Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jeong Hyun Ahn Date

Characterization of Transcription Regulation in *C. elegans*:

The Role of SIG-7 in Coordinating RNA Polymerase Elongation and mRNA Processing

By

Jeong Hyun Ahn
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology

_____
William G. Kelly, Ph.D.
Advisor

_____
Jeremy M. Boss, Ph.D.
Committee Member

_____
Paula M. Vertino, Ph.D.
Committee Member

_____
Victor G. Corces, Ph.D.
Committee Member

_____
Xiaodong Cheng, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Characterization of Transcription Regulation in *C. elegans*: The Role of SIG-7 in Coordinating RNA Polymerase Elongation and mRNA Processing

By

Jeong Hyun Ahn
B.S. Biology 2005
Baylor University
M.S. Biology 2008
Baylor University

Advisor: William G. Kelly, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology
2016

Abstract

Characterization of Transcription Regulation in *C. elegans*:
The Role of SIG-7 in Coordinating RNA Polymerase Elongation and mRNA Processing
By Jeong Hyun Ahn

The elongation phase of transcription by RNA Polymerase II (Pol II) involves numerous events that are tightly coordinated, including RNA processing, histone modification, and chromatin remodeling. RNA splicing factors are associated with the elongating Pol II, and the interdependent coupling of splicing and elongation has been documented in several systems. Much of the coordinated events are mediated by their interactions with the C-terminal domain (CTD) of Pol II. The specificity of the interaction can be obtained by modulating the structure of the CTD by numerous post-translational modifications (PTMs). One class of enzymes that can modulate the CTD is the peptidyl prolyl isomerases (PPIase). Here I characterize a member of the cyclophilin family of PPIases that interacts with Pol II to coordinately regulate transcription elongation and splicing in *C. elegans*. SIG-7 contains multiple functional domains, including the PPI domain and RNA-interacting domains. In embryos depleted for SIG-7, RNA levels for over a thousand zygotically expressed genes are substantially reduced, Pol II elongation is defective, and unspliced mRNAs accumulate. Our findings suggest that SIG-7 plays a central role in both Pol II elongation and co-transcriptional splicing and may provide an important link for their coordination and regulation.

Characterization of Transcription Regulation in *C. elegans*: The Role of SIG-7 in Coordinating RNA Polymerase Elongation and mRNA Processing

By

Jeong Hyun Ahn
B.S. Biology 2005
Baylor University
M.S. Biology 2008
Baylor University

Advisor: William G. Kelly, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology
2016

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Willian G. Kelly, who has been supportive throughout my graduate study. As a mentor, he made me feel comfortable knocking on his door for discussion whenever I need to seek his advice for experiments or other personal matters. He always encouraged me to move forward whenever the times were not looking good. There isn't one word that I can describe how thankful I was to be a part of his lab, but I can surely say that I really enjoyed the days and nights working and felt respected as a colleague. I would also like to thank my committee: Dr. Paula Vertino, Jererm Boss, Victor Corces and Xiaodong Cheng for sincere advice, encouragements and a great patience with my shortcomings.

I also like to thank all of the past and present Kelly lab members for being good friends as well as good colleagues. I will remember many helpful discussions the positive feedbacks we have exchanged. I also like to thank Roger Deal's lab for adding broader perspectives and their expertise that I could have missed if we didn't have a joint lab meeting. Besides the science, they also made my time here more enjoyable.

I like to thank all of my families for their endless support and encouragement throughout the years. I like to thank my parents for keeping my family and I in their prayers. I especially thank my wife, Sara Chang, and my two loving children, Eugene and Yuna, for their incredible understanding for my absence from home for days and nights, and your presence is my motivation to excel.

Table of Contents

CHAPTER1: INTRODUCTION

An overview of transcription regulation

Transcription produces RNA molecules from a DNA template through the enzymatic activities of multiprotein enzyme complex named RNA Polymerase. Unlike prokaryotic organisms which have a single RNA polymerase, three forms of RNA polymerase (I, II, III) are responsible for transcribing different classes of RNAs in eukaryotic organisms (1, 2). Among these, RNA Polymerase II (Pol II) is responsible for transcription of all protein coding genes and some noncoding RNA genes. A defining feature of Pol II is a unique C-terminal domain (CTD) within Rpb1, the largest subunit of the Pol II holoenzyme complex, that is not present in other RNA polymerases (3). The CTD of Pol II is structurally flexible and contains numerous copies of a tandemly repeated heptapeptide with a consensus sequence of Tyr-Ser-Pro-Thr-Ser-Pro-Ser (Y_1-S_2-P_3-T_4-S_5-P_6-S_7). The number of heptapeptide repeats increases with the genomic complexity of an organism ranging from 26 to 52 repeats (4). The residues within each repeat are subjected to different post-translational modifications (PTMs) throughout the transcription cycle (5-8). These PTMs include phosphorylation, methylation and glycosylation. The phosphorylation of Tyr 1, Ser 2, Ser 5, and Ser 7 residues by distinct kinases have been well characterized in many organisms and can serve different functions depending on the residue being modified (4, 8). For example, Tyr 1 phosphorylation by an unknown kinase has been implicated in preventing termination factor binding to CTD, while phosphorylation of Ser 5 by cyclin-dependent kinase 7 (CDK7) is required for the binding of the 5' mRNA capping enzyme (9, 10). Although there is high sequence conservation among the heptapeptide repeats, the CTD in some organisms also contains

non-consensus repeat sequences. One such example is an arginine substitution at position 7 in place of the serine residue. The methylation at Arg 7 of non-consensus repeat 31 of human CTD has been reported to be mediated by coactivator-associated arginine methyltransferase 1 (CARM-1). The exact function of this arginine methylation is not known. Different types of modification can also occur in the same residue. Glycosylation at Ser 5 and Ser 7 have been implicated in opposing the phosphorylation at these residues, and the reduction of *O*-GlcNAc transferase (OGT) by siRNA caused defects in transcription and Pol II occupancy at several B-cell promoters (11, 12). Likewise, different combinations of PTMs at different residues appear to correlate with distinct stages of transcription regulation, which has led to the hypothesis of a "CTD code", in which distinct combinations of phosphorylation marks at CTD may recruit specific factors that are required at particular stages of the transcription cycle (4, 13-17).

Besides the dynamic regulation of phosphorylation status of residues within each heptapeptide, the peptide bond preceding proline at position 3 (Pro3) and 6 (Pro6) can also be regulated by isomerization reaction of peptidyl prolyl isomerase (PPIase), which can change the conformation of the peptide bond preceding proline from *cis* to *trans* or vice versa (18). This kind of structural change can affect the dynamics of neighboring residue's phosphorylation status as well as the interactions between CTD of elongating Pol II and other factors such as splicing factors. I will summarize the findings of peptidyl prolyl isomerase that interact with CTD during transcription as well.

Transcription Initiation/Elongation regulation

The failure to reconstitute transcription *in vitro* using purified Pol II alone led to the discovery of conserved factors called General Transcription Factors (GTFs) that are necessary for a transcription (19). These GTFs, which include TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, form a preinitiation complex (PIC) at the core promoter of the genes with the Pol II holoenzyme (3, 20). The formation of the PIC has been proposed to be achieved by two pathways: a) a stepwise, sequential assembly pathway, or b) an RNA polymerase II Holoenzyme Pathway, where a subset of TFs including TFIID provide an access to the promoter and subsequent recruitment of a preassembled Pol II complex with the remaining TFs and other factors, leading to complete PIC formation at the promoter (19, 21). The assembly of PIC achieved by either pathway commonly requires initial recognition of the core promoter sequence by TFIID, which contains TATA-binding protein (TBP) that bends the TATA-box DNA (22). TFIIB helps to align the TBP bound promoter DNA close to the Pol II active center cleft (23). The assembly of the PIC does not allow transcription initiation to happen right away. A recent structural study of PIC complex in yeast showed that promoter DNA was only associated with GTFs without any contact with Pol II, demonstrating the presence of a fully assembled PIC with the closed complex of the promoter DNA. This supports the previous observation that *in vitro* reconstitution of the PIC complex is not sufficient to initiate the transcription reaction, and that the addition of ATP is required to convert the closed complex to an open complex (24). The transition to open complex requires formation of a 'transcription bubble' within the promoter DNA, achieved by melting of double stranded DNA, and

this requires the ATP-dependent helicase activity of the TFIIH subunit Rad25 (25). The activation of initiation of PIC requires the phosphorylation of Ser-5 of Pol II CTD by another TFIIH subunit CDK7 (26). This phosphorylation of Ser-5 on the Pol II CTD recruits a 5' capping enzyme that adds m7GpppN cap to the 5' end of an emerging nascent RNA, as well as recruiting the Set-1/COMPASS complex, which carries out tri-methylation of H3K4 at nucleosomes near the transcription start site (TSS) (27-30).

Once Pol II is activated for initiation, it can be subjected to further regulation before engaging into productive elongation phase: Pol II "pausing". In this scenario, Pol II translocates ~30-50 bps downstream of the transcription start site (TSS) and pauses. Pol II pausing has been observed using different techniques, and was first observed as a rate-limiting step in heat shock loci in *Drosophila* S2 cells (31, 32). The initial genome wide observation of an enrichment of Pol II and TAF-1 together near the promoter from human fibroblast cells initially was thought to be an evidence that the rate liming step of transcription is the initiation step, since TAF-1 is a subunit of TFIID that is part of PIC complex (33). The accumulation of Pol II near a TSS can result from either Pol II waiting for activation during an initiation step, or from Pol II engaged in elongation but stalled and waiting for regulated release. A study from *Drosophila* combined ChIP-chip and Permanganate footprinting techniques (a technique to detect transcription bubble) to distinguish between these scenarios. They demonstrated that transcriptionally activated Pol II at hundreds of genes was paused near promoter regions, and this pausing mechanism relied on the negative elongation factor (NELF) (34). In human embryonic stem cells, a similar observation of transcriptionally activated and paused Pol II at many developmentally regulated gene promoters were reported supporting the idea that there is

an additional post-initiation regulatory mechanism (29). Using a global run-on and sequencing (GRO-seq) technique that specifically measures the level of transcripts that are generated from transcriptionally active Pol II, it was shown that ~40% of the protein coding genes of mouse embryonic stem cells and embryonic fibroblasts were regulated by this pausing mechanism (35).

Although the occurrence of Pol II pausing in mammals and *Drosophila* are well documented, the mechanisms involved in pausing regulation are still under investigation. Several mechanisms for how pausing is established have been proposed, and excellent reviews are available (36, 37).

Factors affecting Pol II pausing

The conserved motifs found in promoter regions of *Drosophila* genes are well characterized (38, 39). Among them, the initator (Inr), the downstream promoter element (DPE) and the more upstream GAGA elements has been shown to be prevalent in paused genes in *Drosophila* embryos (40, 41). The Inr is a 14bp sequence containing a transcription start site (TSS) that is sufficient for basal transcription in genes without a TATA box, serving as a simple functional promoter (42). DPE is a 7bp sequence that is positioned ~30bp downstream of the TSS of TATA-less promoters in *Drosophila* (43). The spacing between Inr and DPE is critical for binding of $dTAF_{II}60$ and $dTAF_{II}40$, subunits of TFIID, to the promoter of TATA-less promoters in both human and *Drosophila* (43). GAGA elements are DNA sequences rich in $(GA)_n$ often abbreviated as GAGAG motifs found in promoters as well as other regions throughout the Drosophila genome. GAGA elements are recognized by a GAGA factor encoded by *Drosoplila trithorax-like* (*Trl*) gene (44). GAGA factors are known to have both repressive and

active functions, depending on their association with different factors. For example, GAGA factor has been reported to interact with the histone chaperone FACT (facilitate chromatin transcription) complex to induce *Hox* gene expression by destabilizing the nucleosome, making the DNA accessible to the transcription machineries (45). These findings suggest that a particular promoter architecture conferred by these specific DNA elements on paused genes may create a specific landing pad for certain TFs or chromatin factors to reorganize nucleosomes around promoter regions, and to facilitate the recruitment of factors required for pausing. A recent study from *Drosophila* S2 cells showed that the +1 nucleosome is a barrier to Pol II elongation for essentially all genes (46). It may be possible that binding of different transcription factors may affect the nucleosome occupancy of the target gene promoters via mechanisms yet to be discovered. A comparative analysis of GAGA factor and M1BP transcription factor-bound promoters from *Drosophila* genes with paused Pol II indicated that there may be distinct mechanisms of pausing, where the +1 nucleosome directly affected pausing in M1BP bound genes, while GAGA factor bound genes are not affected by the nucleosome occupancy (47).

The DRB sensitivity-inducing factor (DSIF) and the NELF are two well characterized negative factors important for pausing. Both DSIF and NELF co-localize with paused Pol II at the promoter, and their association with Pol II and transcription factors has been proposed to stabilize the paused Pol II complex (48-50). The interaction between DSIF and Pol-II-associated factor 1 (PAF-1) has been shown to be important for regulating elongation in Hela cell extracts (51). Recently, depletion of PAF-1 by RNAi in human cell culture resulted in an increased release of paused Pol II, leading to more

transcription yield and increased CTD Ser-2 phosphorylation of Pol II. The authors proposed that the association of PAF-1 with Pol II stabilizes Pol II in the paused state by preventing interaction with the super elongation complex (SEC) (52). It appears that Pol II pausing requires the coordination of multiple factors in a context dependent manner.

The release of Pol II pausing by P-TEFb

The key factor involved in releasing a paused Pol II into productive elongation is the positive elongation factor b (P-TEFb). Inhibition of P-TEFb by flavopiridol blocks the productive elongation of Pol II in both *Drosophila* and human cells, suggesting that P-TEFb is required for releasing of promoter proximal paused Pol II (53, 54). P-TEFb is a heterodimer complex composed of a catalytic subunit, cyclin-dependent kinase 9 (CDK9), and a regulatory subunit, cyclin T (CYC-T) (55). P-TEFb can be recruited to paused Pol II by several different mechanisms. P-TEFb can associate with c-MYC via its CYC-T subunit during transcription activation and is recruited to c-MYC target gene promoters (56, 57). P-TEFb can also be recruited to target genes directly by its interaction with the BRD4 coactivator or indirectly by interaction with the MED26 subunit of Mediator complex as a component of the super elongation complex (SEC) (58, 59). Once P-TEFb is recruited to target genes, the kinase subunit CDK9 phosphorylates the two negative factors, DSIF and NELF. CDK9 phosphorylates the Spt5 subunit of DSIF, which appears to convert Spt5 into a positive elongation factor that travels with Pol II through the gene body. Phosphorylation of NELF causes NELF's dissociation from elongating Pol II (60-63). It may be possible that phosphorylation of Spt5 weakens the interaction of DSIF with PAF-1, causing PAF-1 dissociation with paused Pol II, and the

association of Pol II and SEC complex is allowed to activate the paused Pol II for elongation. CDK9 can also phosphorylate Ser-2 of the Pol II CTD, which can then recruit the Set2 histone methyltransferase that mediates methylation at lysine 36 of H3 along the gene body during elongation (60, 64, 65).

Factors alleviating the nucleosome barrier during transcription elongation

After Pol II is released from pausing, Pol II has to overcome the nucleosome barrier ahead of elongating Pol II. Examination of the dynamics of Pol II and histone occupancy at inducible *GAL* genes during transcription showed an inverse relationship between Pol II and histone occupancy, and the authors also demonstrated that rapid deposition of histone after each passage of Pol II depends on the activity of the histone chaperone FACT complex (66). The FACT complex is a heterodimer of SPT16/SSRP1 that interacts with the histone H2A/H2B dimer (67, 68). FACT was initially proposed to destabilize the nucleosome to facilitate transcription *in vitro*, but it is now thought that FACT can also assist in reassembly. Similar results were also obtained by other groups showing a similar anti-correlation between transcription activity and nucleosome occupancy in *GAL* genes, but they also showed that transcription-dependent H3 acetylation by GCN5 occurs instead of histone eviction for certain genes. Different mechanisms may thus exist to alter nucleosome dynamics either by eviction, or the acetylation of histones to weaken nucleosome-DNA contacts during transcription activation (69).

Chromatin remodeling complexes have also been implicated in reorganization of the chromatin around promoter regions. A SWI/SNF family remodeler, the

bromodomain-containing ATPase RSC, has been shown to localize to ORFs of transcribed genes *in vivo* and its presence increased elongation in *in vitro* transcription assays. The recruitment of RSC to nucleosomes depends on the acetylation status of the nucleosome by the activity of HATs, both *in vivo* and *in vitro* (70, 71). Another chromatin remodeling protein, Chd1 in *S. cerevisiae,* was also shown to be important for regulating nucleosome dynamics during transcription elongation, acting with factors known to be coupled to elongation, including Spt-5 and PAF-1 (72, 73). These findings demonstrate the need for dynamic nucleosome regulation by concerted action of remodelers and histone chaperones to evict, or capture and reassemble the displaced or altered nucleosomes during transcription elongation.

Besides the challenges posed by nucleosomes during translocation along the DNA template, elongating Pol II also has to cope with the processing of nascent RNA as it is being generated. Newly transcribed pre-mRNAs have to go through multiple processing events during transcription, including 5' capping, splicing and 3' end formation (74). In the following section, I will focus on splicing processes that are coupled to transcription elongation.

Regulation of Co-Transcriptional Splicing

RNA splicing is mediated by large complex, called the spliceosome, composed of ribonucleoprotein (RNP) subunits with many protein cofactors (75, 76). Different components of the spliceosome complex recognize and interact with distinct features of pre-mRNAs during the splicing reaction (Figure 1A) [See these reviews for detailed interactions (75, 77)]. The assembly of the spliceosome complex occurs in a step-wise manner, and the composition of the spliceosome at distinct stages of splicing reaction is

regulated dynamically (Figure 1B). Therefore, these interactions have to be tightly regulated to ensure proper splicing.

The earlier dominant view of this process, which indicated that most splicing occurred post-transcriptionally, has changed since the demonstration that spliceosomes can be observed at splice junctions of nascent RNA by electron microscopy in *Drosophila* chorion genes (78). Although this study provided evidence for co-transcriptional splicing, the functional coupling of transcription and splicing was not addressed. With the recent development of chromatin immunoprecipitation (ChIP) procedures, the co-transcriptional assembly of splicing machinery has been demonstrated in several organisms (79-81). These findings, however, do not exclude the occurrence of post transcriptional splicing, which is actually a dominant mode of splicing in *S. cerevisiae* (82). An *in vitro* study from a human cell line provided a clear example of co-transcriptional splicing, detecting exon-exon junction formation of nascent RNAs before Pol II reached the termination site of genes (83). The development of nascent RNA-seq techniques has verified the occurrence of co-transcriptional splicing at a genome wide level in several studies, and it is now evident that co-transcriptional splicing is a predominant mechanism (84-87). Furthermore, recent studies have indicated that splicing and transcription elongation may be not only temporally linked, but may also be mechanistically coupled and co-regulated.

Chromatin and splicing

Proper packaging of DNA into higher-order structure organizes genomes into functionally distinct structures, termed heterochromatin and euchromatin. Within transcriptionally active euchromatin regions, the organization of nucleosomes can confer

intrinsic properties to genes. Analysis of nucleosome occupancy in mammals and *C. elegans* discovered an evolutionarily conserved nucleosome enrichment in exons compared to introns, and this non-random positioning of nucleosomes does not depend on the transcription status of the genes analyzed (88). The authors argued that the higher GC contents in exon sequence may partially contribute to favorable interactions between exonic DNA sequence and nucleosomes. Independent analysis by another group found similar higher nucleosome occupancy in exons of both human and *C. elegans*, and they reported that histone modification of H3K36me3 may facilitate the exon inclusion during co-transcriptional splicing as they observed higher level of H3K36me3 for included exons while lower level was observed for excluded exons (89). The latter study also reported other histone modifications enriched in exons including H3K79me1, H2BK5me1, H3K27me2, and H3K27me3. However, this study did not provide any mechanistic insight into how such modifications are established and whether they play any role in splicing. There is also evidence for histone variants playing a role in splicing. An analysis of a mammalian specific H2A variant, H2A.Bbd, showed an enrichment of H2A.Bbd in actively transcribed genes (90). Upon depletion of H2A.Bbd, they observed changes in gene expression and a disruption in normal mRNA splicing patterns. The authors further observed specific binding of 15 spliceosome components to H2A.Bbd, correlating with the splicing defect observed.

In addition to non-random nucleosome distribution and incorporation of specific histone variants, specific modifications to nucleosome histones can also impact splicing processes. Dynamic regulation of histone acetylation at K9 and K14 of H3 by histone acetyltransferases (HATs) and deacetylases (HDACs) in intron containing genes were

shown to be important for spliceosome assembly in yeast (91). In HDAC mutants, they observed increased acetylation in coding regions and a correlative persistence of U2 snRNP components Lea1 and Msl1, as well as defects in the association of downstream factors. This study proposed that dynamic regulation of histone acetylation drives some aspect of stepwise spliceosome assembly during transcription. There is also an interesting observation of stimulus dependent epigenetic changes correlating with alternatively spliced exons. A specific local increase of H3K9 acetylation in the chromatin of an alternatively spliced exon 18 in the NCAM gene, in response to depolarization, was shown to promote exon 18 exclusion, and this phenomenon could be reverted upon withdrawal of the stimulus (92). They proposed that local hyperacetylation may enhance the kinetics of Pol II elongation, thereby increasing the frequency of exon skipping. In order to support their rationale, they demonstrated that transcription by a Pol II 'slow' mutant resulted in 6-fold higher occurrence of exon 18 inclusion than WT Pol II, favoring a kinetic model of alternative splicing mechanisms.

Histone methylation also has been shown to affect the efficiency of splicing. The Reinberg lab showed that H3K4me3 is specifically recognized by several factors involved in post-initiation events including the CHD1 chromatin remodeler, U2 snRNP subunit, and PAF1 (93). They further showed that the recognition of H3K4me3 by CHD1 and its association with splicing factor U2snRNP subunit is important for efficient splicing *in vitro* and *in vivo*. The recognition of H3K36me3 by MRG15 and its interaction with polypyrimidine tract-binding protein (PTB) splicing factor at the gene, has also been shown to repress alternative exon inclusion (94). In human and mouse cells, the recruitment of an H3K36me3 methyltransferase, HYPB/Setd2, depends on splicing

events as an inhibitor for splicing significantly reduced gene body region H3K36me3 (95). However, it should be noted that splicing inhibition has been shown to affect Pol II elongation as well as Ser-2 phosphorylation of CTD, which is known to recruit Set-2 to mediate H3K36 methylation (96, 97). Therefore, it is equally possible that reduced Ser-2 phosphorylation due to inhibition of splicing caused the reduced recruitment of Set-2, and therefore reduction in H3K36me3 levels, not splicing itself.

Recent findings suggest that the heterochromatin protein 1 (HP1) family members can play positive roles in transcription (98-100). CBX3, one isoform of HP1 in murine cells, was shown to be enriched within the body of actively transcribed genes particularly at the exon boundaries and was co-immunoprecipitated with phosphorylated forms of Pol II (101). Upon loss of CBX3, they showed 1037 genes displayed RNA splicing defects and a reduction in splicing factor recruitment at tested genes. Interestingly, they did not observe increased H3K9me3 at the genes that are enriched for CBX3 suggesting an H3K9me3-independent role of CBX3. The above data collectively indicate that distinct histone modifications can either directly interact with or indirectly recruit splicing factors to actively transcribed genes to impact the splicing process. However, much further work is needed to better understand the role of chromatin related processes in splicing regulation, since existing studies show correlation, but do not show causation.

The Role of the CTD in Splicing

The CTD of Pol II has been shown to be an important regulator for splicing processes as truncations of the CTD inhibit splicing (102). It was also shown that interactions of splicing factors such as SRp40 and SC35 with the Pol II CTD were lost upon treatment of HeLa cells with DRB, suggesting the phosphorylation of Ser-2 at the

Pol II CTD is critical for CTD/spliceosome interactions (103). There is an ample amount of evidence supporting a recruitment model in which the phosphorylation status of the heptapeptide repeat within the CTD creates a landing pad for many mRNA processing factors, in addition to elongation factors, during transcription.

From a structural perspective, the Pol II CTD can conform to an extended or compact state, depending on the status of phosphorylation at each residue within each heptapeptide (104). When Pol II is unphosphorylated, the CTD is in a more compact state based on electron crystallography (105). Several biochemical experiments showed that phosphorylation of the CTD induces a structural change from a compact state into a more extended conformation that is more sensitive to protease activity (106, 107). These earlier findings about structural flexibility of Pol II CTD fits well with later findings of numerous reported interactions of CTD with many different factors during transcription. I will discuss evidence of co-transcriptional splicing processes that can benefit from structural flexibility of the CTD upon phosphorylation.

The U1 snRNP protein Prp40 was shown to directly interact with hyperphosphorylated CTD via its WW and FF domain (108, 109). Prp40 has been reported to interact with U5 associated protein Prp8 and the branch point-binding protein, so it is possible that Prp40 interactions with the CTD may facilitate an intron bridging process during spliceosome assembly (110, 111). The human ortholog of Prp40, FBP11, was also shown to bind to phosphorylated CTD peptides *in vitro* suggesting an evolutionally conserved function of this splicing factor (112). Polypyrimidine tract binding protein-(PTB) associated splicing factor (PSF) has been shown to interact with both hypo and hyperphosphorylated Pol II (113). Notably, they also showed that PSF and

the $p53^{nrb}$/NonO transcription factor can function as a heterodimeric complex. A PSF interaction with both hypo and hyperphosphorylated Pol II suggests that PSF may get recruited to Pol II during the initiation step and maintain its association to elongating Pol II via its interaction with the $p53^{nrb}$/NonO transcription factor to mediate co-transcriptional splicing. The direct interaction between U2AF65 and CTD has been demonstrated to be phosphorylation dependent, and this interaction was necessary to further recruit PRP19C to the pre-mRNA to activate the splicing reaction (114). The author proposed that the interaction between phosphorylated Pol II CTD and U2AF65 is important for early exon recognition while further recruitment of PRP19C to the CTD via protein interactions mediated through the RS domain of U2AF65 is required for a later activation step of the spliceosome. These examples provide how phosphorylation of CTD may facilitate recruitment of distinct splicing factors at different steps of spliceosome assembly to assist its loading onto newly made nascent RNA during transcription elongation.

The Roles of Peptidyl Prolyl Isomerases During Transcription

As in the examples discussed above, the Pol II CTD can interact with more than 100 factors throughout the transcription cycle (115, 116). How does Pol II achieve such numerous interactions with any specificity? One plausible way is by inducing structural changes through different post-translational modifications. All the residues except prolines within each consensus heptapepetide (Y_1-S_2-P_3-T_4-S_5-P_6-S_7) are subject to different types of modifications mainly phosphorylation, and there are multiple number of repeats ranging from 26 to 52 (4, 8). Theoretically, each Pol II CTD can have 26 to 52 differentially modified heptapeptides, and various combination of each differentially

phosphorylated repeat can generate many possible patterns. In addition to these modifications, each peptide bond preceding the prolines at position 3 and 6 can exist in either a *cis* or *trans* conformation, further increasing the number of possible configuration of Pol II CTD. The variety of combinations of such modifications, and unique secondary structures form the basis of the CTD Code hypothesis, as discussed earlier.

Cis/trans interconversion of the peptide bond preceding the proline is mediated by the activity of peptidyl prolyl isomerases (PPIase). The peptidyl prolyl bond has a partial double bond characteristic making spontaneous interconversion a rate limiting step, and the activity of peptidyl prolyl isomerase can dramatically increase the rate of interconversion by a factor of 10^3-10^6 (18, 117). There are three distinct families with PPIase activity, and their classification is based on specific binding criteria for immunosuppressive drugs that led to the characterization of the first protein of this family in 1984 (118). Later, it was found out that the PPIase activity is irrelevant to its immunosuppressive effect. These three families are the Cyclophilins, the FK506-binding proteins (FKBP), and the Parvulins.

These PPIases are evolutionarily well conserved and ubiquitously expressed. The size of the PPIase repertoire varies from 32 members in *H. sapiens* to 14 members in *S. cerevisiae* [See this reference for details (119)]. Cyclophilins comprise the largest family making up more than 50% of the PPIase repertoire. Each family is composed of proteins with distinct domain architecture in addition to PPIase domain. PPIases are reported to have roles in numerous physiological processes: protein folding, heat shock response, virus assembly, signal transduction, tumor metastasis, cell cycle control, transcription regulation, channel gating, translation and others (118, 120-128). In the following section

I will focus on discussing the roles of PPIases that are reported to interact with Pol II's CTD and hence play roles in transcription regulation.

PPIases and the Pol II CTD

As discussed earlier, the Pol II CTD is a good substrate for PPIase activity, because there are 2 prolines in each of the many heptapeptide repeats. In the vicinity of each proline residues are serine and threonine residues that are subject to dynamic regulation of phosphorylation/dephosphorylation by distinct kinases/phosphatases. A particular *cis/trans* conformation between proline and neighboring residues can either promote or inhibit phosphorylation/dephosphoryation activities. In addition, various conformations of CTD dictated by isomerization reactions of distinct PPIases can affect numerous interactions between Pol II CTD and CTD interacting proteins during the transcription cycle.

The structural flexibility at each proline has been tested in fission yeast *S. pombe* (129). In this study, both proline 3 (Pro 3) and 6 (Pro 6) were independently substituted by alanine. The authors explained that this P3A or P6A substitution limits the flexibility by allowing the peptide bond to adopt only *trans* conformation. In these alanine mutants, both yeast strains showed lethality at 30°C suggesting there is a requirement for interconversion of *cis/trans* conformation at these peptide bonds during normal growth. They also created separate strains carrying P3G or P6G mutations to mimic better conformation flexibility where the *cis/trans* conformation can be equally achieved without the activity of PPIase due to the achiral nature of the substituted glycine residues. Interestingly, the P3G mutation was lethal while the P6G mutation only caused slower

growth, suggesting that there may be a different requirement for structural flexibility at Pro 3 versus Pro 6 within each repeat. It is important to note that conformation flexibility gained by glycine substitution can equally mean a loss of conformation regulation that may be needed for specific interactions. Nevertheless, this study demonstrated that maintenance of a proper balance of *cis* or *trans* conformation at these peptide bonds is important for cell viability, possibly through its role in CTD dependent transcription regulation.

The best characterized PPIases that interact with the Pol II CTD are the Ess1 (*S. cerevisiae*) and its homologs Pin1 (*S. Pombe* and *H. sapiens*), which belong to the parvulin family of PPIase. Ess1/Pin1 is multi-domain protein containing a WW domain and a PPIase domain at its N-terminus and C-termini, respectively (119). An analysis of two *ess1* mutants that harbor mutations within PPIase domain showed a defect in 3' end formation of pre-mRNAs in *S. cerevisiae* (130). This finding demonstrates that the activity of PPIase is functionally important for mRNA processing. The demonstration of Ess1 binding to phosphorylated CTD via its WW domain from several labs provided supporting evidence that the role of Ess1 in mRNA processing is linked to Pol II's CTD (130-133). The observation of genetic interactions between Ess1 and several CTD kinases led to a testable model that Ess1-dependent structural changes may affect the phosphorylation status of the CTD and its corresponding interactions with other factors (134). A functional interaction between Ess1 and the CTD phosphatase Ssu72 has also been reported. This study showed that *ess1* mutants displayed higher levels of Ser-5 hyperphosphorylated forms of Pol II as detected by western blot, suggesting that Ess1 function may be a prerequisite for downstream dephosphorylation activity by Ssu72 at

the Ser-5 residue of the CTD (135). A structural biochemical study of the phosphorylated CTD/ Ssu72 complex confirmed previous findings demonstrating that *cis* conformation of pSer5-Pro6 mediated by Ess1 promotes rapid dephosphorylation of pSer-5 of CTD by Ssu72 (136). It was also shown that dephosphorylation can still occur without the Ess1, but that the phosphatase activity was markedly reduced. A genome wide study showed that Ess1 is required for the Nrd-1 dependent termination of non-coding RNAs by modulating the dephosphorylation activity of Ssu72, which can affect competition between Pcf11 and Nrd1 for CTD binding (127). An analysis of the distributions of Ser-5 phosphorylated Pol II, the 5' capping enzyme Ceg1, TBP, and TFIIB within many genes by ChIP in the *ess1* mutant showed a marked increase near the promoters (126). This data suggest that the activity of Ess-1 facilitates Pol II elongation.

In *S. pombe*, Pin1 was also shown to facilitate the dephosphorylation of pSer-5 by Fcp-1 showing a conserved function in yeast (137). However, the function of Pin1 in humans showed the opposite effect; i.e., the activity of Pin1 is required for blocking the function of Fcp-1. The *pin1* mutant resulted in an accumulation of the hypophosphorylated forms of Pol II likely by an excessive activity of Fcp-1 (138). Another human Pin1 study in HeLa cells showed a similar result of an inhibitory function of Pin1 for the Fcp-1 dependent dephosphorylation of Pol II (139). Although there is a different effect of Pin1 in human compared to yeast, it is consistent that the PPIase activity of Pin1 modulates the CTD phosphorylation level of Pol II in either positive or negative direction.

The Cyclophilin PPIases

Another family of PPIases, the cyclophilins, has also been implicated in the regulation of Pol II. Among many different types of cyclophilin PPIases, this cyclophilin stands out for its unique domain architecture, consisting of an N-terminal PPIase domain, an RNA recognition motif (RRM), and a highly charged arginine-serine rich (RS) domain-like C-terminal end. This RRM and RS combination resembles the serine/arginine-rich (SR) proteins and is a well characterized property of numerous splicing factors (140, 141).

The first characterized gene of this class is *kin241* which is a nuclear-localized protein in *Paramecium tetraurelia*, (142). A mutation in *kin241* resulted in pleiotropic defects, including aberrant cell size and shape, generation time, thermosensitivity and others, and it was shown to be an essential gene (143). Its homolog in *S. pombe,* Rct1 was also shown to be an essential gene, and mutants showed similar pleotropic effects (144). Increased phosphorylation of both CTD Ser-2 and Ser-5 was observed in *rct1* heterozygotes, and defective splicing of meiotic genes was observed. They also showed that Rct1 associates with transcriptionally active chromatin and that transcription of numerous genes was significantly perturbed in *rct-1* heterozygotes. Another study showed that Rct-1 can interact with Cdk-9, Lsk-1 and CTD peptide *in vitro* through its PPIase domain (145).

The *Arabidopsis* homolog AtCyp59 was shown to interact with several SR splicing factors as well as recombinant *Arabidopsis* CTD in vitro (146). Consistent with findings from Rct1in *pombe*, reduced CTD phosphorylation was observed when AtCyp59

was overexpressed. It was also shown that AtCyp59 can bind RNAs *in vitro*. Following up with this finding, the search for RNA targets of AtCyp59 using systematic evolution of ligands by exponential enrichment (SELEX) identified several consensus RNA sequence motifs for AtCyp59 binding such as GGUGCCG, and their analysis showed that 70% of AtCyp59 bound RNA sequences mapped to protein coding genes (147). The authors observed binding of both spliced and unspliced RNAs with AtCyp59 suggesting association of AtCyp59 during transcription. PPIase activity of AtCyp59 was detected *in vitro*, and the binding of RNAs to RRM domain caused a reduction in the PPIase activity of AtCyp59, suggesting an allosteric inhibition of one domain caused by an activity of the other domain.

In my dissertation project, I investigated the role of this conserved PPIase family in *C. elegans*. Prior to my investigations, very little was known about the role of this conserved family of PPIases in this model organism. The *C. elegans* homolog of AtCyp59, Kin24, and Rct-1 is named SIG-7. I performed the first genome-wide study of this class of PPIases, using a GFP tagged recombinant SIG-7 that fully rescued the deletion mutant. As was observed in the other organisms, *sig-7* depletion causes pleotropic mutant phenotypes such as a delay in overall growth rate and other tissue specific developmental defects. I observed that SIG-7 is enriched in transcriptionally active chromatin and associates with Pol II *in vivo*. I found that the pleiotropic defects in *sig*-7 mutants correlates with a global downregulation of transcription. The transcription defects themselves correlate with a Pol II elongation defect, demonstrated by the reduction of Pol II enrichment towards the 3' end of genes downregulated in embryos depleted of SIG-7 function. RNA-seq experiments also revealed a substantially increased

ratio of unspliced transcripts to spliced transcripts, providing an evidence for a role for SIG-7 function in splicing processes. Combining the observations from investigations in multiple organisms, I propose that this class of cyclophilins are recruited to the Pol II CTD through their PPIase domain, and this allows its RRM domain to capture nascent RNAs as they emerge from Pol II. The RNA binding may allosterically induce structural changes that may modulate the activity of its PPIase domain, which may alter the CTD structure and promote or enhance elongation efficiency. The absence of SIG-7 causes defective coupling of co-transcriptional splicing with transcriptional elongation, thereby decreasing the efficiency of both processes. Collectively, the findings from both multidomain parvulin and cyclophilin PPIases provide a good evidence that PPIases help to coordinate the highly complex transcription process via their interactions with and structural modulations of the Pol II CTD. In chapter two, I have provided a detailed characterization of the function of SIG-7 in *C. elegans* that is published in PLOS Genetics as well as other additional information that were not included in that publication.

Further Investigations into Transcriptional Memory

During the course of my investigations into the role of SIG-7 in transcriptional regulation, I also became interested in how the process of transcription creates its own "memory" of gene activity that may guide gene regulation in a heritable manner. In the course of my studies, I observed that the separate transcription-dependent and transcription-independent modes of the chromatin modifications, histone H3 lysine 4 and lysine 36 methylation (H3K4me and H3K36me, respectively) exhibited an

interdependency that was required for normal germline function. In Chapter three of this dissertation, I will present my preliminary results that investigate the synergistic effects that are observed with combined loss of different modes of H3K4 and H3K36 methylation machinery in the germline of *C. elegans*.

Figures

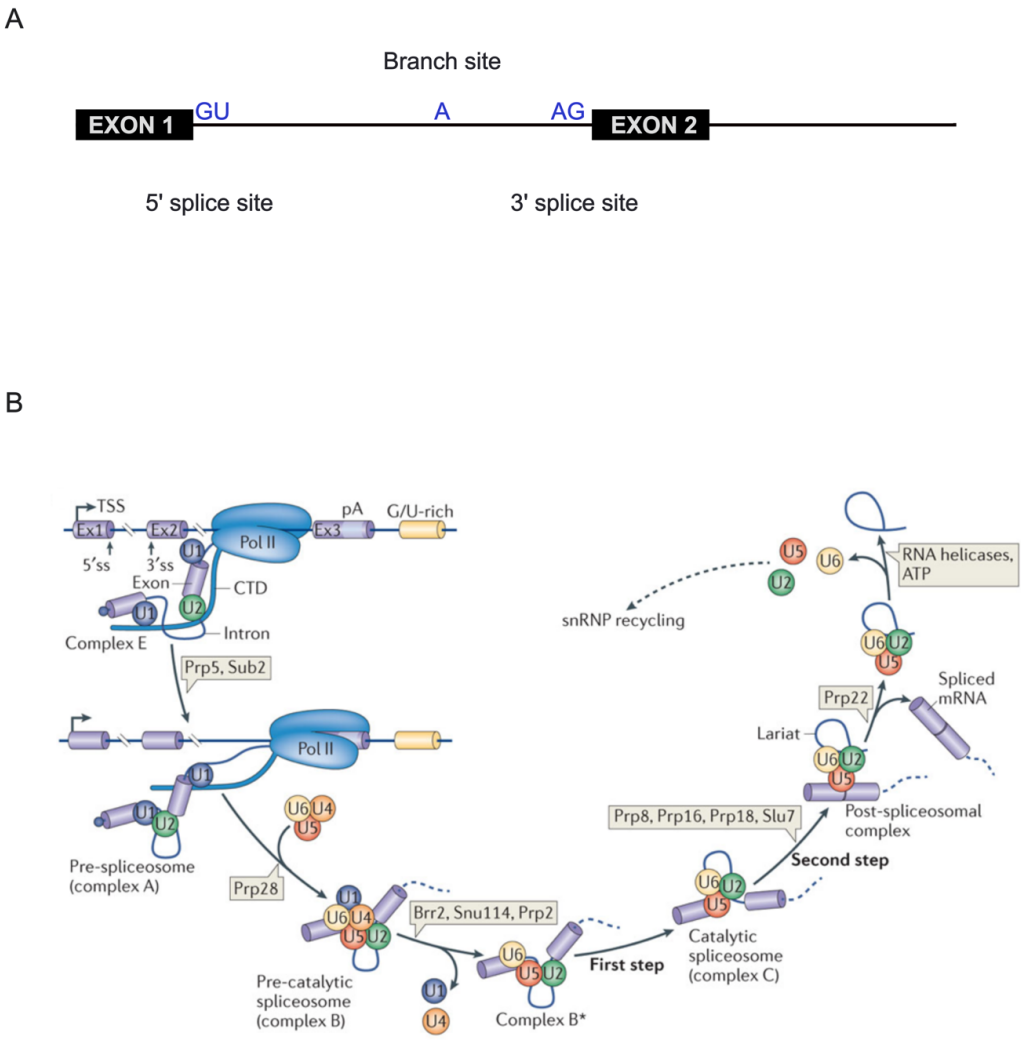


Figure 1. The features of pre-mRNA recognized by spliceosome and step-wise assembly of spliceosome complex

A) The relative position of 5' splice site, branch site and 3' splice sites are indicated. B) The step-wise assembly and dynamic change in composition of each spliceosome complex are illustrated. This figure is reprinted by permission from Macmillan Publishers Ltd:[Nature Reviews: Molecular Cell Biology](75), copyright (2014).

CHAPTER 2: A Conserved Nuclear Cyclophilin is Required for Both RNA Polymerase II Elongation and Co-Transcriptional Splicing in *Caenorhabditis elegans*

Authors: Jeong H. Ahn[1,2], Andreas Rechsteiner[3], Susan Strome[3], and William G. Kelly[1]*

[1]Biology Department, Emory University, Atlanta, GA 30084 USA

[2]Program in Genetics and Molecular Biology, Emory University, Atlanta, GA, 30084 USA

[3]Department of Molecular, Cell and Developmental Biology, University of California Santa Cruz, Santa Cruz, CA 95064

‡ This work has been accepted for publication in PLoS Genetics. With the exception of the bioinformatics analyses, which was performed by Dr. Andreas Rechtsteiner at UCSC, I performed all of the experiments presented in this manuscript.

Abstract

The elongation phase of transcription by RNA Polymerase II (Pol II) involves numerous events that are tightly coordinated, including RNA processing, histone modification, and chromatin remodeling. RNA splicing factors are associated with elongating Pol II, and the interdependent coupling of splicing and elongation has been documented in several systems. Here we identify a conserved, multi-domain cyclophilin family member, SIG-7, as an essential factor for both normal transcription elongation and co-transcriptional splicing. In embryos depleted for SIG-7, RNA levels for over a thousand zygotically expressed genes are substantially reduced, Pol II becomes significantly reduced at the 3' end of genes, marks of transcription elongation are reduced, and unspliced mRNAs accumulate. Our findings suggest that SIG-7 plays a central role in both Pol II elongation and co-transcriptional splicing and may provide an important link for their coordination and regulation.

Introduction

The study of *sig-7* (Silencer In the Germline -7) started with a genetic screen to identify factors required for silencing of repetitive reporter transgenes in the *Caenorhabditis elegans* germline (148). Mutagenized animals were screened for germline GFP expression from a multicopy transgenic array, *ccEx7271*, which is normally silenced in germline tissue (149). One mutant allele recovered in this screen, *sig-7(cc629),* exhibited significant GFP expression in germ cells, but also demonstrated strongly enhanced somatic GFP expression in somatic lineages throughout development, with the increased expression particularly evident in gut, vulva, and uterine tissues (Figure 1). The effect of *sig-7(cc629)* on *ccEx7271* somatic expression suggests that, as in the germ line, this transgenic array is also subject to repression in somatic tissues. The *sig-7(cc629)* mutation also causes a pleiotropic developmental phenotype that is transgene-independent.

Although *sig-7* mutant displays many other transgene-independent developmental phenotypes, we were initially interested in *sig-7*'s role in regulating transgene silencing, particularly its potential involvement in RNA interference (RNAi) pathways. At that time, many RNAi and chromatin factors were discovered to be important in silencing repetitive DNA, and the *ccEx7271* array had been used to screen for such factors (150-155). Therefore, we hypothesized that *sig-7* is a component of an RNAi pathway required for targeting repetitive transgene for silencing in the germline. When we investigated this repetitive transgene array *ccEx7271* by immunostaining for different chromatin modifications, we found out that the array was decorated with heterochromatin marks

when silenced in the germline (See results for details). However, *de novo* euchromatin marks were established when the array was desilenced. This finding suggested that one way of targeting this multicopy repetitive array for proper silencing is via establishing heterochromatin. Based on these preliminary data, we investigated the potential role of *sig-7* in multiple RNAi pathways using available assays, but we could not find any direct evidence for *sig-7* being part of any tested RNAi pathway (Data not shown).

Findings from other labs studying *sig-7* homologs in *Schizosaccharomyces pombe* and *Arabidopsis thaliana* have reported its role in transcription regulation. In order to test whether transcription regulation takes any part in transgene silencing mechanism, we performed RNAi experiment against the largest subunit of RNA Polymerase II (Pol II), *ama-1*, along with *sig-7* and control RNAi using strain carrying *ccEx7271* repetitive transgene array. To our surprise, *ama-1(RNAi)* also resulted in a significant transgene desilencing phenotype similar to *sig-7(RNAi)*. This finding made us shift our focus from investigation of *sig-7's* function in transgene silencing to it role in transcription regulation. The details of how transcription regulation can also affect transgene silencing is discussed in the Results section.

Gene expression regulation is a highly complex process that requires the coordination of many activities, including kinases, DNA and chromatin modifying enzymes, and many classes of noncoding RNAs. (8, 156-158). At the heart of gene expression is the RNA Polymerase II (Pol II) holoenzyme, a multiprotein complex that is dynamically regulated through numerous post translational modifications and the modulation of interactions with general transcription factors and numerous other factors. Kinases regulate transcription progression by modifying the C-terminal domain (CTD) of

Pol II's catalytic subunit and other factors that regulate the transitions accompanying Pol II transcription (7, 24, 37, 159-163). The CTD is composed of a conserved heptapeptide repeat Y^1-S^2-P^3-T^4-S^5-P^6-S^7, and phosphorylation of specific serines, tyrosine and threonines within the repeats correlate with distinct stage of transcription cycle (5, 8, 9, 164-170). Albeit there are some species specific discrepancies, the consensus is that Ser-5 phosphorylation peaks at the 5' end of the gene near the promoter supporting its role in processes related to transcription initiation and shows progressive decrease towards 3' end, while the level of Ser-2 phosphorylation shows progressive increase passed the promoter and peaks at the 3' end of the gene supporting its correlation with processes occurring after transcription initiation step including transcription elongation and termination. For example, Ser-5 phosphorylation by Cdk7, a component of TFIIH basal transcription factor, was shown to recruit 5' capping enzyme that adds a 7-methyl-guanosine cap on the 5' end of emerging pre-mRNA (27). A mutation changing second serine of the CTD repeat to alanine (S2A) caused failure of U2AF65 splicing factor and U2 snRNA as well as reduction in transcription rate (171). A correlation between the modifications of the CTD and distinct mRNA processing steps has been well documented (4, 8, 164, 172).

Specific histone modifications also correlate with discrete steps in Pol II transcription. Active genes have a significant enrichment of histone H3 tri-methylated on lysine 4 (H3K4me3) at transcription start sites, indicating this mark is largely added during promoter interactions, including PIC-associated and early elongation phase events (24, 173-175). Other histone H3 modifications, such as H3 tri-methylation on lysine 36 (H3K36me3) and di-methylation on lysine 79 (H3K79me2), correlate with elongation

(176-178). Although the requirements for**,** and consequences of, CTD phosphorylation and the various histone modifications occurring during the transcription cycle are not completely understood, these conserved modifications have become useful as proxies to assess Pol II regulation and transit through genes in different organisms.

In addition to kinases and histone modifying enzymes, peptidyl proline isomerases (PPIs) can regulate Pol II during transcription progression. The nuclear parvulin family of PPIs direct cis-trans isomerization of prolines in the context of Ser/Thr, such as those found in the Pol II CTD repeat, and their activities are affected by the phosphorylation of Ser/Thr (125, 126, 135, 179). These PPIs are thought to contribute to structural regulation of the CTD, participating in a "CTD code" that controls the recruitment of various factors to Pol II during elongation and transcript processing (135, 136, 180, 181).

The nuclear cyclophilin PPI family, characterized by having an RNA-recognition motif in addition to a PPI domain (RRM), has also been implicated in regulation of Pol II through interactions with the CTD. Members of this highly conserved family include KIN241 in *Paramecium tetraurelia*, AtCyp59 in *Arabidopsis thaliana* and Rct1 in *Schizosaccharomyces pombe* (142, 144, 146, 147). AtCyp59 interacts with Pol II, and its overexpression causes defective regulation of Pol II CTD phosphorylation (146). AtCyp59 also interacts with RNA through its RRM domain and has PPI activity, but whether the PPI domain is required for AtCyp59 function is unclear (147). The *S. pombe* Rct1 interacts also with and affects Pol II CTD phosphorylation (144), yet removal or mutation of the PPI domain has little impact on transgenic rescue of *rct1* mutants*,*

whereas the RRM domain is essential (A. Chang and R. Martienssen, ***personal communication).***

Here we show that SIG-7, a *C. elegans* member of the nuclear cyclophilin family, is essential for normal transcription during embryogenesis. Loss of SIG-7 results in a genome-wide decrease in mRNA production that is correlated with both defective elongation and defective co-transcriptional splicing. Our results identify SIG-7 as a conserved and important factor for the proper coordination of Pol II transcription elongation with co-transcriptional splicing.

Result

Mapping and characterization of *sig-7* mutants

We used the sterile phenotype of *cc629* for positional cloning of its associated genetic locus. Our combined results from polymorphism-based mapping and complementation rescue restricted the *cc629* physical interval to a single predicted gene annotated as F39H2.2 (Figure 2A). We found that dsRNA-mediated interference (RNAi) targeting this locus caused highly penetrant embryonic arrest of progeny when mothers were either injected dsRNA or fed bacteria that produce dsRNA. However, the rare RNAi animals that escaped lethality exhibited post-embryonic defects similar to *cc629* mutants (data not shown). Among these defects is a recessive, fully penetrant hermaphrodite sterility. We observed a clear excess of sperm production – a masculinization-of-the-germ-line (Mog) phenotype – in some, but not all hermaphrodites (Figure 3C). This phenotype results from the failure of the hermaphrodite germ line to transition to a program of oocyte production after an initial period of sperm production during larval

stages (182). *sig-7(cc629)* animals also have a retarded rate of larval development and gross defects in the post-embryonic morphogenesis of their genitalia, e.g., a protruding vulva (Figure 3A). Many *sig-7(cc629)* adults die soon after reaching adulthood from trauma caused by herniation at either the vulva or male tail, as well as a partially penetrant molting defect (not shown).

We sequenced the locus from *cc629* animals and identified a single mutation, a G-to-A base pair transition at the conserved splice acceptor site in the predicted fifth intron of *F39H2.2*, suggesting that *cc629* is an allele that likely affects splicing of this gene. The *F39H2.2(RNAi)* phenotype is more severe than what is observed for the *cc629* allele, however, indicating that *cc629* is not a null allele. Indeed, of ten *F39H2.2* cDNAs from *cc629* animals that were sequenced, intron 5 splicing was only skipped in six, resulting in a transcript predicted to encode a severely truncated protein (Figure 2B). However, the remaining four cDNA clones were not different from wild type. Thus, normal *F39H2.2* splicing can occur at some level in *cc629* animals, despite the mutated intron 5 splice acceptor. This discovery is not unusual, as multiple studies have previously reported the production of both mis-spliced and normally spliced transcripts in *C. elegans* splice acceptor mutants (183-185). In any case, the amount of functional protein produced in *cc629* animals is clearly not enough for normal development, and may be supplemented by maternally loaded protein or mRNA from the heterozygous parent in the F1 homozygotes. Because of the transgene silencing defect in *cc629*, we named the gene *sig-7* (silencer *i*n germline). Because of its homology to cyclophilin-like genes (below), *sig-7* has also been annotated as *cyn-14*.

In the course of these studies, we also obtained a deletion allele, *n5037*, that was isolated in another genetic screen (L. Ma and R. Horvitz, unpublished). *n5037* animals exhibit an early larval stage arrest phenotype, indicating that either SIG-7 has an essential zygotic function limited to post-embryonic development, or that maternal supplies from the heterozygous parent is sufficient to rescue an embryonic requirement but insufficient to support development to later stages. Given the mutant phenotypes and the embryonic lethality observed in *sig-7(RNAi)*, it is likely that there is continuous requirement for SIG-7 during all stages that includes a maternal component during embryogenesis. Indeed, our SIG-7::GFP::3XFLAG expression pattern shows ubiquitous expression in every cell throughout all developmental stage including embryogenesis (Figure 4).

sig-7 encodes a conserved nuclear cyclophilin that is essential for development

SIG-7 possesses an N-terminal peptidyl prolyl isomerase (PPI) domain and an adjacent RNA-recognition motif (RRM) (Figure 2C). The C-terminal region is of low complexity, characterized by the presence of many charged residues including RS and RD dipeptides (Figure 5). SIG-7 is the sole *C. elegans* ortholog of a highly conserved family of PPI and RRM domain-containing nuclear cyclophilins that is found in most eukaryotes from fission yeast to humans, but notably absent from *S. cerevisiae* (186). SIG-7 homologs share greater than 39% overall amino acid identity, most of which is concentrated within the PPI and RRM domains, with the highest degree of sequence identity in the RRM domain (Figure 2D).

In order to study the function of *sig-7* in better detail, it was essential to generate a strain expressing tagged version of SIG-7 that can rescue the mutant phenotype. *sig-7* is

positioned as last gene within a cluster of three genes predicted to comprise operon CEOP1492. *sig-7* is comprised of eight exons, and its open reading frame spans a little over 5 kb. We first tried to generate GFP tagged *sig-7* construct by conventional PCR based cloning strategy. Somehow, PCR reaction always generated about 2kb smaller product than what is predicted in the *C. elegans* genome database. We have sequenced the PCR product and found out that this PCR product contains whole *sig-7* ORF except majority of long intron 1, which is composed of highly repetitive sequences resembling DNA transposon. We are not sure how DNA polymerase has skipped most of intron 1 and continued polymerization reaction. Although this PCR product contained all 8 exons except intron 1, we were concerned that this missing 2kb intron 1 may contain some important regulatory elements that may be important for gene expression. Therefore, we sought for other approach. There was a fosmid clone with pre-engineered GFP::3XFLAG tag on *sig-7* locus available from the TransgeneOme project (187). We found restriction enzyme site that will cut the fosmid into 18kb fragment containing the whole CEOP1292 operon with additional 4.5kb upsteam and 1kb downstream sequences. We integrated this 18kb fragment into defined locus of chromosome II (ttTi5605) using MosSCI technique (188).

After generating this SIG-7::GFP::3XFLAG tagged transgenic line (it will be called SIG-7::GFP), we tested if this SIG-7::GFP can functionally complement loss of function of *sig-7* by crossing this transgenic line into strain carrying *sig-7(n5037)* deletion allele. As expected, this transgene expression fully rescued *sig-7(n5037)* larval arrest phenotype back to nearly WT level. As it was mentioned earlier, *sig-7(cc629)* mutant allele displays other developmental phenotypes like protruding vulva and sterility.

We could not observe any of these phenotypes suggesting this transgene expression is sufficient to replace the endogenous SIG-7 functions. Although we were confident that this functional rescue was solely attributable to expression of SIG-7::GFP, we wanted to verify that it is true. This transgenic line contains an additional copy of four genes including two genes (*F43G9.12* and *syx-17*) in the same operon as well as two additional genes (*mfap-1* and *cpn-1*) located upstream of CEOP1492 operon during cloning procedure due to limited choice of enzymatic sites for digestion from fosmid. To show that functional rescue of mutant phenotype we observes is specifically resulting from the expression of SIG-7::GFP, we performed RNAi experiment targeting for GFP, so that we can specifically knock down expression of SIG-7::GFP. *gfp(RNAi)* in this strain showed 99% embryonic lethality similar to *sig-7(RNAi)* that was done in parallel showing that the functional rescue was due to the expression of SIG-7::GFP, not the expression of duplicated copy of four additional genes that were included in the transgene (Table 1).

SIG-7 Localizes to Transcriptionally Active Chromatin

The studies from SIG-7 homolog in *A. thaliana* AtCyp59 and *S. pombe* Rct1 have been reported to interact with the CTD of Pol II and have roles in transcription regulation (144, 146). SIG-7::GFP localizes to the nucleus in all tissues including germ cells (Figure 4). In adult germ cells, the localization overlaps with DNA in mitotic and meiotic nuclei, but with a somewhat broader distribution than DAPI-stained chromatin (Figure 6). In diakinetic oocytes, however, SIG-7::GFP loses its chromatin association and becomes diffuse within the nucleoplasm (Figure 4D). This transition correlates with the loss of Pol II from chromatin, and the presumed global cessation of transcription in

oogenesis (151). We further examined the correlation between SIG-7 staining and transcriptionally active chromatin in meiotic germ cells. Transcriptional activity is repressed on the X chromosomes relative to autosomes during *C. elegans* meiosis, and the X chromosomes are easily identified by their significantly lower levels of AMA-1 (the catalytic subunit of *C. elegans* Pol II) and H3K4me2, as detected by antibody probes (151). SIG-7::GFP localization in meiotic chromatin exhibits the same pattern as anti-AMA-1 and anti-H3K4me2, i.e., abundant on all autosomes and depleted from the X chromosomes (Figure 6A and B). These data, coupled with the *sig-7(RNAi)* phenotypes, indicate that SIG-7 plays an essential role in development, and this role correlates with active transcription, similar to its plant and fission yeast orthologs.

SIG-7 is essential for normal gastrulation during embryonic development.

We next examined *sig-7(RNAi)* embryos for defects in gastrulation, a sensitive indicator of zygotic transcription defects in embryos. Gastrulation in *C. elegans* consists of the inward migration of a few peripheral cells, including the P4 cell, which is the progenitor of the two primordial germ cells named Z2 and Z3 (189, 190). This process is largely completed by the ~80 cell-stage, with P4 having moved in and subsequently divided to yield its internally localized Z2/Z3 daughters, which are readily identified using antibodies that recognize germ granules (Figure 7A; L4440 control). Zygotic gene activation is required for gastrulation, and disruption of embryonic Pol II or other essential transcription activities results in Z2 and Z3 being born at the periphery instead of the interior of the embryo (Figure 7A; *ama-1(RNAi)*). RNAi targeting of either *ama-1* or *sig-7* caused a highly penetrant gastrulation phenotype, yielding 92.5% and 86.15%

gastrulation-defective embryos, respectively (Figure 7A and B). This suggests that SIG-7 is required for normal zygotic transcription during early embryonic development.

The *sig-7(RNAi)* gastrulation phenotype is indicative of a defect in zygotic transcription, but could be due to inactivation of one or a few genes specifically involved in gastrulation. As a first test for a more widespread defect, we quantified in *sig-7(RNAi)* embryos transcript levels of a panel of genes with strict zygotic expression (191-193). A significant decrease in level was observed for all tested zygotic transcripts in *sig-7(RNAi)* embryos (Figure 7C). This decrease was substantial, albeit not as dramatic as that observed in *ama-1(RNAi)* embryos. Thus, both molecular and phenotypic data indicate that loss of SIG-7 activity leads to significant defects in zygotic transcription, and the effect appears to be widespread.

SIG-7 physically interacts with Pol II *in vivo*

We next tested whether SIG-7 and Pol II interact *in vivo* in *C. elegans*. We tested for physical interactions between SIG-7 with Pol II in *C. elegans* by immunoprecipitating AMA-1 from transgene-rescued *sig-7*(*n5037)* animals, followed by probing western blots of the precipitated material with anti-FLAG antibodies to detect SIG-7::GFP::3xFLAG. SIG-7 was detected in the anti-AMA-1 precipitate (Figure 8). Likewise in reciprocal experiments, AMA-1 was detected in anti-SIG-7 precipitates (Figure 9). These results indicate that SIG-7, like its plant and yeast counterparts, interacts with Pol II *in vivo*. The stoichiometry suggests that this interaction is either between a subset of each protein in the cell and/or is transient and dynamic.

***sig-7* RNAi causes a global decrease in embryonic transcript levels**

To further explore the extent of SIG-7's role in gene expression, we next performed RNA-seq on *sig-7(RNAi)* embryos versus L4440 RNAi control embryos (Figure 10). The results verified that *sig-7* RNAi causes a global defect in embryonic gene expression. Of the 45,627 annotated genes (including non-coding RNAs, etc.), 10,703 had sufficient read representation for further analysis. Of these, 3,045 genes displayed differential expression between *sig-7(RNAi)* and L4440 control RNAi samples with a statistical significance of $q \leq 0.05$ (TableS1). Among these genes, roughly five times more genes exhibited a two-fold or greater decrease in *sig-7(RNAi)* than the number of genes exhibiting a two-fold or greater increase (1549 vs 362, respectively) (Figure 10A). We sorted these genes into gene categories based on published evidence for either zygotic expression during embryonic development ("soma-specific", "embryo–expressed", and "X-linked"), exhibiting "ubiquitous" expression, or having enriched or restricted expression in the germline ("germline-enriched" and "germline-specific", respectively) (194). X-linked genes show a distinct bias for either having weak expression in germ cells or only being expressed in somatic lineages (195). Of the 1549 genes that exhibited at least a two-fold decrease in RNA abundance after *sig-7* RNAi, genes categorized as soma-specific, embryo-expressed, or X-linked were significantly over-represented, and genes categorized as germline-expressed were significantly under-represented (Figure 10B, left panel). This pattern was reversed for the 362 genes that exhibited at least a two-fold increase after *sig-7(RNAi)*: germline-expressed genes, including ubiquitous and germline-enriched genes, were significantly over-represented (Figure 10B, right panel).

These results suggested a potential difference in germline and somatic requirements for SIG-7, which is of interest since other aspects of transcription regulation differ between soma and germline in *C. elegans* (196, 197). However, the germ versus soma differences we observed may also be due to lower efficiency of RNAi in adult germlines compared to the embryos produced by those germlines. We tested 12 genes with known germline-restricted expression by qRT-PCR after extended *ama-1* or *sig-7* RNAi (exposing worms for 55 hours post L3 larval stage). Extended RNAi resulted in significant reduction of germline-expressed genes in both *sig-7* and *ama-1* RNAi adult animals, suggesting that SIG-7 is also required for efficient RNA production in larval and adult germ cells (Figure 11). These longer exposure conditions were not employed in our genomic analysis of embryos, because the resulting sterility in the treated animals prevented analysis of embryonic defects.

We also considered whether the embryonic arrest phenotype was skewing the effect on embryonic loci, since we compared *sig-7(RNAi)* embryos that mostly arrest at ~200-300 cells with control embryos that can continue to develop. The impact of stage differences on our results is probably low. The embryos used in our experiments were isolated from young adults with developing embryos in their uterus; these embryos are highly enriched for stages prior to the *sig-7(RNAi)* arrest point. Indeed, analysis of the developmental stage distributions among embryos from independent RNAi experiments showed the expected bias for early stages (e.g. Figure 12 and data not shown).

We further examined this issue by comparing the affected genes from our experiments with those analyzed in a landmark study examining transcript dynamics in early *C. elegans* embryos (191). The developmental time points assayed in that study

were all earlier than the embryonic arrest caused by *sig-7* RNAi. We focused on three narrowly defined gene sets defined by their dynamics in embryos: "strictly maternal" (expressed in the ovary and degraded in the early embryo), "maternal/embryonic" (expressed in both ovary and embryos), and "strictly embryonic" (expressed only by the embryo with no maternal contribution). Of the "strictly embryonic" genes that showed >2-fold changes in expression with *sig-7* RNAi, 328/339 (96.7%) were down-regulated and only 11/339 were significantly up-regulated (Figure 13). "Maternal/embryonic" and "strictly maternal" genes showed less bias, with 202/310 (65%) and 77/163 (47.3%) showing down-regulation, respectively, with *sig-7* RNAi. The increase of several strictly maternal genes in *sig-7(RNAi)* embryos was verified by qRT-PCR (not shown). The increased abundance of strictly maternal RNAs in *sig-7(RNAi)* embryos may be an indirect effect of defective zygotic transcription-driven development, causing impaired degradation of maternal RNAs (198).

sig-7 RNAi causes changes in RNA processing

Our RNA-seq analyses also suggests a role for SIG-7 in RNA splicing. Cuffdiff algorithm have multiple functions to analyze a several aspects of gene expression pattern difference between samples (199). The analysis we did was limited to differential expression testing. This particular test generates 4 different output files named as; isoform_exp.diff, gene_exp.diff, tss_group_exp.diff, and cds_exp.diff. These files contain results of testing for differences in the level of spliced transcripts, genes, primary transcripts, and coding sequences respectively. All files contained very similar list of genes that are approximately 90% identical to each other (Table S1). This results support

that most of genes that are differentially expressed between L4440 and *sig-7(RNAi)* are reflecting defects in general transcription as well as mRNA processing. Upon closer examination, many of these *sig-7(RNAi)*-dependent "isoform differences" appeared to be caused by decreases in exon reads without a corresponding decrease apparent in intron reads (Figure 14A). Genome-wide analyses also revealed this trend: among genes showing decreased expression in *sig-7(RNAi)* embryos, the average exon read coverage per gene showed the expected significant decrease, but the intron read coverage showed no significant change relative to controls (Figure 14B). Thus, although the amount of RNA for these genes was decreasing, the ratio of intron to exon reads for these RNAs appeared to increase. One limitations with RNA-seq method in respect to analysis for intron read count is that many reads that are not unique to each genomic locus are discarded before analysis. These discarded reads are not just random junk. They are just lacking specificity with statistical confidence, and many of intronic sequences are repetitive or can be occasionally found to have partial matches to multiple loci in the genome when BLAST search is performed. The reason why this can be important is that the chance of capturing the reads from intron region of pre-mRNA in any sample is very low due to the quick nature of mRNA processing occurring during transcription in the cell. So, the information we can obtain regarding the amount of intron read can be limited to only subset that contains specific sequences that can pass the QC step before analysis. One way to overcome this potential technical problem is by confirming the results by qRT-PCR. For qRT-PCR, there is no selection and discarding steps for specific reads. Any cDNA sequences that are generated from present RNA molecules will serve as a template for the reaction. Therefore, we decided to confirm this by performing qRT-PCR

for several genes that are down-regulated from our RNA-seq experiment (Figure 15). We used primer sets that span intron-exon junctions and exon-exon junctions to distinguish unspliced primary transcripts (pre-mRNAs) from spliced, mature mRNAs (mRNAs), respectively. The six embryonic genes tested (*sdz-27, sdz-28*, *epi-1, sqd-1, vet-2, end-1*) confirmed the markedly reduced mRNA levels in *sig-7(RNAi)* embryos observed by RNA-seq. All six also showed significantly increased levels of unspliced RNAs, confirming that the reduced transcripts present in *sig-7(RNAi)* embryos are enriched for defectively processed RNAs (Figure 14A and 15).

In contrast, RNAs from "upregulated" genes, such as strictly maternal genes, showed the opposite trend. RNA-seq results for these genes showed an increase in exon reads, yet their intron reads stayed relatively the same in *sig-7(RNAi)* embryos, suggesting a relative enrichment for spliced RNAs relative to controls (Figure 14B). Indeed, intron sequences could not be detected by qRT-PCR for two maternal genes tested (Figure 15B). This result is also consistent with abnormal persistence of fully spliced maternal products, resulting in an apparent enrichment for exon reads relative to intron reads compared to controls.

Reads from sequences 5' to the first exon of many of these downregulated genes also increased in *sig-7(RNAi)* embryos (Figure 14A and 15A). These reads represent 5' "outrons", which like introns are normally removed from the primary transcripts. In addition to the removal of intron sequences by *cis* splicing of exons, *C. elegans* exhibits co-transcriptional *trans*-splicing, in which a common "splice leader" transcript serves as a 5' splice donor, leading to a common 5' exon that is present on the majority of mRNAs (200-202). In *C. elegans*, approximately 70% of mRNAs are reported to be trans-spliced

(203). The outron reads thus represent 5' nascent transcript sequences that are normally removed by trans-splicing and replaced by splice leader sequences during transcription. Indeed the 5' reads enriched in *sig-7(RNAi)* embryos mark the transcription start sites (TSSs) recently identified by GRO-seq and related methods (173, 204, 205). The retention of RNA-seq reads corresponding to outrons indicates that depletion of SIG-7 causes defects in both *cis*- and *trans*-splicing, the latter of which is only known to occur co-transcriptionally.(206, 207). Since SIG-7 interacts with Pol II, this suggests that SIG-7 plays an important role in transcription-coupled RNA processing events.

sig-7 **RNAi causes a global change in Pol II occupancy and distribution within gene bodies**

Numerous reports indicate that co-transcriptional splicing is mechanistically coupled to Pol II elongation, and that defects in co-transcriptional splicing can affect Pol II elongation (83-85, 208-212). We therefore analyzed the genome-wide distribution of Pol II by anti-AMA-1 ChIP-seq in *sig*-7*(RNAi)* and control embryos. There was strong correlation with the RNA-seq data; i.e., the same genes and gene classes showed similar patterns in both changes in RNA and Pol II occupancy (Figures 16 and 17). We next performed metagene analyses of the Pol II distribution within the bodies of genes separated into five classes based on published expression data (Figure 18). Genes classified as either "soma-specific" or "ubiquitous" showed substantial changes. The 3' enrichment of Pol II observed in controls was significantly reduced in these genes, with 3' depletion the most obvious with the "soma-specific" class (Figure 18). 5' localization

was also reduced for the "soma-specific" class, but the effect was less than the 3' reduction.

In contrast, genes classified as "germline-enriched" showed little change in Pol II distribution. This result indicates that, as with the RNA-seq data, there is a disproportionate effect of *sig-7(RNAi)* on genes expressed in embryos, including an effect on steady-state localization of Pol II within gene bodies. The lack of effect on "germline-enriched" loci is not as easy to ascribe to reduced RNAi efficiency in parental germlines compared to embryos, since many of these genes are also transcribed in the embryos being assayed. It is not obvious why genes generally expressed in germ cells, and in many other tissues, might be less sensitive to loss of SIG-7 function than those expressed only in embryos. This may be related to the different modes of Pol II regulation that we have observed for germline-expressed genes and soma-restricted genes, the latter of which involve tissue-specific modes of gene regulation (196)

A transcription defect could indirectly cause splicing defects by reducing the production of essential splicing factors. This seemed unlikely, since splicing factors in early embryos are available from maternal stores and thus would fall into the class of genes either unaffected or slightly enriched by *sig-7* RNAi. Indeed, our RNA-seq data confirmed this: of 18 conserved *C. elegans* splicing factors (213) for which significant RNA levels could be detected in control embryos, 10 factors showed a slight increase in *sig-7(RNAi)* embryos, and 8 factors showed no significant difference between *sig-7(RNAi)* and control (not shown). Thus, the splicing defects in *sig-7(RNAi)* embryos are unlikely to be due to reduced expression of splicing factors and are instead likely to be directly due to defects in transcription coupled processing.

Depletion of SIG-7 causes a reduction in Pol II isoforms and histone modifications associated with transcription elongation

The decrease in Pol II at the 3' end of gene bodies observed by ChIP-seq suggested that *sig-7* RNAi affects the elongation phase of transcription. The phosphorylation of specific residues in the Pol II CTD correlates with different stages of the transcription cycle; e.g., Ser-5P correlates with initiation and Ser-2P increases with elongation (4, 5, 8, 197, 214). We assessed the relative abundances of these CTD phospho-epitopes in *sig-7(RNAi)* and L4440 control embryos by western blot analysis using monoclonal antibodies specific for the different phosphorylated isoforms of AMA-1 (Figure 19A). We observed similar levels of anti-AMA-1 signal in experimental and control lanes, indicating that SIG-7 depletion has little effect on embryonic AMA-1 protein levels. The amount of hypo-phosphorylated Pol II (hypo-phos; 8WG16) was variable between experiments but often higher in *sig-7(RNAi)* embryos relative to controls. Importantly, whereas the Ser-5P epitope was largely unchanged, a significant decrease in levels of Pol II Ser-2P was consistently observed in *sig-7(RNAi)* embryos (Figure 19A and C). The decrease in Ser2-phosphorylated Pol II is consistent with the decreased 3' Pol II profile observed by ChIP-seq, and indicates that the elongation phase of transcription is altered in *sig-7(RNAi)* embryos.

In yeast, the addition of Ser-5P to the CTD by TFIIH correlates with recruitment of the histone H3K4-specific methyltransferase Set1, which in turn leads to an enrichment of H3K4me3 in nucleosomes near the promoter (30, 215, 216). Elongation and increased phosphorylation of Ser2 in turn correlates with recruitment of the H3K36 methyltransferase Set2 and a resulting enrichment of H3K36me3 within the body of the

gene as elongation proceeds (30, 162, 217). H3K79me2 is also produced within gene bodies during Pol II elongation (218-220). We further examined *sig-7(RNAi)* and control embryos by western blot analysis, using antibodies specific for H3K4me3, H3K79me2, or H3K36me3 and compared these to total histone H3 (Figure 19B and C). We observed a slight decrease in H3K4me3 and a substantial decrease in H3K36me3 and H3K79me2 in *sig-7(RNAi)* embryos. Elongation-dependent histone modifications are also thus disproportionately affected by *sig-7(RNAi)*. We also looked at H3K4me3 and H3K36me3 marks in mixed stage embryos by immuno-staining. Consistent with our western blot result, we observed slight decrease in the level of H3K4me3 most eminently in ~250 cell stage while H3K36me3 levels are observed to be decreased in all stages after ~60 cell stage (Figure 20). The H3K36me3 in the early embryo is predominantly provided by MES-4, a transcription-independent H3K36 methyltransferase, whereas transcription-dependent H3K36me3 predominates in later stages (51, 101). The minimal effect on H3K4me3, a promoter-proximal mark coincident with a more dramatic loss of transcription-dependent H3K36me3, a mark enriched toward the 3'end of transcribed genes, is consistent with a role for SIG-7 in normal Pol II elongation.

In summary, depletion of SIG-7 from *C. elegans* embryos causes a developmental arrest due to widespread defective splicing accompanied by a global decrease in transcription of genes required for normal embryogenesis. This transcription defect correlates with a marked decrease in the 3' distribution of Pol II correlating with decreases in Pol II CTD phospho-epitopes and chromatin modifications that are hallmarks of elongating Pol II. Furthermore, SIG-7 physically associates with Pol II *in vivo*, is enriched in active chromatin in patterns consistent with association with active

transcription, and loss of SIG-7 causes defects in co-transcriptional splicing. SIG-7 is thus required for both transcription and splicing, and may help coordinate both processes to allow for efficient mRNA production in the nucleus.

Revisiting transgene desilencing in *sig-7(RNAi)* or *sig-7(cc629)* mutant

One of many reasons that make *C. elegans* as a great model organism is its simple method to generate transgenic worms by simply injecting expression vectors into gonad. When such expression vectors are injected into a gonad, these vectors go through a recombination process that forms a repetitive array containing several hundred copies of the injected plasmid. However, this repetitive array gets quickly targeted for silencing in the germline. Transgene silencing in the *C. elegans* germline presented obvious difficulties to researchers studying the function of germline-expressed genes, but it also provided an opportunity to discover important biological processes involved in silencing mechanism, such as RNAi pathways and chromatin factors.

In order to understand the mechanism of transgene silencing, we analyzed several well-characterized chromatin modifications. Since *ccEx7271* repetitive array is targeted for silencing, we hypothesized that the repetitive array chromatin will be enriched for heterochromatin marks. Therefore, we performed immunostaining for H3K9me3. As expected, it was enriched for H3K9me3 in maturing oocytes, where transgene arrays can be visualized as small DAPI positive bodies that are smaller than chromosomes (Figure 21). We also analyzed for euchromatin mark H3K4me2. H3K4me2 marks were enriched in all chromosomes, but *ccEx7271* did not show any enrichment (Figure 21). We then wanted to test if these heterochromatin marks will go away when this array is desilenced.

As a positive control, we performed RNAi against known factors including *dcr-1*, *mes-4* and *mut-7*. RNAi performed against all three genes showed consistent results. *ccEx7271* showed an enrichment of H3K9me3 with *de novo* methylation for H3K4me2 (Figure 21). Due to the absence of oocytes from Mog (masculinization-of-the-germ-line) phenotype of *sig-7* mutant or *sig-7(RNAi)*, we could not test directly if this occurred in *sig-7* mutant as well. However, knowing that all these genes tested are known to have distinct functions, we concluded that the changes in chromatin modifications we observed is likely the general result when the *ccEx7271* repetitive array is desilenced.

We also tested if this appearance of euchromatin mark correlate with increased transcription of the reporter genes in ccEx7271 array when it is desilenced. We performed qRT-PCR using primers specific to GFP sequence. As expected, we saw about 1.5 fold increased expression of GFP reporter transcript in *sig-7(RNAi)* compared to *L4440(RNAi)* control (Figure 22A). This result is consistent with stronger GFP reporter expression we observed in *sig-7(RNAi)* or *sig-7(cc629)* mutant worms. However, this presents a conundrum: how would a defect in SIG-7 function, which we've shown is involved in coordinating splicing and transcription elongation, lead to *increased* expression of the GFP transgene? To assess this, we performed RNAi against largest subunit of RNA Pol II, *ama-1*, in parallel to see how generally knocking down transcription affected the GFP reporter expression from *ccEx7271* array. To our surprise, *ama-1(RNAi)* also resulted in a similar level of desilencing to what we observed in *sig-7(RNAi)* (Figure 23). Therefore, a general defect in transcription can result in increased expression of a reporter in a high copy, repetitive array.

One way in which decreased overall transcription could yield increased expression from repetitive arrays is through decreased expression of genes required for heterochromatin formation. Indeed, qRT-PCR for several important genes in RNAi pathway showed reduced transcript abundance for all tested genes except *ergo-1* (Figure 22B). It is thus likely that more copies of the reporter in the array become available for the transcription machinery, which, despite its overall reduction in efficiency, could produce a net increase in reporter expression in germ cells. Therefore, we concluded that the transgene desilencing phenotype of *sig*-7 mutants most likely results from defects in general transcription, where SIG-7 is playing a direct role.

nrde assay reveals that the presence of the genetic balancer chromosome is somehow mimicking the NRDE function via unknown mechanism

As it was discussed in previous section of transgene desilencing result, we became convinced that SIG-7 is not directly functioning as a component of RNAi pathways. However, initial experiments suggested it was involved in what are called "nuclear RNAi" pathways in *C. elegans* (221). However, we have looked at several RNAi pathways using available assays during the course of study. Although most of the tested assays gave inconclusive results for *sig-7*'s involvement in tested RNAi, the result from nuclear RNAi defective (nrde) assay is worth discussing, as the result as this can be a valuable reference for our lab.

In *C. elegans*, there are multiple RNAi pathways carrying out distinct biological functions (222). However, some core components are shared among the distinct RNAi

pathways, so knocking down one RNAi pathway often enhances the efficiency of the other RNAi pathways by increasing the availability of the shared core components.

The Kennedy lab developed a "Nuclear RNAi Defective" (*nrde*) assay that can be easily used to assess whether the candidate gene is a component of a nuclear RNAi pathway, which leads to transcriptional repression, versus the post-transcriptional, cytoplasmic RNAi pathways (96). This assay takes advantage of the "synthetic multi-vulva" genes *lin-15a* and *lin-15b*, in which disruption of *both* is required for a multi-vulva (Muv) phenotype. In this *nrde* assay, RNAi against *lin-15b* represses expression of both *lin-15b* and *lin-15a* transcripts since *lin-15b* is the first gene transcribed from bicistronic pre-mRNA in the nucleus. In an *eri-1* mutant background a separate pathway, termed "endo-siRNA", is disrupted and the nuclear RNAi pathway is enhanced. Thus, whereas in wild-type animals there is little or no detectable Muv with *lin-15b(RNAi)* alone, in an *eri-1* mutant the occurrence of Muv phenotypes approaches 95% (90). In *nrde* mutants, the enhanced Muv penetrance is therefore suppressed, since the nuclear RNAi pathway is also defective.

Interestingly, previous work in our lab had shown that the expression of a repetitive somatic GFP reporter, *rtIs31*, was strongly repressed in *eri-1* and *rrf-3* mutants. In combination with *sig-7*, however, the enhanced repression was suppressed; i.e., the enhanced RNAi pehotype was defective in the *sig-7* mutant (T. Ratliff and W. Kelly, unpublished). This indicated that *sig-7* was operating in an RNAi pathway separable from the pathways in which ERI-1 and RRF-3 operated, which suggested a role in the nuclear RNAi pathway.

Using the *eri-1* enhanced Muv nrde assay, we tested if SIG-7 is involved in nuclear RNAi. Consistent with previous reports, we observed 97% (634/655) of *lin-15b(RNAi)* animals displaying the Muv phenotype in *eri-1(mg366)* animals. Likewise, the Muv penetrance was fully suppressed to 0% (0/386) in *eri-1(mg366)IV;nrde-2(gg91)II* double mutants, since NRDE-2 is an essential factor in nuclear RNAi (Figure 24). We were then excited to observe that 0% (0/43) of *eri-1(mg366)IV:sig-7(cc629)I* double mutants showed the Muv phenotype, indicating a role for SIG-7 in nuclear RNAi. However, we also observed the suppression (1%; 3/323) in *eri-1(mg366)IV;sig-7(cc629)/hT2(I,III)* animals, in which the *sig*-7 mutation is maintained as a heterozygote in a genetically balanced strain. This was surprising, as all other *sig-7* mutant phenotypes are recessive.

Indeed, the suppression of the Muv phenotype turned out to be due to the presence of the genetic balancer, which is marked by an integrated repetitive GFP reporter transgene, that we used to maintain the *sig-7(cc629)* allele. When we repeated the experiment using *sig-7(cc629)* balanced with *unc-29(e193)*, the Muv suppression phenotype in the heterozygous *sig-7* was minimal (86% ; 272/318) compared to the previously observed significant suppression (1% ; 3/323) when *sig-7(cc629)* was balanced with *hT2* genetic balancer. (Figure 24). In addition, we also observed good Muv suppression (10%; 3/31) when *unc-29(e193)* was balanced with *hT2* genetic balancer.

The original suppression we observed in strains balanced by the *hT2* genetic balancer is interesting, and indicates a maternal affect from the balancer, since offspring not inheriting the balancer also showed Muv suppression. However, the results clearly indicated that *sig-7* plays no significant role in nuclear RNAi, and that the original

defects observed in the *eri-1* enhancement of the *rtIs31* reporter are likely due to the indirect effects we observed with repetitive reporter expression.

Discussion

We report an analysis of a highly conserved, multi-domain nuclear cyclophilin, SIG-7, that is required for efficient transcription elongation and transcript splicing in *C. elegans*. Similar to the *S. pombe* and *A. thaliana* SIG-7 orthologs (Rct1 and AtCyp59, respectively), SIG-7 is an essential protein implicated in regulation of the phosphorylation status of important serines in the CTD of Pol II (144, 146). Both maternal supply and zygotic production of SIG-7 is required for normal development at all stages.

Depletion of SIG-7 results in a substantial decrease in embryonic transcripts that correlates with defective mRNA splicing, including co-transcriptional *trans*-splicing. A change in the distribution of Pol II within gene bodies is also observed that, along with reduced Pol II CTD Ser-2 phosphorylation and H3K36 and H3K79 methylation, are consistent with defects in Pol II elongation. The correlation between elongation defects and splicing defects suggests a co-dependency between these two processes in *C. elegans*, with SIG-7 providing an essential link. The extent to which mRNA processing and Pol II elongation are co-dependent in any organism is controversial (82, 212, 223-228). This will likely remain controversial, because it is challenging to experimentally discriminate between a splicing defect causing an elongation defect versus an elongation defect causing a splicing defect.

We favor the conclusion that the primary defect in *sig-7(RNAi)* and mutant embryos is defective splicing, and this defect results in inefficient Pol II elongation.

Among the myriad abnormal developmental defects observed in *sig-7* mutant animals, the Mog (masculinization of germline) phenotype stands out as most likely to be primarily from a splicing defect. The Mog phenotype occurs when germ cells fail to switch from sperm to egg production when *C. elegans* hermaphrodites transition from L4 to adult. The bulk of the regulation of the male-to-female switch in the *C. elegans* germline sex is post-transcriptional (229). Multiple genetic screens for Mog mutants have recovered mutations in splicing factors (230-232). In addition, a recent RNAi screen targeting conserved RNA splicing factors showed that ~40% of them (47/114) resulted in a Mog phenotype, consistent with splicing of key sex-determination regulators being the predominant mechanism controlling germline sex switching in *C. elegans* (231-233). The Kerins et al. study further showed that general splicing factors are also important for controlling the switch from germ cell proliferation to meiotic entry, and we observe a similar germline tumor enhancement effect in *sig-7(cc629)* mutant worms (T. Ratliff and W. Kelly, unpublished). In addition, recent reports have indicated that a primary defect in splicing can cause defective Pol II elongation. Depletion of a known splicing factor, SC35, results in attenuation of Pol II elongation through gene bodies in mammalian cells (234). Furthermore, a recent study in HeLA cells showed that specifically inhibiting splicing, using spliceostatin (SSA) or antisense oligos targeting snRNAs, results in defects remarkably similar to those caused by SIG-7 depletion: early dissociation of Pol II leading to its 3' depletion and decreases in Ser2P (96). It's thus very likely that the primary defect in *sig-7* mutants is defective splicing, which leads to defective elongation and Pol II dissociation from genes.

We thus propose that a key function of SIG-7 is to ensure efficient co-transcriptional splicing, and that the defects in Pol II elongation in *sig-7* mutants/RNAi are mechanistically linked to inefficient splicing.

What is SIG-7's role in transcription regulation? The mechanistic roles of nuclear cyclophilins in any organism remain to be determined, yet clues are provided by our results and those from other studies. The SIG-7-type nuclear cyclophilins all have a conserved RNA-recognition motif (RRM domain) in addition to the peptidyl-prolyl *cis-trans* isomerase (PPI) domain. Studies in *S. pombe* and *A. thaliana* demonstrate a role for Rct1 and AtCyp59, respectively, in regulation of Pol II CTD phosphorylation (144, 146), and *in vitro* binding experiments show that the CTD of Pol II interacts with the PPI domain of Rct1 (145). Yet other studies implicate Rct1's PPI domain as being dispensable for Rct1 function, whereas its RRM domain is essential (A.Y-Chang and R. Martienssen, *personal communicatio*n). The RRMs of both Rct1 and AtCyp59 were shown to bind a motif present in ~70% of all mRNAs, and AtCyp59 interactions with pre-mRNAs were observed, supporting a general role in co-transcriptional RNA processing *in vivo* (147). Thus, a potential model for SIG-7 is that it binds to the CTD of Pol II through its PPI domain and employs its RRM domain to capture emerging RNAs, perhaps to efficiently recruit them to the spliceosome machinery attached to the CTD. The loss of the PPI domain may only affect the efficiency of this coupling. In the absence of SIG-7, the linkage between emerging transcripts and the splicing machinery may be more overtly compromised, leading to decreased splicing efficiency and concomitant slowing of Pol II elongation via unknown mechanisms.

Alternatively, the PPI domain's catalytic function in isomerization of prolines may target the Pol II CTD repeat, which in turn may affect Pol II elongation via CTD structural alterations. The PPI activity may be regulated by RNA binding and/or RNA processing, and this could provide a mechanistic link between elongation and splicing. Indeed, binding of RNA to AtCyp59 in *A. thaliana* affects the isomerase activity of the PPI domain *in vitro* (147). Future studies should investigate the importance of the PPI domain's catalytic activity and a requirement for it to be structurally linked to the RRM domain.

Interestingly, while *sig-7* homologs are found in most eukaryotes, there is no clear homolog of *sig-7* in budding yeast. Introns are relatively rare in budding yeast genes (present in just ~4% of protein-coding genes) and the introns are small in size (235). A comparative genomic study of functionally linked genes in budding and fission yeast identified three cyclophilin family members, including SIG-7/Rct1, among spliceosomal components and related proteins that have been lost from budding yeast (186). SIG-7 function is thus dispensable in an organism with few, small introns to process, as might be predicted for a protein responsible for coordinating transcription elongation with efficient splicing.

Material and Methods

Worm strains and maintenance

C. elegans strains were maintained at 20°C. Worms were grown on NGM (Nematode Growth Medium) plates unless stated otherwise.

Strains: wild type N2 (Bristol), KW1317: *sig-7*(*cc629*) I/*hT2*[*bli-4(e937) let-?(q782) qIs48*](I;III), KW2230: *sig-7*(*n5037*) I/*hT2* [*bli-4*(*e937*) *let-?*(*q782*) *qIs48*](I, III), PD7271: *pha-1(e2123ts*) III; *ccEx7271*(pBK48.1::pC1), KW2309: *sig-7(n5037)I; ckIs33(unc-119, sig-7:gfp:3XFLAG)II, ccEx7291*(pBK48.1::pRF4), *eri-1(mg366)*IV*;nrde-2(gg91)*II, *eri-1(mg366)*IV*;sig-7(cc629)*I*/hT2* [*bli-4*(*e937*) *let-?*(*q782*) *qIs48*](I, III)*, eri-1(mg366)*IV*, eri-1(mg366)*IV*;unc-29(e193)*I*/ hT2* [*bli-4*(*e937*) *let-?*(*q782*) *qIs48*](I, III), *eri-1(mg366)*IV*;sig-7(cc629)*I*/unc-29(e193)*I

sig-7::GFP::3XFLAG transgenic strain

Repetitive sequences within intron 1 of the *sig-7* gene prevented PCR-based cloning of the whole gene. We used a fosmid clone (construct ID: 15087717651452437 A06) containing *sig-7* engineered with a 3' GFP::3XFLAG tag obtained from the TransgeneOme project (187). The fosmid was cut with SphI to generate an 18KB fragment containing the entire operon with *sig-7* and two neighboring genes (CEOP1492). This fragment was blunt ended and inserted into the pCFJ151 MosSCI targeting vector cut with PvuII. This construct, pJA8, was integrated into an LGII MosSCi targeting site (*ttTi5605*) by standard Mos-SCI integration techniques (188).

RNAi-mediated Depletion

RNAi was performed by feeding HT115 bacteria transformed with plasmids expressing dsRNA targeting the corresponding gene, or carrying the empty L4440 RNAi vector for controls.

RNAi embryos: Adult worms were collected from plates and washed with M9 buffer (22mM KH2PO4, 42mM Na2HPO4, 86mM NaCl, 1 mM MgSO4), bleached with sodium hypochlorite (5% bleach with 1.0N NaOH) to isolate embryos. Embryos were placed on NGM (Nematode Growth Medium) plates without food overnight. The synchronized hatched L1s were transferred to plates with OP50 bacteria and grown for 36 hours until the L3 larval stage. L3s were washed with M9 buffer 3 times and transferred to induced RNAi plates (NGM+1mM IPTG+1mM Ampicillin) pre-seeded with bacteria expressing the desired dsRNA. The worms were grown on RNAi plates for 36 hours, after which the gravid adults were washed with M9 buffer and separated from any extruded embryos by filtration through a 40µm mesh cell strainer (Fisher Scientific, #22363547). The adults were bleached as described above to collect *in utero* embryos for analysis.

RNAi Adults (Figure 11): L3 larvae prepared as above were fed dsRNA-expressing bacteria for 55 hours instead of 36 hours and directly processed for total RNA purification and analysis by qRT-PCR.

Immunofluorescence

Intact embryos were fixed in 2.5%PFA/ethanol (236) or methanol/acetone (237) for all immunofluorescence except for those probed with monoclonal antibody H5, which was fixed in methanol/formaldehyde (236). Primary antibodies used were: anti-H3K4me2 (CMA30) 1:1000 END Millipore (203)], anti-P-granules [OIC1D4, 1:5 (237)], anti-H3K36me3 [(CMA333), 1:1000 (236)], anti-Ser2p RNA pol II CTD (H5, 1:500, Covance MMS-129R), anti-GFP (1:1000, Novus NB600-308), anti-AMA-1 (1:10,000, Novus 38520002), and anti-FLAG (M2, 1:1000, Sigma F1804). Secondary antibodies

used were Alexa Fluor 488-conjugated donkey anti-mouse (1:500, Invitrogen R37114) and Alexa Fluor 594-conjugated goat anti-rabbit (1:500, Invitrogen R37117). Samples were mounted in ProLong Gold anti-fade reagent (Life technologies, P36934) and observed under a fluorescence microscope (Leica DMRXA; Hamamatsu Photonics, Hamamatsu, Japan) with Simple PCI software (Hamamatsu Photonics). Image J was used for quantification of raw immunofluorescence intensity (238).

Immunoprecipitation assays

Embryos were collected as described above, washed , and resuspended in 3X volume of ice-cold Hypotonc Triton-X buffer [20 mM Tris–HCl (pH 7.4), 10 mM KCl, 10 mM MgCl2, 2 mM EDTA, 10% glycerol, 1% Triton X-100, 2.5 mM β-glycerophosphate, 1 mM NaF, 1 mM DTT, and Complete protease inhibitors; (239)]. Resuspended embryos were frozen in liquid nitrogen and ground into a fine powder using a mortar and pestle and thawed on ice for 10 min. The suspension was sonicated for 2 min at high setting using a Bioruptor sonicator (Diagenode Inc., Denville, NJ, USA). The salt concentration was then adjusted to either 150mM or 350mM NaCl, and incubated for 30 min with rotation at 4°C. After an additional 2 min sonication, the lysate was centrifuged for 15 min at 13,000g. The supernatant was transferred to new tubes, and 1ml of each lysate supernatant was pre-cleared by incubation with 60 µl of either Protein A (Life Technologies, 10002D) or Protein G Dynabead (Life Technologies, 10004D) for 30 min with rotation at 4°C. 100µl of each pre-cleared lysate was saved as input sample, and the remaining 900ul was used for immunoprecipitation. Either anti-FLAG (Sigma, F1804) or anti-GFP (Novus biological, NB600-308) for SIG-7 IP and anti-AMA-1 (Novus biological, 38520002) for Pol II (AMA-1) IP were added to the lysate (10ug of

antibody/2.5mg of lysates**)** and incubated for 12 hours at 4°C. 60µl of either Protein A or Protein G Dynabeads were added directly to the lysate/antibody mix and incubated at 4°C for 3 hours. Beads were separated from solution using a magnetic bar, washed 2 times for 5 min in Hypotonic Triton-X buffer, and washed twice more with 500mM NaCl hypotonic Triton-X buffer for 10 min at room temperature. For final elution, beads were incubated with 150µl of 2X SDS-PAGE sample buffer for 15 min at room temperature. The final eluates were further analyzed by SDS-PAGE and Western blot.

Protein isolation and Western blot analysis

RNAi-treated embryos were resuspended in 4X volume of RIPA buffer (Thermo Scientific, #89901) and 2X volume of glass beads (Sigma, G8772), and homogenized using a Mini Beadbeater-16 (BioSpec, Bartlesville, OK, USA) for 1 min 3 times at 4°C. The homogenized embryos were incubated at 4°C on a rotator for 30 min and further processed in a Bioruptor sonicator (Diagenode Inc., Denville, NJ, USA) at high setting for 10 min to fragment chromatin. The final lysates were centrifuged at 13,000g for 10 min at 4°C, and the supernatants were collected. The protein concentration was determined using the Bradford reagent (Biorad, #500-0006). Supernatants were mixed with an equal volume of 2X SDS-PAGE sample buffer and denatured for 5 min at 95°C, and equal protein amounts were loaded and run on a 4-20% precast SDS-PAGE gel (Biorad, #456-1094) and transferred to PVDF membrane. Transferred proteins were blocked in 5% milk PBST for 1 hour, incubated with primary antibody overnight, and washed 3 times with 1X PBST for 10 min each. After incubation with secondary antibody for 2 hours at room temperature, the blot was washed 3 times with 1X PBST for 10 min each. The washed membrane was incubated with chemiluminescence reagent

(Thermo Scientific, #34087) for 5 min, and protein bands were visualized with autoradiography film (Genesee Scientific, #30-100). The primary antibodies used are the following: anti-FLAG (Sigma, 1:2000, F1804), anti-Actin (Millipore, 1:10,000, MAB1501), H14 (Covance, 1:3000, MMS-134R), H5 (Covance, 1:3000, MMS-129R), anti-AMA-1 (Novus Biological, 1:5000, 38520002), 8WG16 (Covance, 1:1000, MMS-126R), anti-H3K36me3 (Abcam, 1:1000, Ab9050), anti-H3K79me2 (Abcam, 1:1000, Ab3594), anti-H3K4me3 (Abcam, 1:1000, Ab8580) and anti-H3 (Abcam, 1:20,000, Ab1792). The following secondary antibodies were used: Goat Anti-Rabbit IgG, HRP-conjugated (Millipore, 1:2500, 12-348), Goat Anti-Mouse IgG, Peroxidase-conjugated (Millipore, 1:2500, AP124P), AffiniPure Goat Anti-Mouse IgM, Peroxidase-conjugated (Jackson ImmunoResearch Laboratories, Inc, 1:5000, 115-035-075).

RNA purification and qRT-PCR

Embryos collected after RNAi exposure were washed with M9, and pelleted embryos were resuspended in Trizol (50µl of embryos/300µl of Trizol, Invitrogen), snap frozen in liquid nitrogen, and subjected to 3 freeze/thaw cycles. 62µl of chloroform was added and mixed thoroughly by shaking 10 times and spun down for 15 min at 4°C. Nucleic acids were precipitated with 0.3M acetic acid in 100% isopropanol and resuspended in 100µl of nuclease-free water. Total RNA was purified using RNeasy kit (Qiagen, Valencia, CA, USA) as per the manufacturers' instructions. cDNA was synthesized from 1µg of purified total RNA using iScript select cDNA synthesis kit (Bio-Rad, 170-8896). 50ng of cDNA was used for qPCR using SsoFast reagent (Bio-Rad, 172-5201) on CFX96 Real-Time system (C1000 Thermal Cycler, Bio-Rad). The transcript levels of genes analyzed were first normalized to 18S RNA for each sample, and the normalized transcript levels of

either *sig-7(RNAi)* or *ama-1(RNAi)* experiments were then compared to the transcript levels of L4440 controls to generate ΔΔCT plots of relative transcript levels. The averages of two technical replicates from two biological samples were plotted with standard deviation.

Library preparation and RNA sequencing

Total RNAs were purified as described for qRT-PCR and sent to Axeq Asia (Seoul, Korea) for transcriptome sequencing. 1μg of total RNA was used as starting material, and a sequencing library was prepared using TruSeq Stranded Total RNA Sample Prep Kit after treatment with Ribo-Zero (Human/Mouse/Rat) for rDNA depletion. Library QC was performed using Tapestation D1000 Screen Tape (Agilent) and quantified using KAPA Library Quantification Kit (for Illumina® platforms). Clusters were generated by HiSeq PE (Paired-End) Cluster Kit v3 cBot, and sequencing was done on a HiSeq2000 with 100bp paired-ends using TruSeq SBS v3-HS kit reagents.

Analysis of RNA-seq

RNA-seq reads were quality-checked using FastQC version 0.5.2 to ensure per-base sequence quality, per sequence quality scores, per base sequence content, per base GC content, per sequence GC content, per base N content, sequence length distribution, sequence duplication levels, kmer-content, and that over-represented sequences were within accepted norms. FastQ Quality Trimmer version 1.0 was used to trim reads with less than optimal quality scores. The DE analysis protocol outlined in Trapnell *et. al*. was used to perform the DE analysis (199). The quality filtered reads were mapped to *C. elegans* (ce10) reference genome using TopHat2 version 0.5. TopHat2 internally uses Bowtie2 to map the reads. Mapping results were used to identify splice junctions between

exons. Cufflinks version 2.1.1 was used to assemble transcripts and estimate their abundance. The transcript assembly outputs from Cufflinks were merged into a unified list of transcripts using Cuffmerge. Cuffdiff version 2.0.2 was then used to quantify gene and transcript expression levels and test them for significant differences. Default parameters were utilized for all steps. This analysis was done in part by the Emory Integrated Genomics Core (EIGC), which is subsidized by the Emory University School of Medicine and is one of the Emory Integrated Core Facilities. For Figure 14B, exon and intron coordinates were obtained from WS2220. 99,830 exons with length ≥ 50bp and 69,762 introns of ≥ 50bp were obtained. Custom scripts were used to calculate the average read coverage of exons and introns per gene.

Library preparation from ChIP material for sequencing

Chromatin immunoprecipitation (ChIP) was done as described in (240) with the following modifications. 1) Worms were grown on 15cm peptone rich plates seeded with NA22 bacteria. 2) Samples were sonicated using a Bioruptor sonicator at high setting for 40min (40sec on/ 20sec off). After collection of immunoprecipitated DNA, DNA libraries were prepared as described in (241). DNA libraries were sent to Axeq Asia (Seoul, Korea) for sequencing. Library QC was done using BioAnalyzer High sensitive DNA chip (Agilent). Clusters were generated by HiSeq PE (Paired-End) Cluster Kit v3 cBot, and sequencing was performed on HiSeq2000 with 100bp paired-ends using TruSeq SBS v3-HS kit reagents.

Analysis of ChIP-seq data

Gene set definitions used were as published in (194). Briefly, ubiquitous genes are defined as genes expressed in 4 different tissue-specific SAGE data sets: germline,

neuronal, muscle, and gut (242, 243). The germline-enriched category was defined by (195), although spermatogenesis-specific genes were removed. Germline-specific genes were defined as expressed in the germline either in Reinke et al. or in Wang et al., and were intersected with the strict maternal gene class in Baugh et al.; any genes expressed in any of the somatic SAGE expression data of Meissner 2009 were removed (191, 195, 242, 243). Soma-specific (any) genes were defined as expressed in any of the somatic SAGE data sets of Meissner et al. but not in the germline SAGE data set of Wang et al. or the germline-enriched set of Reinke et al. (195, 242, 243). Embryo-expressed is the "strict embryonic" class defined in Baugh 2004 (191). chrX genes are all the X-linked genes in WS220. Silent genes are mostly serpentine receptors and were defined in (244). The hyper-geometric distribution was used to calculate the significance of the enrichment or depletion of any of the gene sets among the mis-regulated genes in Fig 4B. AMA-1 ChIP-seq data were mapped to WS220 using bowtie (245). MACS2 was used to obtain peak calls for each replicate of L4440 and *sig-7(RNAi)* (246). The broad peak option was found to produce the most appropriate peak calls and a significance cutoff of q=0.05. The peak calls were mapped to WS220 gene annotations. If AMA-1 peak calls of both replicates of a condition overlapped with a gene body, the gene was called bound by AMA-1. Meta-gene profiles were produced using custom R scripts. Genes were aligned at their Transcription Start Site (TSS) and Transcription End Site (TES), and signal over the gene bodies was averaged in 50 bp windows. The 95% confidence interval of the mean is shown with error bars. To normalize reads between samples AMA-1 peak regions from both conditions were removed. The remaining read coverage was scaled genome-wide so the total number of reads was 2 million reads.

Protein sequence alignment

Protein sequences of SIG-7 homologues were obtained from the NCBI protein database. The alignment of homologues was generated with ClustalW2 (247). The conserved protein domain/motif search was done using ScanProsite web-based tool (248, 249). The accession numbers for proteins used for alignment are the following: SIG-7(CAB03088.2), PPIL-4(NP_624311.1), CG5808(AAF56342.1), AtCyp59(NP_175776.2), Rct-1(CAB52803.1) and KIN241(CAC35733.1).

Transgene desilencing assay

PD7271 strain carrying ccEx7271 repetitive array has been used by many researchers to screen for factors required for transgene silencing. We have assessed the degree of desilencing based on the change in intensity of the GFP reporter expression compared to the expression in control RNAi animals. Out of many cell types where this transgene is expressed, we picked the intestinal cells for quantification due to its large size nuclei that make the quantification easy with less variability. Using image J software, we have measured 2-3 nuclei of intestinal cells that are close to vulva from each animal. Then the averaged intensity from multiple animals from each RNAi experiments were compared. We have done student's t-test for statistical significance of the difference among samples.

Acknowledgments

We would like to thank Thomas Ratliff, PhD, who have performed initial characterizations and mapping of the *cc629* allele. We also like to thank Beth Bowman, PhD, for helping us with injection for generating MosSCI line of SIG-7 GFP construct, and Dr. Corces and Takenaka Naomi for providing reagents and advice for CHIP-seq experiments. Lastly, we would like to thank Susan Strome, PhD and Andreas

Rechsteiner, PhD for bioinformatics analysis including CHIP-seq analysis and exhaustive review of this manuscript for publication.

Figures

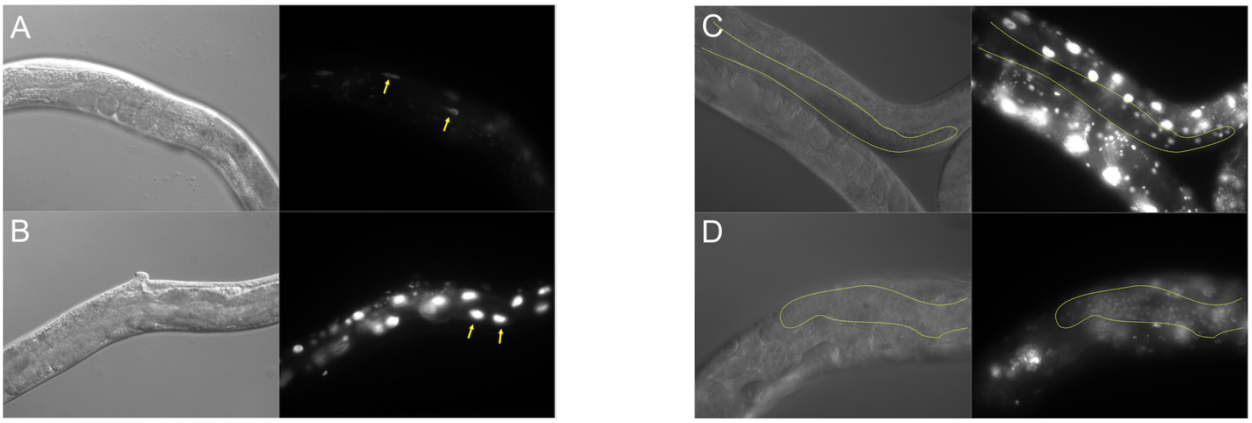
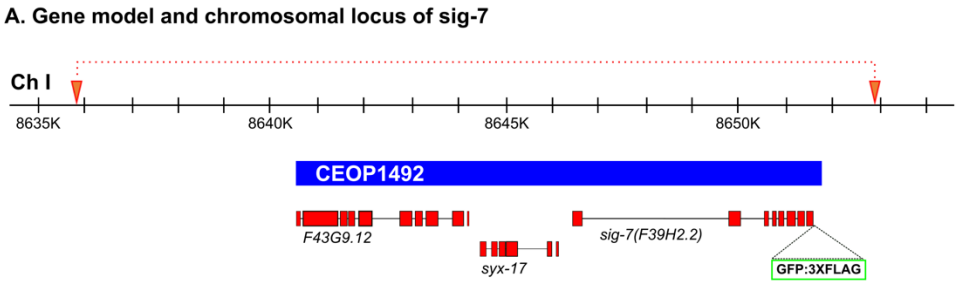
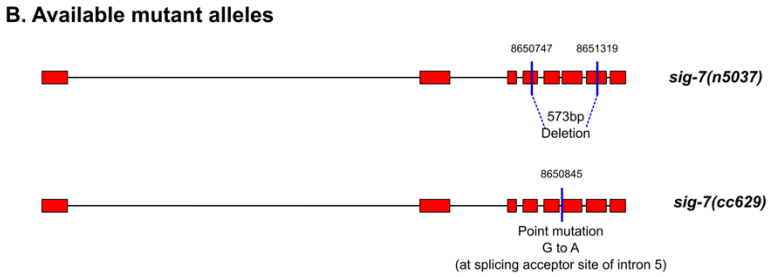
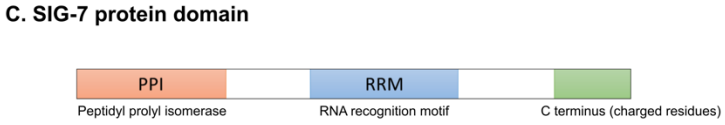


Figure 1. Repetitive transgenes are desilenced in both somatic and germ cells after the *sig-7* RNAi.
The images on the left of each panel are DIC images of the respective GFP image on the right. To assess the degree of desilencing from repetitive transgene, the images in panel A (L4440) and B (*sig-7(RNAi)*) are taken at the same exposure. Panel B (right) shows much brighter GFP expression than the panel A (right). The yellow arrows indicate large gut nuclei. Note that the images in panel C (L4440) and D (*sig-7(RNAi)*) are not taken at the same exposure. One can appreciate the degree of overexposure used to take the image of panel C (L4440) by comparing the GFP expression level to panel A (L4440). The image in Panel C (L4440) was taken at longer exposure time to make sure that the absence of GFP expression from an adult germline is not attributable to exposure difference, but it is rather resulting from disruption of SIG-7 activity by RNAi. The germline expression of desilenced repetitive transgene is easily visualized at normal exposure time (Panel D; *sig-7(RNAi)*). The repetitive array used are *ccEx7271*(Panel A and B) and *ccEx7291*(Panel C and D)







D. *sig-7* orthologs in other species

Protein	Species	Identity(%)
CG5808	*D. melanogaster*	55
PPIL-4	*H. sapiens*	51
Rct-1	*S. pombe*	48
AtCyp59	*A. thaliana*	44
KIN241	*P. tetraurelia*	39
-	*S. cerevisiae*	

Figure 2. The *sig-7* gene, protein, mutant alleles, and orthologs in other species. A) The CEOP1492 operon (blue), the *sig-7* gene with exons (red boxes) and introns (black lines), the position of the GFP::3XFLAG tag, and the genomic region used in the rescuing transgene construct (dotted line with arrowheads). B) The *n5037* deletion and *cc629* splice acceptor mutant alleles are illustrated. C) The domain structure of SIG-7 is illustrated; all three domains are also found in SIG-7 orthologs listed in (D). No homolog has been identified in budding yeast.

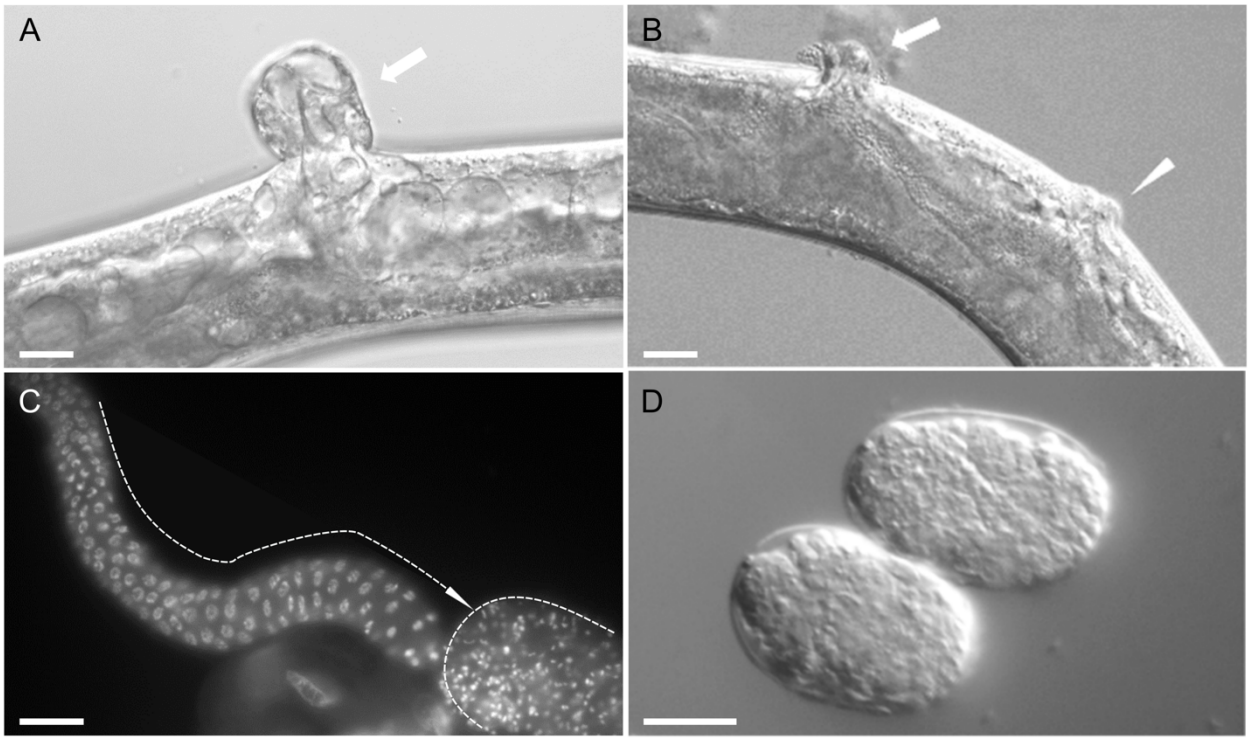


Figure 3. Pleiotropic defects in *sig-7* mutants.
A,B) Homozygous *sig-7(cc629)* adults exhibit a highly penetrant protruding vulva (Pvul, arrow in A) and lower penetrance multiple vulvae (Muv, arrows B). C) Adult hermaphrodites are sterile due to failure to switch from spermatogenesis to oogenesis (masculinization of germline or Mog phenotype), as shown in a DAPI stained whole mount ovary. The dotted line and arrow shows the direction of meiotic progression. The excessive sperm accumulation is outlined. Other phenotypes observed in the *cc629* allele include molting and seam cell defects and enhancement of germline tumors (not shown). D) RNAi depletion of *sig-7* from hermaphrodites starting at the L3 stage resulted in ~95% embryonic lethality (examples of terminal, arrested embryos shown) among their progeny, with rare survivors exhibiting the defects described above. Scale bar = 20um

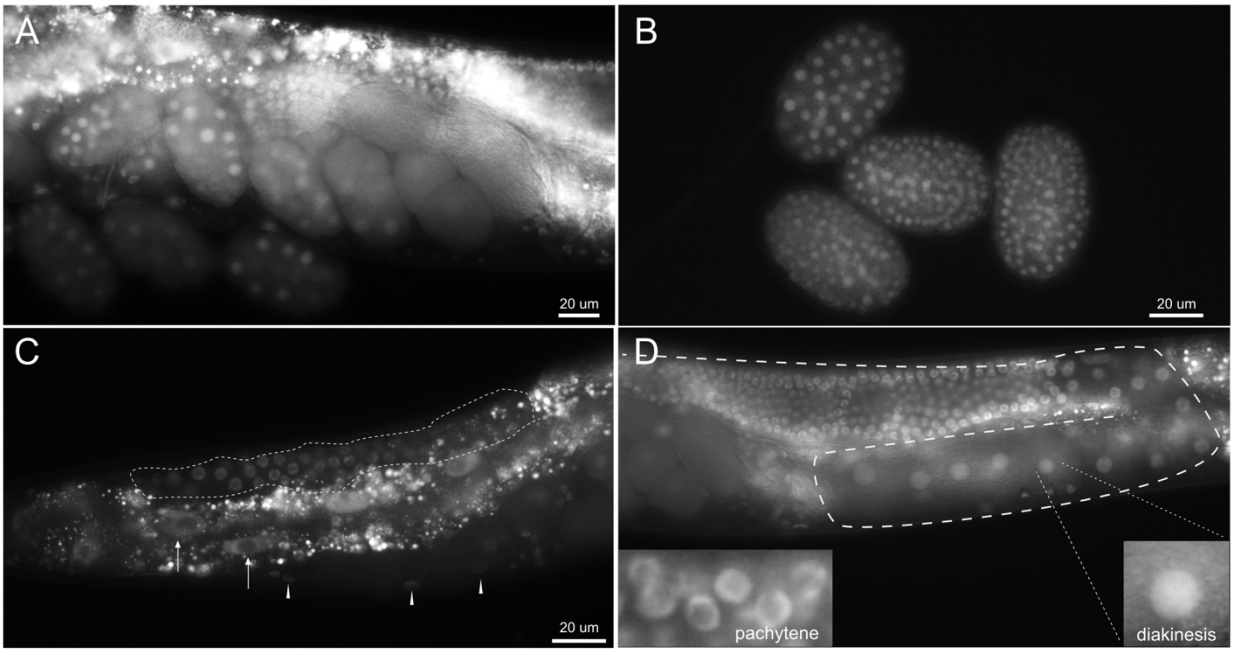


Figure 4. SIG-7 is a ubiquitously expressed nuclear protein.
The *sig-7(n5037)* allele was rescued with a SIG-7::GFP::3XFLAG fusion protein. The GFP expression patterns in live animals are shown. A-D) SIG-7::GFP::3XFLAG is nuclear and observed in all tissues and at all stages, including early embryos (A), late embryos (B), adult somatic cells (C; arrows and arrowheads), and germ cells (C, D; outlined with dotted lines). Small speckles in A, C, D are auto-fluorescent gut granules. The protein is associated with DNA in all germ cells except in transcriptionally inactive diakinesis-stage oocytes, where it becomes dispersed in the nucleoplasm (insets in D). The protein is also associated with DNA in male germ cells at all stages, with faint nuclear signal observed in mature sperm (not shown).

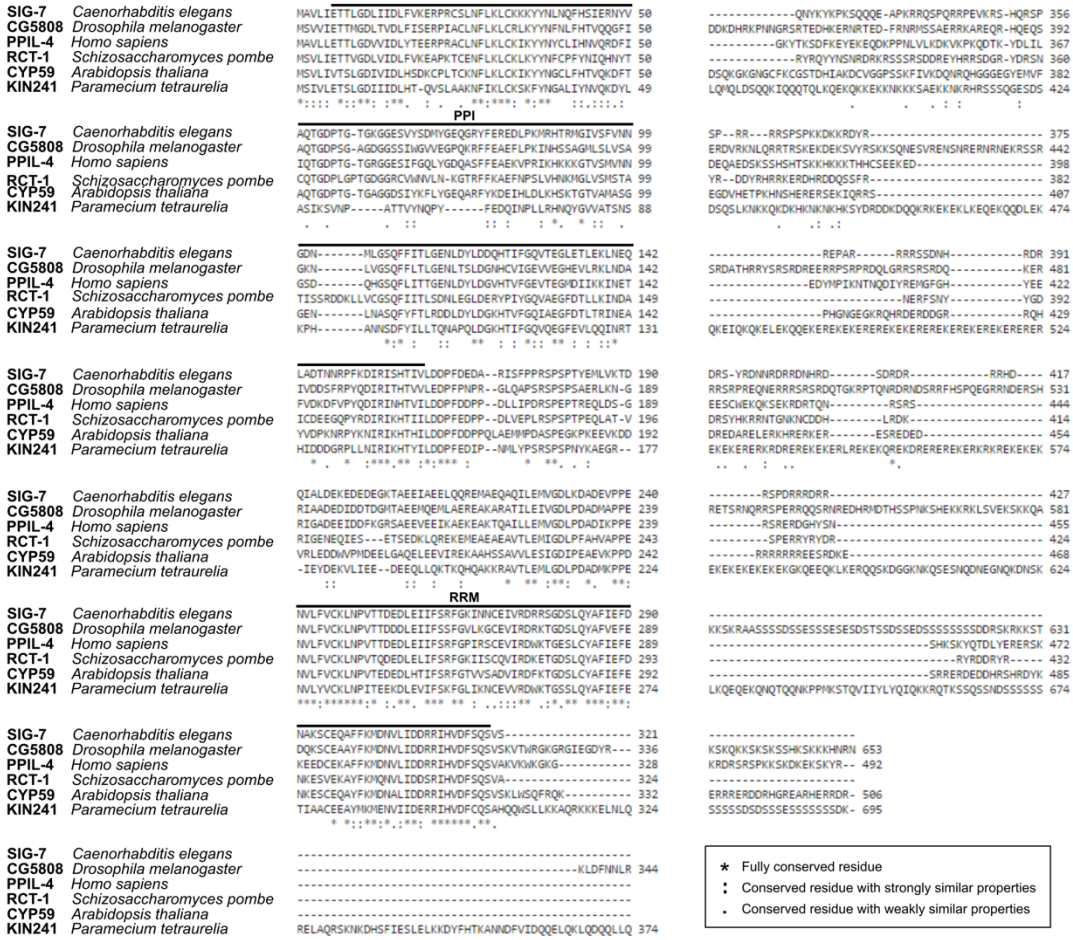


Figure 5. Sequence alignment of SIG-7 orthologs.
Protein sequences were aligned using ClustalW2 [96]. The degree of conservation is noted with different symbols in the bottom row of each alignment. The cyclophilin-type peptidyl-prolyl cis-trans isomerase domain (PPI) and RNA Recognition Motif (RRM) were identified using the ScanProsite web-based tool and are indicated with black solid lines [97, 98]. This alignment shows the high degree of conservation among orthologs within the PPI and RRM domains.

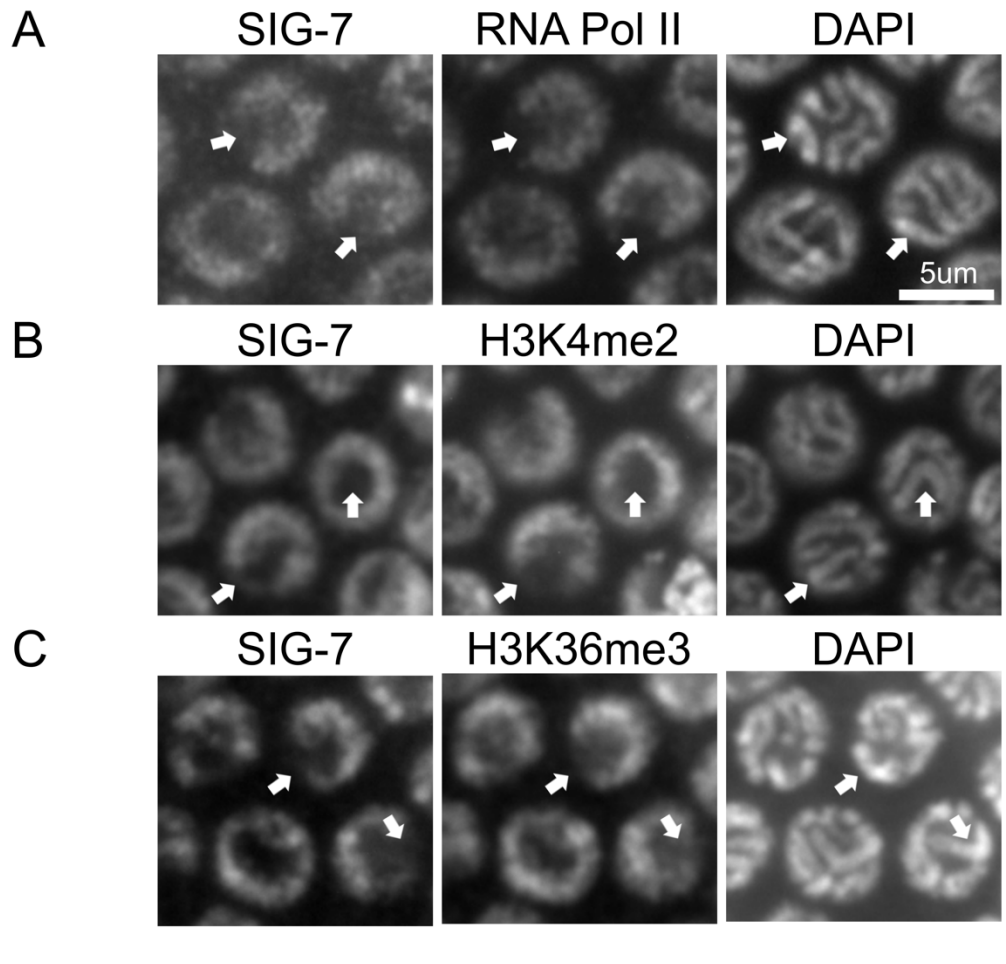


Figure 6. SIG-7 is associated with transcriptionally active chromatin.
A-C) The distribution of SIG-7::GFP in pachytene-stage germ cells was compared to DAPI staining of DNA and the distribution of RNA Polymerase II (A), H3K4me2 (B), and H3K36me3 (C). Arrows point to the pair of X chromosomes, which in germ cells have lower transcriptional activity than the autosomes.

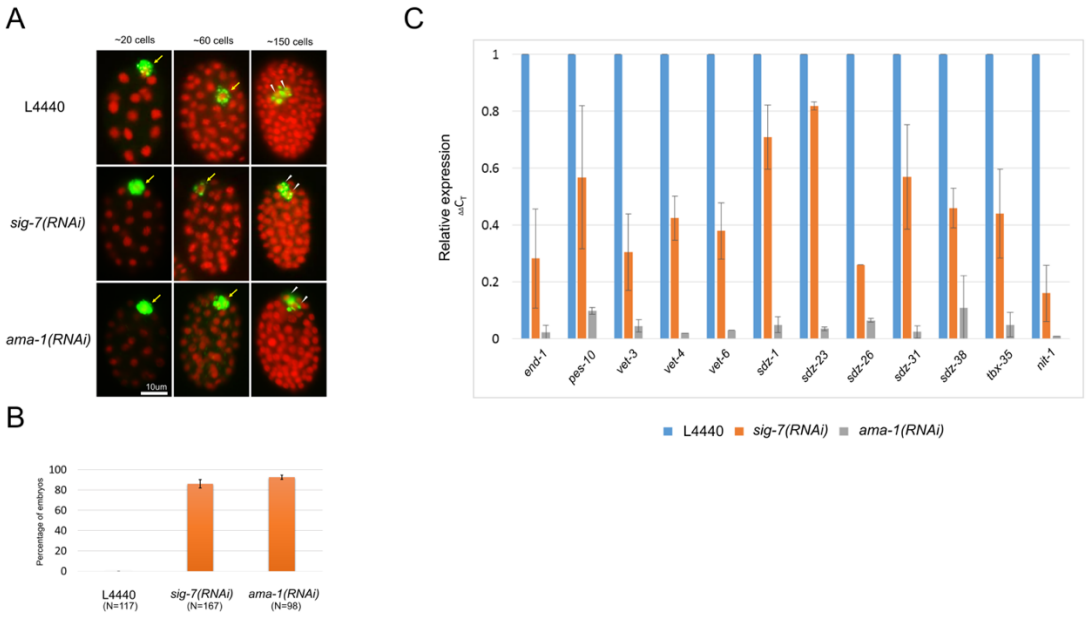


Figure 7. *sig-7* is required for hallmarks of zygotic transcription.
A) RNAi control (L4440), *sig-7(RNAi)* and *ama-1(RNAi)* embryos were fixed and stained with DAPI (red) and antibody OIC1D4, a marker of germ cells (green). In L4440 controls embryos, the germline blastomere P4 is initially at the periphery of the embryo (~20 cells, arrow), migrates to the interior of the embryo (~60 cells, arrow) and then divides to produce the primordial germ cells Z2 and Z3 (~150 cells, arrowheads). In *sig-7(RNAi)* and *ama-1(RNAi)* embryos, both P4 and its daughters Z2 and Z3 remain at the periphery of the embryo. B) Quantification of the % of embryos in which P4 failed to migrate into the interior of the embryo. Error bars = S.D. from two biological replicates. C) The RNA levels of a panel of zygotically expressed genes were measured by qRT-PCR in control (L4440), *sig-7(RNAi),* and *ama-1(RNAi)* embryos. Each sample was normalized to18S RNA levels, and *sig-7(RNAi)* and *ama-1(RNAi)* were plotted relative to L4440 control. Error bars = S.D. from two technical replicates each of two biological replicates.

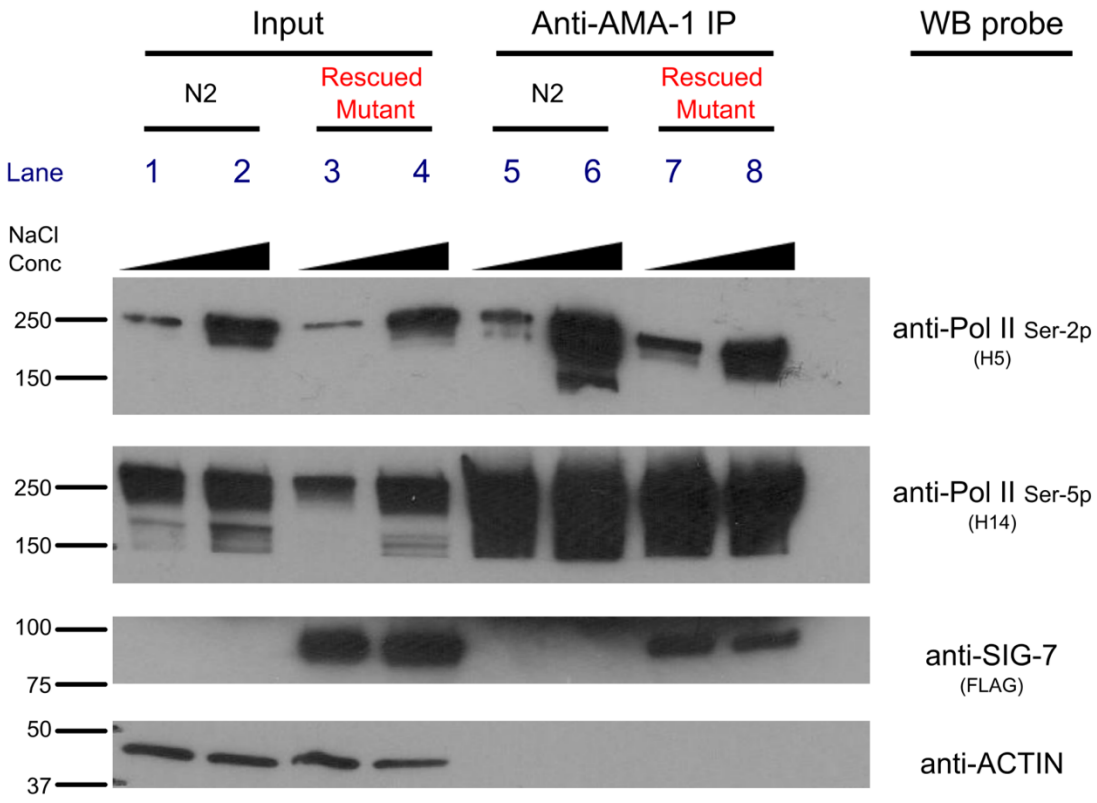


Figure 8. SIG-7 associates with RNA Pol II in vivo.
Embryo lysates were prepared from wild type (N2) or *sig-7(n5037)* mutants rescued by a sig-7::GFP::3XFLAG transgene, using either normal salt (150 mM NaCl; odd lanes) or high salt (420 mM NaCl; even lanes) extraction conditions. RNA Pol II was immunoprecipitated (IP) using an antibody that recognizes all isoforms (anti-AMA-1), and the IP material was probed by western blot using the antibodies indicated. 5% input was used for SDS-PAGE. The blots were also probed with anti-actin both to normalize for the amount of total protein used for the IP (bottom panel, lanes 1-4) and to determine the specificity of the co-IP (bottom panel; lanes 5-8).

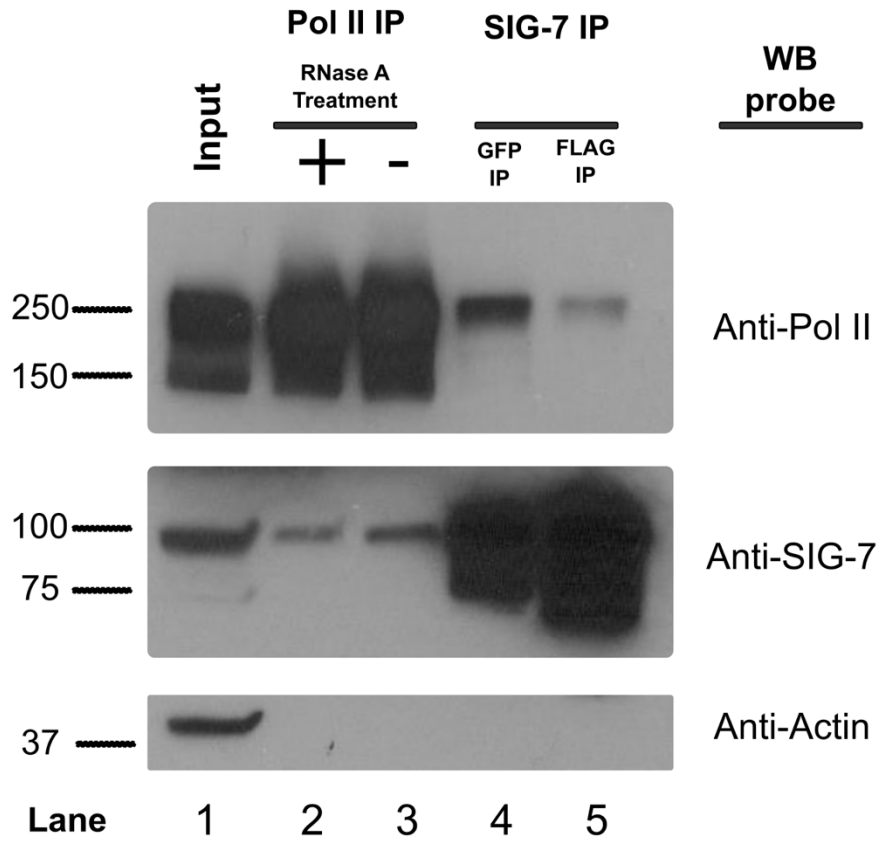


Figure 9. SIG-7 interacts with RNA Pol II *in vivo*.
Lysates from embryos expressing SIG-7::GFP::3XFLAG were divided into equal aliquots and incubated with antibodies specific for RNA Pol II (α-AMA-1; Lanes 2 and 3), GFP (Lane 4), and the FLAG epitope (Lane 5), and the immunoprecipitates were analyzed by western blots probed with anti-Pol II Ser-5P (H14; top), anti-FLAG (anti-SIG-7; middle) and anti-Actin (bottom). The lysate in Lane 2 was incubated with 100ug of RNase A for 30min at RT prior to the IP procedure.

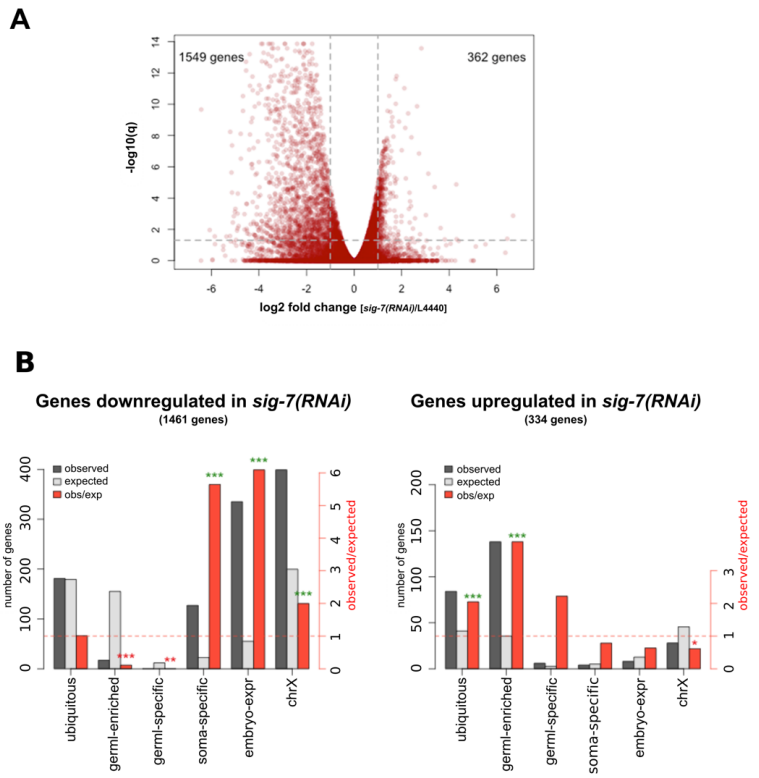


Figure 10. RNA-seq analysis of *sig-7(RNAi)* embryos reveals a global transcription defect.

RNA-seq was performed on RNA purified from *sig-7(RNAi)* and L4440 control RNAi embryos and the profiles compared. A) Volcano plot of normalized RNA-seq results showing the number of genes exhibiting significantly decreased (1549) and increased (362) RNA abundance in *sig-7(RNAi)* compared to L4440 (>2-fold differences with q value < 0.05). (B) Genes exhibiting significantly decreased (left) or increased (right panel) transcript accumulation in *sig-7(RNAi)* compared to L4440 were classified as ubiquitous, germline(germl)-enriched, germline(germl)-specific, soma-specific (broadly defined; soma-any), embryo-expressed (embr-expr), or X-linked (chrX). Gene class definitions are described in Materials and Methods. The number of genes observed and expected for each class (left y-axis), and the ratio of observed/expected (right y-axis) are shown. The gene classes with zygotic expression are highly overrepresented in those showing reduced expression in *sig-7(RNAi)* embryos, while genes with germline expression are highly overrepresented in those showing higher expression in *sig-7(RNAi)* embryos. The significance of the enrichment of observed/expected representation in upregulated or downregulated for each gene class is indicated by asterisks (* = $p \leq 0.01$, ** = $p \leq 10^{-5}$, *** = $p \leq 10^{-10}$)

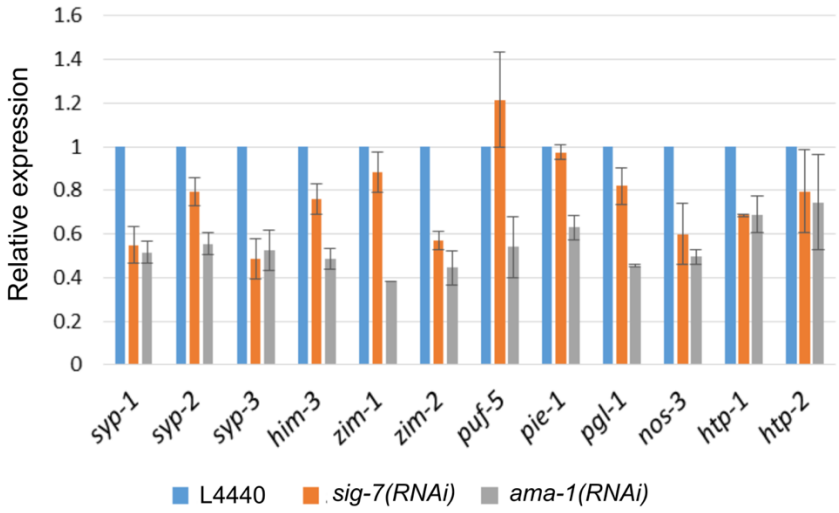


Figure 11. SIG-7 is required for germline transcription.
The abundance of germline-specific transcripts was measured by qRT-PCR in RNAi adults treated with extended RNAi (55 hrs post-L3; see Materials and Method). RNA levels were normalized to 18S RNA levels and plotted relative to L4440 control in each experiment. Error bars = S.D. from two technical replicates each of two biological replicates.

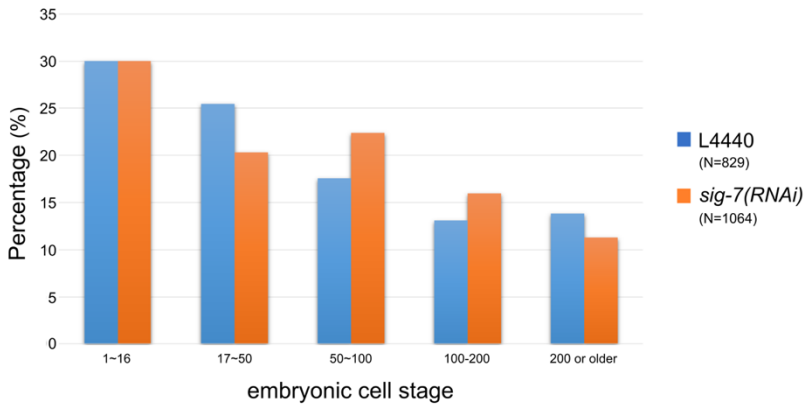


Figure 12. Comparison of embryonic stages present in control vs *sig-7(RNAi)*. Embryos collected from RNAi experiments were fixed and stained with DAPI. The approximate numbers of nuclei in the embryos in a field were binned as indicated. (N= total number of embryos quantified)

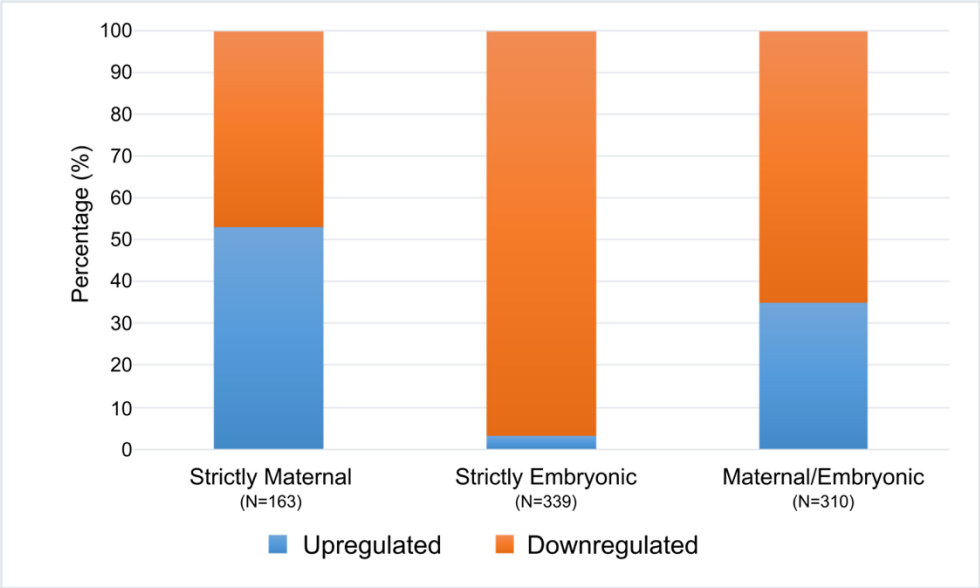


Figure 13. *sig-7(RNAi)* predominantly affects zygotic gene expression.
Of the 1522 down-regulated genes and 389 upregulated in *sig-7(RNAi)* embryos (at least two fold expression difference with q value < 0.05), 607 and 205 genes, respectively, were among those classified by Baugh et al. as Strictly maternal, Strictly Embryonic, or Maternal/Embryonic [99]. The percentages of up- or down-regulated genes that fall into these gene classes are indicated.

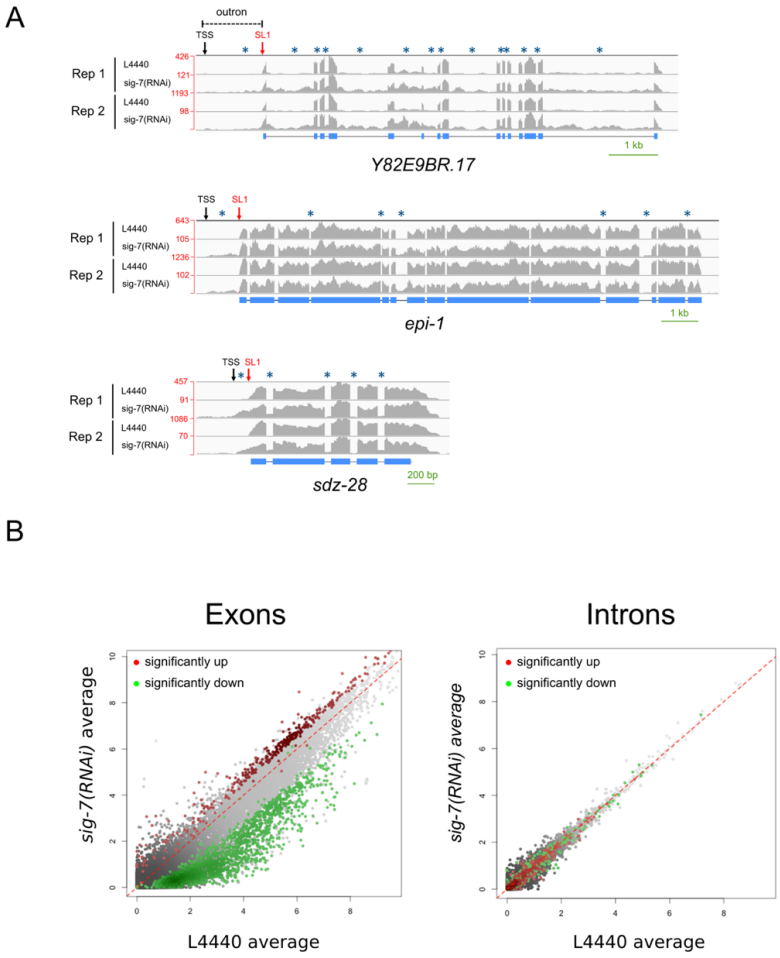


Figure 14. SIG-7 is required for efficient splicing of nascent transcripts.
A) Examples of zygotically expressed genes with splicing defects. The number of aligned reads generated by Tophat are indicated on the left. The y-axis of the *sig-7(RNAi)* reads, which are reduced compared to L4440 controls, is scaled to the exon reads in L4440. An "outron" is the RNA segment removed by trans-splicing of nascent transcript; its sequence corresponds to that between the TSS (Transcription Start Site) and SL1 (Splice Leader 1 acceptor site) of trans-spliced genes. Exons (blue boxes) and introns (black solid lines) are shown under each RNA-seq profile. The relative levels of introns and outrons (indicated with asterisks) compared to exons are significantly higher in *sig-7(RNAi)* compared to L4440 control, reflecting persistence of primary transcripts. B) Average log2 read coverage per gene for exons and introns in *sig-7(RNAi)* vs L4440 is shown. Genes up- and downregulated in *sig-7(RNAi)* compared to L4440 by cuffdiff analysis are shown in red and green, respectively. Exon levels change in the manner expected for mis-regulated genes, while intron levels remain relatively unchanged.

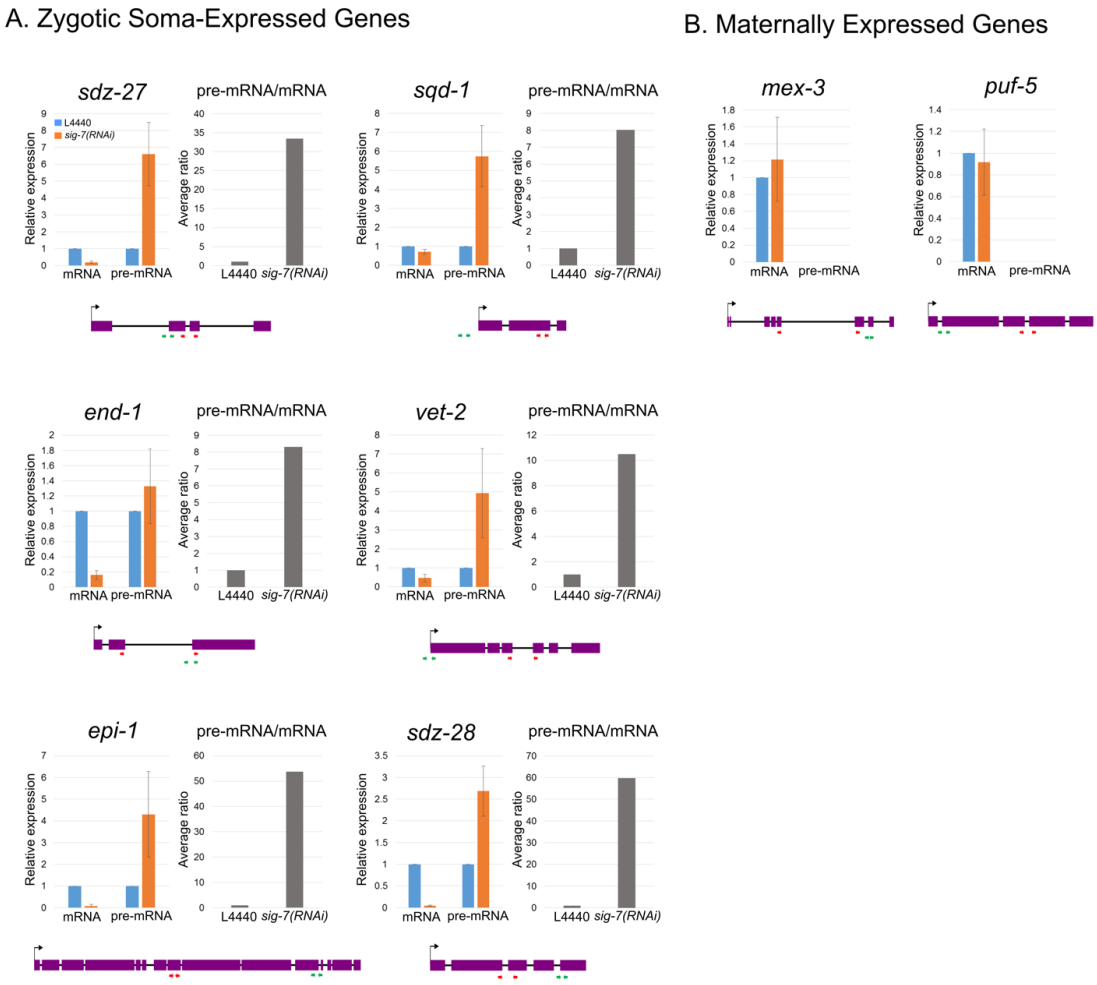


Figure 15. qRT-PCR analysis of the effect of *sig-7(RNAi)* on splicing.
qRT-PCR was performed on several strictly embryonic (A), and strictly maternal genes (B) affected by *sig-7* RNAi. In order to distinguish unprocessed pre-mRNA from mRNA, primers were designed to amplify either outron, introns, or intron-exon junctions as indicated with green arrows. Other primers were designed to amplify only spliced exon-exons of mRNAs as indicated with red arrow. The relative abundances measured by qRT-PCR of respective RNA are shown for each gene as are the ratios of pre-mRNA/mRNA. For maternally expressed genes, no pre-mRNAs were detected. The expression for each was normalized to 18S RNA and plotted relative to L4440 controls in each experiment. Error bars = S.D. from three technical replicates each of four biological replicates.

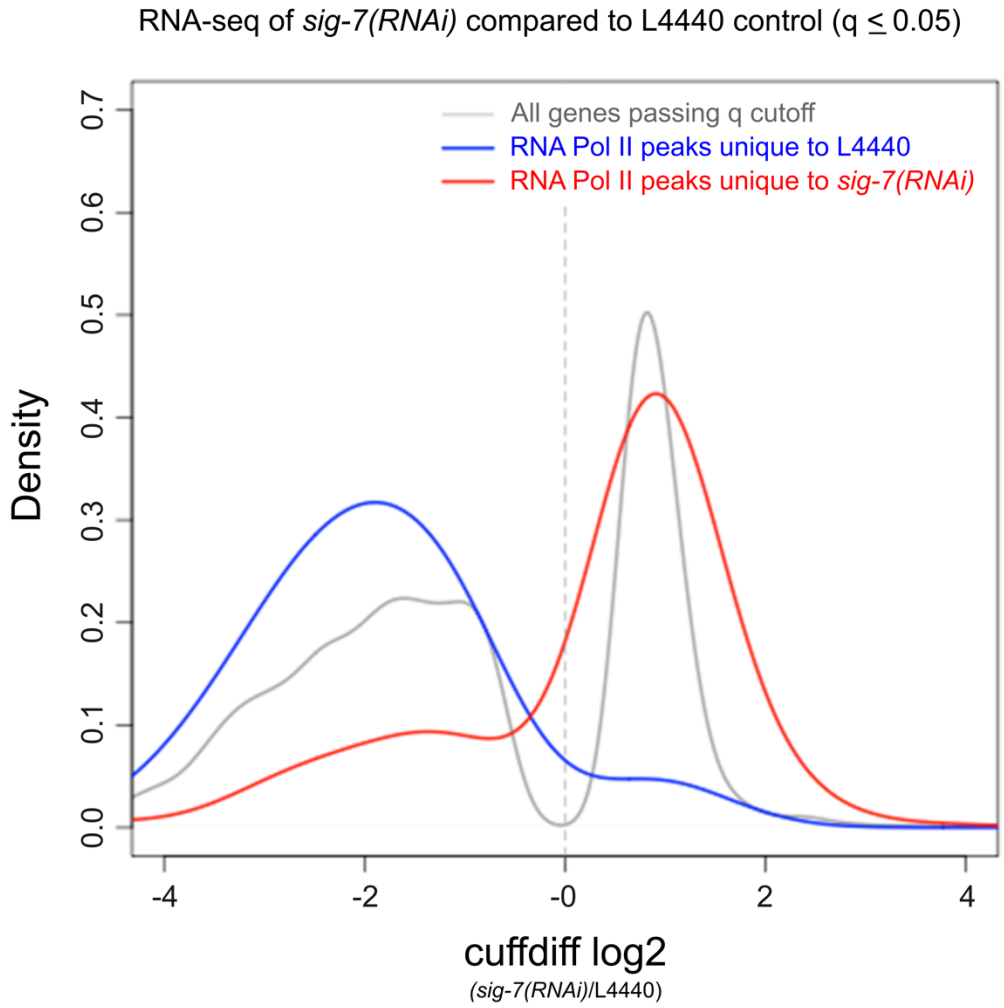


Figure 16. *sig-7(RNAi)*-dependent changes in RNA Pol II occupancy correlate with expression changes.
Genome-wide RNA Pol II (anti-AMA-1 antibody) ChIP-seq was performed in L4440 control RNAi and *sig-7(RNAi)* embryos, and the read density profiles were compared with the RNA-seq results from the experiments described in Figure 4. The comparison shows an excellent correlation between the loss of RNA Pol II from genes with a decrease in RNA abundance, indicating that the changes observed with *sig-7(RNAi)* are associated with a transcription defect.

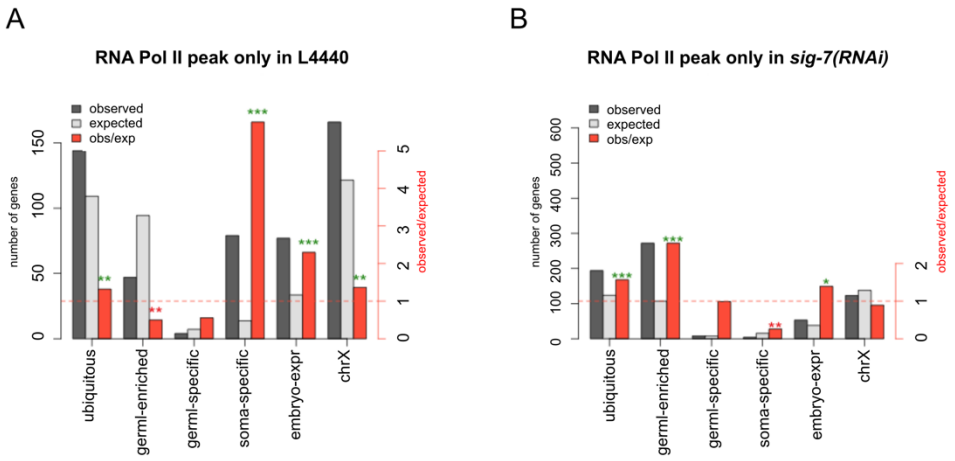


Figure 17. *sig-7(RNAi)*-dependent changes in RNA Pol II occupancy among different gene classes are consistent with defects observed by RNA-seq.
Genes with RNA Pol II occupancy changes in *sig-7(RNAi)* embryos relative to control L4440 embryos were classified and compared as in Figure 4. A) Genes expressed in somatic lineages, embryo-expressed genes, and X-linked genes are overrepresented among genes showing decreased RNA Pol II occupancy in *sig-7(RNAi)* embryos. B) Genes expressed in the germline show an increase in RNA Pol II occupancy in *sig-7(RNAi)* embryos.

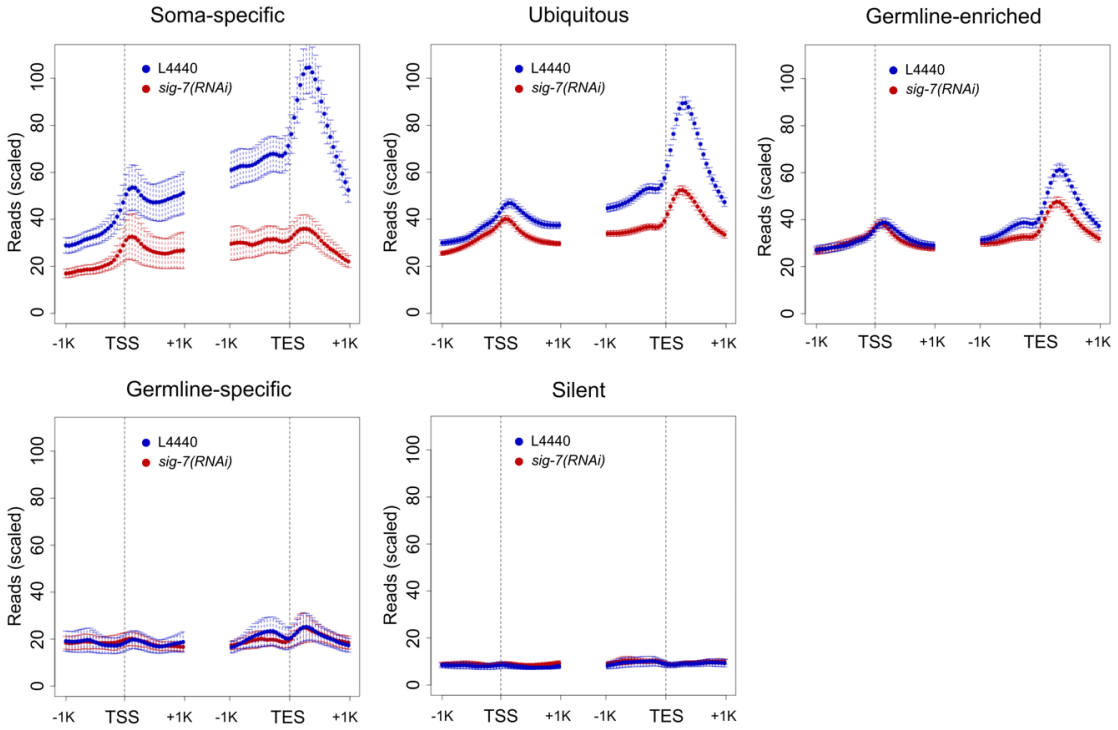


Figure 18. RNA Pol II distribution within genes is altered by *sig-7(RNAi)*. Metagene displays of RNA Pol II ChIP-seq enrichment 1kb upstream and downstream of the Transcription Start Site (TSS) and Transcription End Site (TES) for different classes of genes in L4440 RNAi control (blue) and *sig-7(RNAi)* (red). Genes were categorized as in Figure 4. Graphs illustrate combined results from 2 biological replicates. Error bars indicate the 95% confidence interval of the mean signal, indicated by the circles. Reads were normalized as indicated in Materials and Methods. Note that "germline-enriched" genes include genes expressed in all tissues that exhibit enhanced expression in germ cells.

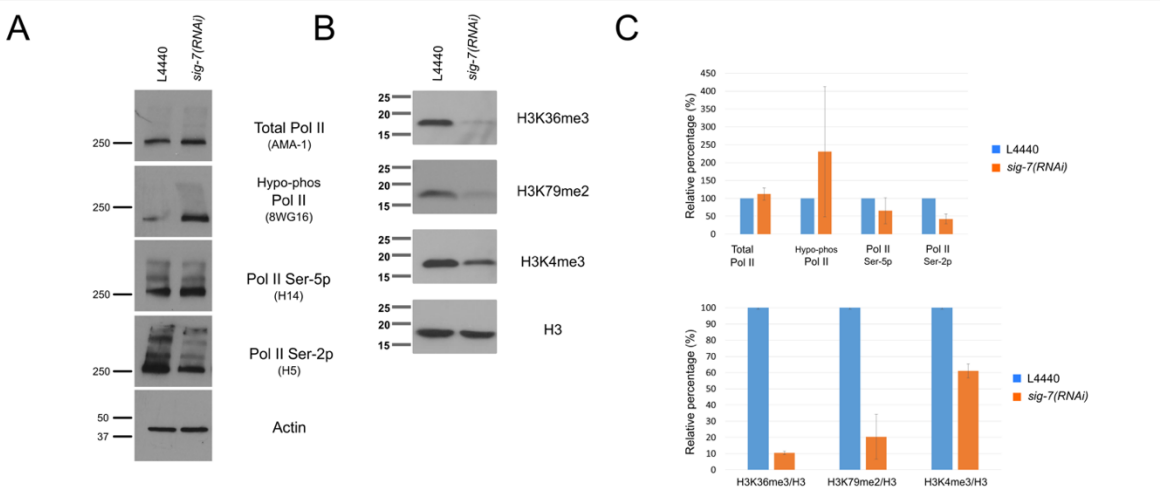


Figure 19. RNA Pol II phosphoepitopes and histone H3 modifications associated with transcription elongation are altered by *sig-7(RNAi).*

A) Equal amounts of lysates from L4440 control RNAi and *sig-7(RNAi)* embryos were analyzed by western blots probed with antibodies against all RNA Pol II isoforms (anti-AMA-1), hypo-phosphorylated Pol II (8WG16), Pol II Ser5-P (H14), and Pol II Ser2-P (H5). Anti-actin antibodies show equivalent protein loading. B) Lysates were analyzed by western blots as in (A) using antibodies against histone H3 modifications associated with transcription elongation, H3K36me3 and H3K79me2, and the 5' end of active genes, H3K4me3. Antibodies against total histone H3 (H3) show similar protein loading. C) Quantification of the probe signals from each RNA Pol II antibody relative to control L4440 in A (Top panel). Quantification of probe signals from each histone modification relative to control L4440 and normalized to total H3 for each sample in B (Bottom panel). Error bars = S.D. from two separate experiments.

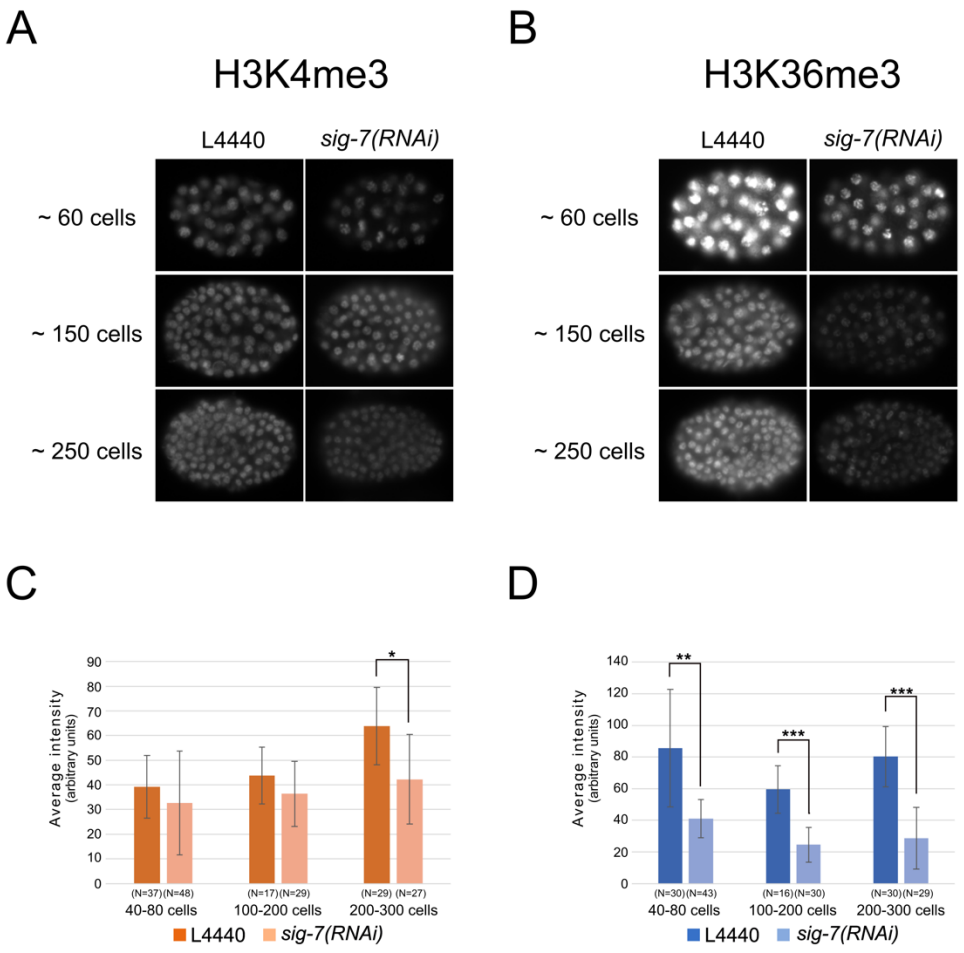


Figure 20. H3K36me3 is more significantly affected than H3K4me3 during embryonic development in *sig-7(RNAi)*.

A) H3K4me3 staining patterns are shown for various embryonic stages. The significant difference is only observed in older embryonic stage between L4440 control and *sig-7(RNAi)*. B) H3K36me3 staining patterns are shown for various embryonic stages. Unlike the H3K4me3 pattern, H3K36me3 shows noticeable difference from earlier stage and continues to be decreased. Quantification of the average intensity of H3K4me3(C) and H3K36me3(D) staining from various stages are shown. N= total number of embryos quantified. Error bars = S.D. The statistical significance of the difference in average intensity is indicated by asterisks (* = $p \leq 0.01$, ** = $p \leq 10^{-5}$, *** = $p \leq 10^{-10}$)

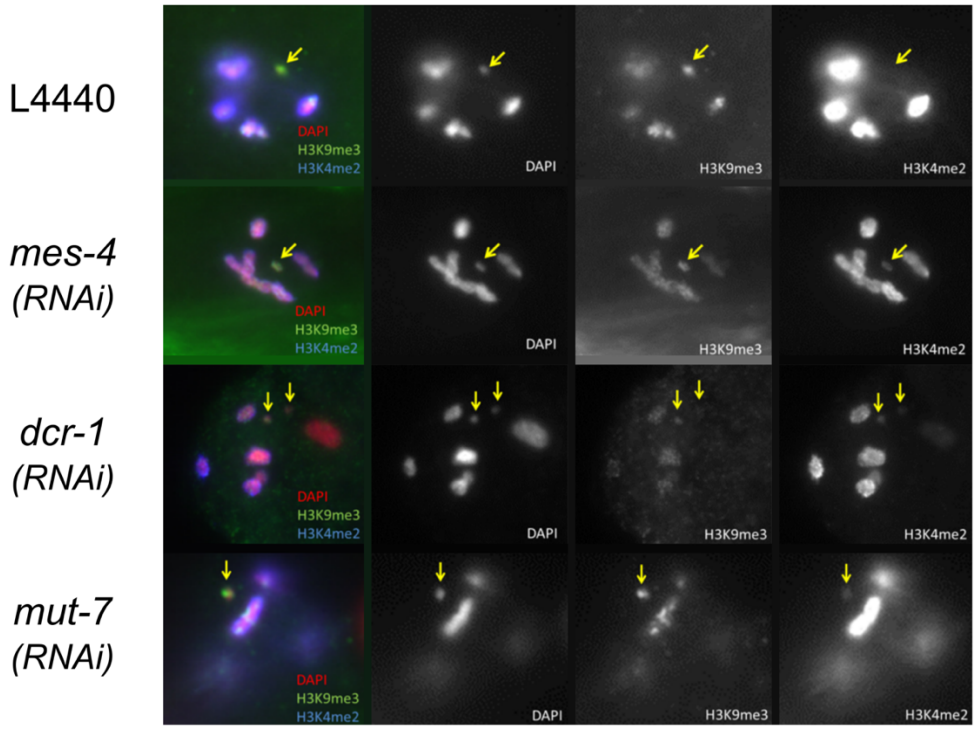


Figure 21. ccEx7271 repetitive array gets *de novo* H3K4 di-methylation mark when it is desilenced in adult oocyte.
The immunostainings for H3K4me2 (euchromatin mark) and H3K9me3 (heterochromatin mark) in diakinetic maturing oocyte are shown after RNAi experiments knocking down genes responsible for repetitive transgene silencing in adult germline. ccEx7271 transgene is indicated by yellow arrow. Note that the image of H3K4me2 staining for L4440 is taken with much longer exposure to make sure H3K4me2 marks are not there when ccEx7271 repetitive array is silenced in the control RNAi.

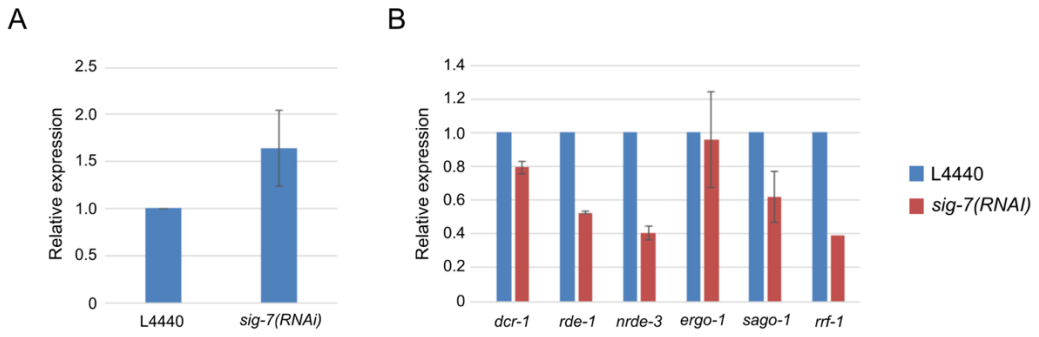


Figure 22. The increased GFP reporter expression is accompanied by decreased expression of RNAi pathway genes in *sig-7(RNAi)* animals.
The abundance of GFP reporter transcript (panel A) and several transcripts of known RNAi pathway genes were measured by qRT-PCR in RNAi adults. RNA levels were normalized to 18S RNA levels and plotted relative to L4440 control in each experiment. Error bars = SD from three technical replicates each of two biological replicates (panel A) and four biological replicates (panel B).

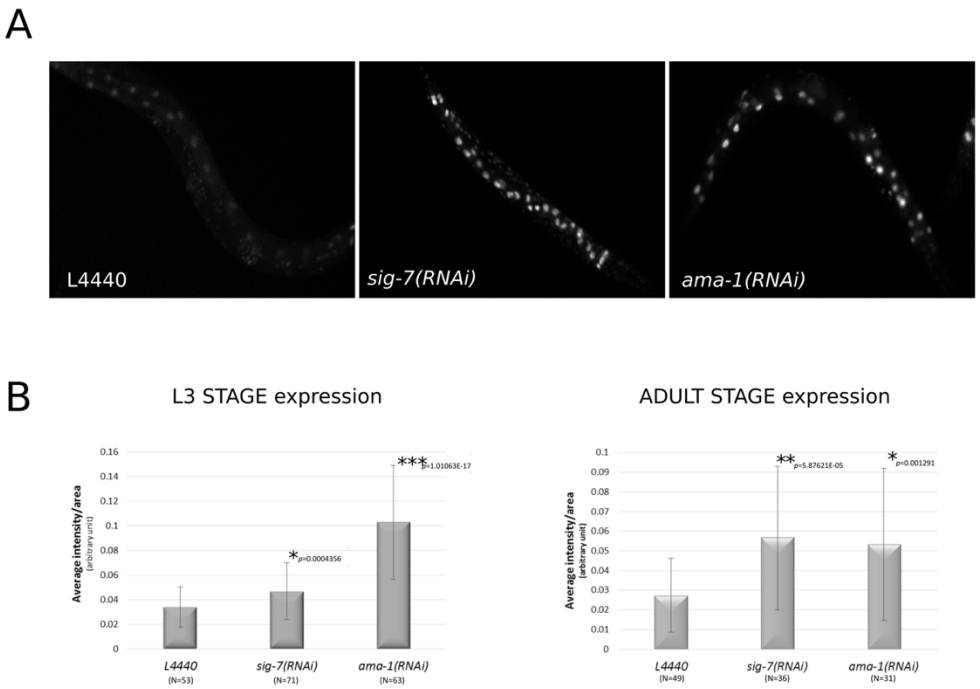


Figure 23. Repetitive transgene array can be desilenced by knocking down RNA Pol II by RNAi

A) Live GFP reporter expression from the *ccEx7271* repetitive array is shown for respective RNAi experiments as indicated. Both *sig-7(RNAi)* and *ama-1(RNAi)* result in similar level of desilencing compared to the control L4440 RNAi. B) The average intensity of live GFP expression are quantified using image J as described in material and method. N= total number of larval or adult worms quantified. Error bars = S.D. The statistical significance of the difference in average intensity is indicated by asterisks (* = $p \leq 0.01$, ** = $p \leq 10^{-5}$, *** = $p \leq 10^{-10}$)

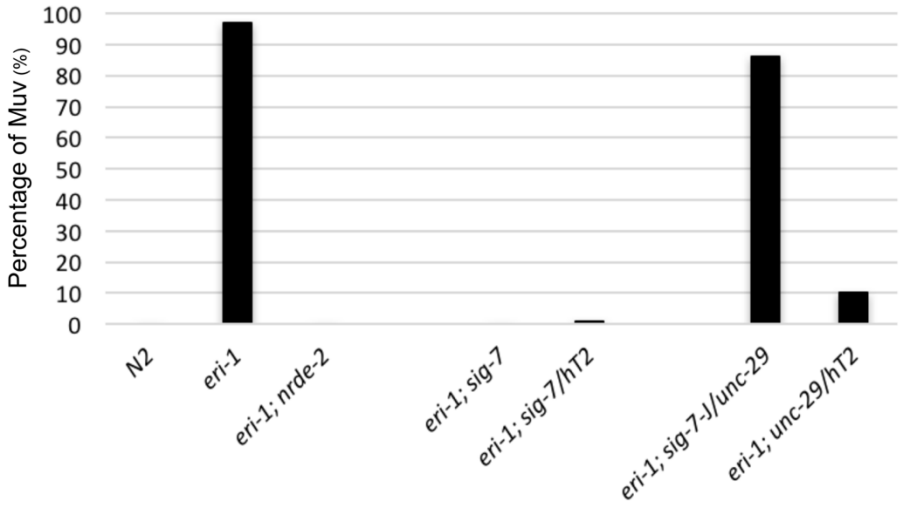


Figure 24. The presence of genetic balancer, not the defects of SIG-7 function, is responsible for suppression of Muv phenotype in *eri-1* background.
The percentage of a multi-vulva (Muv) is plotted for respective strains after RNAi against *lin-15b*.

CHAPTER3: H3K4 and H3K36 Methylation in Germline stem cells and Their Roles in Transgenerational Maintenance of Germline Function

Introduction

A brief history of the emergence of the field of epigenetics

Over the years, biologists have often encountered challenges to explain stable changes of genome function that could not be explained by Mendelian inheritance and have sought additional mechanisms to explain these phenomena. One of the earliest experiments, performed by Holliday and Pugh, indicated that chemical modifications to DNA could affect gene expression (250). Since then, several phenomena have been reported to be controlled by the methylation of bases in DNA. The inactivation of the X chromosome first observed in in female mammalian embryos in 1961 is one good example, in which one copy of the X chromosome is randomly inactivated, and its inactivated state is stably inherited. Several findings indicated that DNA methylation was an essential part of the underlying mechanism (251). The later demonstration of the heritable maintenance of the methylation of injected plasmid DNA into *Xenopus* eggs suggested that there is a mechanism to stably inherit this DNA methylation (252). It was further shown that using hemimethylated DNA, where only one strand of double stranded DNA is methylated, a methylation pattern from one strand can be faithfully copied onto the other strand. This finding led to the later discoveries of the maintenance DNA methyltransferase Dnmt1, and the *de novo* DNA methyltransferases Dnmt3a and Dnmt3b (253-255). These initial studies regarding DNA methylation established the epigenetic field, because these findings demonstrated that modifications like DNA methylation can be stably inherited, similar to how genetic mutations are inherited after each cell division. The field of epigenetics has since grown fast, and many other mechanisms, including chromatin modifying factors/regulators have been discovered to play multiple roles in the

epigenetic regulation. The chromatin-based mechanisms include histone modifications, histone variants, chromatin remodelers and histone chaperones. The coordinated actions of these chromatin factors/regulators have multiple roles in many diverse biological processes, including regulating nuclear architecture (256), modulation of cell signaling pathways (257), maintenance of pluripotency (258), and the establishment and inheritance of epigenetic memory, among others (259-262).

A rationale for the need of maintaining cellular memory

Most multicellular organisms arise from a single cell zygote that results from a fusion of two gametes. In many organisms, the zygote then goes through multiple rounds of cell division called the cleavage stage. The factors driving these early stages often come from maternal contribution, and such cleavage stage embryos are composed of largely homogeneous cells with similar developmental potentials (263). With the onset of zygotic transcription, often referred as zygotic gene activation (ZGA), each cell becomes more distant from each other in terms of its developmental potential. This leads to different lineage specification paths with distinct gene expression programs that can be partially determined and/or maintained by a cell's chromatin architecture (264, 265). Each lineage ultimately gives rise to different tissues and organs of an intact adult animal.

Among all the cells engaging in different developmental paths, only the germ cells will contribute directly to the next generation. The progenitors of embryonic germ cells, called primordial germ cells (PGCs), often undergo partial differentiation during embryogenesis prior to their specification, yet PGCs are thought to retain an underlying pluripotency. PGCs then give rise to a proliferating population of embryonic germ cells,

which then provide a stable pool of germline stem cells (GSCs). GSCs then continue to provide germ cells that further differentiate into meiotic cells and then undergo gametogenesis (266, 267). At the end, these specialized gametes will fuse and become a single cell, totipotent zygote. Therefore, there must be some mechanisms to maintain and/or re-establish pluripotency as germ cells experience differentiation and gametogenesis pathways. For cells to maintain or reprogram back to pluripotency, there must be a form of memory that defines the pluripotent state to instruct the cells. The need for such memory function is not limited to the germ cells (See below for more discussion).

One candidate mechanism that can serve such a memory function is epigenetic regulation. A memory in general implies that there is a preexisting history, and such history is being remembered over time. For cells, transcription activity can be considered as a history of cellular activity. Therefore, epigenetic marks that are coupled to transcription may provide a history of cellular activity, and the established history can be maintained by enzymatic activity in a transcription independent manner. In this Chapter I describe my initial experiments that investigate the roles of H3K4 and H3K36 methylation, and their transcription-dependent and –independent functions, in maintaining normal germline function within and across generations.

The role of DNA methylation in epigenetic memory

The adenosine and cytosine bases in DNA in can be modified by methylation. In mammalian genomes, a methyl group is added to carbon 5 of the pyrimidine ring of cytosine of forming 5-methylcytosine (5mC). This cytosine methylation is catalyzed by

two *de novo* DNA methyltransferases DNMT3A and DNMT3B (253). There is an additional third protein DNMT3L that is catalytically inactive, but was shown to mediate de novo DNA methylation by recruiting DNMT3A and DNMT3B to loci where H3K4 is unmethylated (268). These methylation patterns are then maintained by the DNA methyltransferase DNMT1 during the S phase of a cell cycle, thus faithfully maintaining the previously established methylation patterns (269). The maintenance methylation pattern depends on the cytosine being in the context of CG dinucleotides. A CpG islands (CGI), a short stretch of DNA sequence with more frequent CG appearance than other regions, are usually hypomethylated and often located around the promoter of housekeeping genes that are expressed in all cells, while other non-CGI intergenic regions, including repetitive elements are hypermethylated to suppress expression in human (270). Therefore, DNA methylation near the promoter is considered a repressive mark for gene expression. DNA methylation in plants occurs at CG, CHG, and CHH (H stands for A, C, or T), and altered methylation patterns were shown to be stably inherited through both mitosis and meiosis, and hence transgenerationally (271). However, recent findings suggest there are more complex patterns and distinct mechanisms underlying DNA methylation that can function more than just on and off switch for a gene expression. For example, a comparative analysis of the DNA methylation pattern between embryonic stem cells and differentiated fibroblast cells showed an existence of different classes of genes based on distinct DNA methylation patterns in both cell types (272). They also compared the different clonal populations of both cell types and observed that embryonic stem cell clonal populations showed a more stable ground state of low methylation status without noticeable distinction among them while fibroblast

clonal cells showed a persistence of distinct clonal DNA methylation patterns suggesting DNA methylation pattern might confer a cellular memory function. Another study from human cell lines identified methylation pattern differences between embryonic stem cells and differentiated cells. They observed that the DNA cytosine methylation in IMR90 fibroblast cells occurs strictly in the CG context while 25% of the total DNA cytosine methylation in H1 stem cells occurs in non-CG contexts (mCHG/mCHH) (273). They also showed that this non-CG methylation is specific to stem cells by demonstrating the loss of non-CG methylation upon induced differentiation, and the recovery of non-CG methylation in induced pluripotent stem cells. Another interesting finding from this study was a gene body enrichment of the non-CG methylation that positively correlates with gene expression. A similar finding has been reported from genome-wide DNA methylation analysis in *Arabidopsis*. They found that ~33% of all the annotated genes had gene body methylation, and they also observed a positive correlation between the gene body methylation and expression level of the respective genes (274). Another interesting observation is that gene body methylation is not due to non-CG type methylation that is usually mediated by small-RNA directed methylation (RdDM) pathway. This may suggest that these type of CG gene body methylations have a different function. They also showed that genes containing the CG methylation at their promoters tend to be expressed in a tissue specific manner, therefore different cell types may achieve cell type specific transcription by preventing unnecessary gene expression via DNA methylation. It appears that the location of the methylation seems to confer its distinct conserved functions rather than the types of methylation.

It is possible that the gene body methylation described in above studies may have some unrecognized function. The authors of the works mentioned above proposed that gene body methylation may have some function in splicing processes, since this type of methylation correlates well with the exon enrichment pattern of H3K36 methylation on gene bodies (89, 244). Alternatively, it is also possible that gene body methylations may play a role in transcriptional memory of expressed genes that may contribute to some aspects of epigenetic inheritance. The role of H3K36me3 for providing a germline expression memory, and the specific recognition of DNMT3A for H3K36me3 provide a potential functional interaction that may be important for transmitting the transcription history during development (194, 275).

The Role of Histone Methylation in Epigenetic Memory

The DNA in eukaryotes is organized into a complex structure called chromatin composed of DNA and proteins. The basic building block of chromatin is the nucleosome where 147 base pairs of DNA is wrapped around a histone octamer composed of two copies each of H2A-H2B and H3-H4 dimers (276, 277). The N-terminal "tail" of each histone within the nucleosome contains a flexible structure that is important for the stability of histone core, and are sites for numerous post translational modifications that are important for various biological processes (278-282).

The modifications to histones can have various impacts (283). For example, an acetylation of histone lysine residues can neutralize the positive charge of lysine, therefore it can weaken or change the property of the interactions between the histone tail and DNA. Unlike acetylation, the methylation of histones does not alter the charge of

modified residues. The methylation of lysine can also be mono-, di-, or tri-methylated, and arginine residues can also be methylated by distinct enzymes (284). The added methyl group(s) on the side chain of the targeted lysine or arginine residues, and the position and the level of methylation at the residue being targeted, can serve as a binding platform for proteins that can impart different biological functions. For example, HP1 protein can recognize and bind H3K9me3 to help the establishment and spreading of heterochromatin, while H3K4me3 can be recognized by the TAF3 subunit of TFIID during transcription initiation (285, 286).

So how can such modifications play a stable role in epigenetic memory, especially during DNA replication and cell division? All the cells in multicellular organisms have to go through mitosis during all stages of development regardless of their cell types. Even germ cells arise from germline stem cells initially before they undergo gametogenesis, so virtually all types of cells will be affected if such information is not maintained or transmitted properly during cell division.

One emerging mechanism is called “mitotic bookmarking”. During the interphase of the cell cycle, there is transcription of cell cycle regulators during the transition from G1 to S phase (287). However, as cells enter the mitosis stage of the cell cycle, chromosomes condense, and transcription shuts down. There is also increased phosphorylation of histones, and as a result, transcription factors dissociates from the mitotic chromatin (288). However, some transcription factors and histone modifications have been observed to be stably associated with the condensed chromosomes, and these stably associated factors are proposed to allow quick reactivation of gene transcription once the cells exit mitosis [see these reviews for the complete list of transcription factors

and histone modifications, (289, 290)]. Particularly, H3K4me2 and H3K4me3 at the promoters and gene body regions are enriched at constitutively active genes before and after the mitosis suggesting these marks can persist through cell cycle process despite the dynamic reorganization of the nucleosomes that takes place (291). One interesting observation is the gene body enrichment of H3K4me3, which deviates from the normal promoter restricted enrichment patterns associated with the gene activation. This enrichment pattern may reflect the different biological function than the promoter enriched H3K4me3, perhaps marking the previously transcribed genes serving as an epigenetic memory. Consistent with this idea, a similar type of broader enrichment referred as the Broad H3K4me3 domain was observed to be a conserved feature across numerous species from the re-analysis of H3K4me3 distribution patterns of more than 200 data sets from various organisms (292). Genes with the Broad H3K4me3 domain had a better transcriptional consistency rather than a higher expression level supporting the epigenetic memory function. These distinct H3K4me3 enrichment patterns can be achieved if different methyltransferase complexes are targeted to distinct classes of genes. Indeed, this kind of distinct complexes has been reported very recently from a planarian study (293). Two distinct methyltransferase for H3K4me3, Set-1 and MLL1/2, showed distinct target preferences with its associated distinct enrichment patterns within the stem cells. Therefore, it clearly demonstrates the existence of distinct H3K4methyltransferase complexes marking functionally different classes of genes within the same cell. Whether this kind of mechanism is conserved feature of all species will need further study.

One cell type that certainly needs such mitotic bookmarking is the stem cell populations. During the proliferation stage, stem cells go through constant mitotic divisions while each daughter cells equally maintain pluripotency. Therefore, any factors contributing to pluripotency must be maintained during the clonal expansion of stem cells before they are induced to differentiate into a particular cell type, and any compromised pluripotency can further affect its downstream differentiation process. Several chromatin factors have been shown to be important for stem cell proliferation and differentiation.

Among them, the co-localization of the H3K4 and H3K27 tri-methylation in clusters of homeotic transcription factors and other developmentally important genes, referred to as bivalent or poised chromatin, have gained wide attention for its proposed dual roles of silencing developmental genes in ES cells while priming them for later activation in differentiated cells (294-297). Thus, it is plausible that one methyltransferase may be responsible for establishing H3K4 methylation at bivalent domain while other complexes may regulate the H3K4 methylation at other regions in stem cells.

Numerous publications have identified genes found in bivalent domains, and there have been at least 1165 genes cross-validated in more than two independent studies out of total 6897 genes reported from multiple studies (298). Although characterization of such domains in multiple organisms have been done, we are still far from understanding the mechanism of how these marks are directed to each loci in the genome and the precise biological role of these marks. Several studies have proposed potential mechanism for how bivalent domains may get established. A correlation between binding sites of pluripotency transcription factor and binding sites of WDR-5, a core subunit of

Set-1/MLL complex, and PRC2 complexes have been reported to exist for subsets of genes found in bivalent domain (299, 300). Cfp-1, a component of the Set-1/MLL complex, was shown to tightly correlate with unmethylated CpG islands that are enriched for H3K4me3 in the mouse brain, and *HOTTIP*, a long non-coding RNA, has also been shown to interact with the Set-1/MLL complex to induce *HOXA* expression in human cell lines (301-303). Interestingly, a direct interaction between WDR-5 and Oct-4 has been observed, and a significant overlap among target genes of WDR-5 and key transcription factors for "stemness" including Oct4 has been observed (299). WDR-5 is notably highly expressed in both ES cells and induced pluripotent stem cells (iPSCs) and downregulated in differentiated cells, and loss of WDR-5 causes loss of self-renewal and enhanced differentiation (299). The Tip60–p400 complex of the INO80 chromatin remodeling complex family has been shown to localized to both active and silent genes, but the investigators observed noticeable co-localization of p400 with H3K4me3-marked early differentiation genes in bivalent domain that are suppressed in ES cells (304). These findings suggest that the establishment of histone marks associated with bivalent domains may require coordinated interactions among multiple biological processes.

What are the roles of H3K4 and H3K27 methylation in bivalent domains? Many studies of H3K27 methylation by PRC1/2 complex have provided evidence supporting its proposed function to repress the expression of lineage specification factors (305-307). The PRC complex-mediated repression of lineage specification factors was postulated to be important for the maintenance of pluripotency in ES cells. The tri-methylation of H3K4 in bivalent domains of ES cells has been shown to mark key developmental genes that are poised for transcriptional activation during early development. Repression is

released by loss of H3K27 methylation when ES cells undergo differentiation pathways (308, 309). Depletion of Dpy-30, a component of the SET1/MLL complex in ES cells, did not affect gene expression of most genes nor its ability for self-renewal, but the H3K4me3 mediated by Dpy-30 was shown to be essential at poised developmental genes upon differentiation mediated by RA-induction, suggesting that H3K4me3 mediated by Dpy-30 containing SET1/MLL complexes may have specific role for marking poised developmental genes (310). In the same study, the increased level of H3K27me3 levels at the TSSs of ESC-specific genes (Klf4, Oct4 and Upp1) upon RA-induction, which is thought to be mediated by PRC2 complex, was also dependent on Dpy-30 or RbBP-5 suggesting more complex requirement for a crosstalk between PRC2 complex and SET1/MLL complex.

The requirement for a proper establishment of H3K4 methylation in stem cells is evident. Although it has not been tested whether this establishment of H3K4 methylation, particularly in bivalent domain, requires any means of memory function, the following studies regarding H3K4 methylation suggest it is possible that H3K4 methylation can serve as a memory function.

An observation of the inheritance of the active transcription state has been reported during nuclear transplantation (NT) experiment where isolated somatic nucleus is transplanted into enucleated egg (311). The donated nucleus of an egg obtained by the NT experiment usually gets reprogrammed by cytoplasmic contents of the recipient egg, but they observed a persistence of a donor cell specific gene expression. A further investigation revealed that the lysine 4 residue of H3.3 histone is critical for mediating the memory function of the transcriptional active state (312). Interestingly, this

maintenance of the active state occurs independent of the transcription. A similar finding has also been reported from *Dictyostelium* showing that the memory of transcriptional state is lost by the endogenous mutation changing the lysine 4 of H3 to an alanine (313). In *C. elegans*, our lab has shown that a mutation in *spr-5*, a homolog of the H3K4me2 demethylase LSD1, caused a progressive accumulation of the H3K4me2 in germ cells correlating with the severity of sterility over generation (314). These findings support that proper establishment and maintenance of the H3K4 methylation during development is one important component for transmitting epigenetic memory over a successive generation.

The H3K36 methylation is another histone modification linked to the epigenetic memory function in *C. elegans*. It was shown that H3K36me3 of the germline expressed genes in the adult are maintained in the early embryos by the activity of MES-4, a maintenance methyltransferase for H3K36me3 (194, 236).

In this study, I provide the first evidence for a crosstalk/interdependency between the H3K4 methylation and the H3K36 methylation specifically in the germline stem cells that is essential for a proper germline development. Especially, I describe the importance of an uninterrupted flow of an epigenetic information carried out by the methylation at H3K4 and H3K36 that can occur in two distinct modes, transcription dependent and independent, throughout the development. I would like to explore the significance of this crosstalk during the germline development, particularly its potential function in the epigenetic memory maintenance, and expand our search for more interaction with other histone modifications as well as recently discovered adenine N6-methylation (6mA) in *C. elegans* DNA in the near future.

Result

GFP tagged WDR-5 protein is expressed in adult gonads, and the maternally inherited WDR-5 protein shows a persistence over a few cell divisions during early embryogenesis.

To study the function of WDR-5 dependent H3K4 methylation in the germline stem cells (GSCs), I expressed a tagged version of WDR-5 with expression restricted to the GSCs in adult gonads, 1XFLAG::GFP::WDR-5.1. I obtained the pTS1 plasmid, which contains endogenous *wdr-5.1* genomic sequences with an N terminally inserted *gfp*. I then subcloned this into a plasmid with a *glh-2* promoter and *fbf-2* 3' UTR sequence, which restrict expression and translation, respectively, to the GSCs. I used the MosSCI integration technique to generate a transgenic animal with a single copy insertion (188). It has been shown that 3' UTR is the determining factor for special and temporal gene expression patterns within the gonads. Therefore, I chose the *fbf-2* 3'UTR, which was reported to limit gene expression to the distal region of the gonad, where the GSCs reside (315). I found, however, that the expression pattern was not limited to the GSCs, but instead was expressed throughout the gonad (Figure 2). The 1XFLAG::GFP::WDR-5.1 protein was expressed in adult germline and in one cell embryos, but eventually got diluted during several rounds of cell division in embryos. This pattern is a typical expression pattern of maternally inherited proteins.

I also confirmed that the size and expression was what was predicted for an intact 1XFLAG::GFP::WDR-5 by western blot (Figure 3, Lane 2). As a positive control, I prepared worm lysate from another strain expressing GFP tagged WDR-5 driven by its endogenous promoter and 3'UTR, which is expressed in every cell type (Figure 3, Lane

4). For a negative control, I prepared worm lysate from WT animals lacking GFP expression (Figure3, Lane1). The 1XFLAG::GFP::WDR-5 was detected at the expected size of 68.95kDa (Figure 3).

1XFLAG::GFP::WDR-5 can restore H3K4 methylation in the *wdr-5.1(ok1417)* mutant.

In order to test whether the 1XFLAG::GFP::WDR-5.1 can function as one of the core members of the H3K4 methyltransferase complex, I created a line expressing 1XFLAG::GFP::WDR-5.1 in the *wdr-5.1(ok1417)* mutant. In the *wdr-5.1(ok1417)* gonad, H3K4me3 is significantly reduced in the GSCs. Therefore, I did immunostaining for H3K4me3 in this transgenic line and showed that H3K4me3 methylation is back to normal levels in the *wdr-5.1(ok1417)* background (Figure 4). This result confirms that 1XFLAG::GFP::WDR-5.1 can associate with other components in the methyltransferase complex to restore H3K4 methylation in the GSCs of the *wdr-5.1* mutant.

Both the transcription independent H3K4 methylation and the transcription dependent H3K36 methylation is required for the proper establishment of H3K36 methylation by MES-4 in the GSCs.

H3K4 and H3K36 methylations mediated by WDR-5 and MES-4, respectively, are maintained by transcription independent mechanisms in the GSCs, I was interested in potential crosstalk between these two modifications. I therefore examined H3K4 and H3K36 methylation in the GSCs of the *wdr-5.1(ok1417)* mutant. The loss of H3K4me3

in the GSCs of the *wdr-5.1(ok1417)* mutant did not noticeably affect the gross pattern of H3K36me3 (Figure 5). I then examined *met-1(n4337); wdr-5.1(ok1417)* double mutants to determine the effect of loss of "maintenance" H3K4me3 on MES-4-mediated H3K36me3 in the absence of MET-1 function. In this double mutant, the only methyltransferase for H3K36 is MES-4. I observed a marked increase of H3K36me3 level specifically in the GSCss where H3K4me3 is lost (Figure 5). It is interesting that mutations of both *wdr-5.1* and *met-1* are required to see this increase of H3K36me3 by MES-4, since it is not observed in the *wdr-5* single mutant. It appears that both the transcription independent H3K4 methylation mediated by the WDR-5 complex and the transcription dependent H3K36 methylation mediated by MET-1 are required for normal H3K36 methylation by MES-4 in the GSCs. This data suggests that there may be a novel inter-dependency or a crosstalk between H3K4 and H3K36 methylation, during germline development. The exact mechanism for such interaction needs further studies.

The *met-1(n4337); wdr-5.1(ok1417)* double mutant causes a transgenerational sterility due to various germline defects

As it was described, the *met-1(n4337); wdr-5.1(ok1417)* double mutant displayed an improper increase of H3K36me3 in the GSCs indicating an epigenetic imbalance in these cells. I thus further explored the consequences of this imbalance. I measured the brood size of the *met-1(n4337); wdr-5.1(ok1417)* double mutant compared to the *met-1(n4337)* and the *wdr-5.1(ok1417)* single mutants. As previously reported, the *met-1(n4337)* mutant showed a slight decrease in brood size with no significant change over multiple generations while the *wdr-5.1(ok1417)* mutant showed a progressive decrease in

brood size over generations (Figure 6B). In contrast, the *met-1(n4337); wdr-5.1(ok1417)* double mutant showed a significant reduction in brood size immediately, and produced a completely sterile population after two generations (Figure 6B). I could also replicate the same result by performing RNAi targeting *wdr-5.1* in the *met-1(n4337)* mutant background. Introducing the ckSi35 transgene expressing GFP tagged WDR-5 into the *met-1(n4337);wdr-5.1(ok1417)* double mutants rescued the sterility defect. I then used RNAi targeting GFP in this strain for further experiments, which targeted the rescuing transgene, rather than performing crosses, for subsequent experiments.

The *gfp(RNAi)* exposure from L1 larvae to adult stage was sufficient to cause 100% sterility (Figure 7C). The *gfp(RNAi)* animals displayed a significantly disturbed germline morphology, with reduced gonad size and irregularly organized germ cells (Figure 7A). To further examine the germline phenotype, I performed DAPI staining to visualized the DNA structure. The germ cell nuclei of the *gfp(RNAi)* animal displayed various types of germline development defects. The most frequent defect was a masculinization of germline (*mog*) phenotype where the animals failed to switch from spermatogenesis to oogenesis as adults, yielding adult hermaphrodites with continuous sperm production. I also observed much smaller sized gonads filled with fewer germ cell nuclei, many of which were irregular in size (Figure 7B).

The *wdr-5.1(RNAi)* causes a precocious sterility in the $M^{+}Z^{-}$ *mes-4(ok2326)* mutant

I showed that the H3K4 methylation mediated by WDR-5 complex is partially important for proper MES-4 function during germline development. Therefore, I wanted to see if the loss of WDR-5 function will worsen the *mes-4* phenotype by RNAi. The

mes-4 mutant exhibits maternal effect sterility (*mes*); i.e., *mes-4* homozygous mutants segregated from genetically balanced *mes-4* parental strain are fully fertile in F1 generation due to the maternally provided MES-4 protein acting during an early embryogenesis (Figure 8A). These fertile offspring are maternal + zygotic – (M+Z-) for MES-4 activity. The M^-Z^- offspring of the M+Z- animals completely sterile since they lack even maternal MES-4. I performed *wdr-5.1(RNAi)* from L1 stage of the genetically balanced *mes-4/+* strain, and continued RNAi on their *mes-4* M^+Z^- homozygous offspring (Figure 8C). In contrast to the *mes*-4 M+/Z- RNAi controls, which exhibited no fertility defects, 42% of M^+Z^- *mes-4*;*wdr-5.1(RNAi)* displayed a precocious *mes* sterility phenotype (Figure 8 A and B).

WDR-5 males do not produce cross progeny

In trying to set up a genetic cross between *wdr-5.1(ok1417)* males and *met-1(n4337)* hermaphrodites, I noticed that there was no cross progeny despite numerous attempts and various efforts. In contrast, the reciprocal cross using *met-1(n4337)* males was successful, and this cross was used to generate the *met-1(n4337); wdr-5.1(ok1417)* double mutant for this study. Whether the failure of the genetic cross using the *wdr-5.1(ok1417)* male was due to problems during a spermatogenesis yielding defective sperms or physical defects in mating apparatus in a tail region has not been determined. I do not observe any obvious morphological defect in the *wdr-5.1(ok1417)* males at the gross level, but I did not perform any microscopic level of the analysis examining many aspects of the mating apparatus. The fact that *wdr-5.1(ok1417)* hermaphrodites are fertile, albeit with a decreased brood size, suggests that functional sperm can be produced in

wdr-5.1(ok1417) hermaphrodites. The apparent sex-dependent mating defect I observed for this mutant needs further examination.

Loss of *set-17* or *set-30* does not noticeably affect transcription dependent H3K4 methylation in the adult germline

Our lab and others have shown previously that SET-2, a homolog of yeast Set1, is the methyltransferase that is required to maintain WDR-5-dependent H3K4me3 in the GSCs (316). The methyltransferase responsible for transcription dependent H3K4me3 is not known in *C. elegans*. I therefore examined either mutants for potential H3K4 methyltransferases, or targeted them by RNAi . Two proteins that have been reported to have H3K4 methyltransferase activity *in vitro* are encoded by the genes *set*-17 and *set-30* (317). I performed RNAi against *set-30* in the *wdr-5.1(ok1417)* mutant, and RNAi against *wdr-5.1* in the *set-17(n5017)* mutant. As expected for the loss of WDR-5 alone, in both cases H3K4me3 was significantly reduced in the GSCs, with a minor reduction in meiotic nuclei in the adult gonad (Figure 9, row 3 colum 4). Surprisingly, I did not observe any reduction of H3K4me2 even in the *wdr-5.1(ok1417)* mutant controls. This finding is different from what has been previously reported. However, I also did not observe any significant decrease in either H3K4me2 or H3K4me3 (Figure 9). It is possible that these genes are redundant, so a knock-down of both will be important to examine in future studies.

Discussion

In this study, I report an interaction/crosstalk between H3K4me3 and H3K36me3 that is important for germ cell function. I am particularly interested in understanding the function of transcription independent methylation for both K3K4 and H3K36. Why do early embryos and GSCs share this maintenance mode of these histone modifications? What are in common between these two population of cells? The common property shared between early embryos and GSCs is their 'stemness' or 'pluripotency'. During *C. elegans* embryogenesis, maternal factors including mRNAs and proteins expressed in the parental germline control the early cell divisions and developmental patterning until the onset of a gastrulation, when zygotic gene activation takes over the control of further differentiation programs (318-321). Although some of the early blastomeres have begun cell fate specifications, these cells are still highly plastic and ectopic expression of cell-type inducing transcription factors can convert all cells to a single somatic fate (322-326). However, this kind of developmental plasticity is only observed up to the 80 cell stage (321). This suggests that pluripotency is preserved only during early embryogenesis. This timing correlates well with the when the mode of H3K4 and H3K36 tri-methylation switches from being transcription independent to transcription dependent (Figure1). This observation suggests that both H3K4me3 and H3K36me3 that are maintained in the chromatin at these early stages contribute to maintaining a pluripotent state. What epigenetic patterns are maintained, and how might they be important for pluripotency? MES-4 maintains, in the embryo, H3K36me3 patterns that arrive with the gametes. These patterns match the transcription patterns that are produced by adult germ cells; i.e., in the embryo, MES-4 maintains the H3K36me3 patterns that were established by

transcription in the parental germ cells (194, 327). This suggests that this pattern includes those required for pluripotency, as pluripotency is considered to be a fundamental property of germ cells.

Similarly, H3K4me patterns in the early embryo grossly reflect the patterns observed in adult germ cells (328). Unlike the case with MES-4 and MET-1, the WDR-5 independent H3K4 methyltransferase is as yet unknown, so the patterns of H3K4me3 maintained by WDR-5 alone is not yet known. Identifying the transcription-dependent H3K4 methyltransferase(s) is thus an important goal, and will be one of several aims that we would like to explore further in future studies.

Similar paradigm exists between the GSCs and the differentiating meiotic germ cells in the adult gonad, where the WDR-5 complex and MES-4 are the major transcription independent H3K4me3 and H3K36me3 methyltransferases, respectively, in the GSCs (Figure1). In the adult gonad, H3K4me3 methylation mediated by WDR-5 has been reported to be important for a proper expression of germline expressed genes and repression of somatic genes in the germline (329). This misregulated gene expression caused by the loss of H3K4me3 methylation loosens germ cell identity and allows trans-differentiation of the germline into somatic cell types under certain conditions. The authors proposed that the trans-differentiation is linked to the loss of H3K9 methylation and the loss of PGL-1, a P-granule component that is specific to the germline. However, these phenotypes could be an indirect consequence of the loss of H3K4me3 methylation. First, the H3K4me3 methylation mediated by SET-2/WDR-5.1 complex has been shown to be independent of Pol II activity by our lab (316), so it is not clear whether the misregulation of germline expressed genes is directly linked to the loss of H3K4me3

methylation. Secondly, the major portion of the H3K4me3 methylation mediated by SET-2/WDR-5.1 complex occurs in the GSCs at the distal end of the gonad, so the global loss of H3K9 methylation from the whole gonad does not correlate well with the pattern of the H3K4me3 methylation loss in mutants they have analyzed. Lastly, the trans-differentiation phenotypes are not immediate. The most obvious phenotypes are all seen at F5 generations, which correlates with the generation exhibiting a significant reduction in the germline stem cell pool (316). Although the proposed role of SET-2/WDR-5.1 complex and its phenotypes described in these mutants are very interesting, it requires further investigation to better understand its exact roles and the mechanism for how all these phenotypes make sense. I think the 1XFLAG::GFP::WDR-5.1 transgenic strain and the experimental design I developed in this study will serve well for that purpose. By collecting L3 worms, the only cells expressing the GFP/FLAG tagged WDR5 are GSCs where the majority of H3K4me3 methylation is mediated by SET-2/WDR-5.1 complex. I can then either purify these cells by FACS and perform ChIP-seq for H3K4me3, and/or do anti-FLAG ChIP-seq for WDR-5.1 to identify where this complex is operating in these cells.

Lastly, I found the first example of what appears to be an antagonistic relationship between H3K4me and H3K36me. I found that the H3K36me3 methylation by MES-4 is specifically upregulated in the GSCs of *met-1;wdr-5.1* double mutants, coinciding with where H3K4 methylation is specifically lost. The fact that the loss of the transcription independent H3K4me3 methylation alone is not sufficient to induce such a misregulation of the transcription independent H3K36 methylation raises many interesting possibilities to investigate further.

The loss of H3K4 can cause changes in other chromatin regulators that can now allow MES-4 to bind at more number of loci causing increased H3K36 methylation. Alternatively, the epigenetic information laid out by both the transcription independent H3K4me3 methylation by WDR-5 in the GSCs and continuing or uninterrupted transcription coupled H3K36me3 by MET-1 may be needed for the proper germline transcription history by MES-4 function. In order to test these interesting possibilities, a more thorough comparative analysis of other chromatin modifications including H3K4 and H3K36 in various mutants or combinations of different mutants to see how they are affected in the absence of the other. One potential modification that needs prioritized attention will be the H3K27me3 methylation by MES-2/3/6 complex, a homologous PRC2 complex in *C. elegans*. Like the WDR-5 complex and MES-4, H3K27 methylation by MES-2/3/6 complex occurs in early embryos and in adult germline (330).

We hope that we can learn more details of the function of the histone modifications that act in transcription independent mode and how they are interdependent to each other. Especially, we hope to gain more insight into their potential functions in mediating the transgenerational epigenetic memory using *C. elegans* as a model system.

Material and Method

Worm strains and maintenance

C. elegans strains were maintained at 20 °C and were grown on NGM (Nematode Growth Medium) plates unless it is stated otherwise.

Strains: wild type N2 (Bristol), KW2307: *set-17(n5017),* KW1616: *wdr-5.1(ok1417)*, KW1660: *met-1(n4337)*, KW1083: *cKSi35(unc-119, glh-2p:gfp:1XFLAG:wdr-5.1ORF:fbf-2 3'UTR)II*, KW1087: *cKSi35(unc-119, glh-2p:gfp:1XFLAG:wdr-5.1ORF:fbf-2 3'UTR)II, wdr-5.1(ok1417)III*, KW1119: *met-1(n4337)I*, *cKSi35(unc-119, glh-2p:gfp:1XFLAG:wdr-5.1ORF:fbf-2 3'UTR)II, wdr-5.1(ok1417),* KW1032: a strain expressing GFP tagged WDR-5.1 was a gift from Palladino lab (331).

1XFLAG::GFP::WDR-5.1 transgenic strain generation

To generate the GFP tagged WDR-5.1 construct, I used the *glh-2* promoter to drive germline-restricted expression. The *glh-2* promoter sequence was amplified from the pJL43.1 vector, that has been verified for germline expression in our lab. The sequence of the *gfp::wdr-5.1* ORF was subcloned from the pTS1 vector with the forward primer that contains 1XFLAG sequence. I then cloned in the 3'UTR region of the *fbf- 2* gene to limit the translation of WDR-5.1 to the GSCs in the adult gonad, as previously reported (315). The cloned fragments were ligated together and inserted into the pCFJ151 MosSCI targeting vector. This construct was then integrated into the *ttTi5605* site of the EG6699 strain using the MosSCI integration techniques (188).

RNAi-mediated transcript knockdown

RNAi was performed by feeding animals HT115 bacteria transformed either with plasmids expressing dsRNA targeting the specific gene of interest, or carrying the empty L4440 RNAi vector for RNAi controls. Synchronized L1 larvae were transferred to the induced RNAi plates (NGM+1mM IPTG+1mM Ampicillin) pre-seeded with the bacteria expressing the desired dsRNA and grown until the adult stage for one generation (~65 hrs). The feeding protocol was continued for one more generation to completely knockdown the targeted transcripts in both germline and somatic cells. The F1 adult offspring were then analyzed as described.

Immunofluorescence

Adult gonads were dissected on poly-Lysine coated slides, glass coverslips were placed on the specimens, and then the slides were placed on a dry ice block for 10-20 mins. The Freeze-crack method was performed to separate the coverslip from the slide. The slides were placed in ice cold methanol and acetone for 10 min sequentially, then the slides were washed in PBST for 30-60 min at room temperature. 15 ul of a primary antibody was directly put on to the specimen on the slide and incubated in a humidity chamber overnight at 4°C. The next day, slides were washed in PBST three times for 10 min each. 15ul of a secondary antibody was directly put on to the specimen on the slide and incubated in a humidity chamber for 2-3hrs at room temperature, then washed in the PBST four times for 10 min each. At the third wash, a DAPI was added in the PBST solution. The specimens were mounted in the ProLong Gold anti-fade reagent (Life technologies, P36934) and observed under a fluorescence microscope (Leica DMRXA; Hamamatsu Photonics, Hamamatsu, Japan) with Simple PCI software (Hamamatsu

Photonics). Primary antibodies used were: anti-H3K4me2 (CMA303) 1:500, Millipore], anti-H3K4me3 (CMA304) 1:500, Millipore], anti-H3K36me3 [(CMA333), 1:500 (236)], anti-GFP (1:1000, Novus NB600-308), anti-H3K27me3 (1:500, EMD Millipore 07-449)and anti-FLAG (M2, 1:1000, Sigma F1804). Secondary antibodies used were Alexa Fluor 488-conjugated donkey anti-mouse (1:500, Invitrogen R37114) and Alexa Fluor 594-conjugated goat anti-rabbit (1:500, Invitrogen R37117).

Protein isolation and Western blot analysis

Adults worms were resuspended in 4X volume of RIPA buffer (Thermo Scientific, #89901) and frozen and thawed three times in liquid nitrogen. Samples were incubated at 4°C on a rotator for 30 min and sonicated in a Bioruptor sonicator (Diagenode Inc., Denville, NJ, USA) at the high setting for 15 min to fragment chromatin and worm cuticle. The final lysates were centrifuged at 13,000g for 10 min at 4°C, and the supernatants were collected. The protein concentration was determined using the Bradford reagent (Biorad, #500-0006). Supernatants were mixed with an equal volume of 2X SDS-PAGE sample buffer and denatured for 5 min at 95°C. An equal protein amount was loaded in each lane and run on a 4-20% precast SDS-PAGE gel (Biorad, #456-1094) and transferred to PVDF membrane. Transferred proteins were blocked in 5% milk PBST for 1 hour, incubated with a primary antibody overnight, and washed 3 times with 1X PBST for 10 min each. After incubation with a secondary antibody for 2 hours at room temperature, the blot was washed 3 times with 1X PBST for 10 min each. The washed membrane was incubated with the chemiluminescence reagent (Thermo Scientific, #34087) for 5 min, and protein bands were visualized with autoradiography film (Genesee Scientific, #30-100). The primary antibody used is the following: anti-GFP

(1:1000, Novus NB600-308). The following secondary antibody was used: Goat Anti-Rabbit IgG, HRP-conjugated (Millipore, 1:2500, 12-348).

Brood size assay

20 to 30 L3 larvae of each genotype were individually picked and transferred to 6 cm plates pre-seeded with OP50, grown to adult stage, allowed to lay eggs for 24hrs, and transferred to new plates daily until they stopped laying eggs. The number of viable progeny from the plates were counted and recorded for to determine the brood size of each animal.

Figures

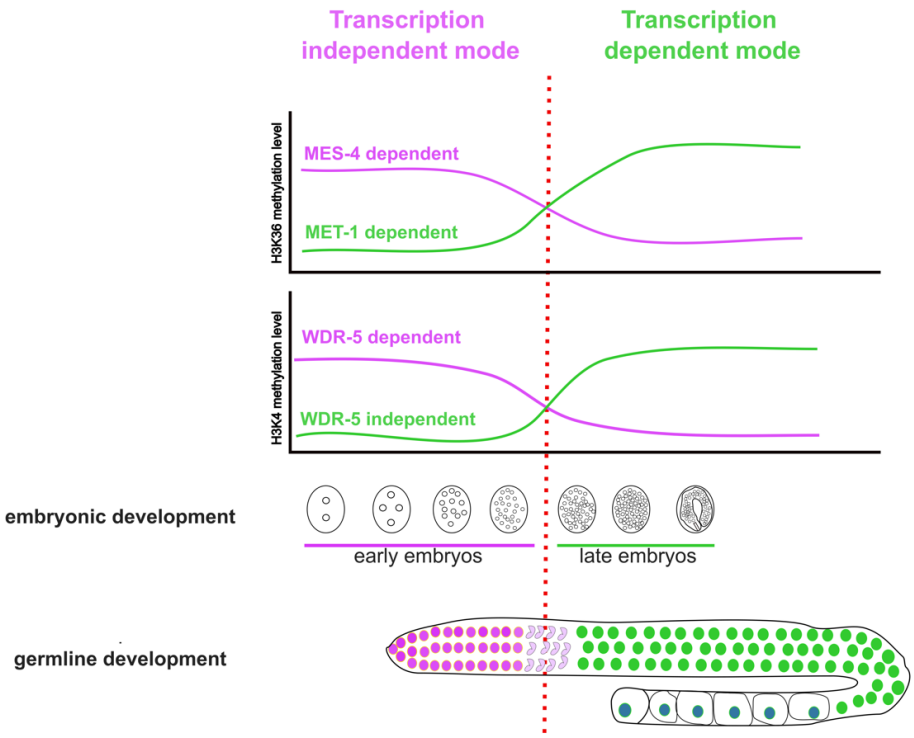


Figure1. Summary of the methylation dynamics of H3K4 and H3K36 during embryonic and an adult germline development

A relative abundance of H3K4 (Bottom) and H3K36 (Top) methylation mediated by different methyltransferase complexes at different stages of either an embryonic or an adult germline development are plotted. Note that both modifications can be added by either transcription independent (Purple) or dependent (Green) mechanisms.

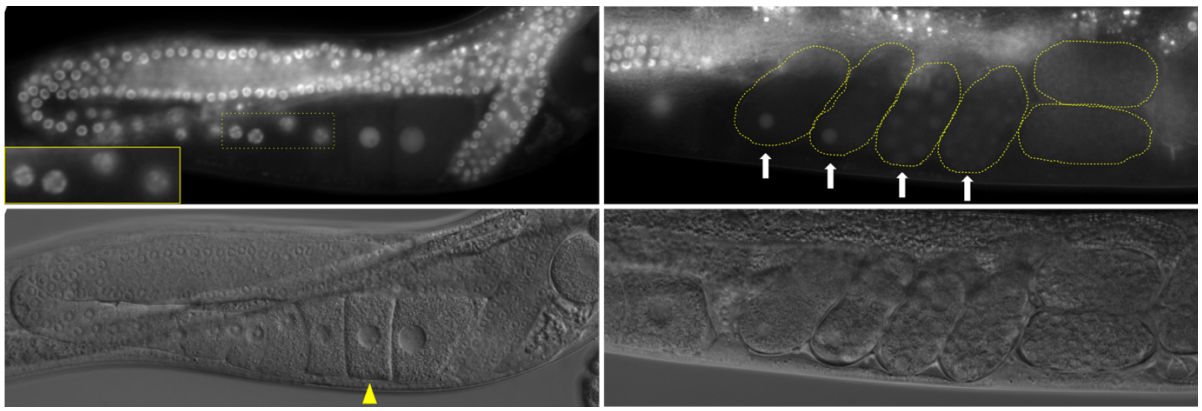


Figure 2. Live WDR-5:GFP expression (ckSi35)
The expression of the GFP tagged WDR-5 in live adult gonads (Top left) and embryos (Top right, yellow dotted circles) are shown with their respective DIC images below. The yellow arrow head indicates an oocyte where WDR-5 protein starts to dissociate from the chromatin and becomes nucleoplasmic. White arrows are indicating the embryos with detectable GFP expression. Note that the intensity of the GFP expression and the age of the embryo is negatively correlated. This pattern is a typical of maternally expressed proteins. See the material and method for details of ckSi35.

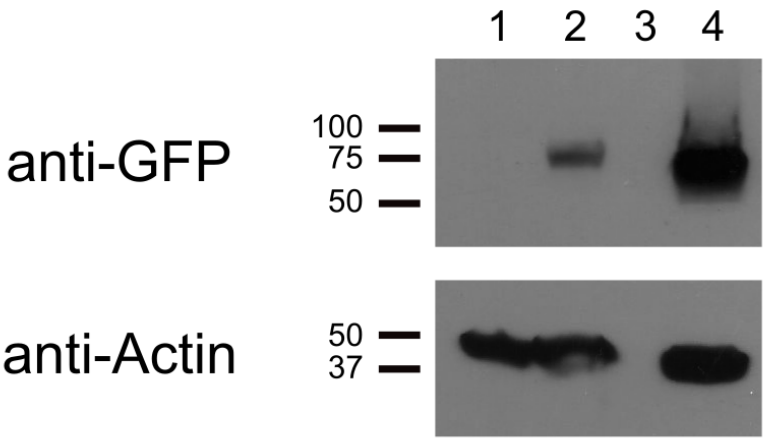


Figure 3. Western blot for GFP confirms the expression of WDR-5.1
Equal amounts of lysates from N2(Lane1), KW1087(Lane2) and KW1032(Lane4) were analyzed by western blots probed with antibodies against GFP and Actin. A lysate from a N2 strain shows no detectable band while lysates from the KW1087 and the KW1032 show very similar sized bands. The expected size of the GFP tagged WDR-5.1 from KW1087 and KW1032 are 68.95kDa and 67.96kDa respectively. Note that the KW1087 strain contains an additional 1X FLAG sequence (~1 kDa) that is not present in the KW1032.

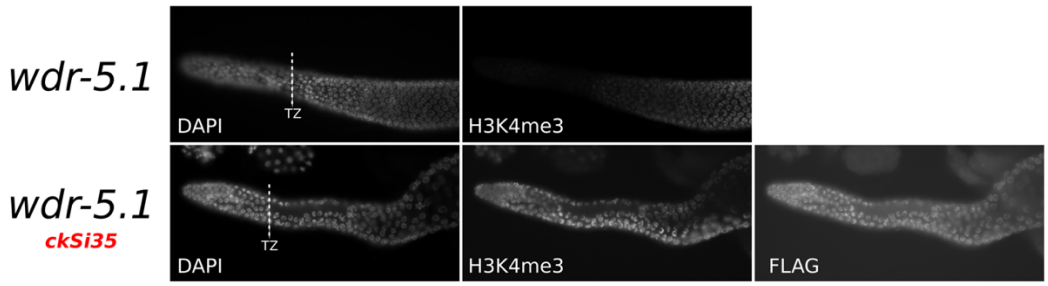


Figure 4. The expression of FLAG::GFP::WDR-5.1 is sufficient to restore the loss of H3K4me3 from the GSCs in the *wdr-5.1* mutant.
The distribution of H3K4me3 in germ cells was compared to a DAPI staining from the strains indicated. The strikingly reduced H3K4me3 distribution in the GSCs of the *wdr-5.1* mutant is restored back to the wildtype level in the *wdr-5.1* mutant expressing the FLAG::GFP::WDR-5.1 proteins from the ckSi35 transgene (See the material and method for details of ckSi35).

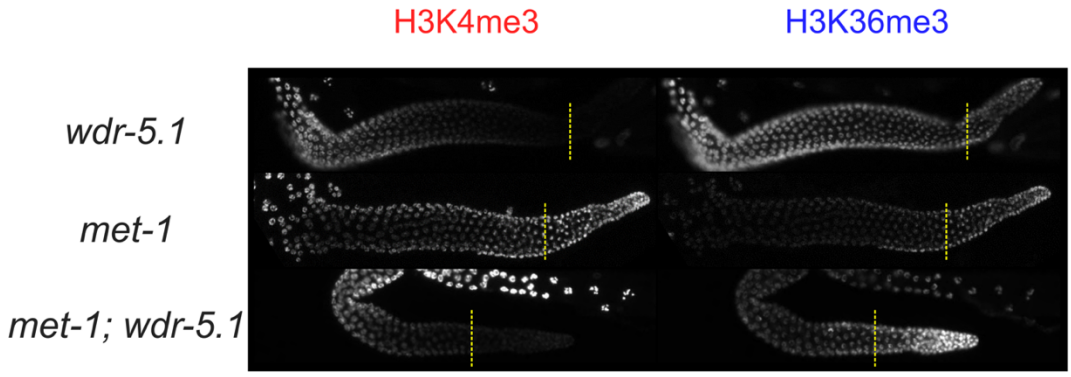
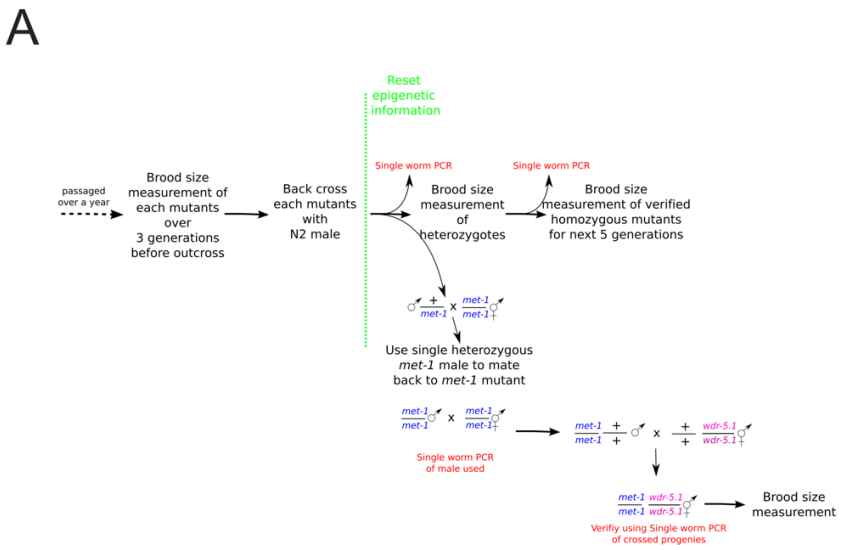


Figure 5. An increase in H3K36me3 is observed in the GSCs of *met-1;wdr-5.1* double mutants.
Dissected adult gonads from the indicated mutants were fixed and stained with antibodies specific for either H3K4me3 or H3K36me3. Note that there is a marked increase in H3K36me3 where H3K4me3 is reduced in the GSCs of the *met-1;wdr-5.1* double mutant. Yellow dotted lines are marking a relative location of the transition zone (TZ), which marks the border between the GSCs and entry into meiosis.



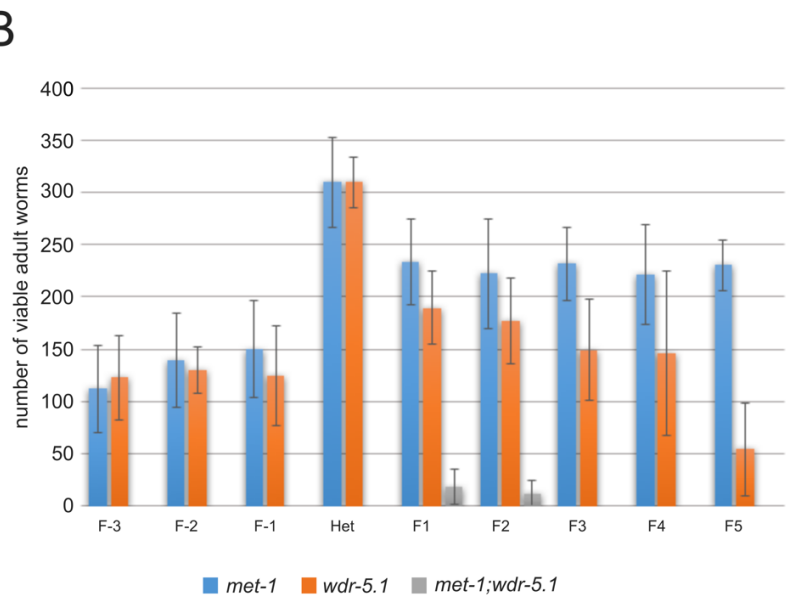


Figure 6. Synergistic defect in the fertility of the *met-1;wdr-5.1* double mutant
A) The schematic of the crosses used to generate the outcrossed single mutants and the double mutants used for brood size assay is shown. B) The brood sizes of each mutant before and after the outcross are shown in comparison. Note that each newly outcrossed single mutants show an increased brood size compared to each single mutant that has been maintained as homozygous for multiple generations.

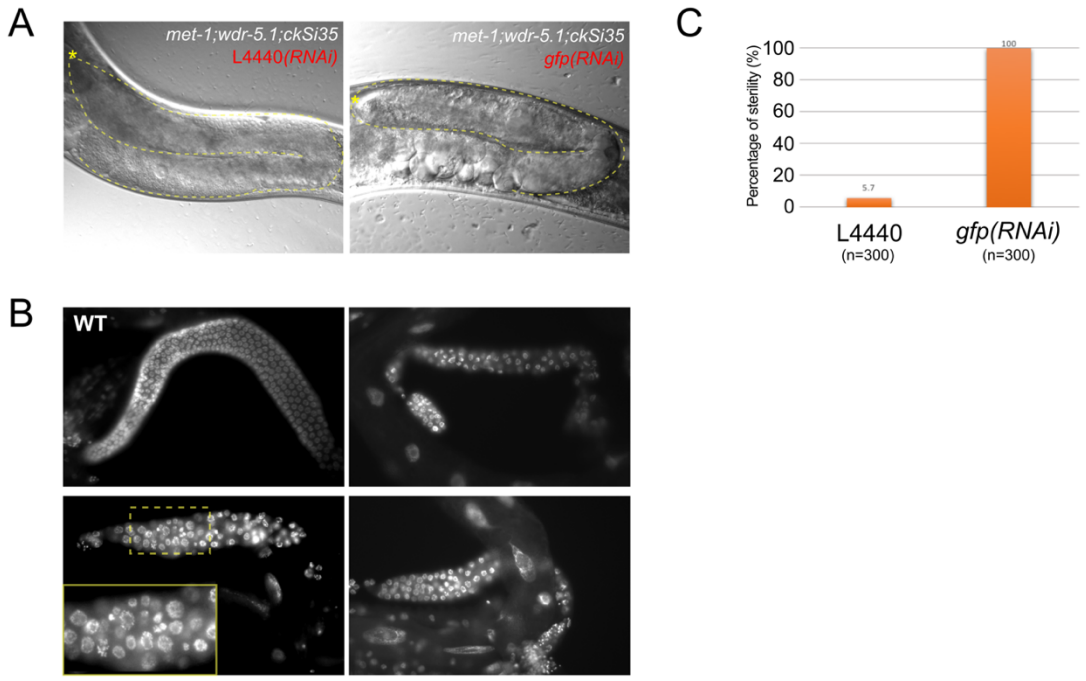


Figure 7. The loss of both *wdr-5.1* and *met-1* causes sterility due to various defects during germline development

A) DIC images of an adult gonad (yellow dotted line) after the L4440 control RNAi and the *gfp(RNAi)* are shown. B) A DAPI staining of a N2 adult gonad is shown on the top left. The other three images are DAPI stained adult gonads of the sterile animals from the *gfp(RNAi)* samples. The most consistent defects observed are a frequent appearance of the irregularly shaped germ cell nuclei and the *mog* phenotype (bottom right) as well as overall decrease in total number of germ cell nuclei in the gonad. C) Percentages of sterile adult animals after RNAi from L1 stage are plotted. Note that the L4440 control RNAi shows 5.7% sterility, which is similar to the sterility frequency observed from the *met-1* mutant alone. N = number of animals quantified. * indicates the approximated position of the DTC located at the distal end of gonad.

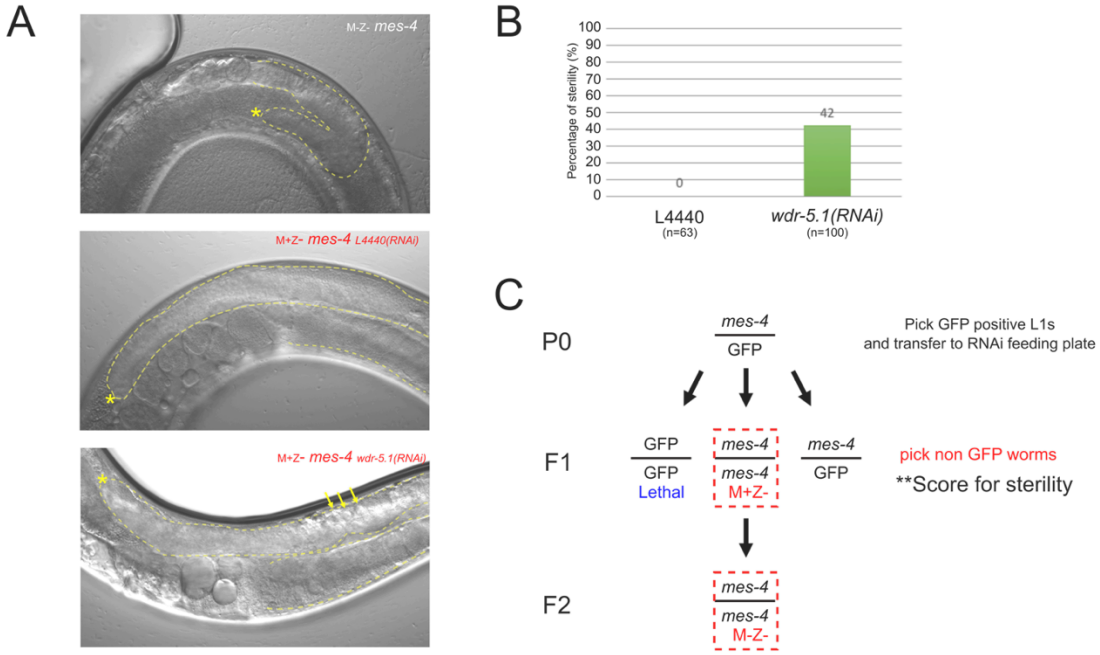


Figure 8. The *wdr-5.1(RNAi)* causes a precocious appearance of the maternal effect sterile (*mes*) phenotype of the *mes-4* mutant (M+Z-)

A) DIC images of an adult gonad (yellow dotted line) for the indicated genotypes are shown. Yellow arrow is pointing at some abnormal germ cell nuclei B) The percentages of the F1 sterile m+z- *mes-4* after RNAi experiment are plotted. C) The schematic of an RNAi experiment is shown. * indicates the approximated position of the DTC located at the distal end of the gonad.

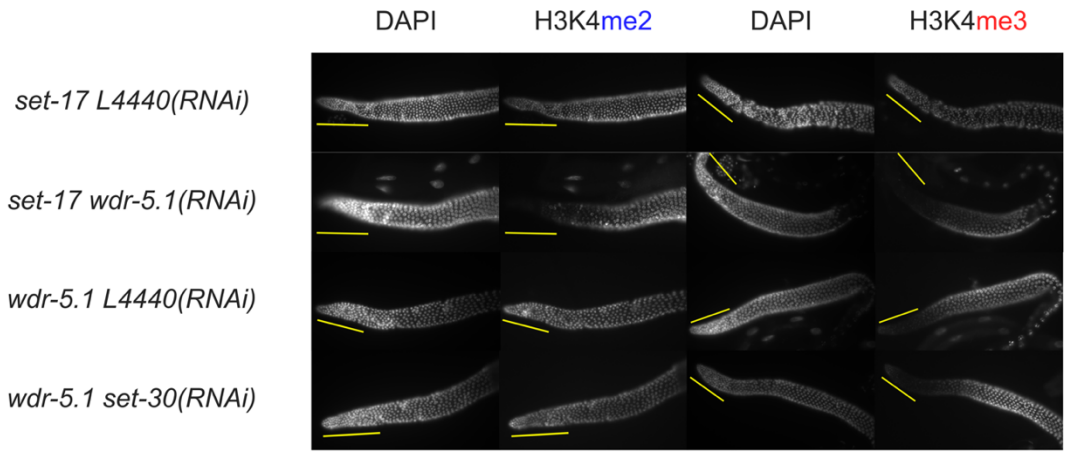


Figure 9. Nether *set-17* nor *set-30* is the major transcription dependent H3K4 methyltransferase in the adult gonad of *C. elegans*
Dissected adult gonads from the indicated mutants/(RNAi) were fixed and stained with the antibody specific for H3K4me2 and H3K4me3. The yellow bars indicate the approximated length of the GSCs based on the DAPI stained nuclei of each gonad.

CHAPTER 4: The significance of my studies and future directions

The significance of SIG-7 studies

We provided the first genome wide *in vivo* description of the role of a nuclear multidomain cyclophilin PPIase, SIG-7, in coordinating splicing and Pol II elongation through its interaction with nascent RNA and CTD of Pol II via its RRM and PPI domains, respectively. I have summarized a potential model of SIG-7's function based on combined data from ours and other model organisms (Figure 1). Such genome wide studies in other organisms have been difficult, due to challenges in generating stable transgenic strains that can replace the endogenous function of SIG-7 homologs. However, our data does not directly reveal the exact mechanism of how SIG-7 carries out its proposed function. Using the CRISPR-Cas9 genome editing system, one can possibly generate mutations in each domain that will inactivate their predicted function, to examine the roles of each independently. It will provide more mechanistic insights into how this class of cyclophilin PPIase function differently than parvulin type PPIase such as Ess1.

It might also be useful to identify SIG-7 interacting partners using immunoprecipitation (IP) followed by Mass-spec identification, as this could provide valuable insights into its direct effects. In initial IP/MassSpec experiments, I have identified some peptides corresponding to interesting proteins such as MOG-2, HRP-2, RNP-5 and RNP-4, which are all involved in some aspects of splicing processes (Table S2). However, these were minor peptides and there were many other peptides observed that suggested the IP was not sufficiently pure. Repeating the IP procedure from nuclear extracts instead of whole embryos may improve this assay and reduce background noise. With our epitope-tagged transgenic line that rescues the deletion phenotype, it is now

possible to use ChIP-seq to capture the location of this enzyme along the gene, and see whether it will localize at the 5' end, gene body or 3' end of the gene. Our initial attempt did not show a consistent result. Only one of three independent experiments yielded sufficient reads with any enrichment pattern, which showed enrichment in exons (data not shown). This kind of information will help us determine which step of the transcription cycle requires the activity of SIG-7.

Another interesting area to study further is the interaction between Pol II CTD and SIG-7. The interaction between Ess1 and CTD is known to be mediated by the WW domain that specifically recognizes the phosphorylated serine. Although I and others have shown that SIG-7 and its homolog can directly interact with Pol II, whether this interaction is phosphorylation dependent or not has not been tested. Therefore, biochemical pull down assays using CTD mimicking peptides, or using endogenous Pol II CTD expressed in bacteria, will be useful to determine if phosphorylation status will affect its interaction with SIG-7. In the similar experimental setting, it will be also informative if SIG-7 and Ess-1 will compete for the binding when only peptide of one repeat length is available compare to peptide with many repeats. If they do compete, then testing whether this competitive binding is phosphorylation dependent or not will also be an important information. All these potential future experiments can help to understand its molecular mechanism of how SIG-7 and its homolog can coordinate the process of splicing and Pol II elongation.

The significance of WDR-5 project

Although many findings from multiple organisms suggest that histone modifications play a critical role in maintaining pluripotency, and also transmitting

epigenetic memory after each cell division during development, we still don't have a clear picture of how they are all connected and influence each other. Although yeast has served as a simple model organism to initially characterize the function and the composition of many chromatin modifying enzyme complexes, it has its own limitation for studying the function of these complexes in the developmental context of multicellular organisms. Likewise, using mammalian models is probably the best in terms of its application to human, the complexity of such model organism often makes it hard to interpret the findings in precise manner due to an increased number of complexes carrying out the same modification. For example, there are multiple WDR-5 containing histone methyltransferase complexes for H3K4 in mammals. Therefore, it can be challenging to distinguish which WDR-5 containing complex is responsible for the phenotypes they observe by knocking down WDR-5 in mammalian systems. In *C. elegans*, we showed that H3K4 tri-methylation can be established in two distinct modes: transcription-dependent and transcription-independent. With the experimental strategy we designed, we can isolate a specific cell type, in this case germline stem cells, where the majority of H3K4 tri-methylation is WDR-5 dependent. The cell type specific analysis of H3K4me3 that is mediated by WDR-5, versus the H3K4me3 that does not depend on WDR-5, will allow identification of the specific targets and its importance for germline stem cell function. Besides this technical advantage, I also provided evidence for a novel antagonistic relationship between H3K4 methylation and H3K36 methylation that has not been described elsewhere. Further examination of this antagonism, and its role in maintaining epigenetic memory and germline function, will be important for

understanding how the “immortal” germline retains its unique capacity for retaining pluripotency despite its differentiation into gametes at each generation.

Figure

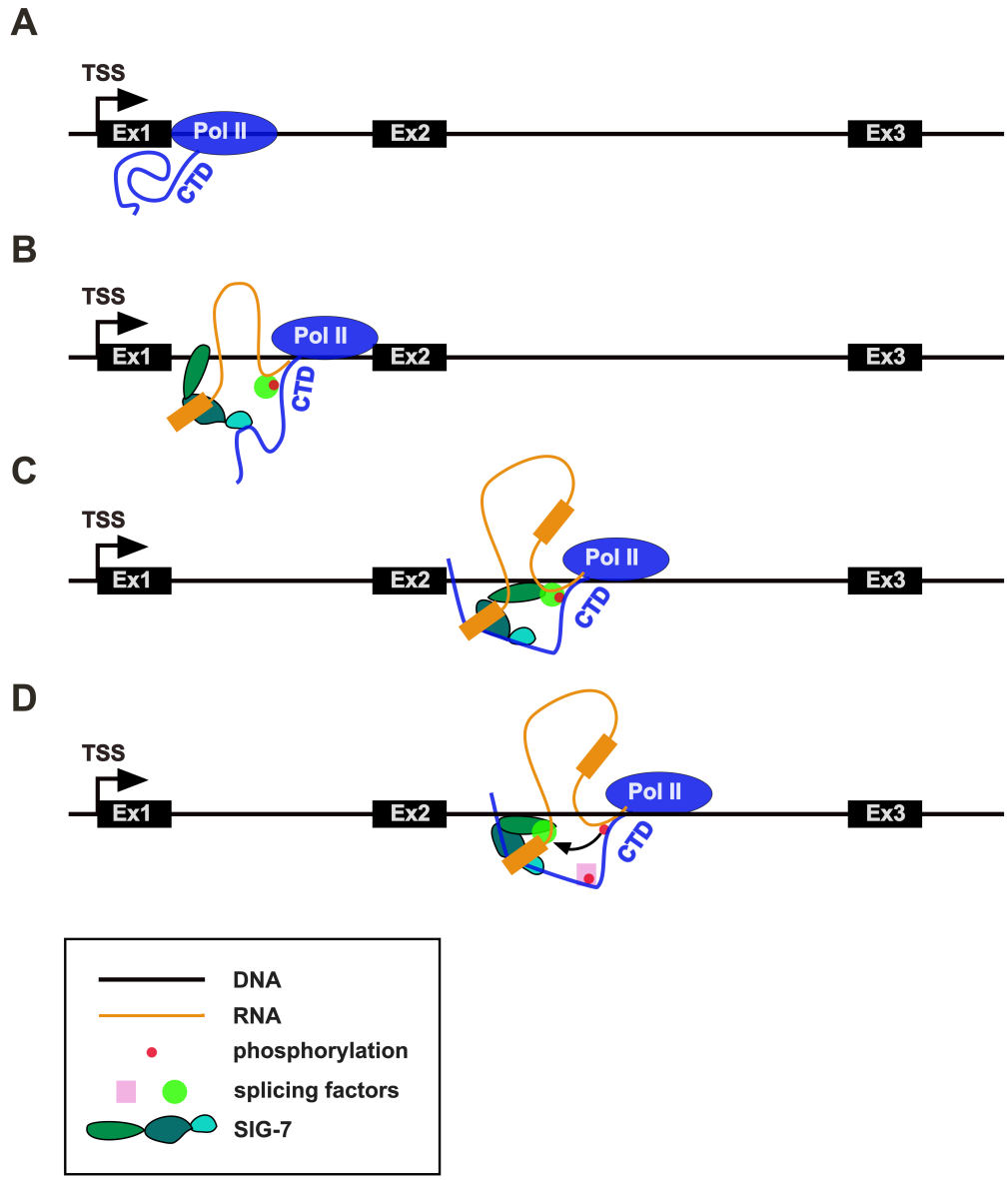


Figure 1. A potential model of SIG-7's function
This is an illustration of how sig-7 may function. A) Before Pol II is engaged into productive elongation, the CTD may conform into a more compact state. B) Upon phosphorylation at numerous heptapeptide repeats, the CTD may conform into more extended structure. This extended structure of the CTD may provide an opportunity for SIG-7 to interact with Pol II using its PPIase domain as well as binding of splicing factors to phosphorylated residues within the repeats. C) Once SIG-7 binds to the CTD, the activity of PPIase domain of SIG-7 may induce structural change of the CTD via its cis/trans isomerization reaction. D) The binding of the pre-mRNA by the RRM domain of SIG-7 may allosterically regulate the PPIase activity, possibly slowing down its PPIase

activity. This may provide an opportunity for SIG-7 to interact with splicing factors bound to the Pol II CTD using its highly charged C-terminal domain and thus assist in linking splicing factors on the Pol II CTD with the emerging pre-mRNAs bound to SIG-7.

References

1. Decker KB, Hinton DM. Transcription Regulation at the Core: Similarities Among Bacterial, Archaeal, and Eukaryotic RNA Polymerases. Annual Review of Microbiology. 2013;67(1):113-39.
2. Roeder RG, Rutter WJ. Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms. Nature. 1969;224(5216):234-7.
3. Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. Nature structural & molecular biology. 2004;11(5):394-403.
4. Egloff S, Dienstbier M, Murphy S. Updating the RNA polymerase CTD code: adding gene-specific layers. Trends in genetics : TIG. 2012;28(7):333-41.
5. Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A, et al. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Molecular cell. 2012;45(2):158-70.
6. Palancade Bt, Bensaude O. Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. European Journal of Biochemistry. 2003;270(19):3859-70.
7. Buratowski S. Progression through the RNA polymerase II CTD cycle. Molecular cell. 2009;36(4):541-6.
8. Heidemann M, Hintermair C, Voss K, Eick D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochimica et biophysica acta. 2013;1829(1):55-62.
9. Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, et al. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science. 2012;336(6089):1723-5.
10. Cho E-J, Takagi T, Moore CR, Buratowski S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes & development. 1997;11(24):3319-26.
11. Ranuncolo SM, Ghosh S, Hanover JA, Hart GW, Lewis BA. Evidence of the Involvement of O-GlcNAc-modified Human RNA Polymerase II CTD in Transcription in Vitro and in Vivo. The Journal of biological chemistry. 2012;287(28):23549-61.
12. Comer FI, Hart GW. Reciprocity between O-GlcNAc and O-Phosphate on the Carboxyl Terminal Domain of RNA Polymerase II. Biochemistry. 2001;40(26):7845-52.
13. Drogat J, Hermand D. Gene-specific requirement of RNA polymerase II CTD phosphorylation. Molecular microbiology. 2012;84(6):995-1004.
14. Tietjen JR, Zhang DW, Rodriguez-Molina JB, White BE, Akhtar MS, Heidemann M, et al. Chemical-genomic dissection of the CTD code. Nature structural & molecular biology. 2010;17(9):1154-61.
15. Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, et al. Gene-specific RNA polymerase II phosphorylation and the CTD code. Nature structural & molecular biology. 2010;17(10):1279-86.
16. Buratowski S. The CTD code. Nature structural & molecular biology. 2003;10(9):679-80.
17. Eick D, Geyer M. The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. Chemical reviews. 2013;113(11):8456-90.

18. Schmid FX. Prolyl isomerase: Enzymatic Catalysis of Slow Protein=Folding Reactions. Annu Rev Biophys Biomol Struct. 1993;22:123-43.
19. Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. Genes & development. 1996;10:2657-83.
20. Conaway RC, Conaway JW. General Initiation Factors for RNA Polymerase II. Annual review of biochemistry. 1993;62(1):161-90.
21. Thomas MC, Chiang C-M. The General Transcription Machinery and General Cofactors. Critical Reviews in Biochemistry and Molecular Biology. 2006;41:105-78.
22. Burley SK, Roeder RG. Biochemistry and Structural Biology of Transcription Factor IID (TFIID). Annual review of biochemistry. 1996;65(1):769-99.
23. Buratowski S, Hahn S, Guarente L, Sharp PA. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell. 1989;56(4):549-61.
24. Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, Azubel M, et al. Architecture of an RNA polymerase II transcription pre-initiation complex. Science. 2013;342(6159):1238724.
25. Guzder SN, Sung P, Bailly V, Prakash L, Prakash S. RAD25 is a DMA helicase required for DNA repair and RNA polymerase II transcription. Nature. 1994;369(6481):578-81.
26. Prelich G. RNA Polymerase II Carboxy-Terminal Domain Kinases: Emerging Clues to Their Function. Eukaryotic Cell. 2002;1(2):153-62.
27. McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, et al. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes & development. 1997;11(24):3306-18.
28. Mandal SS, Chu C, Wada T, Handa H, Shatkin AJ, Reinberg D. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(20):7572-7.
29. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007;130.
30. Ng HH, Robert F, Young RA, Struhl K. Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. Molecular cell. 2003;11:709-19.
31. Rougvie AE, Lis JT. The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell. 1988;54(6):795-804.
32. Gilmour DS, Lis JT. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. Molecular and cellular biology. 1986;6(11):3984-9.
33. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, et al. A high-resolution map of active promoters in the human genome. Nature. 2005;436(7052):876-80.
34. Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, et al. RNA polymerase is poised for activation across the genome. Nature genetics. 2007;39(12):1507-11.

35. Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. Genes & development. 2011;25(7):742-54.
36. Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. Nature reviews Molecular cell biology. 2015;16(3):167-77.
37. Kwak H, Lis JT. Control of transcriptional elongation. Annual review of genetics. 2013;47:483-508.
38. Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental biology. 2010;339(2):225-9.
39. Ohler U, Liao G-c, Niemann H, Rubin GM. Computational analysis of core promoters in the Drosophila genome. Genome biology. 2002;3(12):1-12.
40. Hendrix DA, Hong J-W, Zeitlinger J, Rokhsar DS, Levine MS. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proceedings of the National Academy of Sciences. 2008;105(22):7762-7.
41. Wang YV, Tang H, Gilmour DS. Identification In Vivo of Different Rate-Limiting Steps Associated with Transcriptional Activators in the Presence and Absence of a GAGA Element. Molecular and cellular biology. 2005;25(9):3543-52.
42. Smale ST, Baltimore D. The "initiator" as a transcription control element. Cell. 1989;57(1):103-13.
43. Burke TW, Kadonaga JT. The downstream core promoter element, DPE, is conserved fromDrosophila to humans and is recognized by TAFII60 of Drosophila. Genes & development. 1997;11(22):3020-31.
44. van Steensel B, Delrow J, Bussemaker HJ. Genomewide analysis of Drosophila GAGA factor target genes reveals context-dependent DNA binding. Proceedings of the National Academy of Sciences. 2003;100(5):2580-5.
45. Shimojima T, Okada M, Nakayama T, Ueda H, Okawa K, Iwamatsu A, et al. Drosophila FACT contributes to Hox gene expression through physical and functional interactions with GAGA factor. Genes & development. 2003;17(13):1605-16.
46. Weber CM, Ramachandran S, Henikoff S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Molecular cell. 2014;53(5):819-30.
47. Li J, Gilmour DS. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. The EMBO journal. 2013;32(13):1829-41.
48. Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, et al. NELF and GAGA Factor Are Linked to Promoter-Proximal Pausing at Many Genes in Drosophila. Molecular and cellular biology. 2008;28(10):3290-300.
49. Wu C-H, Yamaguchi Y, Benjamin LR, Horvat-Gordon M, Washinsky J, Enerly E, et al. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. Genes & development. 2003;17(11):1402-14.
50. Guo J, Price DH. RNA polymerase II transcription elongation control. Chemical reviews. 2013;113(11):8583-603.
51. Chen Y, Yamaguchi Y, Tsugeno Y, Yamamoto J, Yamada T, Nakamura M, et al. DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. Genes & development. 2009;23(23):2765-77.

52. Chen Fei X, Woodfin Ashley R, Gardini A, Rickels Ryan A, Marshall Stacy A, Smith Edwin R, et al. PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. Cell. 2015;162(5):1003-15.
53. Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, et al. c-Myc Regulates Transcriptional Pause Release. Cell. 2010;141(3):432-45.
54. Ni Z, Saunders A, Fuda NJ, Yao J, Suarez J-R, Webb WW, et al. P-TEFb Is Critical for the Maturation of RNA Polymerase II into Productive Elongation In Vivo. Molecular and cellular biology. 2008;28(3):1161-70.
55. Price DH. P-TEFb, a Cyclin-Dependent Kinase Controlling Elongation by RNA Polymerase II. Molecular and cellular biology. 2000;20(8):2629-34.
56. Eberhardy SR, Farnham PJ. Myc Recruits P-TEFb to Mediate the Final Step in the Transcriptional Activation of the cad Promoter. Journal of Biological Chemistry. 2002;277(42):40156-62.
57. Kanazawa S, Soucek L, Evan G, Okamoto T, Peterlin BM. c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. Oncogene. 2003;22(36):5707-11.
58. Yang Z, Yik JHN, Chen R, He N, Jang MK, Ozato K, et al. Recruitment of P-TEFb for Stimulation of Transcriptional Elongation by the Bromodomain Protein Brd4. Molecular cell. 2005;19(4):535-45.
59. Takahashi H, Parmely Tari J, Sato S, Tomomori-Sato C, Banks Charles AS, Kong Stephanie E, et al. Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors. Cell. 2011;146(1):92-104.
60. Kim JB, Sharp PA. Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase. Journal of Biological Chemistry. 2001;276(15):12317-23.
61. Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Molecular cell. 2006;21(2):227-37.
62. Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. Dynamics of Human Immunodeficiency Virus Transcription: P-TEFb Phosphorylates RD and Dissociates Negative Effectors from the Transactivation Response Element. Molecular and cellular biology. 2004;24(2):787-95.
63. Chen H, Contreras X, Yamaguchi Y, Handa H, Peterlin BM, Guo S. Repression of RNA Polymerase II Elongation <italic>In Vivo</italic> Is Critically Dependent on the C-Terminus of Spt5. PloS one. 2009;4(9):e6918.
64. Ramanathan Y, Rajpara SM, Reza SM, Lees E, Shuman S, Mathews MB, et al. Three RNA Polymerase II Carboxyl-terminal Domain Kinases Display Distinct Substrate Preferences. Journal of Biological Chemistry. 2001;276(14):10913-20.
65. Li B, Howe L, Anderson S, Yates JR, Workman JL. The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. Journal of Biological Chemistry. 2003;278(11):8897-903.
66. Schwabish MA, Struhl K. Evidence for Eviction and Rapid Deposition of Histones upon Transcriptional Elongation by RNA Polymerase II. Molecular and cellular biology. 2004;24(23):10111-7.

67. Orphanides G, Wu W-H, Lane WS, Hampsey M, Reinberg D. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. Nature. 1999;400(6741):284-8.
68. Orphanides G, LeRoy G, Chang C-H, Luse DS, Reinberg D. FACT, a Factor that Facilitates Transcript Elongation through Nucleosomes. Cell. 1998;92(1):105-16.
69. Kristjuhan A, Svejstrup JQ. Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo. The EMBO journal. 2004;23(21):4243-52.
70. Carey M, Li B, Workman JL. RSC Exploits Histone Acetylation to Abrogate the Nucleosomal Block to RNA Polymerase II Elongation. Molecular cell. 2006;24(3):481-7.
71. Spain Marla M, Ansari Suraiya A, Pathak R, Palumbo Michael J, Morse Randall H, Govind Chhabi K. The RSC Complex Localizes to Coding Sequences to Regulate Pol II and Histone Occupancy. Molecular cell. 2014;56(5):653-66.
72. Park D, Shivram H, Iyer VR. Chd1 co-localizes with early transcription elongation factors independently of H3K36 methylation and releases stalled RNA polymerase II at introns. Epigenetics & Chromatin. 2014;7(1):1-11.
73. Simic R, Lindstrom DL, Tran HG, Roinick KL, Costa PJ, Johnson AD, et al. Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. The EMBO journal. 2003;22(8):1846-56.
74. Bentley DL. Coupling mRNA processing with transcription in time and space. Nature reviews Genetics. 2014;15(3):163-75.
75. Matera AG, Wang Z. A day in the life of the spliceosome. Nature reviews Molecular cell biology. 2014;15(2):108-21.
76. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009;136(4):701-18.
77. Will CL, Lührmann R. Spliceosome Structure and Function. Cold Spring Harbor Perspectives in Biology. 2011;3(7).
78. Osheim YN, O.L. Miller, Jr., Beyer AL. RNP particles at splice junction sequences on Drosophila chorion transcripts. Cell. 1985;43(1):143-51.
79. Görnemann J, Kotovic KM, Hujer K, Neugebauer KM. Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. Molecular cell. 2005;19(1):53-63.
80. Lacadie SA, Rosbash M. Cotranscriptional Spliceosome Assembly Dynamics and the Role of U1 snRNA:5′ss Base Pairing in Yeast. Molecular cell. 2005;19(1):65-75.
81. Listerman I, Sapra AK, Neugebauer KM. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nature structural & molecular biology. 2006;13(9):815-22.
82. Tardiff DF, Lacadie SA, Rosbash M. A Genome-Wide Analysis Indicates that Yeast Pre-mRNA Splicing Is Predominantly Posttranscriptional. Molecular cell. 2006;24(6):917-29.
83. Singh J, Padgett RA. Rates of in situ transcription and splicing in large human genes. Nature structural & molecular biology. 2009;16(11):1128-33.
84. Khodor YL, Rodriguez J, Abruzzi KC, Tang C-HA, Marr MT, Rosbash M. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes & development. 2011;25(23):2502-12.

85. Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome research. 2012;22(9):1616-25.
86. Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, et al. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nature structural & molecular biology. 2011;18(12):1435-40.
87. Nojima T, Gomes T, Grosso Ana Rita F, Kimura H, Dye Michael J, Dhir S, et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell. 2015;161(3):526-40.
88. Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J, et al. Nucleosome positioning as a determinant of exon recognition. Nature structural & molecular biology. 2009;16(9):996-1001.
89. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome research. 2009;19(10):1732-41.
90. Tolstorukov Michael Y, Goldman Joseph A, Gilbert C, Ogryzko V, Kingston Robert E, Park Peter J. Histone Variant H2A.Bbd Is Associated with Active Transcription and mRNA Processing in Human Cells. Molecular cell. 2012;47(4):596-607.
91. Gunderson FQ, Merkhofer EC, Johnson TL. Dynamic histone acetylation is critical for cotranscriptional spliceosome assembly and spliceosomal rearrangements. Proceedings of the National Academy of Sciences. 2011;108(5):2004-9.
92. Schor IE, Rascovan N, Pelisch F, Alló M, Kornblihtt AR. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. Proceedings of the National Academy of Sciences. 2009;106(11):4325-30.
93. Sims RJ, III, Millhouse S, Chen C-F, Lewis BA, Erdjument-Bromage H, Tempst P, et al. Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. Molecular cell. 2007;28(4):665-76.
94. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of Alternative Splicing by Histone Modifications. Science. 2010;327(5968):996-1000.
95. de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, et al. Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nature structural & molecular biology. 2011;18(9):977-83.
96. Koga M, Hayashi M, Kaida D. Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. Nucleic acids research. 2015;43(17):8258-67.
97. Xiao T, Hall H, Kizer KO, Shibata Y, Hall MC, Borchers CH, et al. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes & development. 2003;17(5):654-63.
98. Piacentini L, Fanti L, Berloco M, Perrini B, Pimpinelli S. Heterochromatin protein 1 (HP1) is associated with induced gene expression in Drosophila euchromatin. The Journal of Cell Biology. 2003;161(4):707-14.

99. Font-Burgada J, Rossell D, Auer H, Azorín F. Drosophila HP1c isoform interacts with the zinc-finger proteins WOC and Relative-of-WOC to regulate gene expression. Genes & development. 2008;22(21):3007-23.
100. Kwon SH, Florens L, Swanson SK, Washburn MP, Abmayr SM, Workman JL. Heterochromatin protein 1 (HP1) connects the FACT histone chaperone complex to the phosphorylated CTD of RNA polymerase II. Genes & development. 2010;24(19):2133-45.
101. Smallwood A, Hon GC, Jin F, Henry RE, Espinosa JM, Ren B. CBX3 regulates efficient RNA processing genome-wide. Genome research. 2012;22(8):1426-36.
102. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature. 1997;385(6614):357-61.
103. Misteli T, Spector DL. RNA Polymerase II Targets Pre-mRNA Splicing Factors to Transcription Sites In Vivo. Molecular cell. 1999;3(6):697-705.
104. Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. A structural perspective of CTD function. Genes & development. 2005;19(12):1401-15.
105. Meredith GD, Chang W-H, Li Y, Bushnell DA, Darst SA, Kornberg RD. The C-terminal Domain Revealed in the Structure of RNA Polymerase II. Journal of molecular biology. 1996;258(3):413-9.
106. Laybourn PJ, Dahmus ME. Transcription-dependent structural changes in the C-terminal domain of mammalian RNA polymerase subunit IIa/o. Journal of Biological Chemistry. 1989;264(12):6693-8.
107. Zhang J, Corden JL. Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. Journal of Biological Chemistry. 1991;266(4):2297-302.
108. Morris DP, Greenleaf AL. The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. Journal of Biological Chemistry. 2000;275(51):39935-43.
109. Kang CH, Feng Y, Vikram M, Jeong IS, Lee JR, Bahk JD, et al. Arabidopsis thaliana PRP40s are RNA polymerase II C-terminal domain-associating proteins. Archives of Biochemistry and Biophysics. 2009;484(1):30-8.
110. Kao HY, Siliciano PG. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. Molecular and cellular biology. 1996;16(3):960-7.
111. Abovich N, Rosbash M. Cross-Intron Bridging Interactions in the Yeast Commitment Complex Are Conserved in Mammals. Cell. 1997;89(3):403-12.
112. Allen M, Friedler A, Schon O, Bycroft M. The Structure of an FF Domain from Human HYPA/FBP11. Journal of molecular biology. 2002;323(3):411-6.
113. Emili A, Shales M, McCracken S, Xie W, Tucker PW, Kobayashi R, et al. Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. Rna. 2002;8(9):1102-11.
114. David CJ, Boyne AR, Millhouse SR, Manley JL. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65–Prp19 complex. Genes & development. 2011;25(9):972-83.
115. Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. Genes & development. 2006;20(21):2922-36.

116. Phatnani HP, Jones JC, Greenleaf AL. Expanding the Functional Repertoire of CTD Kinase I and RNA Polymerase II: Novel PhosphoCTD-Associating Proteins in the Yeast Proteome. Biochemistry. 2004;43(50):15702-19.
117. Schiene-Fischer C, Aumüller T, Fischer G. Peptide Bond cis/trans Isomerases: A Biocatalysis Perspective of Conformational Dynamics in Proteins. In: Jackson S, editor. Molecular Chaperones. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 35-67.
118. Göthel FS, Marahiel AM. Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. Cellular and Molecular Life Sciences CMLS. 1999;55(3):423-36.
119. Pemberton TJ, Kay JE. Identification and comparative analysis of the peptidyl-prolyl cis/trans isomerase repertoires of H. sapiens, D. melanogaster, C. elegans, S. cerevisiae and Sz. pombe. Comparative and functional genomics. 2005;6(5-6):277-300.
120. Fischer G, Tradler T, Zarnt T. The mode of action of peptidyl prolyl cis/trans isomerases in vivo: binding vs. catalysis. FEBS letters. 1998;426:17-20.
121. Hanes SD. The Ess1 prolyl isomerase: traffic cop of the RNA polymerase II transcription cycle. Biochimica et biophysica acta. 2014;1839(4):316-33.
122. Fanghänel J, Fischer G. Insights into the catalytic mechanism of peptidyl prolyl cis/trans isomerases. Frontiers in Bioscience. 2004;9:3453-78.
123. Grison A, Mantovani F, Comel A, Agostoni E, Gustincich S, Persichetti F, et al. Ser46 phosphorylation and prolyl-isomerase Pin1-mediated isomerization of p53 are key events in p53-dependent apoptosis induced by mutant huntingtin. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(44):17979-84.
124. Kang CB, Hong Y, Dhe-Paganon S, Yoon HS. FKBP Family Proteins: Immunophilins with Versatile Biological Functions. Neurosignals. 2008;16(4):318-25.
125. Liou YC, Zhou XZ, Lu KP. Prolyl isomerase Pin1 as a molecular switch to determine the fate of phosphoproteins. Trends in biochemical sciences. 2011;36(10):501-14.
126. Ma Z, Atencio D, Barnes C, DeFiglio H, Hanes SD. Multiple roles for the Ess1 prolyl isomerase in the RNA polymerase II transcription cycle. Molecular and cellular biology. 2012;32(17):3594-607.
127. Singh N, Ma Z, Gemmill T, Wu X, Defiglio H, Rossettini A, et al. The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway. Molecular cell. 2009;36(2):255-66.
128. Harrar Y, Bellini C, Faure J-D. FKBPs: at the crossroads of folding and transduction. Trends in Plant Science. 2001;6(9):426-31.
129. Schwer B, Shuman S. Deciphering the RNA Polymerase II CTD Code in Fission Yeast. Molecular cell. 2011;43(2):311-8.
130. Hani J, Schelbert B, Bernhardt A, Domdey H, Fischer G, Wiebauer K, et al. Mutations in a Peptidylprolyl-cis/trans-isomerase Gene Lead to a Defect in 3′-End Formation of a Pre-mRNA inSaccharomyces cerevisiae. Journal of Biological Chemistry. 1999;274(1):108-16.
131. Morris DP, Phatnani HP, Greenleaf AL. Phospho-Carboxyl-Terminal Domain Binding and the Role of a Prolyl Isomerase in Pre-mRNA 3′-End Formation. Journal of Biological Chemistry. 1999;274(44):31583-7.

132. Yaffe MB, Schutkowski M, Shen M, Zhou XZ, Stukenberg PT, Rahfeld J-U, et al. Sequence-Specific and Phosphorylation-Dependent Proline Isomerization: A Potential Mitotic Regulatory Mechanism. Science. 1997;278(5345):1957-60.
133. Myers JK, Morris DP, Greenleaf AL, Oas TG. Phosphorylation of RNA Polymerase II CTD Fragments Results in Tight Binding to the WW Domain from the Yeast Prolyl Isomerase Ess1. Biochemistry. 2001;40(29):8479-86.
134. Wilcox CB, Rossettini A, Hanes SD. Genetic Interactions With C-Terminal Domain (CTD) Kinases and the CTD of RNA Pol II Suggest a Role for ESS1 in Transcription Initiation and Elongation in Saccharomyces cerevisiae. Genetics. 2004;167(1):93-105.
135. Krishnamurthy S, Ghazy MA, Moore C, Hampsey M. Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries. Molecular and cellular biology. 2009;29(11):2925-34.
136. Werner-Allen JW, Lee CJ, Liu P, Nicely NI, Wang S, Greenleaf AL, et al. cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72. The Journal of biological chemistry. 2011;286(7):5717-26.
137. Kops O, Zhou XZ, Lu KP. Pin1 modulates the dephosphorylation of the RNA polymerase II C-terminal domain by yeast Fcp1. FEBS letters. 2002;513(2-3):305-11.
138. Xu Y-X, Hirose Y, Zhou XZ, Lu KP, Manley JL. Pin1 modulates the structure and function of human RNA polymerase II. Genes & development. 2003;17(22):2765-76.
139. Palancade Bt, Marshall NF, Tremeau-Bravard A, Bensaude O, Dahmus ME, Dubois M-F. Dephosphorylation of RNA Polymerase II by CTD-phosphatase FCP1 is Inhibited by Phospho-CTD Associating Proteins. Journal of molecular biology. 2004;335(2):415-24.
140. Philipps D, Celotto AM, Wang QQ, Tarng RS, Graveley BR. Arginine/serine repeats are sufficient to constitute a splicing activation domain. Nucleic acids research. 2003;31(22):6502-8.
141. Long Jennifer C, Caceres Javier F. The SR protein family of splicing factors: master regulators of gene expression. Biochemical Journal. 2009;417(1):15-27.
142. Krzywicka A, Beisson J, Jerka-Dziadosz M, Klotz C. KIN241: a gene involved in cell morphogenesis in *Paramecium tetraurelia* reveals a novel protein family of cyclophilin–RNA interacting proteins (CRIPs) conserved from fission yeast to man. Molecular microbiology. 2001;42(1):257-67.
143. Jerka-Dziadosz M, Garreau de Loubresse N, Beisson J. Development of surface pattern during division in Paramecium. II. Defective spatial control in the mutant kin241. Development. 1992;115(1):319-35.
144. Gullerova M, Barta A, Lorkovic ZJ. Rct1, a Nuclear RNA Recognition Motif-Containing Cyclophilin, Regulates Phosphorylation of the RNA Polymerase II C-Terminal Domain. Molecular and cellular biology. 2007;27(10):3601-11.
145. Skrahina T. The role of *Schizosaccharomyces pombe* cyclophilin *Rct1* in RNA polymerase II transcription. PhD Dissertation, uniwien. 2009.
146. Gullerova M, Barta A, Lorkovic ZJ. AtCyp59 is multidomain cyclophilin from *Arabidopsis thaliana* that interacts with SR proteins and the C-terminal domain of the RNA polymerase II. Rna. 2006;12(4):631-43.

147. Bannikova O, Zywicki M, Marquez Y, Skrahina T, Kalyna M, Barta A. Identification of RNA targets for the nuclear multidomain cyclophilin atCyp59 and their effect on PPIase activity. Nucleic acids research. 2013;41(3):1783-96.
148. Kelly WG, Fire A. Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. Development (Cambridge, England). 1998;125(13):2451-6.
149. Kelly WG, Xu S, Montgomery MK, Fire A. Distinct Requirements for Somatic and Germline Expression of a Generally Expressed Caernorhabditis elegans Gene. Genetics. 1997;146(1):227-38.
150. Kim JK, Gabel HW, Kamath RS, Tewari M, Pasquinelli A, Rual J-F, et al. Functional Genomic Analysis of RNA Interference in C. elegans. Science. 2005;308(5725):1164-7.
151. Kelly WG, Schaner CE, Dernburg AF, Lee M-H, Kim SK, Villeneuve AM, et al. X-chromosome silencing in the germline of C. elegans. Development. 2002;129(2):479-92.
152. Jedrusik MA, Schulze E. Linker histone HIS-24 (H1.1) cytoplasmic retention promotes germ line development and influences histone H3 methylation in Caenorhabditis elegans. Molecular and cellular biology. 2007;27(6):2229-39.
153. Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, Sapetschnig A, et al. Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Molecular cell. 2008;31(1):79-90.
154. Cui M, Kim EB, Han M. Diverse chromatin remodeling genes antagonize the Rb-involved SynMuv pathways in C. elegans. PLoS genetics. 2006;2(5):e74.
155. Wu X, Shi Z, Cui M, Han M, Ruvkun G. Repression of germline RNAi pathways in somatic cells by retinoblastoma pathway chromatin complexes. PLoS genetics. 2012;8(3):e1002542.
156. Voss TC, Hager GL. Dynamic regulation of transcriptional states by chromatin and transcription factors. Nature reviews Genetics. 2014;15(2):69-81.
157. Liu X, Bushnell DA, Kornberg RD. RNA polymerase II transcription: structure and mechanism. Biochimica et biophysica acta. 2013;1829(1):2-8.
158. Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. Nature Reviews Genetics. 2015;16:71-84.
159. Cheung AC, Sainsbury S, Cramer P. Structural basis of initial RNA polymerase II transcription. The EMBO journal. 2011;30(23):4755-63.
160. Fuda NJ, Ardehali MB, Lis JT. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature. 2009;461(7261):186-92.
161. Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature reviews Genetics. 2012;13(10):720-31.
162. Selth LA, Sigurdsson S, Svejstrup JQ. Transcript Elongation by RNA Polymerase II. Annual review of biochemistry. 2010;79:271-93.
163. Kohoutek J. P-TEFb- the final frontier. Cell division. 2009;4:19.
164. Hirose Y, Ohkuma Y. Phosphorylation of the C-terminal Domain of RNA Polymerase II Plays Central Roles in the Intergrated Events of Eukaryotic Gene Expression. J Biochem. 2007;141:601-8.

165. Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, et al. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Molecular and cellular biology. 2009;29(20):5455-64.
166. Kim M, Suh H, Cho EJ, Buratowski S. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. The Journal of biological chemistry. 2009;284(39):26421-6.
167. Lin PS, Tremeau-Bravard A, Dahmus ME. The repetitive C-terminal domain of RNA polymerase II: multiple conformational states drive the transcription cycle. Chemical record. 2003;3(4):235-45.
168. Kim JB, Sharp PA. Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase. The Journal of biological chemistry. 2001;276(15):12317-23.
169. Liu Y, Warfield L, Zhang C, Luo J, Allen J, Lang WH, et al. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. Molecular and cellular biology. 2009;29(17):4852-63.
170. Ping Y-H, Rana TM. DSIF and NELF Interact with RNA Polymerase II Elongation Complex and HIV-1 Tat Stimulates P-TEFb-mediated Phosphorylation of RNA Polymerase II and DSIF during Transcription Elongation. The Journal of biological chemistry. 2001;276(16):12951-8.
171. Gu B, Eick D, Bensaude O. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. Nucleic acids research. 2012.
172. Hsin J-P, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes & development. 2012;26(19):2119-37.
173. Chen RA, Down TA, Stempor P, Chen QB, Egelhofer TA, Hillier LW, et al. The landscape of RNA polymerase II transcription initiation in C. elegans reveals promoter and enhancer architectures. Genome research. 2013;23(8):1339-47.
174. Reinke V, Krause M, Okkema P. Transcriptional regulation of gene expression in C. elegans. WormBook : the online review of C elegans biology. 2013:1-34.
175. Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I, Cheung M-S, et al. Broad chromosomal domains of histone modification patterns in C. elegans. Genome research. 2011;21(2):227-36.
176. Morris SA, Shibata Y, Noma K-i, Tsukamoto Y, Warren E, Temple B, et al. Histone H3 K36 Methylation Is Associated with Transcription Elongation in *Schizosaccharomyces pombe*. Eukaryotic Cell. 2005;4(8):1446-54.
177. Kim S-K, Jung I, Lee H, Kang K, Kim M, Jeong K, et al. Human Histone H3K79 Methyltransferase DOT1L Methyltransferase Binds Actively Transcribing RNA Polymerase II to Regulate Gene Expression. Journal of Biological Chemistry. 2012;287(47):39698-709.
178. Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, Leeuwen Fv, et al. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes & development. 2004:1263-71.
179. Kops O, Zhou XX, Lu KP. Pin1 modulates the dephosphorylation of the RNA polymerase II C-terminal domain by yeast Fcp1. FEBS letters. 2002;513:305-11.

180. Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, et al. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. Genes & development. 2012;26(17):1891-6.
181. Hanes SD. Prolyl isomerases in gene transcription. Biochimica et biophysica acta. 2014.
182. Graham PL, Kimble J. The mog-1 gene is required for the switch from spermatogenesis to oogenesis in Caenorhabditis elegans. Genetics. 1993;133(4):919-31.
183. Aroian RV, Levy AD, Koga M, Ohshima Y, Kramer JM, Sternberg PW. Splicing in Caenorhabditis elegans does not require an AG at the 3' splice acceptor site. Molecular and cellular biology. 1993;13(1):626-37.
184. de Bono M, Zarkower D, Hodgkin J. Dominant feminizing mutations implicate protein-protein interactions as the main mode of regulation of the nematode sex-determining gene tra-1. Genes & development. 1995;9(2):155-67.
185. Jones AR, Schedl T. Mutations in gld-1, a female germ cell-specific tumor suppressor gene in Caenorhabditis elegans, affect a conserved domain also found in Src-associated protein Sam68. Genes & development. 1995;9(12):1491-504.
186. Aravind L, Watanabe H, Lipman DJ, Koonin EV. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(21):11319-24.
187. Sarov M, Murray JI, Schanze K, Pozniakovski A, Niu W, Angermann K, et al. A genome-scale resource for in vivo tag-based protein function exploration in C. elegans. Cell. 2012;150(4):855-66.
188. Frokjaer-Jensen C, Wayne Davis M, Hopkins CE, Newman BJ, Thummel JM, Olesen S-P, et al. Single-copy insertion of transgenes in Caenorhabditis elegans. Nature genetics. 2008;40(11):1375-83.
189. Nance J, Lee J-Y, Goldstein B. Gastrulation in *C. elegans*. Wrombook. 2005:1-13.
190. Powell-Coffman JA, Knight J, Wood WB. Onset of C. elegans Gastrulation Is Blocked by Inhibition of Embryonic Transcription with an RNA polymerase Antisense RNA. Developmental biology. 1996;178:472-83.
191. Baugh LR, Hill AA, Slonim DK, Brown EL, Hunter CP. Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. Development. 2003;130(5):889-900.
192. Broitman-Maduro G, Owraghi M, Hung WWK, Kuntz S, Sternberg PW, Maduro MF. The NK-2 class homeodomain factor CEH-51 and the T-box factor TBX-35 have overlapping function in *C. elegans* mesoderm development. Development. 2009;136(16):2735-46.
193. Robertson SM, Shetty P, Lin R. Identification of lineage-specific zygotic transcripts in early *Caenorhabditis elegans* embryos. Developmental biology. 2004;276(2):493-507.
194. Rechtsteiner A, Ercan S, Takasaki T, Phippen TM, Egelhofer TA, Wang W, et al. The Histone H3K36 Methyltransferase MES-4 Acts Epigenetically to Transmit the Memory of Germline Gene Expression to Progeny. PLoS genetics. 2010;6(9):e1001091.
195. Reinke V, Gil IS, Ward S, Kazmer K. Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans. Development. 2004;131(2):311-23.

196. Bowman EA, Bowman CR, Ahn JH, Kelly WG. Phosphorylation of RNA polymerase II is independent of P-TEFb in the *C. elegans* germline. Development. 2013;140(17):3703-13.
197. Bowman EA, Kelly WG. RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation. Nucleus. 2014;5(3):1-13.
198. Bashirullah A, Cooperstock RL, Lipshitz HD. Spatial and Temporal Control of RNA Stability. PNAS. 2001;98(13):7025-8.
199. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols. 2012;7(3):562-78.
200. Bitar M, Boroni M, Macedo AM, Machado CR, Franco GR. The Spliced Leader Trans-Splicing Mechanism in Different Organisms: Molecular Details and Possible Biological Roles. Frontiers in Genetics. 2013;4.
201. Blumenthal T, Thomas J. *Cis* and *trans* mRNA splicing in *C. elegans*. Trends in Genetics. 1988;4(11):305-8.
202. Thomas J, Lea K, Zucker-Aperison E, Blumenthal T. The spliceosomal snRNAs of *Caenorhabditis elegans*. Nucleic acids research. 1990;18(9):2633-42.
203. Allen MA, Hillier LW, Waterston RH, Blumenthal T. A global analysis of *C. elegans trans*-splicing. Genome research. 2011;21:255-64.
204. Saito TL, Hashimoto S, Gu SG, Morton JJ, Stadler M, Blumenthal T, et al. The transcription start site landscape of C. elegans. Genome research. 2013;23(8):1348-61.
205. Kruesi WS, Core LJ, Waters CT, Lis JT, Meyer BJ. Condensin controls recruitment of RNA polymerase II to achieve nematode X-chromosome dosage compensation. Proudfoot N, editor2013 2013-06-18 18:46:53.
206. Blumenthal T. Trans-splicing and operons in *C. elegans*. WormBook : the online review of C elegans biology. 2012:1-11.
207. Morton JJ, Blumenthal T. RNA processing in C. elegans. Methods in cell biology. 2011;106:187-217.
208. Pandya-Jones A, Black DL. Co-transcriptional splicing of constitutive and alternative exons. Rna. 2009;15(10):1896-908.
209. Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, et al. SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. Molecular cell. 2007;26(6):867-81.
210. CUSTÓDIO N, CARVALHO C, CONDADO I, ANTONIOU M, BLENCOWE BJ, CARMO-FONSECA M. In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. Rna. 2004;10(4):622-33.
211. Das R, Dufu K, Romney B, Feldt M, Elenko M, Reed R. Functional coupling of RNAP II transcription to spliceosome assembly. Genes & development. 2006;20:1100-9.
212. Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu X-D. The splicing factor SC35 has an active role in transcriptional elongation. Nature structural & molecular biology. 2008;15(8):819-26.
213. Zahler AM. Pre-mRNA splicing and its regulation in *Caenorhabditis elegans*. WormBook : the online review of C elegans biology. 2012.
214. Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P. Uniform transitions of the general RNA polymerase II transcription complex. Nature structural & molecular biology. 2010;17(10):1272-8.

215. Hampsey M, Reinberg D. Tails of Intrigue: Phosphorylation of RNA Polymerase II Mediates Histone Methylation. Cell Cycle. 2003;113:429-32.
216. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, et al. Active genes are tri-methylated at K4 of histone H3. Nature. 2002;419(6905):407-11.
217. Zhou Q, Li T, Price DH. RNA polymerase II elongation control. Annual review of biochemistry. 2012;81:119-43.
218. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nature genetics. 2008;40(7):897-903.
219. Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, et al. DOT1L/KMT4 Recruitment and H3K79 Methylation Are Ubiquitously Coupled with Gene Transcription in Mammalian Cells. Molecular and cellular biology. 2008;28(8):2825-39.
220. Crisucci EM, Arndt KM. The Roles of the Paf1 Complex and Associated Histone Modifications in Regulating Gene Expression. Genetics research international. 2011;2011:15.
221. Guang S, Bochner AF, Pavelec DM, Burkhart KB, Harding S, Lachowiec J, et al. An Argonaute Transports siRNAs from the Cytoplasm to the Nucleus. Science. 2008;321(5888):537-41.
222. Fischer SE, Pan Q, Breen PC, Qi Y, Shi Z, Zhang C, et al. Multiple small RNA pathways regulate the silencing of repeated and foreign genes in C. elegans. Genes & development. 2013;27(24):2678-95.
223. Fong N, Bentley DL. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes & development. 2001;15(14):1783-95.
224. de Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J, et al. Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nature structural & molecular biology. 2011;18(9):977-83.
225. Moehle EA, Braberg H, Krogan NJ, Guthrie C. Adventures in time and space. RNA Biology. 2014;11(4):313-9.
226. Brzyżek G, Świeżewski S. Mutual interdependence of splicing and transcription elongation. Transcription. 2015;6(2):37-9.
227. Khodor YL, Menet JS, Tolan M, Rosbash M. Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. Rna. 2012;18(12):2174-86.
228. Brody Y, Neufeld N, Bieberstein N, Causse SZ, Bohnlein EM, Neugebauer KM, et al. The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. PLoS biology. 2011;9(1):e1000573.
229. Zanetti S, Puoti A. Sex Determination in the Caenorhabditis elegans Germline. In: Schedl T, editor. Germ Cell Development in C elegans. Advances in Experimental Medicine and Biology. 757: Springer New York; 2013. p. 41-69.
230. Graham PL, Schedl T, Kimble J. More mog genes that influence the switch from spermatogenesis to oogenesis in the hermaphrodite germ line of Caenorhabditis elegans. Developmental Genetics. 1993;14(6):471-84.
231. Puoti A, Kimble J. The *Caenorhabditis elegans* Sex Determination Gene *mog-1*

Encodes a Member of the DEAH-Box Protein Family. Molecular and cellular biology. 1999;19(3):2189-97.
232. Puoti A, Kimble J. The hermaphrodite spermyoocyte switch requires the *Caenorhabditis elegans* homologs of PRP2 and PRP22. PNAS. 2000;97(7):3276-81.
233. Kerins JA, Hanazawa M, Dorsett M, Schedl T. PRP-17 and the Pre-mRNA Splicing Pathway Are Preferentially Required for the Proliferation Versus Meiotic Development Decision and Germline Sex Determination in *Caenorhabditis elegans*. Developmental Dynamics. 2010;239:1555-72.
234. Barberan-Soler S, Zahler AM. Alternative splicing regulation during C. elegans development: splicing factors as regulated targets. PLoS genetics. 2008;4(2):e1000001.
235. Spingola M, Grate L, Haussler D, Ares MJ. Genome-wide bioinformatic and molecular
analysis of introns in *Saccharomyces cerevisiae*. Rna. 1999;5:221-34.
236. Furuhashi H, Takasaki T, Rechtsteiner A, Li T, Kimura H, Checchi PM, et al. Transgenerational epigenetic regulation of *C. elegans* primordial germ cells. Epigenetics & Chromatin. 2010;3(15):1-21.
237. Strome S, Wood WB. Generation of asymmetry and segregation of germ-line granules in early C. elegans embryos. Cell. 1983;35(1):15-25.
238. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Meth. 2012;9(7):671-5.
239. Klenova E, Chernukhin I, Inoue T, Shamsuddin S, Norton J. Immunoprecipitation techniques for the analysis of transcription factor complexes. Methods. 2002;26(3):254-9.
240. Ercan S, Whittle C, Lieb J. Chromatin immunoprecipitation from *C. elegans* embryos. Protocol Exchange. 2007.
241. Bowman S, Simon M, Deaton A, Tolstorukov M, Borowsky M, Kingston R. Multiplexed Illumina sequencing libraries from picogram quantities of DNA. BMC Genomics. 2013;14(1):466.
242. Meissner B, Warner A, Wong K, Dube N, Lorch A, McKay SJ, et al. An Integrated Strategy to Study Muscle Development and Myofilament Structure in *Caenorhabditis elegans*. PLoS genetics. 2009;5(6):e1000537.
243. Wang X, Zhao Y, Wong K, Ehlers P, Kohara Y, Jones SJ, et al. Identification of genes expressed in the hermaphrodite germ line of C. elegans using SAGE. BMC Genomics. 2009;10:213-.
244. Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. Nature genetics. 2009;41(3):376-81.
245. Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009;10(3):R25.
246. Zhang Y, Liu T, Meyer C, Eeckhoute J, Johnson D, Bernstein B, et al. Model-based Analysis of ChIP-Seq (MACS). Genome biology. 2008;9(9):R137.
247. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8.
248. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-

associated functional and structural residues in proteins. Nucleic acids research. 2006;34(Web Server issue):W362-5.
249. Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. Nucleic acids research. 2013;41(Database issue):D344-7.
250. Holliday R, Pugh JE. DNA Modification Mechanisms and Gene Activity during Development. Science. 1975;187(4173):226-32.
251. Monk M. Methylation and the X chromosome. BioEssays. 1986;4(5):204-8.
252. Harland RM. Inheritance of DNA methylation in microinjected eggs of Xenopus laevis. Proceedings of the National Academy of Sciences of the United States of America. 1982;79(7):2323-7.
253. Okano M, Bell DW, Haber DA, Li E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell. 1999;99(3):247-57.
254. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 1992;69(6):915-26.
255. Vilkaitis G, Suetake I, Klimašauskas S, Tajima S. Processive Methylation of Hemimethylated CpG Sites by Mouse Dnmt1 DNA Methyltransferase. Journal of Biological Chemistry. 2005;280(1):64-72.
256. Bártová E, Krejčí J, Harničarová A, Galiová G, Kozubek S. Histone Modifications and Nuclear Architecture: A Review. Journal of Histochemistry & Cytochemistry. 2008;56(8):711-21.
257. Suganuma T, Workman JL. Signals and Combinatorial Functions of Histone Modifications. Annual review of biochemistry. 2011;80(1):473-99.
258. Young Richard A. Control of the Embryonic Stem Cell State. Cell. 2011;144(6):940-54.
259. Schaefer S, Nadeau JH, Handling Editor Gregory AW. The Genetics of Epigenetic Inheritance: Modes, Molecules, and Mechanisms. The Quarterly Review of Biology. 2015;90(4):381-415.
260. Zaidi SK, Young DW, Montecino MA, Lian JB, van Wijnen AJ, Stein JL, et al. Mitotic bookmarking of genes: a novel dimension to epigenetic control. Nature reviews Genetics. 2010;11(8):583-9.
261. Steffen PA, Ringrose L. What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nature reviews Molecular cell biology. 2014;15(5):340-56.
262. Heard E, Martienssen Robert A. Transgenerational Epigenetic Inheritance: Myths and Mechanisms. Cell. 2014;157(1):95-109.
263. Langley AR, Smith JC, Stemple DL, Harvey SA. New insights into the maternal to zygotic transition. Development. 2014;141(20):3834-41.
264. Lee MT, Bonneau AR, Giraldez AJ. Zygotic Genome Activation During the Maternal-to-Zygotic Transition. Annual review of cell and developmental biology. 2014;30(1):581-613.
265. Hontelez S, van Kruijsbergen I, Georgiou G, van Heeringen SJ, Bogdanovic O, Lister R, et al. Embryonic transcription is controlled by maternally defined chromatin state. Nature communications. 2015;6.

266. Leitch HG, Smith A. The mammalian germline as a pluripotency cycle. Development. 2013;140(12):2495-501.
267. Reik W, Surani MA. Germline and Pluripotent Stem Cells. Cold Spring Harbor Perspectives in Biology. 2015;7(11).
268. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448(7154):714-7.
269. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews Genetics. 2010;11(3):204-20.
270. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(5):1412-7.
271. Iwasaki M, Paszkowski J. Epigenetic memory in plants. The EMBO journal. 2014;33(18):1987-98.
272. Shipony Z, Mukamel Z, Cohen NM, Landan G, Chomsky E, Zeliger SR, et al. Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. Nature. 2014;513(7516):115-9.
273. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315-22.
274. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, et al. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. Cell. 2006;126(6):1189-201.
275. Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, et al. The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. Journal of Biological Chemistry. 2010;285(34):26114-20.
276. Kornberg RD, Lorch Y. Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. Cell. 1999;98(3):285-94.
277. Cutter AR, Hayes JJ. A brief review of nucleosome structure. FEBS letters. 2015;589(20PartA):2914-22.
278. Biswas M, Voltz K, Smith JC, Langowski J. Role of Histone Tails in Structural Stability of the Nucleosome. PLoS Comput Biol. 2011;7(12):e1002279.
279. Arya G, Schlick T. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. Proceedings of the National Academy of Sciences. 2006;103(44):16236-41.
280. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nature biotechnology. 2010;28(8):817-25.
281. Li B, Carey M, Workman JL. The role of chromatin during transcription. Cell. 2007;128(4):707-19.
282. Shilatifard A. Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. Annual review of biochemistry. 2006;75(1):243-69.
283. Kouzarides T. Chromatin Modifications and Their Function. Cell. 2007;128(4):693-705.
284. Gu B, Lee MG. Histone H3 lysine 4 methyltransferases and demethylases in self-renewal and differentiation of stem cells. Cell & Bioscience. 2013;3:39-.

285. Vermeulen M. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell. 2007;131:58-69.
286. Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes & development. 2003;17.
287. Bertoli C, Skotheim JM, de Bruin RAM. Control of cell cycle transcription during G1 and S phases. Nature reviews Molecular cell biology. 2013;14(8):518-28.
288. Lodhi N, Ji Y, Tulin A. Mitotic Bookmarking: Maintaining Post-Mitotic Reprogramming of Transcription Reactivation. Current Molecular Biology Reports. 2016;2(1):10-5.
289. Kadauke S, Blobel GA. Mitotic bookmarking by transcription factors. Epigenetics & Chromatin. 2013;6:6-.
290. Wang F, Higgins JMG. Histone modifications and mitosis: countermarks, landmarks, and bookmarks. Trends in Cell Biology. 2013;23(4):175-84.
291. Kouskouti A, Talianidis I. Histone modifications defining active genes persist after transcriptional and mitotic inactivation. The EMBO journal. 2005;24(2):347-57.
292. Benayoun Bérénice A, Pollina Elizabeth A, Ucar D, Mahmoudi S, Karra K, Wong Edith D, et al. H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency. Cell. 2014;158(3):673-88.
293. Duncan EM, Chitsazan AD, Seidel CW, Alvarado AS. Set1 and MLL1/2 target distinct sets of functionally different genomic loci in vivo. Cell reports. 2015;13(12):2741-55.
294. Sachs M, Onodera C, Blaschke K, Ebata KT, Song JS, Ramalho-Santos M. Bivalent chromatin marks developmental regulatory genes in the mouse embryonic germline in vivo. Cell reports. 2013;3.
295. Vastenhouw NL, Schier AF. Bivalent histone modifications in early embryogenesis. Current opinion in cell biology. 2012;24(3):374-86.
296. Harikumar A, Meshorer E. Chromatin remodeling and bivalent histone modifications in embryonic stem cells. EMBO reports. 2015;16(12):1609-19.
297. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, et al. Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLoS genetics. 2008;4(10):e1000242.
298. Li Q, Lian S, Dai Z, Xiang Q, Dai X. BGDB: a database of bivalent genes. Database. 2013;2013.
299. Ang Y-S, Tsai S-Y, Lee D-F, Monk J, Su J, Ratnakumar K, et al. Wdr5 Mediates Self-Renewal and Reprogramming via the Embryonic Stem Cell Core Transcriptional Network. Cell. 2011;145(2):183-97.
300. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, et al. Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. Cell. 2006;125(2):301-13.
301. Tsai M-C, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, et al. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. Science. 2010;329(5992):689-93.
302. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011;472(7341):120-4.

303. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature. 2010;464(7291):1082-6.
304. Fazzio TG, Huff JT, Panning B. An RNAi Screen of Chromatin Proteins Identifies Tip60-p400 as a Regulator of Embryonic Stem Cell Identity. Cell. 2008;134(1):162-74.
305. Shen X, Liu Y, Hsu Y-J, Fujiwara Y, Kim J, Mao X, et al. EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. Molecular cell. 2008;32(4):491-502.
306. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K. The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation. Molecular and cellular biology. 2007;27(10):3769-79.
307. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature. 2006;441(7091):349-53.
308. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell. 2006;125(2):315-26.
309. Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, et al. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nature structural & molecular biology. 2013;20(9):1093-7.
310. Jiang H, Shukla A, Wang X, Chen W-y, Bernstein BE, Roeder RG. Role for Dpy-30 in ES Cell-Fate Specification by Regulation of H3K4 Methylation within Bivalent Domains. Cell. 2011;144(4):513-25.
311. Ng RK, Gurdon JB. Epigenetic memory of active gene transcription is inherited through somatic cell nuclear transfer. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(6):1957-62.
312. Ng RK, Gurdon JB. Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. Nature cell biology. 2008;10(1):102-9.
313. Muramoto T, Müller I, Thomas G, Melvin A, Chubb JR. Methylation of H3K4 Is Required for Inheritance of Active Transcriptional States. Current Biology. 2010;20(5):397-406.
314. Katz DJ, Edwards TM, Reinke V, Kelly WG. A C. elegans LSD1 Demethylase Contributes to Germline Immortality by Reprogramming Epigenetic Memory. Cell. 2009;137(2):308-20.
315. Merritt C, Rasoloson D, Ko D, Seydoux G. 3' UTRs Are the Primary Regulators of Gene Expression in the *C. elegans* Germline. Current Biology. 2008;18(19):1476-82.
316. Li T, Kelly WG. A Role for Set1/MLL-Related Components in Epigenetic Regulation of the *Caenorhabditis elegans* Germ Line. PLoS genetics. 2011;7(3):e1001349.
317. Greer EL, Beese-Sims SE, Brookes E, Spadafora R, Zhu Y, Rothbart SB, et al. A histone methylation network regulates transgenerational epigenetic memory in C. elegans. Cell reports. 2014;7(1):113-26.

318. Bowerman B. 3 Maternal Control of Pattern Formation in Early Caenorhabditis elegans Embryos. In: Roger AP, Gerald PS, editors. Current Topics in Developmental Biology. Volume 39: Academic Press; 1998. p. 73-117.
319. Newman-Smith ED, Rothman JH. The maternal-to-zygotic transition in embryonic patterning of Caenorhabditis elegans. Current Opinion in Genetics & Development. 1998;8(4):472-80.
320. Ahringer J. Maternal control of a zygotic patterning gene in Caenorhabditis elegans. Development. 1997;124(19):3865-9.
321. Labouesse M, Mango SE. Patterning the C. elegans embryo: moving beyond the cell lineage. Trends in Genetics. 1999;15(8):307-13.
322. Gilleard JS, McGhee JD. Activation of Hypodermal Differentiation in theCaenorhabditis elegans Embryo by GATA Transcription Factors ELT-1 and ELT-3. Molecular and cellular biology. 2001;21(7):2533-44.
323. Page BD, Zhang W, Steward K, Blumenthal T, Priess JR. ELT-1, a GATA-like transcription factor, is required for epidermal cell fates in Caenorhabditis elegans embryos. Genes & development. 1997;11(13):1651-61.
324. Priess JR, Thomson JN. Cellular interactions in early C. elegans embryos. Cell. 1987;48(2):241-50.
325. Zhu J, Fukushige T, McGhee JD, Rothman JH. Reprogramming of early embryonic blastomeres into endodermal progenitors by a Caenorhabditis elegans GATA factor. Genes & development. 1998;12(24):3809-14.
326. Zhu J, Hill RJ, Heid PJ, Fukuyama M, Sugimoto A, Priess JR, et al. end-1 encodes an apparent GATA factor that specifies the endoderm precursor in Caenorhabditis elegans embryos. Genes & development. 1997;11(21):2883-96.
327. Furuhashi H, Kelly WG. The Epigenetics of Germ-line Immortality: Lessons from an Elegant Model System. Development, Growth & Differentiation. 2010;52(6):527-32.
328. Arico JK, Katz DJ, van der Vlag J, Kelly WG. Epigenetic Patterns Maintained in Early *Caenorhabditis elegans* Embryos Can Be Established by Gene Activity in the Parental Germ Cells. PLoS genetics. 2011;7(6):e1001391.
329. Robert Valérie J, Mercier Marine G, Bedet C, Janczarski S, Merlet J, Garvis S, et al. The SET-2/SET1 Histone H3K4 Methyltransferase Maintains Pluripotency in the Caenorhabditis elegans Germline. Cell reports. 2014;9(2):443-50.
330. Bender LB, Cao R, Zhang Y, Strome S. The MES-2/MES-3/MES-6 Complex and Regulation of Histone H3 Methylation in C. elegans. Current Biology. 2004;14(18):1639-43.
331. Simonet T, Dulermo R, Schott S, Palladino F. Antagonistic functions of SET-2/SET1 and HPL/HP1 proteins in C. elegans development. Developmental biology. 2007;312(1):367-83.