UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Simple repetitive RNAs are required for fertility and viability in Drosophila melanogaster

Permalink
https://escholarship.org/uc/item/6x9346xk

Author
Mills, Wilbur Kyle

Publication Date
2018

Peer reviewed|Thesis/dissertation

Simple repetitive RNAs are required for fertility and viability in *Drosophila melanogaster*

By

Wilbur Kyle Mills

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Molecular and Cell Biology
in the
Graduate Division
of the University of California, Berkeley

Committee, in charge:
Professor Gary H. Karpen, Chair
Professor Donald C. Rio
Assistant Professor Nicholas Ingolia
Professor Robert Fischer

Summer 2018

Abstract

Simple repetitive RNAs are required for fertility and viability in *Drosophila melanogaster*

By
Wilbur Kyle Mills
Doctor of Philosophy in Molecular and Cellular Biology
University of California, Berkeley
Professor Gary H. Karpen, Chair

For proper cell function, DNA must be finely packaged into 'chromatin', which is the complex packaging of DNA with proteins and RNA. One state of gene-poor chromatin, termed heterochromatin, is a dense state of chromatin enriched for repeated RNA sequences and defined by repressive histone marks, and is critical for telomere maintenance, pericentromeric cohesion, meiotic segregation, and double strand break repair. Whether or not functions of heterochromatin stem from the underlying DNA, proteins or RNA, or a combination of all three, has largely been unexplored. Recent advances in sequencing technology have identified thousands of RNAs that do not code for proteins, termed non-coding RNAs (ncRNAs), and intriguingly some of them function to modify chromatin organization. Due to difficulties in sequencing RNAs that originate from heterochromatin, most ncRNAs described thus far were found to come from non-heterochromatic regions. Furthermore, due to limitations in studying functionality of ncRNAs, most studies were performed in-vitro and not in whole animals. Therein there lies a great need to identify which heterochromatic RNAs are transcribed and determine if any of these exhibit functions in whole animals. Heterochromatin is composed primarily of repeated DNA sequences, such as transposable elements (TEs), ribosomal DNA and satellites, which are blocks of tandem repeats varying in size from a few repeats to several megabases. Here I identify repetitive satellite ncRNAs that are transcribed from heterochromatin in *Drosophila melanogaster*. The most abundant forms of these satellite ncRNAs contain AAGAG(n) tandem repeats, are maternally loaded and associate with the earliest forms of heterochromatin, which suggests that they have roles in heterochromatin formation. These satellite ncRNAs are also found in later stages of *Drosophila* development in virtually every cell of embryos and larvae, and in addition are enriched in neural tissue and germline cells. In male testes they are enriched in the spermatocytes cells, which prompted us to determine if AAGAG(n) RNA containing satellites contribute to male fertility. We demonstrate that decreasing the levels of these satellites in spermatocytes causes 100% sterility by preventing the production of mature sperm, likely by affecting chromatin organization. These ncRNAs also bind an abundance of proteins involved in or present in spermatogenesis, suggesting that AAGAG(n) containing satellite RNAs affect organization and functions of these proteins. We also find that AAGAG(n) RNA containing satellites bind heterochromatic proteins, and furthermore provide preliminary evidence that these satellites are necessary for viability.

This study thus provides not only some of the first evidence that satellite RNAs are transcribed in *Drosophila*, but that they can also exhibit critical functions. Contrary to the notion that heterochromatin serves mainly to silence RNA, this study suggests that RNAs derived from heterochromatin are critical interactors in the cell.

We also identified long non-coding RNAs (lncRNAs) that associate with heterochromatin proteins and provide detailed subcellular and sub-nuclear distribution of these lncRNAs throughout development. Some of these lncRNAs are expressed only from certain tissues, suggesting they have roles in development. This is an important database that will provide researchers seeking to study these individual lncRNAs a foundation from which to begin to elucidate functions.

Table of Contents

Acknowledgements

I would like to begin by acknowledging my mentor and friend, Gary Karpen, for cultivating my interest in basic science and for providing a fertile, supportive and positive environment. This project would not have been possible without him. He is the best example of a mentor I can think of, from his high level of creativity, uncapped curiosity in all things science and otherwise, and drive in questioning basic assumptions. Most importantly, though, is his insistence on focusing on mentorship and creating the next generation of scientists without regard to his own personal accolades. This project, as are most in his lab, are the result of uncommon freedoms he grants to those to develop their own original ideas, which are critical for making significant advances in biology but are all too rare today.

I would also like to acknowledge mentorship from co-workers and friends in the lab that have guided me through the complexities of navigating techniques and cultivating ideas. Weiguo Zhang, for teaching me old-school cytological techniques, guiding me in developmental studies, and always being around late at night to talk to. Serafin Colmenares, for his patience, extensive help with fly husbandry and cytological techniques, as well as guidance with assay design. Grace Lee, for her enthusiasm for my project and critical help in bioinformatic analysis, parts of my project which would not be possible without her. Sasha Langley, for her insight into bioinformatic analysis. David Acevedo, for putting up with my sense of humor, and helping me manage so many things at once. Aniek Jansen, for her pathological optimism, great ideas and willingness to go on wild adventures outside of lab. I would like to give special acknowledgements to my classmate and friend, Amy Strom, for navigating the process of graduate school with me, helping me with ideas and techniques, and willingness to grab a beer when needed. I would also like to give special acknowledgement to Joel Swenson, a former graduate student, for acting also as a mentor during my first years. To our newest members Shelby Wilson and Collin Hickman, I hope you have as great an experience as I did in the lab.

To my wife, Katy Brown, thank you for putting up with my long hours, supporting me in every way possible, and making all the sacrifices you have made for me.

Finally, I would like to acknowledge members of my thesis committee, Donald Rio, Nicholas Ingolia, and Robert Fischer, for supporting all ideas I presented with insight and enthusiasm, and for keeping me on the right track.

Chapter 1: Introduction: Nuclear organization, heterochromatin and non-coding RNA

In eukaryotic cells, genomic DNA is packaged via a complex arrangement of modified histones, proteins and RNA, termed chromatin, which regulates gene expression and other genome functions. Two major types of chromatin organization define the extremes; euchromatin, which is gene rich and lightly 'packaged' to allow for robust transcription, and heterochromatin, which is densely 'packed' and less transcriptionally active. Heterochromatin can be characterized as two main forms: facultative, which is a transient form of heterochromatin important for selective gene silencing and constitutive heterochromatin, which is a more permanent form found near centromeres and telomeres. Constitutive heterochromatin is localized at peri-centromeric and telomeric regions in *Drosophila*[1], and is mainly composed of repeated sequences, ribosomal DNA and transposable elements, although at least 250 protein coding genes, some essential, are present.[2] Critical roles of heterochromatin include peri-centromeric cohesion,[3] meiotic chromosome segregation,[4] telomere maintenance[5], and double strand break (DSB) repair.[6] Heterochromatin has been shown to form via a network of histone and DNA modifying enzymes, chromatin remodeling complexes, and RNA interference (RNAi).[7][8] RNA interference (RNAi) based mechanisms for heterochromatin formation are well established in *S. pombe*[9], and are implicated in Drosophila[10] and a few other species, [9,10] although RNAi independent mechanisms exist as well. [10][11][12][13]

Heterochromatin formation in *Drosophila* is mediated by methylation of Histone 3 lysine 9 (H3K9) by specific methyltransferases (e.g. Su(var)3-9) in cycles 13 and 14, to produce H3K9me2/3, a major marker of heterochromatin, followed by binding of Heterochromatin Protein 1a (HP1a) to H3K9me2/3 [14,15], which then becomes 'spreading competent'. [16] HP1a in turn recruits more Su(var)3-9, creating more binding sites for HP1a and thus facilitating HP1a and H3K9me2/3 spreading.[17] The RNAi pathway also seems to have a role in heterochromatin maintenance, as Dicer2 is needed for H3K9me2 heterochromatin association [18] and piRNA components are needed for proper HP1a localization and H3K9me2/3 levels.[19] How heterochromatin is initially established in *Drosophila* is largely unknown. There is evidence, however, that prior to H3K9me2/3 deposition, Su(var)3-3 demethylates H3K4me1/2, leading to de-acetylation of H3K9 by Rpd3 and subsequent di and tri methylation of H3K9 in cycles 13 and 14.[20] The abundance and distribution, however, of H3K4me1/2 and H3K9ac in heterochromatic sequences prior to H3K9me2/3 formation, whether this sequence of events is necessary for heterochromatin formation, and more importantly how these enzymes are targeted to heterochromatic sequences are largely unknown.

Others have shown that Piwi is likely a necessary component of heterochromatin formation[21], and that maternally loaded factors are necessary for HP1a recruitment to the 359 bp tandem repeat. [22] Also, HP1a is capable of binding RNA via its hinge domain,[23] and the majority of proteins that bind HP1a[24] have putative RNA binding domains (personal analysis). A series of recent papers have also demonstrated that human and mice SUV39H1 bind single-stranded RNA (ssRNA), that this binding stabilizes SUV39H1 to heterochromatin, and that human SUV39H2 anchors heterochromatin proteins in an RNA-dependent manner.[25][26][25] In mice, reverse repeat transcription of a satellite facilitates paternal DNA sequence association with maternal heterochromatin, which is necessary for progression of embryos past the two cell stage.[27] Also, in mice embryonic stem cells (mESC), a processed lncRNA is critical for guiding complexes to a rDNA promoter to initiate heterochromatin formation around the nucleolus. [28] Taken together, this strongly implicates an RNA dependent targeting mechanism for initiation of heterochromatin formation in *Drosophila*.

It is intriguing that RNAs, in particular non-coding RNAs (ncRNAs), can have functions independent of coding for proteins, and that a major function includes the proper structuring of chromatin. One class of ncRNAs, long-ncRNAs (lncRNAs), which are currently defined as transcripts longer than 200 nucleotides that do not encode small RNA such as ribosomal RNA (rRNA) or transfer RNA (tRNA) and are devoid of an ORF greater than 20 amino acids. lncRNAs regulate transcription by at least four known mechanisms,

including recruitment of silencing factors, guiding and scaffolding chromatin-modifying complexes, titrating proteins away from certain genomic locations, and participating in enhancer like functions.[29] In addition to some lncRNAs having roles in imprinting and dosage compensation[30], lncRNAs such as Xist[31], HOTAIR[32], Kcnqot1[33], AIR[29] and ANRIL[34] have either additional or separate functions in heterochromatin-mediated gene silencing.[29] Xist is one of the most studied lncRNAs, inducing female X chromosome inactivation in mammals by acting as a guide and scaffold, spreading along the X-chromosome in cis and recruiting chromatin-associating silencing factors.[35] Xist itself is regulated by the Tsix ncRNA, which promotes X chromosome counting and prevention of Xist expression on the active X in females.[36-38] Drosophila also utilizes non-coding RNAs as a dosage compensation mechanism, but instead of silencing one X chromosome in females, in males the X chromosome is transcriptionally enhanced. Briefly, roX1 and roX2 ncRNAs bind the Msl2 and Mle protein components of the male specific lethal complex (MSL), to promote H4K16 acetylation (H4K16ac) on the X chromosome.[39]

Other ncRNAs with validated chromatin or nuclear-based functions include repetitive RNAs. In mouse embryonic fibroblasts (MEFs), for instance, a subset of repetitive ncRNAs promote heterochromatin formation and prevent the epithelial-to-mesenchymal transition.[22] In contrast, non-coding transcripts from long-interspersed nuclear element 1 (LINE1) repeats associate with euchromatin and prevent chromatin compaction in mammals.[40] Furthermore, repetitive, GC rich transcripts found in repeat expansion disorders such as amyotrophic lateral sclerosis (ALS) are capable of undergoing protein independent phase-separation,[41] raising the question of whether or not repeat expansion disorders are at least partially driven by pathogenic, phase separating RNAs.

Due to extreme limitations of studying ncRNAs in-vivo in whole animals, most of the studies mentioned above were carried out in cell culture and when appropriate, mechanisms elucidated in-vitro. Furthermore, the majority of ncRNA mechanistic studies were performed on model lncRNAs, i.e. Xist and roX1 and roX2. Not all mechanisms of ncRNAs can be expected to be similar to model lncRNAs, and most importantly functions and mechanisms seen in cell culture do not always translate into functions and mechanisms in-vivo in whole animals, highlighting the critical need to perform studies in whole animals. *Drosophila* is an excellent model organism with which to perform studies in-vivo due to its gene similarity to humans, human disease models, short lifespan, ease of genetic manipulation, and in the case of chromatin studies and genetics, an extensive repertoire of knowledge gained from over 100 years of research. Additionally, thanks to the modENCODE projects [42][43] *Drosophila melanogaster* heterochromatin, at least in terms of sequence, location, and histone mark distribution, is the most detailed of any developing organism.

Pericentromeric heterochromatin flanks centromeres, the sites of kinetochore attachment, which are necessary for mitotic segregation of chromosomes. Centromeres differ in histone modification composition compared to pericentromeric heterochromatin, being enriched for canonical histones H3K4me2, H3K36me3 and centromere specific histones CENP-A and CID.[44] However, similar to pericentromeric heterochromatin, centromeres are composed of blocks of repeats and transposons.[45] Once thought to be transcriptionally inert, it is now clear that some centromeric ncRNAs are critical for CENPA deposition and subsequent kinetochore assembly. [46][47][48] This begs the question of whether or not pericentromeric heterochromatin produces repetitive transcripts, and whether or not these transcripts contribute to heterochromatin function. Going farther, if heterochromatin produces functional transcripts, is it possible that these transcripts are responsible for certain heterochromatic functions, rather than the underlying DNA/chromatin itself? For instance, heterochromatin contributes to genome organization through association of select heterochromatin specific markers,[24] many of which have unstructured, putative RNA binding domains. In addition, heterochromatin formation is driven by HP1a mediated phase separation,[49] HP1a is capable of binding RNA,[23] and others have theorized that RNAs mediate the formation and identity of certain phase separated granules.[47][48] Together, this and other data strongly suggest that RNA is at the very least a critical component in the formation and organization of heterochromatin, and it is tempting to

speculate that established functions of heterochromatin mentioned above depend on heterochromatic RNA. These are fundamental questions that piqued my curiosity early in graduate school and, despite knowing it would be quite challenging, drove me on a journey to study heterochromatic ncRNAs.

Pericentromeric heterochromatin is rich in repetitive sequences such as transposons and ribosomal DNA (rDNA), in addition to satellite DNA. Satellites are highly repetitive DNA blocks that occur primarily in heterochromatin, and in *Drosophila* likely comprise up to 20% of the genome. Large blocks of long tandem repeats include the 359 bp, Responder elements and dodeca satellites. However, most *Drosophila* satellites are composed of short (5-10 bp) simple sequences that extend up to several mega base lengths in tandem, and generally are interspersed with other simple sequences, or transposons such as LINE or SINE elements. The most abundant of these simple sequences include AAGAG, AACATAGAAT and AATAT.[50] AATAACATAG repeats are found predominantly on 2nd and 3rd chromosomes, while the AATAT repeat is found on all chromosomes, but predominantly on the Y and 4th. AAGAG repeats are by far the most abundant in *Drosophila* and in fact are estimated to comprise 5-6% of the genome.[51] AAGAG(n) is found on every chromosome but is most abundant on the Y, 2nd and 4th chromosomes. In chromosome 2, for instance AAGAG(n) spans the centromere and is highly abundant in 2R heterochromatin.[52] It is plausible that AAGAG(n) repeats make up 15% of heterochromatin, given that AAGAG(n) comprises 5-6% of the *Drosophila* genome and that all major blocks of AAGAG(n) repeats are found in heterochromatin, which itself comprises about 30% of the genome.[43] However, major limitations in sequencing technology have thwarted efforts to fully assemble the satellite component of heterochromatin due to their highly repetitive nature. Thus, although most satellites and their general cytological location within heterochromatin have been mapped, the precise sequence and length of long stretches of tandem satellite repeats are still largely unknown.

Previous work demonstrated that transposons and some of the simple repeats described above are transcribed and likely processed into small-RNA forms (<30nt).[53] Traditional RNA sequencing methods, however, have had limited success at identifying longer transcripts generated from tandem satellite repeats. Thus, it is necessary to take alternative approaches to determine which, if any, of the major *Drosophila* satellites are transcribed into longer (>50base) RNAs.

These observations led to several major questions I wished to address during my graduate work. Essentially, I wanted to determine which, if any, ncRNA transcripts are transcribed from heterochromatin, and to quantitate their relative abundances. Assuming that the most abundant transcript or class of transcripts would be the best candidate for further investigation, I then wanted to determine cytologically where and when in development it is transcribed, identify potential functions of this transcript, and if functional, identify the molecular mechanisms.

Considering that most ncRNAs putatively expressed from heterochromatin would be repetitive in nature, combined with the limitations in traditional RNA-seq of repetitive RNAs, I chose to take a non-traditional approach to heterochromatic RNA identification. Basically, I proposed to use RNA-Fluorescent In-Situ Hybridization (RNA-FISH) to the most abundant heterochromatic repeats to identify which of those are transcribed, then for those with demonstrated expression, confirm their presence using northern blotting combined with kmer analysis, (which refers to identifying all possible sequence combinations of a given length), from Illumina deep sequencing.(Chapter 2) For the most predominant satellite RNA, I then sought to characterize sub-nuclear localization patterns in regards to association with heterochromatin, developmental stage and tissue.(Chapter 2) Then, in order to guide functional and mechanistic studies I proposed to determine which proteins these transcripts bind via ncRNA pulldown and mass-spectrometry (ms) (Chapter 2). I then proposed to identify function(s) of this or these satellite ncRNAs by assessing phenotypes in flies depleted for the transcript (Chapters 3 and 4). If functional, I then proposed to investigate mechanisms(s) by assaying whether or not this/these transcript(s) affect organization of these

proteins. Additionally, I sought to identify lncRNAs that associate with heterochromatin and determine their sub-cellular and sub-nuclear distribution (Chapter 5).

Chapter 2. Identification and characterization of AAGAG(n) RNA containing transcripts

RNA-FISH reveals transcription of satellite repeats

I initially used RNA-FISH to determine which, if any, heterochromatic tandem repeats are transcribed in fly embryos. RNA probes were made to repeat regions that are either abundant in heterochromatin, as well as repeat RNAs shown to bind to HP1a.[54] Essentially, out of 36 repeats tested for expression (Table 1) of one strand in 0-8hr Oregon R (OR) embryos, 9 gave visible, nuclear localized expression patterns (data not shown). These repeats consisted of, in the RNA-sequence form, AAGAG(n), AACAC(n), AATAC(n), AAGAGAG(n), AATAGAC(n), AATAC(n), AAGAC(n), SatIII(n) (commonly referred to as 359bp repeat) and AATAACATAG(n).

Table 1. Tandem repeats tested for expression via RNA-FISH

AGCC	AAGCG	ATCGC	AAAG	AATAG	AATT
ACCC	AGCCCC	ATC	AACAG	AATAGAC	AATC
ATTAC	AGCCG	AAACC	AAGGAG	AAGAC	AAGAG*
AGCGG	AACGG	AAACG	AAGAGG	AATAACATAG	ATTAC*
AGC	ACCCC	ATATC	AAGAGAG	AACAC*	AGCC
AACTC	AGTCT	CCCCG	AATAC*	dodeca	359 satellite*

* denotes those tested for expression of the opposite strand

To determine if any of these repeat RNAs associated with heterochromatin marker H3K9me2/3, I then performed co-Immuno Fluorescence RNA-FISH (co-IF RNA-FISH) and looked for co-localization with the heterochromatin-enriched H3K9me2/3 modification. Interestingly, all except AATAACATAG(n) co-localized either partially or significantly with H3K9me2/3. RNAs that associated significantly with H3K9me2/3 included AAGAG(n), AACAC(n), ATTAC(n) AAGAGAG(n), and AATAGAC(n). By far the most abundant RNA repeat according to RNA-FISH signal intensity and prevalence was the AAGAG(n) repeat, which we focused on for further characterization.

AAGAG(n) RNA presence is confirmed via kmer analysis of stranded RNA-seq.

In order to validate that AAGAG(n) RNA is present at the sequence level and not simply an artifact or off-target effect of RNA-FISH, we performed kmer analysis on unmapped, stranded RNA-seq reads generated through the modENCODE project[55] (reads provided by S. Celniker). Due to short Illumina sequencing read sizes, tandem repeats longer that the read size (roughly 45-55 bases after trimming adapters) cannot be properly aligned and annotated since they will not have unique ends. Therefore, in order to see if the longest possible AAGAG(n) transcript was present in the data, we searched, or 'grepped,' for a $\geq$ 45 base stretch of AAGAG RNA (AAGAG in 9 tandem repeats, referred to subsequently as AAGAG(9)) for all embryonic stages. Interestingly, for all stages, including the 0-2hr stage in which most if not all RNA is maternally loaded, this approach identified thousands of reads for AAGAG(9) RNA. "Grepping" from the kmer analysis other repeats tested in RNA-FISH experiments yielded much fewer reads, suggesting that AAGAG(n) RNA may be the most abundant tandem repeat RNA. Surprisingly, we did not find any reads of the antisense transcript CTCTT(5) (25 bases) or longer in any embryonic stage, and in fact a shorter form CTCTT(n) was present in only a few reads as CTCTT(3).

AAGAG(n) RNA is maternally loaded and persists throughout development

RNA-FISH does not allow for determination of transcript length, and additionally repeat RNA size cannot be determined from RNA-seq data if read lengths are shorter than the repeat RNA. Considering that thousands of reads existed for the maximum length of AAGAG RNA (AAGAG(9)) that could possibly be identified in our RNA-seq dataset, combined with knowledge that up to megabase blocks of AAGAG(n)

DNA exist, I theorized that AAGAG(n) RNA existed in forms longer than 45 bases. To address this, I performed northern blots on AAGAG(n) RNA from Oregon R embryos from different stages of embryogenesis and third instar larvae (L3). As Fig. 1 demonstrates, AAGAG(n) RNA is maternally loaded prior to the maternal-zygotic transition (MZT) at cycle 14 in embryos in an approximately 1,000 nucleotide (nt) form. In later stage embryos and third instar larvae (L3), AAGAG(n) RNA is present in approximately 750nt, 500nt and smaller RNA forms. We did not detect the presence of reverse transcript CTCTT(n) in any stage of embryogenesis or L3 by northern blot analysis (data not shown). Combined with our kmer analysis on RNA-seq reads, I conclude that AAGAG(n) RNA is maternally loaded and persists throughout development in a uni-directionally transcribed form.
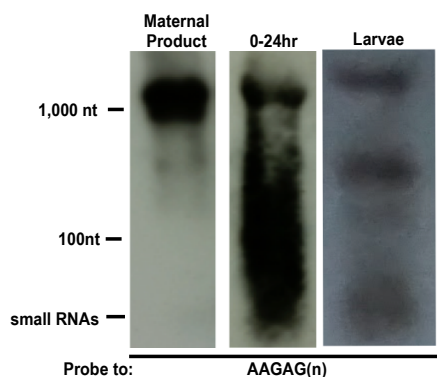

Fig. 1 AAGAG(n) RNA is maternally loaded and persists throughout development. Shown are northern blots hybridized with a probe that binds AAGAG(n) RNA.

AAGAG(n) associates with early heterochromatin in a ssRNA form and is enriched in neural tissue
RNA-seq and northern blot analysis demonstrated that AAGAG(n) RNA exists throughout development, but these methods do not determine if it displays tissue or cell type specific expression patterns. In order to characterize the presence of AAGAG(n) RNA at different stages of early embryos, I performed RNA-FISH prior to (cycles 0-13), during (cycles 13 and 14) and immediately after (cycle 15) the MZT. It is important to note that this technique results in a 'snapshot' of RNA distribution in the nucleus at a given time and is therefore not a representation of all nuclei that have produced or will produce AAGAG(n) RNA. Importantly, we observed that in all cell types analyzed, AAGAG RNA was concentrated in one or a few sites, termed 'foci', rather than being uniformly distributed in the nucleus. Additionally, foci detected can either be sites of nascent transcription or the localization of RNA after movement in *trans*. Regardless, at cycle 10 (10 embryos imaged), none of the embryos contained AAGAG RNA foci, while at cycle 12 (3 embryos imaged) 30% had one or more foci in roughly 5% of nuclei (Fig 2A.). In cycle 13 (3 embryos imaged), 65% of embryos had one or more foci in about 10% of nuclei (Fig. 2B), and in cycle 14 (>30 embryos imaged), 100% of embryos had foci in roughly 20% of nuclei (Fig. 2C). AAGAG(n) RNA detected as foci in these embryos were not expressed selectively from any particular pre-cursor tissue, as foci were found distributed evenly throughout the embryo, and by cycle 12 developmental patterning and tissue differentiation has already begun. Surprisingly, very few AAGAG(n) RNA foci were detected in cycles 15 up until mid-stage embryogenesis. This suggests that foci present in cycle 12 embryos is of maternal origin, AAGAG(n) RNA is transcribed during the MZT (cycle 14) and AAGAG(n) transcription in some of these precursor tissues is abolished or decreased in later stages.
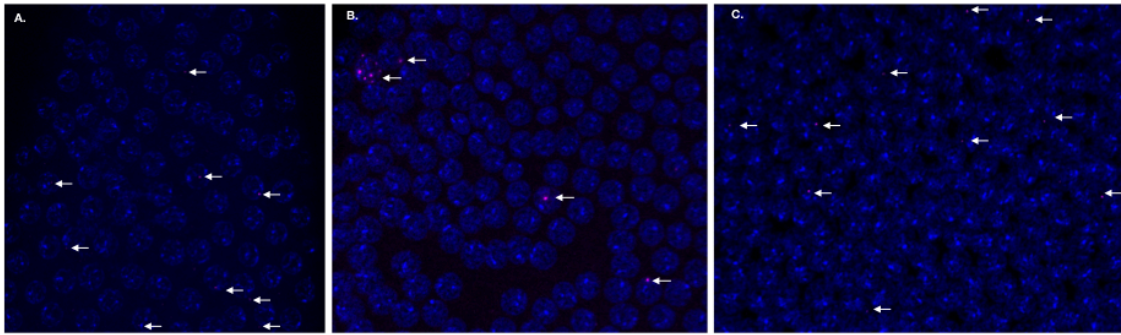
Fig. 2. Examples of AAGAG(n) RNA distribution in cycle 12, 13 and 14 embryos. Projections through embryo nuclei stained with DAPI (Blue). AAGAG(n) RNA foci is shown in magenta and marked with arrows. A.) Cycle 12 embryo B.) Cycle 13 embryo C.) Cycle 14 embryo

Considering that AAGAG(n) RNA foci in cycle 14 embryos was localized to chromatin, we wished to determine if the RNA was present as an RNA/DNA hybrid or in a dsRNA form, as this may potentially guide future mechanistic studies. I therefore performed RNA-FISH to AAGAG(n) on early embryos pre-treated with RNaseIII (which cleaves dsRNA[56]), RNaseH (which cleaves the RNA strand in RNA/DNA hybrids), RNase I (which non-specifically cleaves ssRNA and dsRNA), and RNaseA (which cleaves adjacent to pyrimidines, preferentially in ssRNA, and specifically not between purines such as 5'-AGAAGGGAGAAG [57][58]). Interestingly, AAGAG(n) RNA foci were present in cycle 14 embryos after RNAseH, RNaseIII, or RNAseA treatments, but not in RNase1 treated embryos (Fig. 3). Reduced number of foci and decreased foci intensity were seen in RNaseIII treated embryos compared to controls (data not shown). Additionally, to further confirm that the AAGAG(n) RNA probe hybridized to RNA and not DNA under these conditions, after probe hybridization I treated the samples with either RNaseIII, which should cleave RNA/RNA hybrids or RNAseH, which should cleave the RNA in RNA/DNA hybrids. Under these conditions, foci were not detected after RNAseIII but detected robustly after RNAseH treatment (data not shown), demonstrating that the probe binds only to RNA, and not DNA. Together, I conclude that AAGAG(n) RNA foci in cycle 14 embryos are predominantly present in a ssRNA form, and not present as R-loops (DNA/RNA hybrids).
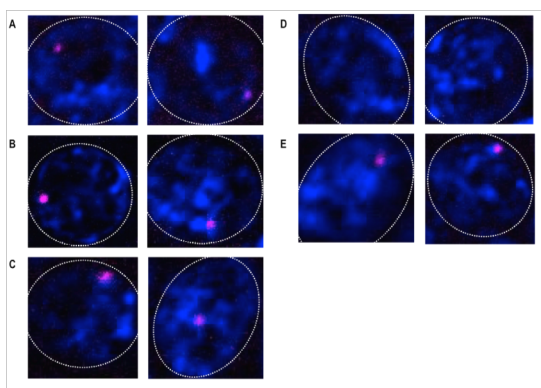
Fig. 3. AAGAG(n) RNA foci after different RNase treatments. Shown are slices of cycle 14 (or 12) nuclei with or without foci from two different embryos. The nuclear periphery is outlined in dotted circles A.) No RNase control. B.) Treated with RNaseIII (left nuclei is cycle 12 and right is cycle 14) C.) RNaseH D.) RNase1 and E.) RNaseA

I next wanted to determine cytologically where AAGAG(n) was present in the nucleus, its localization relative to the bulk of AAGAG(n) DNA, and if it associated with the heterochromatin marker H3K9me3 in early embryos. This was accomplished by performing co-IF RNA/DNA-FISH with H3K9me3 antibody and probes to AAGAG(n) RNA and AAGAG(n) DNA in cycle 11-14 embryos, prior to (cycles 11-13) and during (cycle 14) the MZT. Consistent with the northern blots and kmer analysis data, AAGAG(n) is maternally loaded and present in nuclei prior to cycle 12, before the MZT (Fig. 4). Notably, the few AAGAG(n) RNA foci present in cycle 12 do not co-localize with the majority of AAGAG(n) DNA, which at this stage is largely unorganized and present throughout most of the nucleus. However, as Fig. 4 demonstrates, in cycle 14 during the MZT, AAGAG(n) RNA foci co-localizes with the newly deposited H3K9me3 mark and with AAGAG(n) DNA that is now highly organized and associated with heterochromatin forming around the nucleolus. Though speculative, the coincident timing of H3K9me3 appearance suggests that AAGAG(n) RNA may target component(s) to heterochromatic sequences to promote early heterochromatin formation and/or or in response to formation of a phase separated domain.
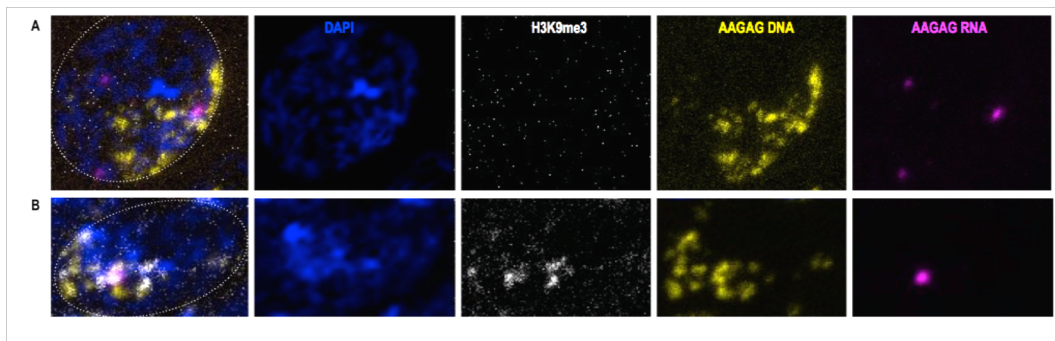
Figure 4. AAGAG(n) RNA foci is present prior to heterochromatin formation and associates with early heterochromatin. Shown are co-IF RNA/DNA-FISH projections of nuclei (outlined in dotted lines) probed for both AAGAG(n) RNA and DNA and stained for heterochromatin marker H3K9me3 and DNA (DAPI). A) cycle 12 nuclei prior to heterochromatin formation and B) early cycle 14 nuclei during the time of heterochromatin formation

In later stages of embryogenesis, AAGAG(n) RNA is highly enriched in the nuclei of the ventral ganglia (neural tissue) in mid-late stage embryos (Fig. 5A) and is nuclear localized in every larval tissue such as brain (Fig. 5B), salivary glands (Fig. 5C) and gut (Fig. 5D). Higher magnification analysis of larval neurons revealed that AAGAG(n) foci co-localize with pericentromeric heterochromatin, which is found around the nucleolus in non-dividing cells (Fig. 5E). Salivary glands contain highly polytenized tissue, in which the pericentromeric regions are highly underrepresented and form what is called a chromocenter.[59] In this tissue I observed that the predominant AAGAG(n) RNA signal is localized to the chromocenter, further confirming that AAGAG(n) RNA associates with heterochromatin, and that the AAGAG(n) foci are not derived from euchromatic arms (Fig. 5F). I conclude that AAGAG(n) RNA, from early embryos to larvae, is nuclear localized and associated with heterochromatin.
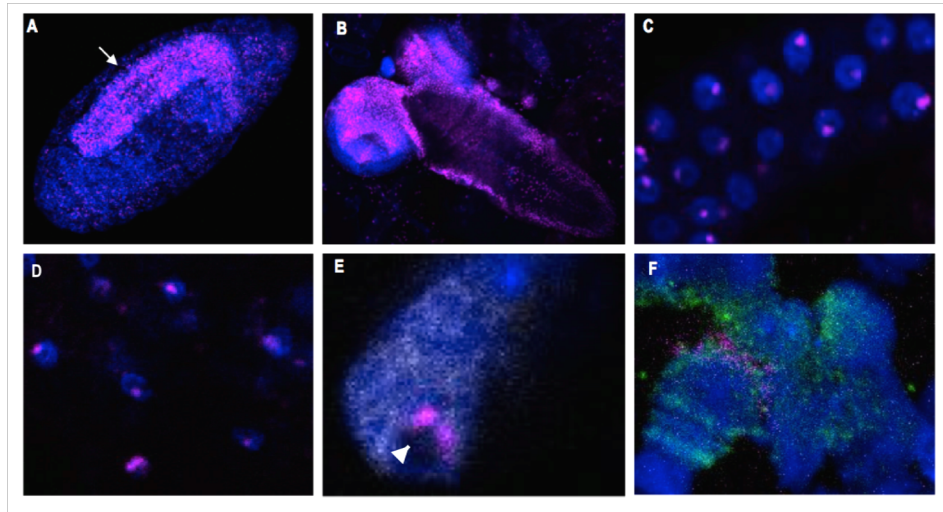
Fig. 5. AAGAG(n) RNA localizations in embryo and larvae. Shown are AAGAG(n) RNA distributions (magenta) throughout development. DNA is indicated with DAPI (Blue) A) Projection of a mid-stage embryo in which AAGAG(n) RNA is highly enriched in the ventral ganglia. B-E) Slices of larval tissue and cells B) Nuclear AAGAG(n) RNA foci in larval brain lobe C) Salivary Glands and D) Gut. E) AAGAG(n) RNA associates with heterochromatin. Arrow points to the nucleolus, around which pericentric heterochromatin is located. Histone are represented in white from antibody staining. F) H3K9me2 (green) marks the chromocenter of a salivary gland squash. Inset is a projection of the chromocenter, which is enriched for AAGAG(n) RNA.

AAGAG(n) transcripts derive from multiple genomic locations

We then wanted to determine the origin of these AAGAG(n) RNA transcript(s), since doing so could yield crucial information for further experimentation. Our first approach to address this, in collaboration with Grace Lee, PhD, utilized Illumina deep sequenced reads courtesy of S. Celniker. Essentially, reads containing one end with at least three repeats of AAGAG(n) and the other end mappable to a region of the genome were assembled into contigs using 'phrap', then blasted against release 6 of the genome.[60] I then manually identified which contigs were next to blocks of AAGAG(n) or AG(n) DNA in the genome. This approach identified one transcribed region in 2R heterochromatin adjacent to a 992 bp block of AAGAG(n) DNA (chr2R:1826680-1826741), in addition to two regions in X chromosome heterochromatin next to either annotated AAGAG(n) or AG(n) rich blocks (chrX:12,660,068-12,660,221; chrX: 12660096-12660145). I then performed co-RNA-FISH to each strand of these regions to see if they co-localized with AAGAG(n) RNA foci in early embryos, the ventral ganglia in later staged embryos, as well as L3 brain lobes. In cycle 13 and 14 embryos, RNA foci was not detected with probes to these unique regions. However, in the ventral ganglia of late stage embryos, probes binding to the AAGAG(n) RNA containing strand of chr2R:1825640-1825699 and chr2R:1,826,690-1,826,893, which together flank the 992bp region of AAGAG(n) DNA, detected foci that co-localized with AAGAG(n) RNA foci (Fig. 6A, top showing chr2R:1825640-1825699 foci). For region chr2R:1825640-1825699, all foci from this region co-localized with AAGAG(n) RNA signal, but not all AAGAG(n) RNA signal (26%) co-localized with chr2R:1825640-1825699 RNA signal, indicating that roughly 74% of AAGAG(n) RNA signal did not come from the transcript containing chr2R:1825640-1825699 RNA. The chr2R:1,826,690-1,826,893 flanking region exhibited RNA co-localization with AAGAG(n) foci in 100% of nuclei with AAGAG(n) RNA signal (data not shown). For

each of these regions, the opposite strand, which would bind to a CTCTT(n) RNA containing strand, were rarely if ever detected as foci in the ventral ganglia. In X heterochromatin, a unique region adjacent to a GA(n) rich block and near A(n) rich low-complexity sequences (chrX:12,660,068-12,660,221), co-localized in both strands approximately 40% with AAGAG(n) foci. For example, 60% of AAGAG(n) foci did not co-localize with chrX:12,660,068-12,660,221 RNA, although all foci from chrX:12,660,068-12,660,221 co-localized with AAGAG(n) in ventral ganglia of embryos. For both of these regions, the strand that would contain AAGAG(n) or AG(n) rich regions exhibited higher signal/noise than the opposite strand (Fig. 6A, bottom). Other unique sequence regions next to AAGAG(n) transcripts found with 'phrap' included chrX:11,830,841-11,831,017 and chrX:22,453,001-22,453,133, although neither strand of exhibited convincing foci in the ventral ganglia in embryos. These results demonstrate that AAGAG(n) RNA foci found in cycle 12-14 embryos are not from these 2R or X genomic regions, but that AAGAG(n) RNA foci found in the ventral ganglia originates at least in part from strand specific transcription of this 2R region and chrX:12,660,068-12,660,221. Surprisingly, in the brain lobe of L3, all regions, regardless of strand, co-localized approx. 100% with AAGAG(n) RNA foci (Fig. 6B-F). This result is perplexing considering that CTCTT(n) foci are never detected in any stage of embryos or larvae, which would suggest that either transcription stops at blocks of pyrimidines, or that transcribed pyrimidine blocks are quickly degraded. Regardless, this demonstrates that all three of these regions are transcribed in L3 brain lobes and co-localize into one focus.

The observation that the 2R region and only one of the X regions are uni-directionally transcribed in embryo ventral ganglia, combined with transcription of all three regions in both strands in L3 brain lobes suggests transcription of these satellites is cell-specific and developmentally regulated. Additionally, the observation that all RNA foci from these regions in brain lobes co-localizes with AAGAG(n) RNA suggests that repetitive heterochromatic transcripts generated from different genomic locations cluster together. This is an interesting observation and suggests that heterochromatic transcripts form sub-heterochromatic foci, although higher resolution imaging and more combinatorial probe labeling will be crucial to understanding the detailed spatial relationship between these transcripts.

We also note that other sources of AAGAG(n) RNA must exist, as RNA from the regions mentioned above were not seen in cycle 12, 13 or 14 embryos, when AAGAG(n) foci is first detected. Therefore, other methods for RNA-seq that do not suffer from PCR bias and short reads, such as Nanopore sequencing (see next) will be required to identify the source of all AAGAG(n) containing transcripts.
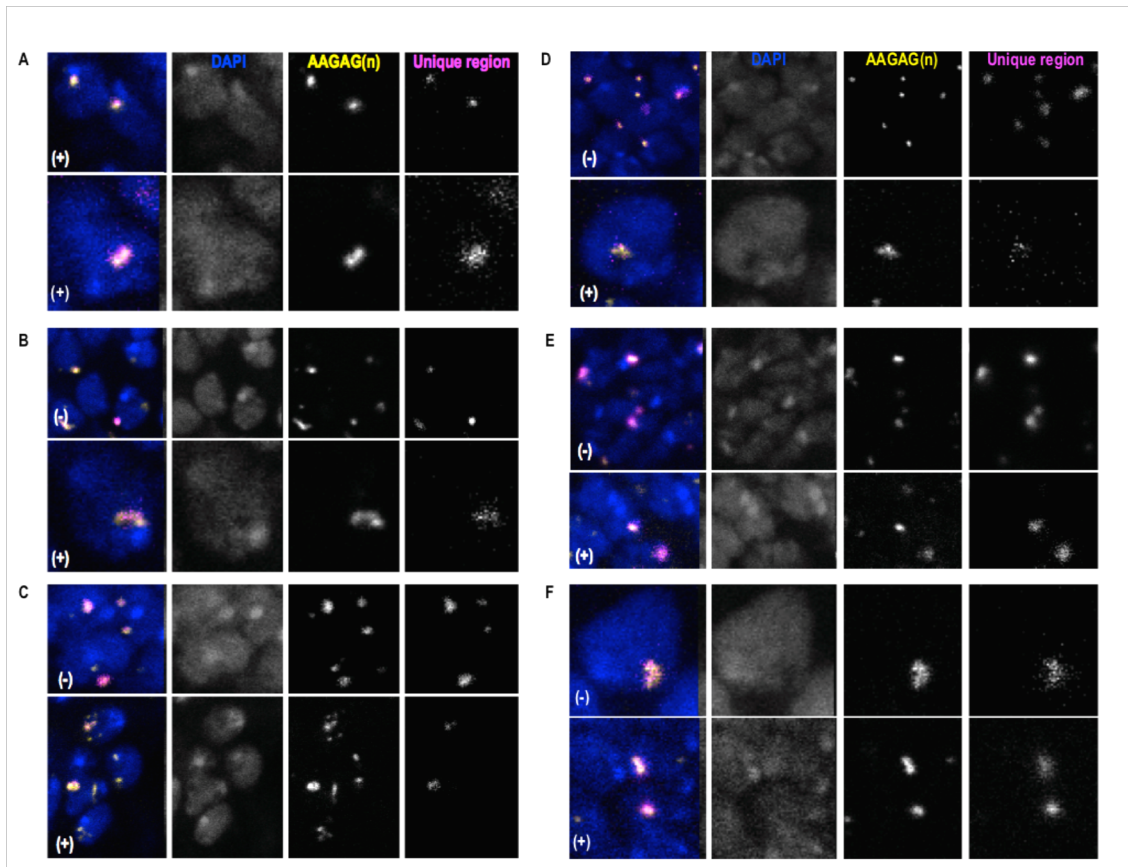
Fig. 6. Heterochromatic Chr2R and ChrX regions containing AAGAG(n) satellites are transcribed in embryos and larval brain. Shown are slices of embryo ventral ganglia or L3 brain lobe nuclei stained with DAPI (blue), and RNA-FISH to AAGAG(n) (yellow) and unique region locations (magenta). 'Unique Region' RNA-FISH was performed with tyramide signal amplification (TSA) as necessitated by the assay, hence the poorer resolution compared to AAGAG(n) RNA, which was detected without TSA. Images labelled (+) were probed to the strand containing AAGAG(n) or AG(n) blocks. Note that 'unique region' probe binds to regions adjacent to AAGAG(n) or AG(n) and not AAGAG(n) or AG(n) sequences themselves. A) Nuclei from late embryo ventral ganglia with RNA-FISH to AAGAG(n) and chr2R:1825640-1825699 (top) or chrX:12,660,068-12,660,221 (bottom). B-F) Slices of nuclei from L3 brain lobes with RNA-FISH to AAGAG(n) and the following: B) Unique region chrR:1825640-1825699 C) (other side of AAGAG(n) repeat); chr2R:1,826,690-1,826,893 D) chrX:11,830,841-11,831,017 E) chrX:12,660,068-12,660,221 F) chrX:22,453,001-22,453,133

Nanopore sequencing

One approach theoretically suited for RNA-seq of long, repetitive RNA or DNA utilizes Oxford Nanopore-based sequencing, which does not require PCR enrichment and is capable of sequencing up to several megabase read lengths. Recently, it was used to assemble the human centromere in the highly repetitive Y-chromosome[61] and has been used to sequence long stretches of repetitive DNA. [59-61] This approach is also capable of sequencing RNA,[62] although the efficiency in sequencing repetitive RNA has not been determined.

We theorized that Nanopore sequencing could help us elucidate the identity and source of other AAGAG(n) RNA transcripts. Nanopore sequencing suffers from high error rates, and therefore it was critical to enrich for AAGAG(n) RNA, which is likely a very small fraction of all non-rRNA, in order to guarantee as many reads as possible. To do this, I essentially used a biotinylated antisense probe to pull out AAGAG(n) RNA

from significant quantities of total larval RNA, performed several washes, then sequenced the resulting pull-down DNA. (see materials/methods for full details).

This approach resulted in an enrichment of AG rich RNAs, some of which are serval kilobases in size, in addition to RNAs with tandem AAGAG repeats (data not shown). One RNA sequence, for instance, was 8.2 kb, composed predominantly of As and Gs, and contained 45 instances of AAGAG. However, only 4 reads had three tandem repeats of AAGAG(n), suggesting that the technique and/or software analysis is inefficient at reading tandemly repetitive regions. After consultation with experts, this appears to be a software issue in which read calling from stretches of tandem repeats is deficient, but that the raw data still accurately detects nucleotide sequence of repetitive regions. We are currently working with pioneers in the Nanopore field to address this issue in order to obtain accurate reads from our dataset. These complications were not unexpected due to the infancy of Nanopore technology combined with what we believe is the first approach to sequence tandemly repetitive RNA. We expect to obtain accurate reads from our dataset in the future in order to identify all sources of AAGAG(n) RNA.

AAGAG(n) RNA binds a diverse set of proteins in larvae

The observation that AAGAG(n) RNA associates with heterochromatin, is present in most stages of development, and is enriched in neural tissue suggested that it may have a role in chromatin organization and potentially development, and therefore we wished to know what proteins it binds in order to guide functional assays and mechanistic studies. Fortunately, several techniques have been developed to assay RNA-protein interactions. Techniques capable of mapping ncRNA contacts with DNA, such as chromatin isolation by RNA-purification (CHiRP-seq) and capture hybridization analysis of RNA targets with deep sequencing (CHARt-seq), are powerful tools that have elucidated mechanisms of lncRNA mediated chromatin organization. [63,64,65,39,66] A variation in these general approaches, termed CHiRP-mass spec (CHiRP-ms), was adapted to identify ncRNA-protein interactions in human cell culture,[67] suggesting that it could be used to determine AAGAG(n) RNA-protein interactions. I subsequently modified this technique for L3 AAGAG(n) RNA pulldown and mass-spec of associated proteins. Though technically challenging, this approach identified 94 proteins that bound AAGAG(n) RNA but not in a scrambled control pull-down, with the exception of actin-87E, the only protein pulled down in the scrambled control (Table 2). Some of the proteins identified include chromatin bound proteins known to be enriched at centromeric sequences, heterochromatin, or associated with heterochromatin proteins, such as Dodeca-satellite binding protein 1 (Dp1) and Tudor-staphylococcal nuclease (Tudor-SN), among others. [24,68] Additionally, AAGAG(n) RNA binds proteins necessary for spermatogenesis and/or proteins enriched in testes, such as Heat shock protein 83 (Hsp83), Heat shock protein cognate 5 (Hsc70-5), Spectrin, Trap1, and Tudor-sn. [69-71] To our knowledge, this is the first utilization of CHiRP-ms in a whole organism, and more importantly, this approach guided further functional experimentation, such as characterizing AAGAG(n) in spermatogenesis.

Table 2. Proteins bound by AAGAG(n) RNA via CHiRP-MS

eEF1alpha1	alphaTub84B	blw	COX4	Gpdh	Pdi	scu	UQCR-C2
ACC	alphaTub85E	btsz	Dp1	GstD4	Pfk	SERCA	Vha100-2
Acon	apolpp	CCT7	eEF1gamma	Hsc70-4	Prm	sesB	Vha26
Act57B	Argk	CG10932	eEF2	Hsc70-5	pug	Sgs3	Vha55
Act5C	Atpalpha	CG15093	eIF4A	Hsp23	PyK	Sgs7	Vha68-2
Act87E	ATPCL	CG1640	Eno	Hsp83	regucalcin	Sgs8	
Actn	ATPsynbeta	CG7461	FASN1	Irp-1A	RpL10Ab	TER94	
Adh	ATPsyngamma	CG7920	Fbp1	Irp-1B	RpL18	Tpi	
AIF	ATPsynO	CG8036	Gapdh1	kdn	RpS3A	Trap1	
Ald	beta-Spec	CG9914	Gdh	l(1)G0156	RpS6	Tudor-SN	
alpha-Spec	betaTub56D	Chd64	GlyP	PCB	Rpt1	UQCR-C1	

Conclusions and Future directions

Here, I have demonstrated that at least a subset of satellite repeats are transcribed and that the most abundant satellite transcripts, satellites containing AAGAG(n) RNA, are maternally loaded, persist throughout development and associate with heterochromatin in all cells. In early embryos prior to the MZT, AAGAG(n) containing RNA is present as foci and associates with the earliest stages of heterochromatin formation in a ssRNA form, suggesting (speculatively) that this RNA targets complexes to heterochromatic sequences in order to nucleate heterochromatin. I also confirm the presence of AAGAG(n) RNA sequence in RNA-seq datasets at all stages of development, and specifically note the absence of CTCTT(n) RNA in these datasets, in northern blots, as well as RNA-FISH, suggesting that AAGAG(n) RNA is exists throughout development in a uni-directional manner.

Considering that AAGAG(n) containing RNAs associate with heterochromatin, it is not surprising that this RNA sequence binds proteins involved in chromosome organization, such as Dp1, Tudor-SN and Heat shock proteins[24,68 72-74]. The presence of proteins important for spermatogenesis/enriched testes in the pulldown was surprising, however, and suggested AAGAG(n) containing RNA could be important for fertility. This dataset was therefore instrumental in guiding functional and mechanistic studies (Chapters 3 and 4).

Additionally, AAGAG(n) RNA present in foci appears to derive from multiple genomic locations, such as 2R and X heterochromatin, in addition to unknown regions such as the foci observed in very early embryos. In L3, transcripts from these 2R and X heterochromatin regions appeared to all co-localize together, along with AAGAG(n) RNA foci, suggesting that heterochromatic transcripts cluster together into sub-heterochromatic foci.

Combined, these observations suggested that AAGAG(n) containing RNAs may be important for heterochromatin organization in somatic as well as germline cells. Also, the presence of AAGAG(n) RNAs in most cells at virtually every stage of development suggested a critical role of these RNAs for other functions. We therefore sought to identify functions of AAGAG(n) RNAs by decreasing these RNAs with RNAi and observing the resulting phenotypes, as described in Chapters 3 and 4.

Materials and Methods:
All images were acquired using a Ziess LSM10 confocal, using a 63X oil objective

<u>RNA-FISH methods:</u>

<u>RNA probe generation for RNA-FISH:</u> Repeat RNA probes were made by using oligo templates with antisense T3 promoters on the 3'ends (Table 3), hybridizing an oligo composed of sense T3 promoter so as to create a double stranded 3' end, transcribing with T3 RNA polymerase and UTP-digoxeginin, or in the case of RNA without Uracil, biotin-ATP. Oligos are listed in Table 3 and were ordered standard desalted from IDT. Reaction conditions were as follows: In a 40ul reaction, 1X RNApol reaction buffer (NEB cat. MO3782), 1mM each final concentration of ATP, GTP, CTP and 0.62 mM UTP, supplemented with 0.35mM final concentration of digoxegenin-11-UTP (Roche cat. 3359247910), 1 Unit Protector Rnase inhibitor (Roche cat. 3335402001), 5uM each of probe template and T3 promoter oligo (5'-AATTAACCCTCACTAAAG), and H20 to 40ul were combined. Reactions were heated to 80ºC, 3min to denature probes, iced 2 min., 4ul (or 200Units) of T3 RNA polymerase (NEB cat. M0378S) added and incubated at 37ºC overnight. 2ul Turbo DNAse (ThermoFisher Cat. AM2238) was then added to degrade DNA templates, incubated at 37ºC 15 min and the reaction stopped by adding 1.6ul of 500mM EDTA. Probes were then purified using standard sodium acetate/ethanol purification. Probe concentration was then assessed using Qubit RNA HS and stored at -80ºC.

<u>RNA-FISH buffers and reagents</u>

PBT solution: 1X PBS and 0.1% Tween-20
Western Blocking Reagent 10X: 10% casein in 100mM maleic acid; 150mM NaCl; pH 7.5. heat at 60ºC 1hr to
PBT block: 1:1 PBT/2X WBR
DEPC solution: 0.1% DEPC in PBT
Hybridization buffer: 50% formamide, 5X SSC, 100 µg/mL heparin, 100 µg/mL sonicated salmon sperm DNA, and 0.1% Tween-20. Filter through a 0.2-µ m filter and store at –20°C

For clarity, the methods for RNA-FISH probe hybridization and detection are numbered below and are referred to in subsequent chapters.

1.) <u>RNA-FISH probe hybridization and primary antibody incubation</u>: RNA probe hybridization for all tissues was carried about according to,[75] steps 10-17 under subheading #3. Samples were then washed one time with PBT then blocked in PBT block 1hr at room temperature. Samples were then processed for either 'non-TSA probe amplification' (2) or 'TSA amplification for RNA-FISH probe detection" (3)

2.) <u>Non-TSA probe detection for RNA-FISH</u>: For 'non TSA amplification' samples were incubated with either mouse anti-digoxigenin coupled to Cy5 or rabbit anti-digoxigenin (with fluorophore tagged secondary antibodies) in PBT block for 1hr at room temperature. After this, samples were incubated 6x's 10min each in PBT block, stained with DAPI 10min, washed 3xs 10min in PBS and mounted in Prolong-Gold antifade mountant (Thermofisher, cat. P36390).

3.) <u>TSA amplification for RNA-FISH probe detection</u>: For samples undergoing 'TSA amplification for RNA-FISH probe detection,' samples were incubated with primary antibody,(1/400 dilution of mouse anti-digoxegenin coupled to biotin (Jackson Immuno Research cat.200-062-156, lot. 123482)), with 0.2 U/ul protector RNAse inhibitor and incubated overnight at 4ºC with nutation. Next, samples were washed 6x's 10min each in PBT block. The next steps are essentially as per 'tyramide signal amplification kit' protocols (ThermoFisher) but with reagents purchased separately: In 75ul, embryos were then incubated with 1:100 streptavidin-HRP (Molecular probes, cat. S911) in PBT block for 1hour at room temperature. Samples were

then washed in 1:1 PBT/2XWBR 6x's 10min each, once with PBT, and 2x's with PBS. Samples were then incubated with Alexa 647 tyramide (TSA™ Reagent, Alexa Fluor® 647 Tyramide cat. T20951) according to Tyramide signal amplification protocols. Essentially, this consisted of adding 1ul of 30% hydrogen peroxide to 200ul tyramide signal kit amplification buffer, then diluting this solution 1/100 in tyramide signal amplification buffer for a final hydrogen peroxide concentration of 0.0015%. This solution was then added to the sample and incubated at room temperature for 1hr in the dark. Samples were then washed 1x with PBS for 10min, stained with DAPI for 10min, washed 4x's with PBS 10min each, and mounted in Prolong Gold Antifade mountant.

RNA-FISH of repeats in embryos: For RNA-FISH of repeat RNAs, 0-8hr Oregon R embryos were collected on apple juice plates, dechorionated and processed according to [75], as per headings 1 and 3 above, with the exception of using 37% formaldehyde stock from Sigma (cat. F1635-500ML). For RNA-FISH of the chr2 and X regions combined with AAGAG(n) RNA-FISH, the unique region probing used protocols 1 and 3, while AAGAG(n) RNA-FISH was performed without TSA amplification, using protocols 1 and 2.

Co-IF DNA/RNA-FISH of AAGAG(n) RNA in embryos: Co-IF RNA/DNA-FISH was performed essentially as described in [76], in which RNA-FISH is performed first, signal detected via tryamide signal amplification, RNAse treated to remove RNA and prevent DNA-FISH probes binding to RNA, and then DNA-FISH performed. Essentially, RNA-FISH was performed as above, but after tryamide signal amplification (Steps 1 and 3) and washing, samples were fixed in 4% formaldehyde. Samples were then washed 3x's in PBS 2min each. RNA was then removed under the following conditions: In a 50ul final volume, 1X Shortcut RNaseIII buffer (NEB cat. M0245S); 1.5ul RNASEIII (neb cat. MO245S), 100ug/ml RNaseA final concentration, and 1X MnCl2 (NEB cat. MO245S) and water to 50ul were added and samples incubated overnight at 4ºC. Samples were then rinsed 3x's in PBT 5min each, rinsed in 1:5, 1:1 and 5:1 mixtures of PBT: RNA hybridization solution for 15min each. Samples were then replaced with hybridization buffer and incubated 15 min. A DNA oligo probe to AAGAG(7) tagged with alexa5 was then diluted in hybridization buffer at 2.5ng/ul, denatured at 70ºC for 3min, then left on ice for 2min. Hybridization solution was removed from the embryos, probe solution added, and the sample denatured at 80ºC for 15min and hybridized overnight at 37ºC with nutation. Samples were then washed 2x's with pre-warmed 37ºC hybridization buffer 10min each. Samples were then washed in 3:1, 1:1, 1:3 hybridization buffer:PBT 15min each at 37ºC. Samples were then washed 2x's in PBT at room temperature 5 min each. Samples were then stained with DAPI 10min, washed once in PBS, and mounted in Pro-Long Gold Antifade mountant.

RNA-FISH of AAGAG(n) in larvae:
This protocol is essentially described in [77]:
A.) 3rd instar larvae were dissected in PBS supplemented with 0.2U/ul Protector RNase Inhibitor. The posterior end of the larvae was removed, then the remaining L3 inverted inside out. The inverted larvae were then transferred to ice cold PBS with 0.2U/ul RNAse inhibitor. Larvae were then fixed in PBT with 4% formaldehyde for 15min, washed 3x's, 5min each, with PBT. Larvae were then incubated with 0.1%(vol/vol) DEPC in PBT for 5min to remove endogenous RNAses. Samples were then rinsed 2x's with PBS. NOTE: Use of TSA amplification in L3 requires removal of endogenous peroxidases and requires the following protocol after DEPC treatment above and rinsing in PBS: In order to quench endogenous peroxidases, samples were incubated in 350 µl (enough to cover all tissue) of 3% H2O2 in PBS 15 min at room temperature and the tube kept open to prevent gas buildup. Samples were then rinsed 2x's with PBT 10min each.
B.) To all larval samples: Larvae were then permeabilized by incubation in 500ul cold 80% acetone in water at -20ºC 10min. Samples were then washed 2x's, 5min per wash with PBT, then post fixed with 4%

formaldehyde in PBT for 5min. Samples were then washed 5x's with PBT 2min each. Samples were then rinsed with 1:1 PBT/RNA hybridization solution, then with 100% RNA hybridization solution, and then stored in hybridization solution at -20ºC until needed. Samples were then processed according to RNA-FISH protocol (1) for probe hybridization and either (2) for non-TSA probe (AAGAG(n) RNA in larvae presented in Fig. 5 E and F, and Fig.6) or (3) for TSA amplification (AAGAG(n) RNA in Fig. 5B-D and unique region RNA in larvae presented in Fig. 6)

RNA-FISH of salivary glands

Larvae were grown, prepped and salivary glands processed as per [78], rehydrated in 95%, 70%, then 30% Ethanol 1min each, then washed 5min in PBT (0.1% Triton X-100 (TX100)). Slides were then fixed again in 3.7% formaldehyde in PBT (0.1% TX100), washed 2x's 3min each in PBT (0.1% TX100) and treated with 0.1% DEPC in PBT (0.1% TX100) and washed one time in PBT (0.1% TX100). The sample was then covered with pre-denatured hybridization solution, covered with a coverslip and incubated at 56ºC in a sealed hybridization chamber for 2 hours. The probe solution was then created by adding 100ng probe in 100ul hybridization solution, heating at 80ºC for 3min, and cooling on ice for 5min. This probe solution was then added to the sample, a coverslip added and sealed with rubber cement, and incubated overnight at 56ºC in a humid box. At 55ºC in a coplin jar, slides were then treated in 50% formamide/PBT (0.1% tx100) 1hr, 25 formamide/PBT (0.1% Tx100) 10min, then 3x's with PBT (0.1%Tx100) 10min each. Once at room temp, samples were blocked in 1:1 PBT/2xWBR, and processed as per larval RNA-FISH using protocol (2):'Non-TSA probe detection'

RNAse treatment of embryos

RNAseIII treatment of embryos after probe hybridization was carried out as follows: After hybridization and washing with PBS, samples were treated in 50ul final volume with 1X RNAseIII buffer, 1.5ul RNaseIII, and 1X $MnCL_2$ at 37ºC for 2 hours. Samples were then processed as per protocol (1) 'RNA-FISH probe hybridization and primary antibody incubation' and protocol (3) 'TSA amplification for RNA-FISH probe detection'

Northern blotting

Non-radioactive, denaturing northern blots were essentially carried out according to Chemiluminescent Nucleic Acid Detection Module Kit (Thermofisher). Essentially, purified RNA was denatured for 3 min at 70ºC in NorthernMax formaldehyde loading dye. Samples were then run on denaturing agarose gels with 6.9% formaldehyde in MOPS buffer. RNA was transferred to (+) charged nylon membranes in an electroblotter using 200mA for 30min. The membrane was then UV crosslinked, prehybridized with ULTRAhyb™ Ultrasensitive Hybridization Buffer (Thermofisher) at 68ºC for 30min. Biotinylated probes at a concentration of 30ng/ml were then added to UltraHyb buffer, and pre-hybridization solution replaced with solution containing probe and hybridized overnight at 68ºC. The next day, membranes were washed and processed according to Chemiluminescent Nuclei Acid Detection kit manual.

Kmer analysis of repeats

Kmer analsysis was performed using Kmc2. Download link: http://sun.aei.polsl.pl/REFRESH/index.php?page=projects&project=kmc&subpage=about and performed using the following commands: kmc_2 -k10 -m24 -t4 -r -cs100000000000 -ci1 -q30 <fastq file> <output_header> <working_dir>

CHiRP-MS

CHiRP-ms was essentially performed as per [63,64] but with the following modifications: Per sample, roughly 8.3 grams of third instar larvae,1,500 pmol antisense probe, 450ul streptavidin beads for RNA isolation and 50ul streptavidin beads for pre-clearing were used. Prior to mass-spec, enriched samples were then Benzonase treated using 10 units Benzonase per 37ug chromatin. For protein precipitation, proteins were precipitated using tri-chloroacetic acid, denatured with urea and DTT, alkylated with iodoacetamide, and quenched with DTT and finally trypsin digested using standard protocols. Proteins were then further concentrated using Omix C18 tips as per manufacturer's instructions.

CHirP-MS probes:

For AAGAG pulldown: 5'-CTCTTCTCTTCTCTTCTCTT/3BioTEG/   Scrambled: 5'-GAGCAATTAACAGCTCCTAA/3BioTEG/

Table 3. Oligos used to make repeat RNA probes

Repeat	Oligo with T3 antisense promoter
CAGC	CAGCCAGCCAGCCAGCCAGCCAGCTCTCCCTTTAGTGAGGGTTAATT
CCCA	CCCACCCACCCACCCACCCACCCACCCATCTCCCTTTAGTGAGGGTTAATT
CATTA	CATTACATTACATTACATTACATTATCTCCCTTTAGTGAGGGTTAATT
CGGAG	CGGAGCGGAGCGGAGCGGAGCGGAGTCTCCCTTTAGTGAGGGTTAATT
CGA	CGACGACGACGACGACGACGACGATCTCCCTTTAGTGAGGGTTAATT
CAACT	CAACTCAACTCAACTCAACTCAACTTCTCCCTTTAGTGAGGGTTAATT
CGAAG	CGAAGCGAAGCGAAGCGAAGCGAAGTCTCCCTTTAGTGAGGGTTAATT
CCCCAG	CCCCAGCCCCAGCCCCAGCCCCAGTCTCCCTTTAGTGAGGGTTAATT
CCGAG	CCGAGCCGAGCCGAGCCGAGCCGAGTCTCCCTTTAGTGAGGGTTAATT
CGGAA	CGGAACGGAACGGAACGGAACGGAATCTCCCTTTAGTGAGGGTTAATT
CACCC	CACCCCACCCCACCCCACCCCACCCTCTCCCTTTAGTGAGGGTTAATT
CTAGT	CTAGTCTAGTCTAGTCTAGTCTAGTTCTCCCTTTAGTGAGGGTTAATT
CATCG	CATCGCATCGCATCGCATCGCATCGTCTCCCTTTAGTGAGGGTTAATT
CAT	CATCATCATCATCATCATCATCATTCTCCCTTTAGTGAGGGTTAATT
CAAAC	CAAACCAAACCAAACCAAACCAAACTCTCCCTTTAGTGAGGGTTAATT
CGAAA	CGAAACGAAACGAAACGAAACGAAATCTCCCTTTAGTGAGGGTTAATT
CATAT	CATATCATATCATATCATATCATATTCTCCCTTTAGTGAGGGTTAATT
CCCCG	CCCCGCCCCGCCCCGCCCCGCCCCGTCTCCCTTTAGTGAGGGTTAATT
GAAA	GAAAGAAAGAAAGAAAGAAAGAAATCTCCCTTTAGTGAGGGTTAATT
CAGAA	CAGAACAGAACAGAACAGAACAGAATCTCCCTTTAGTGAGGGTTAATT
AAGGAG	aaggagaaggagaaggagaaggagaaggagaaggagTCTCCCTTTAGTGAGGGTTAATT
AAGAGG	aagaggaagaggaagaggaagaggaagaggaagaggTCTCCCTTTAGTGAGGGTTAATT
AAGAGAG	aagagagaagagagaagagagaagagagTCTCCCTTTAGTGAGGGTTAATT
AATAC	aatacaatacaatacaatacaatacaatacTCTCCCTTTAGTGAGGGTTAATT
AATAG	aatagaatagaatagaatagaatagTCTCCCTTTAGTGAGGGTTAATT
AATAGAC	aatagacaatagacaatagacaatagacTCTCCCTTTAGTGAGGGTTAATT
AAGAC	aagacaagacaagacaagacaagacaagacTCTCCCTTTAGTGAGGGTTAATT
AATAACATAG	aataacatagaataacatagaataacatagTCTCCCTTTAGTGAGGGTTAATT
AACAC	aacacaacacaacacaacaacacaacacaacacTCTCCCTTTAGTGAGGGTTAATT
dodeca	ACCGAGTACGGGACCGAGTACGGGTCTCCCTTTAGTGAGGGTTAATT
GTGTT	GTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTTCTCCCTTTAGTGAGGGTTAATT
GTAAT	GTAATGTAATGTAATGTAATGTAATGTAATTCTCCCTTTAGTGAGGGTTAATT

|	
GTATT	GTATTGTATTGTATTGTATTGTATTGTATTTCTCCCTTTAGTGAGGGTTAATT
TTAA	ttaattaattaattaattaattaattaattaaTCTCCCTTTAGTGAGGGTTAATT
CAAT	caatcaatcaatcaatcaatcaatcaatcaatTCTCCCTTTAGTGAGGGTTAATT
AAGAG	gagaagagaagagaagagaagagaagagaagagaaTCTCCCTTTAGTGAGGGTTAATT
CTCTT	CTCTTCTCTTCTCTTCTCTTCTCTTCTCTTTCTCCCTTTAGTGAGGGTTAATT
chrR:1825640-1825699 sense	GGC AGT TTA TGT GCG TAC AAC AAC AAC AGG ACT GCA AAC AAA ACA CGA AAC AGA TAT TTT TCT CCC TTT AGT GAG GGT TAA TT
chrR:1825640-1825699	AAA ATA TCT GTT TCG TGT TTT GTT TGC AGT CCT GTT GTT GTT GTA CGC ACA TAA ACT GCCTCT CCC TTT AGT GAG GGT TAA TT
chr2R:1,826,690-1,826,893 sense	tagacacatctacgaagacacaattctacaagaactaaacaacaaaaagtTCT CCC TTT AGT GAG GGT TAA TT
chr2R:1,826,690-1,826,893 as	ACT TTT TGT TGT TTA GTT CTT GTA GAA TTG TGT CTT CGT AGA TGT GTC TATCT CCC TTT AGT GAG GGT TAA TT
chrX:11,830,841-11,831,017 sense	ccaagcttcaggagaaagagaaagaagaaagctttaaacttaaggaaagagaagagagccttaggatTCT CCC TTT AGT GAG GGT TAA TT
chrX:11,830,841-11,831,017 as	CTA AGG CTC TCT TCT CTT TCC TTA AGT TTA AAG CTT TCT TCT TTC TCT TTC TCC TGA AGC TTG GCT T TCT CCC TTT AGT GAG GGT TAA TT
chrX:12,660,068-12,660,221 sense	tcgcacacacacacgcaacacttaggcacacataggagatagagtgagaTCT CCC TTT AGT GAG GGT TAA TT
chrX:12,660,068-12,660,221 as	cgacagacagtaaaattaaacaaactgcggacgcgtgtgacagaactaatccaacttTCT CCC TTT AGT GAG GGT TAA TT
chrX:22,453,001-22,453,133 sense	AAG TTG GAT TAG TTC TGT CAC ACG CGT CCG CAG TTT GTT TAA TTT TAC TGT CTG TCG TCT CCC TTT AGT GAG GGT TAA TT
chrX:22,453,001-22,453,133 as	AAG TTG GAT TAG TTC TGT CAC ACG CGT CCG CAG TTT GTT TAA TTT TAC TGT CTG TCG TCT CCC TTT AGT GAG GGT TAA TT

Antibodies used for IF:

antibody	Supplier; Cat. number	working concentration
rH3K9me3	Abcam; 8898	1/250
mH3K9me2	Active Motif; 39753	1/250
mH2A, H2B,H3,H4	Chemicon; MAB052	1/250

Chapter 3. AAGAG(n) RNA is necessary for male fertility

The abundance of proteins either required for fertility or present in testes that bind AAGAG(n) RNA via CHiRP-MS (Table 2) suggested that AAGAG(n) RNA satellites have a role in fertility. Also, characterization of AAGAG(n) satellite RNAs in larvae led to the observation that these RNAs are highly enriched in testes and ovaries. Furthermore, the Y-chromosome is necessary for male fertility, is entirely heterochromatic and contains several megabase blocks of AAGAG(n) DNA. [50,94] Together, these observations suggested that AAGAG(n) satellite RNAs are involved in male fertility and prompted us to investigate fertility in flies after AAGAG(n) RNAi knockdown.

Spermatogenesis in *Drosophila* occurs via a complex cascade of events facilitated by germline cells and somatic support cells. Essentially, spermatogenesis starts with a cluster of 'hub cells' at the apical end of the testes, which are part of a group of cells important for maintaining normal stem cells.[79] Surrounding the hub are two stem cell populations, germline stem cells (GSC) and somatic stem cells (SSC), wherein SSC's surround and support germ cells, together called a cyst, to promote differentiation.[80] The GSCs divide asymmetrically, producing gonadioblasts (GBs) that begin cell-differentiation. [81] These GB cyst cells then undergo four mitotic divisions with incomplete cytokinesis to produce a cycst of cells called spermatogonia, which then differentiate to produce 16 spermatocytes linked through cytoplasmic bridging. [82] Spermatocytes then undergo S phase, mature during a prolonged G2 phase, and increase substantially in volume.[81] At this stage, termed spermatocytes, the majority of testes specific gene expression occurs, wherein genes not required until later stages are translationally repressed.[82] Some minor transcription occurs in spermatids, and some of these genes are necessary for completion of spermatogenesis. [83] Matured spermatocytes then undergo two rounds of meiosis. After meiosis II, the last two steps, together referred to as spermiogenesis, function to process round spermatids into independent, condensed sperm nuclei. [84,85] In the first step, the 'round spermatids' resulting from complete meiosis II, undergo chromatin compaction, acrosome formation and flagellar elongation and then the second step, termed 'individualization', which allows for removal of cytoplasm and tight condensing and coiling of chromatin[85] (Fig7). Now mature, the sperm are released into the seminal vesicle.
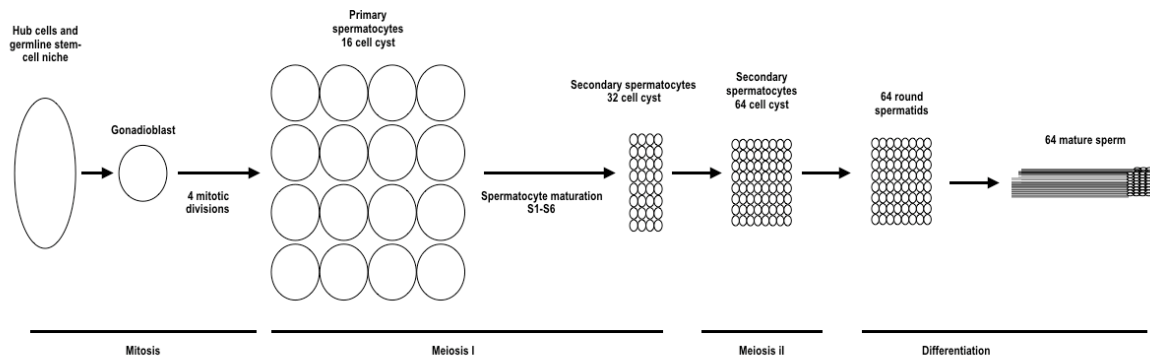
Figure 7. Schematic of spermatogenesis in *Drosophila*. At the apical end of the testes, hub cells, germline and somatic cyst cells coordinate to produce a gonadioblast. The gonadioblast then undergoes 4 mitotic divisions to produce 16 cell cyst spermatocytes. Spermatocytes then mature through several stages, increasing substantially in volume, before collapsing and undergoing two meiotic divisions. The resulting 64 round spermatids then undergo differentiation and individualization to produce mature sperm, that is then channeled to the seminiferous tubule for storage.

Comparatively, little is known about chromatin organization and how it relates to genome function in spermatogenesis. Two of the most striking differences between somatic cells and the male germline are the 200-fold compaction of mature sperm nuclei compared to spermatids and the resulting changes in chromosome volume, as well as the replacement of canonical histones with protamines. After meiosis II, a wave histone H4 acetylation occurs, followed by deposition of the transition protein Mst77f, [86] and finally removal of transition proteins and incorporation of protamines and prtl99c.[87,88] In the last steps of spermatogenesis, termed spermiogenesis, chromatin highly compacts, flagella elongate, cytoplasmic bridges dissolve and cytoplasmic contents are removed. [91,102] Individualization is a step in spermiogeneisis in which the individualization complex (IC) forms at and surrounds the end of spermatid nuclei, moves up the nuclei and flagellum, in turn removing cytoplasmic contents and creating individualized spermatozoon, each with its own plasma membrane. [85]

Heterochromatin proteins Su(var)3-9 and HP1a are essential for GSC maintenance, possibly by regulating levels of Bag of Marbles (bam) expression and entry into GB stage. [89] Effects of targeted knockdown of Su(var)3-9 and HP1a downstream of GSCs, such as in spermatocytes, have apparently not been established, although in the previous study Su(var)3-9 mutant spermatocytes exhibited large nuclei, suggesting that H3K9me2/3 is essential for spermatocyte heterochromatin organization. [89]

Most known mechanisms for *Drosophila* meiotic divisions have been worked out in female oocytes. (reviewed in [90]) Male *Drosophila* meiotic division is less well studied and distinctly different from female *Drosophila* and most other species, most notably that of the absence of recombination. [91] The mechanisms of homolog pairing in spermatocyte remains largely unknown, although it appears that euchromatic homology is the major determinant of chromosomal pairing for autosomes and nucleolus organizers (NORs) located within X heterochromatin for XY pairing. [91,92 93 94] There is also evidence of fiber-like connections between some heterochromatic chromosomes in meiosis such as the X and Y.[91 94] In short, chromosome organization during meiosis, at least for euchromatic regions and heterochromatic sex chromosomal regions, appears critical for proper spermatogenesis, but how this organization is important and the mechanisms involved remain a mystery.

Several genes important for individualization steps have been discovered, [85] including RNAi components such as Dcr and Ago2.[95] The Segregation Distorter (SD) gene complex, which affects the histone to protamine transition and subsequent individualization, is associated with chromosome 2 heterochromatin and Responder elements, [96] indicating that heterochromatin and possibly heterochromatic RNA is important for sperm individualization.

Similar to somatic cells, lncRNAs have critical roles in chromatin organization and cell function,[97] and the piRNA pathway has long been established as a necessity for preventing deleterious effects during spermatogenesis.[98] Regarding potential heterochromatic ncRNA spermatogenesis functions, the Y-chromosome in flies is entirely heterochromatic and composed essentially of transposons and simple tandem repeats. The Y contains at least six fertility factors, two of which encode dynein subunits, [50,94] and contains substantial amounts of AAGAG(n), AAGAC(n) and AATAT(n) repeats.[99] Furthermore, AAGAC(n) repeats are abundantly transcribed and are part of Y-loops, which are large lampbrush-like loops formed from Y-chromosome DNA in primary spermatocytes with unknown function.[100] More recently, antibodies to triplex nucleic acid structures were shown to stain Y-loops in primary spermatocytes,[100] suggesting that RNA is a critical component of Y-loops. Given that AAGAG(n) RNA directly binds proteins involved in spermatogenesis, I suspected that AAGAG(n) RNA is present in spermatocytes as a component of Y-loops and may somehow be linked to fertility factors discovered previously.

**AAGAG(n) RNA is enriched in primary spermatocytes**

To initially determine cytologically where AAGAG(n) is expressed in testes, I performed RNA–FISH in larval and adult testes. This demonstrated that AAGAG(n) RNA is highly abundant in all stages of spermatocyte nuclei but not in other testes cell types (Fig. 8A and data not shown for adult testes). Within spermatocyte nuclei, AAGAG(n) RNA is primarily enriched in regions surrounding DAPI-rich chromosome areas and is located closer to the nuclear periphery than the interior (Fig 8B). I also note the presence of RNA reads $\geq$ AAGAG(15) and CTCTT(15) (75 nucleotide) in testes RNA-seq,[55] validating our observation that long stretches of AAGAG(n) and CTCTT(n) (see later) are transcribed in testes.
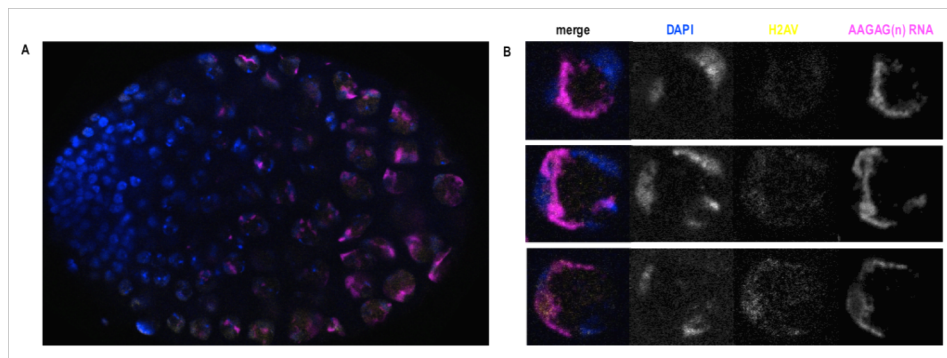
Fig. 8. AAGAG(n) RNA is enriched in primary spermatocytes (magenta). A) Slice of larval testes. H2AV is labeled in yellow and DNA (DAPI) in blue. B) Slices through an S5 spermatocyte.

AAGAG(n) RNAi with Bam-GAL4, but not other germline GAL4 drivers, results in complete male sterility

In order to determine if AAGAG(n) RNA exhibited functions in male testes, we sought to decrease levels via RNAi. This was accomplished by creating genomically inserted small-hairpin RNAs (shRNAs) driven by UAS-tagged, GAL4 driven small hairpin RNAs (shRNAs) to AAGAG(n) RNA, or in the case of a control, a scrambled RNA sequence, using genomic insertion of the pvalium20 vector used for the Transgenic RNAi project (TRiP) at Harvard.[101] Importantly, the scrambled shRNA sequence contained the same percentage of A's and G's as in the test shRNA, but in a random order, to identify phenotypes resulting specifically from AAGAG(n) knockdown versus off-target affects. In order to ascertain if AAGAG(n) RNA is required for normal spermatogenesis, I crossed UAS-shRNA constructs to flies containing Bag of marbles (Bam)-GAL4, which is one of the most effective male germline GAL4 RNAi drivers.[102] Bam is expressed in mitotic germline cysts, which includes late spermatogonia and early spermatocytes.[103] This cross resulted in a minimum 72% average reduction of AAGAG(n) RNA in S5 spermatocytes compared to scrambled control (Fig. 9). Strikingly, none of the AAGAG(n) RNAi with Bam-GAL4 (subsequently referred to regarding testes as AAGAG(n) RNAi or AAGAG(n) knockdown) males were fertile, whereas roughly 100% of scrambled RNAi males, as well as AAGAG(n) RNAi females, were predominantly fertile (Table 4). These results demonstrate that AAGAG(n) RNA is necessary for male fertility and warranted further investigation.
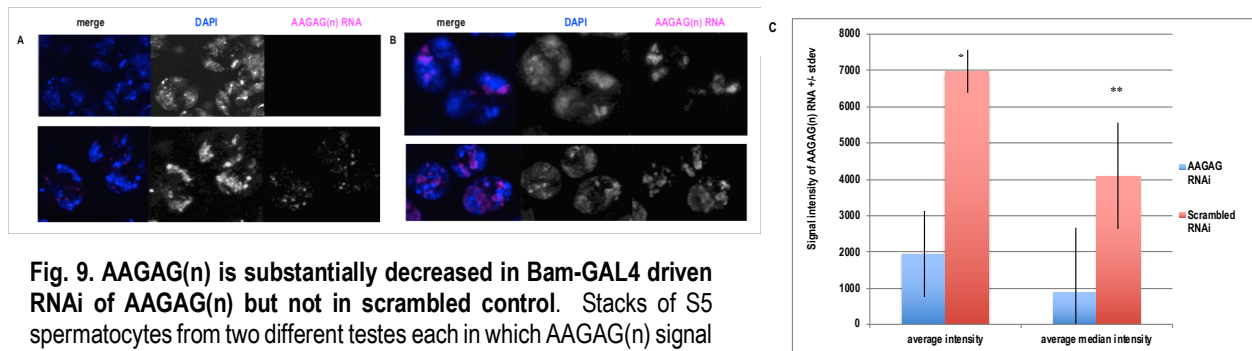
Fig. 9. AAGAG(n) is substantially decreased in Bam-GAL4 driven RNAi of AAGAG(n) but not in scrambled control. Stacks of S5 spermatocytes from two different testes each in which AAGAG(n) signal was imaged with same laser intensities for each genotype. A) Examples of AAGAG(n) foci in S5 spermatocytes in Bam-GAL4 driven AAGAG(n) RNAi and B) Scrambled RNAi. C) Average intensities and average median intensities of AAGAG(n) RNA from stacked images of S5 spermatocytes in Bam-Gal4 AAGAG(n) RNAi and scrambled control. *p=0.001; **p=0.002.

Table 4. Fertility of AAGAG(n) and scrambled RNAi with Bam-GAL4

genotype	% male fertile	+/-stdev	% female fertile	+/-stdev
AAGAG(n) RNAi	0	0	97	2.8
Scramble RNAi	96	3.8	95	5.9

It is possible that undetectable levels of AAGAG(n) RNA existed in cells preceding spermatocyte stages that could have a role in fertility. Therefore, I performed RNAi to AAGAG(n) and scrambled control using GAL4 drivers active in prior stages (ie. somatic cells, GSC's and early germline cysts) (Table 5). [71] However, none resulted in significant levels of sterility (Table 5), demonstrating that AAGAG(n) RNA is required for male fertility in mitotic germline cysts and/or spermatocytes, the stage in which it is first detected. Thus, if present at un-detectable levels in stages prior to spermatocytes, it is not required for fertility.

Table 5: Male Sterility in AAGAG(n) RNAi with different male germline GAL4 drivers

Germline GAL4 RNAi driver	Expression location[71]	% fertile	stdev.
Fascillin	Hub	94.67	16.68
PTC	Soma- CySCs and cyst cells	89.80	4.73
Traffic Jam	Soma- Hub and CySCs	97.22	3.93
Dpp1	Soma- CySCs and early cyst cells	95.83	5.89
Nanos	Germline- GSCs and early germline	83.01	4.74

AAGAG(n) knockdown with Bam-GAL4 prevents mature sperm formation

We were intrigued by these results and wanted to then know what roles AAGAG RNA plays in spermatogenesis. This was addressed by performing more detailed analyses of defective processes in testes of sterile AAGAG(n) RNAi flies. Considering that AAGAG(n) RNA only appears in GB's and

spermatocytes, we theorized that abolishment of AAGAG(n) in these cells with Bam-GAL4 RNAi may prevent spermatocyte maturation to cause a lack of progression to the first meiotic division. Intriguingly, spermatocyte maturation, as well as meiosis I and II appeared normal, in the sense that normal numbers of cysts for each stage were present (data not shown), suggesting that any major visible defects occur after meiosis II.

I then assayed for the production of motile sperm, to determine if fertility defects in the Bam-GAL4 driven AAGAG(n) knockdown males are due to defects during or after fertilization in the embryo, as observed after HP1e depletion[104]. Six-day old Bam-Gal4 AAGAG(n) knockdown testes did not exhibit motile sperm when testes contents were released and imaged with phase contrast microscopy, whereas scrambled knockdown sperm contained many motile sperm. Furthermore, upon fixed tissue staining, I documented a complete absence of mature sperm in the seminiferous tubule in 6-day old males (Fig. 10A), but not in scrambled controls (Fig. 10B). In the basal end of the testes, adjacent to the seminiferous tubules, there is an accumulation of 'debris' in AAGAG(n) knockdown (Fig. 10C) not present in scrambled control (Fig. 10D). Together, this suggested that the major defect that associates with AAGAG(n) RNAi in male testes manifests after meiosis II, but before processes required for sperm motility and movement of mature sperm into the seminiferous tubules.
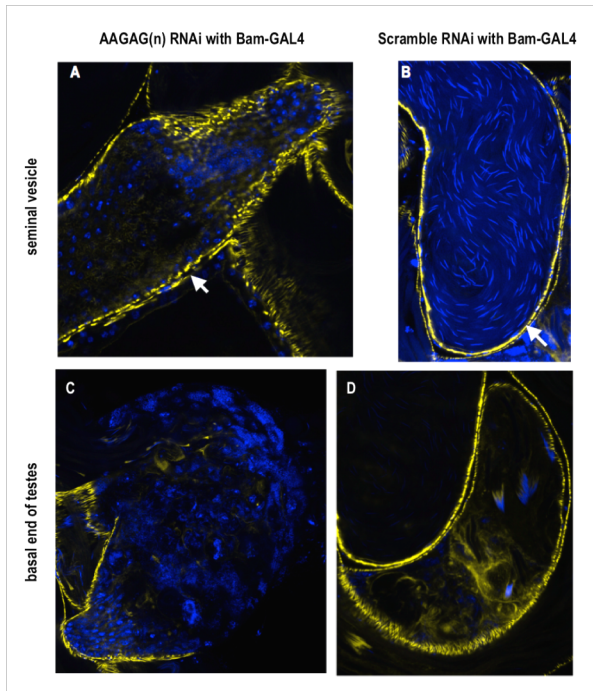
Fig. 10. Sperm is absent in the seminiferous tubules in 6-day old Bam-GAL4 driven AAGAG(n) RNAi testes. Shown are slices of testes stained with DAPI (Blue) and phalloidin, which stains actin (yellow). Arrows point to seminiferous tubules. A) Seminiferous tubules of AAGAG(n) RNAi. B) Seminiferous tubules of 6-day old scrambled RNAi. Note the thin, elongated DAPI signal which represents mature sperm DNA. C) Basal end of AAGAG(n) RNAi testes that have accumulated DNA rich 'debris'. D) Basal end of scrambled RNAi tests. Note the presence of morphologically normal bundles of individualizing sperm DNA and absence of significant amounts of 'debris'

Based on the timing of visible spermatogenesis defects in AAGAG RNAi and the importance of the histone to protamine transition in producing mature sperm, I suspected that these testes might exhibit defects in

these processes. However, similar levels and distributions of H3K4acetylation, transition protein mst77F, as well as protamine B were detected in AAGAG(n) knockdown and scrambled control testes (Fig 11 A-C), suggesting that AAGAG(n) RNA does not affect H3K4 acetylation or the histone to protamine exchange. However, this does not rule out other defects in chromatin organization at this stage that can occur independently of the histone to protamine transition, as observed after depletion of protamine-like 99c (prtl99C).[88]
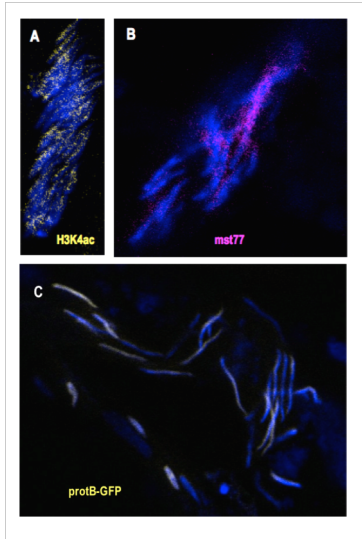
Fig. 11. Histone to protamine exchange is largely not affected in AAGAG(n) RNAi Bam-GAL4 testes. Shown are projections 0-6 hour testes labelled with DAPI (Blue) and stained with the following proteins: A) H3K4ac in canoe stage spermatids. B) mst77 in canoe stage spermatids. C) GFP from protB-GFP in individualizing sperm

Interestingly, in AAGAG(n) knockdown testes, 100% of 0-6 hour old flies exhibit regions of long, broad DAPI staining in regions not present in the scrambled control (Fig. 12A). To our knowledge, this phenotype has not been described in the literature. It is possible that these are decondensed individualizing sperm, considering that they are elongated, broad, striated in nature and appear to be enriched in regions that predominantly exhibit individualizing sperm. However, these DAPI broad regions do not stain for H2Av, mst77F, or protB (Fig 12B and data not shown), which suggests they are not sperm nuclei defective in individualization, but instead may represent uncharacterized waste components associated with other dying cells.
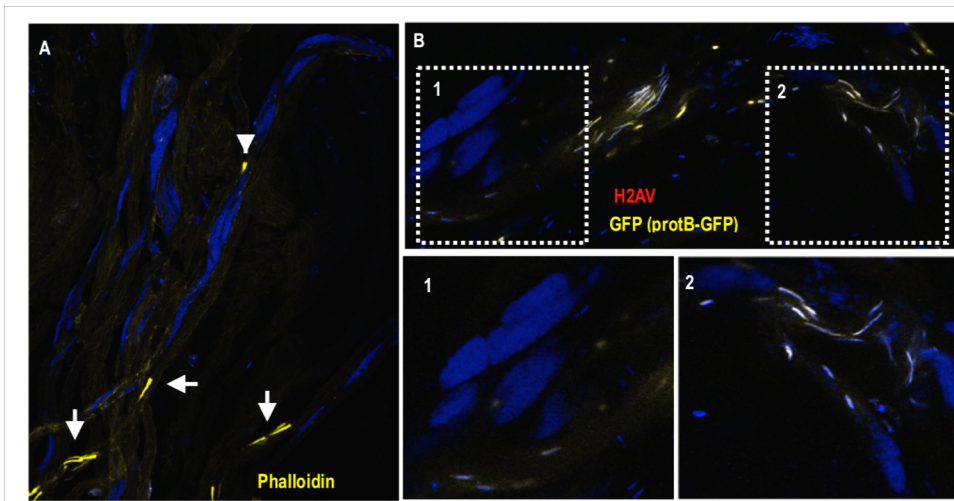
Fig. 12. AAGAG(n) RNAi with Bam-GAL4 0-6hr testes exhibit long, broad regions of DNA. A) Example of long, broad DAPI (Blue) staining structures that appear to be de-condensed individualizing sperm nuclei. Counterstained with phalloidin (yellow). Arrows point to investment cones. B) These broad DNA structures do not exhibit H2AV or prot-B. Below are zoomed imaging on these DNA-broad structures.

In older male flies, such as in 4-7 day old AAGAG(n) RNAi flies, these elongated DAPI poor regions are not present, although DAPI rich and morphologically different 'debris' accumulates at the basal end (Fig.10). Also, in these older AAGAG(n) RNAi testes, individualizing sperm are largely disorganized, have 'lagging

sperm nuclei' and are often composed of less than 64 sperm nuclei within bundles (Fig. 13A). Furthermore, the most mature forms of sperm nuclei in the knockdown that form exhibit a 'kinky' appearance (Fig. 13B), and appear more de-condensed than the normal tight, compact mature sperm nuclei in the seminiferous tubules of the scrambled control (Fig. 13C).
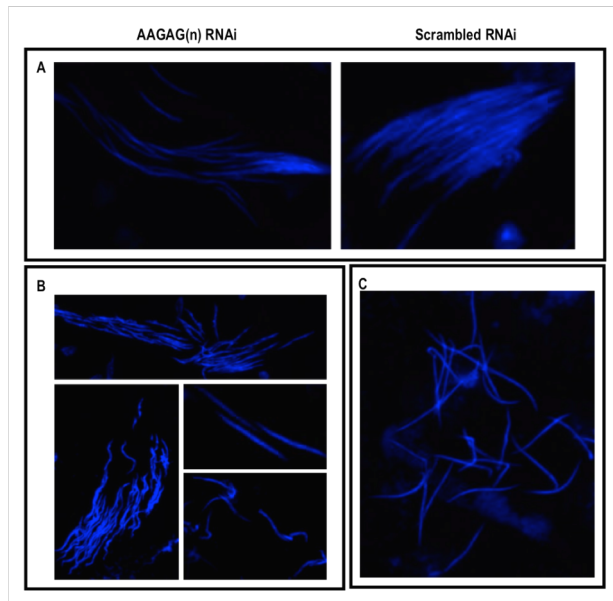
Fig. 13. Examples of individualizing and mature sperm nuclei in 4-7 day old AAGAG(n) and Scrambled RNAi with Bam-GAL4 testes. Shown are stacked images stained with DAPI (blue) A) Early stages of sperm individualization immediately after late canoe stage. Note the tight, organized sperm nuclei bundles in the scrambled RNAi. B) Disorganized, individualizing sperm nuclei that represent the most mature sperm nuclei present in AAGAG(n) RNAi. C) Normal mature sperm found just prior to and in the seminiferous tubules in scrambled RNAi.

The observation that the AAGAG(n) RNAi testes have significantly less than expected sperm nuclei prompted me to determine if there was loss of sperm nuclei during individualization steps. I therefore determined the number of mature sperm nuclei (during and after individualization) per bundle, relative to those present as round spermatids (prior to individualization), in 4-7 day old testes. Although there is a roughly 14% decrease in the most mature stages of sperm nuclei compared to round spermatids in the AAGAG(n) RNAi, I observed a similar decrease in scrambled RNAi (Fig 14), suggesting that loss of individualizing sperm nuclei is not the major driver for fertility defects after depletion of AAGAG(n) RNA. In the testes of old AAGAG(n) RNAi testes, the average number of round spermatids is less than the expected 64 average, suggesting that perhaps slight defects during meiosis occur (Fig 14). It is also possible that the decondensed, DAPI broad structures seen in AAGAG(n) RNAi testes of 0-6 hours are round spermatids that fail to undergo proper nuclear shaping and chromatin condensation, and that the debris found in older AAGAG(n) RNAi testes is an accumulation of defective round spermatids. Regardless, I conclude that AAGAG(n) RNA is required for mature sperm formation, and its depletion leads to disorganized and abnormal individualizing sperm.
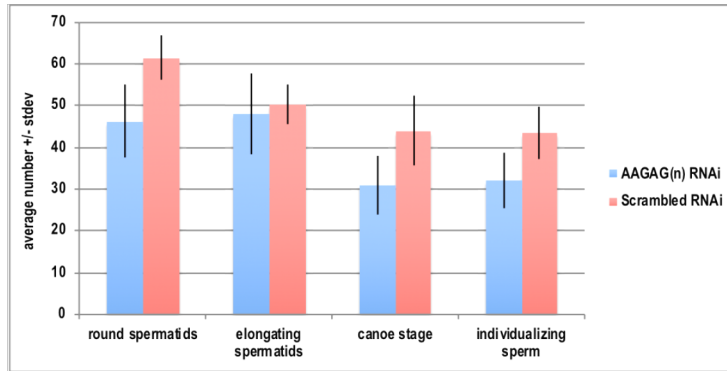
Fig. 14. Counts of nuclei per bundle of elongating and individualizing sperm nuclei in 4-7 day old AAGAG(n) or scrambled RNAi testes with Bam-GAL4. Shown is a graph of average number of round spermatids, elongating leaf stage spermatids, canoe stage and late stage individualizing sperm nuclei from a minimum of four testes from each genotype.

AAGAG(n) containing RNA in spermatocytes does not originate from the Y-chromosome and does not drastically affect Y-loop organization

Considering the importance of the Y-chromosome in male fertility, almost complete heterochromatinization of the Y, predominance of AAGAG(n) DNA on the Y, presence of Y-loops in spermatocytes as well as transcription of AAGAC(n),[100][105] I suspected that the AAGAG(n) RNA transcripts detected in spermatocytes are transcribed from the Y. To test this idea, I created Y-chromosome deficient testes (XO), which produce spermatocytes but are defective in later spermatogenesis steps and performed RNA-FISH to AAGAG(n) RNA in 0-6 hour testes. Contrary to our hypothesis, robust AAGAG(n) RNA is present in XO spermatocytes (Fig. 15), and at slightly higher average intensity levels than in scrambled RNAi (not shown). This demonstrates that at least the majority of AAGAG(n) RNA detected in spermatocytes is not generated from the Y-chromosome. It was still plausible that AAGAG(n) RNA affected Y-loop organization, especially considering that AAGAG(n) RNA is present primarily in the DAPI poor regions of spermatocytes, which is the same general location of Y-loops and Y-loop proteins. [105,106] To test if this was the case, I stained AAGAG(n) knockdown and scrambled control testes with an antibody to X4, a Y-loop binding protein.[106] However, the large-scale organization of X4 is essentially unaffected in AAGAG(n) knockdown testes (Fig. 15), demonstrating that AAGAG(n) RNA abolishment does not significantly affect organization of this Y-loop protein. I conclude that AAGAG(n) RNA does not originate from the Y, and most likely does not affect Y-loop organization.
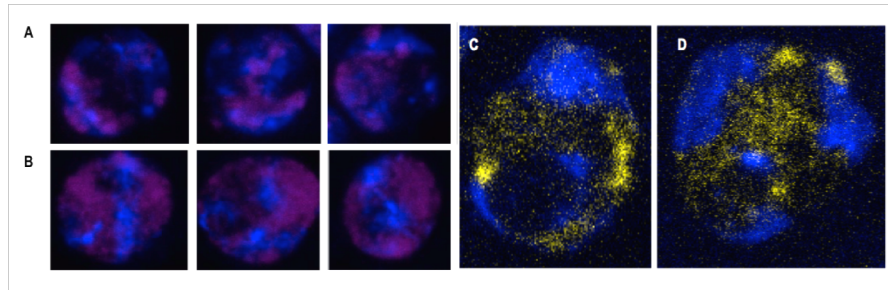
Fig. 15. AAGAG(n) RNA is present in XO teste spermatocytes and does not affect Y-loop protein organization. A) and B) Shown are projections of stacks in which AAGAG(n) signal (magenta) was imaged at same laser intensities in both genotypes. DNA is stained with DAPI (Blue). A) AAGAG(n) RNA in scrambled RNAi with Bam-GAL4. B) AAGAG(n) RNA in XO testes lacking the Y-chromosome. C) and D) Slices of L3 S5 spermatocytes (not scaled with A and B) stained with X4 antibody (yellow) and DAPI (Blue) in Bam-GAL4 RNAi. C) AAGAG(n) RNAi B) Scrambled RNAi

Heterochromatic regions are transcribed in testes

Similar to AAGAG(n) RNA in somatic tissue, we wished to determine the genomic source of AAGAG(n) RNA in testes. However, 'phrap' analysis to identify unique regions transcribed in concert with AAGAG(n) RNA in testes, as was used for somatic cells to identify sources of AAGAG(n) RNA, did not yield any hits. Regardless, I performed RNA-FISH to somatic AAGAG(n) sources, such as chr2R:1826680-1826741; chrX:12,660,068-12,660,221; chrX:12660096-12660145, in addition to CTCTT(n) RNA, to see if these regions were also transcribed in testes and co-localized with AAGAG(n) RNA. Surprisingly, all of these regions and in both strands exhibit signal in the DAPI poor region of spermatocytes. (Fig. 16). I conclude that similar to somatic cells, different satellite regions containing AAGAG(n) RNA are transcribed in spermatocytes
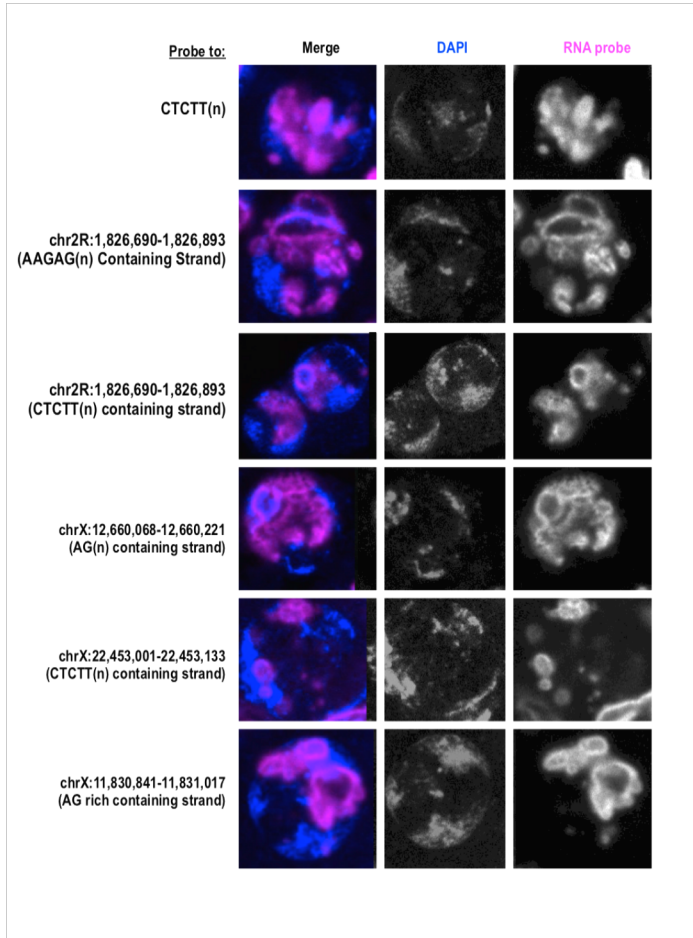
Fig. 16. Regions adjacent to blocks of AAGAG(n) or AG(n), in addition to CTCTT(n) are transcribed in spermatocytes. Shown are slices of spermatocytes probed for regions adjacent to AAGAG(n) or AG(n) blocks, as done for somatic cells. For each X chromosome region, the other strand is also present in spermatocytes but not shown.

Conclusions/Future Directions

Here, I demonstrate that ncRNA(s) containing AAGAG(n) are necessary for spermatogenesis and that abolishment in germline stem cells such as GBs and spermatocytes, but not somatic stem cells, prevents mature sperm formation as demonstrated by lack of sperm nuclei in the seminiferous tubule. Contrary to our expectations, these RNAs do not come from the Y and do not drastically affect organization of one Y-loop protein. Surprisingly, we also document robust expression of heterochromatic regions next to AAGAG(n) or AG(n) blocks from the 2R and X chromosomes, suggesting a possible link between heterochromatic ncRNAs and spermatogenesis.

A plethora of studies over the years have identified proteins that are important for *Drosophila* spermatogenesis, and recently lncRNAs necessary for male fertility were described.[97] This is to our knowledge, however, the first report of tandemly repetitive ncRNAs, very likely derived from satellite DNA, that are necessary for male fertility. It is especially interesting that decreasing AAGAG(n) RNA in primary spermatocytes causes 100% male sterility by affecting later stages of spermiogenesis, perhaps through chromatin organization defects. Whether or not it is the tandem repeat AAGAG(n) RNA alone, other portions of ncRNA attached to AAGAG(n) RNA, or an entire ncRNA transcript containing AAGAG(n) RNA that is important for male fertility is currently unknown. We suspect that is AAGAG(n) RNA itself that is necessary and will test whether or not over expression of 50 and 200 base repeats of AAGAG(n) RNA are sufficient to rescue sterility seen in AAGAG(n) RNA knockdown testes. If rescue of spermatogenesis occurs, this would suggest that in fact the AAGAG(n) RNA component of the functional satellite RNA(s) is the critical factor for this satellite function in spermatogenesis. This would also demonstrate that male sterility in the AAGAG(n) knockdown testes results from removal of repetitive satellite RNA and not the

28

result of unexpected but possible off-target RNAi effects. We stress, however, that off-target effects of AAGAG(n) RNAi are likely to be minimal if not non-existent, considering that our scrambled control shRNA is composed of the same percentage of A's and G's in the passenger strand as in the AAGAG shRNA, the few mRNAs with AAGAG RNA are not decreased in somatic cells with actin-GAL4 driven RNAi (data not shown), mRNAs containing more than 3 AAGAG(n) tandem repeats were not found via kmer analysis in testes (data not shown) and robust knockdown of AAGAG(n) RNA is achieved with AAGAG(n) RNAi but not scrambled RNAi in spermatocytes (Fig. 9).

It is intriguing that defects in spermatogenesis after AAGAG(n) knockdown do not arise until later in spermiogenesis, many steps away from GB's and maturing spermatocytes, when the transcript is first detected. This suggests that these RNAs somehow prime the spermatocyte for a factor that is only necessary later, when sperm nuclei are undergoing shaping and/or individualization. It is also plausible that these RNAs are necessary only for some function within the spermatocytes, and that spermiogenesis serves as a checkpoint to prevent defective cells from organizing into mature sperm. This scenario was suggested in a case in which improperly disjoined X and Y chromosomes lacking nucleolus organizers (NORs) were defective in meiotic pairing.[91][93] The decreases in number of round spermatids per cyst/bundle in AAGAG(n) RNAi testes compared to scrambled control does suggest at least a slight defect in meiosis, and combined with the hypothesis that spermiogenesis is a checkpoint for poorly differentiated sperm chromatin suggests that AAGAG(n) RNAs have a role in chromosome organization. However, the gross organization and number of DAPI rich regions in spermatocytes, number of nuclei in meiosis I and II, as well as general morphology of round spermatids appeared normal, suggesting that AAGAG(n) RNA does not drastically affect meiotic chromosome organization.

We did initially attempt to address AAGAG(n) RNA mediated chromosomal organization in spermatocytes and meiosis using heterochromatic DNA probes but were unsuccessful due to technical challenges (data not shown). It is possible that using a series of dead Cas9 constructs targeted to a series of heterochromatic and euchromatic sequences could address this issue, at least in spermatocytes. The labelling of chromatin regions undergoing later stages of meiosis with dead Cas9, however, seems unlikely with current technology, considering that compacted meiotic DNA is largely inaccessible. Furthermore, it is possible that AAGAG(n) RNA somehow affects components that are critical for proper chromosome segregation and spermiogenesis that we have yet to detect with current technologies.

I identified several proteins important for spermatogenesis and/or enriched in testes that bind AAGAG(n) RNA directly and suspect that AAGAG(n) RNA may facilitate localization of these proteins to accomplish spermatogenesis. AAGAG(n) RNA for instance, binds Tudor-SN (Table 2) and this protein is necessary for spermatogenesis by antagonizing PIWI, while acting synergistically with PIWI in transposon silencing.[70] AAGAG(n) RNA also binds Eya and Heat shock protein 83 (Hsp83), both of which are required for male fertility, in addition to Dodeca binding protein 1 (Dp1 or Ddp1), Heterogeneous nuclear ribonucleoprotein at 27C (Hrb27C), and Heat shock protein 26 (Hsp26), which have not been tested for roles in male fertility. Initial imaging, however, of lines in which these proteins are GFP tagged indicated that all are expressed in testes, suggesting potential roles in male fertility (data not shown). I additionally assayed testes for expression of GFP-tagged proteins that associate with heterochromatin, such as HP1a, Heterogeneous nuclear ribonucleoprotein at 87F (hrb87f), as well as HP1c, which associates with somatic euchromatin, and found all are expressed in testes and are nuclear localized (data not shown). Currently, we are determining whether AAGAG(n) knockdown affects localizations of these proteins in testes, and if so it would suggest a direct role of AAGAG(n) RNA in organizing proteins for proper completion of spermatogenesis.

Besides having a direct role in chromatin organization in spermatogenesis, it is also possible that AAGAG(n) RNA affects expression and/or translation of critical spermatogenesis proteins. For instance, it is possible that satellite RNAs containing AAGAG(n) are processed into siRNAs or piRNAs that affect

expression of nearby genes or TEs via chromatin silencing. In line with this notion, there is evidence of ncRNAs affecting silencing of RNAs in *trans* to achieve spermatogenesis, notably that of the Suppressor of Stellate locus (Su(Ste)) locus from the Y[107], so it is also possible that AAGAG(n) satellite RNAs have *trans* silencing roles as well. It is also possible that these satellite RNAs somehow are involved in antisense translational repression of some proteins by unknown mechanisms. Future experiments will thus address this issue with RNA-seq in AAGAG(n) knockdown, AAGAG(n) overexpression, as well as control testes, to see if any genes are specifically regulated by AAGAG(n) RNA.

Unexpectedly, the majority of AAGAG(n) RNA in spermatocytes does not come from the Y, while heterochromatic regions next to AAGAG or AG blocks from 2R and X chromosomes are transcribed in both strands, along with CTCTT(n) RNA, suggesting bi-directional transcription through these regions. This is intriguing, considering that combined these RNAs are found all throughout the DAPI poor regions in spermatocytes and in the 2R case are produced from an autosome. It is unclear currently what this means, but it could suggest that much of heterochromatin in spermatocytes is transcribed. It will be interesting to determine if transcription of heterochromatic regions away from AAGAG(n) regions occurs in spermatocytes, and which, if any, functions these transcripts have. If significant amounts of heterochromatin are transcribed and localized in the DAPI poor regions where we see the AAGAG(n) containing transcripts, it may suggest that these RNAs serve to localize and sequester proteins necessary for later processes, similar to P-bodies, stress granules and omega speckles. It is also possible that the heterochromatic transcripts detected are processed into smaller RNAs, such as small interfering RNAs (siRNAs) or piRNAs, but attempts to isolate enough RNA from germline cells of testes for northern blotting and subsequent transcript size determination was unsuccessful. New methods for ex-vivo culturing of germline cysts were developed [108] and therefore future experiments will determine sizes of these transcripts via northern blotting of cell-culture derived RNAs. piRNAs are predominantly found in germ cells, and one major established role is transposon and repeat silencing.[109] In *Drosophila*, the majority of PIWI interacting RNA (piRNA) studies have focused on the female germline, although several critical studies showed necessary roles of either piRNAs or PIWI class proteins in male fertility and spermatogenesis. [110] [111] Considering that expression of AAGAG(n) RNA is necessary for spermatogenesis would seem to suggest that, if AAGAG(n) RNA does originate from piRNA clusters, that it has roles beyond that of simply silencing similar repeats. It is also possible (and likely given that several regions next to AAGAG(n) or AG(n) blocks are transcribed) that AAGAG(n) containing RNAs are transcribed from regions that are differentially processed into piRNAs, siRNAs or neither, all of which with different functions. Regardless of the mechanisms of satellite RNA mediated spermatogenesis, this work demonstrates that it is the satellite AAGAG(n) RNA containing transcripts themselves that contribute to spermatogenesis and not simply mechanisms that process these RNAs.

Figure 17. Schematic of AAGAG(n) RNA expression in testes and effects of AAGAG(n) knockdown. A) AAGAG(n) RNA (magenta) is expressed in gonadioblasts and primary spermatocytes but not in other testes cell types. B) Decreasing AAGAG(n) via RNAi in gonadioblasts and primary spermatocytes prevents mature sperm formation but does not substantially affect processes prior to differentiation steps.

Materials and Methods

Fly lines used:

Stock name or genotype	Obtained from	Description
y[1] v[1]:UAS-AAGAGshRNA::	Rainbow Transgenic Flies, Inc	Expresses shRNA under UAS promoter targeting AAGAG(n)
y[1] v[1]:UAS-scramble shRNA:	Rainbow Transgenic Flies, Inc	Expresses shRNA under UAS promoter targeting a random sequence
y[1] w[67c23]; P{w[+mC]=dpp-GAL4.PS}6A/TM3, Ser[1]	Bloomington: 7007	Dpp-GAL4
C(1;Y)1, y[1] w[A738]: y[+]/0 & C(1)RM, y[1] v[1]/0	Bloomington:2494	XO
w[*]; P{w[+mC]=protamineB-eGFP}2/CyO; P{w[+mC]=dj-GFP.S}3/TM3, Sb[1]	Bloomington:58406	ProtB-GFP; Dj-GFP
y[*] w[*]; P{w[+mW.hs]=GawB}NP1233 / CyO, P{w[-]=UAS-lacZ.UW14}UW14	Kyoto: 103948	Fascillin-GALO4
y[*] w[*]; P{w[+mW.hs]=GawB}NP1624 / CyO, P{w[-]=UAS-lacZ.UW14}UW14	Kyoto:104055	Traffic Jam-GAL4
w[*]; P{w[+mW.hs]=GawB}ptc[559.1]	kyoto: 103948	PTC-GAL4
'+; +; nanos-Gal4, dcr2-UAS / TM3 sb	Unknown	Nanos-GAL4
w;;bamGAL4, UAS-dicer2	Unknown	Bam-GAL4

Fertility assay:

Flies containing shRNA to AAGAG(n) or scrambled control were mated to different testes GAL4 drivers (see Table 5) at 25ºC. One male progeny was then allowed to mate with two female Oregon R virgins for 10 days at 25ºC. Male flies were counted as sterile if, after 10 days, the male and at least one female were still alive and no progeny present. Female fertility was calculated as above, with one female RNAi and two OR males.

For analysis of AAGAG(n) RNA levels in male testes without a Y-chromosome, y(1)w(1):: males were mated to C(1)RM, y[1] v[1]/0 females (Bloomington stock # 2494)

RNA-FISH in testes

For RNA-FISH in adult testes, flies were mated at 29ºC and F1 progeny grown at 29ºC. For analysis of 6 day old male progeny, flies were kept with female F1 progeny. Flies were then anaesthetized with C02, testes removed with forceps and placed in 7ul of PBS on (+) charged slides, a RainX-treated coverslip placed over the testes and snap frozen in LiN2. The coverslip was then immediately popped off with a razor blade and slides stored at -80ºC until needed. When ready to process, slides were fixed for 20min in 4% formaldehyde in PBT, washed three times, 5 min each wash, in PBT. Samples were then incubated in 80% cold acetone in PBT for 10min at -20ºC and processed as per RNA-FISH for 'all larval samples" and either (2) for AAGAG detection without TSA amplification or (3) for detection of all other regions with 'TSA amplification'

IF without RNA-FISH in testes

For IF of testes without RNA-FISH, flies were mated 29ºC and F1 progeny grown at 29ºC, then processed according to section 2.2.2 'Formaldehyde fixation for confocal microscopy'[112]

Table 6. Antibody concentrations and sources

antibody	concentration	source

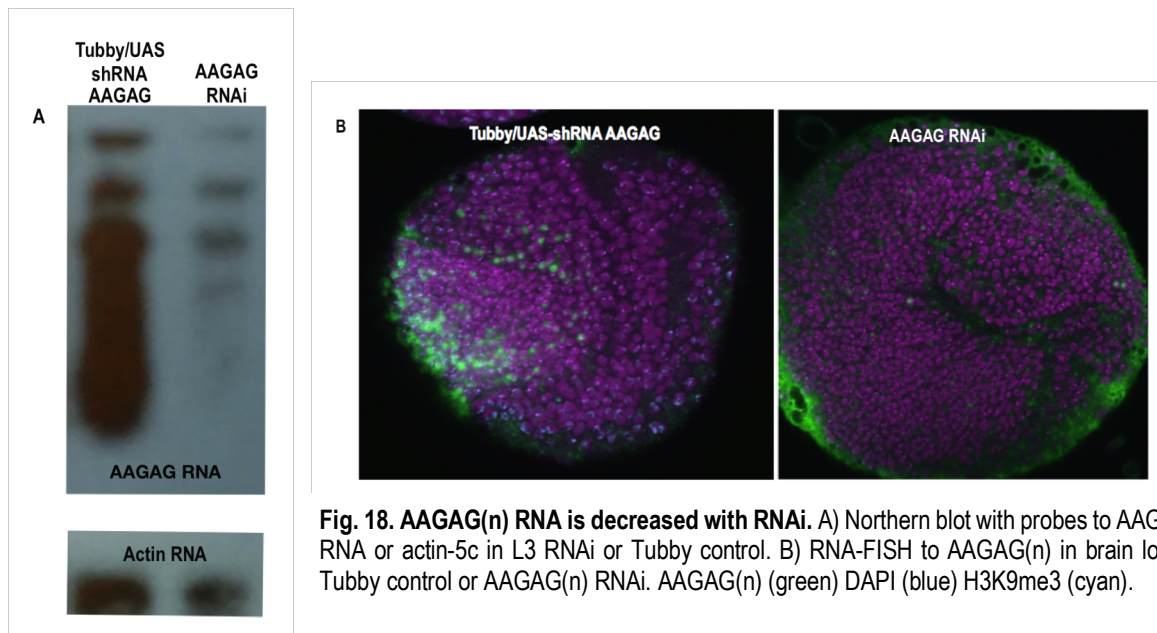| rH2AV	1/100	Lake placid AM318; 9751
gGFP	1/500	Rockalnd 121600-101-215
rH4acetyl	1/200	Millipore 06-598
Mst77F	1/100	Elaine Dunleavey, Phd; NUI Galway, Ireland
X4	1/50	Harald Saumweber, Phd; Humbolt University, Berlin

Chapter 4: AAGAG(n) Satellite RNAs are necessary for viability but do not affect established heterochromatin organization

AAGAG(n) RNA is decreased with targeted RNAi

We next sought to abolish AAGAG(n) RNAs in somatic cells in order to facilitate functional analysis. Elucidating functions and mechanisms of RNA or proteins typically occurs first in cell culture due to ease of perturbation of targets, then followed by validation in whole animals. However, AAGAG(n) RNAs were not expressed in model fly tissue culture cells, such as S2, Kc or BG3 cell lines, guiding us to begin functional analysis in whole animals.

Our first approach at AAGAG(n) RNA knockdown in somatic cells utilized antisense 'LNA GapmeRs', which essentially are highly modified antisense oligos that bind to an RNA and stimulate RNAseH cleavage of the RNA. However, these were in-effective in our hands, in the sense that they produced high toxicity in the early embryo after injection and did not result in a decrease in AAGAG(n) RNA (data not shown). We then wished to assess AAGAG(n) knockdown efficiency in the soma with RNAi by expressing the shRNAs using an actin-GAL4 (act-GAL4) driver, which should be expressed ubiquitously in all cells. This was accomplished by crossing UAS-shRNA females to act-GAL4/Tubby males and collecting L3 F1 progeny and assessing AAGAG(n) RNA levels. This approach resulted in an overall 55% decrease in AAGAG(n) RNA compared to Tubby control (control without act-GAL4 and therefore devoid of AAGAG(n) RNAi) when normalized to actin-5c as a loading control. (Fig. 18). Northern blots to the scrambled control have not yet been completed. To further confirm depletion of AAGAG(n) RNA seen cytologically in nuclei, I performed AAGAG(n) RNA-FISH in larvae after AAGAG(n) RNAi. Figure 18b shows that AAGAG(n) RNA foci are removed after AAGAG RNAi, but not in the Tubby control, demonstrating selective abolishment of AAGAG(n) RNA. This also has not yet been completed in scrambled control.

Fig. 18. AAGAG(n) RNA is decreased with RNAi. A) Northern blot with probes to AAGAG(n) RNA or actin-5c in L3 RNAi or Tubby control. B) RNA-FISH to AAGAG(n) in brain lobes of Tubby control or AAGAG(n) RNAi. AAGAG(n) (green) DAPI (blue) H3K9me3 (cyan).

mRNAs with AAGAG(n) RNA are likely not decreased in RNAi

We were initially concerned that RNAi to AAGAG(n) RNA could potentially affect genes containing AAGAG(n) RNA. To address if this was the case, we searched for all mRNAs containing >3 repeats of AAGAG(n) using Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi). This approach identified one mRNA with

AAGAG(3), CG33080, as well as two with partial AAGAG(3) homology: pip5k59B and peb. None of these mRNAs contain the stretches of AAGAG within exons, but nonetheless we performed qPCR of cDNA using oligos flanking AAGAG(n) in these mRNAs to see if they were decreased in AAGAG(n) RNAi larvae. We have not yet completed biological replicates, but initial quantitative PCR (q-PCR) did not detect any noticeable decrease in these mRNA levels compared to the Tubby control. (data not shown), This demonstrates that we are monitoring effects of AAGAG(n) satellite RNAs and not mRNAs.

AAGAG(n) RNA is necessary for viability

In order to determine if AAGAG(n) RNA is necessary for viability, we decreased AAGAG(n) RNA via actin-GAL4 RNAi and assayed lethality prior to and during pupal stage, rate of eclosion, and lifespan of eclosed adults relative to Tubby control and scrambled RNAi control. When female virgin act-GAL4/Tubby flies were crossed to males containing shRNA-AAGAG, the resulting AAGAG RNAi progeny exhibited roughly 20% lethality compared to Tubby control by pupal stage, suggesting that some AAGAG(n) RNAi embryos or larvae are not viable. When L3 AAGAG(n) RNAi were assayed for lethality, we noticed 10% died before reaching pupal stage, in contrast to Tubby control, suggesting 10% of embryos with AAGAG(n) RNAi die in embryo stage. We did not observe significant lethality in pupae, although 20% of AAGAG(n) RNAi adults died within 3 days after eclosion, and prior to death the adults appeared weak and had difficulties getting un-stuck from the media. Together, this demonstrates that decreasing AAGAG(n) satellite RNA causes embryonic, larval and adult lethality. Surprisingly, scrambled RNAi caused 100% pupal lethality and Tubby pupae from the same cross exhibited 30% pupal, much higher than typically observed for Tubby. This suggests off-target effects of scrambled RNAi, although RNAs, both coding and non-coding, with sequences potentially targeted by scrambled RNAi were not found in RNA-seq datasets. The scrambled shRNA sequence is found in some heterochromatic regions, but kmer analysis of 16-18hr embryos did not identify any putative heterochromatic transcripts containing this sequence. Furthermore, the lethality in Tubby pupae from this cross suggests that the issue is not off-target effects of scrambled RNAi, although theoretically small levels of maternally deposited act-GAL4 could still exist in pupae to cause off-target RNAi. Also, homozygous shRNA-scrambled flies do not exhibit lethality, ruling out toxicity of the DNA sequence itself or ectopic insertion of the shRNA vector elsewhere in the genome. These observations are perplexing and suggest that scrambled RNAi lethality is not due to off-target RNAi effects, but rather from a combination of the shRNA and act-GAL4 or Tubby. In order to determine if lethality seen in AAGAG(n) RNAi flies was due to overwhelming the RNAi pathway, we measured lethality relative to mCherry RNAi. We did not observe significant amounts of lethality at any stage with mCherry RNAi, demonstrating that the lethality observed for AAGAG(n) RNAi is due to knockdown of AAGAG(n) RNA and not from perturbation of the RNAi pathway.

AAGAG(n) RNAs do not affect established heterochromatin organization

Considering that AAGAG(n) RNAs localize in the nucleus to heterochromatic regions, we suspected they may have a role in chromosome organization. To see if they affected large scale heterochromatin organization we assayed for disruption of HP1a-GFP foci in salivary glands of L3, with act-GAL4 driven AAGAG(n) RNAi compared to scrambled control. However, we did not see a noticeable increase in the number of HP1a-GFP foci in salivary glands after AAGAG(n) RNAi, suggesting that heterochromatin organization and integrity was not disrupted in these cells (data not shown). I also did not see a substantial change in signal intensity or organization of H3K9me2 in AAGAG(n) RNAi brain lobes or salivary glands compared to L3 progeny with CyO-Tubby (data not shown). Another method to test if a gene affects heterochromatic organization is assaying for suppression of variegation, which tests whether or not a gene juxtaposed to heterochromatin, termed position effect variegation (PEV), becomes unsilenced after removal of the test factor.[113] This method essentially testes for heterochromatin spreading, and led to initial identification of many heterochromatin proteins.[114] Due to the lethality of the scrambled control, we were not able to accurately compare variegation in the AAGAG(n) RNAi relative to scrambled RNAi, and are

currently using mCherry and GFP RNAi as controls for this. However, in our opinion there does not appear to be a strong suppressor of variegation (su(var)) phenotype in AAGAG(n) RNAi flies compared to known su(var) proteins. Taken together, this suggests that at least in larvae and adults, AAGAG(n) satellite RNAs do not or only minimally affect heterochromatin organization.

AAGAG(n) RNAs do not affect the heat shock response

Human satellite III repetitive non-coding transcripts, as well *Drosophila* non-coding omega-speckle RNAs, are critical for proper heat shock and other stress responses, and essentially function by localizing stress response proteins to hsrω genomic loci. [115,116] The ChiRP-ms screen for AAGAG(n) RNA binding proteins identified association with heat shock proteins, such as Heat shock protein 23 (Hsp23), Heat shock 70 cognate 5 (Hsc 50-C), Heat shock 70 cognate 4 (Hsc 70-4), and Heat shock protein 83 (Hsp83). With the exception of Hsp23, all bind HP1a under different conditions.[24] As their name implies, these proteins are important for heat and stress responses, in addition to other functions. This suggested that AAGAG(n) satellite transcripts may be important for the heat shock stress response. However, AAGAG(n) RNA is neither up-regulated nor differentially localized during heat shock, and AAGAG(n) RNAi larvae are not sensitive to heat shock under two classical heat shock conditions that give either significant developmental delays or 50% death in HSP70 defective flies[117] (data not shown), demonstrating that AAGAG(n) RNA is not a critical interactor in the stress response pathway. Why AAGAG(n) RNA binds these heat shock proteins is thus a mystery, and we are currently determining if AAGAG(n) RNA affects their localization.

AAGAG(n) RNAs may contribute to proper chromosome segregation

The presence of AAGAG(n) RNA(s) prior to and during the MZT suggested that AAGAG(n) RNA(s) might facilitate initial chromosome and/or heterochromatin organization. To initially test if early embryos exhibited substantial defects in chromatin organization after AAGAG(n) depletion by RNAi, I crossed act-GAL4/CyO-Tb-RFP females to males with AAGAG(n) shRNA, then performed AAGAG RNA-FISH combined with DAPI staining in early embryo progeny. This cross should yield half embryos with act-GAL4 driven RNAi and half with Cyo-Tb balancer/ shRNA. Interestingly, roughly half of the embryos in cycle 13 contained nuclei with thread-like DAPI staining between two dividing cells in early Drosophila embryos (Fig. 19). The two nuclei attached via these DNA threads never contained AAGAG(n) RNA foci either on the stretched part or within the nucleus. However, in all cases the embryos that contained nuclei with stretched chromatin also contained AAGAG(n) RNA foci in non-stretched nuclei, suggesting incomplete knockdown of AAGAG(n) in these embryos. Together, this suggests that chromosomal segregation defects occur in nuclei depleted of AAGAG(n) RNA but not in those in which AAGAG(n) RNA is still present. There are many caveats to this experiment, and therefore our conclusions are highly speculative, but at the very least this suggests that AAGAG(n) may have a role in early chromosome segregation. (see Discussion)
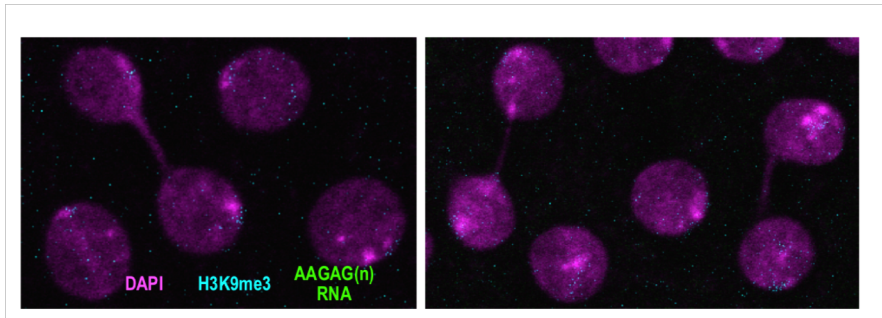
Fig. 19. AAGAG(n) RNA may contribute to proper chromosome segregation. Shown are projections of two embryos entering cycle 13 with putative RNAi knockdown of AAGAG(n) RNA

Conclusions/Future directions

Here I demonstrate that AAGAG(n) RNA is depleted with a ubiquitous RNAi driver, actin-GAL4, and provide preliminary evidence that mRNAs from euchromatic genes containing one or a few tandem AAGAGs are not depleted after AAGAG(n) RNAi. Additionally, and contrary to our hypothesis of AAGAG(n) RNA affecting heterochromatin organization, we demonstrated that AAGAG(n) knockdown does not drastically affect HP1a or H3K9me2 organization in L3. However, our preliminary data does suggest AAGAG(n) RNA may have a role in early chromosome segregation. Demonstrating this definitively will require creating fly constructs that abolish or decrease maternally loaded AAGAG(n) RNA, combined with flies that allow for assaying which early embryos contain act-GAL4 vs Cyo-Tb-RFP balancer. We will address this issue by crossing :shRNA:Dcr2; Nanos-GAL4/H2AV-RFP females to :UAS-GFP:act-GAL4/Cyo-Tb-RFP males, perform live imaging of resulting embryos, and score those with H2AV-RFP and UAS-GFP expression (i.e. those expressing act-GAL4 and creating AAGAG(n) RNAi) for chromosomal segregation defects. If segregation defects occur during mitosis in the AAGAG(n) RNAi line and not in the scrambled control, I will then determine which chromosome(s) exhibit disorganization during mitosis using DNA-FISH, and whether or not AAGAG(n) DNA is associated with this disorganization. These fly constructs will also allow me to determine if AAGAG(n) RNA facilitates early heterochromatin formation, by assaying fixed embryos for GFP expression (thereby expressing act-GAL4 and creating RNAi to AAGAG(n)) and H3K9me2/3 levels and distribution. Additionally, I will use HP1a-RFP instead of H2AV-RFP in the construct listed above to determine if AAGAG(n) affects early HP1a distribution. If early embryos with AAGAG(n) RNAi have lower levels of H3K9me2/3 or altered organization of H3K9me2/3 and/or HP1a, I will conclude that AAGAG(n) RNA has role in early heterochromatin formation. If AAGAG(n) RNA is functional in these regards, then identifying maternally loaded proteins it binds using CHiRP-mass spec will guide mechanistic studies. For instance, if AAGAG(n) RNA facilitates early H3K9me2/3 distribution or HP1a organization, we would suspect that it binds and somehow guides Su(var)3-9 methyltransferase or deacetylases to heterochromatin genomic locations for heterochromatin marking.

The observation that AAGAG(n) RNAs bind heat shock proteins, combined with established roles of repetitive RNAs in heat shock[115,116] led to a reasonable hypothesis that AAGAG(n) RNAs affected the heat

shock response. Clearly, this was not the case. It is currently unclear what, if any, functional association AAGAG(n) RNAs have with heat shock proteins, but many roles for these proteins exist outside of stress and heat shock response, such as cell cycle regulation and Piwi regulation of transposon expression.[118][119][120] It is possible that AAGAG(n) RNA somehow regulates localization of these heat shock proteins involved in these functions or somehow is involved in similar pathways. Additionally, several heat shock proteins bound by AAGAG(n) RNA are crucial for spermatogenesis or are expressed in testes, which suggested that AAGAG(n) could function in spermatogenesis.

The identification of other non-heat shock proteins bound by AAGAG(n) RNA in L3 (Table 2) suggests a role of AAGAG(n) in the organization and/or function of these proteins. Two of these proteins in particular, Tudor-SN and Dodeca binding protein 1 (Dp1), are particularly intriguing, given that each interact with HP1a, associate with heterochromatin, and facilitate heterochromatin structure.[24] [72-74] Though neither have been extensively studied, both Tudor-SN and Dp1 have heterochromatic RNA roles, Tudor-SN being associated with the RNA induced silencing complex (RISC)[121] and Dp1, a vigilin, with A to I modified RNAs.[122] This would suggest that AAGAG(n) containing RNAs recruit the RISC and/or are modified through the ADAR complex. Interestingly, embryos deficient in maternally contributed Dp1 exhibit high levels of chromosomal segregation defects, such as lagging chromosomes and elongated chromatin fibers immediately after telophase[123], very similar to what I observed in embryos after AAGAG(n) RNAi (Fig. 19). This is intriguing and could indicate that AAGAG(n) RNA localizes Tudor-SN and Dp1 to heterochromatic regions to facilitate heterochromatin formation and potentially other functions.

Materials/methods

<u>Northern blotting.</u> Performed as in Chapter 2

<u>Fly lines used:</u>

Stock name	Obtained from	Description
y[1] v[1]:UAS-AAGAGshRNA::	Rainbow Transgenic Flies, Inc	Expresses shRNA under UAS promoter targeting AAGAG(n)
y[1] v[1]:UAS-scramble shRNA:	Rainbow Transgenic Flies, Inc	Expresses shRNA under UAS promoter targeting a random sequence
y[1] w[*]:: P{w[+mC]=Act5C-GAL4}17bFO1/TM6B, Tb[1]	Bloomington: stock 3954	Expresses GAL4 ubiquitously under control of Act5C promoter (P{AyGAL4} with the y[+] FRT cassette flipped out).

<u>RNA-FISH in embryos and larvae:</u> Performed as described in Chapter 2, using the (3) 'TSA amplification protocol'.

<u>Antibodies:</u>
rH3K9me2 (active motif Cat. 39753) used as described in Ch. 2

Satellite RNA General Summary and Future Directions

Our observations that AAGAG(n) or AG(n) containing satellite RNAs are expressed in virtually all fly tissues, that these types of RNAs are maternally loaded and associate with the earliest forms of heterochromatin, combined with our observation that these RNAs contribute to male fertility suggests that satellite RNAs are critical to chromatin functions in flies.

The detection of AAGAG(n) or AG(n) containing RNAs in spermatocytes is novel but not entirely surprising, considering that in fly spermatocytes, roughly 50% of all genes are transcribed[124] and in mice there is a significant upregulation of TE and lncRNA transcription in spermatocytes and round spermatids.[125] In this seminal work, the authors found that endogenous retroviruses (ERVs), which are a class of TEs, form novel lncRNA promoters in post mitotic germ cells, and drive expression of these novel lncRNAs in a strand specific orientation. [125] This led them to theorize that activation of ERVs in the post mitotic spermatogenic cells allows the germline to explore new gene synthesis through the functional evolution of lncRNAs to protein coding genes.[125] Is it possible that perhaps in *Drosophila melanogaster* this same mechanism led to not only to the evolution of lncRNAs into functional proteins, but also allowed for selection of functional satellite RNAs containing AAGAG(n)? It is of course possible that satellite transcripts containing AAGAG(n) RNA identified here are themselves part of TE's and/or ERVs that drive transcription of nearby genes that are necessary for spermatogenesis. It will be interesting to see of over-expression AAGAG(n) RNA in an AAGAG(n) RNAi background can rescue male fertility defects. If so, this not only would demonstrate that it is the AAGAG(n) RNA sequence itself that is the active fertility component, but also that fertility defects seen in AAGAG(n) RNAi are not due to silencing an AAGAG(n) RNA transcribing ERV-derived promoter that drives expression of a necessary protein in *cis*. Even if this was the case it would be the first evidence to our knowledge of a non-coding satellite region driving expression in *cis* of a required fertility factor in flies. We do not think this is the case, considering that RNAi machinery should be decreasing AAGAG(n) RNA satellites only after the act of transcription, although there is a slight possibility that the RNA induced silencing complex (RISC) is recruited the site of transcription to cause silencing of the resulting promoter. Also, the presence of several satellite repeat RNA species from different chromosomes, combined with antisense transcription of CTCTT(n) in testes, however, suggests an RNA centric function of these satellites. It is also plausible that these transcripts, independent of sequence, are required for targeting certain pathways to the genomic locations from which they are transcribed, such as in the case of PIWI. This case, however, also does not rule out a function for the resulting processed RNAs, which once processed by PIWI could be then be necessary for action at some other genomic loci or for protein function. It is also possible that satellite transcripts are robustly transcribed in spermatocytes and function to sequester proteins indiscriminately to specific spermatocyte regions, either targeting them for function at these locations or preventing unwanted functions elsewhere.

In somatic cells, it is perplexing that knockdown of AAGAG(n) containing RNA does not seem to affect organization of heterochromatin markers such as HP1a and H3K9me2, yet this RNA must come from heterochromatin and appears to affect viability. These RNAs did not bind HP1a or Su(var)3-9 in CHiRP-ms, two core components of heterochromatin, yet bound other heterochromatic proteins in addition to proteins not known to associate with heterochromatin. This may suggest that these RNAs do not affect large scale organizations of heterochromatin per se but instead affect sub-heterochromatic components and functions. This theory is based on our observation that somatic AAGAG(n) and AG(n) RNA containing satellite transcripts from different chromosomes, ie those from 2R and X, seem to co-localize into one sub-heterochromatic foci. Even more surprising was the observation that in different stages of development different strands of each of these satellite regions are transcribed and co-localize into one, presumably sub-heterochromatic, foci. This phenomenon of different RNA species co-localizing into one body is reminiscent of p-bodies and stress-granules and leads one to speculate whether or not heterochromatic RNAs form

similar nucleation sites within heterochromatin to localize heterochromatin or other proteins for function. As imaging systems evolve and allow for higher resolution, it is becoming clear that classically defined domains such as heterochromatin are in fact composed of many dynamic and finer sub-domains. [24] It will be interesting to determine if the RNA foci I detected co-localizes with any proteins found to form sub-domains within heterochromatin, such as Hsp83, which also binds AAGAG(n) RNA, and if so, whether or not abolishing these satellite RNAs prevents these subdomain formations. Heterochromatin is also a dynamic and transient form of chromatin at least in part regulated by phase separation, [49] suggesting that heterochromatic RNAs may regulate phases of sub-domain proteins. This leads us to speculate that different regions of heterochromatin are transcribed to regulate phases of the distinct proteins within these regions. Understanding the complex interplay between RNA localizations, RNA binding proteins, phase separation and corresponding functions is bound to be challenging but will be critical to understand how biophysical properties translate into cellular organizations and functions.

It will also be critical to test whether or not functions of satellite RNAs stem from the transcript itself, from promoter start sites, or a combination of the two. This too is bound to be challenging, especially considering that we do not yet have a grasp on the number of satellites transcribed and from where. Hopefully sequencing technologies such as Nanopore sequencing will evolve to allow for highly accurate long-read sequencing so as to identify all satellites transcripts and fully assemble satellite regions of heterochromatin. Once identified, CRISPR tools will likely be instrumental in identifying *cis* vs *trans* roles and promoter vs transcript roles, similar to what has been done recently for select TE's. [126] Whether or not different satellite RNAs share similar conserved functions or contribute to evolution is yet another intriguing mystery. Answering these questions will be especially exciting and will certainly begin to shine light deep in the depths of the dark matter of the genome.

Chapter 5: Identification of lncRNAs associated with heterochromatin and their organismal distribution

Recent RNA profiling through the D. melanogaster modENCODE project showed that many non-coding regions of the annotated genome are transcribed, and subsequently identified 1875 long non-coding RNA (lncRNA) genes, 1500 of which are novel.[127] As time progresses and new technologies emerge, the numbers of annotated lncRNAs continually increase, and indeed a more recent study identified over 600 previously unannotated lncRNAs in Drosophila. [128] Despite the high abundance of lncRNAs in *Drosophila*, human [129]and *C elegans*, [130] only a small fraction of lncRNAs have been extensively characterized and ascribed a functional role[27131], likely due to the relative ease of discovering new transcripts compared to the tedious experimentation needed to study localization and functions of lncRNAs. Therein there is a great need to prioritize characterizing lncRNAs in terms of localization and functions as the field expands. Many lncRNAs have been discovered that affect chromosome organization and function as described in Chapter 1. [28-38] However due to technical limitations and precedence, primarily only those that affect euchromatin or facultative heterochromatin organization have been published. One exception, though, demonstrated that a processed lncRNA is critical for guiding complexes to a rDNA promoter to initiate heterochromatin formation around the nucleolus in mice ESCs, [28] demonstrating that lncRNAs can and in this case do facilitate constitutive heterochromatin formation. To our knowledge, no examples of lncRNA mediated constitutive heterochromatin formation, maintenance or function have been described in *Drosophila*, prompting me to address this issue.

Similar to protein distribution and localization, lncRNA subcellular localization has been linked to function. [132] This suggested that identifying heterochromatin associated lncRNAs could potentially guide us in the characterization of those with heterochromatic function. To address this, our overarching approach was to first determine which, if any, lncRNAs associated with heterochromatin by using RNA immuno-precipitation followed by sequencing (RIP-seq) using heterochromatin proteins as 'bait' in the pulldown. We then proposed to validate co-localization of these lncRNAs with heterochromatin via RNA-fluorescent in-situ hybridization (RNA-FISH) combined with Immuno-fluorescence (IF) of heterochromatin markers H3K9me2 and 3. For those that associated with heterochromatin, we then sought to test them for heterochromatic function through abolishment or 'mutation' of the lncRNA and assay for different heterochromatic functions such as PEV, DNA repair, and silencing, among others.

The first step in this strategy, RIP-seq, was critical for generating a quality data set, and therefore we took great care in assay design. RIP-seq approaches suffer from many complications, most notably that of high background, and in the case of crosslinking, pulldown of background RNAs, complicating efforts to determine RNA-protein interactions. The most stringent methods of determining bonafide RNA-protein interactions involve versions of Cross Linking Immuno-Precipitation (CLIP-Seq), which essentially crosslinks RNA directly to protein residues it binds. However, this technique is not easily amendable to whole animals and even in ideal situations can produce mis-interpretation depending on the RNA localization relative to cross-linkable protein residues. In short, the ideal protein-RNA pulldown protocol employed to determine these interactions depends on the questions being asked. Since we were interested in identifying lncRNAs that associate with heterochromatin, I decided to use a version of non-crosslinking RIP-seq, which identifies RNAs that bind both directly and indirectly to the protein that is pulled-down.

RIP-seq in embryos with heterochromatin proteins identifies associated lncRNAs

My approach to performing RIP-seq for heterochromatin associated lncRNAs consisted of using validated antibodies to pulldown H3K9me2, which likely does not directly bind RNA but should retrieve any associated RNA as well as Shelob, a putative RNA binding protein characterized by our lab that associates with HP1a and affects PEV. (manuscript in preparation) As a positive control I used MSL2, which binds

roX1 and roX2 as described in chapter 1, and as a negative control IGG. Ideally, HP1a would be used as heterochromatic 'bait' in RIP-seq, but antibodies suitable for native pulldown do not exist, and tagged versions would need separate, unavailable controls. However, the approach adopted for RIP-seq was strict, in the sense that it used non-cross linking conditions, a positive control and a stringent negative control consisting of IGG to eliminate from analysis 'sticky' background RNAs that bound beads or IGG protein. Additionally, in order to potentially identify any lncRNA that associated with heterochromatin in each embryonic developmental stage, I used 0-24hr embryos in biological replicates.

This RIP-seq approach resulted in the identification of 631 annotated, non-repeat/TE RNAs that associate significantly (padj<0.05 over control) with Shelob only, 36 with H3K9me2 only, as well as 72 that bind both Shelob and H3K9me2 (not shown). Additionally, most of the mRNAs that bind Shelob fall under GO terms that encompass development such as anatomical structure morphogenesis, system development, post-embryonic organ development, post-embryonic organ development, and metamorphosis, suggesting that Shelob has a role in development mediated by binding RNA. As validation that the approach accurately detected RNA-protein interaction, we note pulldown by MSL2 of roX1, roX2 and MSL2 RNA, reported and validated binders of MSL2. [133] It is important to note that sequencing methodology (ie using random hexamers to make cDNA and PCR amplification) and sequence analysis (necessity of having unique mappable ends) did not allow for the identification of simple satellites from this pulldown, which limited this analysis of repeats to annotated RNAs containing unique sequences mappable to an annotated database. Regardless, both H3K9me2 and Shelob proteins associated with repetitive RNAs such as TEs, although H3K9me2 exhibited more enrichment for most transposons relative to Shelob (Fig 20). Msl2 exhibited <50 fold enrichment (FE), if any enrichment at all, for all transposons, with one exception, Gypsy_1, at 115 FE. Together, this this demonstrates that our RIP-seq approach accurately identified RNA-protein interactions and suggested that it could be instrumental in identifying heterochromatin associated lncRNAs.
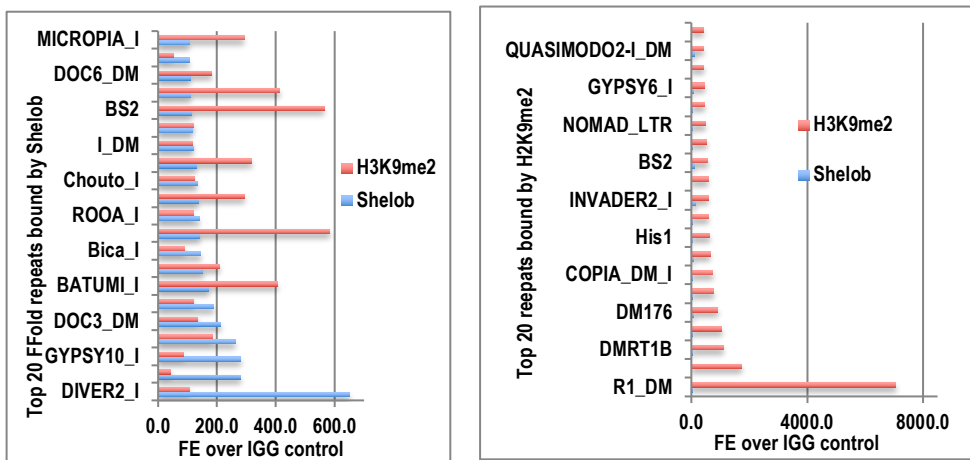
Fig. 20. Non-satellite, non rRNA repetitive RNA bound by Shelob and H3K9me2 normalized to IGG controls. A) Repeats bound by Shelob B) Repeats bound by H3K9me2

Importantly, this screen also identified 17 lncRNAs that associate either with Shelob and/or H3K9me2 (Table 10).

Table 10. lncRNAs that bind Shelob and/or H3K9me2

CR45736	CR42862	CR44717	CR45911	CR43650	CR46003
CR42425	CR43751	CR45347	CR43314	CR45361	CR43314
CR31044	CR44206	CR45736	CR43334	CR45916	

We were also interested in identifying and characterizing lncRNAS that are upregulated in early embryos during the time of heterochromatin formation, since these are lncRNAs that could potentially act as early initiators of heterochromatin formation, but due to the relative abundance of small numbers of cells in early embryos compared to late embryos might not have been detected via RIP-seq. These were determined based on identifying lncRNAs above 10 reads per kilobase Million (RPKM) [134] in 2-4 hour embryos and are listed in table 11.

Table 11. lncRNAs upregulated in 2-4 hour embryos.

Expressed from heterochromatin	Expressed from euchromatin			
CR42653	CR43887	CR32207	CR43973	CR43241
CR42722	roX1	bxd	CR43483	
CR42723	CR34335	CR43482	CR32218	
CR43241	CR43432	CR42491	CR32009	

To then determine cytologically where these transcripts were located during development and if they localized to heterochromatin, I performed co-IF-RNA-FISH with H3K9me2 or 3 and probes to the lncRNA in embryos and larvae. Unfortunately, none of these lncRNAs co-localized strictly with heterochromatin. However, and interestingly, the majority exhibited nuclear localization and developmental patterning. (Fig. 21A and B and table 12). Those with nuclear localization could be classified as having broad or multiple foci, peripheral foci, or single foci (Fig 22 and table 12). A few were localized to the cytoplasm (Fig. 22), and one, CR32009, exhibited both nuclear and cytoplasmic localization. (Fig. 22 and table 12). Additionally, some of these lncRNAs were found in different tissues of L3. (Fig. 23 and table 12). Table 12 gives an overview of localizations and other pertinent information.
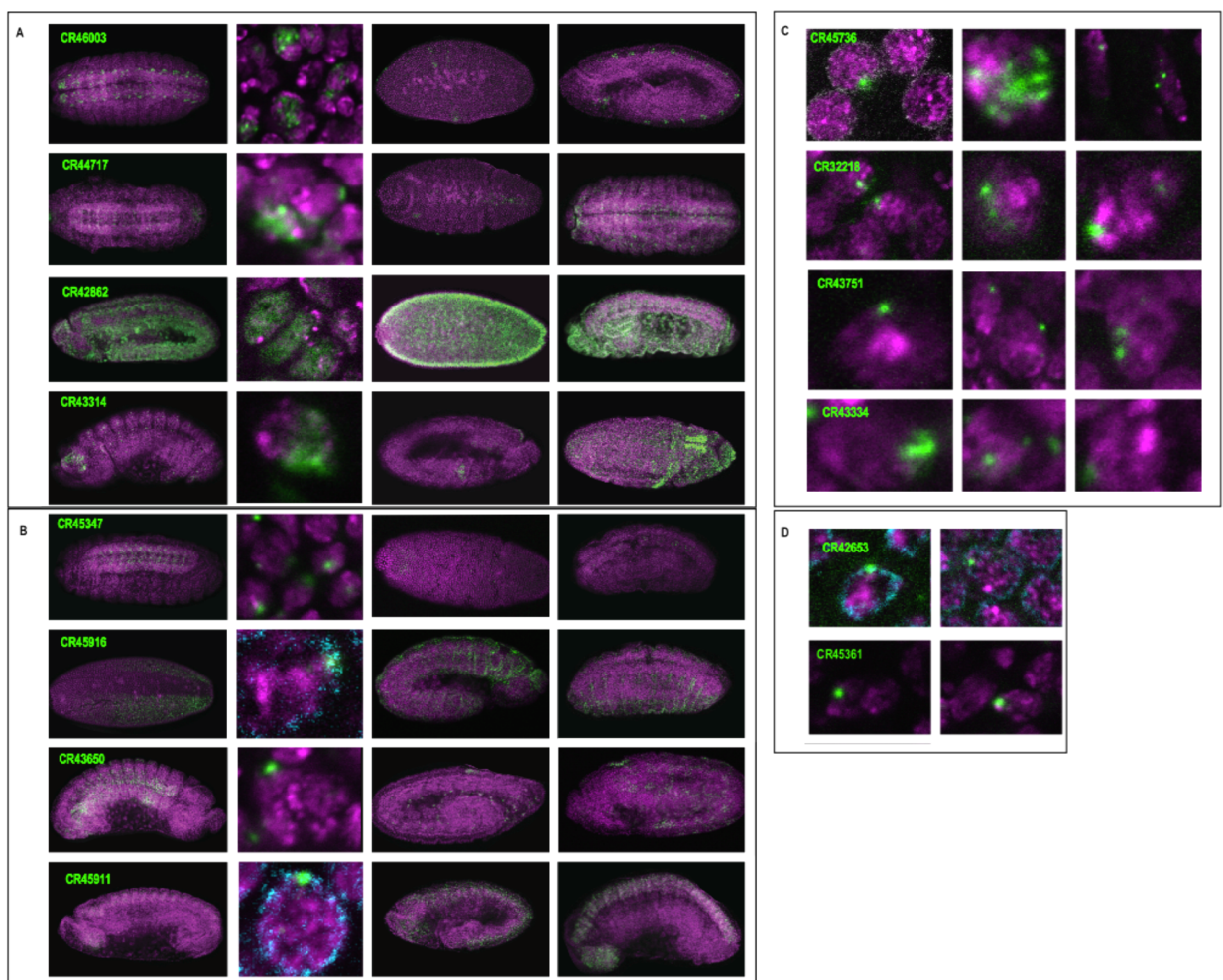
Fig. 21. lncRNAs exhibiting nuclear localization. Shown are slices of embryos and nuclei labelled with DAPI (magenta) and lncRNA (green), and where appropriate Lamin B (cyan). Note that mages are not necessarily to scale. A) lncRNAs that exhibit developmental patterning and are present as multiple or broad nuclear foci. B) lncRNAs that exhibit developmental patterning and are present as one foci per nucleus. Note the appearance of some foci in the nuclear periphery, marked with Lamin B. C) Images of one or two nuclei containing lncRNAs with multiple or broad foci not present in developmental patterns. D) Images of nuclei containing lncRNAs present as single nuclear foci not present in developmental patterns.
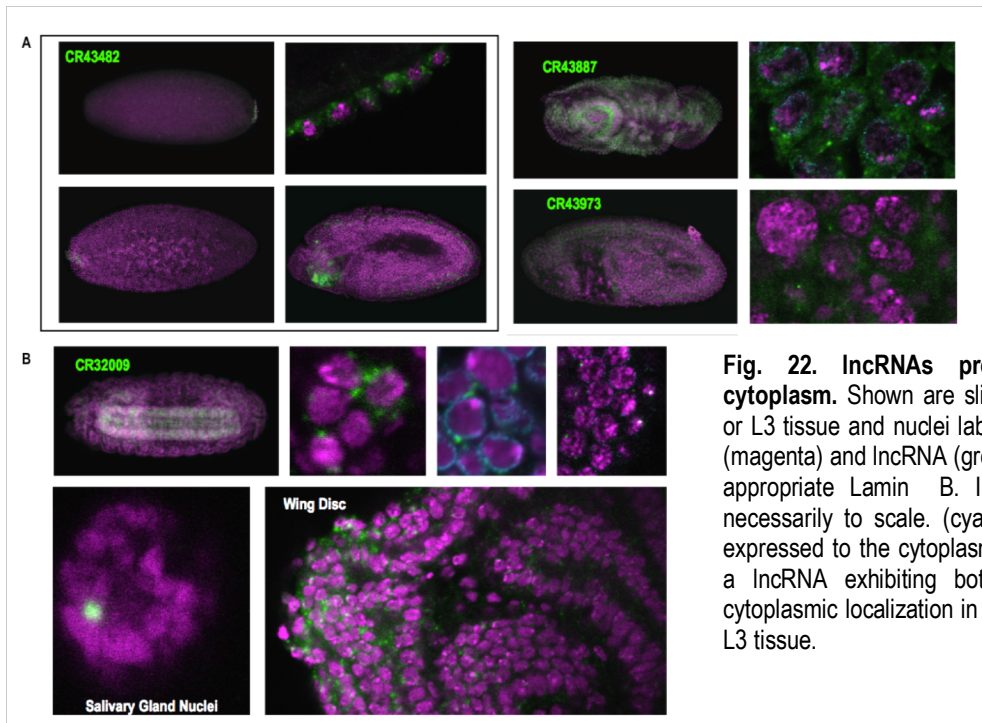
Fig. 22. lncRNAs present in the cytoplasm. Shown are slices of embryos or L3 tissue and nuclei labelled with DAPI (magenta) and lncRNA (green), and where appropriate Lamin B. Images are not necessarily to scale. (cyan). A) lncRNAs expressed to the cytoplasm. B) CR32009, a lncRNA exhibiting both nuclear and cytoplasmic localization in the embryo and L3 tissue.
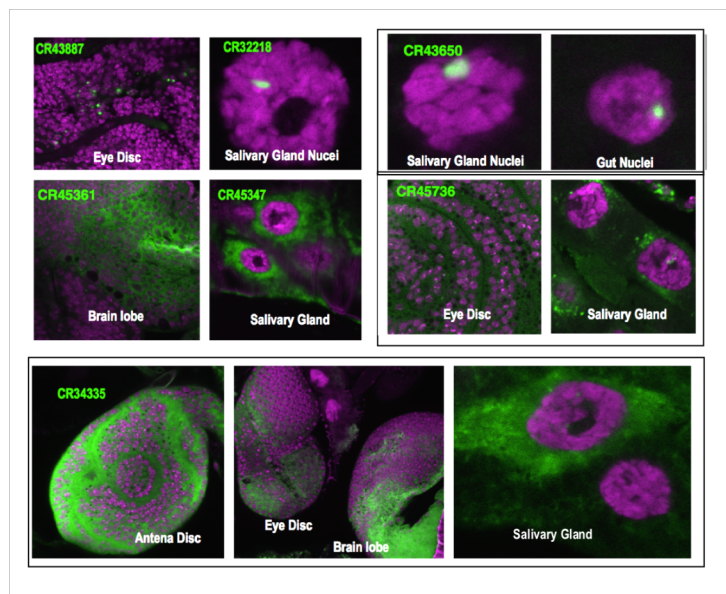
Fig. 23. lncRNAs expressed in L3. Shown are slices of embryos or L3 tissue and nuclei labelled with DAPI (magenta) and lncRNA (green). Images are not necessarily to scale

Table 12. Summary of lncRNA RNA-FISH results

lncRNA	Localization	1-3	4-5	6-7	8-9	10-17	L3 RNA FISH-SIGNAl?	Localization	Salivary gland	gut	optic lobe	eye disc	wing disc	human ortholog? (no)	does flybase have expression data for embryos? (yes, 8-12hr)	Data for L3 expression in Fly base? (yes, CNS)	Has miRNA?
CR46003	Nuclear	No	yes	yes	yes	yes	No	nuclear									
CR44717	Nuclear	NT	yes	yes	yes	yes	NT										
CR42862	Nuclear	No	yes	yes	yes	yes	NT										
CR45911	Nuclear	No	NT	yes	yes	yes	NT										
CR45361	Nuclear	NT	NT	NT	yes	yes	Yes	cytoplasmic			yes			no	no	No	yes
CR45916	Nuclear	No	yes	yes	NT	yes	yes	cytoplasmic			yes			no	yes, 4-10hr	No	yes
CR32218	Nuclear	NT	yes	yes	yes	yes	yes	Nuclear	yes					no	yes	yes	
CR42653	Nuclear	NT	yes	yes	yes	yes	yes	Nuclear	yes					no	no	No	
CR43650	Nuclear	NT	NT	yes	yes	yes	yes	Nuclear		yes	yes			no	yes, 6-24hr	yes, SG	yes
CR43334	Nuclear	NT	NT	yes	yes	yes	yes	Nuclear	no	no	no	yes	no				
CR45911	Nuclear	No	NT	yes	yes	yes	No							yes	yes, 4-12hr	no	yes
CR43314	Nuclear	No	NT	yes	yes	yes	No							no	no	No	yes
CR46003	Nuclear	NT	yes	yes	yes	yes	NT							yes	no	yes	
CR44717	Nuclear	NT	yes	yes	yes	yes	no							yes	yes	yes	
CR42862	Nuclear	No	yes	yes	yes	yes	yes	Nuclear	yes	yes	yes	yes	yes				
CR45916	Nuclear	No	yes	yes	NT	yes	NT										
CR45347	Both	No	yes	yes	yes	yes	yes	Cytoplasmic	yes		no		yes	yes	yes, 14-22hr	no	yes
CR45736	Both	NT	yes	yes	NT	NT	yes	Both	yes		yes	yes		no	no	No	
CR43973	Cytoplasmic	NT	NT	NT	NT	NT	no							no	yes, mainly 0-6hr	yes	
CR43335	Cytoplasmic	NT	yes	NT	NT	NT	yes	Cytoplasmic	yes		yes	yes	yes	no	yes	yes	
CR43482	Cytoplasmic	yes	yes	yes	yes	yes	NT							no	yes, 0-2 hr	yes	
CR43432	Cytoplasmic	No	yes	yes	No	No	yes	Cytoplasmic					yes	no	yes, 4-16hr	yes	
CR43887	Cytoplasmic	NT	yes	NT	NT	NT	No	Nuclear				yes		no	yes, 2-6hr	no	
CR43751	Cytoplasmic	NT	No	No	yes	no	no							no	yes, only 4-10hr	no,	
CR44206	Cytoplasmic	NT	NT	yes	NT	NT	No							yes	no	yes	yes
CR32009	Both	NT	NT	NT	yes	NT	yes	both	yes			yes		no	no	No	

The observation that many of these lncRNAs exhibit nuclear expression during development only in certain tissues was especially exciting and suggested that some of these lncRNAs could exhibit roles in development. Consistent with a known enrichment of lncRNAs in neural tissue of humans and mice, [135] [136] we note at least seven lncRNAs that appear to be primarily expressed from neural tissue, such as CR46003, CR44717, CR43314, CR45347, CR43650, CR45911, and CR32009. Additionally, one, CR43482, is expressed from pole cells in early embryos, which are essentially precursor germline cells. This suggests CR43482 may have a role in germline development. We also note lncRNAs that appear to have multiple and/or broad nuclear foci, such as in the cases of CR46003, CR42862, CR44717, CR43314, CR45736, CR32218, CR43751, and CR43334, which would suggest that these lncRNAs may potentially migrate to other localizations in the nucleus for subsequent function in *trans*. We also note the presence of lncRNAs that appear to co-localize with lamin-B in the nuclear periphery, such as CR45916, CR45911, CR42653, CR45736, and CR32009. With the exception of CR32009, all are strictly nuclear localized and either bind Shelob or H3K9me2 or in the case of CR42653, expressed from heterochromatin. The importance of these localizations to the nuclear periphery are unknown, although it is plausible that they somehow target DNA sequences to lamin, which itself has many functions in gene regulation and chromatin organization. [137] CR42862 appears to be ubiquitously expressed throughout the nucleus in embryos and L3, and notably is essentially devoid in DAPI rich regions, suggesting that it is excluded from heterochromatin. Additionally, flybase.org indicates that CR42862 has a human ortholog, suggesting

potentially functional roles. We note, however, that predicting functions for lncRNAs based on protein orthology is highly speculative, considering the lack of larger coding sequences in lncRNAs.

Also, some lncRNAs tested for localization are pri-miRNAs, which are pre-cursors to micro-RNAs (miRNAs). When making probes to these, I specifically excluded the miRNA RNA region to identify spatial localizations independent of miRNA targets. One of these pri-miRNAS, CR43314, contains the bft miRNA. It is interesting that this lncRNA exhibits broad nuclear patterning in late stage embryos, which is in contrast to single foci found with the other pri-miRNAs. It is unclear what this means, but it could suggest this lncRNA has chromatin organization functions besides those produced from the bft miRNA.

Due to these interesting localizations and in some cases expression patterns that suggested developmental regulation, we were interested if any of these lncRNAs exhibited functions such as PEV and viability. As an initial approach to address this, we obtained all available P-element insertions in these lncRNAs, hoping that P-elements inserted close to promoters or insertion in functional components of lncRNA sequences would abolish function. We therefore obtained flies with p-element insertions to CR44717, CR46003, CR45347, CR42822, and CR43314. However, P-elements to CR46003 and CR42862 caused significantly increased expression of these lncRNAs, with no visible changes in viability or fertility. Two P-elements, one to CR44717 and CR43314, gave inconclusive knockdown of these lncRNAs, likely suggesting that these P-elements do not affect expression of these lncRNAs. These flies also did not exhibit defects in viability, fertility or overall visible phenotypic defects in morphology. However, a P-element in CR45347, which contains mir-4973, a miRNA with unknown function, is homozygous lethal. Additionally, lethality of this P-element has not been reported, and together suggests that mir-4973 is necessary for viability. These results and images of localizations will therefore serve as an important starting point for those wishing to study any of these particular lncRNAs.

Conclusions

Here, I identified a set of lncRNAs that associate with heterochromatin proteins, identified their subcellular localization, in addition to other lncRNAs that are up-regulated in 2-4 hour embryos when heterochromatin is first formed. Importantly, these lncRNAs exhibit diverse patterning in cells and tissue, as some are expressed predominantly or exclusively in specific tissue at different developmental stages. With one exception, CR42862, the localizations of these lncRNAs have not been described, and in addition stage specific expression levels in embryos or specific L3 tissues for most lncRNAs assayed here have not been published in flybase.org. The data presented herein is thus a critical starting point for researchers looking to study the lncRNAs described.

Furthermore, we report that one lncRNA, CR45347, which contains mir-4973, is necessary for viability, as a P-element insertion in this lncRNA, which should abolish mir-4973 expression, is homozygous lethal.

The paucity of sequencing data for some lncRNAs that exhibit robust expression in certain tissues and none in others likely results from a very small fraction of this RNA from whole embryos or larvae. Datasets such as those presented here are thus critical for researchers to identify lncRNAs with particular localizations that may suggest specific functions.

We had hoped to use this screen for heterochromatin lncRNA interactors to identify lncRNAs that affect heterochromatin formation, organization and function. However, we did not find lncRNAs that associated exclusively with heterochromatin ubiquitously in most cells, and therefore the likelihood of these lncRNAs having direct roles on heterochromatin were slim. However, this screen identified many lncRNAs that associate with chromatin, some at different stages of development in specific tissues, suggesting potential roles of these lncRNAs in chromatin organization mediated development. It is not unexpected that P-elements do not decrease and/or abolish most of the lncRNAs we tested for expression in P-element insertions and suggests that using CRISPRi to prevent transcription of lncRNAs should be used for targeted abolishment in future studies. Using this P-element approach to probe for functionality, however,

did identify that CR45347, and potentially mir-4973, is necessary for viability, and warranting further studies. We are currently performing back-crosses to rule out the possibility of another lethal p-element insertion in this fly strain.

Materials/Methods
RIP-seq Protocols:
Essentially, this RIP-seq protocol utilized modified versions of [138,139]I[133] and consists of isolating nuclei from 0-24 hr old Oregon R embryos prior to performing RNA-IP.
Buffers: Digitoxin buffer: 42ug/ul digitonin; 2mM DTT; 2mM MgCl2 in 1X NEH buffer (150mM Nacl; 0.2mMEDTA; 20mM Hepes NaOh, pH 7.4).
KTMHG + 300mM NaCl (20mM Hepes-Koh ph 7.; 140mM Kcl; 2mM MgCl2; 0.2% Tween in DEPC H20) .
RNA wash buffer A (1x PBS; 150mM Nacl; 0.1% T-x100)
Dnase 1 buffer**:** 1x= 10mM Tris pH 7.4; 2.5mM MgCl2; 5mM CaCl2

Table 12. Antibodies used for RIP-seq

antibody	From:	Cat. number	amount
rMSI2	Santa Cruz Biotechnology	Sc-6698; Lot C2216	40ul
rShelob	Lab generated		15ul
mH3K9me2	Abcam	1220	20ul
ProtA Dynabeads	Thermofisher	10001D	90ul
ProtG Dynabeads	Thermofisher	1003D	90ul
rIgG	unkown		20ul
mIgG	unkown		20ul

Embryo collection and processing: Oregon R flies were grown at 25ºC and 0-24hr old embryos dechorionated and flash frozen in liquid nitrogen.
Antibody to Prot A/G Dynabead binding: All subsequent steps were performed in low-retention tubes. Dynabeads were washed in 500ul 3 x's PBS-tween 0.02% pH 7.13 then diluted to 200ul in PBS-Tween per 50ul stock beads. Antibodies were then added and allowed to bind to Dynabeads prot A or G for 10min at RT. Rabbit antibodies were bound to Prot A Dynabeads, while mouse antibodies were bound to Prot G Dynabeads

Main Protocol:
750mg of frozen embryos were diluted in 1.87ml Digitoxin buffer (to allow for better solubilization of plasma membrane) and pippetted up and down for 2 min. Samples were then spun at 1,500g, 10min, 4ºC

Whole Nuclear solubilization: The supernatant was then removed and pellet (now enriched for nuclei) gently diluted in 1.8ml KTMHG + 300mM. Samples were then transferred to a 7ml ground glass dounce and lysed with 4 strokes of the pestle. Samples were then transferred to a 7ml non-ground glass pestle. Residual material from the ground glass pestle was rinsed with 500ul KTM-HG + 300mM NaCl and also transferred into the 7ml non-ground glass pestle. Samples were then lysed again with 50 strokes with pestle B and 50 strokes of pestle A. Then, samples were sonicated with a biorupter in 15ml tubes (1.8ml per tube) 30sec on, 30sec off, 6x's 5 min. Samples were then spun 10,000g, 10min, at 4ºC

Pre-clear and RNA-IP: Samples were pre-cleared with 15ul/sample Prot A or G Dynabeads, washed 3's in PBS-tween 0.02% pH 7.13 for 15 min at 4ºC. Antibody-prot A or G Dynabead conjugates were then added to each sample and incubated 1.5hrs at 4ºC. Beads were then collected via magnet, supernatant removed and beads re-suspended in RNA wash buffer. Samples were then transferred into clean tubes and washed 3x's with 500ul of RNA wash buffer A. After washing, samples were resuspended in 25ul DEPC treated water

<u>DNAse treatment:</u> Samples were then DNased to remove IP'd DNA. Essentially, reactions took place in a 30ul FV, with 1X Turbo DNAse buffer, 1ul Turbo DNAse (megascript), 0.4ul Rnase inhibitor, along with 25ul IP'd sample. These were incubated at 37ºC for 15min. Samples were then washed once in 500ul RNase wash buffer A and resuspended in 120ul DEPC treated H20. RNA was then purified using standard Trizol purification and eluted in 20ul DEPC H20. Purified RNA was then run on an RNA-bioanalyzer to determine sizes.

<u>RNA library creation:</u> Non-stranded RNA-libraries for Illumina sequencing were then made according to the company protocol. Libraries were then sequenced using HiSeq 2000 100SR

<u>Read analaysis:</u> Reads were then filtered using the following command: fastq_quality_filter -Q33 -q 20 -p 90 -i input_file.fastq -o output_filtered_file.fastq. Reads were then mapped to Release 5 of *D. Mel* genome, counted using Htseq and analyzed with Deseq. To identify repeats, I used Bowtie with a lab-generated index.

<u>Probe generation:</u> Essentially, PCR oligos to regions flanking lncRNAs (Table 13) were made with either T3 or T7 promoters on the 5' ends. Regions were then amplified and probes transcribed and labelled with digoxigenin.

Table 13. PCR oligos used to make lncRNA probes

CR43314:1-23+T7	TTAATACGACTCACTATAGGGAGActtgctcgattctttcagtcaac	CR43314:500ish-+T3	AATTAACCCTCACTAAAGGGAGA GTG CTT ACT AAT AGA ACG TTT TGA G
CR42425 (part of su (ste):-+T7	TTAATACGACTCACTATAGGGAGAaagccggtgttcgacagttcctc	CR42425 (part of su (ste):500ish bp-+T3	AATTAACCCTCACTAAAGGGAGA GGT GAG GAC TTG GGC GAT TTA AG
CR42862_NR_048207:exon 4 start-+T3	AATTAACCCTCACTAAAGGGAGA ttactgagtgagtctcaacgt	CR42862_NR_048207:400 bases into exon4-+T3	TTAATACGACTCACTATAGGGAGAATA AAT GAA TGC TCC GCT GTT C
CR43314:exon1 R-CR43314+T3	AATTAACCCTCACTAAAGGGAGA ctgcccttttacaggccaagg	CR43314:exon1 F-+T7	TTAATACGACTCACTATAGGGAGAAAG TCG CAA AGT GGA ACG CA
CR43314:exon 2 R-CR43314+T3	AATTAACCCTCACTAAAGGGAGA ttgcagaacgaggatgacgaactg	CR43314:exon2 F-+T7	TTAATACGACTCACTATAGGGAGAACA ATT TGG TCT CGT CTT GTC
CR43334:exon1 F-CR43334+T3	AATTAACCCTCACTAAAGGGAGA ggcggtgagccggactcagcac	CR43334:exon1 R-+T7	TTAATACGACTCACTATAGGGAGAATA AAT CTT CTT ACA TAC ATA TG
CR45361:F-CR45361+T3	AATTAACCCTCACTAAAGGGAGA gccagaacacagtcgccgtcttc	CR45361:R-+T7	TTAATACGACTCACTATAGGGAGAGGT GGC ACT ACT CAC TTG GAC AG
CR45916:f-CR45916+T3	AATTAACCCTCACTAAAGGGAGA ggtgtgtgctaaagcaatcc	CR45916:r-+T7	TTAATACGACTCACTATAGGGAGAATG GAC ATT CAC TCG ATT GCT C
CR46003:f-CR46003+T3	AATTAACCCTCACTAAAGGGAGA catgcaaatcgcgcataaattc	CR46003:r-+T7	TTAATACGACTCACTATAGGGAGATGG GAT CAG TAG GAT GGA AAA G
CR45736:+T3 F	AATTAACCCTCACTAAAGGGAGA ACG CAT CAC TCT GAA AAT ACA AG	CR45736:+T7 R	TTAATACGACTCACTATAGGGAGAG TTT AAT GTC TAA CTG GTG ATG ATG
CR42425:+T3 F	AATTAACCCTCACTAAAGGGAGA CAA CCC TTT TAG CAC GTG TCA AAA AC	CR42425:+T7 R	TTAATACGACTCACTATAGGGAGAGCA GAT GGG AAC ACT AGA CTC GA
CR43751:+T3 F	AATTAACCCTCACTAAAGGGAGA CAG CCC AGA ATC GTT CCA ATT TAG	CR43751:+T7 R	TTAATACGACTCACTATAGGGAGAGC TGA AGT GCC GGA GGA G
CR44206 :+T3 F	AATTAACCCTCACTAAAGGGAGA ttccgtggatttcaagtggcaa	CR44206 :+T7 R	TTAATACGACTCACTATAGGGAGATAAACCGCAATAGAGCAAAACTGCC
CR44717:+T3 F	AATTAACCCTCACTAAAGGGAGA TAC AAA TTA TTT AAA TCA GAC GTG T	CR44717:+T7 R	TTAATACGACTCACTATAGGGAGAGA CTT TCA TTT TGC CAC A
CR45347 exon 1:+T3 F	AATTAACCCTCACTAAAGGGAGA CGA AGG ACT TCA GTG CTT TTG	CR45347 exon 1:+T7 R	TTAATACGACTCACTATAGGGAGACGC AAT ATT GTG TTT AGT TTT GGG
CR45911:+T3 F	AATTAACCCTCACTAAAGGGAGA TAA AGT AAC TAA CGT AAC GTT AAG C	CR45911:+T7 R	TTAATACGACTCACTATAGGGAGA AGT GTG GTT GTA AAA AGG C
CR43334:+T3 F	AATTAACCCTCACTAAAGGGAGA CTC AGC ACT TGG GAA TTA TAG TGT GT	CR43334:+T7 R	TTAATACGACTCACTATAGGGAGAAGT TAG AGG TGC CAC TTA ATT GC
CR46003:+T3 F	AATTAACCCTCACTAAAGGGAGA GAT CGA GAA AAG TGA TCT TAA CTT ATT	CR46003:+T7 R	TTAATACGACTCACTATAGGGAGAC TCC AGG CAG CAT TAA AAA T
CR43887:+T3 F	AATTAACCCTCACTAAAGGGAGA CTG CAC TGC AAG CGA TGT TC	CR43887:+T7 R	TTAATACGACTCACTATAGGGAGACTT ATA AAT TTC ATT CTC TAT TGC AAA AAA ATC
CR34335:+T3 F	AATTAACCCTCACTAAAGGGAGA CTC ACA GTG TAT CAA GGG TT	CR34335:+T7 R	TTAATACGACTCACTATAGGGAGAGAC TGG AAT GTG AGA ATA ACC
CR32207:+T3 F	AATTAACCCTCACTAAAGGGAGA CCA CCA CCA TCA TTT TTG AC	CR32207:+T7 R	TTAATACGACTCACTATAGGGAGAGG TAA TCA TTT AGC GAA AGA CCT ATA
CR43482:+T3 F	AATTAACCCTCACTAAAGGGAGA CGA AGC ACG AGA GAA AGA AG	CR43482:+T7 R	TTAATACGACTCACTATAGGGAGAC GTG TAT TTG TAT TTA TTG TAT TCG TTT A
CR42653:+T3 F	AATTAACCCTCACTAAAGGGAGA GGT CCT TTA ATG AAA GAC TAA TGA A	CR42653:+T7 R	TTAATACGACTCACTATAGGGAGATAA CAA ATT TAA CGC TGA GCT
CR43973:+T3 F	AATTAACCCTCACTAAAGGGAGA AGT TTC AAT TTA TTA GAA GTG	CR43973:+T7 R	TTAATACGACTCACTATAGGGAGATT CCC ATT ATT TTA TAT GGT
CR43483:+T3 F	AATTAACCCTCACTAAAGGGAGA GAT GTG TTT TGT TTC TTA ACT TCT AAG	CR43483:+T7 R	TTAATACGACTCACTATAGGGAGATAA ATA TGT GCA AAC AAT TGA TTA AAT
CR32218:+T3 F	AATTAACCCTCACTAAAGGGAGA TAG CAG ATA CCT TTT GTA CGT TGA	CR32218:+T7 R	TTAATACGACTCACTATAGGGAGAAAA TGT AAG AAT TTC ACT TCC GTA TG

CR42722:+T3 F	AATTAACCCTCACTAAAGGGAGA GATTCAATGGGAAGTTTTAAATT	CR42722:+T7 R	TTAATACGACTCACTATAGGGAGAATT GAT GCA AGT CAT GTC CT
CR424722 (actually CR42723):+T3 F	AATTAACCCTCACTAAAGGGAGA GGAAGTTTTAAATTGATGAAACT	CR424722 (actually CR42723):+T7 R	TTAATACGACTCACTATAGGGAGACTC CTT TTT AAA TTG ATG CA
CR32009:+T3 F	AATTAACCCTCACTAAAGGGAGA TTCTTGTTCTTGCAAAATAC	CR32009:+T7 R	TTAATACGACTCACTATAGGGAGAA ATT AGT GAC ATT ACT GCA GAT T
CR43241:+T3 F	AATTAACCCTCACTAAAGGGAGA ACATTTTATATTTCTTTCACCA	CR43241:+T7 R	TTAATACGACTCACTATAGGGAGAAAT TTT TAA TAT TAA CAT TAT GTA TG

RNA-FISH

RNA-FISH was performed as per (2) 'TSA amplification' from Chapter 2.

Table 14. Fly strains used

lncRNA affected	Collection	FlyBase Genotype	Stock List Description	FlyBase ID	Stock Number
CR44717	Bloomington Drosophila Stock Center	y^1 w$^-$; Mi{MIC}CR44717^{MI11084}	y[1] w[*]; Mi{y[+mDint2]=MIC}CR44717[MI11084]		56280
CR46003	Bloomington Drosophila Stock Center	w^{1118}; Mi{ET1}MB07895	w[1118]; Mi{ET1}MB07895	FBst0025359	25359
CR46003	Bloomington Drosophila Stock Center	y^1 w$^-$; Mi{MIC}MI02878/TM3, Sb^1 Ser^1	y[1] w[*]; Mi{y[+mDint2]=MIC}MI02878/TM3, Sb[1] Ser[1]	FBst0036157	36157
CR46003	Bloomington Drosophila Stock Center	w^{1118}; Mi{ET1}MB01149	w[1118]; Mi{ET1}MB01149	FBst0024796	24796
CR45347	Bloomington Drosophila Stock Center	y^1 w$^-$; Mi{MIC}MI03642	y[1] w[*]; Mi{y[+mDint2]=MIC}MI03642	FBst0037357	37357
CR42862	Drosophila Genetic Resource Center	y$^-$ w$^-$; PBac{SAstopDsRed}LL05359 P{FRT(w^{hs})}2A P{neoFRT}82B P{Car20y}96E/TM6B, Tb^1	P{FRT} PBac{DsRed}LL05359	FBst0320191	141573
CR433314	Drosophila Genetic Resource Center	y$^-$ w^{67c23}; P{GSV6}GS14711/SM1	y[1] w[67c23]; P{w[+mC]=GSV6}GS14711/SM1	FBst0323189	205923
CR42862	Exelixis at Harvard Medical School	PBac{RB}CR42862e03102	e03102	FBst1014782	e03102

References

1. Riddle, N. C. *et al.* Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res.* **21,** 147–163 (2011).
2. Smith, C. D., Shu, S., Mungall, C. J. & Karpen, G. H. The Release 5.1 Annotation of Drosophila melanogaster Heterochromatin. *Science* **316,** 1586–1591 (2007).
3. Bernard, P. Requirement of Heterochromatin for Cohesion at Centromeres. *Science* **294,** 2539–2542 (2001).
4. Dernburg, A. F., Sedat, J. W. & Hawley, R. S. Direct Evidence of a Role for Heterochromatin in Meiotic Chromosome Segregation. *Cell* **86,** 135–146 (1996).
5. Fanti, L., Giovinazzo, G., Berloco, M. & Pimpinelli, S. The Heterochromatin Protein 1 Prevents Telomere Fusions in Drosophila. *Molecular Cell* **2,** 527–538 (1998).
6. Chiolo, I. *et al.* Double-Strand Breaks in Heterochromatin Move Outside of a Dynamic HP1a Domain to Complete Recombinational Repair. *Cell* **144,** 732–744 (2011).
7. Jenuwein, T. & Allis, C. D. Translating the Histone Code. *Science* **293,** 1074–1080 (2001).
8. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science* **303,** 669–672 (2004).
9. Grewal, S. I. S. & Jia, S. Heterochromatin revisited. *Nature Reviews Genetics* **8,** 35–46 (2007).
10. Gu, T. & Elgin, S. C. R. Maternal Depletion of Piwi, a Component of the RNAi System, Impacts Heterochromatin Formation in Drosophila. *PLoS Genet* **9,** e1003780–14 (2013).
11. Folco, H. D., Pidoux, A. L., Urano, T. & Allshire, R. C. Heterochromatin and RNAi Are Required to Establish CENP-A Chromatin at Centromeres. *Science* **319,** 94–97 (2008).
12. Maison, C. *et al.* The SUMO protease SENP7 is a critical component to ensure HP1 enrichment at pericentric heterochromatin. *Nature Publishing Group* **19,** 458–460 (2012).
13. Almouzni, G. & Probst, A. V. Heterochromatin maintenance and establishment: Lessons from the mouse pericentromere. *Nucleus* **2,** 332–338 (2014).
14. Riddle, N. C. *et al.* Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res.* **21,** 147–163 (2011).
15. Schotta, G. *et al.* Central role of Drosophila SU(VAR)3–9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J.* **21,** 1121–1131 (2002).
16. Canzio, D. *et al.* Chromodomain-Mediated Oligomerization of HP1 Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly. *Molecular Cell* **41,** 67–81 (2011).
17. Mendez, D. L., Mandt, R. E. & Elgin, S. C. R. Heterochromatin Protein 1a (HP1a) partner specificity is determined by critical amino acids in the chromo shadow domain and C-terminal extension. *J. Biol. Chem.* **288,** 22315–22323 (2013).
18. Peng, J. C. & Karpen, G. H. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol* **9,** 25–35 (2006).
19. Pal-Bhadra, M. Heterochromatic Silencing and HP1 Localization in Drosophila Are Dependent on the RNAi Machinery. *Science* **303,** 669–672 (2004).
20. Rudolph, T. *et al.* Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Molecular Cell* **26,** 103–115 (2007).
21. Gu, T. & Elgin, S. C. R. Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in Drosophila. *PLoS Genet* **9,** e1003780–14 (2013).
22. Yuan, K. & O'Farrell, P. H. TALE-light imaging reveals maternally guided, H3K9me2/3-independent emergence of functional heterochromatin in Drosophilaembryos. *Genes & Development* **30,** 579–593 (2016).

23. Muchardt, C. *et al.* Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. *EMBO Rep.* **3,** 975–981 (2002).

24. Swenson, J. M., Colmenares, S. U., Strom, A. R., Costes, S. V. & Karpen, G. H. The composition and organization of Drosophila heterochromatin are heterogeneous and dynamic. *eLife* **5,** 1445 (2016).

25. Shirai, A. *et al.* Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. *eLife* **6,** e25317–23 (2017).

26. Johnson, W. L. *et al.* RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. *eLife* **6,** 1923–4 (2017).

27. Casanova, M. *et al.* Heterochromatin Reorganization during Early Mouse Development Requires a Single-Stranded Noncoding Transcript. *CellReports* **4,** 1156–1167 (2013).

28. Savić, N. *et al.* lncRNA Maturation to Initiate Heterochromatin Formation in the Nucleolus Is Required for Exit from Pluripotency in ESCs. *Stem Cell* **15,** 720–734 (2014).

29. Rinn, J. L. & Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* **81,** 145–166 (2012).

30. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and Functions of Long Noncoding RNAs. *Cell* **136,** 629–641 (2009).

31. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nature Publishing Group* **20,** 1250–1257 (2013).

32. Tsai, M. C. *et al.* Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* **329,** 689–693 (2010).

33. Pandey, R. R. *et al.* Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Molecular Cell* **32,** 232–246 (2008).

34. Yap, K. L. *et al.* Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Molecular Cell* **38,** 662–674 (2010).

35. Engreitz, J. M. *et al.* The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science* **341,** 1237973–1237973 (2013).

36. Luikenhuis, S., Wutz, A. & Jaenisch, R. Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells. *Molecular and Cellular Biology* **21,** 8512–8520 (2001).

37. Sado, T., Hoki, Y. & Sasaki, H. Tsix defective in splicing is competent to establish Xist silencing. *Development* **133,** 4925–4931 (2006).

38. Shibata, S. & Lee, J. T. Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice. *Current Biology* **14,** 1747–1754 (2004).

39. Ilik, I. A. *et al.* Tandem Stem-Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in Drosophila. *Molecular Cell* **51,** 156–173 (2013).

40. Hall, L. L. *et al.* Stable C0T-1 Repeat RNA Is Abundant and Is Associated with Euchromatic Interphase Chromosomes. *Cell* **156,** 907–919 (2014).

41. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature Publishing Group* **546,** 243–247 (2017).

42. Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature Publishing Group* **471,** 480–485 (2011).

43. Hoskins, R. A. *et al.* Sequence finishing and mapping of Drosophila melanogaster heterochromatin. *Science* **316,** 1625–1628 (2007).

44. Sullivan, B. A. & Karpen, G. H. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nature Publishing Group* **11,** 1076–1083 (2004).

45.   Sun, X., Le, H. D., Wahlstrom, J. M. & Karpen, G. H. Sequence Analysis of a Functional Drosophila Centromere. *Genome Res.* **13,** 182–194 (2003).

46.   Blower, M. D. Centromeric Transcription Regulates Aurora-B Localization and Activation. *CellReports* **15,** 1624–1633 (2016).

47.   McNulty, S. M., Sullivan, L. L. & Sullivan, B. A. Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C. *Developmental Cell* **42,** 226–240.e6 (2017).

48.   Rošić, S., Köhler, F. & Erhardt, S. Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *J Cell Biol* **207,** 335–349 (2014).

49.   Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature Publishing Group* **547,** 241–245 (2017).

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

51.   LOHE, A. R. & Brutlag, D. L. Multiplicity of satellite DNA sequences in Drosophila melanogaster. *Proceedings of the National Academy of Sciences* **83,** 696–700 (1986).

52.   Khost, D. E., Eickbush, D. G. & Larracuente, A. M. *Single molecule long read sequencing resolves the detailed structure of complex satellite DNA loci in Drosophila melanogaster.* 1–35 (2016). doi:10.1101/054155

53.   Aravin, A. A. *et al.* The Small RNA Profile during Drosophila melanogaster Development. *Developmental Cell* **5,** 337–350 (2003).

54.   Alekseyenko, A. A. *et al.* Heterochromatin-associated interactions of Drosophila HP1a with dADD1, HIPP1, and repetitive RNAs. *Genes & Development* **28,** 1445–1460 (2014).

55.   Brown, J. B. *et al.* Diversity and dynamics of the Drosophila transcriptome. *Nature Publishing Group* 1–7 (2014). doi:10.1038/nature12962

56.   Nicholson, A. W. Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdiscip Rev RNA* **5,** 31–48 (2014).

57.   Herbert, C., Dzowo, Y. K., Urban, A., Kiggins, C. N. & Resendiz, M. J. E. Reactivity and Specificity of RNase T1, RNase A, and RNase H toward Oligonucleotides of RNA Containing 8-Oxo-7,8-dihydroguanosine. *Biochemistry* **57,** 2971–2983 (2018).

58.   Kelemen, B. R., Schultz, L. W., Sweeney, R. Y. & Raines, R. T. Excavating an active site: the nucleobase specificity of ribonuclease A. *Biochemistry* **39,** 14487–14494 (2000).

59.   Zhang, P. & Spradling, A. C. The Drosophila Salivary Gland Chromocenter Contains Highly Polytenized Subdomains of Mitotic Heterochromatin. *Genetics* **139,** 659–670 (1995).

60.   Hoskins, R. A. *et al.* The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Res.* **25,** 445–458 (2015).

61.   Jain, M. *et al.* Linear assembly of a human centromere on the Y chromosome. *Nat Biotechnol* **36,** 321–323 (2018).

62.   Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nature Publishing Group* 1–10 (2018). doi:10.1038/nmeth.4577

63.   Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161,** 404–416 (2015).

64.   Chu, C., Quinn, J. & Chang, H. Y. Chromatin Isolation by RNA Purification (ChIRP). *JoVE* 1–6 (2012). doi:10.3791/3912

65.   Simon, M. D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature Publishing Group* **504,** 465–469 (2013).

66.   Quinn, J. J. *et al.* Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat Biotechnol* **32,** 933–940 (2014).

67.   Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161,** 404–416 (2015).

68. Huertas, D., Cortés, A., Casanova, J. & Azoŕn, F. Drosophila DDP1, a Multi-KH-Domain Protein, Contributes to Centromeric Silencing and Chromosome Segregation. *Current Biology* **14,** 1611–1620 (2004).

69. Yue, L. *et al.* Genetic analysis of viable Hsp90 alleles reveals a critical role in Drosophila spermatogenesis. *Genetics* **151,** 1065–1079 (1999).

70. Ku, H.-Y., Gangaraju, V. K., Qi, H., Liu, N. & Lin, H. Tudor-SN Interacts with Piwi Antagonistically in Regulating Spermatogenesis but Synergistically in Silencing Transposons in Drosophila. *PLoS Genet* **12,** e1005813 (2016).

71. Demarco, R. S., Eikenes, Å. H., Haglund, K. & Jones, D. L. Investigating spermatogenesis in Drosophila melanogaster. *Methods* **68,** 218–227 (2014).

72. Cortes, A. & Azorin, F. DDP1, a heterochromatin-associated multi-KH-domain protein of Drosophila melanogaster, interacts specifically with centromeric satellite DNA sequences. *Molecular and Cellular Biology* **20,** 3860–3869 (2000).

73. Huertas, D., Cortés, A., Casanova, J. & Azoŕn, F. Drosophila DDP1, a Multi-KH-Domain Protein, Contributes to Centromeric Silencing and Chromosome Segregation. *Current Biology* **14,** 1611–1620 (2004).

74. Cortes, A. *et al.* DDP1, a single-stranded nucleic acid-binding protein of Drosophila, associates with pericentric heterochromatin and is functionally homologous to the yeast Scp160p, which is involved in the control of cell ploidy. *EMBO J.* **18,** 3820–3833 (1999).

75. Legendre, F. *et al.* Whole Mount RNA Fluorescent in situ Hybridization of Drosophila Embryos. *JoVE* 1–8 (2013). doi:10.3791/50057

76. Shpiz, S., Lavrov, S. & Kalmykova, A. in *Programmed Cell Death* **1093,** 161–169 (Humana Press, 2013).

77. Jandura, A., Hu, J., Wilk, R. & Krause, H. M. High Resolution Fluorescent In Situ Hybridization in Drosophila Embryos and Tissues Using Tyramide Signal Amplification. *JoVE* e56281–e56281 (2017). doi:10.3791/56281

78. Cai, W., Jin, Y., Girton, J., Johansen, J. & Johansen, K. M. Preparation of Drosophila polytene chromosome squashes for antibody labeling. *JoVE* e1748–e1748 (2010). doi:10.3791/1748

79. Tulina, N. & Matunis, E. Control of stem cell self-renewal in Drosophila spermatogenesis by JAK-STAT signaling. *Science* **294,** 2546–2549 (2001).

80. Kiger, A. A., White-Cooper, H. & Fuller, M. T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407,** 750 EP —754 (2000).

81. Lim, C., Tarayrah, L. & Chen, X. Transcriptional regulation during Drosophila spermatogenesis. *Spermatogenesis* **2,** 158–166 (2012).

82. White-Cooper, H. Molecular mechanisms of gene regulation during Drosophila spermatogenesis. *Reproduction* **139,** 11–21 (2010).

83. Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F. & White-Cooper, H. Post-meiotic transcription in Drosophila testes. *Development* **135,** 1897–1902 (2008).

84. Fabian, L. & Brill, J. A. Drosophila spermiogenesis: Big things come from little packages. *Spermatogenesis* **2,** 197–212 (2012).

85. Steinhauer, J. Separating from the pack: Molecular mechanisms of Drosophila spermatid individualization. *Spermatogenesis* **5,** e1041345 (2015).

86. Kost, N. *et al.* Multimerization of Drosophila sperm protein Mst77F causes a unique condensed chromatin structure. *Nucleic Acids Research* **43,** 3033–3045 (2015).

87. Rathke, C., Baarends, W. M., Awe, S. & Renkawitz-Pohl, R. Chromatin dynamics during spermiogenesis. *Biochim. Biophys. Acta* **1839,** 155–168 (2014).

88.   Eren-Ghiani, Z., Rathke, C., Theofel, I. & Renkawitz-Pohl, R. Prtl99C Acts Together with Protamines and Safeguards Male Fertility in Drosophila. *CellReports* **13,** 2327–2335 (2015).

89.   Xing, Y. & Li, W. X. Heterochromatin components in germline stem cell maintenance. *Nature Publishing Group* **5,** 17463 (2015).

90.   Hughes, S. E., Miller, D. E., Miller, A. L. & Hawley, R. S. Female Meiosis: Synapsis, Recombination, and Segregation in Drosophila melanogaster. *Genetics* **208,** 875–908 (2018).

91.   McKee, B. D., Yan, R. & Tsai, J.-H. Meiosis in male Drosophila. *Spermatogenesis* **2,** 167–184 (2014).

92.   McKee, B. D. & Karpen, G. H. Drosophila ribosomal RNA genes function as an X-Y pairing site during male meiosis. *Cell* **61,** 61–72 (1990).

93.   McKee, B. D., Wilhelm, K., Merrill, C. & Ren, X.-J. Male Sterility and Meiotic Drive Associated With Sex Chromosome Rearrangements in Drosophila: Role of X-Y Pairing. *Genetics* **149,** 143–155 (1998).

94.   Tsai, J.-H., Yan, R. & McKee, B. D. Homolog pairing and sister chromatid cohesion in heterochromatin in Drosophila male meiosis I. *Chromosoma* **120,** 335–351 (2011).

95.   Wen, J. *et al.* Adaptive regulation of testis gene expression and control of male fertility by the Drosophila hairpin RNA pathway. [Corrected]. *Molecular Cell* **57,** 165–178 (2015).

96.   Larracuente, A. M. & Presgraves, D. C. The selfish Segregation Distorter gene complex of Drosophila melanogaster. *Genetics* **192,** 33–53 (2012).

97.   Wen, K. *et al.* Critical roles of long noncoding RNAs in Drosophila spermatogenesis. *Genome Res.* **26,** 1233–1244 (2016).

98.   Huang, X., Fejes-Tóth, K. & Aravin, A. A. piRNA Biogenesis in Drosophila melanogaster. *Trends in Genetics* **33,** 882–894 (2017).

99.   LOHE, A. R., HILLIKER, A. J. & ROBERTS, P. A. Mapping Simple Repeated Dna-Sequences in Heterochromatin of Drosophila-Melanogaster. *Genetics* **134,** 1149–1174 (1993).

100.   Piergentili, R. & Mencarelli, C. Drosophila melanogaster kl-3 and kl-5 Y-loops harbor triple-stranded nucleic acids. *Journal of Cell Science* **121,** 1605–1612 (2008).

101.   Ni, J.-Q. *et al.* A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat Meth* **8,** 405–407 (2011).

102.   White-Cooper, H. Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the Drosophila testis. *Spermatogenesis* **2,** 11–22 (2012).

103.   Chen, D. & McKearin, D. M. A discrete transcriptional silencer in the bam gene determines asymmetric division of the Drosophila germline stem cell. *Development* **130,** 1159–1170 (2003).

104.   Levine, M. T., Vander Wende, H. M. & Malik, H. S. Mitotic fidelity requires transgenerational action of a testis-restricted HP1. *eLife* **4,** e07378 (2015).

105.   Redhouse, J. L., Mozziconacci, J. & White, R. A. H. Co-transcriptional architecture in a Y loop in Drosophila melanogaster. *Chromosoma* **120,** 399–407 (2011).

106.   Rastelli, L. & Kuroda, M. I. An analysis of maleless and histone H4 acetylation in Drosophila melanogaster spermatogenesis. *Mechanisms of Development* **71,** 107–117 (1998).

107.   Aravin, A. A. *et al.* Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line. *Molecular and Cellular Biology* **24,** 6742–6750 (2004).

108.   Gärtner, S. M. K., Rathke, C., Renkawitz-Pohl, R. & Awe, S. Ex vivo culture of Drosophila pupal testis and single male germ-line cysts: dissection, imaging, and pharmacological treatment. *JoVE* 51868–e51868 (2014). doi:10.3791/51868

109.   Czech, B. & Hannon, G. J. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends in Biochemical Sciences* **41,** 324–337 (2016).

110. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Current Biology* **11,** 1017–1027 (2001).

111. Gonzalez, J., Qi, H., Liu, N. & Lin, H. Piwi Is a Key Regulator of Both Somatic and Germline Stem Cells in the Drosophila Testis. *CellReports* **12,** 150–161 (2015).

112. White-Cooper, H. Spermatogenesis: analysis of meiosis and morphogenesis. *Methods Mol. Biol.* **247,** 45–75 (2004).

113. Elgin, S. C. R. & Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harbor Perspectives in Biology* **5,** a017780–a017780 (2013).

114. Elgin, S. C. R. & Reuter, G. Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in Drosophila. *Cold Spring Harbor Perspectives in Biology* **5,** a017780–a017780 (2013).

115. Singh, A. K. & Lakhotia, S. C. Dynamics of hnRNPs and omega speckles in normal and heat shocked live cell nuclei of Drosophila melanogaster. *Chromosoma* **124,** 367–383 (2015).

116. Goenka, A. *et al.* Human satellite-III non-coding RNAs modulate heat-shock-induced transcriptional repression. *Journal of Cell Science* **129,** 3541–3552 (2016).

117. Gong, W. J. Loss of Hsp70 in Drosophila Is Pleiotropic, With Effects on Thermotolerance, Recovery From Heat Shock and Neurodegeneration. *Genetics* **172,** 275–286 (2005).

118. Tower, J. Heat shock proteins and Drosophila aging. *Exp. Gerontol.* **46,** 355–362 (2011).

119. Gangaraju, V. K. *et al.* Drosophila Piwi functions in Hsp90-mediated suppression of phenotypic variation. *Nat Genet* **43,** 153–158 (2011).

120. Olivieri, D., Senti, K.-A., Subramanian, S., Sachidanandam, R. & Brennecke, J. The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in Drosophila. *Molecular Cell* **47,** 954–969 (2012).

121. Caudy, A. A. *et al.* A micrococcal nuclease homologue in RNAi effector complexes. *Nature Publishing Group* **425,** 411–414 (2003).

122. Wang, Q. Q., Zhang, Z., Blackwell, K. & Carmichael, G. G. Vigilins bind to promiscuously A-to-I-edited RNAs and are involved in the formation of heterochromatin. *Current Biology* **15,** 384–391 (2005).

123. Huertas, D., Cortés, A., Casanova, J. & Azorín, F. Drosophila DDP1, a Multi-KH-Domain Protein, Contributes to Centromeric Silencing and Chromosome Segregation. *Current Biology* **14,** 1611–1620 (2004).

124. Chintapalli, V. R., Wang, J. & Dow, J. A. T. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat Genet* **39,** 715–720 (2007).

125. Davis, M. P. *et al.* Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution. *EMBO Rep.* **18,** 1231–1247 (2017).

126. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Publishing Group* **18,** 71–86 (2017).

127. Brown, J. B. *et al.* Diversity and dynamics of the Drosophila transcriptome. *Nature Publishing Group* 1–7 (2014). doi:10.1038/nature12962

128. Chen, B. *et al.* Genome-wide identification and developmental expression profiling of long noncoding RNAs during Drosophila metamorphosis. *Nature Publishing Group* **6,** 23330 (2016).

129. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22,** 1775–1789 (2012).

130. Ma, X. *et al.* Characterisation of Caenorhabditis elegans sperm transcriptome and proteome. *BMC Genomics* **15,** 168–13 (2014).

131. Wang, K. C. & Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Molecular Cell* **43,** 904–914 (2011).

132. Chen, L.-L. Linking Long Noncoding RNA Localization and Function. *Trends in Biochemical Sciences* **41,** 761–772 (2016).

133. Johansson, A. M., Allgardsson, A., Stenberg, P. & Larsson, J. msl2 mRNA is bound by free nuclear MSL complex in Drosophila melanogaster. *Nucleic Acids Res* **39,** 6428–6439 (2011).

134. Brown, J. B. *et al.* Diversity and dynamics of the Drosophila transcriptome. *Nature Publishing Group* 1–7 (2014). doi:10.1038/nature12962

135. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22,** 1775–1789 (2012).

136. Ramos, A. D., Attenello, F. J. & Lim, D. A. Uncovering the roles of long noncoding RNAs in neural development and glioma progression. *Neurosci. Lett.* **625,** 70–79 (2016).

137. van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169,** 780–791 (2017).

138. Zhao, J. *et al.* Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Molecular Cell* **40,** 939–953 (2010).

139. Lu, Z., Guan, X., Schmidt, C. A. & Matera, A. G. RIP-seq analysis of eukaryotic Sm proteins identifies three major categories of Sm-containing ribonucleoproteins. *Genome Biol.* **15,** R7 (2014).