assembled chromosome 3 sequence in cultured GSCs. Pericentromeric heterochromatin and euchromatic chromosome arms are indicated. (B) Scatter plot comparing average log2 ChIP-seq enrichments for Rhino in ovaries from w^{1118} versus cultured GSCs (1-kb tiles separated into pericentromeric heterochromatin and chromosomal arms; piRNA clusters *38C, 42AB*, and *80F* are shown separately; colored 1-kb tiles correspond to example loci in
). (C) Venn diagram depicting minimal overlap of Rhino bound tiles in cultured GSCs and w^{1118} ovaries. Rhino binding was defined by a binary cutoff of 4-fold enrichment over input in two independent libraries.

functional piRNA pathway with ping-pong piRNA biogenesis and Rhino-marked piRNA clusters. Furthermore, its transcriptional profile closely matches the germline stem cell profile derived from single cell RNA-seq experiments or from sorted germline stem cells derived from *bam* mutant ovaries (Kai et al., 2005, Rust et al., 2020). In agreement with our *in vivo* expression analysis, cultured GSCs express very low levels of Kipferl, and ChIP-seq experiments revealed no detectable Kipferl chromatin occupancy.

Encouraged by this, we performed ChIP-seq for Rhino, HP1a, as well as H3K9me3 in GSCs and compared these to ovarian ChIPseq data. The general distinction of chromosomes into pericentromeric heterochromatin and euchromatic chromosome arms based on HP1 and H3K9me3 holds true also for GSCs, analogous to the ovary data, despite lower overall enrichments of both marks (Figure 60A). Rhino's chromatin pattern, however, differs strongly between GSCs and ovaries. Overall, Rhino enrichments reach similar levels in GSCs and ovaries, but we observed fewer very strongly bound sites in GSCs. Maximum Rhino enrichments in GSCs reached roughly half of maximum ovarian Rhino enrichment values (Figure 60B). Furthermore, piRNA cluster identity differs in GSCs: we observed no detectable binding of Rhino at the prominent ovarian cluster *80F* and found considerably lower Rhino levels at cluster *42AB*. Among the three biggest ovarian piRNA clusters, only cluster *38C* is strongly bound by Rhino in GSCs. Instead, in GSCs Rhino occupies a large number of genomic 1kb tiles in euchromatin, and to a lesser extent in heterochromatin, that do not associate with Rhino in the ovary. Overall, we observe only weak overlap between Rhino bound sites in ovaries and GSCs (Figure 60C).

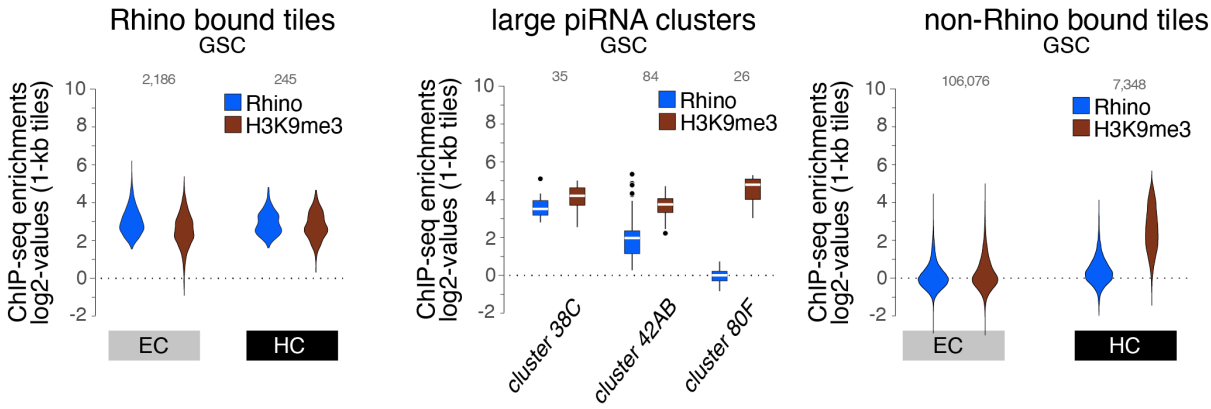


Figure 61: Rhino binds to a subset of H3K9me3-marked chromatin in cultured GSCs. Violin plots showing average log2 fold enrichment of Rhino and H3K9me3 ChIP-seq over input for 1-kb tiles from cultured GSCs. Tiles were grouped into Rhino-bound and non-Rhino-bound based on a cutoff of 4-fold enrichment of Rhino ChIP-seq signal over input in two independent GSC replicate experiment. Rhino-dependent piRNA clusters *38C, 42AB*, and *80F* were analyzed separately (shown as box plots due to low number of tiles).

In line with our observations in ovaries, Rhino-bound chromatin in GSCs was always enriched in H3K9me3 (Figure 61). But unlike in ovaries, where H3K9me3 levels were highly variable across sites with comparable Rhino enrichment levels (Figure 10), we did not observe large variations between K9me3 levels at Rhino-bound regions in GSCs. Furthermore, the majority of cellular H3K9me3-marked heterochromatin was not bound by Rhino, just as in ovaries (Figure 61). In contrast, HP1a bound to H3K9me3 genome wide, irrespective of Rhino occupancy, indicating that Rhino binding does not prevent binding of HP1a to H3K9me3 (Figure 62A). Vice versa, high levels of HP1 prevented binding of Rhino to H3K9me3-marked genomic tiles (Figure 62B). Together, this suggested that upon direct competition, HP1 can out-compete Rhino for the binding to H3K9me3.

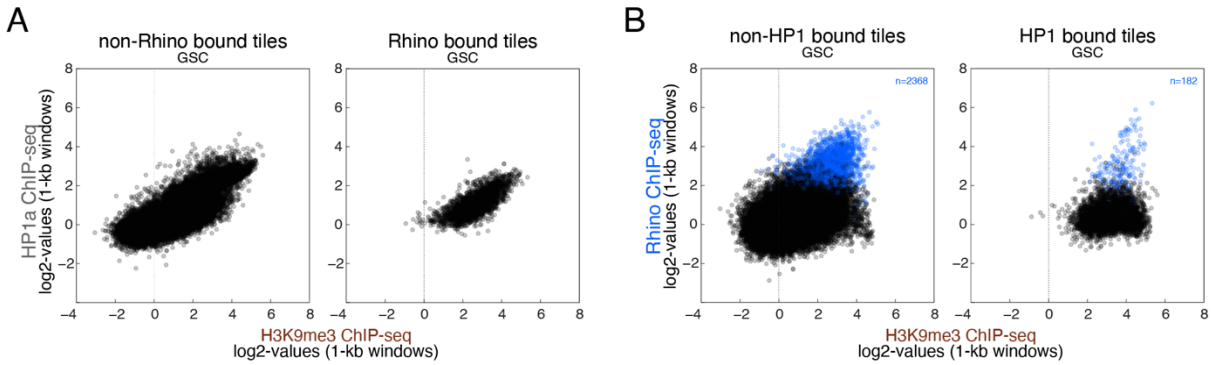


Figure 62: HP1a binds to H3K9me3 irrespective of co-occupancy by Rhino. Scatter plots contrasting the ChIP-seq enrichment of HP1a (A) and Rhino (B) to H3K9me3 levels at sites bound alone or together with the respective other protein. Blue dots in (B) indicate Rhino bound tiles and their number is given as n.

Overall, the differences in Rhino occupancy between GSCs and ovaries often correlated with differences in the underlying H3K9me3 levels (Figure 63). Interestingly, while GSC-specifically Rhino-bound tiles are found on all chromosomes, they are particularly frequent on the X chromosome. Instead, genomic tiles with less Rhino binding in GSCs relative to ovaries are enriched on chromosome 4. While differential H3K9 methylation levels are also enriched on

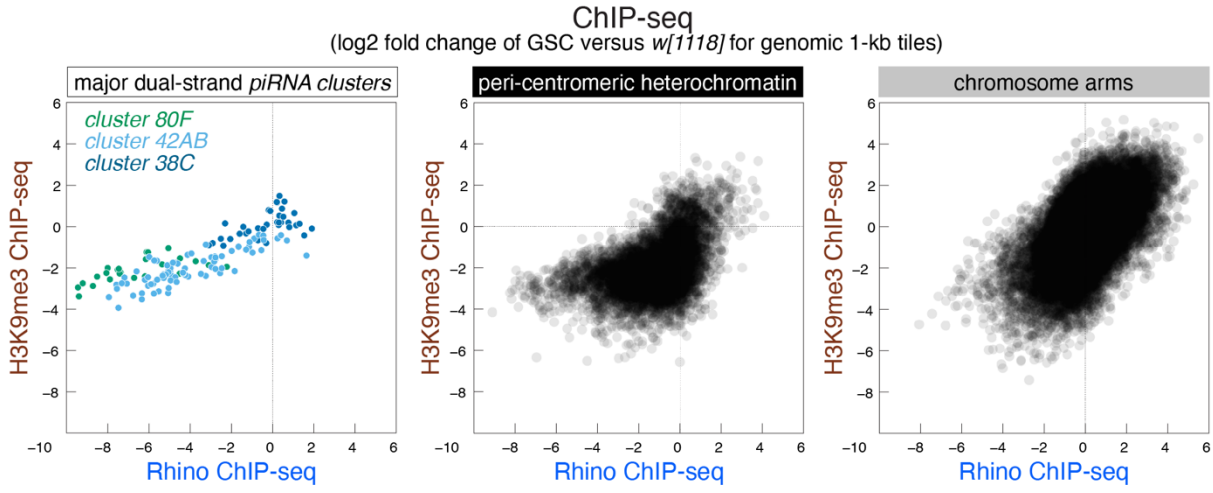


Figure 63: Changes in Rhino occupancy between cultured GSCs and ovaries are often accompanied by corresponding changes in H3K9me3. Scatter plot comparing log2 fold change in ChIP-seq enrichments for H3K9me3 and Rhino between w^{1118} ovaries and cultured GSCs in pericentromeric heterochromatin, chromosomal arms, and at piRNA clusters.

the X chromosome, most of these sites displayed a decrease in H3K9me3 levels in GSCs. On the X chromosome, increased Rhino occupancy is therefore not directly linked to H3K9me3 levels, pointing to potentially higher order differences in chromosome architecture as underlying basis. In this respect, it is interesting to note that the GSC line is haploid for the X chromosome (Tirian et al., in preparation). Maybe the loss of the second copy of the X chromosome induced changes in the chromatin landscape of the remaining copy that are relevant for Rhino binding.

In summary, Rhino occupies different subsets of H3K9me3-marked heterochromatin in GSCs and ovaries. These observations are consistent with previous findings that the piRNA pool is variable across developmental stages of germline cells, and implies a dynamic regulation of Rhino specificity, which is in line with the developmentally regulated expression pattern of Kipferl.

2.4.1 Ectopic Kipferl expression changes the Rhino profiles in cultured GSCs

Considering the developmental dynamics of Kipferl expression with barely detectable signal in early stages of the germarium, we hypothesized that the absence of Kipferl might be causal for the altered Rhino localization pattern in GSCs. Western blow analysis confirmed that, while Rhino levels were elevated in GSC lysate as compared to ovary lysate (relative to HP1a for example), Kipferl levels were detectable but noticeably lower in GSCs (Figure 64). Kipferl was not detectable in immunofluorescence stainings in GSCs, and no Kipferl ChIP-seq enrichment was detected.

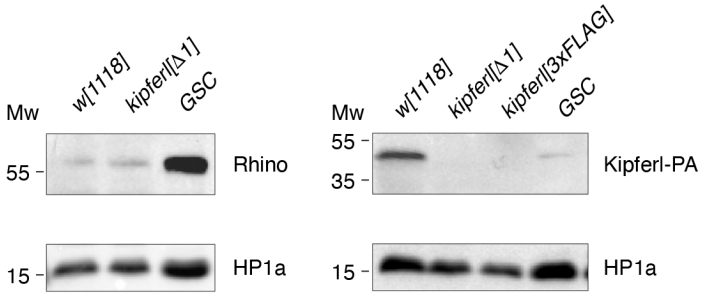


Figure 64: Kipferl levels are low in cultured GSCs. Western blot images showing the protein levels of Rhino (left) and Kipferl (right) in ovaries of the indicated genotype and cultured GSCs. HP1a serves as a loading control.

To investigate the effect of ectopic Kipferl expression in GSCs, we established a stable GSC line expressing internally FLAG-tagged Kipferl under the control of the strong *ubiquitin 63E* promoter (note that the internal FLAG tag disrupts the Kipferl antibody epitope, thereby preventing a direct comparison to wildtype protein levels). Based on ChIP-seq experiments, ectopic Kipferl bound to thousands of sites in GSCs (Figure 65A). These were enriched in the GRGGN Kipferl consensus motif and largely overlapped with sites bound by overexpressed Kipferl in OSCs and in ovaries. At Kipferl binding sites that displayed additional H3K9me3 occupancy, we observed ectopic Rhino recruitment (Figure 65B). This led to the establishment of pronounced Rhino domains at sites that were occupied by Kipferl and Rhino in wildtype ovaries, but not in wildtype GSCs.

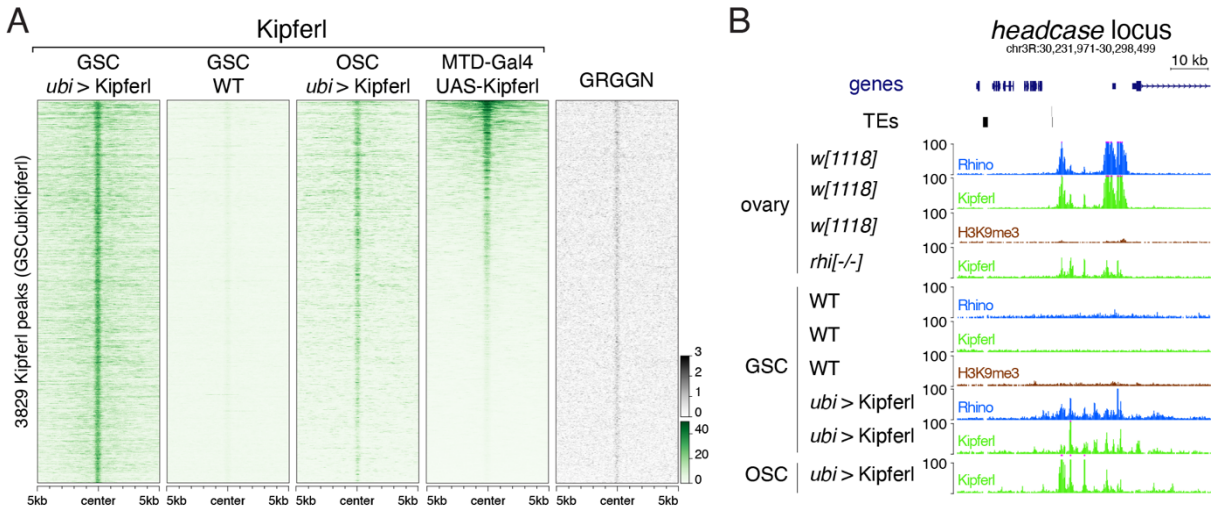


Figure 65: Overexpression of Kipferl in cultured GSCs leads to the binding of GRGGN motifs and the recruitment of Rhino to sites overlapping with local heterochromatin. (A) Heat map depicting Kipferl signal in the indicated genotypes centered on Kipferl peaks detected in cultured GSCs upon overexpression of Kipferl using both endogenous and anti-FLAG antibody. Signal is sorted by Kipferl levels in ovaries overexpressing untagged Kipferl. Motif count is given as number of motifs per non-overlapping genomic 100 bp window. (B) UCSC genome browser tracks depicting indicated ChIP-seq signal (coverage per million reads) at the *headcase* locus.

Furthermore, Kipferl co-occupied all Rhino bound tiles genome-wide upon overexpression in GSCs, indicating that the interaction with Rhino was fully intact also in cultured cells (Figure 66A). We also observed an increase in Rhino signal at certain previously occupied sites upon overexpression of Kipferl, indicating that, similar to our observations in ovaries, Rhino and Kipferl generally re-enforce each other's chromatin association.

To our surprise, overexpression of Kipferl in GSCs failed to establish a Rhino domain at piRNA cluster *80F*, which is not bound by Rhino in GSCs (Figure 66B). This was despite detectable Kipferl enrichment at cluster *80F*, which was however considerably lower than the levels reached in *rhino* mutant ovaries. Maybe a closed conformation of local chromatin at cluster *80F* prevents efficient Kipferl binding in GSCs, and it is the transcription of the neighboring genes upon onset of germ cell differentiation that contributes to the opening of the region,

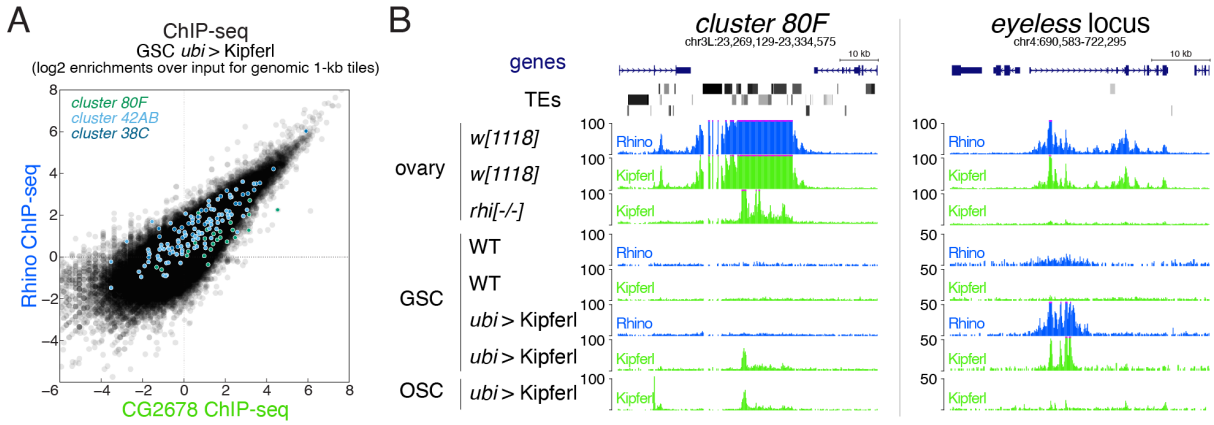


Figure 66: Overexpressed Kipferl co-occupies all Rhino-marked domains in cultured GSCs. (A) Scatter plot comparing the log2 ChIP-seq enrichment for Rhino and Kipferl in cultured GSCs upon overexpression of Kipferl. (B) UCSC genome browser tracks depicting the indicated ChIP-seq signal at piRNA cluster 80F and the eyeless locus as coverage per million reads.

granting access to Kipferl in the ovary. The lack of Rhino recruitment despite Kipferl binding upon overexpression in GSCs might on the other hand be caused by competition with HP1a. Further experiments will be required to investigate the establishment of Rhino at cluster *80F*.

Despite the absence of Kipferl in wildtype GSCs, the Rhino pattern in these cells does not resemble the pattern observed in *kipferl* mutant ovaries. Instead of strong enrichment at few piRNA clusters and at Satellites, Rhino is more evenly distributed in shallow enrichments along chromosome arms and, to a lesser extent, in heterochromatin (Figure 67). This implies either the presence of alternative guidance and stability factors in GSCs, or the absence of Rhino sequestration to Satellite regions. In the latter case, empowered by higher protein levels, Rhino might simply compete with HP1a for H3K9-methylated chromatin. Indeed, Rhino ChIP-seq experiments upon siRNA-mediated HP1a depletion in GSCs showed a marked reorganization for Rhino (Figure 68A). Several regions in pericentromeric heterochromatin previously bound by HP1a were newly occupied by Rhino. In turn, Rhino levels at several

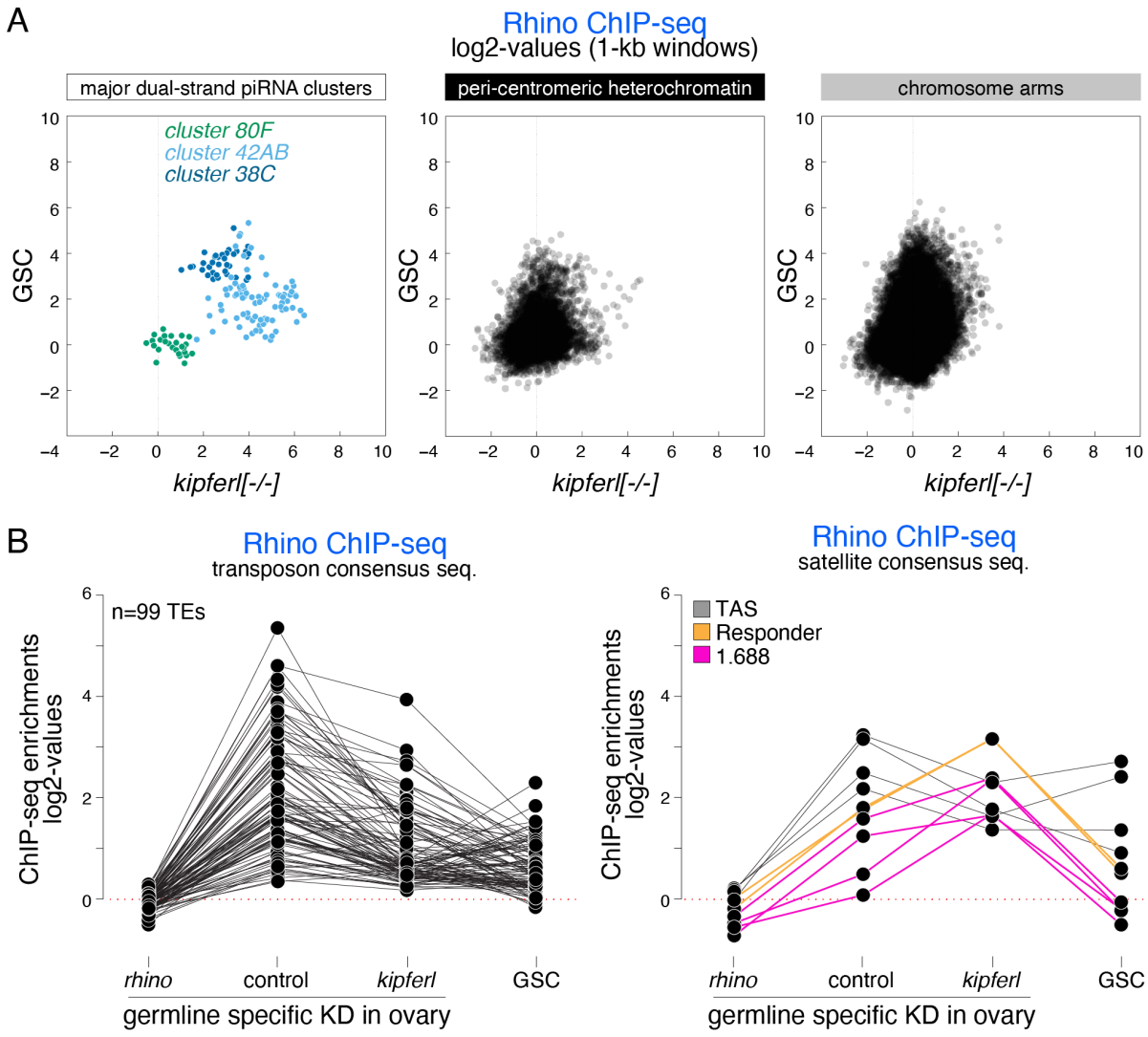


Figure 67: The Rhino pattern in cultured GSCs does not resemble the *kipferl* mutant Rhino pattern. (A) Scatter plot comparing average log2 ChIP-seq enrichments for Rhino in ovaries from *kipferl* mutants versus cultures GSCs (1-kb tiles separated into pericentromeric heterochromatin, chromosomal arms, and piRNA clusters.) (B) Jitter plots depicting the log2-fold Rhino ChIP-seq enrichments on transposon (left) and Satellite (right) consensus sequences in indicated genetic backgrounds.

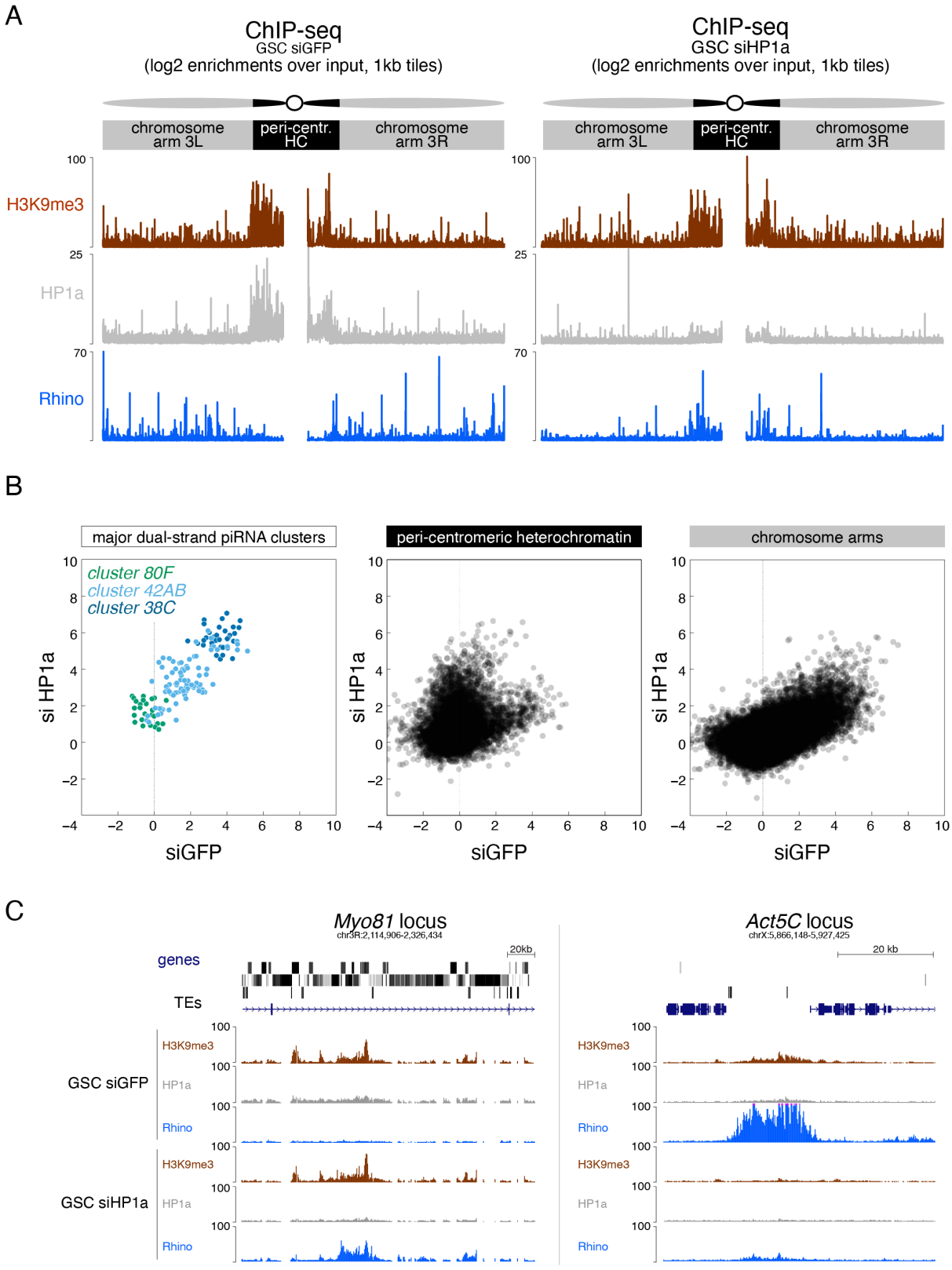


Figure 68: HP1a blocks access of Rhino to H3K9me3 in cultured GSCs. (A) ChIP-seq enrichment (genome unique reads; 1-kb tiles, one representative replicate each) of H3K9me3, HP1a, and Rhino along the assembled chromosome 3 sequence in control (siGFP) and HP1a-depleted cultured GSCs. Pericentromeric heterochromatin and euchromatic chromosome arms are indicated. (B) Scatter plot comparing average log2 ChIP-seq enrichments for Rhino in control versus HP1a-depleted cultured GSCs (1-kb tiles separated into pericentromeric heterochromatin and chromosomal arms; piRNA clusters *38C, 42AB,* and *80F* are shown separately). (C) UCSC genome browser tracks showing the indicated ChIP-seq signal as coverage per million reads at example regions where Rhino occupancy increased (left) or decreases (right) in response to HP1a depletion.

euchromatic Rhino binding sites was reduced (Figure 68B, C). These sites often showed a reduction of H3K9me3 upon HP1 depletion or were characterized by only low levels of H3K9me3 in both control and HP1a-depleted GSCs (Figure 68C, Figure 70). Rhino binding at these sites in wildtype cells might be mediated by H3K9me2, which we did not characterize in GSCs, or due to other factors giving Rhino a competitive advantage against HP1a.

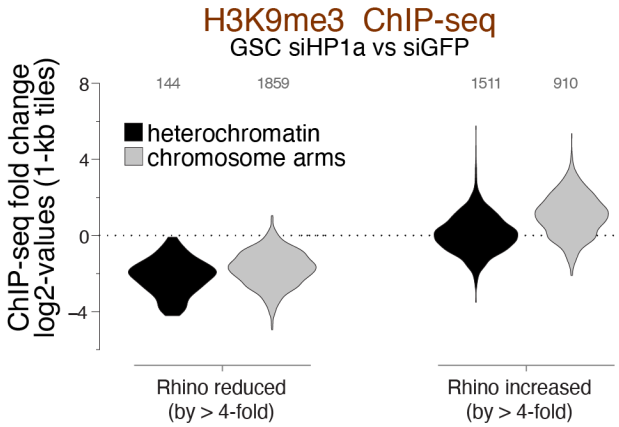


Figure 70: Changes in Rhino levels are often accompanied by changes of H3K9me3 enrichment levels in HP1a-depleted cultured GSCs. Violin plot showing the log2 fold change in H3K9me3 ChIP-seq signal between HP1a-depleted cultured GSCs and control. 1-kb tiles are shown that display at least 4-fold reduction or increase in Rhino binding in either pericentromeric heterochromatin or euchromatic chromosome arms.

Importantly, the drastically rearranged Rhino chromatin binding profile in HP1a-depleted GSCs still did not resemble the ovarian ChIP-seq profile focused around Kipferl peaks (Figure 69). This once again supported the model that Kipferl is required as a sequence specific recruitment factor for Rhino to establish the stereotypical Rhino landscape of the ovary, also in the absence of competing HP1a.

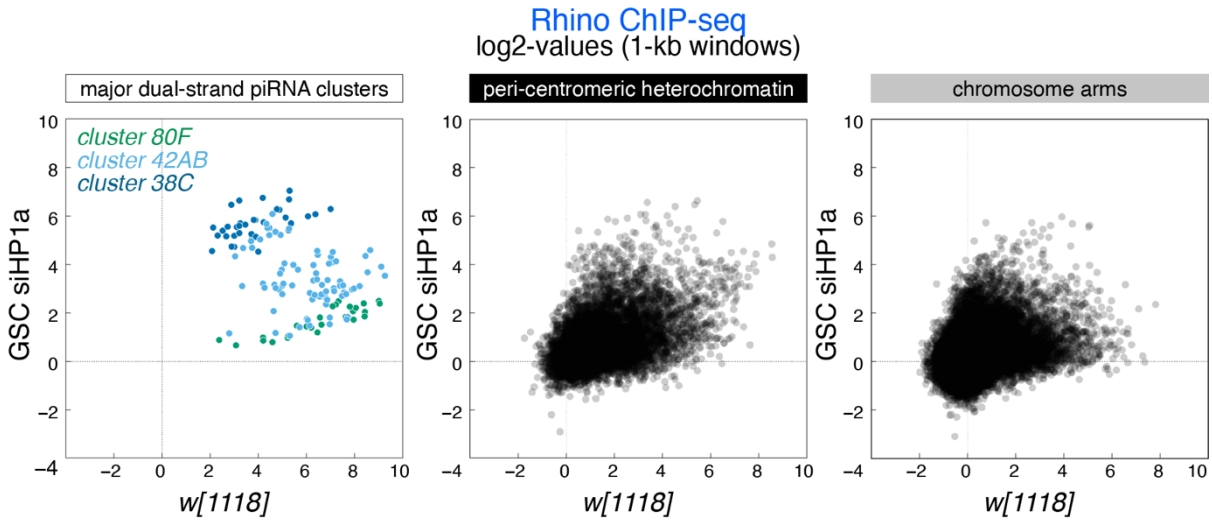


Figure 69: Rhino in HP1a-depleted cultured GSCs does not mimic the ovarian Rhino patten. Scatter plot comparing average log2 ChIP-seq enrichments for Rhino in HP1a-depleted cultured GSCs versus w^{1118} ovaries (1-kb tiles separated into pericentromeric heterochromatin and chromosomal arms; piRNA clusters *38C, 42AB,* and *80F* are shown separately).

Our findings demonstrate that Rhino, in line with the described affinity of its chromo domain to H3K9me3, is capable of binding to general H3K9me2/3-marked regions in the absence of factors actively recruiting it elsewhere. The expression of Kipferl in GSCs was sufficient to induce the formation of Rhino domains at several Kipferl binding sites, further strengthening the model of sequence specific Rhino recruitment via Kipferl.

2.5 Alternative targeting mechanisms for Rhino

Several observations point to the existence of additional guidance factors acting at different sites for Rhino recruitment: Strong remaining Rhino enrichments at piRNA clusters *42AB* and *38C* in the absence of Kipferl imply alternative mechanisms for the establishment of a Rhino domain, independently of Kipferl. The same is true at telomeric HTT and TAS repeats and the adjoining genome-unique flanking regions where Rhino is maintained in the absence of Kipferl. The massive accumulations of Rhino at *Rsp* and *1.688* family Satellites likely also relies on dedicated recruitment factors that allow the stunning sequestration of Rhino to these repeats. In addition to these Kipferl-independent Rhino recruitment mechanisms, our experimental data also indicate the likely existence of Rhino domains that are nucleated by alternative recruitment factors that require the stabilizing function of Kipferl for the stable formation of a larger domain. The following section summarizes preliminary data relating to several potential alternative recruitment routes for Rhino.

2.5.1 Transcription plus heterochromatin

The function of the extended hinge domain that is unique to Rhino is currently unclear. Our experiments indicate that not just the composition, but also the sequence of the hinge is important for proper formation of large Rhino domains (Figure 21). Moreover, the comparison of the hinge length across different HP1 proteins, as well as different *Drosophila* species, indicates that a long hinge is a conserved feature of Rhino proteins, with many species harboring a longer hinge than *melanogaster* Rhino (Figure 71).

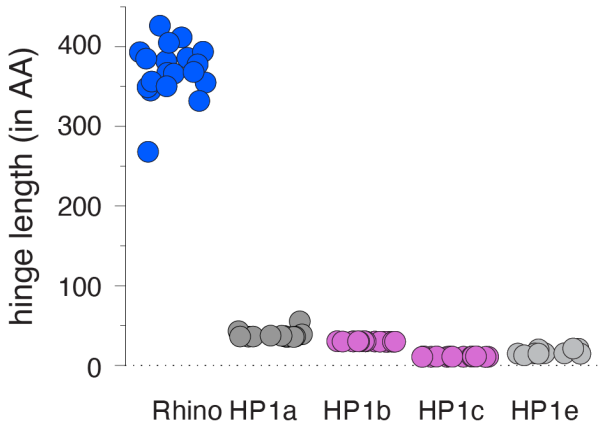


Figure 71: Rhino's hinge is much longer than that of other HP1 proteins. Jitter plot showing the distribution of hinge length for different HP1 protein variants across *Drosophila* species. The outlier towards a shorter hinge length in Rhino proteins corresponds to the *melanogaster* protein.

One intriguing possible explanation for why a long hinge is required for Rhino lies in the chromatin environment at piRNA clusters: While repressed constitutive heterochromatin is very compact and densely packed, ongoing active transcription at sites occupied by Rhino likely requires (or causes) the opening of chromatin. Thereby, the internucleosomal distance might be increased, requiring a longer hinge domain to bridge neighboring nucleosomes.

Rhino would thereby have a competitive advantage over HP1a in binding to transcribed heterochromatic regions.

Curiously, previous work has shown that piRNA clusters *38C* and *42AB*, which maintain Rhino binding in the absence of Kipferl, harbor transposon-derived promoters allowing their Moonshiner-independent transcription. To test if the combination of heterochromatin with transposon-based transcription was sufficient to establish Rhino at *38C* and *42AB*, we introduced the *kipferl* mutation into flies harboring deletions for either the two promoters flanking cluster *38C* or the promoter driving the expression of *42AB*. Neither promoter deletion had an effect on Rhino occupancy at these two clusters, arguing against a simple model of Rhino binding at specialized, transcribed heterochromatin (Figure 72).

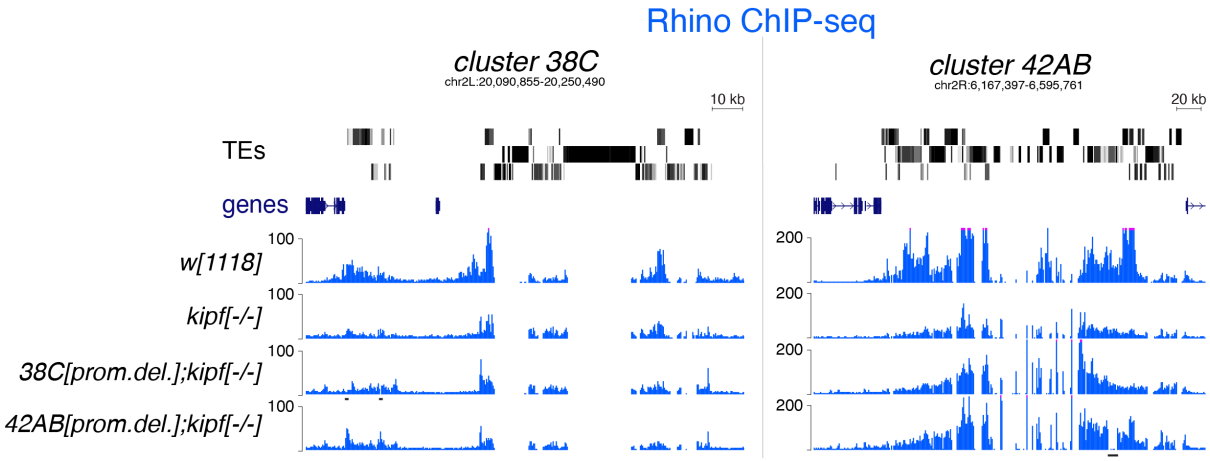


Figure 72: Deletion of cluster promoters does not influence the Rhino binding pattern in *kipferl* mutant ovaries. UCSC genome browser tracks showing Rhino ChIP-seq signal in the indicated genotypes at Kipferl-independent piRNA clusters *38C* and *42AB* as coverage per million reads. Black bars below the respective track indicate location of promoter deletions.

2.5.2 Piwi-mediated Rhino deposition

Rhino binding at standalone transposon insertions is lost upon depletion of Piwi through germline-specific RNAi. At large piRNA clusters, however, Rhino remains stably bound upon Piwi depletion (Mohn et al., 2014). This implied more hard-wired specificity factors acting at big piRNA clusters. In the case of piRNA cluster *80F*, we found that Kipferl acted as the required specificity factor. Intrigued by the observation that the large piRNA clusters *42AB* and *38C* are also largely refractory to loss of Rhino upon Kipferl depletion, we wondered if a combination of Piwi-mediated Rhino specification with Kipferl might be acting in a redundant fashion at these sites. To test the dependency of Rhino on Piwi in the absence of Kipferl, we simultaneously knocked down both Piwi and Kipferl in the germline through Gal4-mediated expression of a double hairpin construct. Interestingly, while neither single depletion of Piwi nor Kipferl affected ovary morphology, simultaneous depletion of both proteins via MTD-Gal4 resulted in highly atrophic ovaries. We observed the same phenotype in a combined knock down of Rhino and Piwi. This indicated that loss of multiple pathway components leads to a more complete collapse of the piRNA pathway system. We overcame this complication

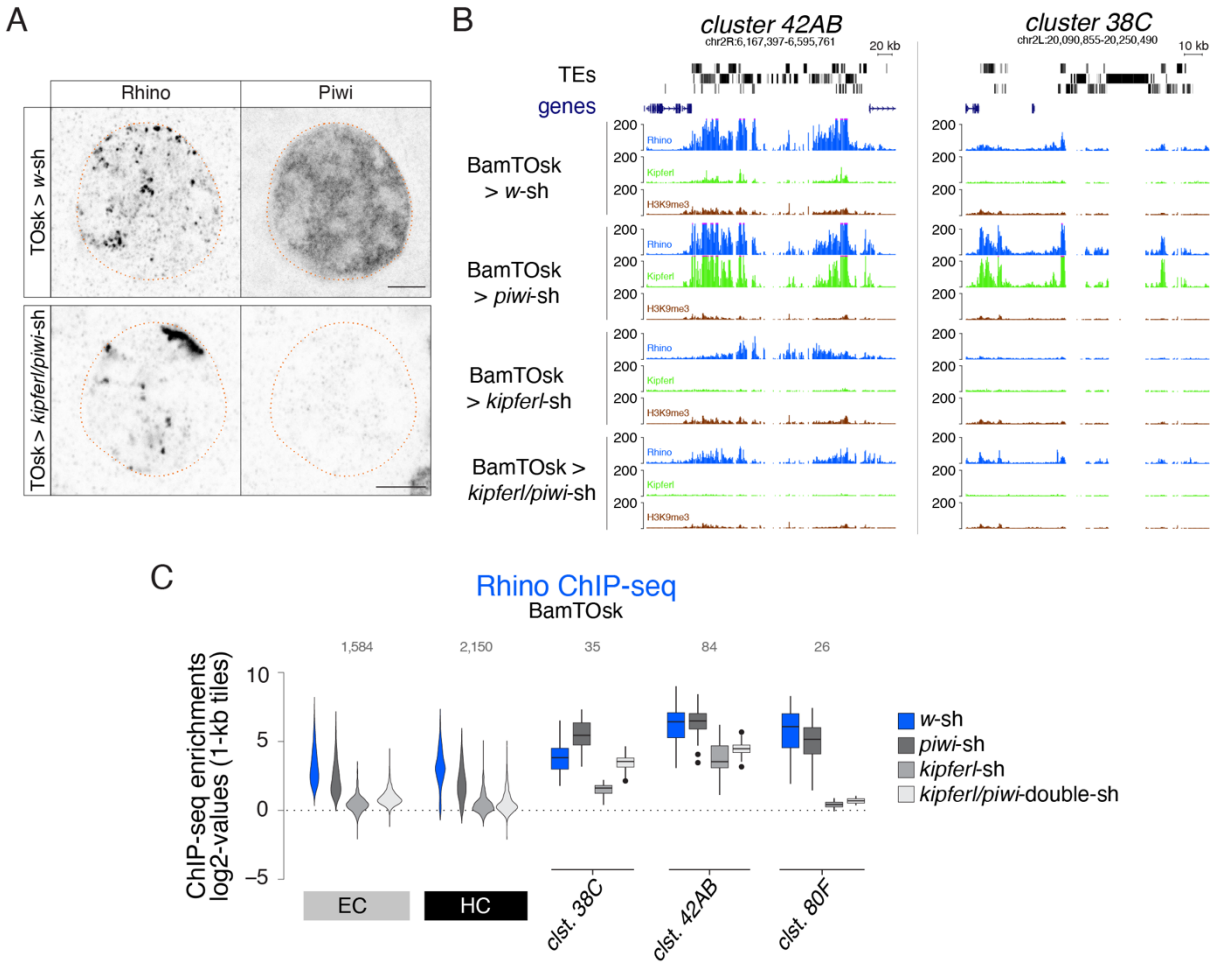


Figure 73: Double depletion of Kipferl and Piwi does not lead to a complete collapse of piRNA clusters *38C* and *42AB*. (A) Confocal microscopy images depicting the distribution and levels of Rhino and Piwi in control and kipferl-piwi-double knock down nurse cell nuclei. (B) UCSC genome browser tracks showing Rhino ChIP-seq signal in the indicated genotypes at Kipferl-independent piRNA clusters *38C* and *42AB* as coverage per million reads. (C) Violin plot depicting the Rhino ChIP-seq enrichment levels for Rhino-bound tiles (defined as previously in w^{1118} ovaries) in the indicated knock downs

through the use of a late germline driver system which does not affect germline stem cells and allows the formation of morphologically normal ovaries (TOsk ElMaghraby et al., 2022). Knock down efficiency for both Kipferl and Piwi was complete after first stage egg chambers, as evidenced by immunofluorescence staining for Piwi and ChIP-seq for Kipferl (Figure 73A, B).

The Kipferl phenotype with characteristic enrichment of Rhino at the nuclear envelope was observed also in the double knock down, and Rhino was broadly depleted from chromatin (Figure 73A, C). Rhino binding at clusters *42AB* and *38C* was largely maintained, although in a slightly altered pattern (Figure 73B). The highly repetitive center of cluster *42AB*, which was most strongly bound by Rhino upon Kipferl depletion, showed a decrease in Rhino binding in the double knock down with Piwi, while the more peripheral regions of the cluster showed increased Rhino binding. H3K9me3 levels at the clusters did not change more than in the Piwi depletion alone, indicating that chromatin changes are unlikely to be responsible for these reorganizations (Figure 73B). Co-depletion of Kipferl and Piwi lead to a reduction of H3K9me3 levels at Satellite repeats as compared to Kipferl depletion alone, which were accompanied

by a reduction of Rhino levels at *1.688* family repeats, but not at the Responder Satellites, indicating that different mechanisms might be at play for recruitment of Rhino to different classes of Satellites (Figure 74).

Our collective data indicate that the role of Piwi at single insertions might be limited to the deposition of H3K9me2/3 which, when in proximity of a nucleation site in the form of a binding site for Kipferl or alternative factors, allows the formation of a Rhino domain at selected insertions. This would offer an explanation for why not all Piwi-targeted single TE insertions are bound by Rhino. We can, however, not exclude that additional or alternative specificity factors are deposited by through Piwi-mediated transcriptional gene silencing.

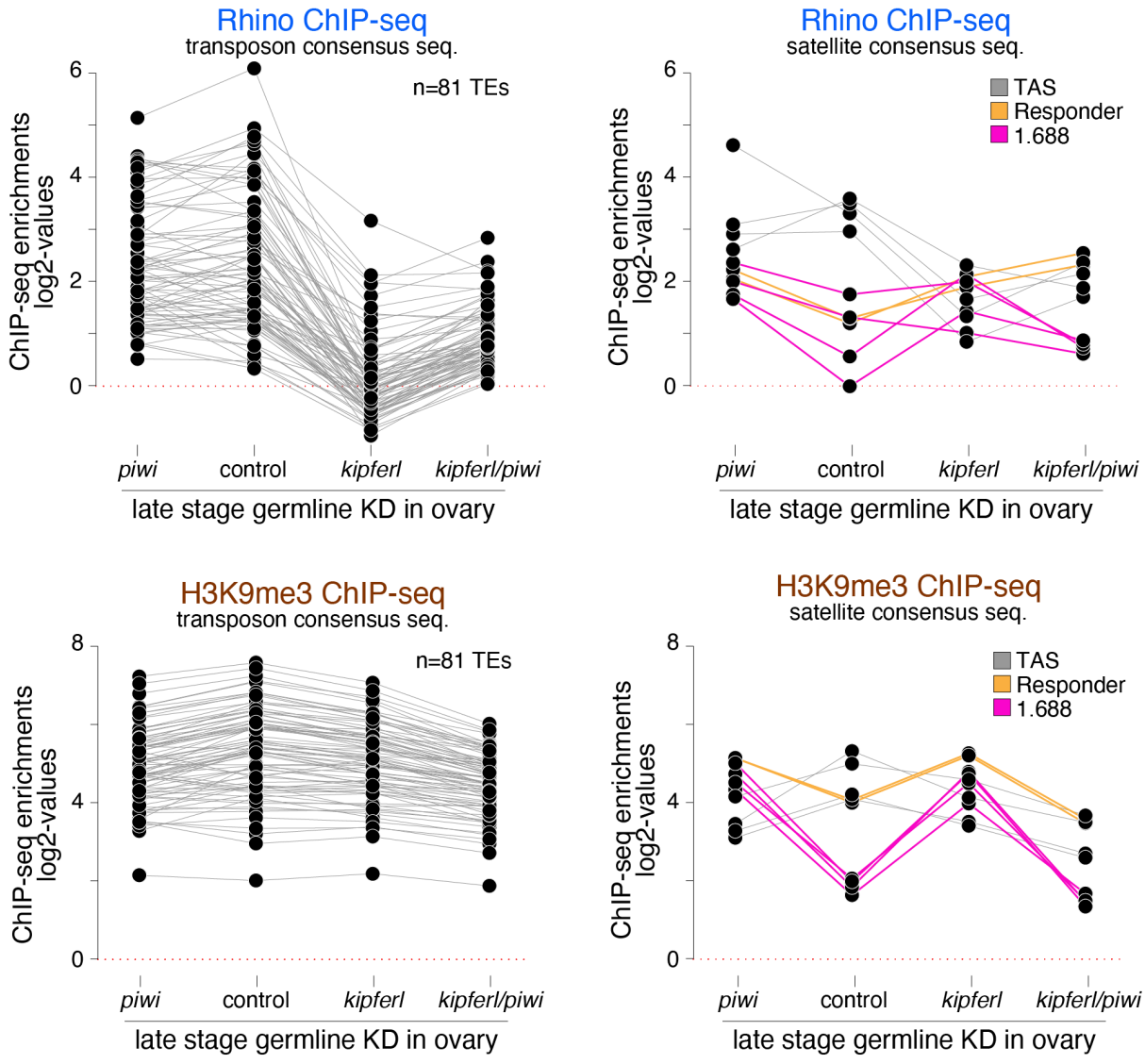


Figure 74: *Rsp* and *1.688* Satellites respond differently to double depletion of Kipferl and Piwi. Jitter plots depicting the log2-fold Rhino and H3K9me3 ChIP-seq enrichments on Rhino-dependent transposon (left) and Satellite (right) consensus sequences in indicated genetic backgrounds.

2.5.3 The ZAD-ZnF protein CG8388 in Rhino specification

Additional, alternative Rhino specificity factors could take many shapes, but might share common features or protein architecture with Kipferl. Besides Kipferl, one more uncharacterized ZAD ZnF protein, CG8388, was identified in our yeast two-hybrid screen as a direct interactor of Rhino (Figure 75A). *CG8388* is expressed at low levels in ovaries and reaches higher expression levels in testes (Figure 75B, (Larkin et al., 2021)). A yeast two-hybrid screen testing Rhino against a library of cDNA fragments obtained from testes was also able to recover CG8388 as a direct interactor of Rhino, further strengthening a connection between the two proteins (Supplementary Table 3).

Depletion of *CG8388* via transgenic RNAi does not lead to an obvious change in Rhino's nuclear distribution, but introduction of a frame shift mutation in *CG8388* led to the accumulation of Rhino at the nuclear envelope in a pattern distinct from the Kipferl phenotype (Figure 75C). Instead of concentrating in large, compact accumulations, Rhino was distributed evenly along the nuclear envelope. We hypothesized that the even distribution in small foci indicates that Rhino remains bound to a large number of loci. ChIP-seq analysis of

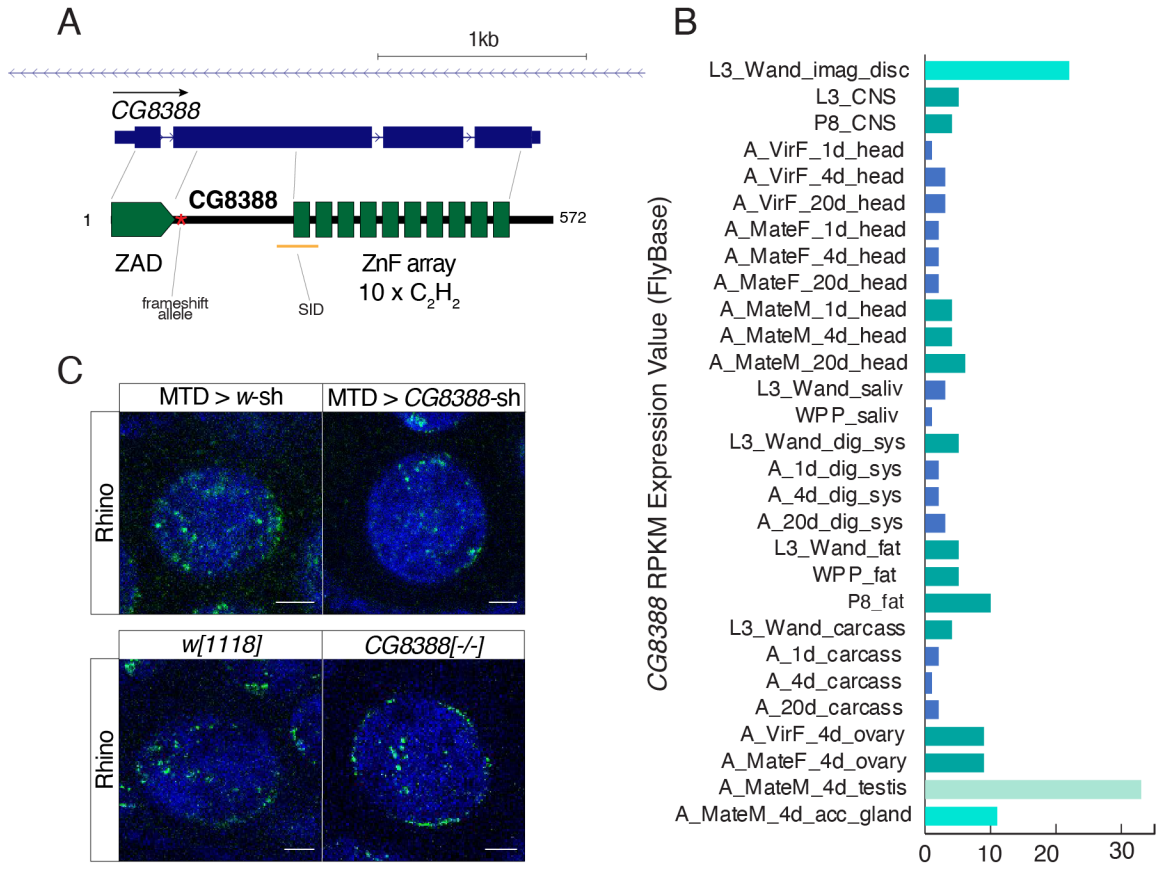


Figure 75: The ZAD ZnF protein CG8388 is expressed in ovaries and testes and influences Rhino's localization pattern. (A) Genomic *CG8388* locus depicting the annotated transcript and CG8388 protein domain architecture. Location of the frame shift mutation (red asterisk) and smallest interacting domain with Rhino (yellow bar, as determined in the yeast two-hybrid screen) are indicated. (B) Bar graph indicating the expression pattern of *CG8388* (Flybase modENCODE Anatomy RNA-Seq; L: larva; P: pupa; A: adult; M: male; F: female). (C) Confocal microscopy images showing the localization of Rhino in representative nurse cell nuclei of ovaries where *CG8388* was depleted with MTD-Gal4-mediated germline knock down (top) or mutated as indicated in panel A (bottom) with the respective controls.

Rhino in the absence of *CG8388* indeed showed that Rhino remained bound at most sites, with merely 30 % of Rhino-bound genomic tiles losing Rhino binding in knock down or frame shift mutants compared to the respective controls. The majority of tiles affected by loss of CG8388 is also dependent on Kipferl, and the reduction in Rhino ChIP-seq enrichment upon depletion of *CG8388* was weaker than we had observed upon loss of Kipferl (Figure 76A, B). The CG8388-dependent sites were not particularly enriched in sites lacking intrinsic, Rhino-independent Kipferl nucleation sites. Importantly, at sites where Rhino was lost in *CG8388* mutant ovaries, Rhino-independent Kipferl binding was maintained (Figure 76C). On a cautionary note, we find very little overlap between Rhino bound genomic 1-kb tiles affected by MTD-Gal4-mediated *CG8388* knock down and the *CG8388* frame shift mutation (Figure 76D). As a CG8388 antibody is not currently available, we are not able to assess the strength of the *CG8388* knock down or the quality of the *CG8388* frame shift alleles, which will be required for conclusive future studies of this protein.

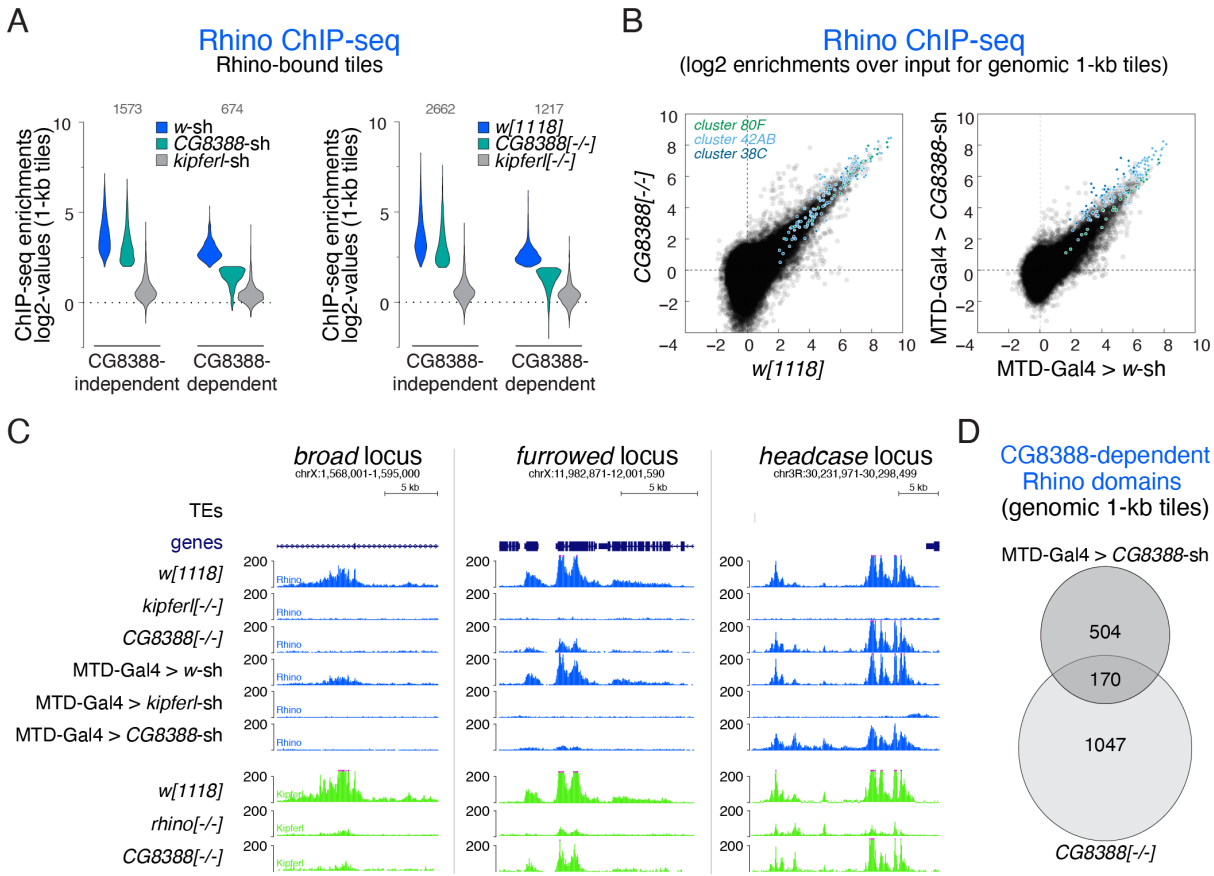


Figure 76: Loss of CG8388 affects the Rhino ChIP-seq pattern in ovaries. (A) Violin plot depicting Rhino ChIP-seq levels at Rhino bound 1-kb tiles (determined in the respective wildtype background by 4-fold Rhino enrichment in two independent samples) in the indicated depletion or frame shift genotypes. Tiles are divided into two groups based on their dependency on CG8388 for Rhino binding in the respective genetic background (binary cutoff at 4-fold log2 enrichment remaining in the absence of CG8388). (B) Scatter plot comparing the Rhino ChIP-seq enrichments in the indicated genotypes. (C) UCSC genome browser tracks showing Rhino and Kipferl ChIP-seq signal in the indicated genotypes as coverage per million reads. (D) Venn diagram depicting the overlap between CG8388-dependent 1-kb tiles found via MTD-Gal4-mediated depletion or frame shift mutation, respectively.

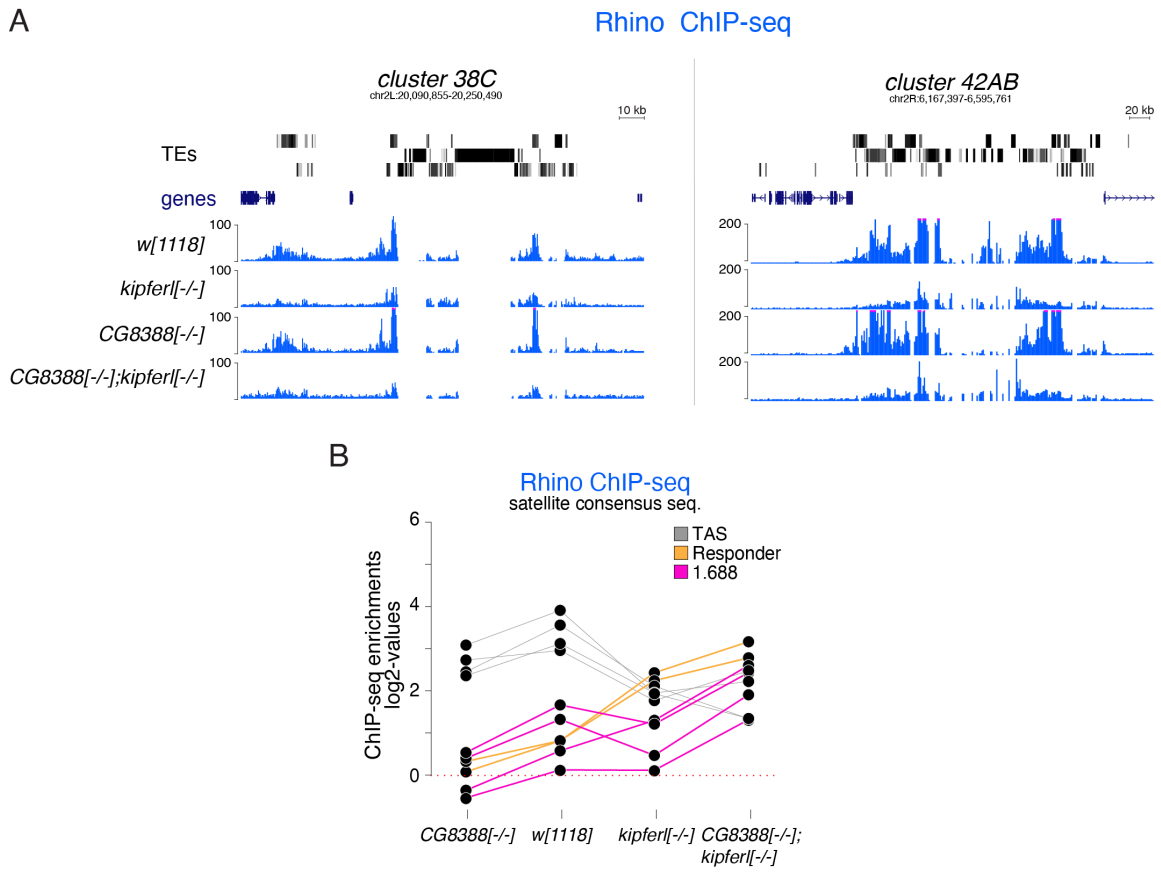


Figure 77: Rhino binding at large piRNA clusters and Satellite repeats in maintained in the double mutant of Kipferl and CG8388. (A) UCSC genome browser tracks depicting Rhino ChIP-seq signal in the indicated genotypes at Kipferl-independent piRNA clusters *38C* and *42AB* as coverage per million reads. (B) Jitter plot showing log2 Rhino ChIP-seq enrichment levels at Satellite consensus sequences in the indicated genotypes.

The large piRNA clusters *42AB, 38C*, and *80F* retained Rhino binding in the absence of CG8388 (Figure 76B). To investigate potential redundancy of CG8388 and Kipferl at *42AB* and *38C*, we combined both full mutants. Rhino levels were slightly reduced at *42AB* and *38C*, but remained strong in the double mutant, indicating that CG8388 does not play a major role in the definition of these two clusters as Rhino domains (Figure 77A). Moreover, Rhino remained strongly bound to Satellite repeats of the *Rsp*, *1.688*, and TAS families in the double mutant. These data indicate that CG8388 shares a redundant role with Kipferl at several genomic sites, but does not act as a guidance factor for Rhino at Satellites of large clusters (Figure 77B). We were unable to express a tagged version of CG8388 in ovaries, precluding an analysis of the native chromatin binding pattern of this protein and its localization in nurse cell nuclei.

2.5.4 Sequestration to Satellites

The strong effect of Kipferl depletion, leaving only the two largest piRNA clusters and large Satellite accumulations intact as Rhino domains, was surprising, given the sequence diversity of all Rhino domains. The dual function of Kipferl as both a recruiter and a stabilizer of Rhino, together with the absence of clear Kipferl peaks at many Kipferl-dependent Rhino domains,

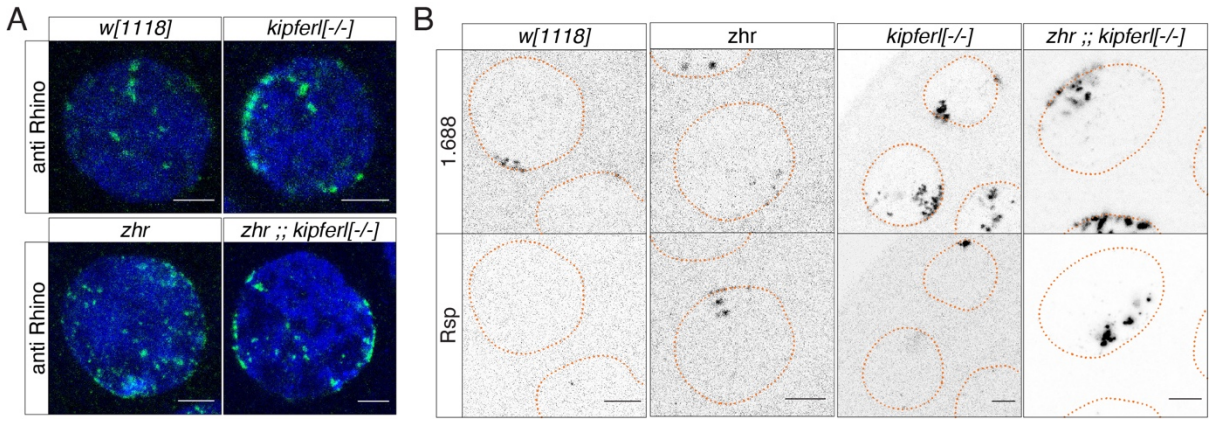


Figure 78: The *zhr* allele has no strong influence on Rhino's immunofluorescence localization pattern. (A) Confocal microscopy images showing the localization of endogenous Rhino in nurse cell nuclei of the respective genotypes. Scale bar: 5 µm. (B) Confocal images showing *Rsp* and *1.688* Satellite RNA FISH signal in nurse cell nuclei of the indicated genotypes. Scale bar: 5 µm. Orange dotted line: nuclear outline based on DAPI.

would support a model where several Rhino binding sites rely on Kipferl not for nucleation of a Rhino domain, but simply for stabilization of Rhino to prevent its sequestration to Satellite repeats. In this scenario, only the strongest Rhino domains would manage to maintain Rhino binding. We wondered if release of large amounts of Rhino from Satellites would allow smaller domains to re-gain Rhino. To this end, we combined the *kipferl* deletion allele with the *zhr* allele, harboring a deletion of the several Mb long array of *1.688* Satellites of the X chromosome, arguably the largest continuous Rhino-bound region in *kipferl* mutant flies (Ferree and Barbash, 2009). Immunofluorescence staining for Rhino in these flies shows large Rhino accumulations at the nuclear envelope, reminiscent of the *kipferl* deletion phenotype. In agreement with the fact that *1.688* family arrays on other chromosomes, as well as the Responder arrays are not affected in these flies, RNA FISH confirmed that both Satellites still

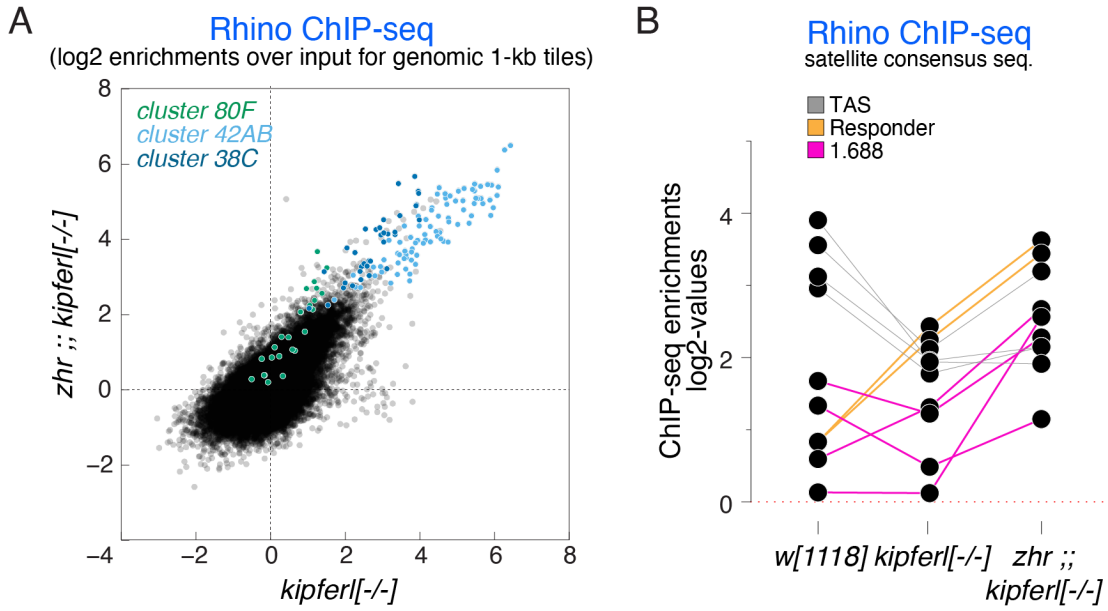


Figure 79: Rhino binding is not recovered at non-Satellite binding sites in the presence of the *zhr* allele. (A) Scatter plot comparing the Rhino ChIP-seq enrichment at genomic 1-kb tiles in the indicated genotypes. (B) Jitter plot depicting the log2 Rhino enrichment at Satellite consensus sequences in the indicated genotypes.

accumulated near the nuclear envelope and that their expression was strongly increased in the absence of Kipferl (Figure 78). ChIP-seq analysis in flies harboring both the *zhr* allele and the *kipferl* deletion allele showed no recovery of Rhino at Kipferl-dependent sites (Figure 79A). Instead, we found an additional increase in Rhino ChIP-seq enrichment at *Rsp* consensus sequences, and to a lesser extent at elements of the *1.688* family of Satellites (Figure 79B). This was in line with the strong RNA FISH signal observed for both of these elements (Figure 78B). These data indicate that the removal of several Mega-bases of *1.688* repeat sequences is not sufficient to release Rhino back to its wildtype binding sites. Whether this is due to strong compensatory sequestration of the remaining Satellite sequences or indicates the direct requirement for Kipferl to stabilize Rhino at chromatin cannot be distinguished based on the current data.

2.6 Importance of Rhino at its various binding sites

The organismal phenotype regarding female fertility of Kipferl's loss of function is much weaker than it is observed for most piRNA pathway factors. Unlike Rhino, loss of Kipferl does not lead to complete sterility. Homozygous *kipferl* mutant stocks are viable for dozens of generations, although viability issues arose with higher frequency for newly generated rescue stocks with increasing age of the parental *kipferl* deletion stock. Different aspects of the *kipferl* mutant phenotype might contribute to the residual fertility of these flies, and the following section summarizes preliminary investigations attempting to separate them.

2.6.1 Rhino at piRNA clusters

The discrepancy in fertility defect between mutants for *kipferl* versus other piRNA pathway proteins might stem from the lower level of transposon reactivation in these flies. This might be achieved through maintained piRNA production from clusters *42AB* and *38C*, allowing the continued silencing of many TEs in *kipferl* mutant ovaries. However, many elements reactivated upon loss of Rhino are not primarily controlled by the large piRNA clusters, putting this hypothesis into question. The deletion of clusters *42AB* and *38C*, both separate and in combination, was recently shown not to impact female fertility (Gebert et al., 2021). A combination of these cluster deletions with the *kipferl* deletion allele had a clear impact on female fertility, but did not lead to complete sterility (Figure 80). Instead, fertility remained at levels similar to those observed in *kipferl* mutant females, arguing against a strong role of the remaining piRNAs derived from *42AB* and *38C* in maintenance of fertility.

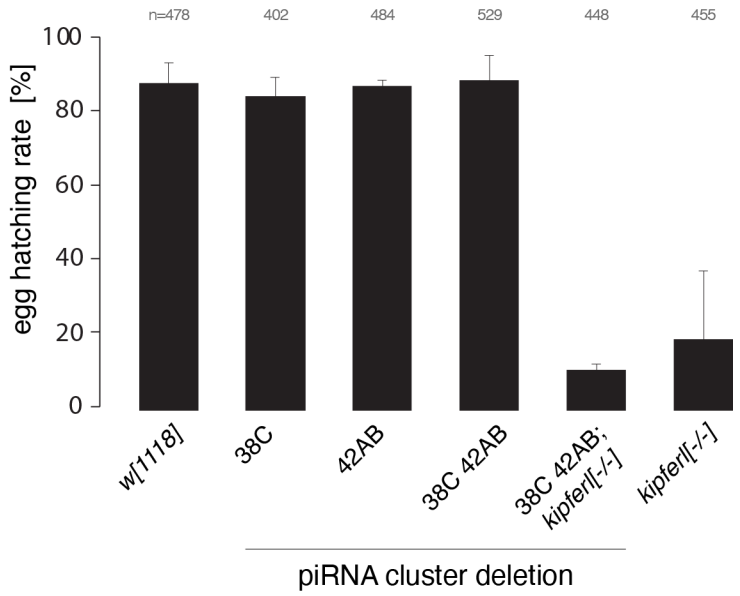


Figure 80: Cluster deletions in combination with the *kipferl* mutation do not lead to complete sterility. Bar graph depicting the hatching rate of eggs laid by females of the indicated genotypes. The total number of eggs is given as n.

2.6.2 Rhino in early versus late stages of oogenesis

An alternative explanation might lay in the developmental regulation of Kipferl, with strong expression levels detected only in differentiated cells of the adult ovary. Kipferl depletion thereby does not affect early stages of germ cell development, where important silencing marks might be deposited at transposon loci that could maintain silencing of most elements even in the absence of ongoing Rhino-dependent piRNA production.

Expression of Rhino through the maternal tubulin promoter allowed for an almost perfectly wildtype-like Rhino localization pattern (Figure 17) and full female fertility, which argued against a crucial role of early pools of Rhino in the germline stem cells (Figure 81A). To investigate if intact remaining Rhino in early developmental stages of the ovary was sufficient to allow a certain degree of transposon silencing and fertility, we made use of RNAi-mediated

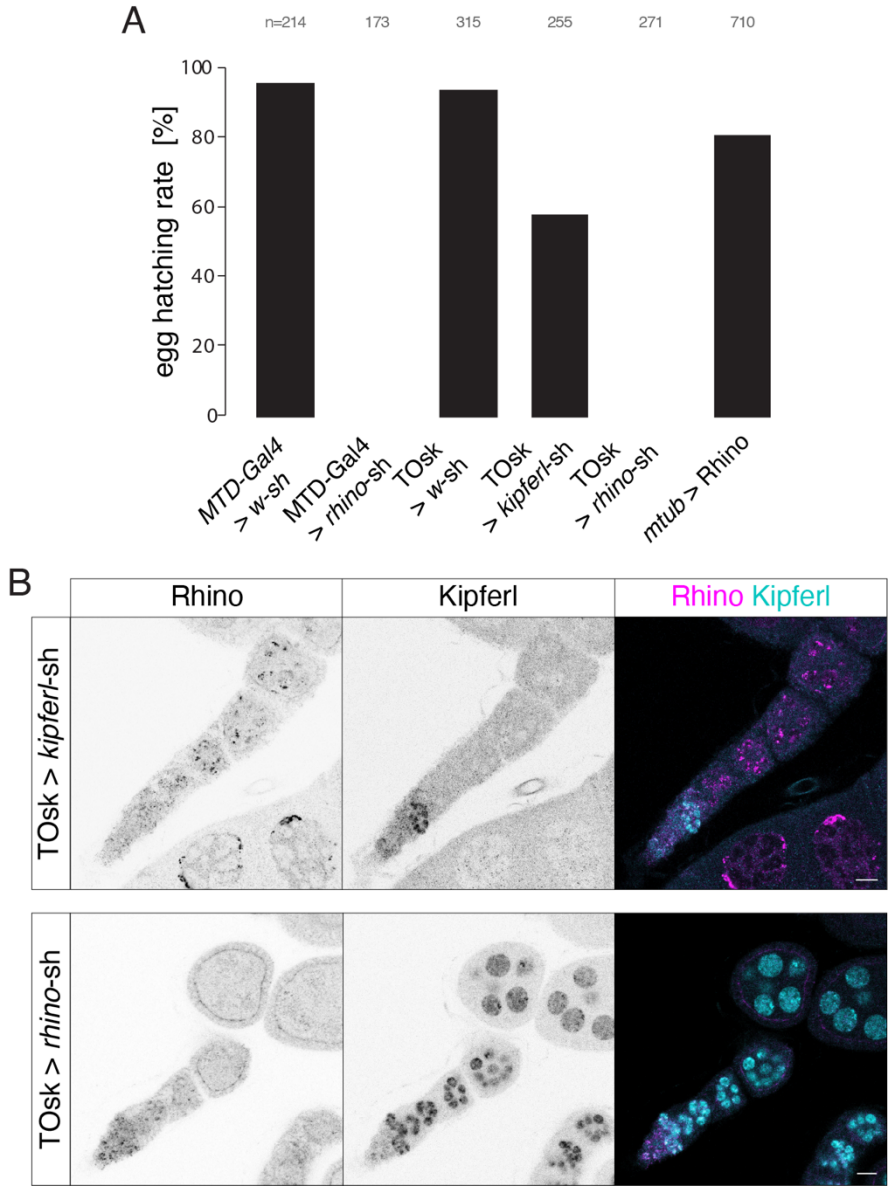


Figure 81: Early Rhino is neither required, nor sufficient for fertility. *(A)* Bar graph depicting the hatching rate of eggs laid by females of the indicated genotypes. The total number of eggs is given as n. (B) Confocal images depicting the signal for Rhino and Kipferl in indicated knock down conditions in the germarium. Scale bar: 10 µm.

depletion via TOsk, which sets on shortly after Kipferl levels increase in developing germaria (ElMaghraby et al., 2022). Depletion of Rhino with TOsk left a narrow window of co-expression of Kipferl with Rhino in early cysts, with otherwise complete depletion of Rhino. (Figure 81B) This indicated that TOsk-mediated Rhino depletion adequately mimics the remaining unperturbed early Rhino pool existing in *kipferl* mutant ovaries. Fertility of these flies was fully ablated, indicating that late Rhino was required to sustain the remaining fertility observed in *kipferl* mutant flies (Figure 81A). Of note, RNA FISH for the *mdg3* element known to be regulated in early stages of oogenesis confirmed that TOsk-mediated Rhino depletion did not lead to upregulation of this transposon (ElMaghraby et al., 2022). Late rescue of a *rhino* mutant via a transgenic copy of Rhino controlled by the maternal tubulin promoter, however, was not able to re-establish silencing of *mdg3* expression (Figure 82). Interestingly, this did not impact female fertility, arguing that Rhino might contribute to fertility independently of TE silencing.

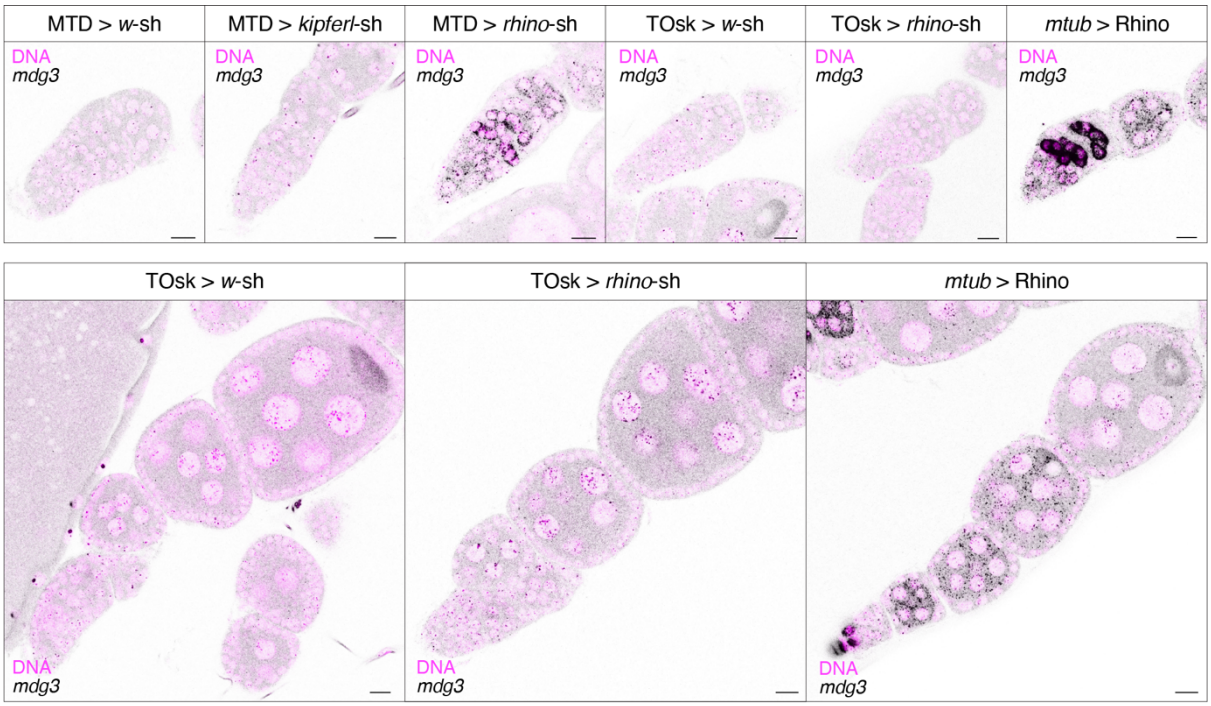


Figure 82: Incomplete repression of transposons is not incompatible with female fertility. Confocal images showing RNA FISH signal of transcripts of the *mdg3* transposon in germaria of the indicated genotypes (top). Whole ovarioles are depicted for genotypes with late depletion or expression of Rhino (scale bar: 10 µm; purple signal: DAPI).

2.6.3 Rhino at Satellites

Finally, while transposon derepression has been previously demonstrated to be causative for developmental arrest due to DNA damage (Klattenhoff et al., 2007), the data presented in this work demonstrates that transposons are not the only repetitive species under the control of piRNAs. The maintained production of piRNAs from Satellites, as well as continued telomere maintenance, might crucially contributed to the increased fitness of *kipferl* mutant flies as compared to full *rhino* mutants.

To investigate this, selective depletion of Rhino from Satellite repeats or telomeres would be required. In order to identify putative guidance factors recruiting Rhino to these regions that remain bound in the absence of Kipferl, we again made use of proximity mediated biotin ligation, combining the TurboID biotin ligase fused to a GFP nano-body with GFP-tagged Rhino in *kipferl* mutants. However, no significant interactors were identified besides the known piRNA cluster factors Deadlock, Cutoff, and Bootlegger. This hindered further investigations of Rhino's role in Satellite maintenance.

3. Discussion

The target spectrum of RNA interference pathways, which silence repetitive elements and aid genome maintenance from yeast to humans, is defined by the sequence of small RNAs guiding Argonaute proteins to all complementary RNA molecules. In the case of the piRNA pathway, a highly diverse set of piRNAs is loaded onto PIWI clade Argonaute proteins to enable the silencing of transposable elements in animal gonads. In several organisms, piRNAs have long been known to originate from so called piRNA clusters (Aravin et al., 2007a, Brennecke et al., 2007, Houwing et al., 2007). These accumulations of transposon sequences serve as heritable and adaptable storage loci for the sequence information required to recognize the large variety of elements that had previously invaded the genome. In the germline of *Drosophila melanogaster*, bi-directionally transcribed piRNA clusters are generally found at the boundary of pericentromeric heterochromatin and euchromatic chromosome arms, are marked by the heterochromatic chromatin mark H3K9me3, and are bound by the heterochromatin protein 1 (HP1) variant Rhino (Brennecke et al., 2007, Klattenhoff et al., 2009). Rhino fundamentally differs from canonical HP1a: instead of linking to repressive downstream factors, Rhino recruits effector proteins that allow the transcription, stabilization, and nuclear export of piRNA precursor RNAs at otherwise repressed heterochromatic loci. Rhino's exclusive binding to H3K9me3 at piRNA clusters is thought to be crucial both for the effective silencing of transposons through the production of an adequate set of piRNAs, but also for the restriction of these activating forces to the correct subset of heterochromatic loci. At the beginning of my doctoral studies, the mechanisms that define Rhino's chromatin binding specificity for piRNA clusters were unknown.

In my doctoral project I set out to understand Rhino's specific binding to piRNA clusters. To this end, I performed an in-depth characterization of Rhino's chromatin binding profile in different genomic backgrounds, identifying a previously undescribed diversity in Rhino domains, extending beyond the well-known piRNA clusters. I further assessed the role of individual protein domains of Rhino for its localization pattern, finding that all parts of the protein were required to allow the full extent of binding to various target sites. It has long been proposed that Rhino must rely on an unknown guidance factor to distinguish its binding sites from general heterochromatin. Using a yeast two-hybrid screen, supported by *in vivo* proximity ligation experiments, we identified a novel Rhino interactor, CG2678/Kipferl. Using fly genetics, imaging, and a variety of molecular biology approaches coupled to next generation sequencing, I uncovered a global redistribution of Rhino in the absence of Kipferl. Without Kipferl, Rhino was lost from most of its chromatin binding sites but accumulated at Satellite repeats. This was accompanied by a profound change in the composition of the ovarian piRNA pool, which ultimately led to the derepression of a hand full of transposable elements, as well as a reduction in female fertility. Importantly, I found that Kipferl binds to chromatin epistatically upstream of Rhino at a large number of Rhino domains genome wide. Kipferl binding sites are enriched in a short Guanosine-rich DNA motif, which matches the predicted specificity of Kipferl's Zinc finger arrays. These data led us to propose a model in

which Kipferl acts as the long elusive guidance factor for Rhino, binding to a consensus DNA motif and recruiting Rhino to its binding sites if they overlap with local heterochromatin (Baumgartner et al., 2022). Interestingly, a recent study by Chen et al. described that Rhino's chromatin binding pattern in *Drosophila* testes, where the piRNA pathway is also active, differs significantly from the ovarian Rhino pattern (Chen et al., 2021). This comes in further support of our model implying Kipferl as a guidance factor for Rhino in ovaries, as Kipferl is not expressed in testes and hence cannot contribute to Rhino's chromatin binding there. Similarly, we observed a drastically different Rhino profile in cultured germline stem cells (GSCs) of the ovary, where Kipferl is not active, and we were able to reproduce Kipferl-dependent Rhino binding in GSCs upon overexpression of Kipferl. We argue that the different chromatin binding profiles for Rhino across tissues and developmental stages, together with few Kipferl-independent Rhino domains in ovaries, imply that other guidance factors for Rhino likely exist.

My dissertation thus provides a first example of a putative DNA-binding protein acting as a specificity factor for the fast-evolving chromatin reader protein Rhino, implying DNA sequence in the specification of heterochromatin sub-compartments (Figure 83). In the following sections, I will discuss the various findings outlined above, as well as more nuanced interdependencies of Kipferl and Rhino, and the implications of my findings for the piRNA pathway field and beyond.

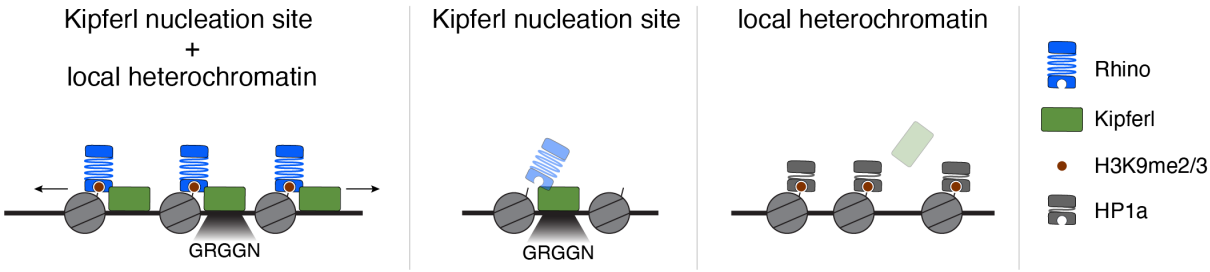


Figure 83: Updated model of piRNA cluster specification for Rhino binding. At sites where Kipferl binding is supported by its consensus GRGGN motif and H3K9me2/3 is present, Rhino is recruited and stabilized, followed by the spreading of both proteins to neighboring sequences. At Kipferl binding sites lacking heterochromatic marks Rhino cannot be stabilized. Heterochromatic regions lacking Kipferl binding sites are occupied by HP1a.

3.1 Diversity of Rhino domains beyond piRNA clusters

Rhino has been extensively described as a protein localizing to the large, dual-strand piRNA cluster of the *Drosophila* germline (Klattenhoff et al., 2007, Klattenhoff et al., 2009, Zhang et al., 2014, Mohn et al., 2014). Previous studies investigated piRNA producing loci genome wide, and found that besides the large piRNA clusters, Rhino also associates with single transposon insertions, where it induced piRNA production on both genomic strands (Mohn et al., 2014, Shpiz et al., 2014). A recent study by Gebert et al. found that large germline piRNA clusters are dispensable for transposon control and female fertility, pointing to an important

role of smaller Rhino domains considering the full female sterility of *rhino* mutants (Gebert et al., 2021).

To determine how Rhino distinguishes its binding sites from regions of general heterochromatin, we therefore first performed a complete and unbiased description of Rhino domains. Unlike previous studies, which had focused their characterization of Rhino on piRNA source loci, we relied solely on Rhino ChIP-seq enrichments for the definition of Rhino domains genome wide. Intrigued by the link between the Rhino profile and the transposon insertion landscape, which is known to be highly strain specific, we characterized Rhino's chromatin binding pattern in three different *Drosophila melanogaster* laboratory strains. This showed that the Rhino profile in ovaries was largely identical between strains. Remarkably, we find that Rhino domains were highly diverse in their size, underlying sequence content, and piRNA output. We found several domains that were strongly enriched in Rhino ChIP-seq signal across strains but showed no sign of transposon sequence content and produced only small amounts of piRNAs. Rhino recruitment at these domains, which had previously been missed by piRNA centered analyses, was hard to explain by maternally deposited piRNAs acting as Rhino specifiers, and implied the existence of additional, locus-specific recruitment mechanisms. Supported by the observation that these sites gave rise to piRNA precursors identified with Nxf3-RIP-seq experiments, we argued that these transposon-free Rhino domains are not to be discarded as experimental or biological artefacts but are functionally identical to piRNA clusters. The failure to produce significant amounts of mature piRNAs from these sites likely stems from the absence of complementarity to pre-existing piRNAs in these transcripts, which is required to initiate piRNA biogenesis (Mohn et al., 2015, Han et al., 2015). We argue that an approximation of Rhino's chromatin association profile based on sequenced piRNAs, which is often used for poorly accessible tissues, does not sufficiently depict the entirety of Rhino's binding profile, and that transposon-free Rhino domains may hold important clues for the investigation of Rhino specifying mechanisms. Indeed, we found that the prominent Rhino domain at the headcase locus, which lacks transposon content, is among the best examples of Rhino domains specified by Kipferl.

3.2 Rhino's protein domains

To understand how Rhino reads out putative specificity cues, we investigated the role of its individual protein domains in Rhino localization. We found that each protein domain of Rhino is required for full functionality of the protein, as evidenced both by female fertility defects and an impaired localization pattern upon replacement of the chromo domain, hinge, or chromoshadow domain with the respective equivalent part of HP1a (Figure 18). This is in agreement with analogous studies comparing other HP1 variant proteins, as well as the described signs of positive selection found in every domain of the Rhino protein (Smothers and Henikoff, 2001, Vermaak et al., 2005). The domain swap approach taken in this study had several limitations, precluding strong conclusions. The dimerization of HP1 proteins complicates the analysis of modified protein variants in the presence of wildtype protein.

Moreover, the effect of feedback mechanisms and diffuse multivalent interactions in both general HP1 biology and in the piRNA pathway cannot be neglected, and likely complicates the analysis of chimeric HP1a-Rhino domain swap constructs.

3.2.1 Chromo domain

The identification of Rhino's chromo domain as interaction point with Kipferl was foreshadowed by an important role of the chromo domain in initial domain swap constructs (Figure 21). Exchange of Rhino's chromo domain with the HP1a counterpart had a strong deleterious effect on ovary morphology. This is unlikely to be related to Kipferl, as neither loss of Kipferl nor loss of Rhino has a negative effect on ovary morphology. Instead, interference with HP1a-mediated heterochromatin integrity is known to lead to the same atrophic ovary phenotype (Casale et al., 2019). How the low-level expression of HP1a chromo domain fused to the Rhino chromoshadow domain can interfere with wildtype HP1a is unclear. The chimeric protein does not seem to associate with H3K9me3-marked chromatin, which should thereby still be available for HP1a to bind. It is possible that the HP1a chromo domain, similar to Rhino, interacts with dedicated specificity factors that regulate chromatin association of HP1a in early stages of heterochromatin establishment. Ectopic, non-functional HP1a chromo domain-containing constructs might titrate these factors away from functional HP1a, leading to downstream defects.

The finding that the chromo domain of Rhino serves as the interaction point with Kipferl uncovers a novel role for chromo domains, which had previously not been shown to interact with other proteins besides methylated histone tails. A dimeric Rhino chromo domain, as well as Rhino variants lacking either the chromoshadow domain or hinge region are able to bind to many Rhino domains, presumably through the interaction with Kipferl. Interestingly, binding of these constructs at Kipferl-independent piRNA clusters *42AB* and *38C* was not detected, indicating that these regions might employ Rhino specification factors that do not act via the chromo domain.

3.2.2 Hinge

The Rhino hinge domain is exceptional across HP1 proteins, as it is several fold longer than that of most other HP1 proteins (Figure 71). The sheer length of the Rhino hinge might be connected to the unique function of this HP1 variant protein: While canonical HP1a resides in compact, densely packed, silent heterochromatin, Rhino bound regions are constantly transcribed, which might result in more relaxed nucleosome spacing and require a longer hinge to link neighboring nucleosomes. Along this line, the hinge domain might serve mostly as a flexible linker between two loosely spaced chromo domains. This is however not in agreement with the observation, that replacement of the Rhino hinge by an unstructured linker impairs Rhino's ability to form defined foci in immunofluorescence imaging, or to support the formation of wildtype level Rhino domains. Rhino's hinge contains long stretches

of multiple Serine residues, as well as a high number of charged amino acids, which might mediate interactions with other proteins or nucleic acids, or serve as targets for regulation via post translational modifications. The hinge domains of *Drosophila* HP1a, as well as human HP1alpha, and to a lesser extent also HP1beta and gamma, have been shown to interact with nucleic acids, which mediates the phase separation behavior of the protein (Keenen et al., 2021, Larson et al., 2017, Strom et al., 2017). Positively charged residues in the hinge region of the *Schizosaccharomyces pombe* HP1 protein Swi6 have been shown to impact the association with H3K9me3 in an RNA-dependent manner (Keller et al., 2012). Post translational modification of HP1 proteins has also been abundantly described, and is crucial to regulate the chromatin association of Tetrahymena HP1 protein Pdd1p (Kataoka et al., 2016). A recent study showed that a chimeric Rhino construct where the hinge domain had been replaced for the corresponding sequence of *Drosophila simulans* Rhino was able to support full fertility and intact piRNA biogenesis (Parhad et al., 2017). This indicates that, despite marked differences in amino acid sequence between the two species, the crucial features of the hinge domain are conserved. Future experiments will be required to delineate which features of the hinge are required for different aspects of Rhino's functionality.

3.2.3 Chromoshadow domain

With the caveats mentioned above, the experiments presented in this study are not suitable to elucidate any role of the chromoshadow domain in determining the localization pattern of Rhino. Similar to the hinge domain, replacement of the chromoshadow domain with an artificial dimerization domain impaired the formation of wildtype level Rhino domains as observed by immunofluorescence stainings as well as ChIP-seq analysis. This might be due to the role of the chromoshadow domain in recruitment of downstream factors. The lack of nascent transcripts induced by Moonshiner and concentrated by the remaining piRNA cluster factors might interfere with higher order interactions and feedback mechanisms required for the full strength of Rhino domains (Andersen et al., 2017). Of note, Parhad et al. recently demonstrated that exchange of the chromoshadow domain of *Drosophila melanogaster* with the *Drosophila simulans* version rendered the Rhino protein non-functional for maintenance of female fertility or piRNA production (Parhad et al., 2017). This was explained by the fact that the interaction interface with Deadlock was not conserved and the protein was unstable (Parhad and Theurkauf, 2019, Yu et al., 2018b, Parhad et al., 2017).

3.3 Interaction mode between Rhino and Kipferl

Our data indicate that Rhino interacts with Kipferl via its chromo domain. To our knowledge, this is the first reported interaction of an HP1 family protein taking place at the chromo domain. Binding of HP1 partner proteins has so far been described mostly for the chromoshadow domain, and to a lesser extent also for the hinge region (Lomberk et al., 2006). As the chromo domain binds to the H3K9-methylated histone tail, an interaction with Kipferl

in the same domain might immediately allow for modulation of Rhino's chromatin binding. At the same time, it is compatible with Rhino's interaction with the downstream factor Deadlock via the chromoshadow domain (Yu et al., 2018a).

According to our results, stable chromatin binding of Rhino at most sites requires simultaneous interaction with H3K9me2/3 as well as DNA-bound Kipferl. How this is achieved at chromatin will be interesting to investigate using biochemical and structural work. Moreover, both Rhino and Kipferl harbor dimerization domains, which are crucial for the proteins' functions. This holds the potential for the formation of larger protein structures at chromatin and might be important for the spreading behavior we observed for the two proteins. Unfortunately, we were unable to purify recombinant Kipferl protein to study its interaction with Rhino *in vitro*. This also precluded any biochemical characterization of Kipferl's DNA binding ability and specificity, which will be required to undoubtedly confirm Kipferl's specificity for the GRGGN motif found in Kipferl ChIP-seq peaks.

3.4 ZAD-Zinc finger proteins

With its N-terminal ZAD, Kipferl is a member of the largest group of ZnF proteins found in insects. Of the 92 members in *Drosophila melanogaster*, only a hand full of ZAD-ZnF proteins have been previously characterized. The functions of these proteins are very diverse, ranging from roles in transcriptional regulation to heterochromatin biology and chromatin organization (Bag et al., 2021, Harms et al., 2000, Kasinathan et al., 2020a, Sabirov et al., 2021a, Sabirov et al., 2021b, Maksimenko et al., 2015, Maksimenko et al., 2020). Many ZAD-ZnF proteins are expressed predominantly in gonads and during early embryogenesis where they often have essential functions (Kasinathan et al., 2020a). It is tempting to speculate, that other ZAD-ZnF proteins might act as sequence-specific guidance factors for Rhino in other tissues, or carry out similar roles for other HP1 protein variants, which are also abundantly found in high diversity in insects and often expressed predominantly in the germline. One example where this might be the case is the uncharacterized ZAD-ZnF protein CG8388, which we identified as an additional hit in yeast two-hybrid screens testing for Rhino interactors in ovaries and in testes, respectively. Indeed, loss of CG8388 leads to a reduction of Rhino ChIP-seq signal at selected domains in ovaries. Interestingly, expression levels of *CG8388* are higher in testes, where Kipferl is not expressed. Preliminary immunofluorescence staining experiments interrogating Rhino localization in *CG8388* mutant testes did not show obvious differences, but more sensitive assays will be required to further study the role of CG8388 in Rhino biology. Of note, the minimal Rhino-interacting domain of CG8388 identified for interacting protein fragments in the yeast two-hybrid screen using a testes-derived cDNA library corresponds roughly to the first Zinc finger within CG8388 (Figure 75). It will be interesting to determine if CG8388 and Kipferl share the same interaction mode with Rhino's chromo domain, and whether their binding to Rhino is mutually exclusive.

The large family of ZAD-ZnF proteins, together with the newly found role of two of these proteins in sequence-specific transposon silencing parallels the highly diverse family of KRAB-ZnF proteins in tetrapods. This protein family radiated in a process that is believed to be fueled by the conflict between host and transposons (Bruno et al., 2019). The adaptation to new transposon sequences drives the fast evolution of KRAB-ZnF proteins, specifically of their DNA binding ZnFs. In ZAD-ZnF proteins, however, signatures of positive selection are not concentrated on ZnF domains (Kasinathan et al., 2020b), indicating that the two families of ZnF proteins experience different evolutionary pressures.

While the KRAB domain links to silencing pathways to induce the repression of underlying chromatin, the ZAD has not been reported to interact with other factors, but acts merely as a dimerization domain, predominantly mediating the formation of homodimers (Jauch et al., 2003, Bonchuk et al., 2021). It seems likely that ZAD ZnF proteins hold diverse functions in many different biological processes, unlike KRAB ZnF proteins that use their shared N-terminal domain to link to a common set of downstream factors.

3.5 Interdependence of Kipferl and Rhino at chromatin

Our collective data indicate that Kipferl not only recruits Rhino to its intrinsic binding sites, but also supports the spreading of Rhino from these nucleation sites to flanking regions, supported by local H3K9me2/3. Interestingly, Kipferl co-occupies the full extent of all Rhino domains in wildtype ovaries, despite its intrinsic chromatin binding sites being only narrow, shallow peaks within extended Rhino domains. Upon initial Rhino recruitment, the two proteins thus appear to engage in mutual stabilization, which facilitates their spreading into larger domains. This stabilizing effect of Kipferl is also apparent at the large piRNA clusters *42AB* and *38C*, despite Rhino binding being largely Kipferl-independent. Here, Rhino is likely recruited by alternative guidance factors, but requires Kipferl for stabilization and spreading to achieve the full extent of binding observed in wildtype ovaries. A similar scenario is also observed at a considerable set of fully Kipferl-dependent Rhino domains, as roughly 40% of these domains do not contain clearly detectable Kipferl nucleation sites. At these loci Rhino and Kipferl bind in a co-dependent manner which might be mediated by low intrinsic affinities of both proteins. Alternatively, additional guidance factors might exist, which allow the recruitment of Rhino, but rely on Kipferl for the stabilization and expansion of a larger Rhino domain.

Together, these observations indicate that Kipferl, while preferentially binding to GRGGN motifs, is able to bind DNA in a more promiscuous manner irrespective of sequence motifs, which allows it to strongly bind to the entirety of all Rhino-bound chromatin. We observed that the boundaries and irregular shapes of Rhino domains often coincide with fluctuations in the GC-content of the underlying DNA. It is possible that Kipferl's ZnFs, while showing highest affinity for their ideal consensus motif, also stably associate with Guanosine-rich sequences in general, facilitating the spreading along GC-rich DNA. It will be interesting to

investigate Kipferl's DNA binding abilities and nucleotide preferences in more detail using *in vitro* purified Kipferl protein to gain closer mechanistic insight.

3.6 Kipferl's function in the piRNA pathway

The evolutionary processes that lead to the emergence of a factor like Kipferl are open for speculation. For several proteins acting in the piRNA pathway, the involvement in the arms race between transposable elements and the host genome has led to the emergence of highly species-specific protein features through strong positive selection (Blumenstiel et al., 2016, Parhad and Theurkauf, 2019, Simkin et al., 2013). For example, the interaction of Rhino with Deadlock was recently shown to be significantly different in the closely related *Drosophila melanogaster* and *Drosophila simulans* species (Parhad and Theurkauf, 2019, Yu et al., 2018b). For Kipferl, which is conserved among *Drosophila* species with stronger divergence outside the *melanogaster* clade, the lack of amino acid polymorphisms at the *kipferl* locus precludes a conclusive analysis of positive selection signatures. However, upon comparison with Kipferl proteins in other *Drosophila* species we do not find pronounced amino acid changes in any of the folded domains, indicating that neither DNA binding nor the Rhino interaction interface changed significantly since the divergence of different *Drosophila* species (Figure 47). In line with this, Parhad et al. found that substituting Rhino's chromo domain with the *Drosophila simulans* counterpart results in a functional protein sustaining fertility, TE silencing, piRNA cluster binding, and piRNA production (Parhad et al., 2017). This indicates that *Drosophila simulans* Rhino is compatible with *Drosophila melanogaster* Kipferl and the pairing of Rhino with Kipferl is more evolutionarily stable than Rhino's interaction with Deadlock.

Nevertheless, the *kipferl* mutant phenotype holds important clues for possible forces that might have led to the evolution of Kipferl, which I discuss below.

3.6.1 Kipferl as guide to find transposon sequences

Loss of Kipferl leads to reduced female fertility. This is likely connected to a loss of Rhino from transposon sequences, which leads to a reduction of antisense piRNAs and subsequent de-repression of multiple transposons. Kipferl therefore seems to be required for Rhino to recognize transposon sequences and mark them for piRNA production. In support of this, we find that the genomic 1-kb tiles showing the highest intrinsic binding of Kipferl often contain sequences of *gypsy-8* and *DMRT1* family transposons. Moreover, these elements are highly enriched in Kipferl's GRGGN consensus motif. The connection of Kipferl to few transposable elements might suggest that Kipferl specifically evolved to target these elements for Rhino-dependent piRNA production. Among the elements with a high number of Kipferl motifs, only the *Max-element* is derepressed in the absence of Kipferl. The *gypsy-8* and *DMRT1* elements, although depending on Rhino for piRNA production, are not upregulated even upon depletion

of Rhino itself, indicating that the genomic copies of these elements are no longer intact and cannot be re-activated. This implies that Kipferl might have evolved to combat an ancient invasion of these now inactive elements, and might explain the lack of ongoing selection on the Kipferl protein.

With its ability to support the spreading of Rhino into unrelated neighboring sequences, Kipferl might allow Rhino to bind also to nearby insertions of elements lacking Kipferl nucleation sites. A prominent example for this mode of action is the piRNA cluster *80F*, where fragments of *gypsy8* and *Rt1A* and *B* elements provide binding sites for Kipferl, which subsequently recruits Rhino and supports its binding also to neighboring transposons.

The Kipferl consensus motif is relatively simple and occurs at many transposon-unrelated sites in the genome. We find that Kipferl indeed binds to chromatin in sharp stand-alone peaks all along chromosomal arms where H3K9me2/3 is absent. According to our model, Rhino recruitment at these sites is only prevented by the lack of H3K9 methylation. We speculate that transposon insertions proximal to Kipferl binding sites, followed by their targeting through Piwi-dependent heterochromatin formation, induces the nucleation of a new Rhino domain at the insertion site. In this manner, potentially active copies of transposons would immediately be converted to new sources of piRNAs, strengthening the silencing of all other copies in *trans*.

It is tempting to speculate that Kipferl, besides being an effective binder of certain transposon sequences, hit a sweet spot of selectivity and promiscuity in its DNA binding specificity. The right frequency and distribution of Kipferl binding sites across the genome, aided by the low expression level of Kipferl, might thereby contribute to the effective control of transposons, irrespective of their sequence. This system would allow for robust Rhino recruitment, at the cost of occasional off-target Rhino domains, when Kipferl motifs and H3K9 methylation coincide at regions lacking transposon sequences. The prominent Rhino domain at the *headcase* locus likely represents an example of such an off-target domain. The lack of piRNA production at this locus is an indication that additional levels of control are at place in the piRNA pathway that prevent piRNA biogenesis from unsuited transcripts, which leaves room for flexibility of Rhino's chromatin binding.

3.6.2 Kipferl as competition factor against Satellites

The dramatic re-localization of Rhino from hundreds of wildtype domains to Satellite arrays in *kipferl* mutant ovaries leads us to envision a second function for Kipferl. The massive sequestration of Rhino to Satellites might indicate an intragenomic conflict, where different genomic loci compete for Rhino binding at the expense of others. Intriguingly, both the *Rsp* and the *1.688* Satellites are known to be involved in genetic conflicts (Larracuente and Presgraves, 2012, Ferree and Barbash, 2009, Chen et al., 2021, Ferree and Prasad, 2012). The largest array of *1.688* Satellite sequences is located at the boarder of centromeric and pericentromeric regions of the X-chromosome and spans more than 10 Mbp. This array of

359bp element repeats acts as a hybrid lethality locus in crosses between *Drosophila melanogaster* and *Drosophila simulans* (Ferree and Barbash, 2009, Chen et al., 2021). Females lacking this repeat generate non-viable offspring when mated to wildtype males, while no defects occur in the reciprocal cross. This phenotype was found to be due to a requirement for maternally deposited small RNAs (siRNAs or piRNAs) to prevent mitotic catastrophe caused by uncontrolled *1.688* repeats (Usakin et al., 2007). It was proposed that the proper packaging of this Mega-basepair long, repetitive sequence in early stages of embryogenesis, mediated by small RNA pathways, might be crucial for the faithful pairing and segregation of chromosomes (Wei et al., 2021). High levels of piRNA production from *1.688* sequences would thereby increase the chances of proper segregation of long Satellite arrays, and these selfish elements might have evolved mechanisms to bias the cellular machinery towards this purpose.

The *Rsp* Satellite has been identified genetically as part of the Segregation Distorter system, a meiotic drive system in males (Larracuente and Presgraves, 2012). The mechanisms underlying the selfish properties of *Rsp* are largely unclear to date, but the production of piRNAs from *Rsp* sequences might hint at a possible involvement of piRNAs in its regulation.

Both *Rsp* and *1.688* Satellites have recently been described to give rise to piRNAs in an RDC/Moonshiner-dependent manner in both ovaries and testes (Wei et al., 2021). If piRNA production from *Rsp* and *1.688* Satellites is important to maintain high copy numbers of *1.688* and *Rsp* repeats during embryogenesis and gametogenesis, respectively, Satellites might have evolved means to force piRNA production from their own loci. One way to do so, would be the direct recruitment of Rhino to genomic Satellite arrays. The discovery of Kipferl in this study indicates that direct recruitment of Rhino via sequence-specific factors is possible, which would likely be a very efficient strategy for the highly repetitive Satellites.

As evidenced by the *kipferl* mutant phenotype, disproportional sequestration of Rhino and its associated transcription and export machinery to Satellites results in the loss of Rhino from loci relevant to transposon control. We speculate that Kipferl might have evolved out of a necessity to counteract this sequestration by selfish Satellite repeats and to stabilize Rhino at its diverse genomic target loci. Kipferl's affinity for Guanosine-rich sequences optimally opposes the AT-rich Satellite sequences. This preference for opposing sequence content might in this scenario be more central to Kipferl's mode of action than the optimal targeting of certain transposon sequences.

3.7 Kipferl-independent Rhino recruitment across time and tissues

Despite the close inter-dependence of Rhino and Kipferl we find that Kipferl's spatio-temporal expression profile is distinct from that of Rhino: Kipferl levels are readily detectable only upon germ cell differentiation in the ovary, with very low levels in germline stem cells and cystoblasts and no protein detected in testes (Figure 25). In line with this, we find that the Rhino profile in cultured GSCs differs strongly from the pattern described for ovaries, and

similar observations were recently made by Chen et al. who characterized Rhino in testes (Chen et al., 2021). Kipferl's specific expression pattern therefore likely contributes to the distinct Rhino landscapes in these different niches. Interestingly, Chen et al. also report a developmentally regulated chromatin binding pattern of Rhino during spermatogenesis: Germline stem cells and early spermatogonia show Rhino binding at several piRNA clusters (Chen et al., 2021). Of note, cluster *80F* is not bound by Rhino in testes, in line with the lack of Kipferl in testes. In later stages of spermatogenesis, Chen et al observe strong reorganization of Rhino on a single X-linked locus called *AT-chX* (Chen et al., 2021). The Rhino binding profile thus seems to be highly regulated also in testes.

Several observations thus imply the existence of additional Rhino guidance factors: (I) the existence of rare Kipferl-independent Rhino domains in ovaries at piRNA clusters *42AB* and *38C*, (II) domains of strong interdependence of Rhino and Kipferl, (III) the sequestration of Rhino to Satellite repeats in the absence of Kipferl, (IV) and the spatio-temporal regulation of Rhino's chromatin binding pattern. With the ZAD-ZnF protein CG8388 we identified a likely candidate for such a factor that might be active in testes. How many additional guidance factors exist to contribute to the full diversity and extent of Rhino's chromatin binding, and whether they all belong to the family of ZAD-ZnF proteins remains to be investigated. Moreover, it will be interesting to see if these factors share Kipferl's functions as a broad stabilizer of Rhino, or if Kipferl is especially important, holding a dual role in both guidance and support of Rhino's chromatin association.

The physiological relevance of different Rhino localization patterns in different tissues and developmental stages is unclear at this point, but might be rooted in the different requirements for transposon repression. Transposable elements have adapted to different niches within the reproductive organs of *Drosophila*, with different expression patterns in time and space, and the piRNA pathway likely adapted to silence the right element in the right cell type. We find that overexpression of Kipferl in early stages of the germarium did not have a negative impact on female fertility, implying that the lack of Kipferl in these cells might not be important (Figure 46). It is possible that the more loosely defined Rhino distribution as observed in cultured GSCs might nevertheless be beneficial in undifferentiated cells of the germline for reasons that are yet to be uncovered.

3.8 Kipferl and previous models of Rhino specification

The discovery of a sequence-specific component in the definition of Rhino's chromatin binding profile presents a major advance in the understanding of piRNA cluster specification. Previous research conducted over the last decades has identified other important factors in this process. Here I discuss how they might interact with the mechanisms identified in my doctoral work.

3.8.1 Maternally provided piRNAs mark Rhino domains across generations

The most important component known to be involved in Rhino deposition are maternally deposited piRNAs, which carry the heritable information about previously Rhino-bound genomic locations across generations. It was shown that Rhino binding at single transposon loci, as well as at reporter transgenes mimicking repetitive transposon sequences depends on maternally provided Piwi-piRNA complexes (Casier et al., 2019, Akulenko et al., 2018, Komarov et al., 2020, Olovnikov et al., 2013). This is most likely mediated by the heterochromatic marks deposited through the silencing machinery downstream of Piwi. A recent study further reported that targeting by antisense small RNAs was sufficient to convert the Rhino-independent uni-strand germline piRNA cluster *20A* into a Rhino domain giving rise to bi-directional piRNA production, which was inheritable through the maternal lineage for several generations (Luo et al., 2022). Together, these findings undoubtably demonstrate an important role of maternally provided piRNAs in Rhino specification.

However, studies from the Kalmykova lab demonstrate that only a relatively small subset of transposon insertions is bound by Rhino, although many other insertions are also targeted by piRNAs as evidenced by Piwi-dependent heterochromatin in their flanking regions (Akulenko et al., 2018). The authors observed the same picture for repetitive reporter constructs, whose conversion into piRNA producing Rhino domains depends strongly on the genomic location of their insertion. Despite abundant maternal piRNAs targeting both transposon and reporter sequences, certain genomic locations seem to be predisposed for the formation of a Rhino domain, while others are refractive to this conversion. This implied the existence of context-dependent influences that aided Rhino recruitment at certain genomic locations, but not at others.

With Kipferl we present the first factor capable of providing an explanation for the locus-specific conversion of local heterochromatin into Rhino domains. We find evidence that transposon insertions occurring close to pre-existing Kipferl binding sites are efficiently bound by Rhino and propose that other Rhino guidance factors likely act in the same manner at other genomic sites. Importantly, Kipferl does not influence the H3K9 methylation status, but rather stabilizes Rhino at sites where H3K9me2/3 is provided through parallel pathways. In this manner, Kipferl's intrinsic chromatin binding pattern likely contributes to the establishment of Rhino binding where it coincides with sites of Piwi-mediated heterochromatin formation. This model proposes that the role of maternally inherited Piwi/piRNA complexes in Rhino specification lies in the deposition of H3K9me2/3 at target sites, but that Kipferl and putative other specificity factors then lead to the recruitment of Rhino to a subset of piRNA-targeted regions. This provides a possible explanation for the relatively low fraction of transposon insertions bound by Rhino, despite piRNA targeting at many other sites.

Of note, the antisense small RNA-induced conversion of cluster *20A* into a Rhino domain as described by Chen et al. occurs proximal to a strong intrinsic Kipferl peak at the beginning of the cluster, and it would be interesting to see if Kipferl is indeed required for this conversion. The context-specific conversion of an *I*-element containing reporter construct into Rhino

domains described by Akulenko and colleagues (Akulenko et al., 2018) cannot conclusively be explained by Kipferl binding sites, suggesting that alternative guidance factors likely play a more important role at these sites.

3.8.2 Rhino binding is restricted by H3K9 demethylation

Recent results by Casier et al. indicate that Rhino's chromatin occupancy is restricted by the activity of the histone lysine demethylase Kdm3 (Casier et al., 2022). Loss of Kdm3 activity led to an increase in H3K9me2/3 at ectopic sites and allowed the conversion of a previously described reporter locus, *BX2*, into an active, Rhino-dependent piRNA cluster. The authors report that, along with *BX2*, several additional loci were ectopically bound by Rhino after an increase of H3K9 methylation was observed at these regions. Not all newly methylated sites were bound by Rhino, however. It is interesting to speculate that this observation might be explained by our model, where Rhino binding at new H3K9me3-domains depends on the presence of a nearby binding site of Kipferl or alternative specificity factors.

3.8.3 Feedback of transcription and export machinery influences Rhino specification

The chromatin binding profile of Rhino was recently described to be regulated by the transcription and export (TREX) complex composed of the hexameric THO complex and the DEAD box helicase UAP56 (Zhang et al., 2018). The TREX complex plays a conserved role in nuclear mRNA export from yeast to humans. It was previously shown to colocalize with Rhino in nuclear foci corresponding to piRNA clusters and to bind piRNA precursor RNAs, which was remarkable considering that nuclear export of piRNA precursors happens via a dedicated alternative pathway centered around Nxf3, a germline-specific paralog of the general mRNA export adaptor Nxf1 (ElMaghraby et al., 2019, Zhang et al., 2012a, Kneuss et al., 2019). Nxf3 forms a stable complex with the conserved NXF partner protein Nxt1, the piRNA pathway factor Bootlegger, and UAP56. Recruitment of Nxf3 and UAP56 to Rhino foci in nurse cells was shown to rely strongly on Bootlegger, which is itself recruited by Deadlock, establishing a seemingly clear recruitment hierarchy at piRNA clusters (ElMaghraby et al., 2019). The mechanism and timing of THO recruitment, however, has not been conclusively clarified. Collective data from the previous studies demonstrates that, despite the essential role of THO/UAP56 in mRNA export, certain allelic combinations of UAP56 deficiencies as well as mutants for THO subunits Thoc5 and Thoc7 allow the development of adult, but sterile flies with piRNA pathway-specific defects (Zhang et al., 2018, Zhang et al., 2012a). Zhang et al find that these mutations disrupt the colocalization of other THO subunits with Rhino foci in nurse cell nuclei (Zhang et al., 2018). This is accompanied by a reduction of piRNA production, as well as increased transposon expression. Interestingly, the authors also describe an effect of mutations in Thoc5/7 and UAP56 on Rhino's chromatin binding pattern, with reduced Rhino levels at cluster *42AB*, but increased binding at several other domains. Comparison with our own ChIP-seq data, however, suggest strong discrepancies between wildtype Rhino ChIP-seq

data sets from different labs, where the Rhino profiles in Thoc5/7 and UAP56 deficient ovaries more closely resemble the w^{1118} Rhino profile observed in this work than the accompanying w^{1} data set. This implies that unfortunate genotype comparisons might confound the conclusive analysis of these data. Thus, the role of the TREX complex at piRNA clusters awaits further clarification.

3.8.4 piRNA precursor export and the Kipferl phenotype

The Nxf3-dependent accumulation of Rhino-bound chromatin at the nuclear envelope in *kipferl* mutant ovaries suggests that the piRNA precursor export pathway is tightly connected to precursor transcription. We can envision that the export of RNA molecules still associated with underlying chromatin, either through co-transcriptional RNA export or strong RNA retention of accumulated TREX complexes, might serve an important protective function: piRNA precursors lack most signatures of mature mRNAs, and therefore likely require shielding from nuclear RNA degradation. Moreover, the bi-directional transcription of Rhino-dependent piRNA source loci inherently means that these sites are excellent targets for Piwi/piRNA mediated silencing, which might compete with piRNA precursor export and contribute to the requirement for a highly efficient export pathway.

3.9 What drives Rhino's fast evolution?

The "Red Queen" Hypothesis states that continued adaptation on both sides of a conflict is required for the system to remain in equilibrium (Van Valen, 1973). In a conflict between host and parasite, the host's defense system thereby constantly evolves to recognize and fight the parasite, which is in turn forced to adapt and overcome this defense, forcing the host to adapt again in a continuous cycle of innovation. This fast evolution of the involved factors leads to higher than expected differences between closely related species and is referred to as positive selection.

Along with several other piRNA pathway genes, such as *aub*, *armi*, *spn-E* and *maelstrom*, Rhino has been described to be under strong positive selection, suggesting its involvement in a genomic arms race (Wang et al., 2020, Parhad and Theurkauf, 2019). The conflict of the *Drosophila* germline genome with transposable elements is thought to be the major driving force of the evolution of piRNA pathway factors, although doubts have been raised about this connection in recent years based on several observations that I summarize in the following.

Unlike other examples of host-parasite conflicts, where positive selection acts on the direct interface of host and parasite proteins, the fastest evolving members of the piRNA pathway are involved in piRNA biogenesis rather than direct transposon contact (Blumenstiel et al., 2016). This might reflect the strength of the pathway, relying on RNA sequence complementarity rather than protein-protein interaction for transposon recognition. Targeting of long transposon transcripts via highly abundant and diverse piRNAs puts an

immense pressure on transposons to change their entire nucleotide sequence, coding and non-coding, while maintaining their functionality. It was therefore suggested that transposon proteins might adapt to mimic piRNA pathway factors to disrupt interactions within the pathway, which would then be counteracted by compensatory mutations in the involved proteins. A potential example of this might be the highly species-specific interaction of Rhino and Deadlock, which was shown to stem from compensatory amino acid substitutions in both proteins (Yu et al., 2015a, Yu et al., 2018b). However, no piRNA pathway factor-mimicking transposon proteins have been described to date.

The classical "Red Queen" scenario implies that both sides of an ongoing conflict undergo repeated rounds of innovation, leading to positive selection in both parasite and host proteins (Van Valen, 1973). This is reflected for example in the fast evolution of viruses continuously evading the immune system. Blumenstiel and colleagues point out, however, that *Drosophila* transposons show hardly any sign of positive selection, and are remarkably stable over long evolutionary distances (Blumenstiel et al., 2016). Transposon proteins thus don't seem to participate in the same arms race as the piRNA pathway.

Moreover, the copy number of transposons is far below the high load of parasites in an ongoing viral infection, suggesting that the evolutionary pressure through transposons is unlikely to explain the high degree of selection observed in the piRNA pathway (Blumenstiel et al., 2016).

These observations led to the suggestion that the piRNA pathway is involved not just in the conflict with transposable elements, but also other arms races. A connection to viral defense was suggested, but is not supported by available data, which indicates no defect in viral defense for piRNA pathway mutants. Other hypotheses suggest that cycles of high or low transposon load might have led the piRNA pathway to fluctuate between states of high defense efficiency during active invasions, versus high specificity in times of low transposon burden (Blumenstiel et al., 2016). According to this model, the avoidance of auto-immunity through off-target effects at endogenous genes might contribute to the evolution of piRNA pathway genes.

3.9.1 The role of Rhino at Satellite repeats

The strong potential connection of Rhino to Satellite biology described in this work might offer an additional explanation of the evolutionary pressure acting on the piRNA pathway. Satellite sequences, although not known to encode protein features, form complex and highly species-specific assemblies in centromeric regions of virtually all organisms. They are among the fastest evolving sequences known (Thakur et al., 2021). A conflict with Satellite sequences would explain the requirement for piRNA pathway factors to evolve more rapidly.

Our findings demonstrate that Satellite repeats of the *Rsp* and *1.688* families have the capacity to sequester Rhino at very high efficiency in the absence of Kipferl, and that their transcripts are processed into piRNAs with an astonishing efficiency. Observations by Wei et

al. indicate that high levels of piRNAs complementary to these Satellites are likely beneficial for the maintenance of these sequences across generations (Wei et al., 2021). By hijacking different steps of the piRNA pathway, Satellites could push the pathway to higher production of Satellite-targeting piRNAs, thereby ensuring their maintenance. Unlike transposons, whose activity and expansion are restricted by the piRNA pathway, Satellites would have opposing needs, requiring piRNA-mediated heterochromatic packaging for their continued propagation. It is interesting to speculate that this requirement for higher piRNA production from their sequences might be a potential explanation for the accumulation of positive selection signatures in piRNA biogenesis factors, rather than the silencing branch of the pathway.

Without any encoded proteins, the evolutionary potential of Satellites is restricted to their nucleotide sequence, which might be adapted to recruit pre-existing transcription factors or RNA-binding proteins. Of note, with several piRNA biogenesis proteins as well as PIWI proteins themselves exhibiting a strong bias toward Uridine-containing substrates, *Rsp* and *1.688* transcripts with their high AT-content serve as ideal piRNA precursors.

This allows us to speculate that the silencing of transposons is not the only function of the piRNA pathway in *Drosophila*, but that Satellite maintenance is an equally important function. While excessive sequestration of the piRNA pathway machinery to selfish Satellite repeats is likely undesirable as it impairs the control of transposable elements through the pathway, proper packaging of Satellite sequences is crucial for their successful segregation in early embryonic cycles of mitosis.

Despite strong sequestration of Rhino to Satellite repeats and a strong decrease in piRNAs antisense to a large variety of transposons, fertility in *kipferl* mutant females is only decreased to 20-40%. While this might be due to the intact Rhino pool in early stages or maintained Rhino binding at clusters *42AB* and *38C*, our experiments indicate that neither early Rhino, nor clusters *42AB* and *38C* are required – or sufficient, in the case of early Rhino – to maintain female fertility. It is therefore tempting to speculate that the reason for the remaining fertility of *kipferl* mutant females lies in the continued maintenance of Satellite repeats, which is lost in *rhino* mutants. A distinction of Rhino's function in transposon silencing and Satellite maintenance is currently hindered by a lack of genetic tools to specifically disrupt Rhino binding at Satellites while leaving transposon silencing intact, and will be an interesting point of future investigation.

3.10 Implications beyond the piRNA pathway and *Drosophila*

Besides marking an important step in the understanding of the *Drosophila* piRNA pathway, the identification of Kipferl as a co-factor for Rhino in piRNA cluster recognition holds implications beyond the piRNA pathway. The cooperation of a DNA binding protein with a chromatin reader for the sequence-specific recognition of a subset of constitutive

heterochromatin represents a novel concept, that might be used in other biological processes and systems.

We show that Kipferl interacts with the chromo domain of Rhino. To our knowledge, this is the first report of a protein-protein interaction with a chromo domain that goes beyond histone tail binding. This challenges the prevalent view that HP1 proteins interact with their partner proteins only through the chromoshadow domain. Our attempts to identify novel Rhino interactors via co-immunoprecipitation experiments coupled to mass spectrometry failed to identify Kipferl or CG8388, indicating that these interactions taking place at the direct interface with chromatin and DNA can be hard to detect with conventional approaches. This might be one reason why such interactions have not been detected previously. Our findings can serve as a novel incentive for the dedicated search for proteins acting in a similar manner.

Known examples where putative DNA-binding ZnF proteins have been described for their role in HP1 recruitment have been described in different organisms: The *Drosophila* HP1 variants HP1b and c rely on their cofactors Row and Woc for chromatin binding (Font-Burgada et al., 2008). In mouse embryonic stem cells, HP1beta/gamma associate with the chromatin remodeler CHD4 and the activity-dependent neuroprotective protein (ADNP) to form the ChAHP complex, which binds to selected chromatin sites and is believed to regulate gene expression and genome organization (Ostapcuk et al., 2018, Kaaij et al., 2019). In both cases, the interaction with ZnF proteins enables the H3K9me3-independent recruitment of HP1 variant proteins to sites where their function has not been fully elucidated. This mode of recruitment is distinct from Kipferl's function in the specification of Rhino domains, which crucially relies on underlying H3K9me2/3 coinciding with Kipferl binding sites. An example where DNA sequence motifs instruct the formation of a specialized repressive chromatin environment is the Polycomb-regulated facultative heterochromatin in *Drosophila*. Here, so called Polycomb response elements (PRE) act as nucleation sites for PcG recruitment and establishment of repressive chromatin through methylation of H3K27, as well as ubiquitylation of H2AK119 (Srinivasan and Mishra, 2020). A further example is a recently described sequence motif sufficient for inheritance of heterochromatin in *Schizosaccharomyces pombe* (Wang et al., 2021). While the recognition of a sequence element is followed by the subsequent establishment of repressive chromatin via histone tail modifications in these examples, however, we find no evidence that Kipferl's chromatin binding affects the histone modification pattern at chromatin. Instead, a new reader protein is stabilized at pre-existing H3K9me2/3 deposited by parallel pathways.

It will be interesting to see if other chromatin reader proteins rely on similar recruitment mechanisms to specify chromatin sub-compartments. Interestingly, HP1 proteins have recently been shown to have undergone massive radiation in Diptera, with certain species encoding up to 20 HP1 variant proteins (Helleu and Levine, 2018). The function of most of these proteins is unknown, and it is easy to speculate that many of them will require additional specificity factors besides chromatin marks to determine their target loci at chromatin. As discussed above, also the ZAD-ZnF family of proteins is highly diverse in insects

and makes up one of the largest families of transcription factors in *Drosophila*. It is therefore conceivable, that other pairs of HP1 and ZAD-ZnF proteins will be shown to cooperate in the future, although similar roles could certainly also be taken by other types of DNA-binding proteins. Furthermore, future research will hopefully show whether this concept expands beyond the HP1 family of proteins, and if similar mechanisms are employed in other organisms.

4. Outlook

My PhD project aimed to understand how the specific chromatin binding pattern of the HP1 protein variant Rhino at piRNA clusters is achieved. The data presented in this work represents the first unbiased and complete characterization of Rhino binding sites in multiple genetic backgrounds and highlights the diversity of Rhino domains reaching far beyond stereotypical piRNA clusters, which had been the focus of Rhino-related research in the past. More importantly, the discovery of Kipferl, a putative sequence-specific DNA binding ZnF protein, as a guidance and stability factor for Rhino critically enhances our understanding of how the complex chromatin binding pattern of Rhino is established. For the first time, it allows us to answer how Rhino selects a subset of cellular H3K9me2/3-marked heterochromatin for binding. However, several questions remain unanswered.

Kipferl's intrinsic chromatin binding profile fully explains roughly 60% of Rhino domains in ovaries including the large piRNA cluster *80F*. At the two largest piRNA clusters *42AB* and *38C*, however, Rhino binding does not depend on Kipferl, implying the existence of other guidance factors for Rhino. The same holds true at many domains where Rhino and Kipferl seem to equally depend on one-another for chromatin binding. Specification of these sites likely requires additional factors that remain to be discovered. Moreover, Rhino was shown to bind to chromatin in a developmentally regulated pattern in testes, where Kipferl is absent. We envision the involvement of additional specificity factors in this process. How many factors share Kipferl's role as Rhino guidance factors, and how much their functions overlap at various loci remains to be determined. In addition, analogous studies investigating potential guidance factors for other HP1 variant proteins will be of great interest for general heterochromatin research.

One interesting candidate for Rhino specification in testes is the ZAD-ZnF protein CG8388, which we identified as Rhino interactor. Initial attempts to determine the localization of an additional copy of FLAG-tagged CG8388 in ovaries or testes were unsuccessful so far, and future overexpression experiments or endogenous tagging should be undertaken to characterize the localization of CG8388 in respect to Rhino and Kipferl, as well as its behavior in the absence of Rhino. We observed a mild effect of CG8388 depletion on ovarian Rhino binding, and it will be interesting to extend these studies to testes.

Among the most fascinating Kipferl-independent Rhino binding sites are the Satellite arrays of the *Rsp* and *1.688* families. The role of Rhino in Satellite biology is currently unclear. Our collective evidence suggests that a conflict with selfish Satellites might contribute to the fast evolution of piRNA pathway components and promises interesting new insights into transposon-independent roles of the piRNA pathway. We were unable to identify Rhino interactors that might be involved in recruitment to Satellites. This hinders the specific interference with this process, which will be instrumental to disentangle Satellite biology from transposon silencing.

Despite a potential connection of Rhino to fast evolving Satellites, Kipferl does not show signs of fast evolution. This might be an indication that Kipferl represents a relatively old and stable solution to an evolutionary problem that later evolved independently of the robust Rhino recruitment mechanisms set up through Kipferl. In this light, it would be interesting to further investigate the role of Kipferl in other *Drosophila* species, as well as the cross-species compatibility of Kipferl and Rhino.

Our ChIP-seq data for Kipferl in *rhino* mutant ovaries, corroborated by Kipferl overexpression experiments in OSCs and GSCs, strongly indicate that Kipferl's ZnFs preferentially bind to GRGGN sequence motifs. This direct binding of Kipferl to DNA remains putative, however, pending *in vitro* binding assays with purified ZnF arrays. Moreover, the interaction between Kipferl's ZnF#4 and the Rhino chromo domain awaits further characterization. The exact residues of Rhino involved in this interaction are still unknown. In addition, the influence of DNA binding, H3K9me3 binding, as well as dimerization of both Kipferl and Rhino on the interaction of the two proteins would be of great interest. Complementary to *in vitro* studies of the interaction between Kipferl and Rhino, additional Kipferl variant constructs could further be designed to test the importance of the precise arrangement of DNA binding ZnFs and the Rhino-interacting ZnF#4 *in vivo*. This might provide insight into the mechanistic details not just of Rhino recruitment to Kipferl nucleation sites, but also the codependent spreading of both proteins along neighboring, non-motif containing sequences.

Finally, the validity of our Kipferl nucleation site model should be demonstrated by a proof of concept experiment. To this end, ectopic Kipferl binding sites close to pre-existing, non-Rhino bound heterochromatic regions could be introduced, which are expected to result in the association of Rhino with the respective locus. In an inverse scenario, the induction of heterochromatin formation at Kipferl stand-alone peaks (eg. through tethering of histone methyl transferases) should result in the formation of a Rhino domain. Alternatively, the artificial recruitment of Rhino to a completely different set of heterochromatic loci could further be induced through the fusion of Kipferl's ZnF#4 to an orthogonal array of ZnFs with altered sequence-specificity.

Without a doubt, the identification of Kipferl triggers a multitude of new questions that were behind closed doors before. Addressing the points mentioned above will not only advance our understanding of how the chromatin reader protein Rhino specifically localizes to a subset of cellular heterochromatin, but holds the potential to uncover new roles of the piRNA pathway in DNA Satellite maintenance in *Drosophila*. In addition, mechanistic insights gained from the investigation of Kipferl and Rhino might extend to other factors in entirely different areas of biology.

5. Materials and Methods

Fly strains and husbandry

All fly stocks were kept at 25°C with 12h dark/light cycles. Fly strains with genotypes, identifiers, and original sources are listed in Supplementary Table 4 and strains published in the course of this study are available from VDRC (http://stockcenter.vdrc.at/control/main). For ovary dissections, flies were aged for 2–6 days, and kept on apple juice plates with fresh yeast paste for two days.

Generation of transgenic and mutant fly strains

Frame-shift mutant alleles for *CG2678* and *rhino,* whole locus *CG2678* deletion for RMCE, endogenously tagged HP1b and HP1c, as well as fly strains harboring short hairpin RNA (shRNA) expression cassettes for germline knock down, an overexpression constructs for *CG2678,* or TurboID-vhhGFP-3xHA-NLS (cloned from on Addgene #107171) under the *eggless* enhancer were generated as described in Baumgartner et al. 2022 Frame-shift mutant alleles for CG8388 were generated through co-injection of plasmids pBS-Hsp70-Cas9 (Addgene #46294) and pU6-BbsI-chiRNA (Addgene #45946) modified to express sgRNAs in isogenised *white* embryos. sgRNA sequences are given in Table 5 Transgenic flies harboring GFP tagged wildtype or engineered Rhino constructs were generated via insertion of desired tag sequences under the control of the *rhino* or *maternal tubulin* promoter region and the *vasa* 3'UTR into the attP40 landing site (Markstein et al., 2008) in flies harboring a *rhino* frame shift mutation on the same chromosome. *CG2678* rescue constructs were introduced into the *CG2678* whole locus deletion flies as described in Baumgartner et al. 2022 Successful cassette exchange was monitored via loss of dsRed in eyes and the orientation of the inserted construct was verified by PCR. Rescue constructs harbor an internal 3xFLAG tag at residue S161, as neither amino- nor carboxy-terminal tagging of *CG2678* yielded fully functional protein. Exceptions are ZAD^{delta}, ZAD^{GCN4}, and ZAD^{linker} constructs, which harbor an N-terminal 3xFLAG tag or 3xFLAG included in an artificial linker, respectively.

Antibody generation

Mouse monoclonal antibodies against His-tagged CG2678 (aa M2-K188) and His-tagged Rhino (full length) were generated by the Max Perutz Labs Antibody Facility as described in Baumgartner et al. 2022 Polyclonal antibodies were raised against a CG2678 peptide (aa R171-I190) at Eurogentec. Antibodies and the antigens used to generate them are described in Supplementary Table 6.

Cell lines

Drosophila ovarian somatic cells (OSC) cells were cultured as previously described (Niki et al., 2006, Saito et al., 2009). *Drosophila* germline stem cells (GSCs) were cultured under the same conditions with small modifications in transfection protocols (Tirian et al., in preparation). Stable OSC and GSC lines expressing internally tagged CG2678 under control of the *ubi63E* promoter and an SV40 3'UTR were generated by integration into an RMCE landing site.

Western Blot

Western blot experiments were performed as described in Baumgartner et al. 2022 To assess protein levels of Rhino and Kipferl in ovaries, 5 pairs of ovaries were mechanically disrupted in 30 uL lysis buffer (20 mM Tris pH 8.0, 1% Triton X-100, 2 mM $MgCl_2$, Benzonase, protease inhibitors) using a plastic pestle. To compare protein levels between ovaries and cell lines, the respective samples were lysed in RIPA buffer. Protein concentrations were determined via Bradford assay to allow equal loading, and SDS-PAGE gel electrophoresis was performed according to standard procedures. Primary antibodies were incubated at 4°C overnight in 5% milk, secondary antibodies for 1 hr at RT and the blots were developed using ECL (BioRad). Antibodies and dilutions are listed in the Supplementary Table 6.

RNA Fluorescent In Situ Hybridization

RNA FISH experiments were conducted as described in Baumgartner et al. 2022 piRNA clusters *42AB* and *38C*, as well as *HMS-Beagle*, *Max*, *diver*, *3S18, and mdg3* transposons were detected using Stellaris probes (Biosearch Technologies). Probe sequences are published by Baumgartner et al or are available through the Brennecke laboratory in the case of *mdg3*. RNA FISH for *1.688* and *Rsp* Satellites was performed using a single fluorescent oligo or an in-house labelled probe set of 48 oligos, respectively (Wei et al., 2021, Gaspar et al., 2017). Images are given as Z-stack across a maximum of 2 micrometers.

Immunofluorescence staining of ovaries and testes

Immunofluorescence stainings of ovaries and testes were performed as previously described (Baumgartner et al., 2022). Briefly, 5-10 ovary pairs or testes were dissected into ice cold PBS and subsequently incubated in fixation solution (4% formaldehyde, 0.3% Triton X-100, 1x PBS) for 20 minutes at room temperature on a nutator. Fixed ovaries were washed 3x 5 minutes in PBX and blocked with 0.1% BSA for 30 minutes, all at room temperature with rotation. Primary antibodies were incubated at 4°C overnight in the presence of 0.1% BSA followed by three 5-minute washes in PBX. Incubation with fluorophore-coupled secondary antibodies was done overnight at 4°C, with 3 subsequent washes in PBX, including DAPI in the first wash to stain DNA (1:50,000 dilution). The final wash buffer was carefully removed before mounting in ∼40 µL DAKO mounting medium. The samples were imaged on a Zeiss LSM-880 Axio Imager confocal-microscope and image processing was done using FIJI/ImageJ (Schindelin et al., 2012). Images are given as Z-stack across a maximum of 2 micrometer. All relevant antibodies and dilutions are listed in the Supplementary Table 6.

Yeast Two-Hybrid Assay

Screening for Rhino interactors: Yeast two-hybrid screening for Rhino interactors in ovaries and testes was performed by Hybrigenics Services, S.A.S., Paris, France (https://www.hybrigenics-services.com) as described in Baumgartner et al. 2022 Upon screening of Rhino against a cDNA library obtained from *Drosophila* ovaries with 65 million interactions screened, 173 colonies were selected on a medium lacking tryptophan, leucine,

and histidine. Screening against a cDNA library obtained from *Drosophila* testes yielded 53 sequenced colonies from 108 million tested interactions.

Validation and interaction mapping: Yeast strains were grown in YPD or SC selective medium at 30˚C and individual interactions were assayed as described previously (Baumgartner et al., 2022). Briefly, assayed proteins were fused to the activation domain (AD) and DNA-binding domain (DBD) of the Gal4 transcription factor and transformed into yeast strains PJ694A (AD) and PJ694α (DBD). Individually transformed colonies were selected, picked, and mated. Dillution series were spotted on selective (-LTH) plates to assay interactions. Parallel plating on non-selective (-LT) plates controlled for presence of both plasmids. Yeast strains used in this study are listed in Supplementary Table 7.

Biotin proximity labeling

Proximity-dependent biotin labelling of proteins in the proximity of Rhino was performed as described in Baumgartner et al. 2022 In brief, flies expressing GFP-Rhino were crossed to flies expressing low levels of TurboID biotin ligase fused to GFP-nanobody and the progeny was fed with yeast paste containing 100 µM biotin (Sigma) on apple juice plates in cages for 16h prior to ovary dissection (adapted from (Roux et al. 2018). Flies expressing GFP fused to a nuclear localization signal were crossed to GFP-nanobody-TurboID flies as a negative control. 100 µl of ovaries were washed once with ice-cold PBS and dounced (6 times) in 1.2 ml of pre-extraction buffer (10 mM Tris-HCl pH 7.5, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 0.5% NP40, 10% glycerol, cOmplete Protease Inhibitor Cocktail (Roche)). Incubation at 4 °C with nutation for 15 min and followed by centrifugation for 5 min at 20,000 g at 4 °C was used for mild pre-extraction of cytoplasmic contaminants. The nuclear-enriched fraction was resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% SDS, 0.5% Na-Deoxycholate, 1% Triton-X, 1 mM DTT, Benzonase, cOmplete Protease Inhibitor Cocktail (Roche)) and homogenized before incubation at 4 °C with nutation for 2h. The lysate was cleared twice by centrifugation at 18,000 g for 10 min and mixed with 100 µL magnetic Pierce Streptavidin beads (ThermoFischer) pre-equilibrated in lysis buffer, followed by overnight incubation at 4 °C. The beads were washed once with lysis buffer for 10 min at 4 °C, with 2% SDS (10 min at room temperature), and 10 minutes each with wash buffers 1 (50mM HEPES pH 7.5, 500 mM NaCl, 1 mM EDTA, 0.1% Na-Deoxycholate, 1% Triton-X) and 2 (10mM Tris-HCl pH 7.5, 250 mM LiCl, 1 mM EDTA, 0.5% Na-Deoxycholate, 1% NP40) for a minimum of 10 min at 4 °C, followed by 5 washes without detergent (20mM Tris-HCl pH 7.5, 137mM NaCl) before further downstream mass spectrometry analysis according to previous reports (Batki et al., 2019).

Scoring of embryo hatching rates

Female fertility was measured as previously described (Baumgartner et al., 2022). Briefly, 10 virgin females were collected and aged for 2-3 days and mated with three w^{1118} males for at least 24h before scoring. The hatching rate as percentage of hatched eggs from total was determined 30 hours after egg laying (25 degrees) onto apple juice plates. Plates with less than 50 eggs were disregarded in the analysis. Wildtype females were included in every

experiment as control.

Definition and Curation of 1 kb Genomic Tiles

The four assembled chromosomes of the *Drosophila melanogaster* genome (dm6 assembly (Larkin et al., 2021)) were split into non-overlapping 1-kb tiles with annotations for piRNA clusters and filtering for mappability and background artefacts as described in Baumgartner et al. 2022. 116'000 genomic 1-kb tiles were analyzable according to these criteria.

Heterochromatin and Euchromatin Definitions Used in This Study

We used H3K9me3 ChIP-seq data from wildtype ovaries to define the extent of pericentromeric heterochromatin and euchromatic chromosome arms. The heavily H3K9me3 covered pericentric regions of the assembled chromosomes, as well as the entire chromosome 4 were classified as heterochromatic, while the rest was annotated as euchromatic (Baumgartner et al., 2022). Detailed coordinates can be found in Supplementary Table 1. For analyses of 1-kb tiles, piRNA clusters *42AB*, *38C*, and *80F* were not included into either heterochromatin or euchromatin, but were analyzed separately as reference loci. Small genome contigs not assembled into the four major chromosomes were excluded from all analyses.

ChIP-Seq

ChIP was performed as previously described (Lee et al., 2006). For ChIP from OSCs or GSCs 5-10 million cells were crosslinked with 1.8% formaldehyde in PBS for 10 min at room temperature, quenched with Glycine, and lysed. For ChIP from ovaries, 150 µl of ovaries were dissected into ice-cold PBS, crosslinked, quenched, rinsed in PBS and flash frozen in liquid nitrogen after removing all PBS. Frozen ovaries were disrupted in PBS using a dounce homogenizer, centrifuged at low speed and the pellet was resuspended in lysis buffer. Sonication (Bioruptor) resulted in DNA fragment sizes of 200–800 bp. Immunoprecipitation with antibodies specified in Supplementary Table 6 was done with rotation overnight at 4°C in 350–700 µl total volume using 1/3 to 1/4 of chromatin per ChIP. 40 µl Dynabeads (equal mixture of Protein G and A, Invitrogen) were then added and incubated for 1 hr at 4˚ with rotation. After extensive washes, immuno-precipitated protein-DNA complexes were eluted with 1% SDS, treated with RNAse-A, decrosslinked overnight at 65°C, and proteins were digested with proteinase K before clean-up using ChIP DNA Clean & Concentrator columns (Zymo Research). Barcoded libraries were prepared according to manufacturer's instructions using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB), and sequenced on a HiSeqV4, NextSeq550, or NovaSeqSP (Illumina) (Baumgartner et al., 2022).

RNA-Seq

Strand-specific RNA seq was performed as described previously (Zhang et al., 2012c). In brief, total RNA was extracted from 5–10 ovaries from 7-day old flies using Trizol (Invitrogen). Total RNA was purified using RNAeasy columns (QIAGEN) and six micrograms of total RNA were subjected to polyA selection and subsequent fragmentation, reverse transcription, and library

preparation according to manufacturer's instructions using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) for sequencing on an Illumina NovaSeqSP instrument (Baumgartner et al., 2022).

Small RNA-Seq

Small RNA cloning was performed as described in (Grentzinger et al., 2020). In brief, ovaries were lysed applied to TraPR ion exchange spin columns for isolation of Argonaute-sRNA complexes. sRNAs were subsequently purified using Trizol and subjected to ligations of 3' and 5' barcoded adapters containing 4 random nucleotides at the ends to reduce ligation biases, reverse transcribed, PCR amplified, and sequenced on an Illumina NextSeq550 instrument (Baumgartner et al., 2022).

Computational Analysis

ChIP-Seq Analysis: ChIP-seq reads were processed as described in Baumgartner et al. 2022 In brief, reads were trimmed to remove the adaptor sequences and to adjust all reads to 50 bp irrespective of sequencing mode. Reads were mapped to the dm6 genome using Bowtie (version.1.3.0), allowing up to three mismatches. Genome unique reads were mapped to 1-kb tiles and a pseudocount of 1 was added after normalization to library depth, before enrichment over input values were determined. Each ChIP-seq sample was adjusted with a correction factor determined from median input levels and median background levels to reach median background enrichment of 1 to correct for unequal ChIP efficiency (Baumgartner et al., 2022). To classify genomic regions into Rhino domains and non-Rhino domains, we used a binary cutoff of 4-fold enrichment calculated from two independent replicate experiments of the relevant wildtype genotypes. This cutoff corresponds to a p-value of < 0.05. Kipferl-only 1-kb tiles were defined by lack of Rhino enrichment (below 4-fold) and significant enrichment of Kipferl over Rhino (Z-score = 3). Replicates were averaged for analyses of genomic 1-kb tiles.

ChIP-seq peak calling: We used MACS2 (Zhang et al., 2008) with --broad --broad-cutoff 0.1 for Kipferl and Rhino in wildtype ovaries due to the broad extent of Rhino/Kipferl domains (Baumgartner et al., 2022). The 'narrow peak' setting was used for the remaining experiments. Peaks mapping to genomic contigs outside the four main chromosomes were discarded, and peaks were filtered for a score of 50 (broad peaks; $p<10^{-5}$) and 30 (narrow peaks; $p<10^{-3}$). Kipferl-Rhino shared versus Kipferl-only peaks were distinguished as described previously (Baumgartner et al., 2022) by using bedtools intersect. Grouping into heterochromatic and euchromatic peaks was done by intersection with heterochromatin coordinates outlined above. Kipferl DNA binding motifs were recovered from the top ~3000 summits of Rhino-independent Kipferl peaks (achieved through a score cutoff of 7 on summits, Baumgartner et al., 2022) using HOMER (Heinz et al., 2010). Heatmaps display one representative replicate and were produced through deeptools (Ramirez et al., 2016).

ChIP-seq analysis on transposon consensus sequences: Mapping of ChIP-seq reads to transposon consensus sequences was performed as described previously (Baumgartner et al.,

2022). Genome mapping reads longer than 23 nucleotides were mapped to TE consensus sequences using bowtie allowing up to 3 mismatches as described in Baumgartner et al. 2022 Reads mapping to multiple elements were assigned to the best mapping position. Reads mapping to multiple positions were randomly distributed. Library depth normalized ChIP and input reads, respectively, were averaged over all nucleotide positions of each element to give one value per element. ChIP-seq enrichment was calculated after adding a pseudo count of 1 and adjusted using sample-specific correction factors determined from background 1 kb tiles to reach median background enrichments of 1 (Baumgartner et al., 2022). Corrected per-base enrichment was calculated for TE ChIP-seq profiles and replicates were averaged.

Motif instances: Occurrences of Kipferl DNA binding motifs were determined using PWMScan (ccg.epfl.ch/pwmtools /pwmscan.php) on the dm6 genome and on TE consensus sequences. For display in heatmaps, cumulative motif counts on both genomic strands were intersected with non-overlapping 100 bp windows (Baumgartner et al., 2022).

smallRNA-Seq Analysis: Analysis of smallRNA-seq reads was performed as described in Baumgartner et al. 2022 Briefly, raw reads were trimmed for linker sequences and the 4 random nucleotides flanking the small RNA before mapping to the *Drosophila melanogaster* genome (dm6), using Bowtie (version.1.3.0) with 0 mismatches allowed. Genome mapping reads were intersected with Flybase genome annotations (r6.40) using Bedtools to allow the removal of reads mapping to rRNA, tRNA, snRNA, snoRNA loci and the mitochondrial genome. For TE mappings, all genome mappers were used allowing no mismatches. Reads mapping to multiple elements were assigned to the best match. Reads mapping equally well to multiple positions were randomly distributed. Libraries were normalized to 1 Mio miRNA reads. For the 1 kb tile analysis, a pseudocount of 1 was added after normalization to library depth and correction for the mappability of the respective 1-kb tile. Tiles with fewer than 10 mapping piRNAs in all libraries were disregarded for sRNA analysis to avoid distortion due to very low abundant piRNAs. For calculation of piRNAs mapping to TEs, sense and antisense piRNAs were kept separate, and counts were normalized to transposon length. For classification of tiles and transposons into somatic, Rhino-independent, and Rhino-dependent source loci, the soma index was determined as the log2 ratio of somatic (Piwi-IP in *piwi* GLKD, normalized to library depth) and germline (GL-Piwi IP, normalized to library depth) piRNAs mapping to each tile or TE (Mohn et al., 2014). Classification by Kipferl-dependency of TEs was achieved by a binary cutoff of at a 2-fold reduction in antisense piRNA levels in *kipferl* knock down compared to control (Baumgartner et al., 2022).

RNA-Seq Analysis: For RNA-seq analysis, genome matching reads were randomized in order and quantified using Salmon (v.1.7.0) as described previously (Baumgartner et al., 2022). For the analysis we used the FlyBase transcriptome (r6.40) masked for sequences similar to transposons. TE-consensus sequences were added to the FlyBase transcriptome in sense and antisense orientation to include both strands in the analysis. For gene expression visualization Salmon results were further processed to GeTMM values using edgeR (v3.34.0), and for

differential gene expression analysis Salmon results were processed using DeSeq2 (v1.32.0) (Baumgartner et al., 2022).

TE Insertion Calling: Euchromatic TE insertions were extracted from insertions called previously (Mohn et al., 2014), and nearby insertions were merged using bedtools merge -d 100, as these mostly corresponded to the same insertions called at slightly different positions (Baumgartner et al., 2022). Further, all insertions overlapping UCSC repeat masker track annotations were discarded. For analysis of non-Kipferl nucleation site containing, but Rhino-dependent TEs, the resulting list was subsequently filtered for elements with no Kipferl enrichment in *rhino* knock out ovaries and at least 10-fold difference in piRNA levels between control and *rhino* MTD-Gal4 mediated germline knock down (Baumgartner et al., 2022). This retrieved 285 euchromatic solo TE insertion sites in the genome of our experimental MTD-Gal4 strains (Baumgartner et al., 2022).

Supplementary Tables

Table 1: Chromosomal coordinates of heterochromatic and euchromatic compartments used in this study.

chr	start	end	compartment
2L	0	22115000	euchromatin
2L	22115001	23513712	heterochromatin
2R	0	5812495	heterochromatin
2R	5812496	25286936	euchromatin
3L	0	23006900	euchromatin
3L	23006901	28110227	heterochromatin
3R	0	4174279	heterochromatin
3R	4174280	32079331	euchromatin
4	0	1348131	heterochromatin
X	0	21510000	euchromatin
X	21510001	23542271	heterochromatin

Table 2: Putative Rhino interactors identified in yeast two-hybrid screen against ovarian cDNA library.

gene name	nr of in frame colonies
Del	83
Kap-alpha3	19
CG2678	12
E(bx)	9
Nelf-A	5
Br140	3
msn varB	3
CG9114	2
CrebB	2
ttk	2
mtRNApol	2
RpL22	2
Mrtf	1
CG2662	1
CG43333	1
CG44774	1
CG5098	1
CG7922	1
CG8388	1
CG9114	1
Cfp1	1
GATAd	1
HP1c	1
Nup98-96	1
Taf1	1
msn varA	1

Table 3: Putative Rhino interactors identified in yeast two-hybrid screen against testes derived cDNA library.

gene name	nr of in frame colonies
CG17765	8
CG44774	7
CG8388	4
CG14668	3
CG11060	2
CG14232	2
CG14402	2
CG2841	2
Got1	2
Tom70	2
betaTub85D	2
CG14190	1
CG17010	1
CG31523	1
HP1c	1
Kapalpha3	1
S-Lap1	1
prd	1
sm	1

Table 4: Fly strains used in this study.

Genotype	Source
w^{1118};;;	Bloomington stock 3605
w;;; (*iso1*)	Susan Celniker
Cog-GAL4; NGT-GAL4; nos-GAL4;	Bloomington stock 31777
w; ; pW20>w_sh[attP2]/TM3, Sb;	Mohn *et al.*, 2014, VDRC-ID 313772
w; ; pW20>rhi_sh[attP2]/TM3, Sb;	Mohn *et al.*, 2014, VDRC-ID 313156
w; ; pW20>CG2678_sh2[attP2]/TM3, Sb;	Baumgartner *et al.,* 2022
w; ; pGLKD>piwi_sh2[attP2]/TM3,Sb;	Senti *et al.,* 2015, VDRC-ID 313199
w;; CG2678[Δ1](dsRed+)/TM3,Sb;	Baumgartner *et al.,* 2022
w;; CG2678[Δ2](dsRed+)/TM3,Sb;	Baumgartner *et al.,* 2022
w;; CG2678[fs1]/TM3,Sb;	Baumgartner *et al.,* 2022
w;; CG2678RMCE[S161-3xFLAG]/TM3,Sb;	Baumgartner *et al.,* 2022
w; pRhino>3xFLAG/V5/Precission/GFP-Rhino[attP40],rhi[g2m11]/CyO ; CG2678[Δ1](dsRed+),nxf3[A2-2]/TM3,Sb ;	Baumgartner *et al.,* 2022
w; pRhino>3xFLAG/V5/Precission/GFP-Rhino[attP40],rhi[g2m11]/CyO ; ;	Baumgartner *et al.,* 2022
w; pRhino>3xFLAG/V5/Precission/GFP-Rhino[attP40],rhi[g2m11]/CyO ; CG2678[Δ1](dsRed+)/TM3,Sb ;	Baumgartner *et al.,* 2022
w; rhi[18-7]/CyO; ;	Andersen *et al*., 2017, VDRC-ID 313488
w; rhi[g2m11]/CyO ;;	Baumgartner *et al.,* 2022
w;; pUASz>CG2678[attP2];	Baumgartner *et al.,* 2022
w;; CG2678RMCE[ΔZnFarray1-S161-3xFLAG]/TM3,Sb;	Baumgartner *et al.,* 2022
w;; CG2678RMCE[ΔZnFarray2-S161-3xFLAG]/TM3,Sb;	Baumgartner *et al.,* 2022
w;; CG2678RMCE[3xFLAG-ΔZAD]/TM3,Sb;	Baumgartner *et al.,* 2022
w;; CG2678RMCE[3xFLAG-ZAD::GCN4]/TM3,Sb;	Baumgartner *et al.,* 2022
w;; CG2678RMCE[3xFLAG-ouibZAD]/TM3,Sb;	Baumgartner *et al.,* 2022
w;; CG2678RMCE[ΔZnF#4-S161-3xFLAG]/TM3,Sb;	Baumgartner *et al.,* 2022
w; pRhino>3xFLAG/V5/Precission/GFP-Rhino(CSD::GCN4)[attP40],rhi[g3m13]/CyO ;;	Baumgartner *et al.,* 2022
w; pRhino>3xFLAG/V5/Precission/GFP-Rhino(art.hinge)[attP40],rhi[g3m13]/CyO ;;	Baumgartner *et al.,* 2022
w;; pRhino>3xFLAG/V5/Precission/GFP-Su(var)2-5[attP2];	Baumgartner *et al.,* 2022
w, 3xFLAG/V5/Precission/GFP-HP1b; ; ;	Baumgartner *et al.,* 2022
w;; 3xFLAG/V5/Precission/GFP-HP1c ;	Baumgartner *et al.,* 2022
w; peggless>TurboID-linker-vhhGFP-3xHA-NLS[attP40]/CyO;;	Baumgartner *et al.,* 2022

w; ; 3xFLAG/V5/Precission/GFP(replacing CG13741 CDS)+NLS [attP2]/TM3,Sb;	ElMaghraby, Andersen et al. 2019
w; pRhino>3xFLAG/V5/Precission/GFP-RRH[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pRhino>3xFLAG/V5/Precission/GFP-RHH[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pRhino>3xFLAG/V5/Precission/GFP-RHR[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pRhino>3xFLAG/V5/Precission/GFP-HHR[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pRhino>3xFLAG/V5/Precission/GFP-HRR[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pRhino>3xFLAG/V5/Precission/GFP-HRH[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pRhino>3xFLAG/V5/Precission/GFP-HRH[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pmtub>3xFLAG/V5/Precission/GFP-NTEswap[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pmtub>3xFLAG/V5/Precission/GFP-RRH[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pmtub>3xFLAG/V5/Precission/GFP-RHH[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pmtub>3xFLAG/V5/Precission/GFP-RHR[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pmtub>3xFLAG/V5/Precission/GFP-HHR[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pmtub>3xFLAG/V5/Precission/GFP-HRR[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pmtub>3xFLAG/V5/Precission/GFP-HRH[attP40],rhi[g2m11]/CyO ; ;	this thesis
w; pmtub>3xFLAG/V5/Precission/GFP-NTEswap[attP40],rhi[g2m11]/CyO ; ;	this thesis
w;; pUASz>3xFLAG-HP1aCDtandem-GFP[attP2];	this thesis
w;; pUASz>3xFLAG-HP1cCDtandem-GFP[attP2];	this thesis
w;; pUASz>3xFLAG-RhinoCDtandem-GFP[attP2];	this thesis
w;; pW20>CG2678-piwi_doublesh[attP2];	this thesis
w;; pW20>CG2678-CG8388_doublesh[attP2];	this thesis
w;; pW20>Rhi-piwi_doublesh[attP2];	this thesis
y1; pW20>CG8388_sh[attP40]/CyO;;	Bloomington stock 57545
w; CG8388[fs1]/CyO;;	this thesis*
w; CG8388[fs1]/CyO; CG2678[Δ1](dsRed+)/TM3,Sb;	this thesis
w; 42AB[prom.del]/CyO; CG2678[Δ1](dsRed+)/TM3,Sb;	this thesis

w; 38C[prom.del]/CyO; CG2678[Δ1](dsRed+)/TM3,Sb;	this thesis
w; 38C[deletion]/CyO;;	Gebert et al., 2021
w; 42AB[deletion]/CyO;;	Gebert et al., 2021
w; 38C[deletion], 42AB[deletion]/CyO;;	Gebert et al., 2021
w; 38C[deletion], 42AB[deletion]/CyO; CG2678[Δ1](dsRed+)/TM3,Sb;	this thesis (piRNA cluster deletion fly from Gebert et al., 2021)

* CG8388g2m2-1: indel -32; Sequence:

ATG--------------------------------ACGGGAGCACAAGCAGCGCGGCTAT

Table 5: Sequences of guide RNAs used in this study.

targeted locus	guide number	sequence	purpose
Rhino	g2	GACCAAGATTTGGTCGCTGA	indel
Rhino	g3	GTCCCCAGGTTCTGGTGAAG	indel
CG2678	g1	GTACAAATGATCAGTGCGA	RMCE left
CG2679	g2	GAAGGCATTAAGTAGCATG	RMCE right
CG2678	g3	GAACCGGAAAGCATTCTGCA	indel
HP1b	g2	CACAATGGCCGAATTCTCAG	N-terminal endogenous tagging
HP1c	g2	GATGCGCTCCACCACGAAGT	N-terminal endogenous tagging
CG8388	G2	GATGCTGGCCCATGTACGA	indel

identifier	published	comment	source	use	dilution	species	name
CG2678#2_4 P39glyc	Baumgartner et al., 2022	raised against Kipferl peptide R171-I190	Eurogentec	ChIP	7 uL per IP	rabbit	anti-CG2678#2
M4 3A6-3E1-E11	Baumgartner et al., 2022	raised against Kipferl amino acids M2-K188	Monoclonal Antibody Facility	WB	1/250	mouse	anti-CG2678 monoclonalM4
M3 2C5-3C3	Baumgartner et al., 2022	raised against Kipferl amino acids M2-K188	Monoclonal Antibody Facility	IF	1/500	mouse	anti-CG2678 monoclonalM3
6B7-F2	Baumgartner et al., 2022	raised against full length denatured Rhino protein	Monoclonal Antibody Facility	WB	1/10	mouse	anti-Rhino monoclonal
Rhino#1_357 3gly	Mohn et al., 2014	-	Eurogentec	ChIP, IF	5 uL per IP, 1/1000	rabbit	anti-Rhino polyclonal
5B5-6D7-3H10	Andersen et al., 2017	-	Monoclonal Antibody Facility	IF	1/100	mouse	anti-Deadlock
8E4-F1	ElMaghraby, Andersen et al., 2019	-	Monoclonal Antibody Facility	IF	1/50	mouse	anti-Nxf3
active motif 39161	-	-	Active motif	ChIP, IF	5 uL per IP, 1/100	rabbit	Histone H3K9me3
mAbcam 1220	-	-	Abcam	ChIP	5 uL per ChIP	mouse	Anti-Histone H3 (di methyl K9)
DSHB C1A9	-	-	DSHB	ChIP, IF	5 uL per IP, 1/100	mouse	C1A9 HP1 (Su(var)2-5)
Sigma F1804-1MG	-	-	Sigma Aldrich	ChIP, IF	2 uL per IP, 1/2000	mouse	ANTI-FLAG(R) M2

Table 6: List of antibodies used in this study.

Table 7: *Saccharomyces cerevisiae* strains used in this study.

name	genotype	comments	mother strain
PJ694α	*Matα, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, GAL2-ADE2, LYS2::GAL1-HIS3, met2::GAL7-lacZ*	(Mohn et al., 2014)	-
PJ694A	*MatA, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, GAL2-ADE2, LYS2::GAL1-HIS3, met2::GAL7-lacZ*	(Mohn et al., 2014)	-
YGS13	*pOAD*	(Mohn et al., 2014)	YGS2 (MatA)
YGS15	*pOAD-rhino*	(Mohn et al., 2014)	YGS2 (MatA)
YGS16	*pOBD-rhino*	(Mohn et al., 2014)	YGS1 (Matα)
YGS19	*pOAD-HP1a*	(Mohn et al., 2014)	YGS2 (MatA)
YGS39	*pOAD-rhino(1-100)*	(Mohn et al., 2014)	YGS2 (MatA)
YGS41	*pOAD-rhino(101-300)*	(Mohn et al., 2014)	YGS2 (MatA)
YGS43	*pOAD-rhino(301-418)*	(Mohn et al., 2014)	YGS2 (MatA)
YLB56	*pOBD-CG2678-ZF4nolinker*	(Baumgartner et al., 2022)	YGS1 (Matα)
YLB59	*pOBD-CG2678-190-311*	(Baumgartner et al., 2022)	YGS1 (Matα)
YLB60	*pOAD-CG2678FL*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB61	*pOAD-CG2678frag1*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB62	*pOAD-CG2678frag2*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB63	*pOAD-CG2678frag3*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB64	*pOAD-CG2678frag4*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB65	*pOAD-CG2678frag5*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB66	*pOAD-CG2678-ZF1-2*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB67	*pOAD-CG2678-ZF1-3*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB68	*pOAD-CG2678ZF2-4*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB69	*pOAD-CG2678ZF3-4*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB71	*pOAD-CG2678ZF4nolinker*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB73	*pOAD-ouibZAD*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB76	*pOAD-Rhi19-85*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB77	*pOAD-CG2678-190-311*	(Baumgartner et al., 2022)	YGS2 (MatA)
YLB78	*pOBD-CG2678frag1*	(Baumgartner et al., 2022)	YGS1 (Matα)
YLB79	*pOAD-HP1CD*	(Baumgartner et al., 2022)	YGS2 (MatA)

References

AASLAND, R. & STEWART, A. 1995. The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1. *Nucleic Acids Res,* 23**,** 3163-73.

ABAD, J. P., AGUDO, M., MOLINA, I., LOSADA, A., RIPOLL, P. & VILLASANTE, A. 2000. Pericentromeric regions containing 1.688 satellite DNA sequences show anti-kinetochore antibody staining in prometaphase chromosomes of Drosophila melanogaster. *Mol Gen Genet,* 264**,** 371-7.

AKULENKO, N., RYAZANSKY, S., MORGUNOVA, V., KOMAROV, P. A., OLOVNIKOV, I., VAURY, C., JENSEN, S. & KALMYKOVA, A. 2018. Transcriptional and chromatin changes accompanying de novo formation of transgenic piRNA clusters. *RNA,* 24**,** 574-584.

ALTEMOSE, N., LOGSDON, G. A., BZIKADZE, A. V., SIDHWANI, P., LANGLEY, S. A., CALDAS, G. V., HOYT, S. J., URALSKY, L., RYABOV, F. D. & SHEW, C. J. 2022. Complete genomic and epigenetic maps of human centromeres. *Science,* 376**,** eabl4178.

ANDERSEN, P. R., TIRIAN, L., VUNJAK, M. & BRENNECKE, J. 2017. A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature,* 549**,** 54-59.

ARAVIN, A. A. & BOURC'HIS, D. 2008. Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes Dev,* 22**,** 970-5.

ARAVIN, A. A., HANNON, G. J. & BRENNECKE, J. 2007a. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *science,* 318**,** 761-764.

ARAVIN, A. A., SACHIDANANDAM, R., BOURC'HIS, D., SCHAEFER, C., PEZIC, D., TOTH, K. F., BESTOR, T. & HANNON, G. J. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell,* 31**,** 785-99.

ARAVIN, A. A., SACHIDANANDAM, R., GIRARD, A., FEJES-TOTH, K. & HANNON, G. J. 2007b. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science,* 316**,** 744-7.

BAG, I., CHEN, S., ROSIN, L. F., CHEN, Y., LIU, C. Y., YU, G. Y. & LEI, E. P. 2021. M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity. *Nat Commun,* 12**,** 4170.

BATEMAN, J. R., LEE, A. M. & WU, C.-T. 2006. Site-specific transformation of Drosophila via φC31 integrase-mediated cassette exchange. *Genetics,* 173**,** 769-777.

BATKI, J., SCHNABL, J., WANG, J., HANDLER, D., ANDREEV, V. I., STIEGER, C. E., NOVATCHKOVA, M., LAMPERSBERGER, L., KAUNECKAITE, K., XIE, W., MECHTLER, K., PATEL, D. J. & BRENNECKE, J. 2019. The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat Struct Mol Biol,* 26**,** 720-731.

BAUMGARTNER, L., HANDLER, D., PLATZER, S., DUCHEK, P. & BRENNECKE, J. 2022. The Drosophila ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. *bioRxiv***,** 2022.05.09.491178.

BIESSMANN, H., PRASAD, S., SEMESHIN, V. F., ANDREYEVA, E. N., NGUYEN, Q., WALTER, M. F. & MASON, J. M. 2005. Two distinct domains in Drosophila melanogaster telomeres. *Genetics,* 171**,** 1767-1777.

BLUMENSTIEL, J. P., ERWIN, A. A. & HEMMER, L. W. 2016. Focus: Epigenetics: What drives positive selection in the drosophila pirna machinery? the genomic autoimmunity hypothesis. *The Yale journal of biology and medicine,* 89**,** 499.

BOEKE, J. D., GARFINKEL, D. J., STYLES, C. A. & FINK, G. R. 1985. Ty elements transpose through an RNA intermediate. *Cell,* 40**,** 491-500.

BONCHUK, A., BOYKO, K., FEDOTOVA, A., NIKOLAEVA, A., LUSHCHEKINA, S., KHRUSTALEVA, A., POPOV, V. & GEORGIEV, P. 2021. Structural basis of diversity and

homodimerization specificity of zinc-finger-associated domains in Drosophila. *Nucleic Acids Res,* 49**,** 2375-2389.

BOURGEOIS, Y. & BOISSINOT, S. 2019. On the population dynamics of junk: a review on the population genomics of transposable elements. *Genes,* 10**,** 419.

BOURQUE, G., BURNS, K. H., GEHRING, M., GORBUNOVA, V., SELUANOV, A., HAMMELL, M., IMBEAULT, M., IZSVÁK, Z., LEVIN, H. L. & MACFARLAN, T. S. 2018. Ten things you should know about transposable elements. *Genome biology,* 19**,** 1-12.

BRASHER, S. V., SMITH, B. O., FOGH, R. H., NIETLISPACH, D., THIRU, A., NIELSEN, P. R., BROADHURST, R. W., BALL, L. J., MURZINA, N. V. & LAUE, E. D. 2000. The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. *The EMBO journal,* 19**,** 1587-1597.

BRAYER, K. J. & SEGAL, D. J. 2008. Keep your fingers off my DNA: protein–protein interactions mediated by C2H2 zinc finger domains. *Cell biochemistry and biophysics,* 50**,** 111-131.

BRENNECKE, J., ARAVIN, A. A., STARK, A., DUS, M., KELLIS, M., SACHIDANANDAM, R. & HANNON, G. J. 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell,* 128**,** 1089-103.

BRUNO, M., MAHGOUB, M. & MACFARLAN, T. S. 2019. The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu Rev Genet,* 53**,** 393-416.

CACCHIONE, S., CENCI, G. & RAFFA, G. D. 2020. Silence at the end: how Drosophila regulates expression and transposition of telomeric retroelements. *Journal of molecular biology,* 432**,** 4305-4321.

CAIZZI, R., CAGGESE, C. & PIMPINELLI, S. 1993. Bari-1, a new transposon-like family in Drosophila melanogaster with a unique heterochromatic organization. *Genetics,* 133**,** 335-345.

CANZIO, D., LARSON, A. & NARLIKAR, G. J. 2014. Mechanisms of functional promiscuity by HP1 proteins. *Trends in cell biology,* 24**,** 377-386.

CANZIO, D., LIAO, M., NABER, N., PATE, E., LARSON, A., WU, S., MARINA, D. B., GARCIA, J. F., MADHANI, H. D. & COOKE, R. 2013. A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature,* 496**,** 377-381.

CASACUBERTA, E. 2017. Drosophila: Retrotransposons Making up Telomeres. *Viruses,* 9.

CASALE, A. M., CAPPUCCI, U., FANTI, L. & PIACENTINI, L. 2019. Heterochromatin protein 1 (HP1) is intrinsically required for post-transcriptional regulation of Drosophila Germline Stem Cell (GSC) maintenance. *Scientific reports,* 9**,** 1-12.

CASIER, K., AUTAA, J., GUEGEN, N., DELMARRE, V., MARIE, P. P., RONSSERAY, S., CARRE, C., BRASSET, E., TEYSSET, L. & BOIVIN, A. 2022. The histone demethylase KDM3 prevents auto-immune piRNAs production in Drosophila. *bioRxiv*.

CASIER, K., DELMARRE, V., GUEGUEN, N., HERMANT, C., VIODÉ, E., VAURY, C., RONSSERAY, S., BRASSET, E., TEYSSET, L. & BOIVIN, A. 2019. Environmentally-induced epigenetic conversion of a piRNA cluster. *Elife,* 8.

CASTRO-DIAZ, N., ECCO, G., COLUCCIO, A., KAPOPOULOU, A., YAZDANPANAH, B., FRIEDLI, M., DUC, J., JANG, S. M., TURELLI, P. & TRONO, D. 2014. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes & development,* 28**,** 1397-1409.

CHANG, C.-H., CHAVAN, A., PALLADINO, J., WEI, X., MARTINS, N. M., SANTINELLO, B., CHEN, C.-C., ERCEG, J., BELIVEAU, B. J. & WU, C.-T. 2019. Islands of retroelements are major components of Drosophila centromeres. *PLoS biology,* 17**,** e3000241.

CHEN, P., LUO, Y. & ARAVIN, A. A. 2021. RDC complex executes a dynamic piRNA program during Drosophila spermatogenesis to safeguard male fertility. *PLoS genetics,* 17**,** e1009591.

CHEN, Y.-C. A., STUWE, E., LUO, Y., NINOVA, M., LE THOMAS, A., ROZHAVSKAYA, E., LI, S., VEMPATI, S., LAVER, J. D. & PATEL, D. J. 2016. Cutoff suppresses RNA polymerase II termination to ensure expression of piRNA precursors. *Molecular cell,* 63**,** 97-109.

CHUNG, H.-R., LÖHR, U. & JÄCKLE, H. 2007. Lineage-specific expansion of the zinc finger associated domain ZAD. *Molecular biology and evolution,* 24**,** 1934-1943.

CHUNG, H.-R., SCHÄFER, U., JÄCKLE, H. & BÖHM, S. 2002. Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in Drosophila. *EMBO reports,* 3**,** 1158-1162.

CHUONG, E. B., ELDE, N. C. & FESCHOTTE, C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics,* 18**,** 71-86.

CORNEO, G., BIANCHI, P., GINELLI, E. & POLLI, E. 1966. Heterogeneity of base composition of DNA extracted from human leukaemic leucocytes. *European Journal of Cancer (1965),* 2**,** 307-312.

COSBY, R. L., CHANG, N.-C. & FESCHOTTE, C. 2019. Host–transposon interactions: conflict, cooperation, and cooption. *Genes & development,* 33**,** 1098-1116.

CRICK, F., BARNETT, L., BRENNER, S. & WATTS-TOBIN, R. J. 1961. General nature of the genetic code for proteins.

CZECH, B., PREALL, J. B., MCGINN, J. & HANNON, G. J. 2013. A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway. *Mol Cell,* 50**,** 749-61.

DE VANSSAY, A., BOUGÉ, A.-L., BOIVIN, A., HERMANT, C., TEYSSET, L., DELMARRE, V., ANTONIEWSKI, C. & RONSSERAY, S. 2012. Paramutation in Drosophila linked to emergence of a piRNA-producing locus. *Nature,* 490**,** 112-115.

DOHENY, J. G., MOTTUS, R. & GRIGLIATTI, T. A. 2008. Telomeric position effect—a third silencing mechanism in eukaryotes. *PloS one,* 3**,** e3864.

EICKBUSH, T. H. & EICKBUSH, D. G. 2007. Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics,* 175**,** 477-485.

EISSENBERG, J., MORRIS, G., REUTER, G. & HARTNETT, T. 1992. The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. *Genetics,* 131**,** 345-52.

EISSENBERG, J. C. & ELGIN, S. C. 2000. The HP1 protein family: getting a grip on chromatin. *Current opinion in genetics & development,* 10**,** 204-210.

EISSENBERG, J. C. & ELGIN, S. C. 2014. HP1a: a structural chromosomal protein regulating transcription. *Trends in Genetics,* 30**,** 103-110.

ELMAGHRABY, M. F., ANDERSEN, P. R., PUHRINGER, F., HOHMANN, U., MEIXNER, K., LENDL, T., TIRIAN, L. & BRENNECKE, J. 2019. A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. *Cell,* 178**,** 964-979 e20.

ELMAGHRABY, M. F., TIRIAN, L., SENTI, K. A., MEIXNER, K. & BRENNECKE, J. 2022. A genetic toolkit for studying transposon control in the Drosophila melanogaster ovary. *Genetics,* 220.

ENCODE, P. C. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57.

FAN, J. Y., RANGASAMY, D., LUGER, K. & TREMETHICK, D. J. 2004. H2A. Z alters the nucleosome surface to promote HP1α-mediated chromatin fiber folding. *Molecular cell,* 16**,** 655-661.

FEDOROFF, N. 1979. On spacers. *Cell,* 16**,** 697-710.
FEDOROFF, N. V. 2012. Transposable elements, epigenetics, and genome evolution. *Science,* 338**,** 758-767.
FEDOTOVA, A. A., BONCHUK, A. N., MOGILA, V. A. & GEORGIEV, P. G. 2017. C2H2 Zinc Finger Proteins: The Largest but Poorly Explored Family of Higher Eukaryotic Transcription Factors. *Acta Naturae,* 9**,** 47-58.
FERREE, P. M. & BARBASH, D. A. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. *PLoS biology,* 7**,** e1000234.
FERREE, P. M. & PRASAD, S. 2012. How can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. *Genetics Research International,* 2012.
FESCHOTTE, C. 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics,* 9**,** 397-405.
FILIPOWICZ, W., JASKIEWICZ, L., KOLB, F. A. & PILLAI, R. S. 2005. Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr Opin Struct Biol,* 15**,** 331-41.
FISCHLE, W., WANG, Y., JACOBS, S. A., KIM, Y., ALLIS, C. D. & KHORASANIZADEH, S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & development,* 17**,** 1870-1881.
FITZGERALD-HAYES, M., CLARKE, L. & CARBON, J. 1982. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell,* 29**,** 235-244.
FLAVELL, A. J. 1995. Retroelements, reverse transcriptase and evolution. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology,* 110**,** 3-15.
FONT-BURGADA, J., ROSSELL, D., AUER, H. & AZORIN, F. 2008. Drosophila HP1c isoform interacts with the zinc-finger proteins WOC and Relative-of-WOC to regulate gene expression. *Genes Dev,* 22**,** 3007-23.
FRIEDMAN, J. R., FREDERICKS, W. J., JENSEN, D. E., SPEICHER, D. W., HUANG, X.-P., NEILSON, E. G. & RAUSCHER, F. J. 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes & development,* 10**,** 2067-2078.
GASPAR, I., WIPPICH, F. & EPHRUSSI, A. 2017. Enzymatic production of single-molecule FISH and RNA capture probes. *RNA,* 23**,** 1582-1591.
GEBERT, D., NEUBERT, L. K., LLOYD, C., GUI, J., LEHMANN, R. & TEIXEIRA, F. K. 2021. Large Drosophila germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. *Molecular cell,* 81**,** 3965-3978. e5.
GRENTZINGER, T., OBERLIN, S., SCHOTT, G., HANDLER, D., SVOZIL, J., BARRAGAN-BORRERO, V., HUMBERT, A., DUHARCOURT, S., BRENNECKE, J. & VOINNET, O. 2020. A universal method for the rapid isolation of all known classes of functional silencing small RNAs. *Nucleic acids research,* 48**,** e79-e79.
GREWAL, S. I. S. & ELGIN, S. C. R. 2007. Transcription and RNA interference in the formation of heterochromatin. *Nature,* 447**,** 399-406.
GUNAWARDANE, L. S., SAITO, K., NISHIDA, K. M., MIYOSHI, K., KAWAMURA, Y., NAGAMI, T., SIOMI, H. & SIOMI, M. C. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science,* 315**,** 1587-90.
GUTBROD, M. J. & MARTIENSSEN, R. A. 2020. Conserved chromosomal functions of RNA interference. *Nature Reviews Genetics,* 21**,** 311-331.
HALE, T. K., CONTRERAS, A., MORRISON, A. J. & HERRERA, R. E. 2006. Phosphorylation of the linker histone H1 by CDK regulates its binding to HP1α. *Molecular cell,* 22**,** 693-699.

HAN, B. W., WANG, W., LI, C., WENG, Z. & ZAMORE, P. D. 2015. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science,* 348**,** 817-821.

HARMS, E., CHU, T., HENRION, G. & STRICKLAND, S. 2000. The only function of Grauzone required for Drosophila oocyte meiosis is transcriptional activation of the cortex gene. *Genetics,* 155**,** 1831-1839.

HAYASHI, R., SCHNABL, J., HANDLER, D., MOHN, F., AMERES, S. L. & BRENNECKE, J. 2016. Genetic and mechanistic diversity of piRNA 3'-end formation. *Nature,* 539**,** 588-592.

HEIKKINEN, E., LAUNONEN, V., MÜLLER, E. & BACHMANN, L. 1995. The pvB370 BamHI satellite DNA family of the Drosophila virilis group and its evolutionary relation to mobile dispersed genetic pDv elements. *Journal of molecular evolution,* 41**,** 604-614.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

HELLEU, Q. & LEVINE, M. T. 2018. Recurrent amplification of the heterochromatin protein 1 (HP1) gene family across Diptera. *Molecular Biology and Evolution,* 35**,** 2375-2389.

HENIKOFF, S., AHMAD, K. & MALIK, H. S. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science,* 293**,** 1098-1102.

HERMANT, C., BOIVIN, A., TEYSSET, L., DELMARRE, V., ASIF-LAIDIN, A., VAN DEN BEEK, M., ANTONIEWSKI, C. & RONSSERAY, S. 2015. Paramutation in Drosophila requires both nuclear and cytoplasmic actors of the piRNA pathway and induces Cis-spreading of piRNA production. *Genetics,* 201**,** 1381-1396.

HOUWING, S., KAMMINGA, L. M., BEREZIKOV, E., CRONEMBOLD, D., GIRARD, A., VAN DEN ELST, H., FILIPPOV, D. V., BLASER, H., RAZ, E. & MOENS, C. B. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell,* 129**,** 69-82.

HOYT, S. J., STORER, J. M., HARTLEY, G. A., GRADY, P. G., GERSHMAN, A., DE LIMA, L. G., LIMOUSE, C., HALABIAN, R., WOJENSKI, L. & RODRIGUEZ, M. 2022. From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science,* 376**,** eabk3112.

IYENGAR, S. & FARNHAM, P. J. 2011. KAP1 protein: an enigmatic master regulator of the genome. *Journal of Biological Chemistry,* 286**,** 26267-26276.

IYER, L. M., ABHIMAN, S. & ARAVIND, L. 2011. Natural history of eukaryotic DNA methylation systems. *Progress in molecular biology and translational science,* 101**,** 25-104.

JACOBS, S. A. & KHORASANIZADEH, S. 2002. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science,* 295**,** 2080-2083.

JAGANNATHAN, M., CUMMINGS, R. & YAMASHITA, Y. M. 2018. A conserved function for pericentromeric satellite DNA. *Elife,* 7**,** e34122.

JAGANNATHAN, M. & YAMASHITA, Y. M. Function of junk: pericentromeric satellite DNA in chromosome maintenance. Cold Spring Harbor symposia on quantitative biology, 2017. Cold Spring Harbor Laboratory Press, 319-327.

JAKUBCZAK, J. L., ZENNI, M. K., WOODRUFF, R. & EICKBUSH, T. H. 1992. Turnover of R1 (type I) and R2 (type II) retrotransposable elements in the ribosomal DNA of Drosophila melanogaster. *Genetics,* 131**,** 129-142.

JAMES, T., EISSENBERG, J., CRAIG, C., DIETRICH, V., HOBSON, A. & ELGIN, S. 1989. Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila. *Eur J Cell Biol,* 50**,** 170-80.

JAMES, T. C. & ELGIN, S. 1986. Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. *Molecular and cellular biology,* 6**,** 3862-3872.

JANSSEN, A., COLMENARES, S. U. & KARPEN, G. H. 2018. Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol,* 34**,** 265-288.

JAUCH, R., BOURENKOV, G. P., CHUNG, H.-R., URLAUB, H., REIDT, U., JÄCKLE, H. & WAHL, M. C. 2003. The zinc finger-associated domain of the Drosophila transcription factor grauzone is a novel zinc-coordinating protein-protein interaction module. *Structure,* 11**,** 1393-1402.

JEHN, J., GEBERT, D., PIPILESCU, F., STERN, S., KIEFER, J. S. T., HEWEL, C. & ROSENKRANZ, D. 2018. PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Communications Biology,* 1**,** 137.

JENUWEIN, T. & ALLIS, C. D. 2001. Translating the histone code. *Science,* 293**,** 1074-1080.

JONES, K. W. & ROBERTSON, F. W. 1970. Localisation of reiterated nucleotide sequences in Drosophila and mouse by in situ hybridisation of complementary RNA. *Chromosoma,* 31**,** 331-345.

KAAIJ, L. J., MOHN, F., VAN DER WEIDE, R. H., DE WIT, E. & BÜHLER, M. 2019. The ChAHP complex counteracts chromatin looping at CTCF sites that emerged from SINE expansions in mouse. *Cell,* 178**,** 1437-1451. e14.

KABI, M. & FILION, G. J. 2021. Heterochromatin: did H3K9 methylation evolve to tame transposons? : Springer.

KAI, T., WILLIAMS, D. & SPRADLING, A. C. 2005. The expression profile of purified Drosophila germline stem cells. *Dev Biol,* 283**,** 486-502.

KAPITONOV, V. V., HOLMQUIST, G. P. & JURKA, J. 1998. L1 repeat is a basic unit of heterochromatin satellites in cetaceans. *Molecular biology and evolution,* 15**,** 611-612.

KAPITONOV, V. V. & JURKA, J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics,* 9**,** 411-412.

KARPEN, G. H. & SPRADLING, A. C. 1992. Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics,* 132**,** 737-753.

KASINATHAN, B., COLMENARES III, S. U., MCCONNELL, H., YOUNG, J. M., KARPEN, G. H. & MALIK, H. S. 2020a. Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in Drosophila. *Elife,* 9**,** e63368.

KASINATHAN, B., COLMENARES, S. U., 3RD, MCCONNELL, H., YOUNG, J. M., KARPEN, G. H. & MALIK, H. S. 2020b. Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in Drosophila. *Elife,* 9.

KATAOKA, K., NOTO, T. & MOCHIZUKI, K. 2016. Phosphorylation of an HP1-like protein is a prerequisite for heterochromatin body formation in Tetrahymena DNA elimination. *Proceedings of the National Academy of Sciences,* 113**,** 9027-9032.

KAWAOKA, S., HAYASHI, N., KATSUMA, S., KISHINO, H., KOHARA, Y., MITA, K. & SHIMADA, T. 2008. Bombyx small RNAs: genomic defense system against transposons in the silkworm, Bombyx mori. *Insect biochemistry and molecular biology,* 38**,** 1058-1065.

KEENEN, M. M., BROWN, D., BRENNAN, L. D., RENGER, R., KHOO, H., CARLSON, C. R., HUANG, B., GRILL, S. W., NARLIKAR, G. J. & REDDING, S. 2021. HP1 proteins compact DNA into mechanically and positionally stable phase separated domains. *Elife,* 10.

KELLER, C., ADAIXO, R., STUNNENBERG, R., WOOLCOCK, K. J., HILLER, S. & BÜHLER, M. 2012. HP1Swi6 mediates the recognition and destruction of heterochromatic RNA transcripts. *Molecular cell,* 47**,** 215-227.

KIDWELL, M. G. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica,* 115**,** 49-63.

KIT, S. 1961. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *Journal of molecular biology,* 3**,** 711-IN2.

KLATTENHOFF, C., BRATU, D. P., MCGINNIS-SCHULTZ, N., KOPPETSCH, B. S., COOK, H. A. & THEURKAUF, W. E. 2007. Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Developmental cell,* 12**,** 45-55.

KLATTENHOFF, C., XI, H., LI, C., LEE, S., XU, J., KHURANA, J. S., ZHANG, F., SCHULTZ, N., KOPPETSCH, B. S., NOWOSIELSKA, A., SEITZ, H., ZAMORE, P. D., WENG, Z. & THEURKAUF, W. E. 2009. The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell,* 138**,** 1137-49.

KNEUSS, E., MUNAFÒ, M., EASTWOOD, E. L., DEUMER, U.-S., PREALL, J. B., HANNON, G. J. & CZECH, B. 2019. Specialization of the Drosophila nuclear export family protein Nxf3 for piRNA precursor export. *Genes & development,* 33**,** 1208-1220.

KOFLER, R. 2019. Dynamics of transposable element invasions with piRNA clusters. *Molecular Biology and Evolution,* 36**,** 1457-1472.

KOFLER, R. 2020. piRNA clusters need a minimum size to control transposable element invasions. *Genome biology and evolution,* 12**,** 736-749.

KOMAROV, P. A., SOKOLOVA, O., AKULENKO, N., BRASSET, E., JENSEN, S. & KALMYKOVA, A. 2020. Epigenetic Requirements for triggering heterochromatinization and piwi-interacting RNA production from transgenes in the drosophila germline. *Cells,* 9**,** 922.

LACHNER, M., O'CARROLL, D., REA, S., MECHTLER, K. & JENUWEIN, T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature,* 410**,** 116-20.

LARKIN, A., MARYGOLD, S. J., ANTONAZZO, G., ATTRILL, H., DOS SANTOS, G., GARAPATI, P. V., GOODMAN, J. L., GRAMATES, L. S., MILLBURN, G. & STRELETS, V. B. 2021. FlyBase: updates to the Drosophila melanogaster knowledge base. *Nucleic acids research,* 49**,** D899-D907.

LARRACUENTE, A. M. 2014. The organization and evolution of the Responder satellite in species of the Drosophila melanogaster group: dynamic evolution of a target of meiotic drive. *BMC evolutionary biology,* 14**,** 1-12.

LARRACUENTE, A. M. & PRESGRAVES, D. C. 2012. The selfish Segregation Distorter gene complex of Drosophila melanogaster. *Genetics,* 192**,** 33-53.

LARSON, A. G., ELNATAN, D., KEENEN, M. M., TRNKA, M. J., JOHNSTON, J. B., BURLINGAME, A. L., AGARD, D. A., REDDING, S. & NARLIKAR, G. J. 2017. Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature,* 547**,** 236-240.

LE THOMAS, A., ROGERS, A. K., WEBSTER, A., MARINOV, G. K., LIAO, S. E., PERKINS, E. M., HUR, J. K., ARAVIN, A. A. & TOTH, K. F. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev,* 27**,** 390-9.

LE THOMAS, A., STUWE, E., LI, S., DU, J., MARINOV, G., ROZHKOV, N., CHEN, Y.-C. A., LUO, Y., SACHIDANANDAM, R. & TOTH, K. F. 2014. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes & development,* 28**,** 1667-1680.

LEE, D. H., RYU, H. W., KIM, G. W. & KWON, S. H. 2019. Comparison of three heterochromatin protein 1 homologs in Drosophila. *Journal of cell science,* 132**,** jcs222729.

LEE, T. I., JOHNSTONE, S. E. & YOUNG, R. A. 2006. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc,* 1**,** 729-48.

LEVINE, M. T., MCCOY, C., VERMAAK, D., LEE, Y. C., HIATT, M. A., MATSEN, F. A. & MALIK, H. S. 2012. Phylogenomic analysis reveals dynamic evolutionary history of the Drosophila heterochromatin protein 1 (HP1) gene family. *PLoS Genet,* 8**,** e1002729.

LEVINE, M. T., VANDER WENDE, H. M. & MALIK, H. S. 2015. Mitotic fidelity requires transgenerational action of a testis-restricted HP1. *Elife,* 4**,** e07378.

LOHE, A. R. & BRUTLAG, D. L. 1986. Multiplicity of satellite DNA sequences in Drosophila melanogaster. *Proceedings of the National Academy of Sciences,* 83**,** 696-700.

LOHE, A. R., HILLIKER, A. & ROBERTS, P. 1993. Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. *Genetics,* 134**,** 1149-1174.

LOMBERK, G., WALLRATH, L. & URRUTIA, R. 2006. The heterochromatin protein 1 family. *Genome biology,* 7**,** 1-8.

LONG, E. O. & DAWID, I. B. 1980. Repeated genes in eukaryotes. *Annual review of biochemistry,* 49**,** 727-764.

LOSADA, A. & VILLASANTE, A. 1996. Autosomal location of a new subtype of 1.688 satellite DNA ofDrosophila melanogaster. *Chromosome Research,* 4**,** 372-383.

LU, X., WONTAKAL, S. N., EMELYANOV, A. V., MORCILLO, P., KONEV, A. Y., FYODOROV, D. V. & SKOULTCHI, A. I. 2009. Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. *Genes & development,* 23**,** 452-465.

LUO, Y., HE, P., KANRAR, N., TOTH, K. F. & ARAVIN, A. 2022. Maternally inherited siRNAs initiate piRNA cluster formation. *bioRxiv*.

LYTTLE, T. W. 1991. Segregation distorters. *Annual review of genetics,* 25**,** 511-581.

MAKSIMENKO, O., BARTKUHN, M., STAKHOV, V., HEROLD, M., ZOLOTAREV, N., JOX, T., BUXA, M. K., KIRSCH, R., BONCHUK, A. & FEDOTOVA, A. 2015. Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. *Genome research,* 25**,** 89-99.

MAKSIMENKO, O., KYRCHANOVA, O., KLIMENKO, N., ZOLOTAREV, N., ELIZAROVA, A., BONCHUK, A. & GEORGIEV, P. 2020. Small Drosophila zinc finger C2H2 protein with an N-terminal zinc finger-associated domain demonstrates the architecture functions. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* 1863**,** 194446.

MARIE, P. P., RONSSERAY, S. & BOIVIN, A. 2017. From embryo to adult: piRNA-mediated silencing throughout germline development in Drosophila. *G3: Genes, Genomes, Genetics,* 7**,** 505-516.

MARKSTEIN, M., PITSOULI, C., VILLALTA, C., CELNIKER, S. E. & PERRIMON, N. 2008. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet,* 40**,** 476-83.

MARTIENSSEN, R. & MOAZED, D. 2015. RNAi and heterochromatin assembly. *Cold Spring Harb Perspect Biol,* 7**,** a019323.

MASON, J. M. & BIESSMANN, H. 1995. The unusual telomeres of Drosophila. *Trends Genet,* 11**,** 58-62.

MATTHEWS, K. A., MILLER, D. F. & KAUFMAN, T. C. 1989. Developmental distribution of RNA and protein products of the Drosophila α-tubulin gene family. *Developmental biology,* 132**,** 45-61.

MCCULLERS, T. J. & STEINIGER, M. 2017. Transposable elements in Drosophila. *Mobile genetic elements,* 7**,** 1-18.

MELTERS, D. P., BRADNAM, K. R., YOUNG, H. A., TELIS, N., MAY, M. R., RUBY, J. G., SEBRA, R., PELUSO, P., EID, J. & RANK, D. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome biology,* 14**,** 1-20.

MÉREL, V., BOULESTEIX, M., FABLET, M. & VIEIRA, C. 2020. Transposable elements in Drosophila. *Mobile DNA,* 11**,** 1-20.

MITRA, R., FAIN-THORNTON, J. & CRAIG, N. L. 2008. piggyBac can bypass DNA synthesis during cut and paste transposition. *The EMBO journal,* 27**,** 1097-1109.

MOHN, F., HANDLER, D. & BRENNECKE, J. 2015. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science,* 348**,** 812-817.

MOHN, F., SIENSKI, G., HANDLER, D. & BRENNECKE, J. 2014. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. *Cell,* 157**,** 1364-79.

MOSCHETTI, R., PALAZZO, A., LORUSSO, P., VIGGIANO, L. & MASSIMILIANO MARSANO, R. 2020. "What you need, baby, I got it": Transposable elements as suppliers of cis-operating sequences in Drosophila. *Biology,* 9**,** 25.

MOTAMEDI, M. R., VERDEL, A., COLMENARES, S. U., GERBER, S. A., GYGI, S. P. & MOAZED, D. 2004. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell,* 119**,** 789-802.

NELSON, J. O., SLICKO, A. & YAMASHITA, Y. M. 2021. The retrotransposon R2 maintains Drosophila ribosomal DNA repeats. *bioRxiv*.

NIELSEN, P. R., NIETLISPACH, D., MOTT, H. R., CALLAGHAN, J., BANNISTER, A., KOUZARIDES, T., MURZIN, A. G., MURZINA, N. V. & LAUE, E. D. 2002. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature,* 416**,** 103-107.

NIKI, Y., YAMAGUCHI, T. & MAHOWALD, A. P. 2006. Establishment of stable cell lines of Drosophila germ-line stem cells. *Proc Natl Acad Sci U S A,* 103**,** 16325-30.

NURK, S., KOREN, S., RHIE, A., RAUTIAINEN, M., BZIKADZE, A. V., MIKHEENKO, A., VOLLGER, M. R., ALTEMOSE, N., URALSKY, L. & GERSHMAN, A. 2022. The complete sequence of a human genome. *Science,* 376**,** 44-53.

O'SHEA, E. K., KLEMM, J. D., KIM, P. S. & ALBER, T. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science,* 254**,** 539-544.

OHNO, S. So much'junk'DNA in our genome. Evolution of Genetic Systems, Brookhaven Symp. Biol., 1972. 366-370.

OLOVNIKOV, I., RYAZANSKY, S., SHPIZ, S., LAVROV, S., ABRAMOV, Y., VAURY, C., JENSEN, S. & KALMYKOVA, A. 2013. De novo piRNA cluster formation in the Drosophila germ line triggered by transgenes containing a transcribed transposon fragment. *Nucleic acids research,* 41**,** 5757-5768.

ORGEL, L. E. & CRICK, F. H. 1980. Selfish DNA: the ultimate parasite. *Nature,* 284**,** 604-607.

OSTAPCUK, V., MOHN, F., CARL, S. H., BASTERS, A., HESS, D., IESMANTAVICIUS, V., LAMPERSBERGER, L., FLEMR, M., PANDEY, A., THOMA, N. H., BETSCHINGER, J. & BUHLER, M. 2018. Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. *Nature,* 557**,** 739-743.

OZATA, D. M., GAINETDINOV, I., ZOCH, A., O'CARROLL, D. & ZAMORE, P. D. 2019. PIWI-interacting RNAs: small RNAs with big functions. *Nature Reviews Genetics,* 20**,** 89-108.

PADEKEN, J., METHOT, S. P. & GASSER, S. M. 2022. Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nature Reviews Molecular Cell Biology***,** 1-18.

PALMER, D. K., O'DAY, K., WENER, M. H., ANDREWS, B. S. & MARGOLIS, R. L. 1987. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *The Journal of cell biology,* 104**,** 805-815.

PANE, A., JIANG, P., ZHAO, D. Y., SINGH, M. & SCHÜPBACH, T. 2011. The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. *The EMBO journal,* 30**,** 4601-4615.

PARDUE, M. L. & GALL, J. G. 1970. Chromosomal localization of mouse satellite DNA. *Science,* 168**,** 1356-1358.

PARHAD, S. S. & THEURKAUF, W. E. 2019. Rapid evolution and conserved function of the piRNA pathway. *Royal Society Open Biology,* 9**,** 180181.

PARHAD, S. S., TU, S., WENG, Z. & THEURKAUF, W. E. 2017. Adaptive evolution leads to cross-species incompatibility in the piRNA transposon silencing machinery. *Developmental cell,* 43**,** 60-70. e5.

PARO, R. & HOGNESS, D. 1991. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc Natl Acad Sci U S A,* 88**,** 263-7.

PELISSON, A., SONG, S. U., PRUD'HOMME, N., SMITH, P. A., BUCHETON, A. & CORCES, V. G. 1994. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *Embo J,* 13**,** 4401-11.

PÉREZ-GONZÁLEZ, C. E. & EICKBUSH, T. H. 2002. Rates of R1 and R2 retrotransposition and elimination from the rDNA locus of Drosophila melanogaster. *Genetics,* 162**,** 799-811.

PIKAARD, C. S. & MITTELSTEN SCHEID, O. 2014. Epigenetic regulation in plants. *Cold Spring Harb Perspect Biol,* 6**,** a019315.

PIMPINELLI, S., BERLOCO, M., FANTI, L., DIMITRI, P., BONACCORSI, S., MARCHETTI, E., CAIZZI, R., CAGGESE, C. & GATTI, M. 1995. Transposable elements are stable structural components of Drosophila melanogaster heterochromatin. *Proceedings of the National Academy of Sciences,* 92**,** 3804-3808.

PIMPINELLI, S. & DIMITRI, P. 1989. Cytogenetic analysis of segregation distortion in Drosophila melanogaster: the cytological organization of the Responder (Rsp) locus. *Genetics,* 121**,** 765-772.

PIPPADPALLY, S. & VENKATESH, T. 2020. Deciphering piRNA biogenesis through cytoplasmic granules, mitochondria and exosomes. *Archives of Biochemistry and Biophysics,* 695**,** 108597.

PRYDE, F. E., GORHAM, H. C. & LOUIS, E. J. 1997. Chromosome ends: all the same under their caps. *Current opinion in genetics & development,* 7**,** 822-828.

QI, H., WATANABE, T., KU, H. Y., LIU, N., ZHONG, M. & LIN, H. 2010. The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. *J Biol Chem,* 286**,** 3789-97.

QUÉNERCH'DU, E., ANAND, A. & KAI, T. 2016. The piRNA pathway is developmentally regulated during spermatogenesis in Drosophila. *Rna,* 22**,** 1044-1054.

RAE, P. M., BARNETT, T. & MURTIF, V. L. 1981. Nontranscribed spacers in Drosophila ribosomal DNA. *Chromosoma,* 82**,** 637-655.

RAMIREZ, F., RYAN, D. P., GRUNING, B., BHARDWAJ, V., KILPERT, F., RICHTER, A. S., HEYNE, S., DUNDAR, F. & MANKE, T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res,* 44**,** W160-5.

ROBERT, X. & GOUET, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research,* 42**,** W320-W324.

ROUND, E. K., FLOWERS, S. K. & RICHARDS, E. J. 1997. Arabidopsis thaliana centromere regions: genetic map positions and repetitive DNA structure. *Genome research,* 7**,** 1045-1053.

ROZEWICKI, J., LI, S., AMADA, K. M., STANDLEY, D. M. & KATOH, K. 2019. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic acids research,* 47**,** W5-W10.

ROZHKOV, N. V., HAMMELL, M. & HANNON, G. J. 2013. Multiple roles for Piwi in silencing Drosophila transposons. *Genes Dev,* 27**,** 400-12.

RUDD, M. K. & WILLARD, H. F. 2004. Analysis of the centromeric regions of the human genome assembly. *Trends in genetics,* 20**,** 529-533.

RUST, K., BYRNES, L. E., YU, K. S., PARK, J. S., SNEDDON, J. B., TWARD, A. D. & NYSTUL, T. G. 2020. A single-cell atlas and lineage analysis of the adult Drosophila ovary. *Nature communications,* 11**,** 1-17.

RYAN, D. P. & TREMETHICK, D. J. 2018. The interplay between H2A. Z and H3K9 methylation in regulating HP1α binding to linker histone-containing chromatin. *Nucleic acids research,* 46**,** 9353-9366.

RYDER, E. & RUSSELL, S. 2003. Transposable elements as tools for genomics and genetics in Drosophila. *Briefings in Functional Genomics,* 2**,** 57-71.

SABIROV, M., KYRCHANOVA, O., POKHOLKOVA, G. V., BONCHUK, A., KLIMENKO, N., BELOVA, E., ZHIMULEV, I. F., MAKSIMENKO, O. & GEORGIEV, P. 2021a. Mechanism and functional role of the interaction between CP190 and the architectural protein Pita in Drosophila melanogaster. *Epigenetics Chromatin,* 14**,** 16.

SABIROV, M., POPOVICH, A., BOYKO, K., NIKOLAEVA, A., KYRCHANOVA, O., MAKSIMENKO, O., POPOV, V., GEORGIEV, P. & BONCHUK, A. 2021b. Mechanisms of CP190 Interaction with Architectural Proteins in Drosophila Melanogaster. *Int J Mol Sci,* 22.

SAITO, K., INAGAKI, S., MITUYAMA, T., KAWAMURA, Y., ONO, Y., SAKOTA, E., KOTANI, H., ASAI, K., SIOMI, H. & SIOMI, M. C. 2009. A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. *Nature,* 461**,** 1296-9.

SAITO, K., ISHIZU, H., KOMAI, M., KOTANI, H., KAWAMURA, Y., NISHIDA, K. M., SIOMI, H. & SIOMI, M. C. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes Dev,* 24**,** 2493-8.

SAITO, K., SAKAGUCHI, Y., SUZUKI, T., SIOMI, H. & SIOMI, M. C. 2007. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev,* 21**,** 1603-8.

SANDOVAL-VILLEGAS, N., NURIEVA, W., AMBERGER, M. & IVICS, Z. 2021. Contemporary transposon tools: a review and guide through mechanisms and applications of sleeping beauty, piggyBac and Tol2 for genome engineering. *International journal of molecular sciences,* 22**,** 5084.

SANULLI, S., TRNKA, M., DHARMARAJAN, V., TIBBLE, R., PASCAL, B., BURLINGAME, A., GRIFFIN, P., GROSS, J. & NARLIKAR, G. 2019. HP1 reshapes nucleosome core to promote heterochromatin phase separation. *Nature,* 575**,** 390-394.

SATO, K. & SIOMI, M. C. 2020. The piRNA pathway in Drosophila ovarian germ and somatic cells. *Proceedings of the Japan Academy, Series B,* 96**,** 32-42.

SAUNDERS, W., CHUE, C., GOEBL, M., CRAIG, C., CLARK, R., POWERS, J., EISSENBERG, J., ELGIN, S., ROTHFIELD, N. & EARNSHAW, W. 1993. Molecular cloning of a human homologue of Drosophila heterochromatin protein HP1 using anti-centromere autoantibodies with anti-chromo specificity. *J Cell Sci,* 104 (Pt 2)**,** 573-82.

SAWAMURA, K., YAMAMOTO, M. & WATANABE, T. K. 1993. Hybrid lethal systems in the Drosophila melanogaster species complex. II. The Zygotic hybrid rescue (Zhr) gene of D. melanogaster. *Genetics,* 133**,** 307-313.

SAWAMURA, K. & YAMAMOTO, M.-T. 1997. Characterization of a reproductive isolation gene, zygotic hybrid rescue, of Drosophila melanogaster by using minichromosomes. *Heredity,* 79**,** 97-103.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods,* 9**,** 676-82.

SCHNABL, J., WANG, J., HOHMANN, U., GEHRE, M., BATKI, J., ANDREEV, V. I., PURKHAUSER, K., FASCHING, N., DUCHEK, P. & NOVATCHKOVA, M. 2021. Molecular principles of Piwi-mediated cotranscriptional silencing through the dimeric SFiNX complex. *Genes & development,* 35**,** 392-409.

SCHNABLE, P. S., WARE, D., FULTON, R. S., STEIN, J. C., WEI, F., PASTERNAK, S., LIANG, C., ZHANG, J., FULTON, L. & GRAVES, T. A. 2009. The B73 maize genome: complexity, diversity, and dynamics. *science,* 326**,** 1112-1115.

SCHOELZ, J. M. & RIDDLE, N. C. 2022. Functions of HP1 proteins in transcriptional regulation. *Epigenetics & Chromatin,* 15**,** 1-15.

SENTMANAT, M. F. & ELGIN, S. C. 2012. Ectopic assembly of heterochromatin in Drosophila melanogaster triggered by transposable elements. *Proceedings of the National Academy of Sciences,* 109**,** 14104-14109.

SHPIZ, S. & KALMYKOVA, A. 2014. Analyses of piRNA-mediated transcriptional transposon silencing in Drosophila: nuclear run-on assay on ovaries. *Methods Mol Biol,* 1093**,** 149-59.

SHPIZ, S., RYAZANSKY, S., OLOVNIKOV, I., ABRAMOV, Y. & KALMYKOVA, A. 2014. Euchromatic transposon insertions trigger production of novel Pi-and endo-siRNAs at the target sites in the drosophila germline. *PLoS genetics,* 10**,** e1004138.

SIENSKI, G., BATKI, J., SENTI, K. A., DONERTAS, D., TIRIAN, L., MEIXNER, K. & BRENNECKE, J. 2015. Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. *Genes Dev,* 29**,** 2258-71.

SIENSKI, G., DONERTAS, D. & BRENNECKE, J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell,* 151**,** 964-80.

SIMEONE, A., VOLPE, A. L. & BONCINELLI, E. 1985. Nucleotide sequence of a complete ribosomal spacer of D. melanogaster. *Nucleic acids research,* 13**,** 1089-1101.

SIMKIN, A., WONG, A., POH, Y. P., THEURKAUF, W. E. & JENSEN, J. D. 2013. Recurrent and recent selective sweeps in the piRNA pathway. *Evolution: International Journal of Organic Evolution,* 67**,** 1081-1090.

SINGH, P., MILLER, J., PEARCE, J., KOTHARY, R., BURTON, R., PARO, R., JAMES, T. & GAUNT, S. 1991. A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. *Nucleic Acids Res,* 19**,** 789-94.

SINGH, P. B. & GEORGATOS, S. D. 2002. HP1: facts, open questions, and speculation. *Journal of structural biology,* 140**,** 10-16.

SMOTHERS, J. F. & HENIKOFF, S. 2000. The HP1 chromo shadow domain binds a consensus peptide pentamer. *Current Biology,* 10**,** 27-30.

SMOTHERS, J. F. & HENIKOFF, S. 2001. The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins. *Molecular and cellular biology,* 21**,** 2555-2569.

SRINIVASAN, A. & MISHRA, R. K. 2020. Genomic organization of Polycomb Response Elements and its functional implication in Drosophila and other insects. *Journal of biosciences,* 45**,** 1-16.

STROM, A. R., EMELYANOV, A. V., MIR, M., FYODOROV, D. V., DARZACQ, X. & KARPEN, G. H. 2017. Phase separation drives heterochromatin domain formation. *Nature,* 547**,** 241-245.

SUN, X., LE, H. D., WAHLSTROM, J. M. & KARPEN, G. H. 2003. Sequence analysis of a functional Drosophila centromere. *Genome research,* 13**,** 182-194.

SZAKMARY, A., REEDY, M., QI, H. & LIN, H. 2009. The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. *J Cell Biol,* 185**,** 613-27.

THAKUR, J., PACKIARAJ, J. & HENIKOFF, S. 2021. Sequence, chromatin and evolution of satellite DNA. *International Journal of Molecular Sciences,* 22**,** 4309.

THIERY, J.-P., MACAYA, G. & BERNARDI, G. 1976. An analysis of eukaryotic genomes by density gradient centrifugation. *Journal of molecular biology,* 108**,** 219-235.

THIRU, A., NIETLISPACH, D., MOTT, H. R., OKUWAKI, M., LYON, D., NIELSEN, P. R., HIRSHBERG, M., VERREAULT, A., MURZINA, N. V. & LAUE, E. D. 2004. Structural basis of HP1/PXVXL motif peptide interactions and HP1 localisation to heterochromatin. *The EMBO journal,* 23**,** 489-499.

TODESCHINI, A.-L., TEYSSET, L., DELMARRE, V. & RONSSERAY, S. 2010. The epigenetic trans-silencing effect in Drosophila involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. *PloS one,* 5**,** e11032.

TROJER, P. & REINBERG, D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? *Molecular cell,* 28**,** 1-13.

USAKIN, L., ABAD, J., VAGIN, V. V., DE PABLOS, B., VILLASANTE, A. & GVOZDEV, V. A. 2007. Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in Drosophila melanogaster ovaries. *Genetics,* 176**,** 1343-1349.

VAGIN, V. V., SIGOVA, A., LI, C., SEITZ, H., GVOZDEV, V. & ZAMORE, P. D. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science,* 313**,** 320-4.

VAKOC, C. R., MANDAT, S. A., OLENCHOCK, B. A. & BLOBEL, G. A. 2005. Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. *Molecular cell,* 19**,** 381-391.

VAN VALEN, L. 1973. A new evolutionary law.

VERDEL, A., JIA, S., GERBER, S., SUGIYAMA, T., GYGI, S., GREWAL, S. I. & MOAZED, D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science,* 303**,** 672-6.

VERDEL, A. & MOAZED, D. 2005. RNAi-directed assembly of heterochromatin in fission yeast. *FEBS Lett,* 579**,** 5872-8.

VERMAAK, D., HENIKOFF, S. & MALIK, H. S. 2005. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. *PLoS genetics,* 1**,** e9.

VERMAAK, D. & MALIK, H. S. 2009. Multiple roles for heterochromatin protein 1 genes in Drosophila. *Annu Rev Genet,* 43**,** 467-92.

VILLASEÑOR, R., PFAENDLER, R., AMBROSI, C., BUTZ, S., GIULIANI, S., BRYAN, E., SHEAHAN, T. W., GABLE, A. L., SCHMOLKA, N. & MANZO, M. 2020. ChromID identifies the protein interactome at chromatin marks. *Nature biotechnology,* 38**,** 728-736.

VOLPE, A. M., HOROWITZ, H., GRAFER, C. M., JACKSON, S. M. & BERG, C. A. 2001. Drosophila rhino encodes a female-specific chromo-domain protein that affects chromosome structure and egg polarity. *Genetics,* 159**,** 1117-1134.

VOLPE, T. A., KIDNER, C., HALL, I. M., TENG, G., GREWAL, S. I. & MARTIENSSEN, R. A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science,* 297**,** 1833-7.

WALTER, M., JANG, C., KASRAVI, B., DONATH, J., MECHLER, B., MASON, J. & BIESSMANN, H. 1995. DNA organization and polymorphism of a wild-type Drosophila telomere region. *Chromosoma,* 104**,** 229-241.

WANG, G., MA, A., CHOW, C.-M., HORSLEY, D., BROWN, N. R., COWELL, I. G. & SINGH, P. B. 2000. Conservation of heterochromatin protein 1 function. *Molecular and cellular biology,* 20**,** 6970-6983.

WANG, L., BARBASH, D. A. & KELLEHER, E. S. 2020. Adaptive evolution among cytoplasmic piRNA proteins leads to decreased genomic auto-immunity. *PLoS genetics,* 16**,** e1008861.

WANG, S. H. & ELGIN, S. C. 2011. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc Natl Acad Sci U S A,* 108**,** 21164-9.

WANG, X., PAULO, J. A., LI, X., ZHOU, H., YU, J., GYGI, S. P. & MOAZED, D. 2021. A composite DNA element that functions as a maintainer required for epigenetic inheritance of heterochromatin. *Molecular Cell,* 81**,** 3979-3991. e4.

WEI, X., EICKBUSH, D. G., SPEECE, I. & LARRACUENTE, A. M. 2021. Heterochromatin-dependent transcription of satellite DNAs in the Drosophila melanogaster female germline. *Elife,* 10**,** e62375.

WELLAUER, P. K. & DAWID, I. B. 1977. The structural organization of ribosomal DNA in Drosophila melanogaster. *Cell,* 10**,** 193-212.

WELLS, J. N. & FESCHOTTE, C. 2020. A field guide to eukaryotic transposable elements. *Annual review of genetics,* 54**,** 539.

WESSLER, S. R. 2006. Transposable elements and the evolution of eukaryotic genomes. *Proceedings of the National Academy of Sciences,* 103**,** 17600-17601.

WESTERMANN, S., CHEESEMAN, I. M., ANDERSON, S., YATES III, J. R., DRUBIN, D. G. & BARNES, G. 2003. Architecture of the budding yeast kinetochore reveals a conserved molecular core. *The Journal of cell biology,* 163**,** 215-222.

WOLFE, S. A., NEKLUDOVA, L. & PABO, C. O. 2000. DNA recognition by (Cys2His2) zinc finger proteins. *Annual review of biophysics and biomolecular structure,* 29**,** 183.

WU, C.-I., LYTTLE, T. W., WU, M.-L. & LIN, G.-F. 1988. Association between a satellite DNA sequence and the Responder of Segregation Distorter in D. melanogaster. *Cell,* 54**,** 179-189.

YAMANAKA, S., SIOMI, M. C. & SIOMI, H. 2014. piRNA clusters and open chromatin structure. *Mobile DNA,* 5**,** 1-12.

YE, J. & EICKBUSH, T. H. 2006. Chromatin structure and transcription of the R1-and R2-inserted rRNA genes of Drosophila melanogaster. *Molecular and cellular biology,* 26**,** 8781-8790.

YE, Q., CALLEBAUT, I., PEZHMAN, A., COURVALIN, J.-C. & WORMAN, H. J. 1997. Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR. *Journal of Biological Chemistry,* 272**,** 14983-14989.

YU, B., CASSANI, M., WANG, M., LIU, M., MA, J., LI, G., ZHANG, Z. & HUANG, Y. 2015a. Structural insights into Rhino-mediated germline piRNA cluster formation. *Cell research,* 25**,** 525-528.

YU, B., LIN, Y. A., PARHAD, S. S., JIN, Z., MA, J., THEURKAUF, W. E., ZHANG, Z. Z. & HUANG, Y. 2018a. Structural insights into Rhino-Deadlock complex for germline piRNA cluster specification. *EMBO Rep,* 19.

YU, B., LIN, Y. A., PARHAD, S. S., JIN, Z., MA, J., THEURKAUF, W. E., ZHANG, Z. Z. & HUANG, Y. 2018b. Structural insights into Rhino-Deadlock complex for germline piRNA cluster specification. *EMBO reports,* 19**,** e45418.

YU, Y., GU, J., JIN, Y., LUO, Y., PREALL, J. B., MA, J., CZECH, B. & HANNON, G. J. 2015b. Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science,* 350**,** 339-42.

YUAN, Y.-W. & WESSLER, S. R. 2011. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proceedings of the National Academy of Sciences,* 108**,** 7884-7889.

ZHANG, F., WANG, J., XU, J., ZHANG, Z., KOPPETSCH, B. S., SCHULTZ, N., VREVEN, T., MEIGNIN, C., DAVIS, I. & ZAMORE, P. D. 2012a. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell,* 151**,** 871-884.

ZHANG, F., WANG, J., XU, J., ZHANG, Z., KOPPETSCH, B. S., SCHULTZ, N., VREVEN, T., MEIGNIN, C., DAVIS, I., ZAMORE, P. D., WENG, Z. & THEURKAUF, W. E. 2012b. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell,* 151**,** 871-84.

ZHANG, G., TU, S., YU, T., ZHANG, X.-O., PARHAD, S. S., WENG, Z. & THEURKAUF, W. E. 2018. Co-dependent assembly of Drosophila piRNA precursor complexes and piRNA cluster heterochromatin. *Cell reports,* 24**,** 3413-3422. e4.

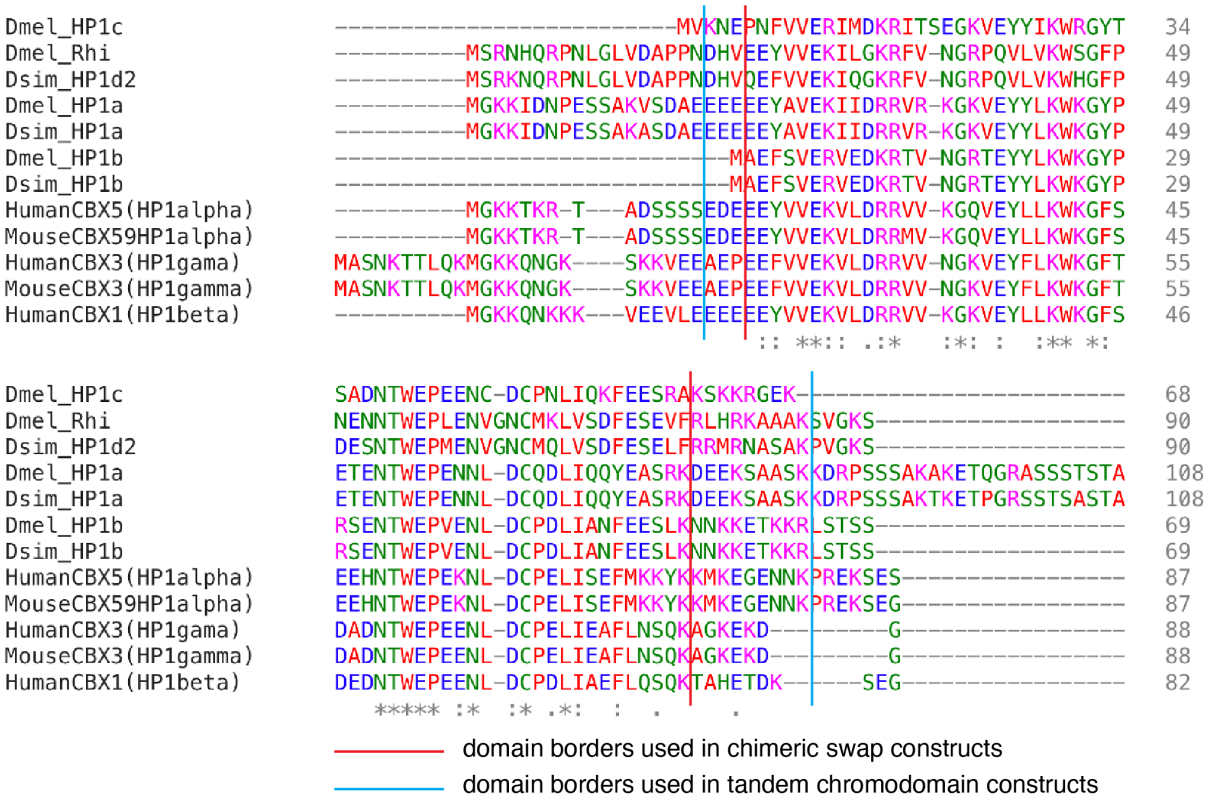
ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol,* 9**,** R137.

ZHANG, Z., THEURKAUF, W. E., WENG, Z. & ZAMORE, P. D. 2012c. Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence,* 3**,** 9.

ZHANG, Z., WANG, J., SCHULTZ, N., ZHANG, F., PARHAD, S. S., TU, S., VREVEN, T., ZAMORE, P. D., WENG, Z. & THEURKAUF, W. E. 2014. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell,* 157**,** 1353-1363.

ZHANG, Z., XU, J., KOPPETSCH, B. S., WANG, J., TIPPING, C., MA, S., WENG, Z., THEURKAUF, W. E. & ZAMORE, P. D. 2011. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol Cell,* 44**,** 572-84.

Appendix



Multiple sequence alignment of HP1 protein variant chromodomains generated through Clustal Omega (1.2.4).