An analysis of the essential chromatin factor Zinc Finger Protein 1 (ZFP-1), and a study of the involvement of RNAi factors in histone processing in *Caenorabditis elegans*

Daphne Christina Anastasiades Avgousti

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

© 2012

Daphne C. Avgousti

All Rights Reserved

ABSTRACT

An analysis of the essential chromatin factor Zinc Finger Protein 1 (ZFP-1), and a study of the involvement of RNAi factors in histone processing in *C. elegans*

Daphne C. Avgousti

The formation of chromatin defines when and where genes will be expressed, from the histone proteins forming the nucleosome and their many post-translational modifications, to the immense number of proteins that bind these modifications. This thesis is comprised of two projects: the first is an analysis of the essential plant homeodomain-containing protein called zinc finger protein 1 (ZFP-1); the second is a study into how RNA interference (RNAi) factors are involved in histone production in *C. elegans.* I investigate the physical and biological properties of the PHD fingers of ZFP-1 and find that 1) they are essential for viability and 2) they specifically bind to methylated lysine 4 on histone H3. This study has expanded our understanding of the molecular nature of ZFP-1, the *C. elegans* ortholog of AF10, which has a role in chromosomal translocations promoting leukemia. I also determine that the RNAi factors CSR-1, EGO-1 and DRH-3 are required for proper histone production in *C. elegans*. Severe histone depletion results from the knockdown of these RNAi pathway components, which explains both the phenotypes of sterility and chromosome segregation defects in early embryos that are associated with mutants of these factors. This discovery explains the well-known, but poorly understood, phenotypes of the RNAi mutants and provides the first evidence for RNAi positively affecting gene expression.

TABLE OF CONTENTS

List of Figures and Tables

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Appendix I:

Appendix II:

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Alla Grishok for invaluable time, instruction, insight and most of all helping me mature as a scientist. My years in the Grishok lab have undoubtedly had many upward and downward turns, but Alla's strength and perseverance have always kept the lab going.

Next I would like to thank my committee members, Dr. Larry Shapiro, Dr. Art Palmer and Dr. Richard Mann for their guidance and support over the years. Larry especially has been instrumental in my understanding of work with proteins and I greatly appreciate that his door and his lab were always open to me. Art has also been one of the most reliable and nicest professors I have had the pleasure to know. Art's experience, wit, and advice have impacted my scientific development in ways that are hard to express.

In the Biochemistry Office, I would also like to thank departmental administrators Ed Johnson and Rachel Hernandez for always being available and willing to help for every endless tedious question I seemed to find. I thank them for their patience and confidence in me and all the students in our department.

I would also like to thank Dr. Lori Sussel and the long-term NIH Endocrinology Training Grant (DK07328). Not only did this grant provide me with funding and travel expenses for conferences for multiple years, but Lori has also been instrumental in guiding me towards post-doctoral work.

In the Grishok lab, I would like to particularly thank Lisa Kennedy and Germano Cecere for sharing time, space, tears and laughs with me over the past 5 years. It has been a pleasure to work with these two brilliant colleagues who have taught me both science and endurance.

Without good friends, no one can survive graduate school. I'd like to say a special thank you to Julia Brasch for her unending support both scientifically and emotionally. Julia has been a hidden gem who always manages to find the silver lining in a cloud. I'd also like to thank Elaine Zhang and Cate Jensen for being around to listen and be supportive.

I would like to thank my parents, George and Ana Maria Anastasiades, whose patience and drive have inspired me to pursue this degree.

Finally, I'd like to thank my husband, Andreas, for endless love, support, guidance, and patience. I would never be here without him by my side.

To my husband, Andreas

Είσαι η ζωή μου.

Chapter 1: Introduction

This chapter is divided into several sections that will introduce the various topics that this thesis will cover. Although perhaps the topics will seem somewhat unrelated, it is certainly true that many a thesis has started out in one direction and taken several different turns along the way. When I joined the Grishok lab, my original proposed thesis plan was to investigate the biochemical properties of an RNAi-promoting chromatin factor called zinc finger protein 1, ZFP-1. In the process of investigating both the biological and physical properties of this protein, I discovered an unusual phenotype produced by the knockdown of RNAi components in *C. elegans*: the depletion of histone proteins. This specific depletion explained the known sterility and chromosome segregation phenotypes associated with RNAi mutants but was contrary to the published model (Claycomb et al 2009). Given the importance of describing these new findings, towards the end of my graduate work I devoted the majority of my efforts into this project. The result has been two very different manuscripts. Future work will be needed to understand the precise yet complex relationship between chromatin and RNAi, specifically in the germline of *C elegans*.

Chromatin structure

The existence of the combination of nucleic acid and protein called chromatin has been known for well over a century (reviewed in Olins & Olins, 2003). However, it was not until the mid-twentieth century when the structure of DNA was proposed that chromatin structure could begin to be investigated. Since DNA replication, transcription of RNA and maintenance of the genome are critical for maintaining cellular viability, it stands to reason that understanding proteins that wrap up DNA in the cell and control access to the DNA are of paramount importance for the understanding of the biology of a eukaryotic cell.

DNA is tightly wrapped around a core of histone proteins, which makes up the nucleosome (Luger et al, 1997; Oudet et al, 1975; Schaner et al, 2003; Van Holde et al, 1980). This is the basic unit of chromatin, which contains approximately 147 base pairs of DNA wrapped around one nucleosome; it is artistically depicted by Grandville (Figure 1). Nucleosomes are subsequently compacted from 'beads on a string' to a condensed fiber to fit in the nucleus (Kornberg & Lorch, 1999; Olins & Olins, 2003). The nucleosome core is made up of an octamer of histone proteins containing two of each histone H2A, H2B, H3 and H4. These histones are positively charged as they contain a range of 10-30% arginine and lysine residues (Horn & Peterson, 2002) and are highly conserved from yeast to humans (Figure 2).

The N-terminal tails of the core histone proteins extend outward from the nucleosome beyond the wrapped DNA. These tails are rich in arginine and lysine residues that are sites for post-translational modifications (PTMs) including methylation, acetylation, phosphorylation and ubiquitylation among many others (Horn & Peterson, 2002; Rando & Chang, 2009; Zhang & Reinberg, 2001). The PTMs are deposited, recognized and modified by a vast array of proteins and have been associated with different types of active and silent transcription. It has been

proposed that there is a 'histone code' that controls the expression of various genomic regions. Histone modifications have been termed 'epigenetic' marks, suggesting that they are heritable and control gene expression across generations although not defined in the genetic code. However, whether a distinct code exists and whether or not histone modifications are heritable is still under debate within the field (Ptashne, 2007; Rando & Chang, 2009; Rando & Verstrepen, 2007; Smith & Shilatifard, 2010; Taverna et al, 2007; van Steensel, 2011).

There are histone tail PTMs associated with active transcription, such as H3K4me3 and H3K36me3, while others are associated with silenced transcription, such as H3K9me3 and H3K27me3 (Rando & Chang, 2009; Zhang et al, 2012) (Figure 2, highlighted). The different combinations of methylation and acetylation marks on histone tails are thought to recruit effector proteins that will maintain a given state of chromatin. For example, H3K27me is bound by Polycomb (Pc) proteins to help maintain silencing during differentiation (Grimaud et al, 2006; Smith & Shilatifard, 2010). Conversely, H3K4 methylation, which is deposited by the COMPASS complex (Smith & Shilatifard, 2010) is associated with active transcription and has been shown to counteract Pc silencing of developmental genes (Eissenberg & Shilatifard, 2010). H3K4 methylation is also found on histones associated with many other genes, most commonly at the 5′ end, where it is highly correlated with active transcription although its precise purpose is unknown (Kim & Buratowski, 2009; Rando & Chang, 2009).

The COMPASS complex (complex of proteins associated with Set1) consists of several components that are well conserved from yeast to humans (Smith et al, 2011; Takahashi et al, 2011). The core methyltransferase is a SET-domain (Su (var) Enhancer of Zeste and Trithorax) containing protein is known as MLL (mixed lineage leukemia) in humans (Eissenberg & Shilatifard, 2010; Takahashi et al, 2011). Other components of the COMPASS complex include

RbBP5, Ash2L and WDR5, which are conserved in the different COMPASS-like complexes from yeast to humans (Eissenberg & Shilatifard, 2010; Li & Kelly, 2011; Takahashi et al, 2011). It has been shown that multiple components of the complex are required for complete trimethylation of H3K4 (Patel et al, 2011) and, even more surprisingly, the components together without MLL can carry out methylation of H3K4 (Patel et al, 2011). However, the precise regulation of the different degrees of methylation and the components needed for their deposition has yet to be understood.

There are several methyltransferase enzymes responsible for depositing methylation on histone tails, the majority of which contain SET domains. For example, the SET-domain containing SUV39 family of histone methyltransferases (HMTs) carries out methylation of H3K9. Methylation of H3K27 and H3K36 are carried out by PRC2 and SETD2 respectively, two enzymes that also contain the catalytic SET domain (Zhang et al, 2012).

Another type of methyltransferase domain, distinct from the SET-domain containing enzymes, is DOT1 (disrupter of telomeric silencing) (Nguyen & Zhang, 2011). This is the only histone methyltransferase found to methylate H3K79, which is unusual in that it resides in the globular core of the histone H3, in contrast to other histone tail PTMs that are generally found on the protruding N-terminal tails. DOT1 has been shown to play an essential role in transcriptional elongation as well as in the development of leukemia (Chang et al, 2010). Regulation of H3K79 methylation has also been linked to Wnt signaling in a DOT1-containing complex called DotCom (Mohan et al, 2010a).

Chromatin-Binding Proteins

There are many types of domains that bind specifically to methylated lysine residues on histone tails. They can be divided into three groups as follows: the tudor family, which contains tudor and chromodomains; the zinc family that contains the plant homeodomains and other zinc-binding types, and the WD40 group, which includes a more divergent set of proteins with a barrel shape (Khorasanizadeh, 2011). The plant homeodomain (PHD) family is the largest one and is capable of recognizing different histone tail methylation states. The interactions of PHD fingers and histone tails are highly regulated and are very precise. PHD fingers can distinguish between mono-, di- and trimethylation of a single lysine residue, and take into account the methylation or acetylation states of nearby lysines on the same histone tail (Musselman & Kutateladze, 2011; Sanchez & Zhou, 2011).

The PHD finger is a small domain of approximately 50-80 amino acid residues and contains a zinc-binding motif that is common to many chromatin and DNA-associated proteins. The PHD fold consists of two zinc ions anchored by a Cys_4-His-Cys_3 motif in cross brace topology (Musselman & Kutateladze, 2009; Sanchez & Zhou, 2011) (Figure 3). The majority of PHD fingers bind to methylated lysines specifically on histone H3. For example, BPTF (Li et al, 2006) and ING2 (Peña et al, 2006) PHD fingers both bind to H3K4me3, whereas BHC80 binds to unmethylated H3K4 (Lan et al, 2007).

Although the binding of PHD fingers to histone tails has been known since 2006 (Doyon et al, 2006; Li et al, 2006), the effect of this binding on gene expression is varied. The function of PHD fingers seems to be that of an anchor within a protein or complex, which can stimulate catalytic activity of the same protein or another subunit of the complex. In some cases, PHD fingers also bind to the product of enzymatic activity, such as BHC80, which binds to

unmethylated H3K4. This is the product of demethylation of its complex, LSD-1 (Lan et al, 2007).

The type of residue that PHD fingers bind does not dictate whether that binding will lead to activation or repression of specific target genes. For example, PHD fingers of ING2 and ING5 both bind to H3K4me3, but lead to opposite outcomes (Doyon et al, 2006; Musselman & Kutateladze, 2011). ING2 is part of an HDAC complex leading to gene repression where ING5 associates with a HAT complex that leads to gene activation (Doyon et al, 2006). Therefore, it seems that the downstream function of a PHD finger is determined by the complex or protein in which it resides.

Another layer of complexity of histone PTMs is the various combinations that can occur and lead to different outcomes. For example, the PHD finger of PHF8 binds to H3K9me2/3 and promotes demethylation carried out by the Jumonji domain of the same protein. However, it has been shown that if the histone H3 tail is also trimethylated at K4, then the demethylation of K9 occurs 12-fold more efficiently (Horton et al, 2010). This piece of evidence supports the idea of cross-talk of histone-binding proteins, in this case, in *cis*.

Histone modifications and the *C. elegans* germline

The maintenance of histone modifications is extremely important to ensure the correct expression of specific genes at various time points of development. *C. elegans* is a very useful model organism to study genetics and development, especially the changes in histone modifications throughout development. Due to the efforts of many labs and the modENCODE project (Gerstein et al, 2010; Washington et al, 2011), we have genome-wide localization datasets for several chromatin-binding proteins. Importantly, histone modifications throughout the genome at different stages of development have been mapped. These datasets have only been made publicly available very recently, but we have had the opportunity to collaborate with labs involved in the modENCODE project that have shared with us invaluable antibodies and datasets.

There have been various studies on the histone methylation state of H3K4 in *C. elegans*, linking the COMPASS complex to aging (Greer et al, 2010) and epigenetic gene regulation in the germline (Li & Kelly, 2011). It is known that proper oocyte development requires precise maintenance of H3K4me3 although the exact mechanism for this is unknown (Schaner, 2006).

The *C. elegans* germline represents a timeline of gamete development that has been widely used to understand the roles of various histone modifiers in meiosis and germline maintenance (Furuhashi et al, 2010; Strome, 2005). Adult worms can be dissected to release intact gonads that can be stained for various histone PTMs or germline proteins (Figure 4). The distal germline begins with the mitotic zone where primordial germ cells are maintained. Moving proximally, the cells progress from mitosis to meiosis in what is known as the 'transition zone', which is followed by meiosis in the pachytene region progressing towards oocyte development in the adult gonad (Figure 4).

The germline carries a distinct pattern of histone modifications involving marks associated with both repression and activation, which can be easily visualized with the use of histone specific antibodies in immunostaining. Generally, somatic genes tend to be silenced in the germline whereas germline-specific genes are expressed. This balance of gene expression in the germline is essential for fertility in adults (Schaner, 2006; Strome, 2005). The *C. elegans* germline provides a unique model to study the role of essential proteins in development and has been be greatly utilized in the work presented in this thesis.

Histone production

Besides carrying various covalent PTMs, histone proteins need to be produced in a controlled way in order to ensure sufficient amounts of histones at the time of replication, during S-phase (Jaeger et al, 2005). Thus, the production of core histone proteins is coupled to the cell cycle and is dependent on replication. In order for precise regulation of histone production to occur, the cell has evolved to treat the 3′UTR of histone genes differently than most other transcripts (Davila Lopez & Samuelsson, 2007). Replication-dependent histone mRNAs end with a highly conserved stem-loop structure, unlike all other mRNAs that end in a poly(A) tail (Dominski & Marzluff, 2007; Marzluff et al, 2008). The addition of a poly(A) tail to the vast majority of mRNAs is carried out by the cleavage and polyadenylation machinery and is dependent on the presence of a conserved hexanucleotide polyadenylation signal (PAS). The 3′end of the mRNA following the PAS is removed and a poly(A) tail is added (Marzluff et al, 2008).

Conversely, in the case of replication-dependent histone mRNAs, the 3′end is cleaved after a conserved stem-loop. This cleavage is promoted by a small nuclear RNA, U7, that binds

to a complementary conserved sequence known as the histone downstream element (HDE) (Davila Lopez & Samuelsson, 2007; Marzluff et al, 2008). U7 is a component of a small nuclear ribonucleoprotein (snRNP) that consists of a heptameric ring with five Sm proteins and two U7 snRNP-specific Sm-like proteins, LSM11 and LSM10. The cleavage complex contains cleavage and polyadenylation specificity factor subunits CSPF73 and CPSF100, in addition to a conserved stem-loop binding protein (SLBP) (Marzluff et al, 2008). SLBP is also responsible for efficient export of the histone mRNA from the nucleus to the cytoplasm and its proper translation (Allard et al, 2005; Sullivan et al, 2009).

Histone mRNA processing has been studied in *Xenopus*, *Drosophila* and mammals (Dominski et al, 2005; Gorgoni et al, 2005; Sanchez & Marzluff, 2002; Thelie et al, 2012). A recent study of the conservation of the histone processing machinery showed that neither U7, nor Lsm11 are conserved in the *Caenorhabditis* genus (Davila Lopez & Samuelsson, 2007), although little is known about alternative factors involved in histone mRNA processing in the worm.

The homolog of SLBP, called CDL-1 (cell death lethal), does exist in *C. elegans* and has been shown to be required for germline fertility and embryonic development (Keall et al, 2007; Kodama et al, 2002; Pettitt et al, 2002). In addition to CDL-1, the depletion of any single histone protein can result in sterility and chromosome segregation defects (Kodama et al, 2002; Sönnichsen et al, 2005). Given the importance of histone proteins in DNA replication, it is beneficial to understand the divergent mechanisms of histone mRNA processing in *C. elegans*. In fact, depletion of RNAi factors results in the phenotypes of sterility and chromosome segregation defects as well (Claycomb et al, 2009). This prompted us to investigate histone abundance in the RNAi mutants, as will be shown in Chapter 3.

RNA interference

RNA interference was first discovered in *C. elegans* as a mechanism by which injected double stranded RNA (dsRNA) could induce the silencing of targeted mature mRNA (Fire et al, 1998). The classical RNA interference pathway involves a double-stranded RNA (dsRNA) precursor that is cleaved by the RNase III type endoribonuclease, Dicer, into smaller, 21-22 nucleotide-long RNAs. These small RNAs are then loaded onto Argonaute proteins and can either bind to target mRNAs by base-pairing and mediate the cleavage of target mRNAs, or serve as primers for RdRPs to create more dsRNA in an amplification step that exists in plants, fission yeast and nematodes (Azimzadeh Jamalkandi & Masoudi-Nejad, 2011; Ketting, 2011).

Although the cleavage abilities of Argonaute proteins have been characterized in other systems (Cenik & Zamore, 2011; Tolia & Joshua-Tor, 2007), the downstream steps of exogenous RNAi in *C. elegans* remain poorly understood. There is evidence for both post-transcriptional gene silencing (PTGS) (Fire et al, 1998; Montgomery et al, 1998) as well as transcriptional gene silencing (TGS) (Burton et al, 2011; Grishok, 2005; Gu et al, 2012; Guang et al, 2010). The most well-studied example of RNAi-TGS pathway is the mechanism of silencing of centromeric repeat regions in *S. pombe* (Grewal, 2010; Halic & Moazed, 2010). Here, heterochromatin formation is dependent both on siRNAs as well as on H3K9 methylation. First, dsRNA is made by bi-directional transcription of pericentromeric repeats. dsRNA is cleaved by Dicer to generate siRNAs that are then loaded onto an Argonaute protein in the RNA-induced transcriptional silencing complex, RITS. This complex recruits the RNA dependent RNA polymerase complex (RDRC), which promotes heterochromatin formation by recruiting histone methyltranferase, Clr4, and heterchromatin protein 1 (HP1). RDRC becomes anchored to chromatin via chromodomain binding to H3K9me (Halic & Moazed, 2010; Motamedi et al, 2004; Verdel &

Moazed, 2005). This pathway exists endogenously and is required for proper chromosome segregation in *S. pombe.*

In *C. elegans,* a large number of endogenous siRNAs (endo-siRNAs) have been cloned, the majority of which are antisense to protein-coding mRNAs (Ambros et al, 2003; Claycomb et al, 2009; Gu et al, 2009; Pak & Fire, 2007; Ruby et al, 2006), in contrast to those in *S. pombe* that map to pericentromeric repeat regions. One large class of endo-siRNAs, called 22G RNAs, preferentially begin with guanosine and are 22 nucleotides in length (Claycomb et al, 2009; Gu et al, 2009).

C. elegans has the most Argonaute proteins of any species known at 27, and although the function of many of these still remains unclear (Boisvert & Simard, 2008; Hutvagner & Simard, 2008), a few are known to be worm-specific and have been termed WAGOs (Gu et al, 2009). In many systems, as described above, RdRPs work in concert with Dicer to synthesize siRNAs with 5′ monophosphate residues (Lee & Collins, 2007). In nematodes, RdRPs catalyze *de novo* synthesis of siRNAs that have a 5′ triphosphate (Pak & Fire, 2007) and are directly loaded, without Dicer activity, onto Argonaute proteins (Gu et al, 2009). The majority of the WAGOs in *C. elegans* are involved in the silencing of transposons, pseudogenes and cryptic loci and generally function towards genome protection (Gu et al, 2009). In contrast, CSR-1 is bound to 22G-siRNAs that map specifically to a large number of protein-coding regions (Claycomb et al, 2009). The synthesis of endo-siRNAs in the CSR-1 22G-RNA pathway is dependent on the RdRP EGO-1 (Claycomb et al, 2009; Smardon et al, 2000) and the dicer-related helicase DRH-3 (Duchaine et al, 2006; Nakamura et al, 2007). Mutants of *ego-1*, *drh-3*, and *csr-1* have similar germline and embryo defects. Therefore, it was proposed that EGO-1, DRH-3 and CSR-1 were involved in chromosome segregation in *C. elegans* (Claycomb et al, 2009), although the effect of

siRNAs on the global architecture of the holocentric chromosomes in the worm is not clear. Work from our lab has found that the CSR-1 RNAi pathway may in fact be involved in gene regulation in somatic tissues (unpublished data), and that the chromosome segregation defects seen in the RNAi mutants are due to the depletion of histone proteins (see Chapter 3).

There have been several genome-wide screens utilizing the ability of RNAi to silence genes (Kamath & Ahringer, 2003). These aimed to discover factors affecting many different processes in *C. elegans* from neuron development (Poole et al, 2011) to longevity (Samuelson et al, 2007). Additionally, there have been several genome-wide screens to identify factors affecting RNAi in *C. elegans* in order to better understand the function of this process as a gene silencing mechanism (Dudley, 2002; Kim, 2005; Vastenhouw et al, 2006). One protein that was identified as a factor promoting both exogenous RNAi and endogenous RNAi-TGS in *C. elegans* is Zinc Finger Protein 1 (ZFP-1) (Dudley, 2002; Grishok et al, 2005; Kim et al, 2005), which is the focus of Chapter 2.

Leukemia and AF10

ZFP-1, which contains two N-terminal PHD fingers, is the ortholog of mammalian AF10 (acute lymphoblastic leukemia 1-fused gene from chromosome 10) (Chaplin et al, 1995). AF10 is best known as a fusion partner with MLL in the MLL-AF10 oncogene that results in leukemia. Acute leukemias caused by MLL fusion proteins are associated with poor prognoses and high mortality rates (Moneypenny et al, 2006; Ross, 2008; Stone, 2002). MLL fusion oncogenes are responsible for more than half of the leukemia cases in infants (Moneypenny et al, 2006; Ross, 2008) and are implicated in some instances of acute leukemia in adults (Stone, 2002). The most common fusion partners of MLL are nuclear proteins AF4, AF9, AF10, ENL and ELL (Krivtsov

& Armstrong, 2007; Mohan et al, 2010b). In the MLL-AF10 fusion, the C-terminus of AF10 is fused to the N-terminus of MLL, which results in the loss of the PHD fingers of AF10 and the methyltransferase domain of MLL. The fusion protein binds to MLL target sites via an AT hook in MLL that is retained in the chimera, while the C-terminus of AF10 recruits the histone methyltransferase DOT1L leading to misregulation of target gene expression (Mohan et al, 2010b; Okada et al, 2005). MLL-AF10 fusions account for approximately 3% of lymphoid leukemias and 15% of myeloid leukemias (Mohan et al, 2010b).

Mutation in *zfp-1* was found to affect the transcript levels of many targets, by microarray, close to 250 of which overlapped with another RNAi-promoting factor called RDE-4, a double-stranded RNA-binding protein (Grishok et al, 2008). Subsequently, *rde-4* and *zfp*-1 mutant worms were shown to have reduced lifespan and enhanced susceptibility to oxidative stress and pathogens due to the misregulation of stress-related genes (Mansisidor et al, 2011). It is important to note that previous studies on *zfp-1* in *C. elegans* have utilized the loss-of-function allele that retains the N-terminal portion of the protein. The N-terminus of AF10/ZFP-1 contains two conserved PHD fingers that are predicted to bind to chromatin and have been shown to mediate oligomerization (Forissier et al, 2007; Linder et al, 2000). We find that this N-terminal region of ZFP-1 is essential for development (see Chapter 2).

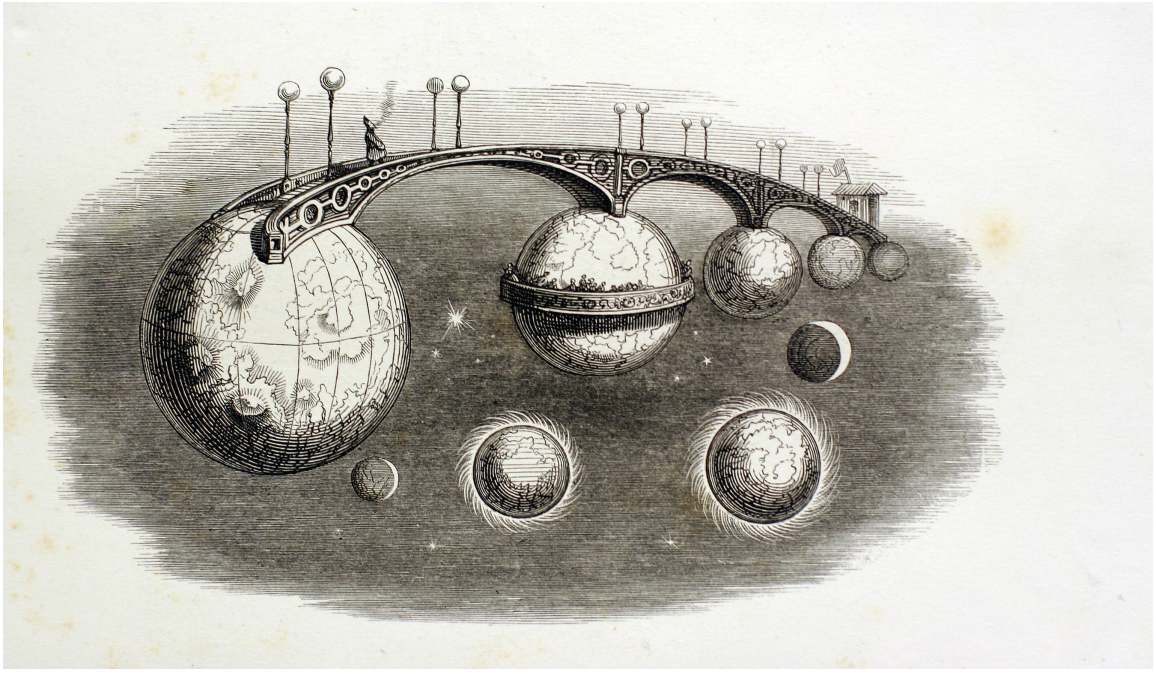


Figure 1: An artistic depiction of chromatin structure. '*Les pont des planetes*' by Grandville, J. J. reproduced from *Un Atre Monde* (H. Fournier, Paris 1844).

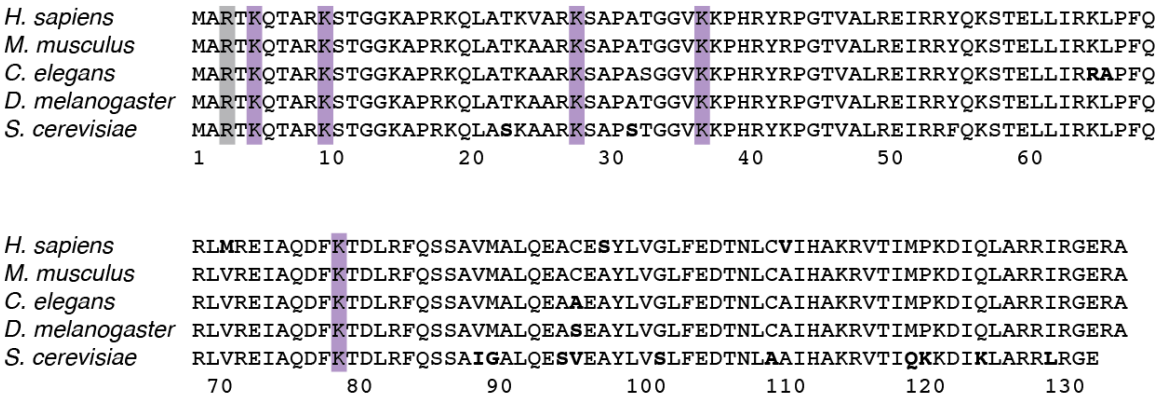


Figure 2: Alignment of histone H3 from 5 species as shown. As all histones, the protein is extremely conserved. Any varying residues are highlighted in bold. Lysine residues that are commonly methylated, K4, K9, K27, K36 and K79 are highlighted in purple. R2, whose methylation status is also important, is highlighted in grey.

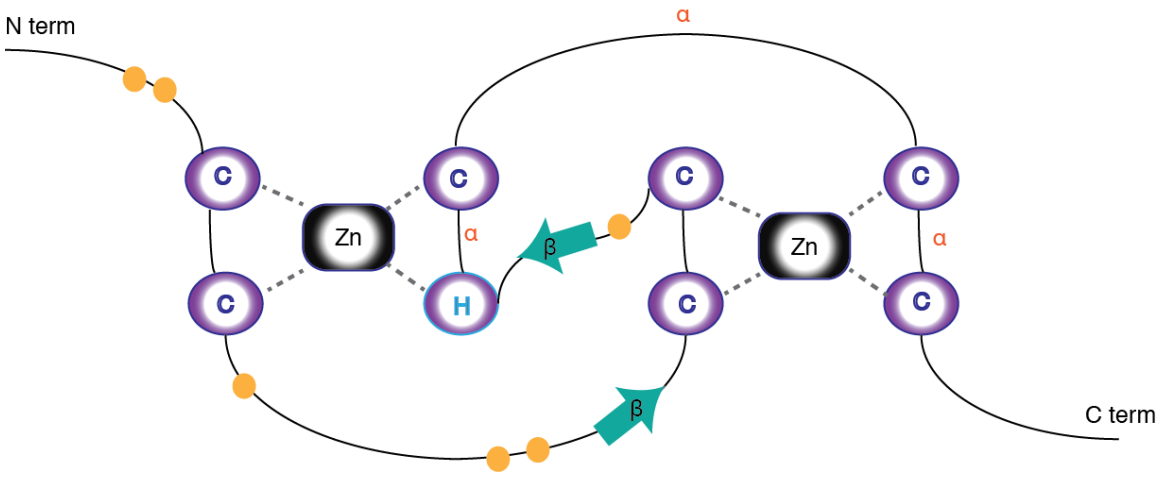


Figure 3: Schematic of the PHD finger motif showing two coordinated zinc ions within the brace position of the Cys_4-His-Cys_3 motif. Orange circles indicate regions that have been shown to be responsible for recognizing the methylation states of H3K4 in various proteins. Arrows indicate generally conserved β-sheets; helices are indicated with an α. Adapted from Sanchez & Zhou, 2011.

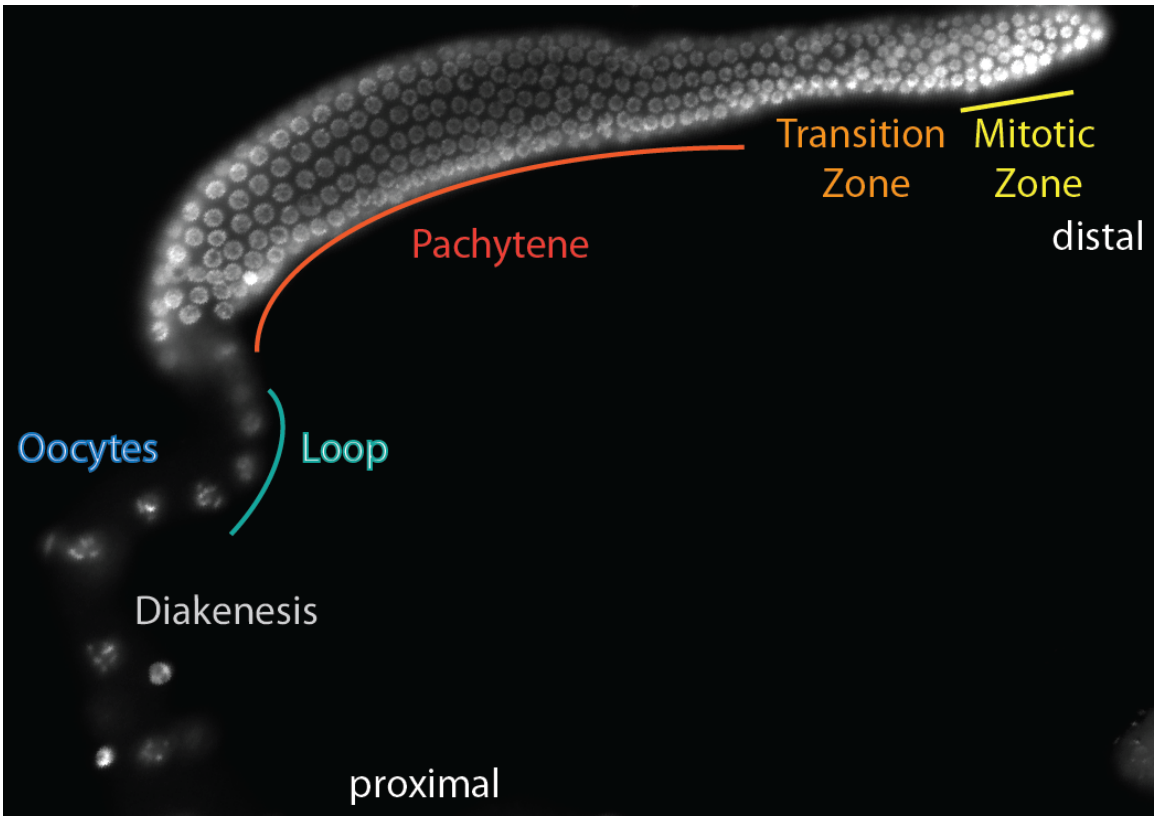


Figure 4: Dissected gonad arm from *C. elegans* adult hermaphrodite stained with DAPI to visualize DNA. The distal germline begins with the mitotic zone (yellow) where primordial germ cells are maintained. Moving proximally, the cells progress from mitosis to meiosis in what is known as the 'transition zone' (orange), which is followed by meiosis in the pachytene region (red) progressing towards oocyte development in the adult gonad. Image taken on a Zeiss AxioImager Z1 microscope at 200x total magnification, adapted from WormAtlas.

References

Allard P, Yang Q, Marzluff WF, Clarke HJ (2005) The stem-loop binding protein regulates translation of histone mRNA during mammalian oogenesis. *Dev Biol* **286:** 195-206

Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D (2003) MicroRNAs and other tiny endogenous RNAs in C. elegans. *Curr Biol* **13:** 807-818

Azimzadeh Jamalkandi S, Masoudi-Nejad A (2011) RNAi pathway integration in Caenorhabditis elegans development. *Functional & integrative genomics* **11:** 389-405

Boisvert ME, Simard MJ (2008) RNAi pathway in C. elegans: the argonautes and collaborators. *Current topics in microbiology and immunology* **320:** 21-36

Burton NO, Burkhart KB, Kennedy S (2011) Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 19683-19688

Cenik ES, Zamore PD (2011) Argonaute proteins. *Curr Biol* **21:** R446-449

Chang MJ, Wu H, Achille NJ, Reisenauer MR, Chou CW, Zeleznik-Le NJ, Hemenway CS, Zhang W (2010) Histone H3 lysine 79 methyltransferase Dot1 is required for immortalization by MLL oncogenes. *Cancer Res* **70:** 10234-10242

Chaplin T, Ayton P, Bernard OA, Saha V, Della Valle V, Hillion J, Gregorini A, Lillington D, Berger R, Young BD (1995) A novel class of zinc finger/leucine zipper genes identified from the molecular cloning of the t(10;11) translocation in acute leukemia. *Blood* **85:** 1435-1441

Claycomb JM, Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, Conte D, Jr., Mello CC (2009) The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139:** 123-134

Davila Lopez M, Samuelsson T (2007) Early evolution of histone mRNA 3' end processing. *RNA* **14:** 1-10

Dominski Z, Marzluff WF (2007) Formation of the 3′ end of histone mRNA: Getting closer to the end. *Gene* **396:** 373-390

Dominski Z, Yang XC, Purdy M, Marzluff WF (2005) Differences and similarities between Drosophila and mammalian 3' end processing of histone pre-mRNAs. *RNA* **11:** 1835-1847

Doyon Y, Cayrou C, Ullah M, Landry AJ, Cote V, Selleck W, Lane WS, Tan S, Yang XJ, Cote J (2006) ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol Cell* **21:** 51-64

Duchaine TF, Wohlschlegel JA, Kennedy S, Bei Y, Conte D, Jr., Pang K, Brownell DR, Harding S, Mitani S, Ruvkun G, Yates JR, 3rd, Mello CC (2006) Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. *Cell* **124:** 343-354

Dudley NR (2002) Using RNA interference to identify genes required for RNA interference. *Proceedings of the National Academy of Sciences* **99:** 4191-4196

Eissenberg JC, Shilatifard A (2010) Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Dev Biol* **339:** 240-249

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391:** 806-811

Forissier S, Razanajaona D, Ay A-S, Martel S, Bartholin L, Rimokh R (2007) AF10-dependent transcription is enhanced by its interaction with FLRG. *Biol Cell* **99:** 563-571

Furuhashi H, Takasaki T, Rechtsteiner A, Li T, Kimura H, Checchi PM, Strome S, Kelly WG (2010) Trans-generational epigenetic regulation of C. elegans primordial germ cells. 1-21

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorrakrai K, Agarwal A, Alexander RP, Barber G, Brdlik CM, Brennan J, Brouillet JJ, Carr A, Cheung MS, Clawson H, Contrino S, Dannenberg LO, Dernburg AF, Desai A, Dick L, Dose AC, Du J, Egelhofer T, Ercan S, Euskirchen G, Ewing B, Feingold EA, Gassmann R, Good PJ, Green P, Gullier F, Gutwein M, Guyer MS, Habegger L, Han T, Henikoff JG, Henz SR, Hinrichs A, Holster H, Hyman T, Iniguez AL, Janette J, Jensen M, Kato M, Kent WJ, Kephart E, Khivansara V, Khurana E, Kim JK, Kolasinska-Zwierz P, Lai EC, Latorre I, Leahey A, Lewis S, Lloyd P, Lochovsky L, Lowdon RF, Lubling Y, Lyne R, MacCoss M, Mackowiak SD, Mangone M, McKay S, Mecenas D, Merrihew G, Miller DM, 3rd, Muroyama A, Murray JI, Ooi SL, Pham H, Phippen T, Preston EA, Rajewsky N, Ratsch G, Rosenbaum H, Rozowsky J, Rutherford K, Ruzanov P, Sarov M, Sasidharan R, Sboner A, Scheid P, Segal E, Shin H, Shou C, Slack FJ, Slightam C, Smith R, Spencer WC, Stinson EO, Taing S, Takasaki T, Vafeados D, Voronina K, Wang G, Washington NL, Whittle CM, Wu B, Yan KK, Zeller G, Zha Z, Zhong M, Zhou X, Ahringer J, Strome S, Gunsalus KC, Micklem G, Liu XS, Reinke V, Kim SK, Hillier LW, Henikoff S, Piano F, Snyder M, Stein L, Lieb JD, Waterston RH (2010) Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330:** 1775-1787

Gorgoni B, Andrews S, Schaller A, Schumperli D, Gray NK, Muller B (2005) The stem-loop binding protein stimulates histone translation at an early step in the initiation pathway. *RNA* **11:** 1030-1042

Greer EL, Maures TJ, Hauswirth AG, Green EM, Leeman DS, Maro GS, Han S, Banko MR, Gozani O, Brunet A (2010) Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in C. elegans. *Nature* **466:** 383-387

Grewal SIS (2010) RNAi-dependent formation of heterochromatin and its diverse functions. *Current Opinion in Genetics & Development* **20:** 134-141

Grimaud C, Negre N, Cavalli G (2006) From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. *Chromosome Res* **14:** 363-375

Grishok A (2005) Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. *Genes & Development* **19:** 683-696

Grishok A, Hoersch S, Sharp PA (2008) RNA interference and retinoblastoma-related genes are required for repression of endogenous siRNA targets in Caenorhabditis elegans. *Proc Natl Acad Sci USA* **105:** 20386-20391

Grishok A, Sinskey JL, Sharp PA (2005) Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. *Genes Dev* **19:** 683-696

Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, Fire A (2012) Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet* **44:** 157-164

Gu W, Shirayama M, Conte D, Jr., Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, Chen CC, Chaves DA, Duan S, Kasschau KD, Fahlgren N, Yates JR, 3rd, Mitani S, Carrington JC, Mello CC (2009) Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. *Mol Cell* **36:** 231-244

Guang S, Bochner AF, Burkhart KB, Burton N, Pavelec DM, Kennedy S (2010) Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* **465:** 1097-1101

Halic M, Moazed D (2010) Dicer-Independent Primal RNAs Trigger RNAi and Heterochromatin Formation. *Cell* **140:** 504-516

Horn PJ, Peterson CL (2002) Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science* **297:** 1824-1827

Horton JR, Upadhyay AK, Qi HH, Zhang X, Shi Y, Cheng X (2010) Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat Struct Mol Biol* **17:** 38-43

Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9:** 22-32

Jaeger S, Barends S, Giege R, Eriani G, Martin F (2005) Expression of metazoan replication-dependent histone genes. *Biochimie* **87:** 827-834

Kamath RS, Ahringer J (2003) Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* **30:** 313-321

Keall R, Whitelaw S, Pettitt J, Müller B (2007) Histone gene expression and histone mRNA 3' end structure in Caenorhabditis elegans. *BMC Molecular Biology* **8:** 51

Ketting RF (2011) The Many Faces of RNAi. *Developmental Cell* **20:** 148-161

Khorasanizadeh S (2011) Recognition of methylated histones: new twists and variations. *Curr Opin Struct Biol* **21:** 744-749

Kim JK (2005) Functional Genomic Analysis of RNA Interference in C. elegans. *Science* **308:** 1164-1167

Kim JK, Gabel HW, Kamath RS, Tewari M, Pasquinelli A, Rual J-F, Kennedy S, Dybbs M, Bertin N, Kaplan JM, Vidal M, Ruvkun G (2005) Functional genomic analysis of RNA interference in C. elegans. *Science* **308:** 1164-1167

Kim T, Buratowski S (2009) Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* **137:** 259-272

Kodama Y, Rothman JH, Sugimoto A, Yamamoto M (2002) The stem-loop binding protein CDL-1 is required for chromosome condensation, progression of cell death and morphogenesis in Caenorhabditis elegans. *Development* **129:** 187-196

Kornberg RD, Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98:** 285-294

Krivtsov AV, Armstrong SA (2007) MLL translocations, histone modifications and leukaemia stem-cell development. *Nature reviews Cancer* **7:** 823-833

Lan F, Collins RE, De Cegli R, Alpatov R, Horton JR, Shi X, Gozani O, Cheng X, Shi Y (2007) Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448:** 718-722

Lee SR, Collins K (2007) Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat Struct Mol Biol* **14:** 604-610

Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*

Li T, Kelly WG (2011) A role for Set1/MLL-related components in epigenetic regulation of the Caenorhabditis elegans germ line. *PLoS Genet* **7:** e1001349

Linder B, Newman R, Jones LK, Debernardi S, Young BD, Freemont P, Verrijzer CP, Saha V (2000) Biochemical analyses of the AF10 protein: the extended LAP/PHD-finger mediates oligomerisation. *J Mol Biol* **299:** 369-378

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389:** 251-260

Mansisidor AR, Cecere G, Hoersch S, Jensen MB, Kawli T, Kennedy LM, Chavez V, Tan M-W, Lieb JD, Grishok A (2011) A Conserved PHD Finger Protein and Endogenous RNAi Modulate Insulin Signaling in Caenorhabditis elegans. *PLoS Genet* **7:** e1002299

Marzluff WF, Wagner EJ, Duronio RJ (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature Reviews Genetics* **9:** 843-854

Mohan M, Herz HM, Takahashi YH, Lin C, Lai KC, Zhang Y, Washburn MP, Florens L, Shilatifard A (2010a) Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes Dev* **24:** 574-589

Mohan M, Lin C, Guest E, Shilatifard A (2010b) Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nature reviews Cancer* **10:** 721-728

Moneypenny CG, Shao J, Song Y, Gallagher EP (2006) MLL rearrangements are induced by low doses of etoposide in human fetal hematopoietic stem cells. *Carcinogenesis* **27:** 874-881

Montgomery MK, Xu S, Fire A (1998) RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 15502-15507

Motamedi MR, Verdel A, Colmenares SU, Gerber SA, Gygi SP, Moazed D (2004) Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119:** 789-802

Musselman CA, Kutateladze TG (2009) PHD fingers: epigenetic effectors and potential drug targets. *Mol Interv* **9:** 314-323

Musselman CA, Kutateladze TG (2011) Handpicking epigenetic marks with PHD fingers. *Nucleic Acids Res* **39:** 9061-9071

Nakamura M, Ando R, Nakazawa T, Yudazono T, Tsutsumi N, Hatanaka N, Ohgake T, Hanaoka F, Eki T (2007) Dicer-related drh-3 gene functions in germ-line development by maintenance of chromosomal integrity in Caenorhabditis elegans. *Genes Cells* **12:** 997-1010

Nguyen AT, Zhang Y (2011) The diverse functions of Dot1 and H3K79 methylation. *Genes & Development* **25:** 1345-1358

Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM, Su L, Xu G, Zhang Y (2005) hDOT1L links histone methylation to leukemogenesis. *Cell* **121:** 167-178

Olins DE, Olins AL (2003) Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* **4:** 809-814

Oudet P, Gross-Bellard M, Chambon P (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4:** 281-300

Pak J, Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in C. elegans. *Science* **315:** 241-244

Patel A, Vought VE, Dharmarajan V, Cosgrove MS (2011) A novel non-SET domain multi-subunit methyltransferase required for sequential nucleosomal histone H3 methylation by the mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem* **286:** 3359-3369

Peña PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature*

Pettitt J, Crombie C, Schümperli D, Müller B (2002) The Caenorhabditis elegans histone hairpin-binding protein is required for core histone gene expression and is essential for embryonic and postembryonic cell division. *J Cell Sci* **115:** 857-866

Poole RJ, Bashllari E, Cochella L, Flowers EB, Hobert O (2011) A Genome-Wide RNAi Screen for Factors Involved in Neuronal Specification in Caenorhabditis elegans. *PLoS Genet* **7:** e1002109

Ptashne M (2007) On the use of the word 'epigenetic'. *Curr Biol* **17:** R233-236

Rando OJ, Chang HY (2009) Genome-wide views of chromatin structure. *Annu Rev Biochem* **78:** 245-271

Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. *Cell* **128:** 655-668

Ross JA (2008) Environmental and genetic susceptibility to MLL-defined infant leukemia. *J Natl Cancer Inst Monogr***:** 83-86

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell* **127:** 1193-1207

Samuelson AV, Klimczak RR, Thompson DB, Carr CE, Ruvkun G (2007) Identification of Caenorhabditis elegans genes regulating longevity using enhanced RNAi-sensitive strains. *Cold Spring Harb Symp Quant Biol* **72:** 489-497

Sanchez R, Marzluff WF (2002) The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro. *Mol Cell Biol* **22:** 7093-7104

Sanchez R, Zhou MM (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36:** 364-372

Schaner C (2006) Germline chromatin. *WormBook*

Schaner CE, Deshpande G, Schedl PD, Kelly WG (2003) A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. *Dev Cell* **5:** 747-757

Smardon A, Spoerke JM, Stacey SC, Klein ME, Mackin N, Maine EM (2000) EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. *Curr Biol* **10:** 169-178

Smith E, Lin C, Shilatifard A (2011) The super elongation complex (SEC) and MLL in development and disease. *Genes Dev* **25:** 661-672

Smith E, Shilatifard A (2010) The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Mol Cell* **40:** 689-701

Sönnichsen B, Koski LB, Walsh A, Marschall P, Neumann B, Brehm M, Alleaume A-M, Artelt J, Bettencourt P, Cassin E, Hewitson M, Holz C, Khan M, Lazik S, Martin C, Nitzsche B, Ruer M, Stamford J, Winzi M, Heinkel R, Röder M, Finell J, Häntsch H, Jones SJM, Jones M, Piano F, Gunsalus KC, Oegema K, Gönczy P, Coulson A, Hyman AA, Echeverri CJ (2005) Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* **434:** 462-469

Stone RM (2002) The difficult problem of acute myeloid leukemia in the older adult. *CA Cancer J Clin* **52:** 363-371

Strome S (2005) Specification of the germ line. *WormBook : the online review of C elegans biology***:** 1-10

Sullivan KD, Mullen TE, Marzluff WF, Wagner EJ (2009) Knockdown of SLBP results in nuclear retention of histone mRNA. *RNA* **15:** 459-472

Takahashi YH, Westfield GH, Oleskie AN, Trievel RC, Shilatifard A, Skiniotis G (2011) Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 20526-20531

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14:** 1025-1040

Thelie A, Pascal G, Angulo L, Perreau C, Papillier P, Dalbies-Tran R (2012) An oocyte-preferential histone mRNA stem-loop-binding protein like is expressed in several mammalian species. *Molecular reproduction and development*

Tolia NH, Joshua-Tor L (2007) Slicer and the argonautes. *Nature chemical biology* **3:** 36-43

Van Holde KE, Allen JR, Tatchell K, Weischet WO, Lohr D (1980) DNA-histone interactions in nucleosomes. *Biophysical journal* **32:** 271-282

van Steensel B (2011) Chromatin: constructing the big picture. *EMBO J* **30:** 1885-1895

Vastenhouw NL, Brunschwig K, Okihara KL, Muller F, Tijsterman M, Plasterk RH (2006) Gene expression: long-term gene silencing by RNAi. *Nature* **442:** 882

Verdel A, Moazed D (2005) RNAi-directed assembly of heterochromatin in fission yeast. *FEBS Letters*

Washington NL, Stinson EO, Perry MD, Ruzanov P, Contrino S, Smith R, Zha Z, Lyne R, Carr A, Lloyd P, Kephart E, McKay SJ, Micklem G, Stein LD, Lewis SE (2011) The modENCODE Data Coordination Center: lessons in harvesting comprehensive experimental details. *Database : the journal of biological databases and curation* **2011:** bar023

Zhang X, Wen H, Shi X (2012) Lysine methylation: beyond histones. *Acta biochimica et biophysica Sinica* **44:** 14-27

Zhang Y, Reinberg D (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev* **15:** 2343-2360

Chapter 2:

The conserved PHD1-PHD2 domain of ZFP-1/AF10 is a discrete functional module essential for viability in *C. elegans*

Daphne C. Avgousti, Germano Cecere and Alla Grishok

Adapted from Avgousti *et al*, in revision, MCB

Abstract

Plant homeodomain (PHD) type zinc fingers play an important role in recognizing chromatin modifications and recruiting regulatory proteins to specific genes. A specific module containing a conventional PHD finger followed by an extended PHD finger exists in the mammalian AF10 protein among a few others. AF10 has mostly been studied in the context of the leukemic MLL-AF10 fusion protein, which lacks the N-terminal PHD fingers of AF10. Although this domain of AF10 is the most conserved region in the protein, its biological significance has not been elucidated. Here, we use biochemical and genetic approaches to study the PHD1-PHD2 region of the *C. elegans* ortholog of AF10, zinc finger protein 1 (ZFP-1). We demonstrate that the PHD1-PHD2 region is essential for viability and this domain as a multimer *in vivo*. This region of ZFP-1 is also able to bind nucleosomes *in vitro*. The second extended PHD finger is responsible for multimerization of the protein while the first PHD finger interacts with methylated H3K4. We find that introduction of a point mutation in the N terminus of PHD1 in the context of PHD1-PHD2 alters the specificity of this domain by reducing binding to methylated and increasing binding to unmethylated H3K4. Moreover, we show that ZFP-1 localizes to H3K4 methylation-enriched promoters of actively expressed genes genome-wide, and that H3K4 methylation is important for this localization in the embryo. Finally, in mutants of the COMPASS complex, which are lacking proper H3K4me, we find ZFP-1 to be both mislocalized in the germline and specifically depleted at target gene promoters. We predict that the essential biological role of the PHD1-PHD2 module of ZFP-1/AF10 is connected to the regulation of actively expressed genes during early development.

Introduction

The role of chromatin-binding proteins in gene expression regulation and development is being increasingly recognized. While the important developmental function of a few conserved chromatin regulators, such as the Polycomb group of proteins, has been known for some time, (reviewed in Grimaud et al, 2006), there are a number of conserved proteins with predicted chromatin-binding properties whose potential role in development has yet to be revealed. Acute Lymphoblastic Leukemia 1-Fused gene from chromosome 10 (AF10) (Chaplin et al, 1995) is one such gene.

AF10 is best known as a fusion partner of Mixed Lineage Leukemia (MLL) in the MLL-AF10 oncogene (Chaplin et al, 1995). Acute leukemias caused by MLL fusion proteins are associated with poor prognoses and high mortality rates (Moneypenny et al, 2006; Ross, 2008; Stone, 2002). MLL fusion oncogenes are responsible for more than half of the leukemia cases in infants (Moneypenny et al, 2006; Ross, 2008) and are implicated in some instances of acute leukemia in adults (Stone, 2002). The most common fusion partners of MLL are nuclear proteins AF4, AF9, AF10, ENL and ELL (Krivtsov & Armstrong, 2007; Mohan et al, 2010b). In the MLL-AF10 fusion, the C-terminus of AF10 is fused to the N-terminus of MLL. MLL-AF10 fusions account for approximately 3% of lymphoid leukemias and 15% of myeloid leukemias (Mohan et al, 2010b). In the fusion protein, the C-terminus of AF10 inappropriately recruits the H3K79 methyltransferase Dot1-like (DOT1L) to MLL targets, which contributes to oncogenesis (Okada et al, 2005). AF10 and DOT1L have been recently shown to exist in the same complex, DotCom (Mohan et al, 2010a). Although the C-terminus of AF10 has been studied in the context of its interaction with DOT1L, very little is known about the biological function of the N-terminal PHD fingers of AF10, which are missing in the MLL-AF10 fusion protein. In fact, the

PHD fingers are the most highly conserved region of the protein, with 70% sequence identity between *C. elegans* and human (Figure 1C), (Chaplin et al, 1995; Saha et al, 1995).

It is clear from the primary sequence alignment that, in addition to the PHD fingers, the linker region between the two predicted PHD fingers is also highly conserved and is likely to be important for the function of the protein (Figure 1C). The linker region together with the second PHD finger is often called the 'extended PHD finger' and has been implicated in mediating oligomerization of AF10 (Forissier et al, 2007; Linder et al, 2000). Modules consisting of a canonical PHD finger directly followed by a conserved linker region and a second PHD finger, or extended PHD finger, have been described in a few other proteins, including Jade-1 (Zhou et al, 2004). The linker region between the PHD fingers of Jade-1 was shown to be important for its interaction with the von Hippel-Lindau (VHL) tumor suppressor, which is a functional component of the E3 ligase complex (Zhou et al, 2004). Thus, the extended PHD finger of AF10 has the potential to serve as a protein-protein interaction domain.

Although PHD zinc fingers were described some time ago (Schindler et al, 1993) and were recognized as distinct protein domains, PHD-containing proteins became of particular interest following the discovery of their histone tail-binding properties. PHD fingers can often discriminate between methylation states of lysine residues of histone tails, such as lysine 4 on histone H3, H3K4 (Musselman & Kutateladze, 2011; Sanchez & Zhou, 2011; Taverna et al, 2007). This ability of PHD fingers makes them important components of chromatin-modifying proteins and complexes. In some cases, they serve to recruit such complexes to specific genomic regions. For example, both PHD fingers in the Jade-1 protein were shown to interact with histone H3 (Saksouk et al, 2009) and to facilitate the recruitment of histone H4 acetyltransferase HBO1 to chromatin. In other cases, PHD fingers bind to the products of a modification reaction

at specific loci within the genome. For example, the PHD finger of BHC80, a member of the lysine-specific demethylase (LSD1) complex (Lan et al, 2007), binds specifically to the unmethylated H3K4 that results from demethylation by LSD1.

Here we investigate the chromatin-binding properties and biological function of the *C. elegans* ortholog of AF10, ZFP-1, and specifically its N-terminal PHD fingers. We demonstrate that the highly conserved N-terminal region of ZFP-1, which contains the PHD fingers and is retained in the *zfp-1(ok554)* deletion mutant, is essential for viability. This conclusion is based on the fact that only transgenes containing the PHD1-PHD2 region of the protein can rescue the lethality caused by the loss of one copy of *zfp-1(ok554)*. Additionally, we find that the extended PHD2 is responsible for the multimerization of the protein both *in vitro* and *in vivo,* in agreement with previous studies on AF10 (Forissier et al, 2007; Linder et al, 2000), and that PHD1 confers chromatin-binding specificity of the protein by preferentially interacting with methylated H3K4. We find that PHD1-PHD2 also binds nucleosomes *in vitro* and that a mutation in PHD1 can alter its specificity for H3K4me. Finally, we show that methylated H3K4 contributes to the localization of ZFP-1 to the promoters of its target genes enriched in H3K4 methylation. Together, our data suggest that the conserved PHD1-PHD2 region of ZFP-1, found in the long isoform preferentially expressed in the germline and embryos, contributes to the regulation of ZFP-1 target genes essential for fertility and early development. We find that nearly half (46%) of all ZFP-1 target genes are also enriched in H3K4me2/3 at the promoters. This proportion increases to 83% when considering only germline-specific ZFP-1 targets. We predict that these genes are likely to be regulated by ZFP-1 in the germline and early embryo, and that their abnormal expression leads to the lethality seen in animals that lose one copy of *zfp-1(ok554)*.

Results

The long isoform of ZFP-1 is specifically expressed in the germline

In *C. elegans*, there are two isoforms of the ZFP-1 protein predicted by the current genome assembly (WS228): a long isoform that encodes the full-length ZFP-1 protein and a short isoform that is missing the N-terminal PHD fingers (Figure 1A and 1B). The *zfp-1* locus is complex, and predicted transcripts include those encoding *C. elegans* AF10 ortholog (F54F2.2a and F54F2.2c) and a transcript encoding an unrelated protein (F54F2.2b). In the existing deletion mutant for the *zfp-1* gene, *zfp-1*(*ok554*) (Cui et al, 2006), exon 4 is spliced to exon 7, as confirmed by RT-PCR (Figure 1A), leading to a frame shift. The resulting 239 amino acid protein product is predicted to retain the first 189 amino acids, including both N-terminal PHD fingers (Figure 1B) (Cui et al, 2006). A protein-protein interaction between AF10 and DOT1L occurs through the conserved leucine zipper (LZ) in the C-terminal portion of AF10 (Okada et al, 2005). This interaction is predicted to be conserved in *C. elegans* but would not occur in *zfp-1*(*ok554*) expressing the truncated ZFP-1 protein, which is missing the relevant LZ. The *ok554* mutant is a loss-of-function allele that displays slow growth and protruded vulva phenotypes (Cui et al, 2006). Also, *zfp-1*(*ok554*) has a reduced lifespan (Mansisidor et al, 2011; Oh et al, 2006) and enhanced susceptibility to oxidative stress and pathogens (Mansisidor et al, 2011). These phenotypes of the *zfp-1(ok554)* mutant are likely to be associated with the lack of the interaction between ZFP-1 and a *C. elegans* Dot1 family methyltransferase.

In order to investigate the N-terminal region of ZFP-1, we raised an antibody that recognizes a peptide sequence in the linker region between the two N-terminal PHD fingers and thus only the long isoform (Figure 1C, blue box). This antibody was not suitable for Western blotting. However, we detected expression of the truncated long ZFP-1 isoform in the *zfp-*

1(ok554) mutant by immunostaining dissected gonads (Figure 2A panel 1). This signal was diminished upon *zfp-1(RNAi)* (Figure 2A, panel 2), indicating that it is due to the residual protein expressed in the *zfp-1(ok554)* mutant. We also used an anti-ZFP-1 C-terminal antibody (Mansisidor et al, 2011) to detect the full-length ZFP-1 by the same method (Figure 2B, panel 1). The immunostaining with both antibodies showed a distinct pattern in maturing oocyte nuclei in the germline (Figure 2A, 2B: panels 1, arrows). This pattern is missing in *zfp-1(ok554)* mutants when staining with the C-terminal specific antibody, confirming that there is no C-terminus of ZFP-1 present in *zfp-1(ok554)* (Figure 2B, panel 2).

We used homologous recombination in bacteria to introduce a GFP tag at the C-terminus of ZFP-1 on the 35kb fosmid (Mansisidor et al, 2011). This construct retained all regulatory sequences present at the *zfp-1* locus and we observed nuclear expression of ZFP-1::GFP at all developmental stages of *C. elegans* (Figure 3). Consistent with immunostaining, ZFP-1::GFP localization is seen at the condensed chromosomes in oocytes (Figure 3). The ZFP-1::GFP protein binds to all six condensed chromosomes.

Our immunostaining data indicate that the long isoform of ZFP-1, full-length or truncated, is expressed in the germline. However, the C-terminal antibody recognizes both isoforms of the protein (Figure 4), and the GFP tag, visualized on chromosomes in oocytes (Figure 3), is present on the C-terminus of both ZFP-1 isoforms. To clarify whether one or both ZFP-1 isoforms are present in the germline we constructed another transgenic strain, *zfp-1(ok554)*; [armEx14], in which we introduced a FLAG tag downstream of the PHD fingers followed by a stop codon, leaving the endogenous short isoform intact on the fosmid as it has an alternate promoter (armEx14 (ZFP-1::PHD::FLAG, ZFP-1 short) Figure 2C, schematic). We then carried out immunostaining with either an anti-FLAG antibody to visualize the PHD

fingers, or the anti-ZFP-1 C-terminal antibody (Mansisidor et al, 2011) to visualize the short isoform (Figure 2C). We found that the PHD fingers are expressed in the germline (Figure 2C, panel 1) while the short isoform is not (Figure 2C, panel 2). The failure to detect expression of the short ZFP-1 isoform in the germline is not due to poor sensitivity of the antibody since we readily detected germline ZFP-1 expression using this antibody in wild type worms (Figure 2B, panel 1). The signal from the anti-FLAG antibody is also specific because immunostaining of wild type worms with this antibody produced no signal (not shown). This signal also confirms expression of the armEx14 transgene in the germline. In addition, we detected production of the short ZFP-1 isoform from the armEx14 fosmid by Western blotting performed on larva extracts (Figure 4). Our immunostaining results are consistent with the expression patterns of the short and long isoforms of ZFP-1 determined by Western, where the long isoform is predominantly expressed in embryos and in adult worms with developed germline tissue and the short isoform is more abundant at the larval stages (Figure 4). These experiments indicate that only the long ZFP-1 isoform is expressed in the germline and localizes to chromosomes there. Moreover, the N-terminal portion of ZFP-1 retains the localization properties of the full-length protein. Taken together, these immunostaining results indicate that the long isoform of ZFP-1, or its truncated version as in *zfp-1(ok554)*, is the only isoform expressed in the germline.

In order to determine the functional significance of ZFP-1 expression in the germline we performed RNAi experiments. The knockdown of the remaining portion of ZFP-1 in *zfp-1(ok554)* by RNAi led to low penetrant germline defects (not shown). In order to enhance the efficiency of RNAi we performed *zfp-1(RNAi)* experiments in the sensitized *eri-1(mg366)* genetic background (Kennedy et al, 2004). We used a strain expressing H2B::GFP in the germline to visualize chromosomes (Vastenhouw et al, 2006). The sterility phenotype was

indeed enhanced with 25% (5/20) of the worms exhibiting oogenesis defects and no embryos produced in either one or both gonad arms (Figure 5A, panel 1 and 5B), as compared to control worms with no defects. We also observed proximal germline tumors (Figure 5A, panel 2) indicative of the expansion of mitotic cells in place of normal oocytes undergoing meiosis. In *zfp-1(ok554)*, the PHD fingers are retained and the worms do not have germline defects. In the knockdown of *zfp-1* by RNAi in a sensitized background, however, we do see germline defects. Therefore, the protein containing the PHD1-PHD2 region, whether in the long isoform in wild type worms or in the truncated version in *zfp-1(ok554)*, has a role in promoting normal germline development and fertility.

ZFP-1 PHD fingers are essential for viability

In order to further analyze the biological role of the PHD1-PHD2 module of the ZFP-1 protein we performed experiments aimed at reducing its zygotic function. The *zfp-1(ok554)* mutant worms, in which the N-terminal portion of the long ZFP-1 protein is retained, are short-lived (Mansisidor et al, 2011; Oh et al, 2006) and stress-sensitive (Mansisidor et al, 2011). These phenotypes are linked to the misregulation of genes during larval development and specifically to the increased expression of the PDK-1 kinase (Mansisidor et al, 2011). In order to reduce *zfp-1* function, we decided to put the *zfp-1(ok554)* allele in trans over a chromosome with a deletion that includes the *zfp-1* locus. We took advantage of the deficiency allele on chromosome III, nDf17 (Ellis et al, 1991), which is balanced by chromosome qC1[*dpy-19(e1259) glp-1(q339)*] (Edgley et al, 2006). First, we separated the nDf17 and qC1 chromosomes by crossing the nDf17/qC1 hermaphrodites with males containing chromosome III marked with GFP (rhIs4) (Lim et al, 1999), (Figure 6A). The resulting cross progeny males

either carried the deficiency chromosome in trans to the GFP-marked chromosome III or the qC1 chromosome in trans to the GFP marker (green F1 Figure 6A). These animals were phenotypically indistinguishable. Next, we crossed these F1 males with *zfp-1(ok554)* hermaphrodites and analyzed their progeny (F2 in Figure 6A) by counting the number of GFP(+) and GFP(-) cross progeny. For crosses with qC1/GFP males, we expected an equal number of GFP(+): *zfp-1(ok554)/*GFP and GFP(-): *zfp-1(ok554)/*qC1 animals. By crossing nDf17/GFP males with *zfp-1(ok554)* hermaphrodites, we investigated whether the phenotype of nDf17/*zfp-1(ok554)* cross progeny would be more severe than that of *zfp-1(ok554)* (Figure 6A). We expected an equal number of GFP(+): *zfp-1(ok554)/*GFP and GFP(-): *zfp-1(ok554)/*nDf17 animals if reducing the dosage of *zfp-1(ok554)* did not affect development, or fewer GFP(-) animals if the *zfp-1(ok554)/*nDf17 genotype caused lethality. We determined the genotype of the F1 males used for the crosses with *zfp-1(ok554)* by the presence or absence of the qC1 phenotypes (Dumpy and Sterile) (Edgley et al, 2006) among the F3 generation (Figure 6A). We found that there were far fewer GFP(-) worms among the cross progeny of nDf17/GFP males (Figure 6C, row 1), indicating that *zfp-1(ok554)*/nDf17 animals did not survive. The few surviving GFP(-) worms were healthy and likely represent recombinants (Figure 6C, asterisks). We observed arrested deformed larvae that likely represent *zfp-1(ok554)*/nDf17 animals. We conclude that a single copy of *zfp-1(ok554)* is not sufficient for viability.

To verify that the lethality of *zfp-1(ok554)*/nDf17 is indeed due to a lack of sufficient amounts of ZFP-1, we repeated the crosses using various transgenic strains made in the *zfp-1(ok554)* background (Figure 6B). We found that transgenes containing the long isoform of ZFP-1, either in the context of a fosmid (armEx5(ZFP-1::GFP)) or a plasmid (pLong(ZFP-1L::FLAG::GFP)) were able to rescue the lethality of *zfp-1(ok554)*/nDf17 in the F2 generation

(Figure 6C, rows 2 and 3). However, a transgene containing only the short isoform (pShort(ZFP-1S::FLAG::GFP)) without the N-terminal PHD fingers, failed to rescue (Figure 6C, row 4). Importantly, the fosmid-based armEx14 transgene expressing the PHD1-PHD2 module in addition to the short ZFP-1 isoform (Figure 2C, Figure 4, and Figure 6B), but not the long isoform, was also able to rescue the *zfp-1(ok554)*/nDf17 lethality (Figure 6C, row 5). Together, these results demonstrate that the most conserved N-terminal region of ZFP-1, which contains the PHD fingers, is sufficient for viability.

The experiments described above indicate a zygotic requirement for PHD1-PHD2, since *zfp-1(ok554)/zfp-1(ok554)* mothers of *zfp-1(ok554)*/nDf17 worms do not have fertility problems. However, our RNAi experiments indicate that PHD1-PHD2 may also have an essential function in the germline. The transgenic strains that rescued *zfp-1(ok554)*/nDf17 lethality differ in that the fosmid-based armEx5(ZFP-1::GFP) and armEx14 (ZFP-1::PHD::FLAG, ZFP-1) were made by the bombardment technique, allowing germline expression, while the plasmid-based strain pLong(ZFP-1L::FLAG::GFP) was made by DNA injection and did not express full-length ZFP-1 in the germline. Therefore, the surviving *zfp-1(ok554)*/nDf17; [pLong(ZFP-1L::FLAG::GFP)] adults should be rescued in the soma, but not in the germline. Indeed, these animals did not produce progeny (Figure 4A star, Figure 4C right column, star). Notably, *zfp-1(ok554)*/nDf17 animals also carrying fosmid-based transgenes were fertile, although their brood size was lower compared to *zfp-1(ok554)/zfp-1(ok554)*. This data indicates that germline expression of the rescuing transgenes is required for rescue in the following generation. The somatically expressed pLong transgene rescues zygotically for one generation but rescued worms do not produce progeny. In contrast, fosmids expressed in the soma and germline (armEx5 and armEx14) rescue the first generation, thus allowing these worms to produce progeny, which

indicates a maternal and zygotic rescue. Therefore, we conclude that PHD1-PHD2 is needed both maternally and zygotically for normal fertility and development in *C. elegans*.

Extended PHD finger of ZFP-1 is responsible for multimerization of the protein *in vitro* and *in vivo*

To further investigate the properties of the PHD fingers of ZFP-1, we expressed the recombinant protein corresponding to the N-terminal region of ZFP-1 (PHD1-PHD2) tagged with cleavable MBP (Maltose Binding Protein) in *E. coli*. First, we affinity-purified PHD1-PHD2::MBP using amylose beads, and then we cleaved the MBP tag with TEV protease. We used analytical size exclusion chromatography to separate PHD1-PHD2 from the MBP tag and observed that the molecular weight of the PHD1-PHD2 protein (28kDa) appeared larger than that of MBP (45kDa) (Figure 7A, top). We confirmed that the peaks seen by gel filtration corresponded to either PHD1-PHD2 or MBP by SDS-PAGE (Figure 7A, bottom). Consistent with previous studies on AF10 (Forissier et al, 2007; Linder et al, 2000), we found that ZFP-1 is likely to exist as a multimer *in vitro.* However, these results do not reflect the shape of ZFP-1 and cannot exclude the possibility than an odd shape of this protein may contribute to the higher apparent molecular weight. Previous studies carried out cross-linking experiments showing results consistent with tetramerization (Linder et al, 2000); our findings agree.

Next, we constructed a homology model of PHD1-PHD2 of ZFP-1 to better understand the region in between the two PHD fingers (Figure 7B). This model predicts a likely metal ion-binding site in the linker region (blue region, pink highlights) in addition to the zinc ion binding sites within the predicted PHD1 (green) and PHD2 (orange) (Figure 1C, residues with asterisk, Figure 7B in pink). The linker region together with PHD2 has been shown to mediate the

multimerization of AF10 *in vitro* (Linder et al, 2000). To confirm this for ZFP-1, we used analytical size exclusion chromatography for PHD2 and the linker region (extended PHD2, Figure 7B grey dashed box), without PHD1, and calibrated the column with known standards (Figure 7C). We found that the linker together with PHD2 of ZFP-1 has an apparent weight of four times its actual size (14kDa) (Figure 7C). Therefore, we conclude that this region is likely to be solely responsible for the apparent multimerization of the protein.

To confirm the multimerization of ZFP-1 through the N-terminal PHD region *in vivo,* we took advantage of the armEx14 transgenic line expressing PHD1-PHD2 tagged with FLAG. Expression of PHD1-PHD2::FLAG rescued the lethality of *zfp-1(ok554)/*nDf17, as described earlier (Figure 4C, row 5). We introduced the PHD1-PHD2::FLAG transgene into the wild type background, immunoprecipitated the tagged protein with an anti-FLAG antibody, and used an antibody recognizing the N-terminal region of ZFP-1 generated by modENCODE (see Materials and Methods) for Western blot analysis (Figure 5D). We observed a band corresponding to PHD1-PHD2::FLAG at approximately 28kDa as well as a large band corresponding to the full-length endogenous ZFP-1 protein at 95kDa (Figure 5D). This suggests that the PHD1-PHD2::FLAG protein interacts with the long isoform of ZFP-1 through the PHD1-PHD2 region. Immunoprecipitation of ZFP-1::GFP with anti-GFP antibodies, followed by Western blotting with anti-ZFP-1 N-terminal antibodies (modENCODE), showed that ZFP-1::GFP interacts with the full length endogenous protein (Figure 7F). Thus, the multimerization properties *in vivo* indeed apply to the entire protein, consistent with the extended PHD2 being responsible for self-binding. Overall, our results indicate that the N-terminal region of ZFP-1, and specifically the linker region and PHD2 (extended PHD), is responsible for the multimerization of the protein both *in vitro* and *in vivo*.

To further understand the biological significance of the PHD1-PHD2 region, we tested its interaction with whole nucleosomes. We reconstituted nucleosomes in collaboration with the Greene Lab (Visnapuu & Greene, 2009) and carried out gel-shift assays with increasing concentrations if PHD1-PHD2. Using a DNA sequence with high affinity nucleosomes to reconstitute nucleosomes around histone octamers (Visnapuu & Greene, 2009), we observed a gel shift with increasing concentration of PHD1-PHD2 (Figure 7E). Interestingly, these nucleosomes do not have modified histone tails, which indicates that the PHD1-PHD2 region itself is capable of binding to naked nucleosomes *in vitro*. If this interaction takes place *in vivo*, and given that the PHD1-PHD2 domain forms tetramers, we can speculate that this region is involved in higher-order chromatin organization at specific loci in the genome. Together with our genetic experiments described earlier, these findings demonstrate that PHD1-PHD2 represents a discrete protein module with an essential developmental function.

Histone H3 Lysine 4 methylation contributes to the localization of ZFP-1 to chromatin

The first PHD finger of ZFP-1 represents a conventional PHD finger (Zhou et al, 2004) similar to PHD fingers found in chromatin-binding proteins (Figure 8A). A large number of PHD fingers have been shown to interact with the N-terminal tails of histone H3 (Musselman & Kutateladze, 2011; Sanchez & Zhou, 2011; Taverna et al, 2007). Interestingly, some PHD fingers are specific to methylated H3K4 (Li et al, 2006; Pena et al, 2006; Shi et al, 2006; Wysocka et al, 2006), while others, such as the PHD finger of BHC80, a subunit of H3K4 demethylase LSD-1, interact with the unmethylated H3K4 (Lan et al, 2007). To investigate whether the PHD1 of ZFP-1 interacts with histone tails, we expressed recombinant PHD1::GST and carried out binding assays with biotinylated histone tails. We found that PHD1::GST

specifically binds to H3K4me2 under the same conditions that BHC80-GST specifically interacts with unmethylated histone H3K4 (Lan et al, 2007) (Figure 8B). We also found that PHD2 alone binds to histone tails less specifically (Figure 8C), which may represent the situation *in vivo*. The ability of PHD2 to bind histone tails nonspecifically may contribute to the binding of PHD1-PHD2 to unmodified nucleosomes *in vitro* (Figure 7E). Nevertheless, in the context of PHD1-PHD2, the specificity of PHD1 allows the protein to preferentially bind methylated H3K4 (Figure 8B, panel 3).

The N-terminus of the BHC80 PHD finger has been shown to directly bind the unmethylated lysine 4 on histone H3 through a charge interaction with an aspartic acid residue, D4 (Lan et al, 2007). We observed a similarity between the N-terminus of PHD1 of ZFP-1/AF10, which has a KEM motif, and the N-terminus if BHC80 PHD finger, which has an HED motif (Figure 8A). We hypothesized that the methionine residue in ZFP-1 PHD1 interacts with methyl groups on H3K4. If this was the case, a mutation from methionine to aspartic acid could alter the specificity of ZFP-1 PHD1. Indeed, we found that the M4D mutation in the context of PHD1-PHD2 affects the histone-binding specificity of this domain, preventing binding to H3K4me3. Instead it promotes a stronger interaction with unmethylated H3K4 compared to wild type PHD1-PHD2 (Figure 8B, compare panels 3 and 4). Interestingly, this mutation does not affect the ability of PHD1-PHD2 to bind any other histone tails (data not shown); this is consistent with a model where PHD2 binds non-specifically to nucleosomes and PHD1 confers specificity.

H3K4 methylation at promoters correlates with gene activity (reviewed in (Campos & Reinberg, 2009; Gardner et al, 2011; Rando & Chang, 2009)). ZFP-1 also shows localization to promoters (Mansisidor et al, 2011) and binds to methylated H3K4 *in vitro,* as shown above. To

confirm that the *in vitro* binding specificity of PHD1-PHD2 to methylated H3K4 is relevant *in vivo*, we compared the list of ZFP-1 target genes generated by ChIP/chip (Mansisidor et al, 2011) with genes that have an H3K4me2/3 peak at their promoters (Gu & Fire, 2010). We found that nearly half of all ZFP-1 peaks corresponded to the annotated peaks of H3K4me2/3 (1174 out of 2565, p value = 8.275E-267, Figure 8C, top), indicating that this overlap is highly significant. In contrast, ZFP-1 target genes were significantly underrepresented at H3K9me-rich regions (overlap of 140 out of 2565, p = 3.5E-90, Figure 8C, bottom). Of the germline-specific targets of ZFP-1, 83% (73/88) are also enriched in H3K4 methylation compared to 46% genome wide (see Appendix I for gene lists). This suggests that the role of ZFP-1 binding to H3K4me enriched targets is more relevant for germline genes. Considering our genetic data indicating an essential germline function of PHD1-PHD2, it is possible that this function is directly linked to H3K4me.

To further investigate the role of H3K4me with ZFP-1 in the germline, we took advantage of existing mutants of members of the COMPASS complex, which is responsible for methylation of H3K4 (Krogan et al, 2002; Li & Kelly, 2011; Poole et al, 2011; Simonet et al, 2007). *C. elegans* mutants in COMPASS complex component homologs of SET-2 (*set-2(ok952)*) and WDR5 (*wdr-5.1(ok1417)*) have been shown to be deficient in H3K4me (Simonet et al, 2007). However, other homologs had not been well characterized. We obtained mutant alleles of ASH-2 (*ash-2(tm1905)*) and RbBP5 (*lsy-15(tm3463)*) (Poole et al, 2011), and found them to be severely depleted of H3K4me3, both globally (Figure 9B) and in the germline (Figure 9A). Upon staining with anti-ZFP-1 C-terminal antibodies (Mansisidor et al, 2011), we found visibly less ZFP-1 in the gonads of these mutants (Figure 9C). To determine whether this depletion was due to low levels of ZFP-1 protein or due to poor localization leading to a more diffuse signal, we performed Western blot analysis on total worm lysates of the COMPASS mutants adults

(Figure 9D). Since the total ZFP-1 protein and specifically the germline-expressed long ZFP-1 isoform were not depleted in the COMPASS mutant adults, we posit that the apparent reduction in the immunostaining signal is due to mislocalization of the ZFP-1 protein. Furthermore, we find that in *glp-4(bn2)* temperature-sensitive mutant adults grown at the restrictive temperature of 15°C, and therefore missing germline tissue (Beanan & Strome, 1992), the amount of ZFP-1 long isoform seen by Western is significantly diminished (Figure 9E). This suggests that the majority of the ZFP-1 long isoform signal seen by Western (Figure 4 and Figure 9D) comes from the germline tissue. This data increases the likelihood that ZFP-1 is mislocalized, rather than depleted, in the germline of COMPASS mutant adults.

To test whether the localization of ZFP-1 to the promoters of its target genes is dependent on the PHD1 finger specificity for H3K4me, we performed ChIP-qPCR on integrated transgenic lines that contain either the long isoform of ZFP-1 with the PHD finger region (pLong) or the short isoform lacking this region (pShort) (see Figure 6B for schematic). We found that in embryos, the long isoform is enriched at the promoter of a known ZFP-1 target gene, *pdk-1* (Mansisidor et al, 2011), but not at a non-target gene, *lys-7* (Mansisidor et al, 2011), (Figure 7A). We also confirmed the same pattern at several other promoter and coding regions of known ZFP-1 ChIP/chip targets, such as *daf-16* and *egl-30*. The result was that the long isoform of ZFP-1 containing the PHD fingers is enriched at the promoters but not at the coding regions (Figure 10A). Also, the short ZFP-1 isoform does not localize to target promoters in the embryo, although both long and short isoforms are expressed at this stage (Figure 10B), with the long isoform being predominant (Figure 4). Conversely, at the L4 larval stage, both the long and short isoforms are enriched at the target promoters (Figure 10C), suggesting that the short

isoform has other mechanisms for chromatin localization in the larva, where it is predominantly expressed (Figure 4).

To test whether the localization of ZFP-1 to target gene promoters in the embryo was dependent on H3K4me, we repeated the ZFP-1 ChIP-qPCR experiment in the *lsy-15(tm3463)* mutant background. The *ash-2(tm1905)* mutant worms show severe developmental problems and could not be grown in sufficient quantities for ChIP experiments. We confirmed that the severe depletion of H3K4me3 in *lsy-15(tm3463)* seen by Western and immunostaining (Figure 9A and B) is also specific to the tested target genes (Figure 10D). We found that in the *lsy-15(tm3463)* background, the enrichment of ZFP-1 at the *pdk-1, daf-16* and *egl-30* promoters was lost, indicating that H3K4 methylation contributes to its chromatin localization (Figure 10E).

Discussion

Our genetic and biochemical analyses have defined the conserved PHD1-PHD2 module of ZFP-1/AF10 as a multimerizing, chromatin-binding domain with a discrete function essential for viability. The conservation of this region between the human AF10 protein and *C. elegans* ZFP-1 was noted in the initial study describing the translocation of AF10 in acute myeloid leukemias (Chaplin et al, 1995) and thereafter (Saha et al, 1995). However, the biological significance of this conserved module has not been revealed until now.

The AF10 protein function has been mostly studied in the context of its oncogenic variants MLL-AF10 (Krivtsov & Armstrong, 2007; Mohan et al, 2010b) and CALM-AF10 (Caudell & Aplan, 2008). The oncogenic potential of the MLL-AF10 fusion protein is linked to the recruitment of DOT1L H3K79 methyltransferase by the leucine zipper and octapeptide motifs of AF10 present in the C-terminal portion of the protein (Okada et al, 2005). This leads to inappropriately high levels of H3K79 methylation and overexpression of MLL-AF10 target genes, such as MEIS1 and HoxA9, which significantly contribute to the oncogenic process (Krivtsov & Armstrong, 2007; Mohan et al, 2010b). However, a normal copy of the AF10 gene is also present in cancer cells. Since endogenous AF10 exists in complex with DOT1L and contributes to the normal levels of global H3K79 methylation (Mohan et al, 2010a), oncogenic fusion forms are likely to compete with wild type AF10 for interaction with DOT1L, which has been shown for CALM-AF10 (Lin et al, 2009). This may cause aberrant expression of genes normally regulated by AF10, further contributing to cancer progression. Our study suggests that the PHD1-PHD2 module within the ZFP-1/AF10 protein possesses an autonomous function, which is likely to consist of binding chromatin and recruiting other factors necessary for gene

expression regulation. Therefore, the PHD1-PHD2 and leucine zipper may recruit distinct partners and the balance of these interactions is likely to affect normal cell function.

Tissue-specific expression analyses of human AF10 detected mRNA expression in peripheral blood lymphocytes, thymus, ovary and testes (Chaplin et al, 1995). The murine ortholog of AF10 is also expressed in the brain and kidneys, as seen by northern blot (Linder et al, 1998), and expression in the white matter of the cerebellum was determined by *in situ* hybridization (Linder et al, 1998). To date, there have been no studies addressing the developmental role of AF10 using genetic knockdown techniques in mice. AF10 was found to co-immunoprecipitate with β-catenin in the HEK293T cell line (Mohan et al, 2010a) and with Tcf4 and β-catenin in mouse intestinal crypt cells (Mahmoudi et al, 2010). Since these interactors mediate downstream transcriptional effects of the Wnt signaling pathway, siRNA knockdown experiments of AF10 in tissue culture demonstrated that many downstream targets of Wnt signaling had reduced levels of expression in both AF10-depleted colorectal cancer cells an in HEK293T cells (Mahmoudi et al, 2010). Similarly, lower expression of Wingless targets was obtained after reducing expression of the *Drosophila* ortholog of AF10 (*Alhambra*) by RNAi in fly larvae (Mohan et al, 2010a) and after morpholino-depletion of AF10 in zebrafish (Mahmoudi et al, 2010). Although the potential involvement of ZFP-1/AF10 in regulation of Wnt targets may significantly contribute to the developmental role of this protein, it is important to note that the above studies connected both AF10 and DOT1L to Wnt signaling. Our discovery of the essential developmental function of the PHD1-PHD2 module suggests that it is independent of the role of ZFP-1/AF10 in the recruitment of DOT1L to chromatin. Also, ZFP-1 expression is much broader than the expression of β-catenin homologs in *C. elegans* (Eisenmann, 2005), indicating that its targets are not limited to those that are regulated by Wnt pathways.

Although transcription from the AF10 locus has not been studied extensively in human or mouse, expression of at least two major isoforms was previously detected in flies (Bahri et al, 2001; Linder et al, 2001; Perrin et al, 2003), similarly to our findings in *C. elegans*. Consistent with our results, high expression of the PHD-containing long isoform was noted in the oocytes and embryos (Bahri et al, 2001; Linder et al, 2001), and this isoform was also found to associate with euchromatin on polytene chromosomes of salivary glands (Perrin et al, 2003). The existing loss-of-function mutants of *Alhambra* (also called *Dalf* and dAF10) appear to affect expression of the C-terminal portion of both isoforms of the protein (Bahri et al, 2001; Perrin & Dura, 2004). However, these mutant alleles may still allow expression of the PHD1-PHD2 module. Nonetheless, although *Drosophila Alhambra* mutant larvae are very retarded in growth and soon die (Bahri et al, 2001), the analogous *C. elegans* mutant *zfp-1(ok554)*, where only PHD1-PHD2 function is retained, is slow-growing but viable (Cui et al, 2006; Mansisidor et al, 2011). Therefore, it appears that the function of the C-terminal portion of AF10 that interacts with DOT1L is more essential for development in flies than nematodes. We predict that the conserved PHD1-PHD2 module may have an additional essential function both in flies and mammals.

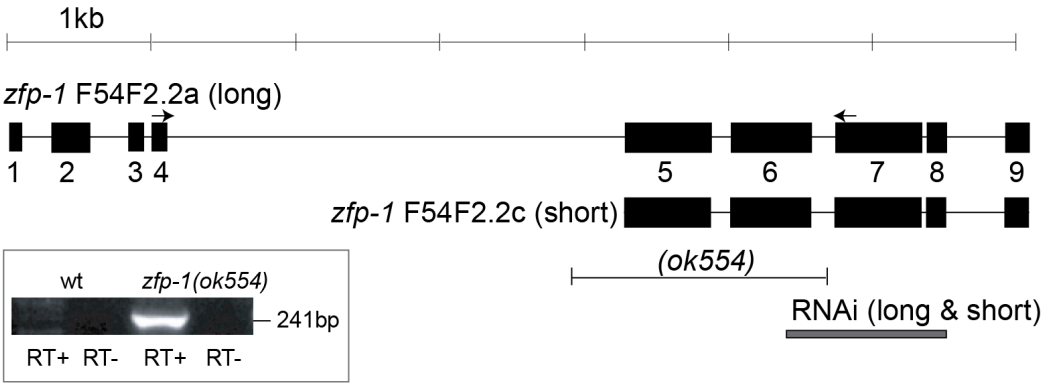
Since the long isoform of ZFP-1 is exclusively expressed in the adult germline and is predominantly expressed in the embryo, we predict that its localization to the H3K4 methylation-enriched promoters contributes to regulation of genes essential for the early development of nematodes. We have shown that PHD1 contributes to the interaction with H3K4me and that extended PHD2 is likely responsible for the multimerization of the protein. Although DOT1L methyltransferase is a known interacting partner of AF10 (Mohan et al, 2010a; Okada et al, 2005), it interacts with the C-terminal region of the protein. Since PHD1-PHD2 is essential for

viability, we predict that additional interacting partners of ZFP-1 regulate ZFP-1 targets during early development.

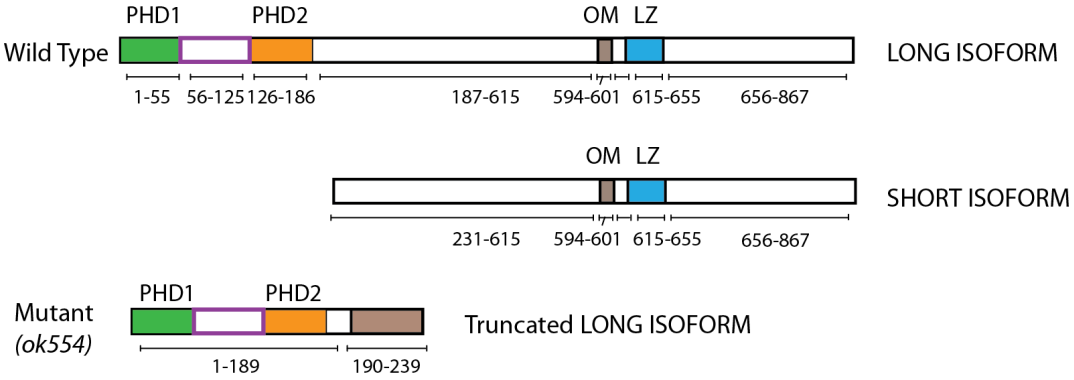
In additon, 83% of germline-specific genes targeted by ZFP-1 are also enriched in H3K4 methylation at the promoters (see Appendix I Table 1 for gene list). This suggests that the interaction of ZFP-1 with H3K4me at target gene promoters may be more relevant for germline specific targets, consistent with the essential role of the PHD fingers in the germline.

Our work has highlighted the structural autonomy of the PHD1-PHD2 domain. Its extended PHD2 finger possesses multimerization properties that are rare among PHD fingers. Future structural studies will determine the precise details of residues in the extended PHD2 responsible for the multimerization of the protein and residues determining the specificity of PHD1 to H3K4me. Possibilities for the connection between the structure and function of this domain are intriguing. We can speculate that PHD1-PHD2 may recruit effector proteins and regulate gene expression; or perhaps it is the case that the histone binding of PHD1-PHD2 contributes to proper gene regulation by conferring an autonomous role for this domain in configuring higher-order chromatin structure. Future work on the structure of the entire PHD1-PHD2 region would help to elucidate the biological functionality of this portion of the protein. See Chapter 4 for an in-depth discussion.

A *zfp-1* gene locus,chromosome III



B ZFP-1 Protein isoforms



C

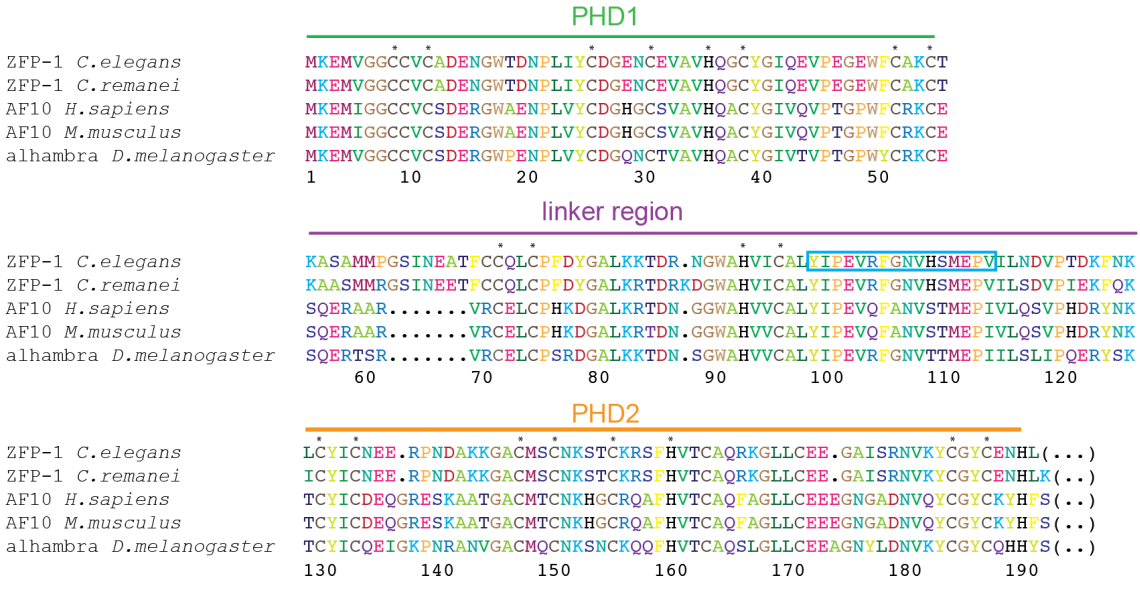


Figure 1. ZFP-1 contains two N-terminal PHD fingers that are highly conserved. (A) The ZFP-1 locus showing its two predicted isoforms and the deletion mutant *ok554*. The long isoform, F54F2.2a, contains all 9 exons whereas the short isoform, F54F2.2c, contains only exons 5-9. Arrows indicate primers used to detect exon 4 to exon 7 splicing in the *zfp-1(ok554)* mRNA, which results in a 241bp RT-PCR product, as shown. All RNAi experiments done in this study target the 3' mRNA region as indicated, leading to reduction in both long and short isoforms. (B) The 867 amino acid ZFP-1 protein has two predicted plant homeodomains (PHD) in the N-terminus (amino acids 1-189) and an octapeptide motif and leucine zipper (OMLZ, amino acids 594-655) in the C-terminus. The short isoform protein is missing the N-terminal PHD fingers but retains the OMLZ in the C-terminus. The mutant allele *ok554* expresses a truncated version of the long isoform, which retains the first 4 exons (amino acids 1-190), but leads to a frame shift that expresses an additonal 49 amino acid fragment out of frame and a premature stop codon. (C) Primary sequence alignment of the PHD1-PHD2 region of *C. elegans* ZFP-1 with homologs in four other species as indicated. Asterisks indicate the cysteine and histidine residues that are predicted to coordinate zinc ions. The blue box indicates the peptide used to raise an N-terminus specific antibody to ZFP-1.

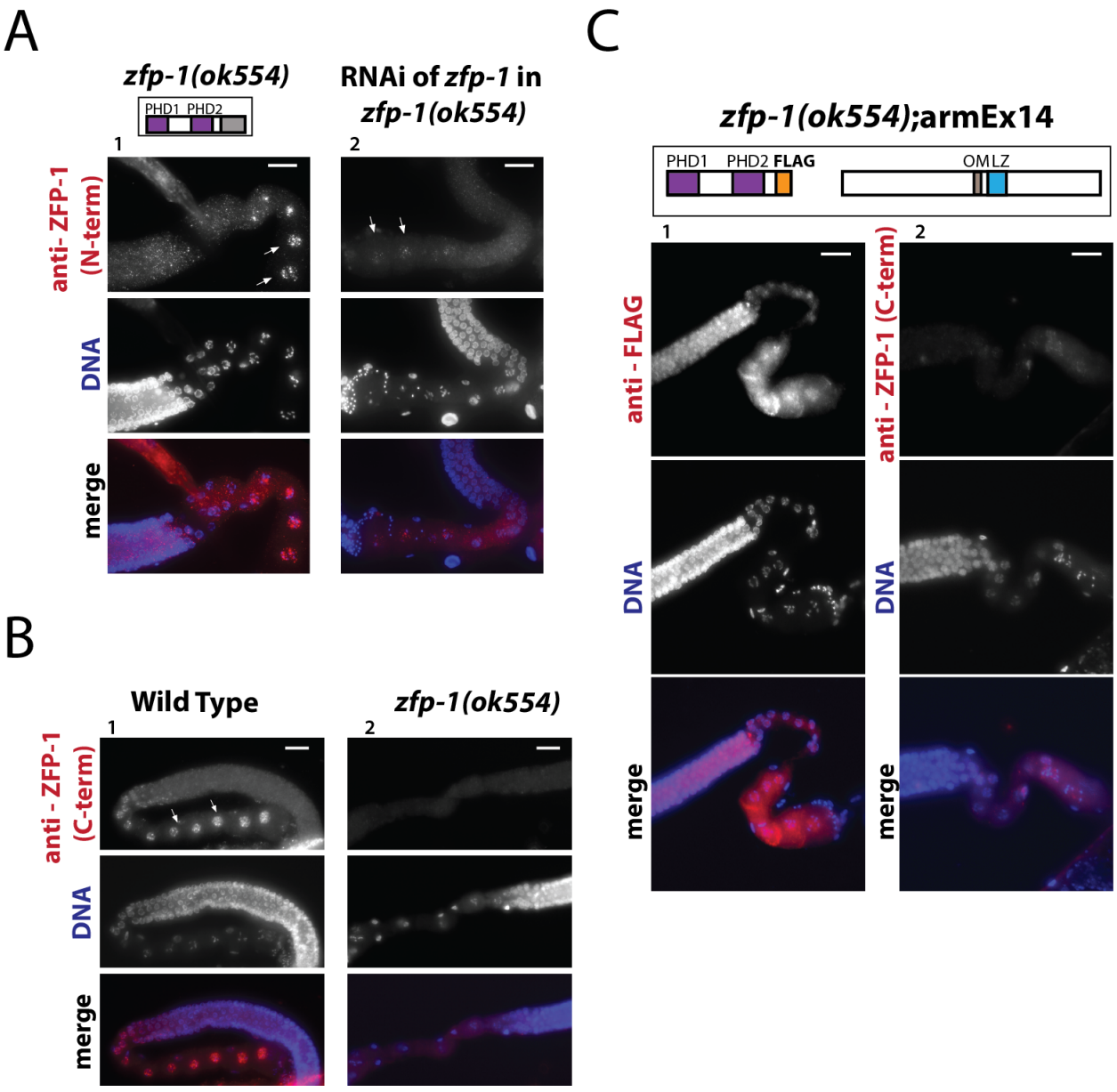


Figure 2. The long isoform of ZFP-1 is specifically expressed in the germline. (A) Immunostaining of dissected gonads from adult worms with anti-ZFP-1 N-terminus-specific antibody (pink) and DNA (blue, DAPI). Panel 1 shows the pattern of ZFP-1 localization to oocyte nuclei (arrows) while panel 2 shows ZFP-1 depletion after knockdown by RNAi in *zfp-1(ok554)*. (B) Immunostaining with anti-ZFP-1 C-terminus-specific antibody in wild type worms

compared to *zfp-1(ok554)*. The C-terminal antibody shows the same pattern as the N-terminal antibody in wild type worms, but no signal in *zfp-1(ok554)*. (C) Immunostaining in transgenic worms expressing the truncated long isoform with a FLAG tag as well as the short isoform in the *zfp-1(ok554)* background. Panel 1 shows anti-FLAG while panel 2 shows anti-C-terminal ZFP-1 staining, indicating that the truncated long isoform is expressed in the germline but the short isoform is not. All scale bars are 20μm.

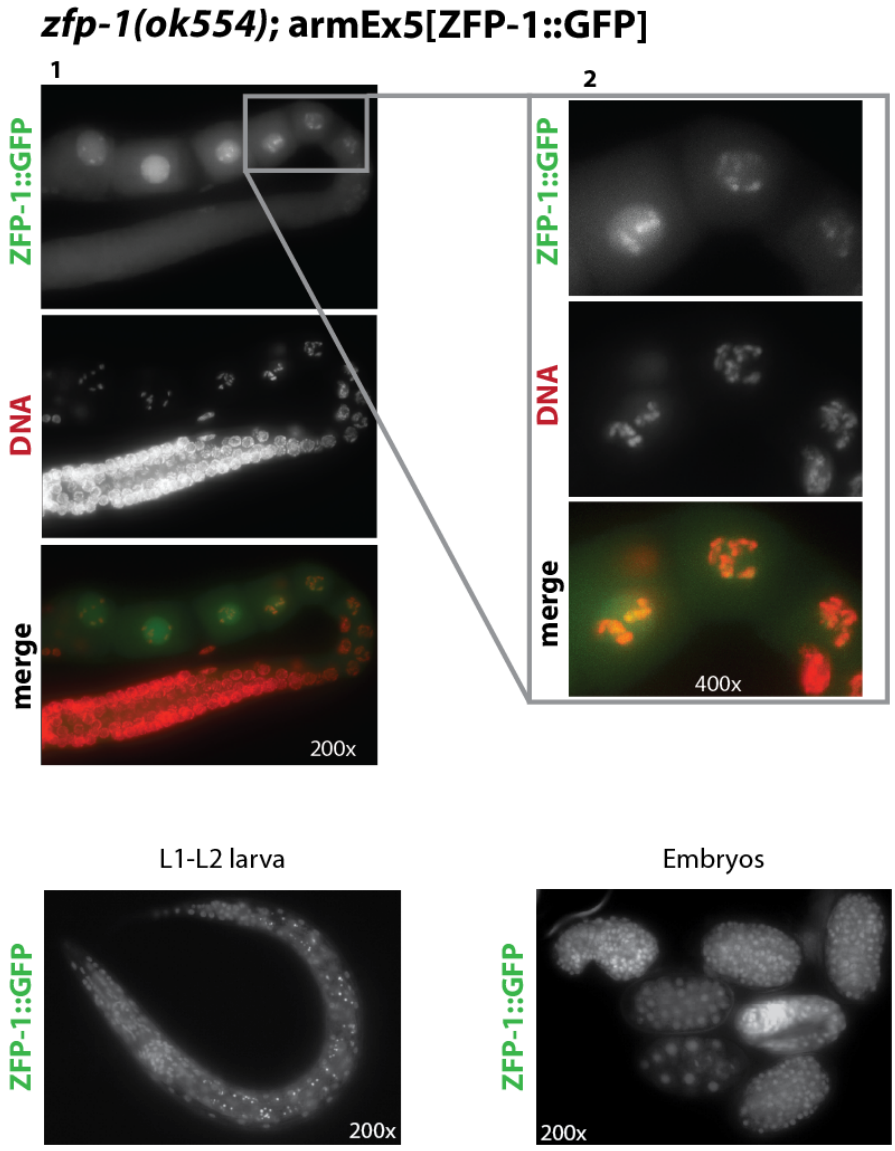


Figure 3: ZFP-1::GFP expression at different developmental stages in *C. elegans* and localization to chromosomes in maturing oocytes. Dissected adult gonad is shown with ZFP-1::GFP (green) and DNA (red, DAPI). Zoom of oocytes is shown to visualize the localization to condensed chromosomes at this stage. ZFP-1::GFP has been observed on all 6 chromosomes. Mixed stage embryos and L1-L2 larva are shown to represent ZFP-1::GFP expression at all developmental stages (Mansisidor et al, 2011).

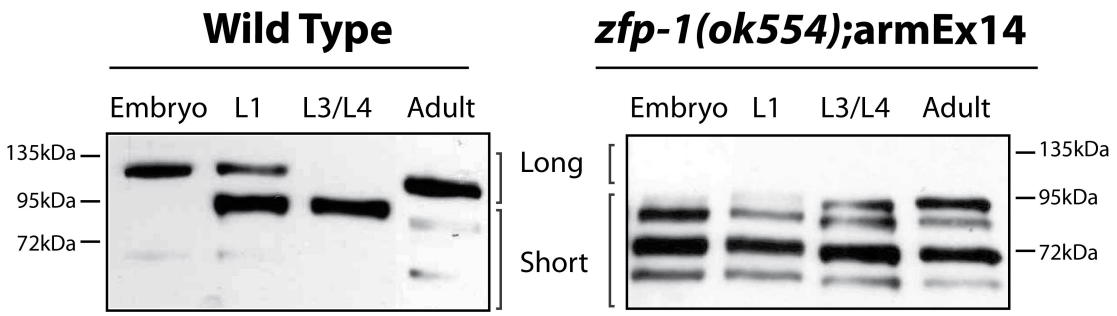
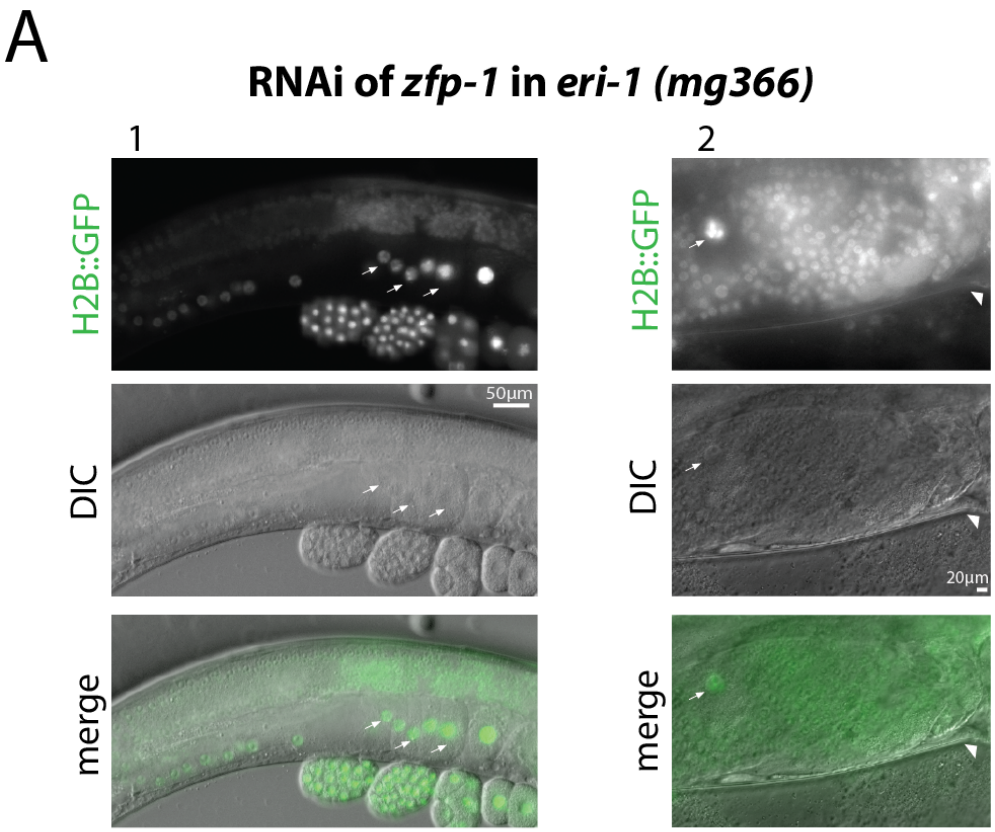


Figure 4: The long isoform of ZFP-1 is predominately expressed in embryos and adult worms. Western blotting experiments, left: antibodies specific to the C-terminus of ZFP-1 (Mansisidor et al, 2011) recognize both isoforms of ZFP-1 in multiple stages, as indicated. In the embryo and adult stages, the long isoform is primarily expressed, in contrast to the larval stages where the short isoform is predominant. The top band represents the long isoform (approximately 95kDa), as indicated, and the short isoform (approximately 65kDa) is represented by the lower bands. It is apparent from this Western blot that there are multiple bands smaller than 95 kDa. In order to investigate whether these bands represent truncations of the long isoform or post-translational modifications of the short isoform, we repeated this experimend in the *zfp-1(ok554);*armEx14 transgenic line (see Figure 2C for schematic), which expresses the N-terminal PHD fingers and the short isoform, but no long isoform. The SDS-PAGE (right) has been run out longer leading to better separation of the bands in addition to a longer exposure of the Western blot to visualize all of the bands. It is clear that there is no long isoform band and that the multiple smaller bands are likely due to post-translational modifications or splice variants of the short isoform.


A
RNAi of zfp-1 in eri-1 (mg366)
1
2
H2B::GFP
DIC
merge
50µm
20µm

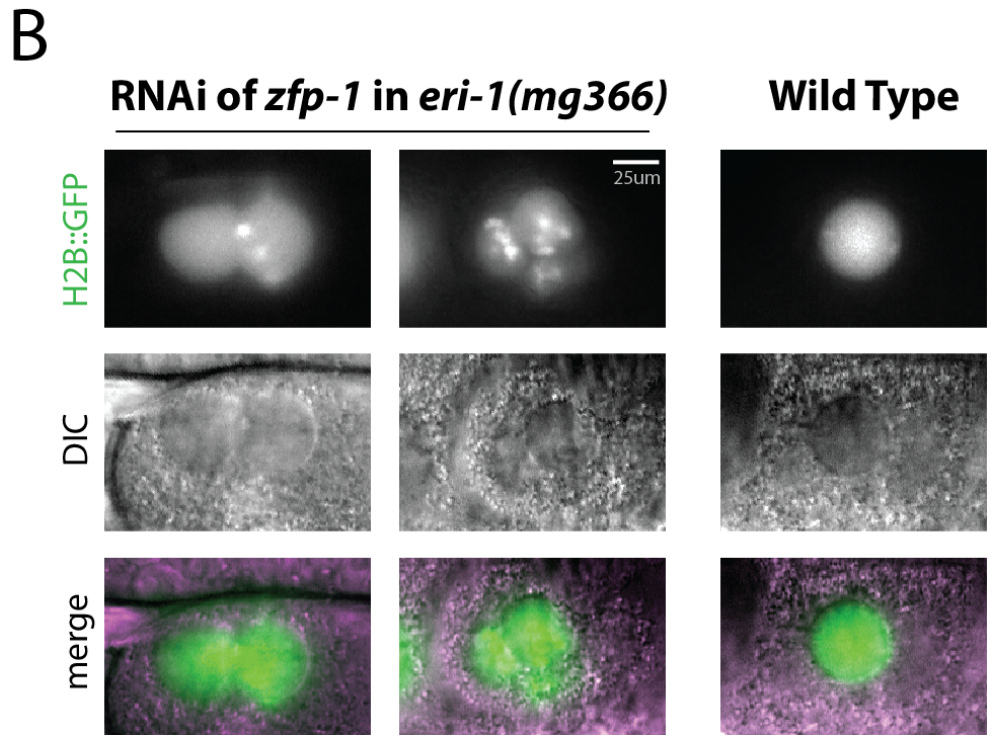
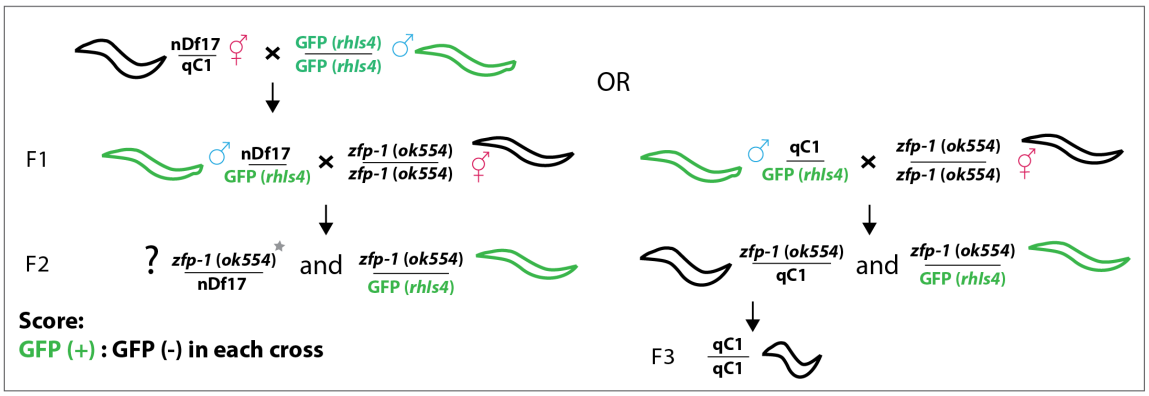
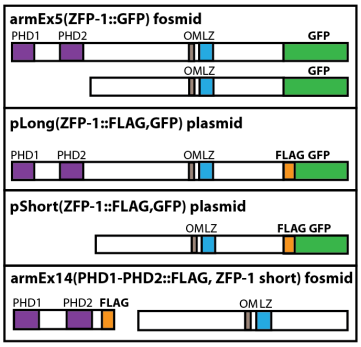

B
RNAi of zfp-1 in eri-1(mg366)
Wild Type
H2B::GFP
DIC
merge
25um

Figure 5. Knockdown of ZFP-1 by RNAi in *eri-1(mg366);* [*pie-1*::GFP::H2B] leads to germline and oogenesis defects. (A) Panel 1 shows malformed oocytes, indicated by arrows, in one gonad arm. Embryos are produced from a second, healthy, gonad arm (not shown). Panel 2 shows a proximal germline tumor in the gonad in place of developing oocytes. Arrowheads point to the vuvla. Chromatin is visualized by H2B::GFP expressed from a construct driven by the *pie-1* promoter (Vastenhouw et al, 2006). (B) Two examples of deformed oocytes are shown; in one case two nuclei have failed to separate (left) and in the other case three nuclei are merged (right). H2B::GFP (Praitis et al, 2001) is used to visualize the nuclei. A wild type nucleus is shown for reference; scale bar is 25 μm.

A



B



C

		qC1 crosses		nDf17 crosses			
		GFP (rhIs4)	non GFP	GFP (rhIs4)	non GFP	Rescue?	F3 Progeny?
1	*zfp-1(ok554)*	29	28	21	2*	N/A	N/A
2	*zfp-1(ok554)*;armEx5	27	25	15	22	+	+
3	*zfp-1(ok554)*;pLONG	33	29	16	10	+	-
4	*zfp-1(ok554)*;pSHORT	26	25	22	3*	-	N/A
5	*zfp-1(ok554)*;armEx14	24	21	27	21	+	+

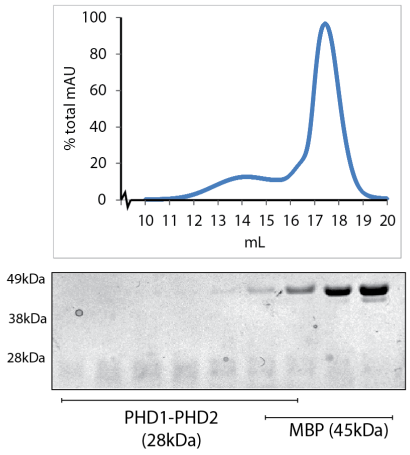
Figure 6. ZFP-1 PHD fingers are essential for viability. (A) Diagram of crossing *zfp-1(ok554)* to a deficiency allele, nDf17. nDf17/qC1 hermaphrodites are crossed to males with chromosome III GFP marker, rhIs4. F1 males are then crossed to *zfp-1(ok554)* hermaphrodites and the F2 progeny are scored for GFP(+) and GFP(-). The phenotype of F1 males is determined by whether the F3 progeny contain qC1/qC1 homozygous worms that exhibit the dumpy phenotype. (B) Various transgenic arrays used to rescue the lethality caused by *zfp-1(ok554)*/nDf17. The armEx5 fosmid contains a C-terminal GFP tag on both ZFP-1 isoforms. The pLong and pShort plasmids express each respective isoform with C-terminal FLAG and GFP tags. The armEx14 fosmid, as shown in Figure 2C, expresses a truncated long isoform with a C-terminal FLAG tag as well as the short isoform. (C) Numbers of GFP(+) and GFP(-) F2 cross-progeny for crosses

with and without the trangenes in (B). qC1 and nDf17 crosses as determined by the F3 progeny are distinguished here. Transgenes containing the N-terminal PHD fingers are able to rescue the lethality caused by *zfp-1(ok554)/*nDf17 but the short isoform alone is not. Rescued F3 *zfp-1(ok554)*/nDf17, as indicated by the grey star in (A), were able to produce progeny by fosmid (germline expression), but not plasmid rescue (no germline expression). Numbers shown represent at least 5 crosses. Asterisks indicate that these worms are likely to be recombinants.

Not shown on schematic: *elt-2*::GFP on the X chromosome was present in nDf17/qC1 hermaphrodites and their male progeny for aiding in identification of cross progeny with *zfp-1(ok554)*.

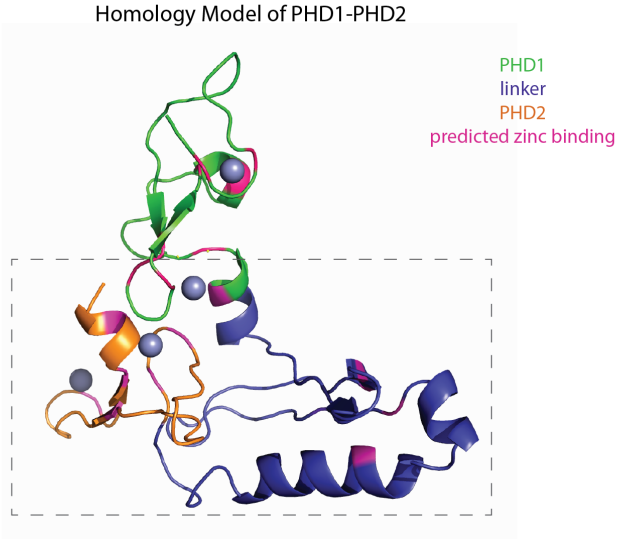
A

Gel Filtration Purification of PHD1-PHD2



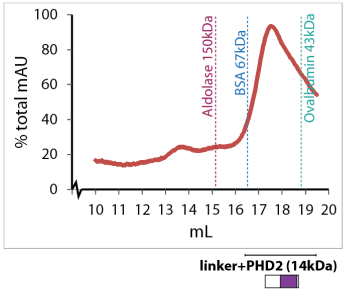
B

Homology Model of PHD1-PHD2



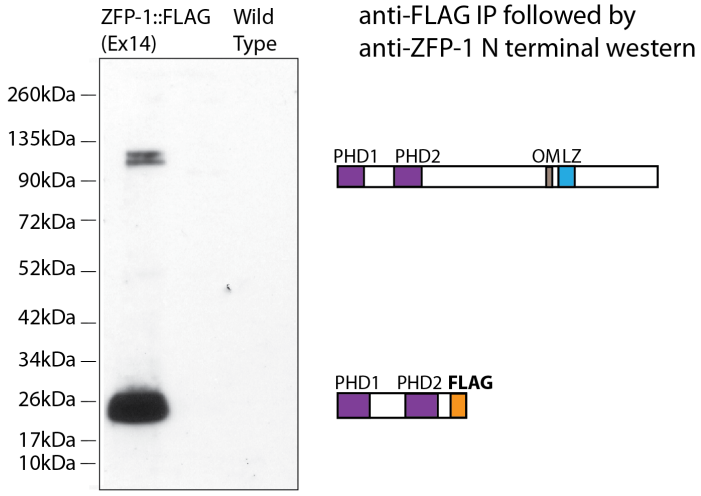
C

Gel Filtration Purification of PHD2+linker



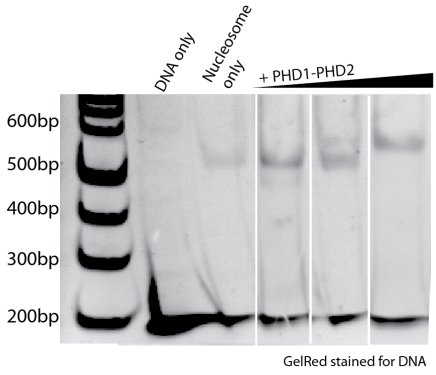
D

PHD1-PHD2 co-IP with full length endogenous ZFP-1



E

Native Gel of Nucleosome Gel Shift Assay



F

ZFP-1::GFP co-IP with full length endogenous ZFP-1

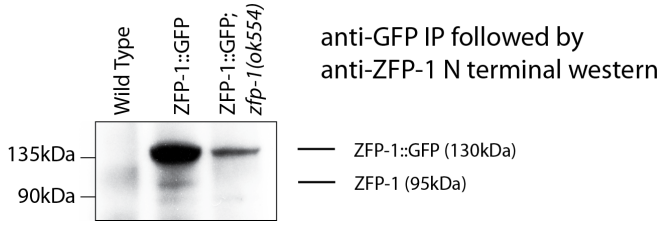


Figure 7. The extended PHD finger of ZFP-1 is responsible for multimerization both *in vitro* and *in vivo*. (A) Analytical size exclusion chromatography separating PHD1-PHD2 of ZFP-1 from maltose binding protein (MBP) and showing that the apparent molecular weight of the 28kDa PHD1-PHD2 is larger than that of the 45kDa MBP. We confirmed that the peaks correspond to expected protein sizes by SDS-PAGE and were not part of the void volume.
(B) Homology model of PHD1-PHD2. PHD1 is highlighted in green, PHD2 in orange and the linker region in blue, as indicated. Predicted zinc ion binding residues are highlighted in pink with the zinc ions represented as grey spheres. Grey dashed box indicates the linker region with PHD2, or extended PHD2. (C) Analytical size exclusion chromatography of the extended PHD2, grey box in (B). The apparent molecular weight of this 14kDa protein is between the known standards of BSA (67 kDa) and ovalbumin (43kDa), which were used to calibrate the column. This indicates that the extended PHD2 is likely to be a tetramer. We confirmed that the peak seen is due to the extended PHD2 protein by SDS-PAGE and was not part of the void volume (not shown). (D) Immunoprecipitation of PHD1-PHD2::FLAG in armEx14 present in the wild type background (schematic in Figure 6C) with anti-FLAG antibody, followed by Western blotting against the N-terminal region of ZFP-1 (antibody from modENCODE). Two bands corresponding to PHD1-PHD2::FLAG (28kDa) expressed from the transgene and the endogenous long isoform (95kDa) are apparent, indicating that the two proteins interact *in vivo*. (E) PHD1-PHD2 binds to nucleosomes *in vitro*. GelRed-stained native gel of nucleosomes incubated with increasing concentrations of PHD1-PHD2. 'DNA only' indicates the 200bp sequence used to reconstitute the nucleosomes from histone octamers and 'nucleosome only' shows the band corresponding to a nucleosome monomer (Visnapuu & Greene, 2009). Increasing concentrations of PHD1-PHD2 leads to a gel-shift showing an interaction between

PHD1-PHD2 and the nucleosome monomer. (F) Immunoprecipitation of ZFP-1::GFP in the wild type or *zfp-1(ok554)* background, as indicated, with anti-GFP antibody, followed by Western blotting against the N-terminal region of ZFP-1 (antibody from modENCODE). Two bands corresponding to ZFP-1::GFP (130kDa) expressed from the transgene and the endogenous long isoform (95kDa) are apparent in the wild type background, but not *zfp-1(ok554)* background. This indicates that the full-length ZFP-1::GFP interacts with the endogenous full length ZFP-1, which is not present in *zfp-1(ok554)*.

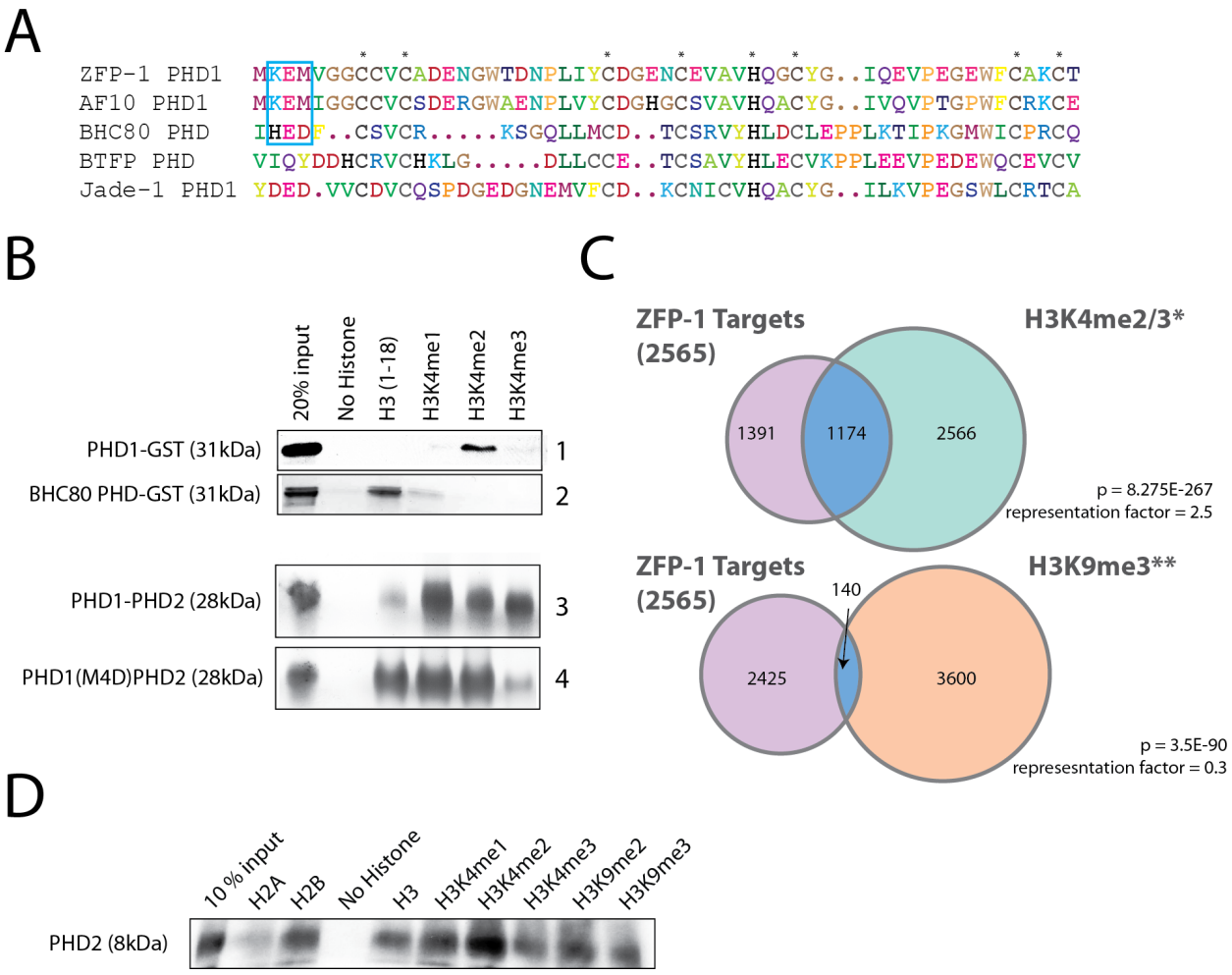


Figure 8. PHD1 of ZFP-1 specifically binds H3K4me. (A) Primary sequence alignment of ZFP-1 PHD1 with the PHD fingers of indicated proteins. Asterisks indicate conserved cysteine and histidine residues that are predicted to coordinate zinc ions. The blue box indicates motif responsible for interaction with unmethylated H3K4 (Lan et al, 2007). (B) *In vitro* binding assay of PHD1-GST with biotinylated histone H3 tails, as indicated. PHD1-GST specifically binds H3K4me2, while BHC80 PHD-GST binds unmethylated H3K4, consistent with previous findings (Lan et al, 2007), as shown by coomassie staining after SDS-PAGE (panels 1 and 2). PHD1-PHD2 preferentially binds all states of methylated H3K4 as seen by silver staining after SDS-PAGE (panel 3). M4D mutation in PHD1 in the context of PHD1-PHD2 causes a shift in

specificity from binding mono-, di- and trimethylated H3K4 but not unmethylated H3K4, to binding mono-, di- and unmethylated H3K4 but not trimethylated H3K4 (panel 4). (C) Venn diagram showing overlap of ZFP-1 ChIP/chip targets described in (Mansisidor et al, 2011) with H3K4me2/3 peaks from Gu & Fire, 2010 or H3K9me3 from Gu et al, 2012. 1174 out of 2565 ZFP-1 bound genomic regions also have H3K4me2/3 peaks, which represents a significant enrichment (top); however, only 140 out of 2565 ZFP-1 targets also have H3K9me3, representing a significant depletion, see Appendix I for gene lists. (D) PHD2 alone binds all histone tails without specificity. This is not due to a charge interaction because at the physiological pH of 7.4, at which the binding assays were carried out, PHD2 alone has a charge of +3.1.

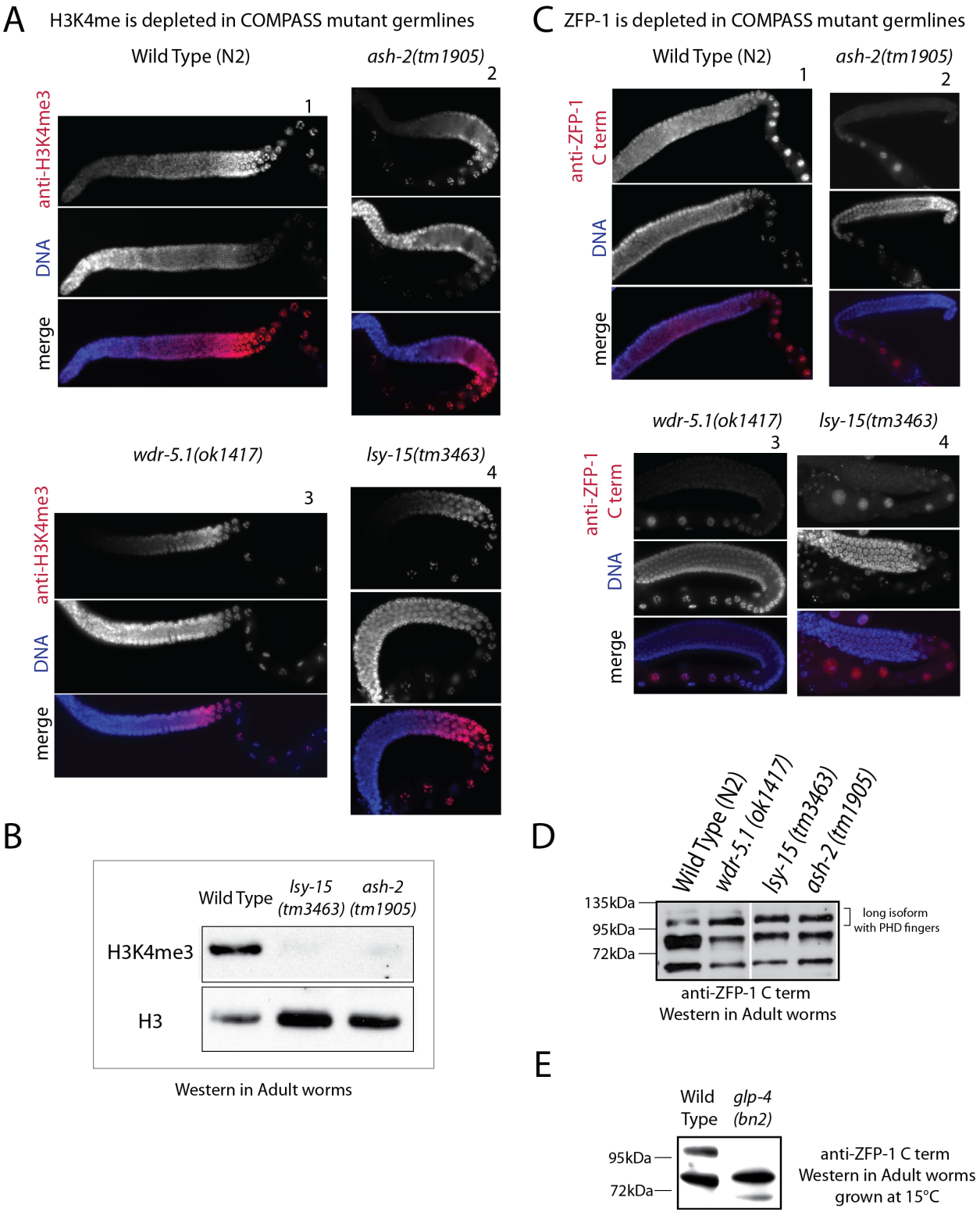

A
H3K4me is depleted in COMPASS mutant germlines
Wild Type (N2)
ash-2(tm1905)
wdr-5.1(ok1417)
lsy-15(tm3463)
1
2
3
4
anti-H3K4me3
DNA
merge
B
Wild Type
lsy-15 (tm3463)
ash-2 (tm1905)
H3K4me3
H3
Western in Adult worms
C
ZFP-1 is depleted in COMPASS mutant germlines
Wild Type (N2)
ash-2(tm1905)
wdr-5.1(ok1417)
lsy-15(tm3463)
anti-ZFP-1 C term
DNA
merge
D
Wild Type (N2)
wdr-5.1 (ok1417)
lsy-15 (tm3463)
ash-2 (tm1905)
135kDa
95kDa
72kDa
long isoform with PHD fingers
anti-ZFP-1 C term
Western in Adult worms
E
Wild Type
glp-4 (bn2)
95kDa
72kDa
anti-ZFP-1 C term
Western in Adult worms
grown at 15°C

Figure 9. COMPASS mutants have depleted ZFP-1 localization in the germline.

(A) H3K4me3 is depleted in the germlines of COMPASS mutants. Immunostaining of dissected adult gonads with anti-H3K4me3 antibodies (pink) and DNA (DAPI, blue), is shown for *ash-2(tm1905)* (panel 2), *wdr-5.1(ok1417)* (panel 3), and *lsy-15(tm3463)* (panel 4), as compared to wild-type (N2) (panel 1). (B) H3K4me3 is depleted in total worm lysates of *ash-1(tm1905)* and *lsy-15(tm3463)* as compared to wild-type (N2). H3 is shown as a loading control. (C) ZFP-1 appears depleted in dissected adult mutant gonads of COMPASS mutants. Immunostaining of dissected adult gonads with anti-ZFP-1 antibodies (pink) and DNA (DAPI, blue), is shown for *ash-2(tm1905)* (panel 2), *wdr-5.1(ok1417)* (panel 3), and *lsy-15(tm3463)* (panel 4), as compared to wild-type (N2) (panel 1). All images in this figure were taken on a Zeiss AxioImager Z1 microscope at 200x total magnification. (D) Western blot analysis of total adult worm lysates from *wdr-5.1(ok1417), lsy-15(tm1905),* and *ash-2(tm3463)* with anti-ZFP-1 C-terminal specific antibodies (Mansisidor et al, 2011). Total levels of ZFP-1 protein are not changing across the COMPASS mutants. (E) ZFP-1 long isoform is depleted in *glp-4(bn2)* sterile mutant adults compared to wild-type. The long isoform band appears to be missing in *glp-4(bn2)* animals grown at the restrictive temperature (Beanan & Strome, 1992). This shows that the band corresponding to the long isoform seen by Western is largely due to the germline tissue.

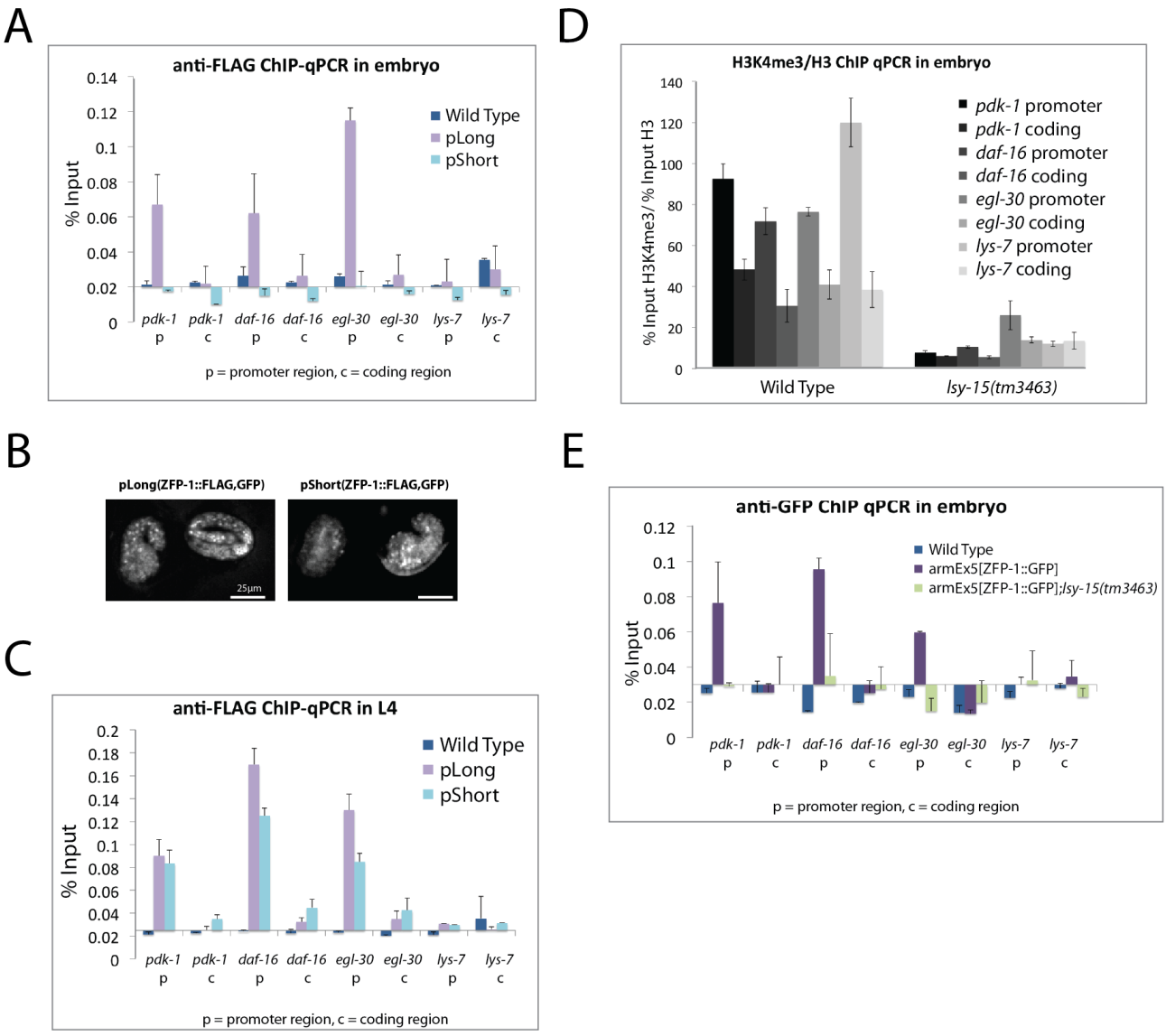


Figure 10. H3K4me contributes to ZFP-1 localization to target gene promoters.

(A) Chromatin immunoprecipitation using anti-FLAG antibodies followed by qPCR in transgenic lines carrying pLong (ZFP-1 long::FLAG,GFP, purple bars) or pShort (ZFP-1 short::FLAG,GFP, light blue bars) (see Figure 6C for schematic) as compared to non-transgenic wild type worms (dark blue bars). pLong is enriched at the promoters of the known target genes *pdk-1, daf-16* and *egl-30* (Mansisidor et al, 2011) but not at the coding regions of these genes or a non-target, *lys-7*, in embryos. Error bars for all bar charts represent the standard deviation from

the mean from at least four biological replicas. (B) pLong and pShort are both expressed in late stage embryos as seen by GFP. Scale bar is 25μm. (C) Both long and short isoforms of ZFP-1 localize to the promoters of target genes at the L4 stage. Anti-FLAG ChIP, as in (A), in L4 worms showing enrichment of both long and short isoforms at the promoters of target genes *pdk-1*, *daf-16*, and *egl-30* but not the coding regions or a non-target, *lys-7*. (D) ChIP-qPCR of H3K4me3 normalized to H3 in wild type embryos as compared to *lsy-15(tm3463)*, the mutant of the COMPASS component homolog RbBP5. H3K4me3 is severely reduced in *lsy15(tm3464)*, as expected. (E) ChIP using anti-GFP antibodies followed by qPCR in the armEx5(ZFP-1::GFP) transgenic line either in a wild type (purple bars) or in a *lsy-15(tm3463)* background (green bars). The enrichment of ZFP-1 at the *pdk-1, daf-16* and *egl-30* promoters, as seen in (A), is lost in the *lsy-15(tm3463)* background, which lacks proper H3K4me, as seen in (D).

Author Contributions: The experiments presented in this chapter were conceived and designed by D.A. and A.G. The fosmid-based transgenes were created by G.C. and the co-immunoprecipitation experiments as well as Western blotting against ZFP-1 in wild-type, armEx14 and COMPASS mutant strains were carried out by G.C. All other experiments were carried out and analyzed by D.A.

Materials and Methods

C. elegans mutants and transgenic strains

Worms were maintained at 20°C using standard methods as previously described (Mansisidor et al, 2011). The following mutants were used: LGII: *lsy-15(tm3463), ash-2(tm1905)*, LGIII: *zfp-1(ok554), wdr-5.1(ok1417),* CB4681 [nDf17/qC1 *dpy-19(e1259)*; *glp-1(q339)*], pkIs32 (*pie-1*::GFP::H2B), LGIV: *eri-1(mg366).*

Transgenic lines armEx5 [ZFP-1::GFP,unc119(+)] (Mansisidor et al, 2011) and armEx14 [ZFP-1 PHD1-PHD2::FLAG,unc119(+)] were made by co-bombardment of both the fosmid of interest and the plasmid pMM016b (AddGene) for *unc-119(ed3)*III rescue using microparticle bombardment with a PDS-1000 Hepta Apparatus (Bio-Rad). Transgenic lines containing pLong[ZFP-1long::FLAG,GFP] and pShort[ZFP-1short::FLAG,GFP] were made by injection using 0.5ng/µl of plasmid co-injected with the PRF4 plasmid (containing *rol-6*) at a concentration of 120ng/µl. For ChIP experiments with plasmid lines, transgenes were integrated and subsequent lines were outcrossed six times.

RNA extraction and RT-PCR

RNA extraction and RT-PCR was carried out as described in (Mansisidor et al, 2011). Primers used to detect *ok554* deletion with sequence 5'AGGAGCTTTCACGTAACCTG forward primer and 5'GTGCATTTGCCATAAGAAGAC reverse primer.

Immunostaining

Adult worms were hand picked and placed in PBS (137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4) with 0.01% Tween and 25mM sodium azide in a glass petri dish. Worms were dissected using a scalpel and fixed with 2% paraformaldehyde for 1 hour at room temperature and post-fixed in ice-cold methanol for 5 minutes. Worms were transferred to glass tubes and

blocked with 3% (w/v) BSA in PBS. Blocked worms were then incubated overnight at 4°C with various anti ZFP-1 antibodies diluted 1:8000 or anti-FLAG diluted 1:300, washed 3 times with PBST and incubated for 1 hour at room temperature with anti-rabbit Alexa Fluor 555, or anti-mouse Alexa Fluor 568 (Molecular Probes) diluted 1:200. Worms were then mounted on polylysine slides (LabScientific, Inc) using Mounting solution with DAPI (Vector Laboratories) and all images were taken on Zeiss AxioImager Z1. Results shown are representative of at least five biological replicates.

Fosmid recombineering and plasmid constructs

Recombineering of the WRM0629bD09 fosmid containing the ZFP-1 locus was carried out as described in (Mansisidor et al, 2011). We generated a derivative fosmid construct to express the N-terminal portion of ZFP-1 maintaining the PHD fingers followed by a FLAG tag and a stop codon to create the armEx14 line.

DNA sequence containing the promoter and 3'UTR of the *zfp-1* gene was obtained from the PCR performed on *C. elegans* N2 genomic DNA in multiple cloning steps and inserted into plasmid pPD95.75 (Addgene). The pLong construct was created using 6kb upstream of the *zfp-1* start site as the promoter, while the pShort construct was created using intron 4 of *zfp-1* as the promoter. Coding sequences of *zfp-1* were made using cDNA libraries created by reverse transcription of total worm RNA from N2 (Bristol) worms. Primer sequences and plasmids available upon request.

RNA interference

All RNAi experiments were carried out by feeding at 20°C. For immunostaining, *zfp-1(ok554)* L1 worms were fed on RNAi food and harvested as adults, dissected and stained. For RNAi defects, *eri-1(mg366);*[*pie-1*::GFP::H2B] L4 worms were fed dsRNA-producing bacteria and

their progeny were scored for germline defects. Both experiments utilized the JA:F54F2.2 clone for *zfp-1(RNAi)* from the *C. elegans* RNAi feeding library (Kamath et al, 2003) and grew bacteria containing this clone via standard methods.

Immunoprecipitation

Eggs harvested from gravid worms grown in liquid culture were washed three times with M9 and treated with a solution of 2% paraformaldehyde in M9 buffer for 30 minutes to cross-link any interacting proteins and DNA. Worms were washed once with cold 0.1M Tris-HCl pH 8.0 and twice more with cold M9. Worms were then resuspended in 1ml of Extraction Buffer 300 [50mM Tris-HCl pH 8.0, 300mM NaCl, 1mM EDTA, 0.1% NP-40, protease inhibitor cocktail (Fermentas)] and sonicated at 20% amplitude for 30 seconds on, 30 seconds off in iced water for eight rounds using the Branson 500 Sonic Dismembrator. Samples were then centrifuged to remove cell debris at 4°C, 12000 rpm (Eppendorf AG 5424) and total protein content was quantified using the Bradford Dye Reagent (BioRad). 3mg of total protein lysate was used for each immunoprecipitation reaction and incubated with 5µg of anti-FLAG antibody (M2, Sigma), or anti-GFP antibody (3E6, Invitrogen) for 1 hour rotating at 4°C. Magnetic beads (Dynabeads, Invitrogen) were washed in extraction buffer 300 and then transferred to the antibody-protein extract solution and incubated for 1 hour rotating at 4°C. Beads were then washed once with cold TSE 150 buffer and three times with cold TSE 500 buffer [0.1% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl pH 8.0, 150 or 500 mM NaCl]. Finally, beads were resuspended in SDS sample buffer (Invitrogen), heated at 95°C for 10 minutes, and resolved by Western blotting.

Western Blotting

Samples were resolved via SDS-PAGE on a 4-12% gel (Invitrogen NuPAGE Bis-Tris) and transferred to a 4.5µm nitrocellulose membrane (BioRad) at 100mA for 60 minutes for ZFP-1 proteins, or a 12% gel and transferred to a 2µm nitrocellulose membrane (Thermo Scientific) at 100mA for 30 minutes for histone proteins. The membrane was blocked using 3% (w/v) BSA in PBS with 0.01% Tween for 1 hour, followed by overnight incubation at 4°C with one of the following antibodies: anti-ZFP-1 N-terminal antibody from modENCODE (45140002, Novus Biologicals), anti-ZFP-1 C-terminal antibody from (Mansisidor et al, 2011), anti-H3K4me3 (Millipore 06-755) or anti-H3 (Millipore 05-928) diluted 1:2000 in PBST-3% BSA. The membrane was then washed 3 times with PBST and incubated for 1 hour at room temperature with anti-rabbit HRP-conjugated secondary antibody (Perkin Elmer) diluted 1:5000 in PBST-3% BSA and visualized by SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) using a Series 2000A Film Processor (TIBA). Results shown are representative of at least three biological replicates.

Bacterial Protein Production

DNA sequence containing the coding region of the *zfp-1* gene was obtained from the PCR performed on *C. elegans* N2 cDNA library created by reverse transcription of total worm RNA using random decamer primers. The cDNA of PHD1 of *zfp-1* was cloned in frame with a C-terminal GST tag into the pET28c plasmid (Addgene). PHD1-PHD2 or only extended PHD2 cDNA was cloned in frame with an N-terminal maltose binding protein (MBP) tag into the pMALP2X plasmid (NEB). Primer sequences and plasmids available upon request.

For protein expression, *Escherichia coli* BL21 pLysS cells (Invitrogen) were transformed with prepared vectors as described and grown at 37°C until an OD_{600} of 0.6 had been reached, with

shaking at 200 rpm. To induce protein expression, we added 100 μM IPTG as well as 50μM $ZnCl_2$ for proper formation of zinc fingers, and lowered the temperature to 18°C. After 18 h, cells were harvested by centrifugation at 6000***g*** for 15 min. Pelleted bacteria were resuspended in lysis buffer [200 mM NaCl, 20 mM Tris–Cl (pH 7.5) and 50 μM $ZnCl_2$] and lysed for 6 min by sonication in 15 s intervals with a 45 s rest in between pulsing. Cell debris was spun down at 4°C and 20,000***g*** for 1 h, and GST tagged proteins were extracted from the clear lysate by flowing over a glutathione sepharose beads (GE Healthecare). Beads were subsequently washed with 40 column volumes of lysis buffer and GST-tagged proteins eluted by adding 25mM glutathione in lysis buffer. The MBP-tagged proteins were purified similarly only using amylose resin (NEB) for affinity purification. The MBP tag was cut enzymatically by adding TEV protease to a final concentration of 2μg/ml to the elution. Proteins were incubated overnight at 4°C in dialysis or until full cleavage of the proteins was achieved as determined by SDS-PAGE.

Analytical size exclusion chromatography

Analytical size-exclusion chromatography was performed at 4°C using a Superose 6 10/20 column (GE Healthcare). After the equilibration of the column in buffer [200 mM NaCl, 20 mM Tris–Cl (pH 7.5) and 50 μM $ZnCl_2$] at a flow rate of 0.3 mL/min, 200 μL of purified protein was injected and eluted at the same flow rate while the UV signal at 280 nm was monitored. We collected 0.5 mL fractions for SDS-PAGE analysis with subsequent Coomassie staining to detect proteins used in this experiment. Void volume was determined to be 8 mL.

Homology Model

The homology model of PHD1-PHD2 was constructed using the SWISS-MODEL server (Arnold et al, 2006). PHD1 was modeled separately from PHD1-PHD2 and the structure shown is based on the PHD finger of WSTF (Pascual et al, 2000), zinc ions shown come from the

structure of WSTF. PHD1-PHD2 was modeled as a whole, however PHD1 was removed from this as it was less resolved and the model of the extended PHD2 (Linker with PHD2) was left and merged with the PHD1 model based on WSTF. PHD1-PHD2 was based on the ADD domain of the ATRX protein (Argentaro et al, 2007) and the zinc ions shown in PHD2 come from this structure.

Pull-down assays

In vitro binding assays of PHD1-GST, PHD1-PHD2 or BHC80-PHD-GST were carried out using a panel of biotinylated histone tails as described in (Lan et al, 2007).

Chromatin Immunoprecipitation (ChIP)

Chromatin immunoprecipitation was carried out as described in (Mansisidor et al, 2011) using the same primers to quantify promoter and coding regions of *pdk-1* and *lys-7*. The following primers were used to detect promoter and coding sequences of *daf-16* and *egl-30* as indicated: *daf-16* promoter (forward 5'GATTCTCCCTCTCCGTTCAC reverse 5'GTGATGAAGAAGGTGGTCTC), *daf-16* coding (forward 5'CGAACTCAATCGAACCTCTC reverse 5'CAATATCACTTGGAATTGCTGG), *egl-30* promoter (forward 5'CGTGAGTTTGAGTGTCTCTG reverse 5'GATTAGGTTGTGGCTTTCCC) and *egl-30* coding (forward 5'AGTTTGGTAACGAATCAGAGGA reverse 5'CCAGAATCCTCCCATAGCTC).

References

Argentaro A, Yang JC, Chapman L, Kowalczyk MS, Gibbons RJ, Higgs DR, Neuhaus D, Rhodes D (2007) Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proc Natl Acad Sci U S A* **104:** 11939-11944

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22:** 195-201

Bahri SM, Chia W, Yang X (2001) The Drosophila homolog of human AF10/AF17 leukemia fusion genes (Dalf) encodes a zinc finger/leucine zipper nuclear protein required in the nervous system for maintaining EVE expression and normal growth. *Mech Dev* **100:** 291-301

Beanan MJ, Strome S (1992) Characterization of a germ-line proliferation mutation in C. elegans. *Development* **116:** 755-766

Campos EI, Reinberg D (2009) Histones: annotating chromatin. *Annu Rev Genet* **43:** 559-599

Caudell D, Aplan PD (2008) The role of CALM-AF10 gene fusion in acute leukemia. *Leukemia* **22:** 678-685

Chaplin T, Ayton P, Bernard OA, Saha V, Della Valle V, Hillion J, Gregorini A, Lillington D, Berger R, Young BD (1995) A novel class of zinc finger/leucine zipper genes identified from the molecular cloning of the t(10;11) translocation in acute leukemia. *Blood* **85:** 1435-1441

Cui M, Kim EB, Han M (2006) Diverse chromatin remodeling genes antagonize the Rb-involved SynMuv pathways in C. elegans. *PLoS Genet* **2:** e74

Edgley ML, Baillie DL, Riddle DL, Rose AM (2006) Genetic balancers. *WormBook***:** 1-32

Eisenmann DM (2005) Wnt signaling. *WormBook***:** 1-17

Ellis RE, Jacobson DM, Horvitz HR (1991) Genes required for the engulfment of cell corpses during programmed cell death in Caenorhabditis elegans. *Genetics* **129:** 79-94

Forissier S, Razanajaona D, Ay AS, Martel S, Bartholin L, Rimokh R (2007) AF10-dependent transcription is enhanced by its interaction with FLRG. *Biol Cell* **99:** 563-571

Gardner KE, Allis CD, Strahl BD (2011) OPERating ON Chromatin, a Colorful Language where Context Matters. *J Mol Biol*

Grimaud C, Negre N, Cavalli G (2006) From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. *Chromosome Res* **14:** 363-375

Gu SG, Fire A (2010) Partitioning the C. elegans genome by nucleosome modification, occupancy, and positioning. *Chromosoma* **119:** 73-87

Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, Fire A (2012) Amplification of siRNA in Caenorhabditis elegans generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat Genet* **44:** 157-164

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperlen P, Ahringer J (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421:** 231-237

Kennedy S, Wang D, Ruvkun G (2004) A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. *Nature* **427:** 645-649

Krivtsov AV, Armstrong SA (2007) MLL translocations, histone modifications and leukaemia stem-cell development. *Nature reviews Cancer* **7:** 823-833

Krogan NJ, Dover J, Khorrami S, Greenblatt JF, Schneider J, Johnston M, Shilatifard A (2002) COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem* **277:** 10753-10755

Lan F, Collins RE, De Cegli R, Alpatov R, Horton JR, Shi X, Gozani O, Cheng X, Shi Y (2007) Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448:** 718-722

Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442:** 91-95

Li T, Kelly WG (2011) A role for Set1/MLL-related components in epigenetic regulation of the Caenorhabditis elegans germ line. *PLoS Genet* **7:** e1001349

Lim YS, Mallapur S, Kao G, Ren XC, Wadsworth WG (1999) Netrin UNC-6 and the regulation of branching and extension of motoneuron axons from the ventral nerve cord of Caenorhabditis elegans. *J Neurosci* **19:** 7048-7056

Lin YH, Kakadia PM, Chen Y, Li YQ, Deshpande AJ, Buske C, Zhang KL, Zhang Y, Xu GL, Bohlander SK (2009) Global reduction of the epigenetic H3K79 methylation mark and increased chromosomal instability in CALM-AF10-positive leukemias. *Blood* **114:** 651-658

Linder B, Gerlach N, Jackle H (2001) The Drosophila homolog of the human AF10 is an HP1-interacting suppressor of position effect variegation. *EMBO Rep* **2:** 211-216

Linder B, Jones LK, Chaplin T, Mohd-Sarip A, Heinlein UA, Young BD, Saha V (1998) Expression pattern and cellular distribution of the murine homologue of AF10. *Biochim Biophys Acta* **1443:** 285-296

Linder B, Newman R, Jones LK, Debernardi S, Young BD, Freemont P, Verrijzer CP, Saha V (2000) Biochemical analyses of the AF10 protein: the extended LAP/PHD-finger mediates oligomerisation. *J Mol Biol* **299:** 369-378

Mahmoudi T, Boj SF, Hatzis P, Li VS, Taouatas N, Vries RG, Teunissen H, Begthel H, Korving J, Mohammed S, Heck AJ, Clevers H (2010) The leukemia-associated Mllt10/Af10-Dot1l are Tcf4/beta-catenin coactivators essential for intestinal homeostasis. *PLoS Biol* **8:** e1000539

Mansisidor AR, Cecere G, Hoersch S, Jensen MB, Kawli T, Kennedy LM, Chavez V, Tan MW, Lieb JD, Grishok A (2011) A Conserved PHD Finger Protein and Endogenous RNAi Modulate Insulin Signaling in Caenorhabditis elegans. *PLoS Genet* **7:** e1002299

Mohan M, Herz HM, Takahashi YH, Lin C, Lai KC, Zhang Y, Washburn MP, Florens L, Shilatifard A (2010a) Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes Dev* **24:** 574-589

Mohan M, Lin C, Guest E, Shilatifard A (2010b) Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nature reviews Cancer* **10:** 721-728

Moneypenny CG, Shao J, Song Y, Gallagher EP (2006) MLL rearrangements are induced by low doses of etoposide in human fetal hematopoietic stem cells. *Carcinogenesis* **27:** 874-881

Musselman CA, Kutateladze TG (2011) Handpicking epigenetic marks with PHD fingers. *Nucleic Acids Res* **39:** 9061-9071

Oh SW, Mukhopadhyay A, Dixit BL, Raha T, Green MR, Tissenbaum HA (2006) Identification of direct DAF-16 targets controlling longevity, metabolism and diapause by chromatin immunoprecipitation. *Nat Genet* **38:** 251-257

Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM, Su L, Xu G, Zhang Y (2005) hDOT1L links histone methylation to leukemogenesis. *Cell* **121:** 167-178

Pascual J, Martinez-Yamout M, Dyson HJ, Wright PE (2000) Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J Mol Biol* **304:** 723-729

Pena PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442:** 100-103

Perrin L, Bloyer S, Ferraz C, Agrawal N, Sinha P, Dura JM (2003) The leucine zipper motif of the Drosophila AF10 homologue can inhibit PRE-mediated repression: implications for leukemogenic activity of human MLL-AF10 fusions. *Mol Cell Biol* **23:** 119-130

Perrin L, Dura JM (2004) Molecular genetics of the Alhambra (Drosophila AF10) complex locus of Drosophila. *Mol Genet Genomics* **272:** 156-161

Poole RJ, Bashllari E, Cochella L, Flowers EB, Hobert O (2011) A Genome-Wide RNAi Screen for Factors Involved in Neuronal Specification in Caenorhabditis elegans. *PLoS Genet* **7:** e1002109

Praitis V, Casey E, Collar D, Austin J (2001) Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. *Genetics* **157:** 1217-1226

Rando OJ, Chang HY (2009) Genome-wide views of chromatin structure. *Annu Rev Biochem* **78:** 245-271

Ross JA (2008) Environmental and genetic susceptibility to MLL-defined infant leukemia. *J Natl Cancer Inst Monogr***:** 83-86

Saha V, Chaplin T, Gregorini A, Ayton P, Young BD (1995) The leukemia-associated-protein (LAP) domain, a cysteine-rich motif, is present in a wide range of proteins, including MLL, AF10, and MLLT6 proteins. *Proc Natl Acad Sci U S A* **92:** 9737-9741

Saksouk N, Avvakumov N, Champagne KS, Hung T, Doyon Y, Cayrou C, Paquet E, Ullah M, Landry AJ, Cote V, Yang XJ, Gozani O, Kutateladze TG, Cote J (2009) HBO1 HAT complexes target chromatin throughout gene coding regions via multiple PHD finger interactions with histone H3 tail. *Mol Cell* **33:** 257-265

Sanchez R, Zhou MM (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36:** 364-372

Schindler U, Beckmann H, Cashmore AR (1993) HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J* **4:** 137-150

Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Peña P, Lan F, Kaadige MR, Lacoste N, Cayrou C, Davrazou F, Saha A, Cairns BR, Ayer DE, Kutateladze TG, Shi Y, Côté J, Chua KF, Gozani O (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442:** 96-99

Simonet T, Dulermo R, Schott S, Palladino F (2007) Antagonistic functions of SET-2/SET1 and HPL/HP1 proteins in C. elegans development. *Dev Biol* **312:** 367-383

Stone RM (2002) The difficult problem of acute myeloid leukemia in the older adult. *CA Cancer J Clin* **52:** 363-371

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14:** 1025-1040

Vastenhouw NL, Brunschwig K, Okihara KL, Muller F, Tijsterman M, Plasterk RH (2006) Gene expression: long-term gene silencing by RNAi. *Nature* **442:** 882

Visnapuu M-L, Greene EC (2009) Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. *Nature Structural & Molecular Biology* **16:** 1056-1062

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C, Allis CD (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442:** 86-90

Zhou MI, Wang H, Foy RL, Ross JJ, Cohen HT (2004) Tumor suppressor von Hippel-Lindau (VHL) stabilization of Jade-1 protein occurs through plant homeodomains and is VHL mutation dependent. *Cancer Res* **64:** 1278-1286

Chapter 3:

CSR-1 RNAi Pathway Positively Regulates Histone Expression in *C. elegans*

Daphne C. Avgousti, Santhosh Palani, Yekaterina Sherman, Alla Grishok

Adapted from Avgousti *et al*, under review EMBO J

Abstract

Endogenous small interfering RNAs (endo-siRNAs) have been discovered in many organisms, including mammals. In *C. elegans*, depletion of germline-enriched endo-siRNAs found in complex with the CSR-1 Argonaute protein causes sterility and defects in chromosome segregation in early embryos. We have discovered that knockdown of either *csr-1,* the RNA-dependent RNA polymerase (RdRP) *ego-1,* or the dicer-related helicase *drh-3,* leads to defects in histone mRNA processing, resulting in severe depletion of core histone proteins. The maturation of replication-dependent histone mRNAs, unlike that of other mRNAs, requires processing of their 3'UTRs through an endonucleolytic cleavage guided by the U7 snRNA, which is lacking in *C. elegans*. We found that CSR-1-bound antisense endo-siRNAs match histone mRNAs and mRNA precursors. Consistently, we demonstrate that CSR-1 directly binds to histone mRNA in an *ego-1* dependent manner using biotinylated 2′-O-methyl RNA oligonucleotides. Moreover, we demonstrate that increasing the dosage of histone genes rescues the lethality associated with depletion of CSR-1 and EGO-1. This role of RNAi in histone mRNA processing is the first demonstration of its positive effect on gene expression.

Introduction

C. elegans contains a large number of endogenous siRNAs, the majority of which are antisense to protein-coding mRNAs and are loaded by Argonaute (AGO) proteins (Ambros et al, 2003; Claycomb et al, 2009; Gu et al, 2009; Pak & Fire, 2007; Ruby et al, 2006). Endo-siRNAs produced by RNA-dependent RNA Polymerases (RdRPs) (Aoki et al, 2007; Pak & Fire, 2007; Ruby et al, 2006) are termed 22G-RNAs and can be divided into classes based on the AGOs to which they bind (Claycomb et al, 2009; Gu et al, 2009). The Argonaute CSR-1 was shown to preferentially bind gene-specific endo-siRNAs (Claycomb et al, 2009). The synthesis of endo-siRNAs in the CSR-1 22G-RNA pathway is dependent on the RdRP EGO-1 (Claycomb et al, 2009; Smardon et al, 2000) and the dicer-related helicase DRH-3 (Duchaine et al, 2006; Nakamura et al, 2007). Depletion of EGO-1, CSR-1 or DRH-3 results in similar sterility and defects in chromosome segregation in early embryos (Claycomb et al, 2009; Duchaine et al, 2006; Nakamura et al, 2007; She et al, 2009; Smardon et al, 2000; Yigit et al, 2006). Based on the established model in *S. pombe* (Halic & Moazed, 2009; Lejeune & Allshire, 2011), it was proposed that the CSR-1 22G-RNA pathway aids in the proper compaction of the holocentric chromosomes of the worm (Claycomb et al, 2009). However, in fission yeast, siRNAs match pericentromeric repeat regions (Halic & Moazed, 2009), in contrast to *C. elegans* siRNAs, which were shown to primarily match protein-coding genomic regions (Claycomb et al, 2009), including histone genes (Grishok et al, 2008).

Here, we demonstrate that knockdown of *csr-1, ego-1* or *drh-3* leads to misprocessed histone mRNA, resulting in severe depletion of core histone proteins, and that expression of transgene-encoded histone transcripts rescues the lethality of the RNAi mutants. We propose that this depletion of core histones is responsible for the sterility and chromosome segregation

defects of the CSR-1 RNAi pathway mutants. In most organisms, specific 3'UTR processing of histone mRNAs includes a cleavage by the Cleavage and Polyadenylation Specificity Factor subunit CPSF73 just downstream of a conserved stem-loop. This stem-loop is recognized by a conserved stem-loop binding protein (SLBP), while the histone downstream element (HDE) sequence is recognized by the U7 snRNA (Marzluff et al, 2008). Surprisingly, in *C. elegans,* U7 and its binding partner Lsm11, as well as the HDE sequence normally present at 3′UTRs of histone genes, are missing (Davila Lopez & Samuelsson, 2008). This avails the possibility for direct involvement of endo-siRNAs in histone mRNA 3′end processing. Consistent with this possibility, we show that CSR-1 interacts with 2′-O-methyl oligonucleotides that mimic the histone mRNA 3′UTR sequences in an *ego-1*-dependent, i.e. siRNA-dependent, manner. We also show that CSR-1 interacts with histone mRNA in an *ego-1*-dependent manner. These results strongly suggest that CSR-1 regulates histone mRNAs directly. We conclude that CSR-1-dependent endo-siRNAs play a positive role in the regulation of core histone expression in *C. elegans*.

Results

CSR-1-bound siRNAs directly target histone genes

Previous studies reported no global changes in the expression of the genes targeted by the CSR-1 endogenous RNAi pathway in *csr-1*(*tm892*) mutant adults (Claycomb et al, 2009; Gu et al, 2009). Our analysis of the microarray data performed on *csr-1*(*tm892*) by Claycomb *et al*. (Claycomb et al, 2009) showed, as previously reported, that the majority of CSR-1 targets (65%), defined by their antisense endo-siRNAs enriched in CSR-1-IP, had no change in gene expression in adults (Figure 1A and B). Interestingly, we found that the ratio of downregulated genes to

upregulated genes among the CSR-1 targets was significantly (10-fold, $p < 0.0001$) higher compared to that of the global gene expression profile (Figure 1A and B). However, most of the downregulated genes could be a consequence of comparing sterile to gravid adults. Nevertheless, downregulation of a few critical CSR-1-targeted germline genes could be causal in inducing the phenotypes observed in *csr-1*, *ego-1* and *drh-3* mutants. Among the downregulated genes, we found the majority of the histone genes (66%) to be depleted in *csr-1(tm892)* mutants (Figure 1C). This is significant because downregulation of just one core histone gene by RNAi has previously been shown to be sufficient to cause chromosome segregation and sterility phenotypes (Kodama et al, 2002; Sonnichsen et al, 2005). We hypothesize that since histones are essential for chromatin compaction, defects in their production could underlie the phenotypes observed in the RNAi mutants.

We observed that antisense endo-siRNAs targeting all core histone genes were enriched in CSR-1-IP libraries (Claycomb et al, 2009) (Figure 1D), indicating a possible direct role in histone mRNA regulation by CSR-1 endo-siRNAs. The 3′ends of histone mRNAs were mapped to a few nucleotides downstream of the stem-loop (Keall et al, 2007) (Figure 1F, green arrows), which is consistent with tiling microarray data (Spencer et al, 2011). However, polyadenylation signals (PAS) are also present at 3'UTRs of histone genes (Keall et al, 2007; Mangone et al, 2010) (Figure 1E and F), and it has recently been discovered that a small portion of *C. elegans* and mammalian histone mRNAs are polyadenylated (Jan et al, 2011; Mangone et al, 2010; Shepard et al, 2011). Unprocessed histone mRNAs that accumulate in mutants deficient in processing are also known to be polyadenylated (Godfrey et al, 2006; Ideue et al, 2009). It has been proposed that the production of polyadenylated histone messages precedes the histone mRNA-specific processing (Mangone et al, 2010). Such a two-step mechanism for histone

mRNA biogenesis would allow more inputs for potential regulation. The stem-loop and PAS are located in close proximity (~30nt) within 3′UTRs of *C. elegans* histone messages (Figure 1F). We therefore searched for endo-siRNAs antisense to these portions of 3′UTRs, which would have the potential to guide the cleavage reaction. Indeed, we found *ego-1*-dependent endo-siRNAs (Claycomb et al, 2009) that match 3′UTRs in antisense orientation opposite to the mapped 3′ends of mature histone mRNAs (Keall et al, 2007) (Figure 1F). These analyses suggest the possibility of a direct role of the CSR-1 endo-siRNA pathway in histone mRNA biogenesis.

CSR-1 binds to histone mRNA in an *ego-1*-dependent manner

To confirm that CSR-1 binds to endo-siRNAs specific to the 3′UTR of histone genes, we designed a 34-nucleotide-long 2′-O-methyl RNA oligonucleotide identical to the region of the 3′UTR in between the stem-loop and the PAS of H2A pre-mRNA (Figure 1F, black). We incubated the biotinylated 2′-O-methyl RNA oligonucleotides with total worm lysates, performed a pull-down with streptavidin beads according to established protocols (Aoki et al, 2007; Hutvágner et al, 2004), and investigated whether CSR-1 was present in the pulled-down material by Western with anti-CSR-1 antibodies (Claycomb et al, 2009). We found an enrichment of CSR-1 in pull-downs using the H2A 3′UTR sequence, but not in experiments with an unrelated sequence (Figure 2A, top panel). This enrichment was lost upon *ego-1(RNAi),* indicating that the binding of CSR-1 to the H2A sequence was dependent on siRNAs produced by EGO-1 (Figure 2A, bottom panel). This experiment shows that CSR-1 is capable of binding to the 3′UTR of H2A pre-mRNA, and specifically to the region that is cleaved during its processing.

To validate the interaction of CSR-1 with the H2A mRNA sequence further, we designed 2′-O-methyl RNA oligonucleotides antisense to the coding region of H2A and repeated the pull-down experiment (see Materials and Methods for details). The antisense H2A oligos were also able to pull down CSR-1 in an *ego-1*-dependent manner (Figure 2B), indicating that CSR-1 interacts with the histone mRNA through its bound siRNAs. We confirmed that H2A mRNA was enriched in the pull-down by qRT-PCR (Figure 2C). These results demonstrate that CSR-1 interacts with histone mRNA through the base pairing of its bound siRNAs. CSR-1 may then facilitate cleavage of the 3′UTR downstream of the stem-loop or it may act to recruit other cleavage factors for the proper processing of the histone mRNA.

CDL-1 (SLBP) and CSR-1 are strongly enriched in the nuclei of maturing oocytes

One factor necessary for the processing, stabilization and efficient translation of histone mRNA is SLBP (Marzluff et al, 2008). Downregulation of *C. elegans* SLBP (CDL-1) by RNAi causes a severe depletion of histones as well as adult sterility and chromosome segregation defects in early embryos (Kodama et al, 2002; Pettitt et al, 2002). To better understand the connection between CDL-1 function and its localization, we created transgenic worms expressing CDL-1 with a C-terminal GFP tag under the endogenous promoter. A previous study described strong expression of a GFP reporter driven by the *cdl-1* (i. e. *R06F6.1*) promoter in the dividing cells of embryo, larva, and adult (Pettitt et al, 2002). However, germline expression was not reported, probably due to transgene silencing in this tissue. We observed expression of the CDL-1::GFP translational fusion protein in adult germline and embryos (Figure 3A). Notably, there is a strong enrichment of CDL-1::GFP in the nuclei of developing oocytes (Figure 3A, enlarged region). This is consistent with findings in other animals, as SLBP has been shown

to be present in maturing oocytes in mammals and frogs (Allard et al, 2005; Gorgoni et al, 2005; Thelie et al, 2012) and to be required maternally in flies (Sullivan et al, 2001). Interestingly, CSR-1 is present in the peri-nuclear P-granules in the distal germline (Claycomb et al, 2009; Updike & Strome, 2009), but becomes nuclear in maturing oocytes (Claycomb et al, 2009). Immunostaining of dissected adult germlines with anti-CSR-1 antibodies (Aoki et al, 2007) revealed that CSR-1 is enriched in oocyte nuclei at the same stage as CDL-1::GFP (Figure 3B). This signal is specific to CSR-1 as it is lost in the *csr-1(tm892)* null mutant (Figure 3B). Since CSR-1 directly binds to histone mRNAs, we propose that it may have a role, along with CDL-1, in histone mRNA processing in oocytes to ensure that early embryos undergoing rapid cell divisions before the onset of zygotic transcription are supplied with histone proteins and mRNA. In early embryos, CDL-1 is likely to be involved in histone mRNA translation, as has been found in other systems (Allard et al, 2005; Sullivan et al, 2001).

Interestingly, some phenotypes of the RdRP mutant *ego-1* can be related to histone biogenesis defects. The germline defects of *ego-1* mutants have been extensively characterized and include a remarkable extension of the transition zone (Smardon et al, 2000) (transition from mitosis to meiosis), enriched in nuclei in premeiotic S-phase (Kimble & Crittenden, 2005). This phenotype is consistent with histone biogenesis defects, which were shown to result in S-phase extension and cell cycle delay in cultured cells (Ideue et al, 2009; Wagner et al, 2005).

In summary, both phenotypic and expression data are consistent with the possible role of RNAi in histone biogenesis in the germline.

Knockdown of RNAi pathway components leads to a depletion in total histone mRNA and an accumulation of unprocessed histone mRNA

Although *cdl-1* has been shown to be required for proper histone expression in *C. elegans* (Kodama et al, 2002; Pettitt et al, 2002), accumulation of unprocessed histone mRNA has not been previously reported for the CDL-1 knockdown (Keall et al, 2007). To measure the level of histone mRNA misprocessing, we designed primers (Table 1) to specifically detect total histone mRNA (primers tF_1, tR_1) and unprocessed histone mRNA (primers uF_2, uR_2) for all core histone messages (Figure 4A). To test whether misprocessing of histone mRNA could lead to the defects observed in *cdl-1*(*RNAi*), we measured the levels of unprocessed and total histone mRNA in young adults, at a stage prior to the emergence of the sterility and chromosome segregation phenotype. Indeed, we found increased levels of the unprocessed mRNA and a depletion of total mRNA for all the core histones in *cdl-1*(*RNAi*) compared to wild-type (empty vector, EV) (Figure 4B). To evaluate the role of endogenous RNAi in the processing of histone mRNA, we quantified the level of misprocessing after the knockdown of endogenous RNAi components. We observed a signature very similar to that of *cdl-1*(*RNAi*) in *csr-1*(*RNAi*) and *ego-1*(*RNAi*) (Figure 4C and D) suggesting that these RNAi factors regulate histone mRNA processing. We did not achieve efficient knockdown of DRH-3 by RNAi; therefore, we performed the mRNA analysis in the *drh-3*(*tm1217*) null mutant and found similar misprocessing (Figure 4E). We also confirmed that all the RNAi treatments resulted in sterility of mature adults. To verify that the sterility we observed in the RNAi knockdown was due to histone mRNA depletion, and not *vice-versa*, we used RNAi against SMC-4, a subunit of the mitotic condensin complex whose knockdown also results in sterility and chromosome segregation defects, specifically in anaphase bridging (Hagstrom et al, 2002). We did not observe any histone mRNA depletion in *smc-*

4(*RNAi*) (Figure 4F). This signature of mRNA misprocessing of all core histones, as quantified by the ratio of the change in total mRNA to the change in unprocessed mRNA, was statistically significant for *cdl-1*(*RNAi*), *csr-1*(*RNAi*), *ego-1*(*RNAi*) and *drh-3*(*tm1217*), but not for *smc-4*(*RNAi*) compared to wild-type (*p*-values are provided in Figure 4G). This suggests that misprocessing of replication-dependent histone mRNA can cause the defects observed after knockdown of RNAi pathway components.

Histone proteins are severely depleted in the knockdown of RNAi pathway components

To evaluate the depletion of core histone proteins, we performed Western blot analyses using extracts of worms depleted of *csr-1* and *ego-1* by RNAi as well as extracts of *drh-3*(*tm1217*) null mutants. Extracts from *cdl-1*(*RNAi*) and *smc-4*(*RNAi*) were included as positive and negative controls, respectively. Consistent with our mRNA results, the levels of H2A (50-75% depletion compared to wild type) and H2B histones (~50% compared to wild type) were dramatically reduced in *csr-1* and *ego-1*(*RNAi*) and in *drh-3*(*tm1217*), similar to the depletion observed in *cdl-1*(*RNAi*) worms (Figure 5A and B). H2A and H2B levels were not reduced in *smc-4*(*RNAi*) (Figure 5A). In a different approach, we stained dissected gonads from young adult worms treated with *csr-1* and *ego-1* RNAi and *drh3*(*tm1217*) mutants with anti-H2B antibodies to confirm that the depletion takes place in the germline and therefore should influence the histone pool in the early embryos. As expected from our Western blot results, we detected a severe depletion of H2B in the *cdl-1*, *csr-1* and *ego-1* RNAi and *drh-3*(*tm1217*), but not in *smc-4*(*RNAi*) (Figure 5C).

Although depletion of just one core histone has been shown to be sufficient to cause sterility and chromosome segregation phenotypes (Kodama et al, 2002; Sonnichsen et al, 2005),

we wanted to determine whether H3 and H4 were also affected. Our mRNA results indicate that all four histones should be similarly affected (Figure 4). We found, by using the same biological samples for Westerns, that one antibody specific to H4 reveals a severe depletion (~75%) in *csr-1(RNAi)* and *ego-1(RNAi)* samples while another shows no difference in H4 levels under the knockdown conditions (Figure 6, compare A and B). Thus, it appears that using more sensitive antibodies may preclude the detection of any difference in protein levels due to saturation of the signal.

Since RNAi targeting histone genes in *C. elegans* was shown to cause chromosome segregation defects (Kodama et al, 2002; Sonnichsen et al, 2005), our findings of severe histone depletion in *csr-1* and *ego-1* RNAi and *drh-3*(*tm1217*) point to it as the most likely cause of their published cell division phenotypes (Claycomb et al, 2009; Duchaine et al, 2006; Nakamura et al, 2007; Yigit et al, 2006).

A higher level of expression of histone proteins is also needed during DNA endoreduplication in the intestinal cells of *C. elegans* (McGhee, 2007). Consistently, we find by immunostaining with anti-CSR-1 antibodies (Aoki et al, 2007) that CSR-1 is present in the intestinal nuclei of adult worms (Figure 7A). Previous studies reported a lower level of CSR-1 protein expression in somatic tissues compared to the germline, although *csr-1* mRNA specific to the long isoform of the protein is produced in the soma at high levels (Claycomb et al, 2009). Our results are consistent with these findings. Furthermore, we find that upon *csr-1(RNAi),* H2B is depleted in intestinal nuclei, confirming that the dependence of histone production on the RNAi pathway is not limited to the germline (Figure 7B).

P-granule formation is not affected by the knockdown of CDL-1

In addition to sterility and chromosome segregation defects, the endogenous RNAi pathway components CSR-1, EGO-1 and DRH-3 have also been shown to affect P-granule formation in the *C. elegans* gonad (Updike & Strome, 2009). Next, we asked whether sterility in *cdl-1*(*RNAi*) correlated with the disruption of P-granule morphology as well, by staining for PRG-1, a PIWI protein found in P-granules (Batista et al, 2008). We did not observe any defects in P-granule morphology in dissected gonads from *cdl-1*(*RNAi*) young adults compared to *csr-1*(*RNAi*) (Figure 8). Since CSR-1-bound endo-siRNAs produced by EGO-1 regulate additional germline genes (Maniar & Fire, 2011) while CDL-1 has a specific and conserved role in histone mRNA processing, these results indicate that histone mRNA misprocessing can cause sterility in *cdl-1*(*RNAi*) without affecting P-granule morphology. Therefore, histone mRNA misprocessing alone is sufficient to cause sterility in RNAi pathway mutants.

Overexpression of transgenic histones rescues the sterility caused by the knockdown of RNAi pathway components

Next, we tested whether the sterility of RNAi mutants can be rescued by increased histone production from transgenes. We chose two pairs of histone genes: H2A/H2B and H3/H4 (*his-59-62* locus) arranged in head-to-head orientation and designed constructs for rescue experiments. The constructs included promoters, coding regions and truncated 3'UTRs of histone genes ending after the conserved stem-loop (see Methods for experimental details). Our results described above suggest that the processing of histone mRNAs in *C. elegans* is dependent on the RNAi pathway. We hypothesized that endo-siRNAs antisense to 3'UTRs of histone genes and mapping downstream of the stem-loop (Figure 1F) regulate histone mRNA-specific

processing, which includes a cleavage between the stem-loop and PAS. Since an increase in unprocessed histone mRNA resulted in the depletion of total mRNA (Figure 4) and histone protein levels (Figure 5), we hypothesized that the removal of the 3'UTR downstream of the stem-loop, including the PAS, should stimulate stem-loop-dependent histone expression. Therefore, we tested whether the constructs containing the stem-loop but lacking the PAS in the 3'UTRs would bypass the requirement of RNAi for histone production.

We examined the expression levels of the transgene-encoded histone mRNAs in several transgenic lines and selected high (*armEx149*) and low (*armEx152*) expressing lines for further experiments. There was a notable increase in the total levels of histone proteins in the high expressing *armEx149* line compared to the low expressing control *armEx152* line (Figure 9A, compare lanes 1 and 6, Figure 11C, compare lanes 1 and 3). Next, we downregulated *csr-1*, *ego-1* and *cdl-1* by RNAi in high and low-expressing transgenic lines. We initiated RNAi by feeding young larva (L1) and observed survival of their progeny. *csr-1(RNAi)*, *ego-1(RNAi)* and *cdl-1(RNAi)* induced penetrant embryonic lethality in the low-expressing transgenic control line (Figure 9B purple bars, $p< 0.0001$). However, *csr-1(RNAi)* and *ego-1(RNAi)* failed to induce significant embryonic lethality in the strain expressing high levels of transgenic histone mRNAs (Figure 9B, green bars). *cdl-1(RNAi)* lethality was not rescued in this transgenic background, consistent with the requirement for the stem-loop binding protein in nuclear export and translation of histone mRNAs (Allard et al, 2005; Marzluff et al, 2008). The resistance of the *armEx152* line to *csr-1(RNAi)* and *ego-1(RNAi)* is not the result of a general insensitivity to RNAi in this strain (Figure 10A and B).

Although the total levels of histone mRNA and histone proteins decreased in the high-expressing line treated with *csr-1(RNAi)* (Figure 9A, compare lanes 1 and 5, Figure 10C), the

expression level of transgene-encoded messages did not change under these conditions (Figure 10D), nor were transgene encoded histone mRNAs polyadenylated (Figure 10E). Due to the overexpression of transgenic histones from the array, we found a significant difference in the expression of histone proteins in the rescued transgenic line treated with *csr-1(RNAi)* compared to the line with low transgene expression, as confirmed by Western (Figure 9A, compare lanes 5 and 7) and immunofluorescence analyses for *csr-1(RNAi)* and *ego-1(RNAi)* (Figure 9C). This difference in histone protein levels in the germlines of rescued versus sensitive transgenic worms under *csr-1(RNAi)* and *ego-1(RNAi)* conditions is completely consistent with either the survival or the lethality of their progeny shown in Figure 6B.

To confirm that the lack of the PAS in the transgenic array contributed to RNAi-independent histone biogenesis, we re-introduced a PAS sequence into the 3'UTRs of H2A and H2B genes on the transgenic array. This was done so that the plasmid backbone sequence was present in place of sequences normally targeted by siRNAs. We selected two lines expressing the transgene at a higher levels than *armEx152* (Figure 11A) and tested whether these arrays were able to rescue the lethality phenotype caused by *csr-1(RNAi)*. We found that the lines containing the PAS signal showed significantly lower levels of rescue compared to *armEx149* line, (compare 90% to 26% for *armEx180*, and 90% to 6% for *armEx184* (p value = 0.001, Figure 11B). We note that these lines had a higher level of rescue compared to *armEx152,* consistent with some level of histone expression in the germline. However, this increase is minimal as it was not possible to see any difference in histone protein levels between *armEx152* and *armEx180* by Western (Figure 11C). Therefore, although the transgenes containing the PAS express mRNA (Figure 11A), this does not translate into a significant increase in total histone protein, which explains the lack of rescue in these lines. The caveat of these experiments is that

none of the transgenic lines designed to express histone mRNA with the stem loop and PAS separated by plasmid sequences produced RNA at the levels similar to that of *armEx149* (Figure 11A). The stability of unprocessed histone mRNAs appears to be lower compared to that of properly processed ones; note the decrease in total mRNA levels in samples containing abundant unprocessed transcripts (Figure 4B-E). This increased destabilization of unprocessed histone transcripts has been observed in the maternal mutant embryos of SLBP in flies, indicating an active degradation process during oogenesis (Sullivan et al, 2001). Since the *C. elegans* SLBP homolog is enriched in oocytes, it is likely that a similar degradation mechanism of unprocessed transcripts takes place in the *C. elegans* germline as well. Therefore, it may not be possible to generate transgenic lines with high expression of RNAs mimicking the unprocessed histone mRNA precursors.

We conclude that histone messages containing both the stem-loop and the PAS sequence, but no siRNA sites between them, appear to be expressed and utilized poorly compared to those without the PAS sequence, which suggests the need for an RNAi-dependent processing step.

Discussion

In most organisms, maturation of mRNAs coding for replication-dependent core histones, unlike other mRNAs, requires the processing of their 3'UTRs through an endonucleolytic cleavage between the conserved stem-loop and the specific element recognized by U7 snRNA (Marzluff et al, 2008). Although the vast majority of mature histone mRNAs are not polyadenylated, recent deep-sequencing studies discovered polyadenylated histone mRNAs in *C. elegans* (Jan et al, 2011; Mangone et al, 2010) and in mammalian cells (Shepard et al, 2011). Hence, it was proposed that addition of polyA precedes histone mRNA-specific cleavage events

(Mangone et al, 2010). Since nematodes lack the U7 snRNA (Davila Lopez & Samuelsson, 2008), it is possible that endo-siRNAs assume its role in guiding histone mRNA processing in *C. elegans* either by directly cleaving the transcript through the slicer activity of CSR-1 (Aoki et al, 2007) or by recruiting other conserved cleavage factors specific to 3' end processing. We have shown that CSR-1 directly interacts with histone mRNAs in an *ego-1*-dependent manner and that the knockdown of CSR-1 leads to histone mRNA misprocessing that results in severe histone protein depletion. We further demonstrated that the overexpression of transgenic histone proteins rescues the lethality caused by *csr-1* and *ego-1* RNAi. Together, these results indicate that an increase in core histone expression is sufficient to rescue the lethality of the CSR-1 pathway mutants. Therefore, we propose that it is the lack of adequate histone production that leads to the published phenotypes in these RNAi mutants (Claycomb et al, 2009; Duchaine et al, 2006; Nakamura et al, 2007; She et al, 2009; Smardon et al, 2000; Yigit et al, 2006).

Higher levels of histone expression in *C. elegans* correlate well with the presence of actively dividing cells, with almost all histone expression in adult worms taking place in the germline (Keall et al, 2007). We observe the highest levels of CDL-1 (SLBP) expression in the oocytes and early embryos, consistent with the need for histone production during early development. CSR-1 is also highly enriched in the oocytes and its depletion in the germline phenocopies the depletion of CDL-1.

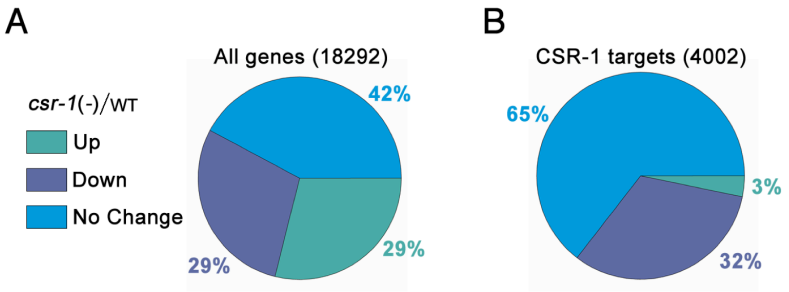
As mentioned earlier, one possibility for the function of RNAi in histone processing is a direct role of CSR-1 in the endonucleolytic cleavage of histone 3'UTRs since it has been shown to possess endonuclease cleavage activity *in vitro* (Aoki et al, 2007). Our findings of the endo-siRNA-dependent binding of CSR-1 to a 2′-O-methyl RNA oligonucleotide identical to the region of the 3′UTR in between the stem-loop and the PAS strongly support this possibility.

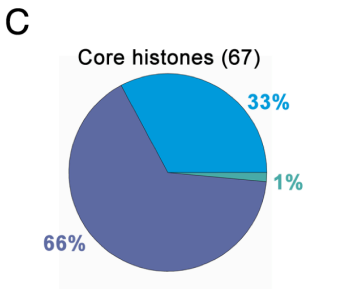
Overall, our results and the existing published data (Jan et al, 2011; Mangone et al, 2010) are consistent with a sequential model of histone mRNA processing where the messages are initially polyadenylated, used by the RdRP EGO-1 as templates for antisense endo-siRNA production, and finally cleaved by CSR-1, guided by *ego-1*-dependent endo-siRNAs matching the 3'UTR between the stem-loop and the PAS. At this time, we also cannot exclude the possibility that the RNAi machinery serves to recruit other cleavage factors to process the histone mRNA.

The production of the polyadenylated histone pre-mRNA intermediate could help in the recruitment of factors required for further processing or, alternatively, could attract proteins promoting pre-mRNA degradation in the oocytes, such that only properly processed messages ready for translation are deposited into the zygote and early embryo.

Interestingly, the majority of CSR-1-bound endo-siRNAs matching other messages serves to regulate transcriptional elongation in somatic tissues (Cecere et al, under review). It is possible that regulation of histone genes by CSR-1-bound endo-siRNAs includes both transcriptional control and a direct role in histone mRNA processing. The transcriptional role could be served by siRNAs antisense to the coding regions while the processing role fits siRNAs matching the 3′UTR sequences in the model described earlier. Also, there are two isoforms of the CSR-1 protein (Claycomb et al, 2009), consistent with diverse functions.

In conclusion, our study has revealed a novel positive role of RNAi in histone mRNA processing and demonstrated that the severe developmental phenotypes of the RNAi pathway mutants in *C. elegans* are due to insufficient production of core histone proteins.





D

Histone	Fold change in CSR-1-IP endo-siRNAs compared to wild-type input
H2A	3.49
H2B	2.54
H3	2.33
H4	1.90

F

his-61 (H2A)

stem-loop PAS
AACGGCCCTCTTTAGGGCCACAATTATCTGAAATCCTAACTATCAATTGAATAAA
CCGGTGTTAATAGACTTTAGGA
GTGTTAATAGACTTTAGGATTG
TAATAGACTTTAGGATTGATAG

his-8 (H2B)

stem-loop PAS
AACGGCCCTCTTTAGGGCCACATTTGTGAAAATCCTTTATCAATGCAAATATTCGTATAAT
ATCCCGGTGTAAACACTTTTAG
TGTAAACACTTTTAGGAAATAG
ACACTTTTAGGAAATAGTTACG

his-17 (H3)

stem-loop PAS
AACGGCCCTCTTTAGGGCCACATATCTATGATCCATCGAATTTACTTTGAATAAA
CCGGTGTATAGATACTAGGTAG
GTGTATAGATACTAGGTAGCTT
TAGATACTAGGTAGCTTAAATG

his-60 (H4)

stem-loop PAS
AACGGCCCTCTTTAGGGCCACAAATATCCAAAATCCTAATTTTTCTTTGCTAATAAA
TCCCGGTGTTTATAGGTTTTAG
GTGTTTATAGGTTTTAGGATTA
TATAGGTTTTAGGATTAAAAAG

Figure 1: CSR-1-associated siRNAs target replication-dependent core histone genes.

(A-C) Gene expression profile of *csr-1*(*tm892*) adults compared to wild type. Pie charts showing the distribution of no change, up- (>1.5-fold) and downregulated (<1.5-fold) genes for the whole genome (A), CSR-1 targets (B) and core histones (C) (microarray data by Claycomb et al, 2009. The majority of all genes and CSR-1 target genes showed no change in gene expression in the *csr-1* mutant, but core histone genes were predominantly down. (D) Fold change in total number of endo-siRNAs corresponding to each core histone in CSR-1-IP compared to wild-type input (deep sequencing data by Claycomb et al, 2009). Fold change is calculated from total endo-siRNA counts for all genes corresponding to each histone. (E) Schematic of a typical histone gene (to scale): an upstream 5'UTR followed by the coding region, typically without introns, and a downstream 3'UTR containing a stem-loop forming sequence (pink) followed by the poly(A) signal (PAS, green). (F) Endo-siRNAs (red) present in wild-type libraries, and not found in the *ego-1* mutant library, mapping to the histone 3'UTR adjacent to the stem-loop (pink) for one example of each core histone gene [*his-61* (H2A), *his-8* (H2B), *his-17* (H3) and *his-60* (H4)] are shown. The dark green arrows indicate the previously detected 3'UTR ends of the mature histone mRNAs from Keall *et al* (Keall et al, 2007).

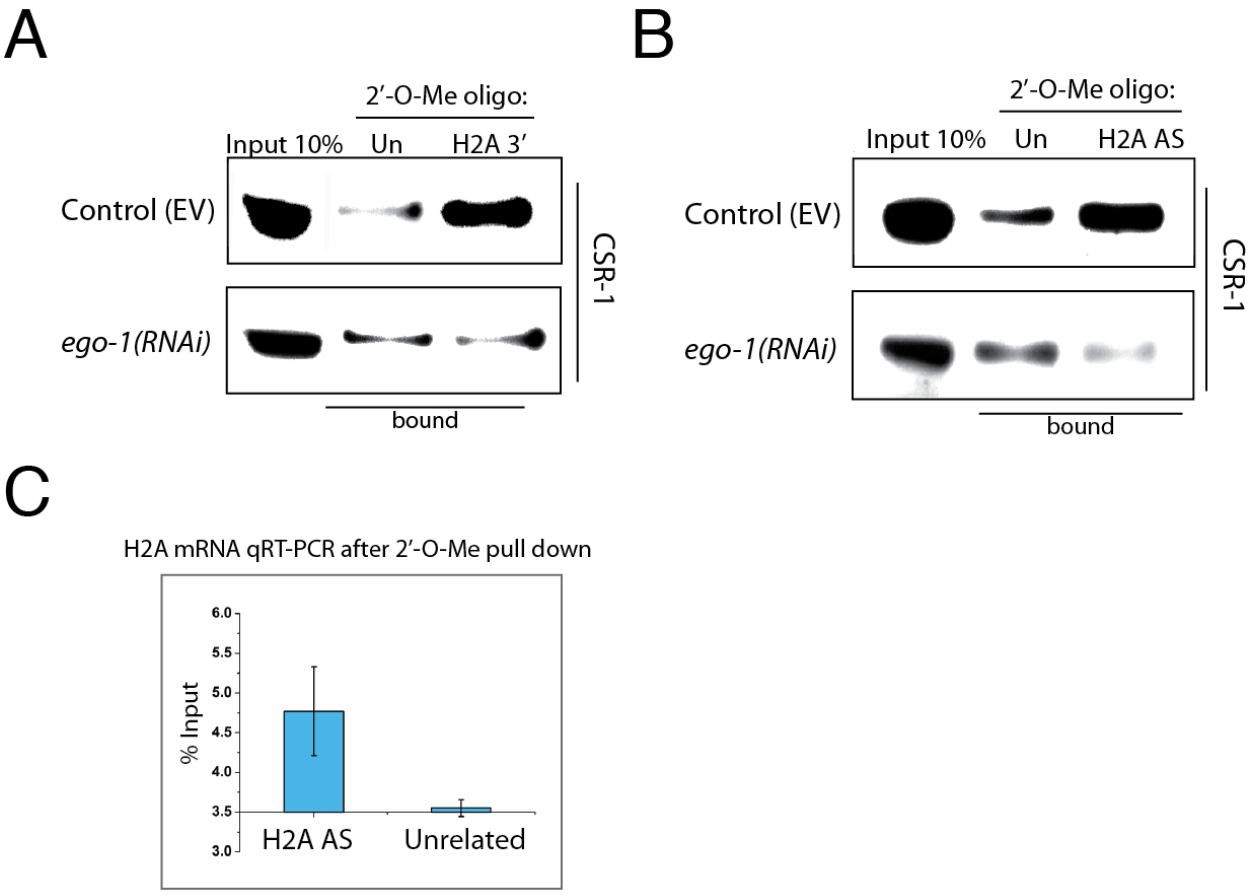


Figure 2: CSR-1 interacts with histone mRNA (A) Western blot analysis with anti-CSR-1 antibodies (Claycomb et al, 2009) of pull down experiments using biotinylated 2′-O-methyl oligonucleotides. 'Un' indicates an unrelated sequence used as a negative control where 'H2A 3'' is the sense strand of the 3′UTR of H2A in between the stem-loop and PAS signal (Figure 1F, black sequence). The top panel shows enrichment of CSR-1 using the H2A sequence but not the unrelated sequence. The bottom panel shows the loss of this enrichment upon knockdown of *ego-1* by RNAi. (B) The same as in (A), only using biotinylated 2′-O-methyl oligonucleotides antisense (AS) to the coding region of H2A. The top panel shows enrichment of CSR-1 using the H2A antisense sequence, but not the unrelated sequence. The bottom panel shows loss of this

enrichment upon knockdown of *ego-1* by RNAi. (C) Pull down using biotinylated 2′-O-methyl RNA oligonucleotides antisense to the coding region of H2A, as done in (B), leads to an enrichment of H2A mRNA when compared to pull down with an unrelated sequence.

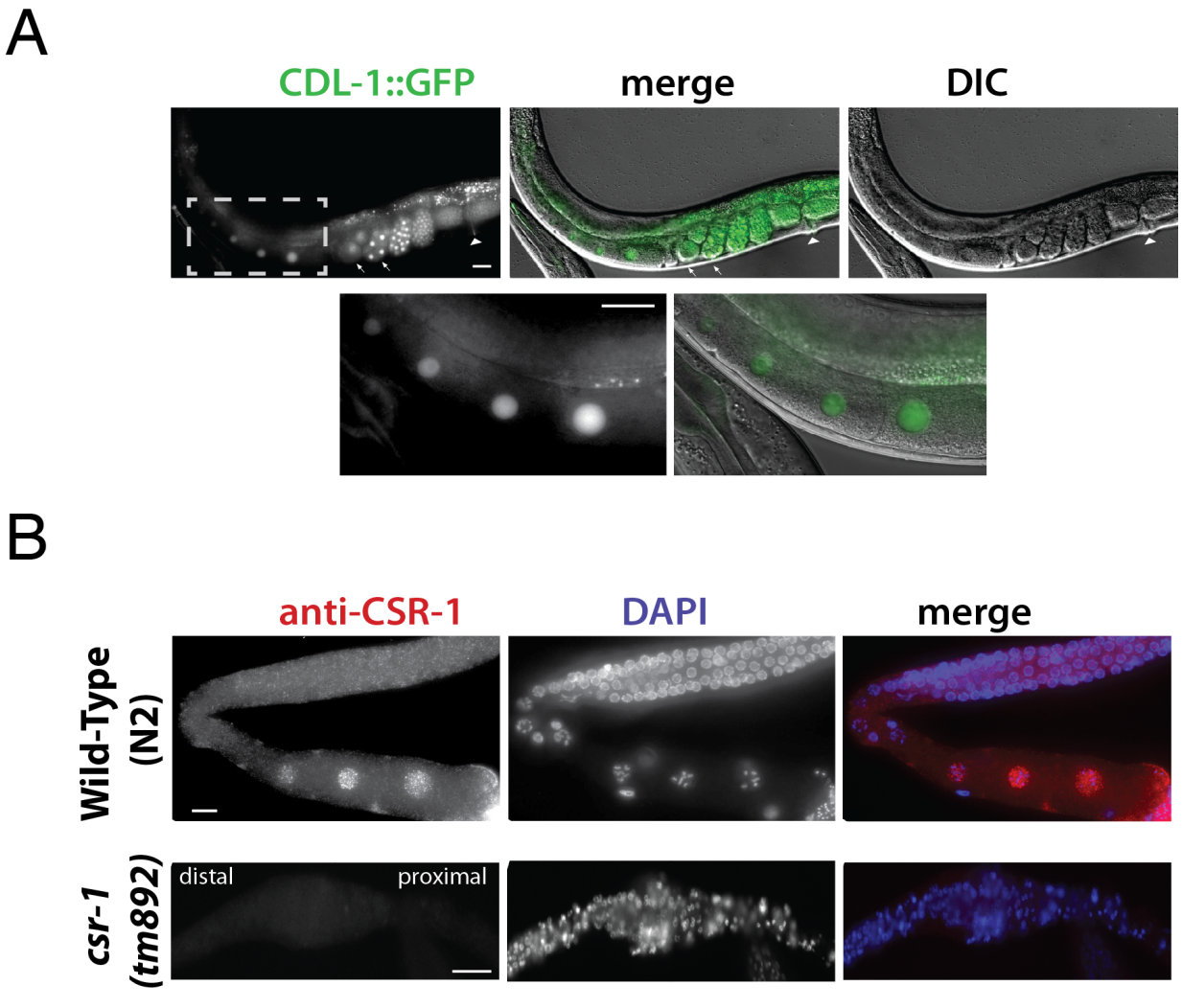
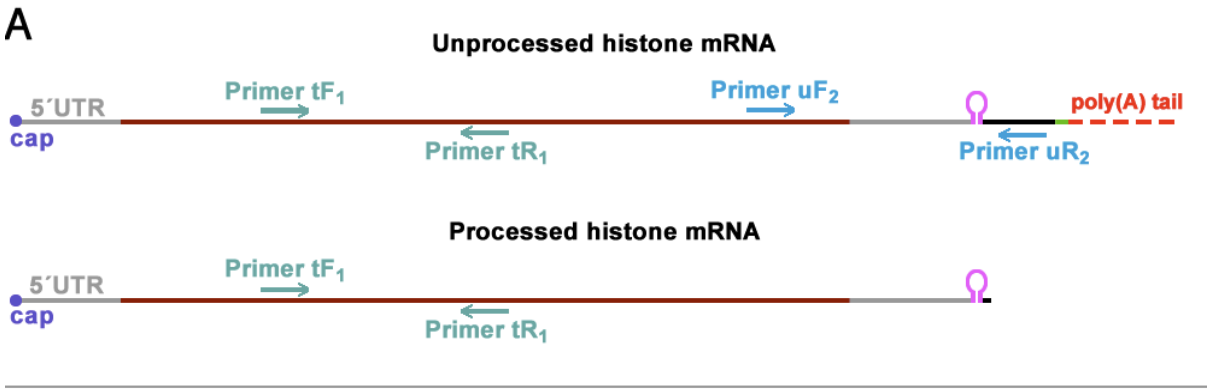
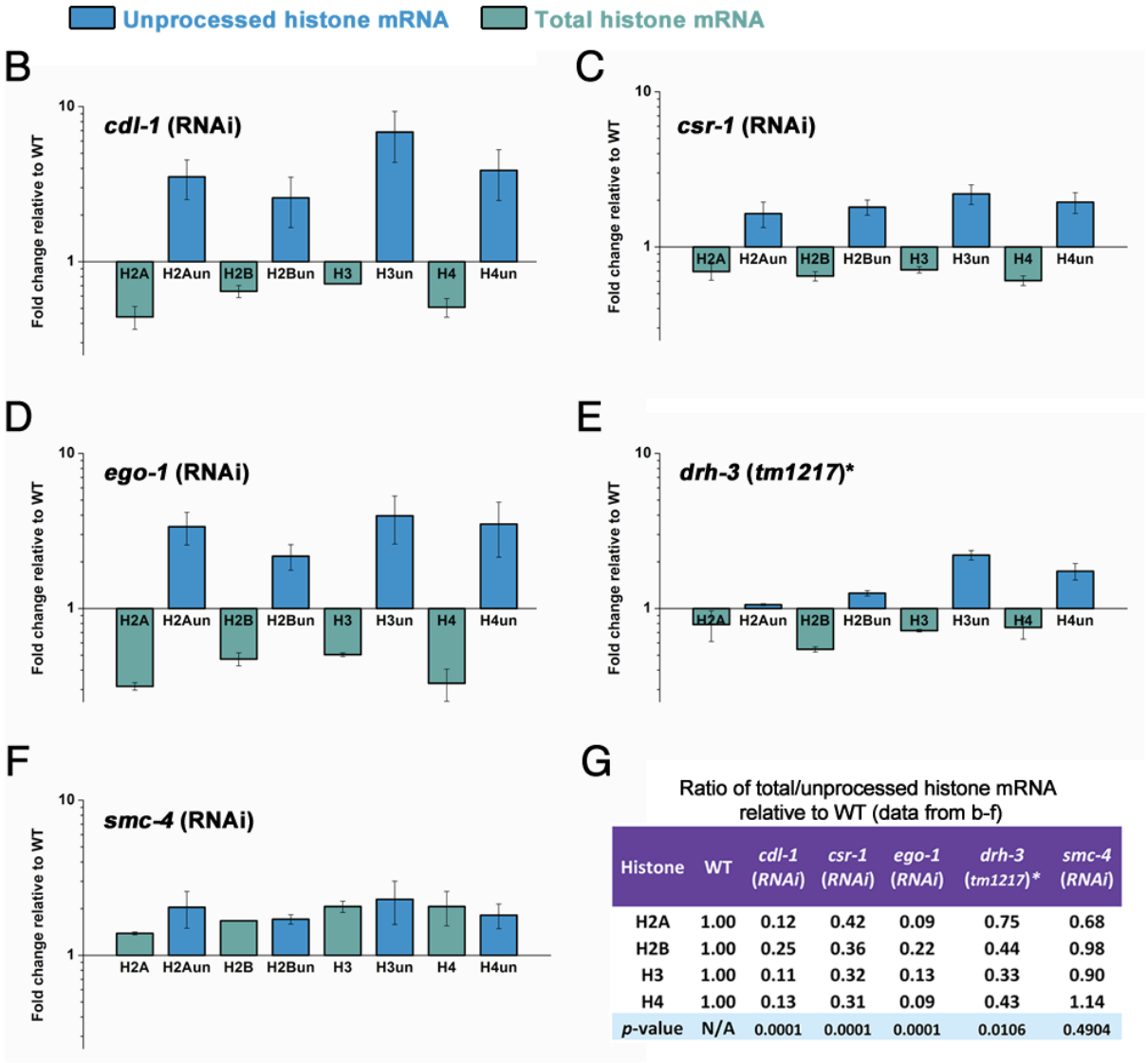


Figure 3: CSR-1 shows a nuclear localization in maturing oocytes, similarly to CDL-1::GFP (A) CDL-1::GFP is expressed in the germline and early embryos. The arrows show embryos in the first two cell divisions while the arrowhead indicates the vulva. The grey dotted box shows the enlarged region in the top panel to better visualize expression in the nuclei of maturing oocytes. (B) Immunostaining of dissected adult worm gonads with anti-CSR-1 antibodies (pink) (Aoki et al, 2007) and DNA (DAPI, blue). The signal is strongest in the maturing oocytes. CSR-1 immunostaining gives no signal in *csr-1(tm892)* mutant germlines. All scale bars represent 20µm.

Primer	Sequence
H2A total forward	TTCATCGTATTCTTCGCAAAGG
H2A total reverse	CAATTCGAGAACCTCAGCAG
H2A unprocessed forward	TCTTCCAAATATCCAAGCTG
H2A unprocessed reverse	CAATTGATAGTTAGGATTTCAGAT
H2B total forward	CCAAAGGACGGAAAGAAGAG
H2B total reverse	AAGACATCATTGACGAAGGAG
H2B unprocessed forward	GTTCGTTTGATTCTTCCAGGAG
H2B unprocessed reverse	CGAACATGTAGTTTTGTAAGGATTTC
H3 total forward	AAATCCACCGAACTTCTTATCC
H3 total reverse	CTTTGGCATAATGGTGACTC
H3 unprocessed forward	TTATGCCAAAGGATATCCAG
H3 unprocessed reverse	AATGGATTTTCAGTAAATTGTG
H4 total forward	GGTTCTCCGTGATAACATCC
H4 total reverse	GGTTTCCTCGTAGATCAATCC
H4 unprocessed forward	AAGAACTCTGTACGGATTCG
H4 unprocessed reverse	GCAAAGAAAAATTAGGATTTTGG
H2A transgene forward	TCTTCCAAATATCCAAGCTG
H2A transgene reverse	GGATACGCTAACAACTTGG
H2A transgene reverse (PAS)	TTGCATGCCTGCAGGTCGACAT

Table 1: Primers used for RT-qPCR





G

Ratio of total/unprocessed histone mRNA relative to WT (data from b-f)

Histone	WT	*cdl-1* (*RNAi*)	*csr-1* (*RNAi*)	*ego-1* (*RNAi*)	*drh-3* (*tm1217*)*	*smc-4* (*RNAi*)
H2A	1.00	0.12	0.42	0.09	0.75	0.68
H2B	1.00	0.25	0.36	0.22	0.44	0.98
H3	1.00	0.11	0.32	0.13	0.33	0.90
H4	1.00	0.13	0.31	0.09	0.43	1.14
p-value	N/A	0.0001	0.0001	0.0001	0.0106	0.4904

Figure 4: Knockdown of CSR-1 RNAi pathway components results in the distinct signature of core histone mRNA misprocessing. (A) Schematic diagram of unprocessed and processed histone mRNAs showing the primers used to detect total (unprocessed and processed) mRNA (tF_1 and tR_1) and unprocessed mRNA only (uF_2 and uR_2) by RT-qPCR in young adult worms. Primer sequences are provided in Table 1. (B-E) Total mRNA (green) of core histones decreased and unprocessed mRNA (blue) increased in *cdl-1, csr-1* and *ego-1* RNAi and *drh-3*(*tm1217*) as compared to appropriate wild-type controls. (F) *smc-4*(*RNAi*) showed no decrease in total histone mRNA. (G) The ratio of the change in total to the change in unprocessed histone mRNA levels for all core histones was significantly lower in *cdl-1* ($p < 0.0001$), *csr-1* ($p < 0.0001$) and *ego-1* ($p < 0.0001$) RNAi and *drh-3*(*tm1217*) ($p = 0.01$) when compared to wild-type. This effect was not seen in *smc-4*(*RNAi*) ($p = 0.49$). Error bars represent the standard deviation from the mean from three independent experiments.

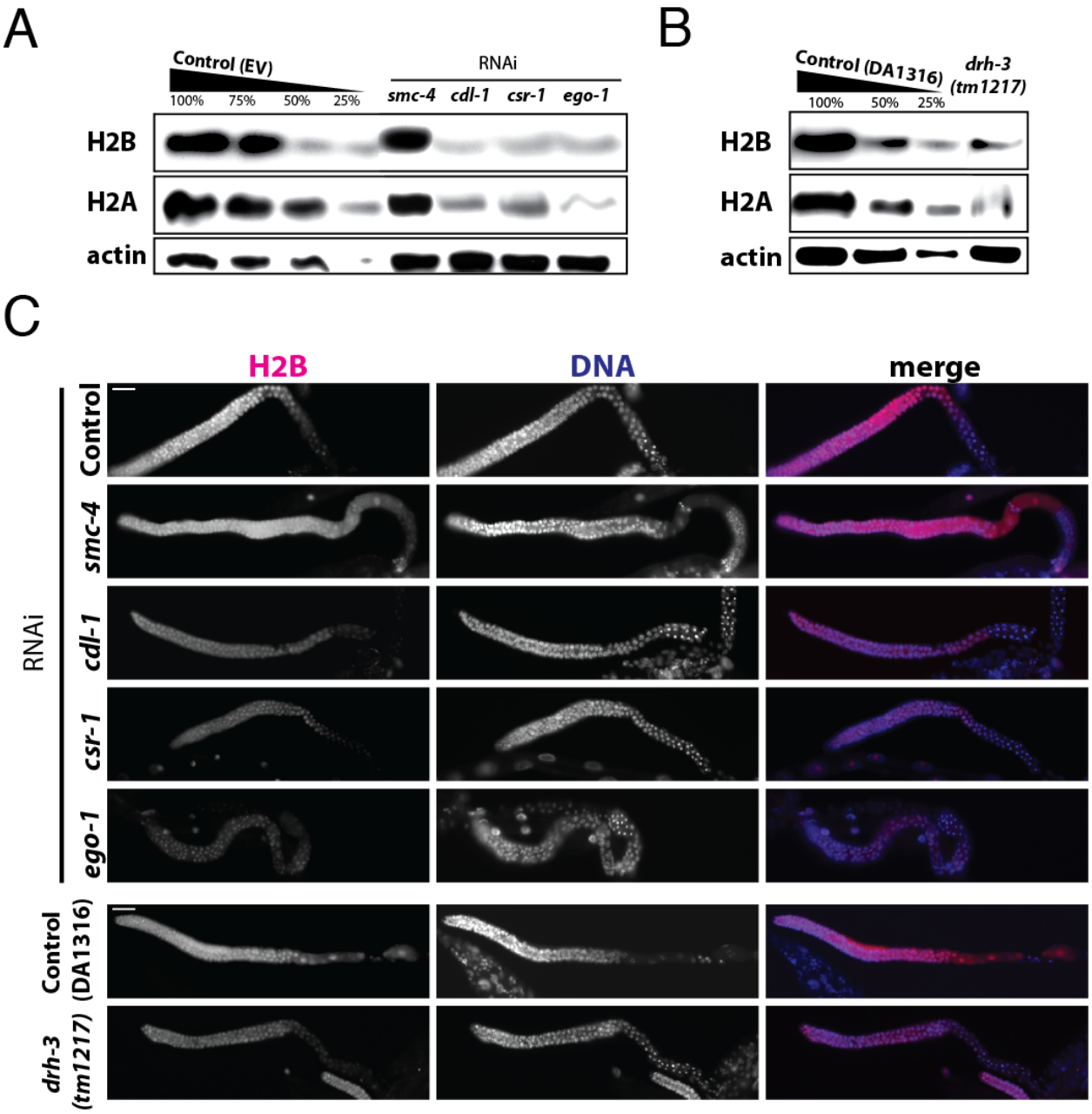


Figure 5: Knockdown of CSR-1 RNAi pathway components results in severe depletion in H2A and H2B proteins. (A) Western blot results with total young adult worm lysates: core histones H2A and H2B showed severe depletion in *cdl-1, csr-1* and *ego-1* RNAi, but not in *smc-4* RNAi when compared to empty vector control (EV). Actin is shown as a loading control. (B) Same as (a) for the *drh-3*(*tm1217*) null mutant with its congenic wild type (DA1316) as the

control. (C) Histone H2B immunostaining of dissected gonads from young adults (pink) shown with DNA (DAPI, blue). H2B was depleted in *cdl-1, csr-1* and *ego-1* RNAi, but not in *smc-4* RNAi, consistent with Western blotting. Similarly, H2B was depleted in *drh-3*(*tm1217*), but not in the control (DA1316). Scale bar is 20µm.

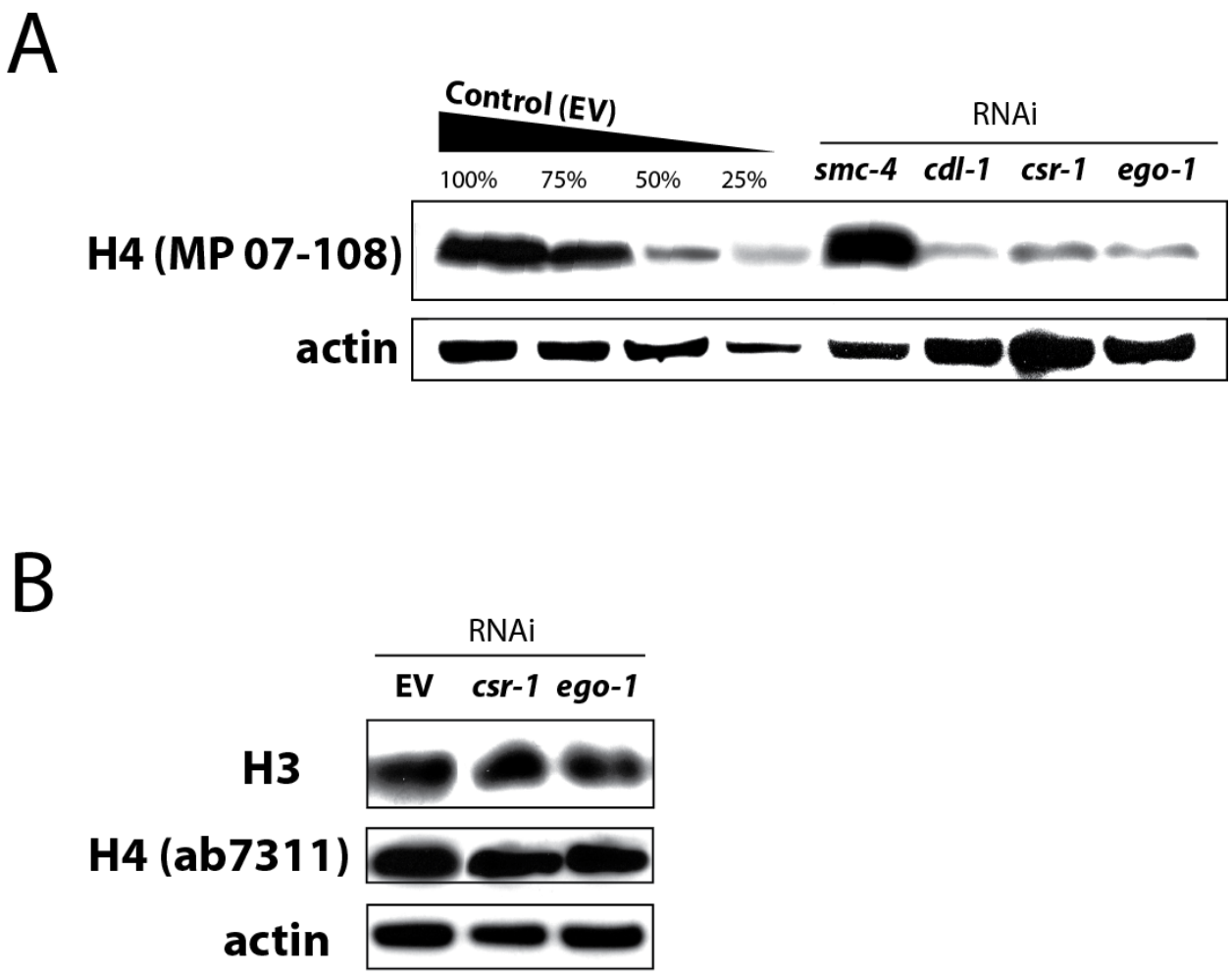


Figure 6: Anti-H4 antibodies have varying sensitivities. (A) Western blot analysis of core histone H4 in young adult total worm lysates showing 50% depletion of H4 in *cdl-1, csr-1* and *ego-1* RNAi compared to control, empty vector (EV), or *smc-1(RNAi)* using Millipore antibody 07-108. (B) The same samples are analyzed by Western blot against core histones H3 and H4 showing no apparent depletion of H4 with Abcam antibody ab7311.

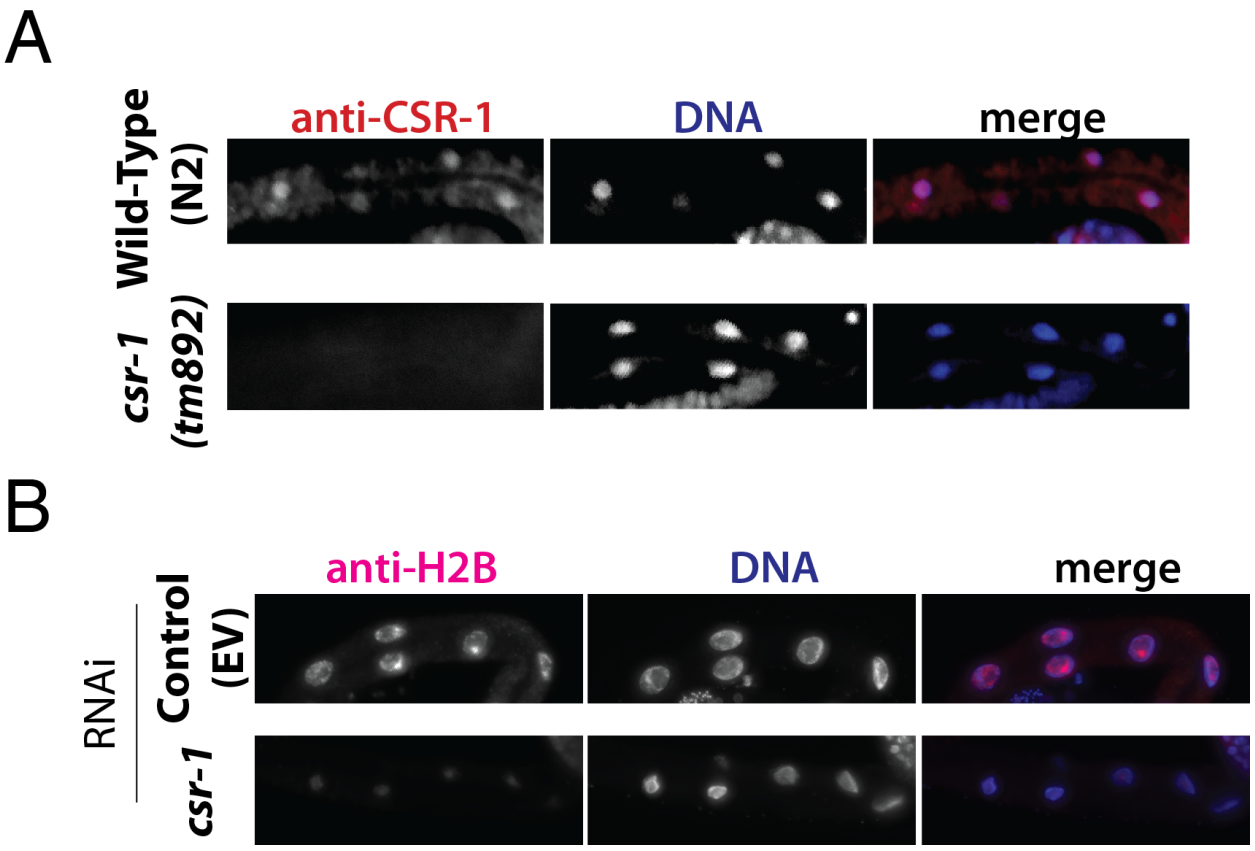


Figure 7: CSR-1 is expressed and affects histone production in gut cells. (A) CSR-1 is expressed in intestinal nuclei as seen by immunostaining with anti-CSR-1 antibodies (Aoki et al, 2007) in wild type worms, but not in *csr-1(tm892)* sterile adults. (B) Histone H2B immunostaining of dissected intestinal nuclei from young adults (pink) shown with DNA (DAPI, blue). H2B is depleted in *csr-1(RNAi)* compared to empty vector control in somatic cells as well as the germline cells shown in Figure 5.

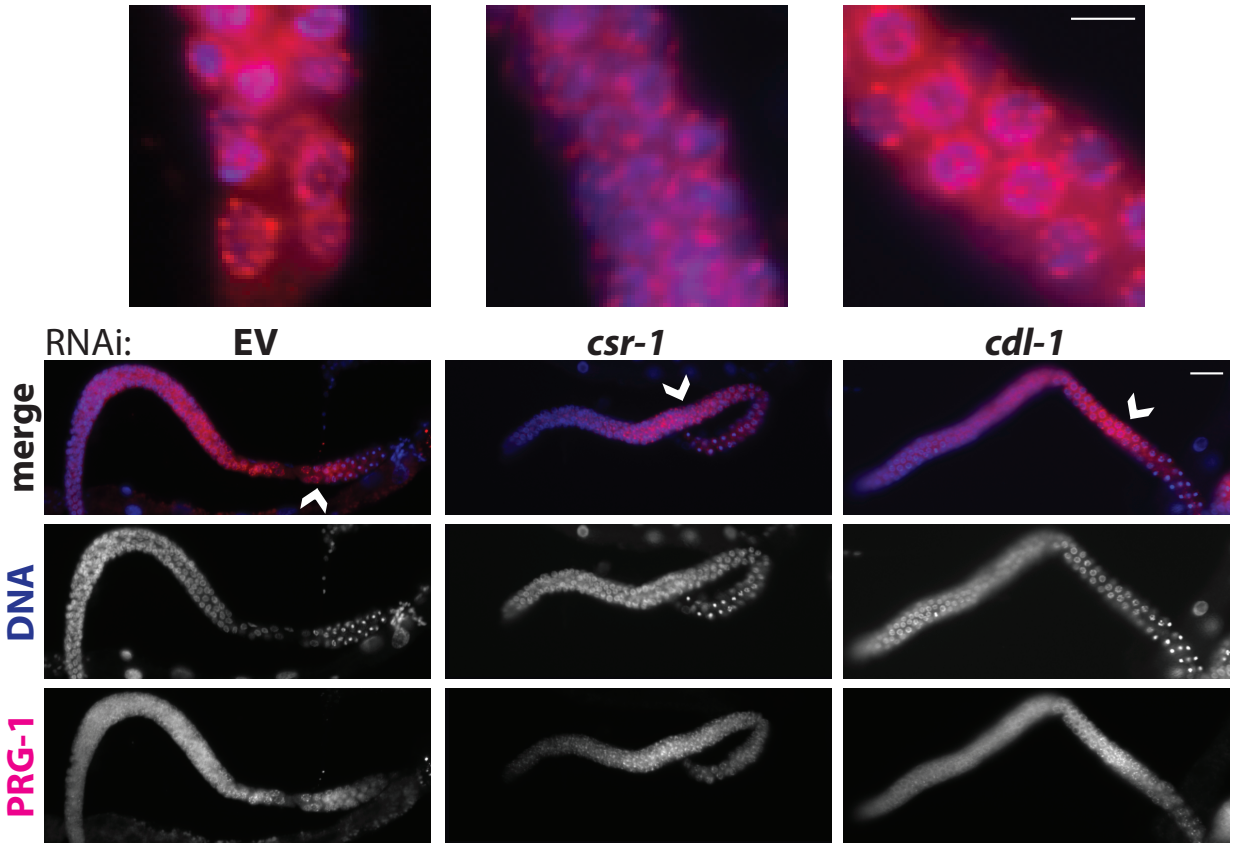


Figure 8: *csr-1(RNAi)* P-granule defects are not related to histone misregulation. PRG-1 immunostaining of dissected gonads from young adults (pink) shown with DNA (DAPI, blue) to visualize P-granule morphology in *csr-1* and *cdl-1* RNAi (anti-PRG-1 antibodies from Batista et al, 2008). Knockdown of *cdl-1*(*RNAi*) does not affect P-granule morphology despite sterility, indicating that the P-granule phenotype in *csr-1* is independent of the histone processing defect. Scale bar is 20 µm. Arrowheads indicate enlarged regions, scale bar is 5µm.

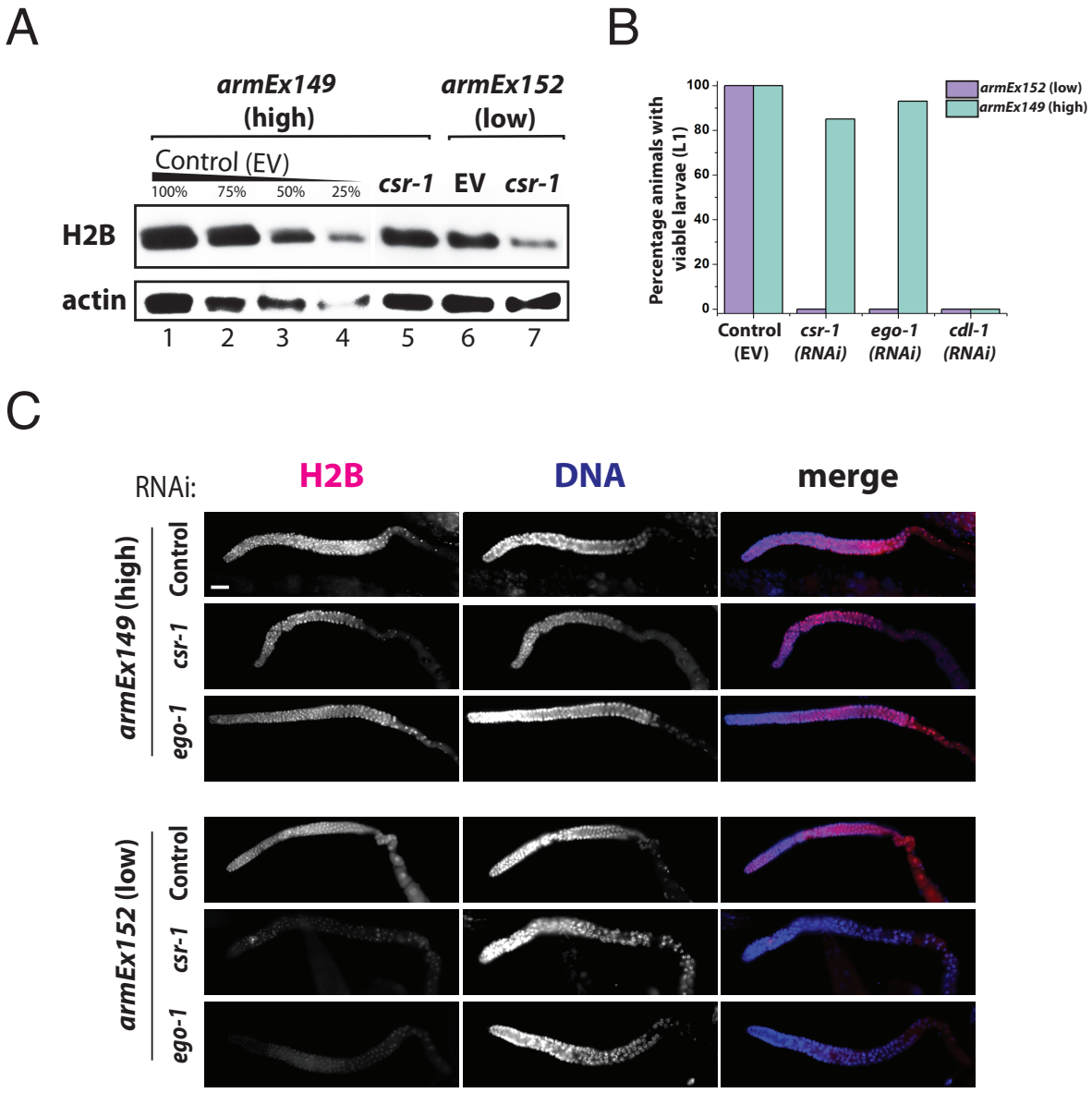
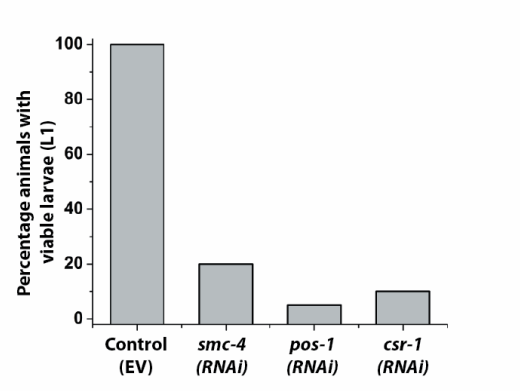


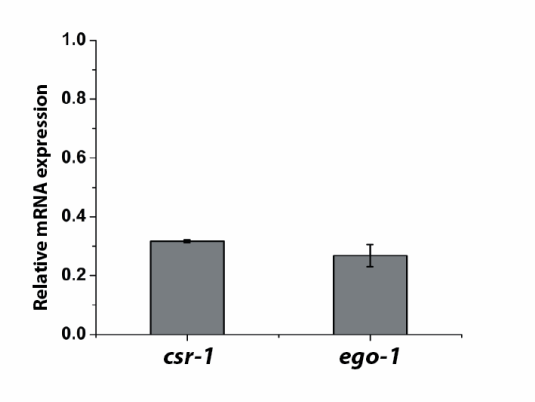
Figure 9: Expression of core histones from a transgenic array rescues the sterility phenotype arising from the knockdown of CSR-1 RNAi pathway components. (A) Western blot results with young adult worm lysates: total histone H2B protein levels were reduced in both the low-expressing transgenic line *armEx152* and high-expressing line *armEx149* upon *csr-1(RNAi)* compared to control RNAi; however total H2B levels were higher in the *armEx149*, accounting for a considerable persistence of H2B expression in *csr-1(RNAi)*-treated animals.

Actin is shown as a loading control. (B) Knockdown of *csr-1* or *ego-1* by RNAi, but not the control treatment (empty vector, EV), caused lethality in the progeny of the treated worms in the low-expressing transgenic line *armEx152* (purple) (n>60). However, in the high-expressing line *armEx149* (green), knockdown of *csr-1* or *ego-1* by RNAi did not cause lethality since greater than 80% of RNAi-treated hermaphrodites produced viable broods (n>60, p< 0.0001). Knockdown of *cdl-1* by RNAi in either transgenic line resulted in progeny lethality (n>60). Combined results of three independent experiments are shown. (C) Histone H2B immunostaining of dissected gonads from young adults (pink) shown with DNA (DAPI, blue). H2B is not depleted in *csr-1(RNAi)* or *ego-1(RNAi)* in *armEx149* (high transgene expression) compared to empty vector control, but is depleted in *csr-1(RNAi)* and *ego-1(RNAi)* in *armEx152* (low transgene expression), consistent with the Western blotting shown in Figure 6A. Scale bar is 20μm.

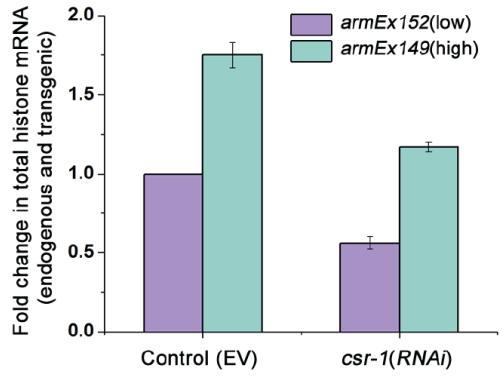
A


Percentage animals with viable larvae (L1)
100
80
60
40
20
0
Control (EV)
smc-4 (RNAi)
pos-1 (RNAi)
csr-1 (RNAi)

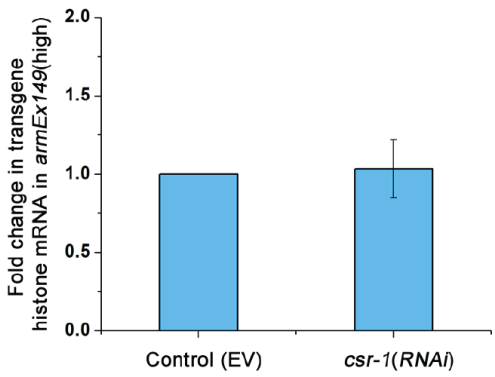
B


Relative mRNA expression
1.0
0.8
0.6
0.4
0.2
0.0
csr-1
ego-1

C


Fold change in total histone mRNA (endogenous and transgenic)
armEx152(low)
armEx149(high)
2.0
1.5
1.0
0.5
0.0
Control (EV)
csr-1(RNAi)

D


Fold change in transgene histone mRNA in armEx149(high)
2.0
1.5
1.0
0.5
0.0
Control (EV)
csr-1(RNAi)

E

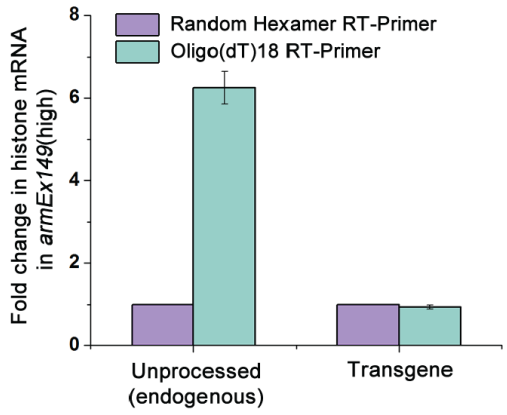

Fold change in histone mRNA in armEx149(high)
Random Hexamer RT-Primer
Oligo(dT)18 RT-Primer
8
6
4
2
0
Unprocessed (endogenous)
Transgene

Figure 10: Expression analysis and functionality of RNAi in *armEx149* transgenic histone line. (A) Knockdown of *pos-1* and *smc-4* results in sterility in armEx149 indicating that RNAi is functional in the germline of these lines. (B) *csr-1* and *ego-1(RNAi)* lead to 70% reduction in mRNA in armEx149 indicating that rescue is not due to inefficient knockdown of the RNAi targets. (C) Histone mRNA quantifications in (C-E) are shown for histone H2A. Unprocessed (endogenous) histone mRNA in *armEx149* showed ~ 6-fold enrichment when reverse transcription reaction was performed using oligo(dT) 18-mers (green) selective for polyadenylated messages compared to random hexamer primers (purple). Transgenic histone mRNA did not show enrichment in the oligo(dT) reverse transcription reaction as it lacks the poly(A) signal. (D) Total histone mRNA levels were reduced upon *csr-1*(RNAi) in both *armEx152* and *armEx149* strains compared to control RNAi. (E) No significant difference in transgenic histone mRNA was observed in *armEx149* upon *csr-1*(RNAi) compared to control RNAi.

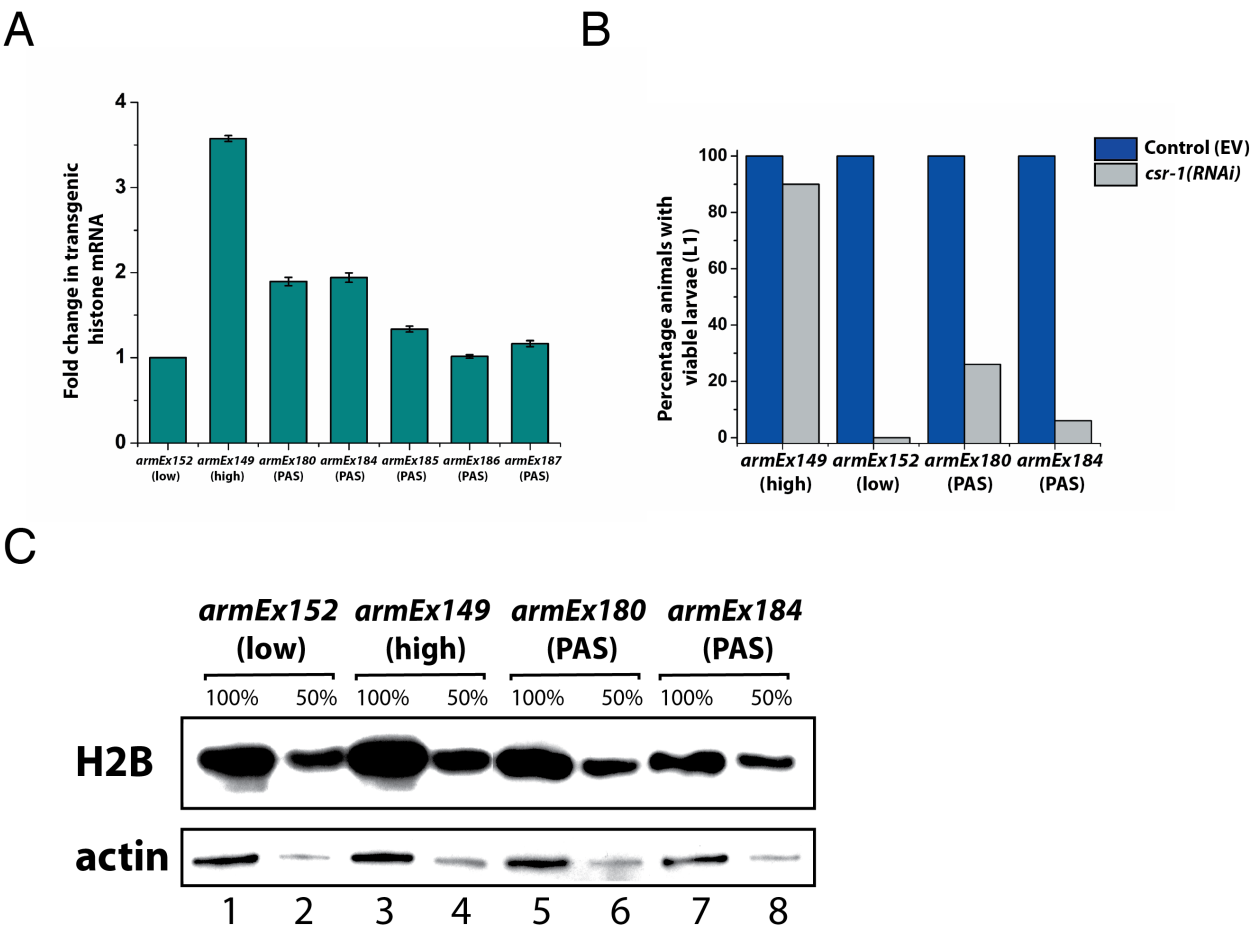


Figure 11: Histone transgenes containing the PAS sequence and no siRNA sites between the stem-loop and PAS do not produce high levels of histone proteins and do not rescue *csr-1(RNAi)* lethality. (A) Transgenic lines containing the PAS signal express transgenic histone mRNA. PAS containing lines *armEx180*, *armEx184, armEx185, armEx186,* and *armEx187* express transgenic histone at varying levels. *armEx180* and *armEx184* express transgenic histone ~2-fold higher than *armEx152.* Primers used to detect the transgenic histone mRNA were specific and did not produce any signal in the control strain without the transgenic array (NOTE: This primer is different than the one used to quantify transgenic histone in Figure 10, see Table 1 and Materials and Methods). Histone mRNA quantification is shown for histone H2A. (B) Knockdown of *csr-1* by RNAi fails to cause lethality in *armEx149* (>80% viability, as

in Figure 9B) but leads to lethality in *armEx152* (0% as in 9B). Introduction of the PAS sequence results in a significant decrease in viability after *csr-1(RNAi)* in *armEx180* (26%, p value < 0.001) and *armEx184* (6%, p value < 0.001) compared to *armEx149*. (B) Western blot analysis of young adult worm lysates for *armEx149*, *armEx152*, *armEx180*, and *armEx184,* showing an increase in total H2B in *armEx149* as compared to the other three. 100% and 50% amounts are shown to give relative abundance. Actin is shown as a loading control.

Author Contributions: The experiments presented in this chapter were conceived and designed by D.A., S.P., and A.G. The bioinformatics were performed by S.P. and both S.P. and Y.S. performed RT-qPCR analysis and histone transgene construction. All other experiments were carried out and analyzed by D.A.

Materials and Methods

C. elegans Strains

Worms were cultured according to (Brenner, 1974). Strains used in this study: **GR1373:** *eri-1*(*mg366*) IV, **WM216:** *avr-14(ad1302) drh-3(tm1217)* I/*hT2 [bli-4(e937) let-?(q782) qIs48(myo-2::gfp; pes-10::gfp; ges-1::gfp)]* (I; III); *avr-15(ad1051) glc-1(pk54)* V, **DA1316**: *avr-14(ad1302)* I*; avr-15(ad1051) glc-1(pk54)* V, **AGK482**: *unc-119*(*ed3*) III; *eri-1*(*mg366*) IV, **AGK458**: *unc-119*(*ed3*) III; *eri-1*(*mg366*) IV; *armEx149*(*his-59,60,61,62*; *unc-119*(+)), **AGK461**: *unc-119*(*ed3*) III; *eri-1*(*mg366*) IV; armEx152(*his-59,60,61,62*; *unc-119*(+)), **AGK516:** *unc-119*(*ed3*) III; *eri-1*(*mg366*) IV; armEx180(*his-59,60,61(PAS),62(PAS)*; *unc-119*(+)), **AGK520:** *unc-119*(*ed3*) III; *eri-1*(*mg366*) IV; armEx184(*his-59,60,61(PAS),62(PAS)*; *unc-119*(+)), **AGK537:** *unc-119*(*ed3*) III; armEx199(*pcdl-1*::CDL-1::GFP; *unc-119(+)).*

Bioinformatics

Gene expression profile for all genes, CSR-1 targets and core histones in *csr-1*(*tm892*) compared to wild-type (Figure 1A-C) was constructed from the microarray data by Claycomb *et al*., 2009 (Claycomb et al, 2009). CSR-1 targets (identified in Claycomb *et al., 2009*) (Claycomb et al, 2009) were classified based on endo-siRNA enrichment in CSR-1-IP small RNA library compared to wild-type input library (Figure 1D). Endo-siRNA sequences used in Figure 1F are from Claycomb *et al.,* 2009 (Claycomb et al, 2009).

RNA interference

RNAi feeding experiments were all done at 20°C in *eri-1*(*mg366*) sensitized background. RNAi was done by feeding on large NGM plates (150mm diameter) in order to obtain sufficient amount of worms for RNA extraction, staining and Western blotting from the same experiment. We used the following clones from the *C. elegans* feeding library (Kamath et al, 2001): *cdl-1*

(JA: R06F6.1), *csr-1* (JA: F20D12.1), *ego-1* (JA: F26A3.3), *smc-4* (JA: F35G12.8), and *pos-1* (JA:F52E1.1). Bacteria containing each RNAi clone were cultured in Luria Broth (LB) containing 50µg/mL ampicillin at 37°C for 8-14 hours in 350mL. The cultures were then spun at 4000rpm (Beckman J2-21) and resuspended in 3.5mL of LB. 1mL of the concentrated RNAi food was then spread on a large NGM plate containing 1 mM IPTG and 50 µg/mL ampicillin and left to dry. To perform the RNAi, approximately 20,000 synchronized L1 worms were placed on each plate of the RNAi media and allowed to grow for 60-70 hrs at 20^0C. Worms were collected as young adults prior to the formation of eggs for each type of analysis: Western blotting, RNA extraction and immunostaining.

RNA extraction and Quantitative RT-PCR

A phenol-chloroform extraction and ethanol-precipitation were performed to extract RNA using TRI Reagent (MRC), followed by a DNase (Ambion) treatment at 37°C for 1 hr. cDNA was prepared using RevertAid (Fermentas) reverse transcriptase, oligo dTs and random hexamers at 42°C for 1 hr. qPCR was performed on cDNA on the Eppendorf Mastercycler ep realplex thermal cycler using Quantifast SYBR Green PCR Kit (Qiagen). Thermocycling was done for 50 cycles in a 25 µl reaction containing 12.5 µl SYBR Green, 0.15 µl of 100 µM forward and reverse primers (Table 1), 5 µl cDNA and 7.2 µl dH2O. Primers were designed to amplify either processed or unprocessed cDNA based on the respective 3'UTRs for each histone type. Each experiment was performed in triplicate for each biological replica using Actin as a normalizing standard. mRNA expression levels were analyzed using the comparative Ct method.

Western Blotting

Synchronized worm populations were sonicated in 1M TSE solution to ensure extraction of all histone proteins at 20% output for 30 seconds on, 30 seconds off for eight rounds using the

Branson 500 Sonic Dismembrator. Samples were then centrifuged to remove cell debris at 4°C, 12000 rpm (Eppendorf AG 5424) and total protein content was quantified using the Bradford Dye Reagent (BioRad). Equal amounts of total lysate for each sample were then resolved via SDS-PAGE on a 12% gel (Invitrogen NuPAGE Bis-Tris) and transferred to a 2μm nitrocellulose membrane (Thermo Scientific) at 100mA for 30 minutes. The membrane was blocked using 3% (w/v) BSA in PBS with 0.01% Tween for 1 hour, followed by overnight incubation at 4°C with one of the following antibodies: anti-H2B (abcam ab1790), anti-H2A (abcam ab13923), anti-H3 (Millipore 05-928), anti-H4 (Millipore 07-108 or abcam ab7311), anti-CSR-1 (Claycomb et al, 2009) and anti-actin (Millipore MAB1501R) diluted 1:2000 in PBST-3% BSA. The membrane was then washed 3 times with PBST and incubated for 1 hour at room temperature with either anti-rabbit or anti-mouse HRP-conjugated secondary antibody (Perkin Elmer) diluted 1:5000 in PBST-3% BSA and visualized by SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) using a Series 2000A Film Processor (TIBA). Results shown are representative of at least three biological replicates.

Immunostaining

Young adult worms were hand picked and placed in PBS (137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4) with 0.01% Tween and 25mM sodium azide in a glass petri dish. Worms were dissected using a scalpel and fixed with 2% paraformaldehyde for 1 hour at room temperature and post-fixed in ice-cold methanol for 5 minutes. Worms were transferred to glass tubes and blocked with 3% (w/v) BSA in PBS. Blocked worms were then incubated overnight at 4°C with anti-H2B antibody (abcam ab1790) diluted 1:8000 or anti-PRG-1 (Batista et al, 2008) diluted 1:8000 or anti-CSR-1 (Aoki et al, 2007) diluted 1:4000, washed 3 times with PBST and incubated for 1 hour at room temperature with anti-rabbit Alexa Fluor 555 (Molecular Probes)

diluted 1:300. Worms were then mounted on polylysine slides (LabScientific, Inc) using Mounting solution with DAPI (Vector Laboratories) and all images were taken at 20x on Zeiss AxioImager Z1 with 50ms exposure for red channels, and 10ms exposure for blue channels. Results shown are representative of at least three biological replicates.

Transgene plasmid construction

DNA sequence containing the promoter, coding region and 3'UTR (including the stem-loop) of *his-61* and *his-62* genes was obtained from the PCR performed on *C. elegans* N2 genomic DNA with

5'seqGCCGCGGTCGACATAATTGTGGCCCTAAAGAGGGCCGTTGGGTTCGGTTGGGTT TTTAAGAAAAAAGTTC forward primer and

5'seqGATCAAGGATCCATATTTGTGGCCCTAAAGAGGGCCGTTGGGTTCGGTTGGTTA GTTTTCAGCCAATGGC reverse primer.

pPD95.75 (Addgene) plasmid and the *his-61,62* PCR product were digested with *BamHI* and *SalI* restriction enzymes and ligated to form pPD95.75w/*his-61,62* plasmid. DNA sequence containing the promoter, coding region and 3'UTR (including the stem-loop) of *his-59* and *his-60* genes was obtained from the PCR performed on *C. elegans* N2 genomic DNA with

5'seqGCCGCCGGTACCATATTTGTGGCCCATAAGAGGGCCGTTGGGTTCGGTAATTGG TTTTGATGAAAACAG forward primer and

5'seqAAGAGTGGGCCCTAAATTGTGGCCCTAAAGAGGGCCGTTGGGTTCGGTGAGTTT TGAGTTGAAGCTTAC reverse primer.

pPD95.75w/*his-61,62* plasmid and the *his-59,60* PCR product were digested with *KpnI* and *ApaI* restriction enzymes and ligated to form the final plasmid pPD95.75w/*his-59,60,61,62*.

To introduce the PAS sequence into the above plasmid, the QuikChange Site-Directed Mutagenesis Kit (Stratagene) was used with the following primer pairs:

5'CAACTTGGAAATGAAATTTTATTGCATGCCTGCAGGTCG and

5'GTTGAACCTTTACTTTAAAATAACGTACGGACGTCCAGC for *his-61*

5'CCCCGGGATTGGCCAAAGGAAATAAAGGTATGTTTCGAATG and

5'GGGGCCCTAACCGGTTTCCTTTATTTCCATACAAAGCTTAC for *his-62*

pPD95.75w/*his-59,60,61,62* or pPD95.75w/*his-59,60,61*(PAS)*,62*(PAS) were injected into AGK482: *eri-1*(*mg366*)IV; *unc-119*(*ed3*)III at 1 ng/µl, together with N2 genomic DNA (30 ng/µl) and pMM016B [*unc-119*(+)] (1 ng/µl). Transgenic lines were generated based on the rescue of *unc-119* mutant phenotype. In the rescued lines, transgene histone expression was quantified by qRT-PCR.

CDL-1::GFP was also created using the pPD95.75 (Addgene) plasmid. The endogenous promoter and coding region were amplified from N2 genomic DNA using the following primers:

5'CATCCCGGGGACGTCCATCTCAACTCGCAATC forward

5'GTAGGTACCCTCTTGTCATCGTCATCTTTATAATCGTGCGACGACATCTTGGAGAA

G reverse, including a FLAG-tag in frame.

The 3.5kb insert was then digested with *SmaI* and *KpnI* restriction enzymes and ligated into a similarly digested pPD95.75. The resultant CDL-1::GFP plasmid was co-bombarded with the 112 plasmid pMM016b (AddGene) for *unc-119(ed3)*III rescue using microparticle bombardment 113 with a PDS-1000 Hepta Apparatus (Bio-Rad). Transgenic lines were selected by *unc-119* phenotypic rescue. The germline expressing transgenic line is called AGK537, expressing transgene *armEx199*.

Statistics

Statistical significance between all genes and CSR-1 targets in Fig. 1A,B was calculated based on the Z-test of two-proportion. Statistical significance for the ratio of the change in total to the change in unprocessed histone mRNA levels for all core histones relative to wild-type in Figure 3G was performed based on one sample t-test with the hypothetical mean of 1. Statistical significance in Figure 5 was calculated based on z-test of two proportions.

2′-O-methyl RNA Pull Down Experiments

Pull down experiments were conducted according to (Hutvágner et al, 2004). The only modification to the existing protocol was to incubate two separate antisense oligonucleotides for the pull down of the mRNA as opposed to a direct complement to the desired small RNA itself. The results were analyzed by Western blot. The following oligonucleotides were used:

H2A UTR: 5'GGCCACAAUUAUCUGAAAUCCUAACUAUCAAUUG

Unrelated sequence: 5'UCUUUACGCUAACAACUUGGAAAUGAAAUAAGCU

H2A antisense 1: 5'UCUUGGCUUUGCCUCCCUUUCCACGUCCAGACAU

H2A antisense 2: 5'AUUGGAGUCCGGCUCUUGAUGAGCGGGACUUGGC

References

Allard P, Yang Q, Marzluff WF, Clarke HJ (2005) The stem-loop binding protein regulates translation of histone mRNA during mammalian oogenesis. *Dev Biol* **286:** 195-206

Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D (2003) MicroRNAs and other tiny endogenous RNAs in C. elegans. *Curr Biol* **13:** 807-818

Aoki K, Moriguchi H, Yoshioka T, Okawa K, Tabara H (2007) In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. *Embo J* **26:** 5007-5019

Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, Kasschau KD, Chaves DA, Gu W, Vasale JJ, Duan S, Conte D, Jr., Luo S, Schroth GP, Carrington JC, Bartel DP, Mello CC (2008) PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. *Molecular cell* **31:** 67-78

Brenner S (1974) The genetics of Caenorhabditis elegans. *Genetics* **77:** 71-94

Claycomb JM, Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, Conte D, Jr., Mello CC (2009) The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139:** 123-134

Davila Lopez M, Samuelsson T (2008) Early evolution of histone mRNA 3' end processing. *RNA* **14:** 1-10

Duchaine TF, Wohlschlegel JA, Kennedy S, Bei Y, Conte D, Jr., Pang K, Brownell DR, Harding S, Mitani S, Ruvkun G, Yates JR, 3rd, Mello CC (2006) Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. *Cell* **124:** 343-354

Godfrey AC, Kupsco JM, Burch BD, Zimmerman RM, Dominski Z, Marzluff WF, Duronio RJ (2006) U7 snRNA mutations in Drosophila block histone pre-mRNA processing and disrupt oogenesis. *RNA* **12:** 396-409

Gorgoni B, Andrews S, Schaller A, Schumperli D, Gray NK, Muller B (2005) The stem-loop binding protein stimulates histone translation at an early step in the initiation pathway. *RNA* **11:** 1030-1042

Grishok A, Hoersch S, Sharp PA (2008) RNA interference and retinoblastoma-related genes are required for repression of endogenous siRNA targets in Caenorhabditis elegans. *Proc Natl Acad Sci U S A* **105:** 20386-20391

Gu W, Shirayama M, Conte D, Jr., Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, Chen CC, Chaves DA, Duan S, Kasschau KD, Fahlgren N, Yates JR, 3rd, Mitani S, Carrington JC, Mello CC (2009) Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. *Mol Cell* **36:** 231-244

Hagstrom KA, Holmes VF, Cozzarelli NR, Meyer BJ (2002) C. elegans condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. *Genes Dev* **16:** 729-742

Halic M, Moazed D (2009) 22G-RNAs in transposon silencing and centromere function. *Mol Cell* **36:** 170-171

Hutvágner G, Simard MJ, Mello CC, Zamore PD (2004) Sequence-specific inhibition of small RNA function. *PLoS Biol* **2:** E98

Ideue T, Hino K, Kitao S, Yokoi T, Hirose T (2009) Efficient oligonucleotide-mediated degradation of nuclear noncoding RNAs in mammalian cultured cells. *RNA* **15:** 1578-1587

Jan CH, Friedman RC, Ruby JG, Bartel DP (2011) Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. *Nature* **469:** 97-101

Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J (2001) Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans. *Genome biology* **2:** RESEARCH0002

Keall R, Whitelaw S, Pettitt J, Muller B (2007) Histone gene expression and histone mRNA 3' end structure in Caenorhabditis elegans. *BMC Mol Biol* **8:** 51

Kimble J, Crittenden SL (2005) Germline proliferation and its control. *WormBook***:** 1-14

Kodama Y, Rothman JH, Sugimoto A, Yamamoto M (2002) The stem-loop binding protein CDL-1 is required for chromosome condensation, progression of cell death and morphogenesis in Caenorhabditis elegans. *Development* **129:** 187-196

Lejeune E, Allshire RC (2011) Common ground: small RNA programming and chromatin modifications. *Curr Opin Cell Biol* **23:** 258-265

Mangone M, Manoharan AP, Thierry-Mieg D, Thierry-Mieg J, Han T, Mackowiak SD, Mis E, Zegar C, Gutwein MR, Khivansara V, Attie O, Chen K, Salehi-Ashtiani K, Vidal M, Harkins TT, Bouffard P, Suzuki Y, Sugano S, Kohara Y, Rajewsky N, Piano F, Gunsalus KC, Kim JK (2010) The landscape of C. elegans 3'UTRs. *Science* **329:** 432-435

Maniar JM, Fire AZ (2011) EGO-1, a C. elegans RdRP, modulates gene expression via production of mRNA-templated short antisense RNAs. *Curr Biol* **21:** 449-459

Marzluff WF, Wagner EJ, Duronio RJ (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* **9:** 843-854

McGhee JD (2007) The C. elegans intestine. *WormBook : the online review of C elegans biology***:** 1-36

Nakamura M, Ando R, Nakazawa T, Yudazono T, Tsutsumi N, Hatanaka N, Ohgake T, Hanaoka F, Eki T (2007) Dicer-related drh-3 gene functions in germ-line development by maintenance of chromosomal integrity in Caenorhabditis elegans. *Genes Cells* **12:** 997-1010

Pak J, Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in C. elegans. *Science* **315:** 241-244

Pettitt J, Crombie C, Schumperli D, Muller B (2002) The Caenorhabditis elegans histone hairpin-binding protein is required for core histone gene expression and is essential for embryonic and postembryonic cell division. *J Cell Sci* **115:** 857-866

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell* **127:** 1193-1207

She X, Xu X, Fedotov A, Kelly WG, Maine EM (2009) Regulation of heterochromatin assembly on unpaired chromosomes during Caenorhabditis elegans meiosis by components of a small RNA-mediated pathway. *PLoS Genet* **5:** e1000624

Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17:** 761-772

Smardon A, Spoerke JM, Stacey SC, Klein ME, Mackin N, Maine EM (2000) EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. *Curr Biol* **10:** 169-178

Sonnichsen B, Koski LB, Walsh A, Marschall P, Neumann B, Brehm M, Alleaume AM, Artelt J, Bettencourt P, Cassin E, Hewitson M, Holz C, Khan M, Lazik S, Martin C, Nitzsche B, Ruer M, Stamford J, Winzi M, Heinkel R, Roder M, Finell J, Hantsch H, Jones SJ, Jones M, Piano F, Gunsalus KC, Oegema K, Gonczy P, Coulson A, Hyman AA, Echeverri CJ (2005) Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* **434:** 462-469

Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, McWhirter RD, Petersen S, Sreedharan VT, Widmer C, Jo J, Reinke V, Petrella L, Strome S, Von Stetina SE, Katz M, Shaham S, Ratsch G, Miller DM, 3rd (2011) A spatial and temporal map of C. elegans gene expression. *Genome Res* **21:** 325-341

Sullivan E, Santiago C, Parker ED, Dominski Z, Yang X, Lanzotti DJ, Ingledue TC, Marzluff WF, Duronio RJ (2001) Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. *Genes & development* **15:** 173-187

Thelie A, Pascal G, Angulo L, Perreau C, Papillier P, Dalbies-Tran R (2012) An oocyte-preferential histone mRNA stem-loop-binding protein like is expressed in several mammalian species. *Molecular reproduction and development*

Updike DL, Strome S (2009) A genomewide RNAi screen for genes that affect the stability, distribution and function of P granules in Caenorhabditis elegans. *Genetics* **183:** 1397-1419

Wagner EJ, Berkow A, Marzluff WF (2005) Expression of an RNAi-resistant SLBP restores proper S-phase progression. *Biochem Soc Trans* **33:** 471-473

Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC (2006) Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127:** 747-757

Chapter 4: Discussion and Future considerations

This chapter is divided into three main sections. The first part is an in depth discussion of the data presented in Chapter 2 on ZFP-1. Second, is a presentation of some preliminary work on the connection between ZFP-1 and RNAi in *C. elegans*. The third is a discussion of the work presented in Chapter 3 on the involvement of RNAi in histone production. My thesis work has contributed to the basic understanding of the molecular characteristics of ZFP-1 demonstrating that the PHD fingers of ZFP-1 are essential, and showed that RNAi factors are directly involved in promoting histone production in *C. elegans*. The latter contribution to the field is of notable importance as it has highlighted not only what is lacking in the published model, but also the significance of openness to new possibilities during experimental work.

Part 1:

ZFP-1 PHD1-PHD2 is an unusual type of zinc finger domain

Most PHD fingers, as described in Chapter 1, have specific binding capabilities for a methylation state of a lysine residue on the N-terminal tail of histone H3. There are many proteins that have multiple PHD finger domains within them, for example MLL has three PHD fingers (Mohan et al, 2010), however, they tend to be spread out within the protein. PHD1-PHD2 of ZFP-1 is unusual in that not only are the two PHD fingers close together, but there is a highly conserved linker region between them, as described in Chapter 2. We showed that ZFP-1 PHD1-PHD2 has specificity for binding methylated H3K4, and that this region is essential for viability in *C. elegans.*

Primary sequence alignment of ZFP-1 PHD1 with other PHD fingers highlights the conserved cysteine and histidine residues that are characteristic of the PHD finger motif (Figure 1 and Chapter 1 Figure 3, Chapter 2 Figure 1). To take a closer look at other conserved residues, we aligned ZFP-1 PHD1 with PHD fingers known to bind H3K4me3 compared to those that bind H3K4me0 (Figure 1). The residues that have been found to interact directly with the unmethylated K4 (Sanchez & Zhou, 2011) (Figure 1, blue highlights) are also somewhat conserved in ZFP-1 PHD1. Conversely, when comparing to PHD fingers that are known to bind H3K4me3, the residues that create the aromatic cage known to interact with the methyl groups on H3K4 are not conserved in ZFP-1 PHD1 (Figure 1, orange highlights). Therefore, although ZFP-1 PHD1 binds to H3K4me2, its primary sequence is more similar to PHD fingers that prefer unmethylated H3K4.

The methylation state of the second arginine residue on histone H3, R2, has also been known to affect the binding specificity of PHD fingers to H3K4 methylation (Di Lorenzo &

Bedford, 2011). For example, any type of methylation on R2 greatly reduces the binding affinity of the BPTF PHD finger to H3K4me3 (Li et al, 2006), whereas the PHD finger of PYGO is not affected by R2 methylation (Fiedler et al, 2008). We obtained a doubly-modified H3 tail peptide, H3R2meK4me2, from the Bedford lab (Iberg et al, 2007) and tested ZFP-1 PHD1 binding. The methylation state of R2 did not affect ZFP-1 PHD1 binding to H3K4me2. It is likely that the N-terminus of PHD1 is interacting with K4, and thus the R2 residue may not fit into a binding pocket as it does in BPTF. This is consistent with the primary sequence alignment of ZFP-1 PHD1 with other PHD fingers whose amino acids interacting with H3R2 are known (Figure 1, green highlights). The residues that have been shown to be sensitive to R2 methylation are not conserved in ZFP-1 PHD. However, we cannot exclude that ZFP-1 PHD1 will be more sensitive to R2 methylation in a biological context. Also, as has been seen for several other PHD fingers, it is possible that the binding of ZFP-1 to H3K4me2 is affected by another histone tail PTM that we have yet to uncover.

PHD1 of ZFP-1 is a canonical histone-binding PHD finger in that it has a specific binding preference for H3K4me2, as compared to PHD2, which has no apparent specificity for any histone tail PTMs. However, PHD1 is still unusual in that it clearly prefers the dimethylated state of H3K4. PHD fingers that are specific to H3K4 tend to bind either di- and trimethyalted H3K4 by forming an aromatic cage around the methyl groups, or unmethylated H3K4 by charge interactions with the ammonium group of K4. Thus, methylation-specific PHD fingers usually bind to trimethyalted lysines stronger than dimethylated lysines, which is not the case with PHD1 of ZFP-1. However, the specificity of PHD1 for H3K4me2 is lost in the context of PHD1-PHD2, suggesting that the chromatin binding properties of this double-PHD domain are more complex *in vivo*.

Double-PHD fingers have been described in a limited number of proteins. The most well-known are the double plant homeodomain finger proteins (DPF1-3), which contain two PHD fingers directly in tandem, with no linker region. The two PHD fingers in DPF proteins tend to have different and specific binding preferences; for example in DPF3, one PHD finger binds to H3K4me3, while the other binds to H3K14ac on the same histone tail (Ishizaka et al, 2012; Musselman & Kutateladze, 2011). This PHD finger arrangement differs from that in ZFP-1 in that there is no linker region between the two PHD fingers, and they both interact with specific modifications on the same histone tail. In ZFP-1, not only is there a linker region that is highly conserved, but also the second PHD finger is not specific to any histone tail modifications as far as we know. This lack of specificity of PHD2 points towards the possibility of a more structural role for PHD2, rather than specific histone tail binding. We also found that this region is responsible for the multimerization of ZFP-1 (Chapter 2 Figure 7), which may help explain why it does not have the ability to bind specific histone tail PTMs.

Given that PHD1 binds to H3K4me2 but PHD2 nonspecifically binds all types of histone tails, one could imagine a scenario where PHD2 acts as an anchor on the nucleosome and also allows for the homo-multimer of ZFP-1 proteins to come together at the chromatin. PHD1 could then serve to read the methylation state of H3K4 and subsequently promote downstream activity of an unknown binding partner. Since PHD1-PHD2 is able to bind nucleosomes *in vitro*, without any histone tail modifications, it seems likely that when no methylation is available the affinity of PHD2 for nucleosomes substitutes for the affinity of PHD1 for H3K4me2. This is also what we observe in the M4D mutation of PHD1 in the context of PHD1-PHD2 (Chapter 2, Figure 8). In this case, the mutation has blocked binding of PHD1(M4D)-PHD2 to H3K4me3, but now allows binding to umethylated H3K4. This is likely due to the non-specificity of PHD2, in

addition to the shifted specificity of PHD1. However, additional quantitative and mutational analyses of these associations are needed to reach definite conclusions. It would be interesting to test the affinity of PHD1-PHD2 for modified nucleosomes with additional mutations in PHD1 and PHD2.

One protein with similarity to PHD1-PHD2 of ZFP-1 is Jade-1, which contains a canonical PHD finger followed by an extended PHD finger (Zhou et al, 2004). The linker region specifically was shown to bind the tumor suppressor VHL, a binding that greatly stabilizes Jade-1 (Zhou et al, 2004). We observed that ZFP-1 PHD1-PHD2 is quite unstable *in vitro* and in cell lysates. It is very likely that PHD1-PHD2 also has a binding partner similar to VHL that greatly stabilizes it *in vivo*.

The family of Jumonji domain 2 (JMJD2)-containing histone demethylases also contains a two PHD finger motif with a similarly-spaced linker region. Interestingly, the linker region of this protein, as well as that of Jade-1, contains conserved histidine and cysteine residues (Figure 2, grey boxes). These residues are predicted to coordinate another zinc ion, as described in Chapter 2. The fact that a PHD1-PHD2 motif, with predicted zinc-binding residues within the linker region, is conserved across proteins with varying functions suggests the possibility that this motif has a specific role, which is more complex than simply PHD finger-binding to methylated lysine residues.

In addition to stabilizing PHD fingers, binding partners have also been shown to affect the specificity of PHD fingers for histone tail PTMs. For example, the PHD finger of PYGO has been found to bind H3K4me2, a binding that is greatly increased by the HD region of BCL9 (Fiedler et al, 2008; Miller et al, 2010). This was the first time another protein was shown to

affect the binding ability of a PHD finger. It is possible that such a binding partner may affect the specificity of ZFP-1 PHD2 *in vivo*, although further study is needed to identify such a factor.

Understanding the essential function of PHD1-PHD2

We have shown that the PHD fingers of ZFP-1 are preferentially expressed in the germline within the long isoform of ZFP-1 (Chapter 2 Figure 2). We have also shown that the PHD fingers of ZFP-1 are required maternally and zygotically (Chapter 2 Figure 6). This means that expression of the PHD fingers in the maternal germline and in the embryos is required for viable progeny to develop. However, we do not yet know the specific molecular function that leads to this requirement. Given that most PHD fingers are associated with enzymatic domains or complexes that carry out PTMs on histone tails, it is likely that some such complex is also involved in the essential role of ZFP-1. Our lab has found that ZFP-1 negatively regulates expression of target genes in somatic tissues through regulation of H2B ubiquitylation (Cecere et al, under review). This regulation seems to be independent of the PHD fingers since they are retained in the *zfp-1(ok554)* mutant, which was used in previous studies (Mansisidor et al, 2011; Cecere et al, under review).

To further understand the essential role of the PHD1-PHD2 region of ZFP-1 retained in the *zfp-1(ok554)* mutant, we attempted mutational analysis *in vivo*. Mutations in either PHD finger have been difficult for us to achieve due to the instability of the protein. The M4D mutation (Chapter 2 Figure 8B) shifts the specificity of PHD1 away from the methylated state. However, when we introduced this mutation in ZFP-1 on the pLONG plasmid for *in vivo* studies, the mutated ZFP-1 was still able to rescue the lethality caused by the haplo-insufficiency of *zfp-1(ok554)*. This suggests that the mutation was not severe enough to disrupt the essential

biological function of the protein and/or perhaps the dependence of ZFP-1 localization on H3K4me is not the only factor relevant to ZFP-1's essential role.

The binding affinity of ZFP-1 to chromatin may also be different depending on the different levels of chromatin compaction. One possibility for the molecular function of ZFP-1 may be that the PHD fingers form a type of scaffold that sets up a higher order chromatin state. This would be consistent with the fact that only small amounts of ZFP-1 seem to be required for viability. As shown in Chapter 2, knockdown of *zfp-1* by RNAi results in low penetrant germline defects in RNAi-treated animals (Chapter 2 Figure 5) although their progeny do survive. Loss of one zygotic copy of *zfp-1(ok554)*, in contrast, leads to lethality. This change in dosage suggests that there is likely to be a threshold amount of ZFP-1 needed at a specific time in development and that falling below this threshold leads to lethality.

H3K4me is a necessary component of the *C. elegans* germline maintenance and development (Schaner, 2006; Schaner et al, 2003). Currently, there is some debate over whether H3K4me affects transcription or is a consequence of transcription. RNA Polymerase II has been shown to recruit the COMPASS complex to sites of active transcription, suggesting that the methylation is deposited to mark where transcription has taken place (Eissenberg & Shilatifard, 2010; Rando & Chang, 2009). However, it has also been shown that H3K4me exists on active genes independent of RNA Polymerase II activity, uncoupling H3K4me from transcription (Gregory et al, 2007; Rando & Chang, 2009; Vastenhouw et al, 2010). For this reason, it is difficult to predict the function of H3K4me in the germline in general, and consequently the biological reason for ZFP-1 binding H3K4me in the germline remains unclear. H3K4me has been proposed to serve as a sort of 'epigenetic memory' mark across multiple generations in *Dictyostelium* (Muramoto et al, 2010), although the heritability of histone PTMs in general is

still debated (Ptashne, 2007; Rando & Chang, 2009; Rando & Verstrepen, 2007; Smith & Shilatifard, 2010; Taverna et al, 2007; van Steensel, 2011). Nevertheless, if the methylation state of H3K4 is heritable, this could be consistent with a model where ZFP-1 is involved in a higher-order chromatin structure involving H3K4me that is inherited and required for development.

ZFP-1, H3K4 methylation and regulation of target genes

We have found an enrichment of H3K4me on ZFP-1 targets that are germline specific when compared to targets genome-wide (83% versus 46%). This enrichment could suggest that H3K4me is more important for gene regulation by ZFP-1 in the germline compared to somatic tissues. It is important to note that the *zfp-1(ok554)* mutant still retains expression of the PHD1-PHD2 region in the germline (Chapter 2 Figure 2A). Therefore, to investigate the effect of loss of this portion, RNAi-mediated knockdown would need to be carried out in the *zfp-1(ok554)* mutant background to knock down the remaining portion.

We attempted expression analysis of selected germline-specific target genes after knockdown by RNAi of *zfp-1* in wild type, in *zfp-1(ok554)*, or in the sensitized *zfp-1(ok554);eri-1(mg366)* strain. Although 70-80% knockdown of *zfp-1* mRNA was achieved, as measured by qRT-PCR, the relative expression of 26 different target genes was extremely variable between experiments. There was no direct correlation between the knockdown of *zfp-1* and the target gene expression levels. We suspect that the variability of RNAi in each animal does not allow for a consistent knockdown of *zfp-1* that would lead to a clear understanding of gene expression regulation. Moreover, we do not know whether the level of knockdown of *zfp-1* mRNA correlates with a similar decrease in protein levels. Although we were able to detect the PHD1-PHD2 fragment expression in *zfp-1(ok554)* by immunofluorescence, our antibodies do not allow

its detection by Western making it difficult to estimate the level of protein knockdown after *zfp-1(RNAi)*. Furthermore, since *zfp-1* was identified as an RNAi-promoting factor, knockdown of *zfp-1* by RNAi may not be very efficient. While it would be enlightening to use the *zfp-1*/nDf17 animals, which are missing one copy of *zfp*-1 for expression analysis, they do not survive past the first larval stage and it is not possible to obtain large enough quantities of these worms for biochemical analysis.

If H3K4me is absolutely required for ZFP-1 function in the germline, then one could imagine that in the COMPASS mutants *zfp-1* targets should also be misregulated. However, the COMPASS complex affects methylation of many more genes than just *zfp-1* targets. As expected, COMPASS mutant worms have many developmental and growth problems due to the misregulation of many genes. Although we tried, it was also impossible to get consistent results for upregulation or downregulation of ZFP-1 targets in the COMPASS mutants.

Future approaches for understanding PHD1-PHD2 of ZFP-1

Creating mutations in PHD1-PHD2 and looking for a disruption of multimerization or H3K4me binding will facilitate the understanding of the molecular and biological function of this protein. This can be done using two approaches: First, one can create multiple mutations in *zfp-1* on the pLONG plasmid, use these various plasmids to create different transgenic lines and repeat the *zfp-1(ok554)* cross to the deficiency allele to determine which mutation is unable to rescue the phenotype. This would not be a trivial amount of work but would give direct biological significance of the mutations created. The major caveat of this approach is that in the *zfp-1(ok554)* mutant, endogenous wild-type PHD1-PHD2 is still expressed. If the mutation created in the pLONG plasmid does not disrupt multimerization, then the transgenic protein will

interact with the endogenous protein, which may affect the binding properties of the multimer of ZFP-1. A different mutagenesis approach could be to mutate the entire PHD1 in the context of the pLONG plasmid. For example, substituting PHD1 to a different PHD finger altogether, such as that of BHC80, but leaving PHD2 intact would allow the protein to multimerize, but the H3K4 specificity would be altered. Expression of this chimeric PHD1-PHD2 protein would likely have a dominant effect. This would allow readout of the necessity of H3K4me binding of PHD in a biological context, including possible expression analysis.

A second approach would be to mutate the protein and test histone binding and multimerization *in vitro*. We attempted this, although it proved difficult since the PHD1-PHD2 protein expresses poorly in the constructs we used in bacteria. Any mutations that disrupt the multimerization are also likely to cause the protein to precipitate given that it is already quite unstable. In order for more efficient *in vitro* analysis to be carried out, expression of the PHD1-PHD2 region would have to be carried out in a different cell type (S2 or mammalian). There may also be post-translational modifications of ZFP-1 *in vivo* that are required for its stability that are not properly created in bacterial expression systems.

In order to understand the stability and function of the ZFP-1 protein, it is essential to identify the binding partners of the PHD1-PHD2 region. It is likely that, similar to Jade-1 and VHL (Zhou et al, 2004), ZFP-1 has a binding partner that enhances its stability and perhaps binding specificity. One possible approach could be immunoprecipitation of the ZFP-1 PHD1-PHD2::FLAG protein followed by mass-spectrometry of any interacting proteins. Another specific experiment could be ChIP-chip analysis of the genome-wide chromatin binding of PHD1-PHD2::FLAG using anti-FLAG antibodies. This would allow us to differentiate the targets of PHD1-PHD2 as compared to the short or the long isoforms of ZFP-1.

Part 2:

ZFP-1 PHD1 interacts with the N-terminus of CSR-1 *in vitro*

The long isoform of ZFP-1, containing the PHD fingers, is expressed in developing oocytes. Although both isoforms of the protein are expressed in the nuclei of most cell types, the germline is the only tissue where the short isoform is not expressed in favor of the PHD finger-containing long isoform. This is perhaps the most likely cell-type in which to identify an interacting partner specific to PHD1-PHD2 of ZFP-1. In Chapter 3, I described that both CDL-1::GFP and CSR-1 are also localized to the nucleus at this particular stage of development. While many proteins are expressed in the developing oocyte, there is evidence to support that CSR-1 and CDL-1 work together to process histone messages. Given that this is also the only tissue where the remaining portion of ZFP-1 expressed in *zfp-1(ok554)* can be visualized, there is also a possibility that CSR-1 may be one of the interacting partners of ZFP-1.

CSR-1 has two isoforms, a long isoform that contains the Argonaute specific domains (PAZ, PIWI-like) including a longer N terminal region (Figure 3A, light purple box), and a short isoform that is missing this region. The N-terminus of the long isoform of CSR-1, including the PAZ domain, was used as an antigen to raise antibodies by Aoki *et al* (Aoki et al, 2007). These antibodies have been used for all CSR-1 immunostaining experiments shown in this thesis. We were able to pull down the N-terminus of CSR-1-MBP with ZFP-1 PHD1-GST but not with GST alone or BHC80-GST *in vitro* (Figure 3B). This was not confirmed *in vivo* and it is difficult to say whether this binding may occur in the oocytes. Co-immunoprecipitation experiments did not yield conclusive results suggesting that if ZFP-1 does interact with CSR-1 *in vivo*, it is likely to be a transient binding.

To better understand the long-isoform specific N-terminal region of CSR-1, we created a homology model, based on the recent structure of human Argonaute2 (Schirle & Macrae, 2012) (Figure 3C). The predicted structure shows two anti-parallel ß-sheets protruding from the rest of the protein. This region may be able to interact with another protein, such as ZFP-1 PHD1. However, as with all homology models not based on experimental evidence, further experiments are needed to test these predictions.

The germline expression pattern of CSR-1 places it in the P-granules in the distal germline (Figure 3D, arrows), nuclear in maturing oocytes, on the chromosomes in late oocytes (Figure 3D, arrowhead)(Claycomb et al, 2009). We have also detected CSR-1 in the intestinal cell nuclei (Chapter 3 Figure 7). ZFP-1 is also present on the chromosomes in maturing oocytes (Chapter 2 Figure 3), and given the interaction of ZFP-1 PHD1 with N-terminus of CSR-1 *in vitro,* it is possible that these two proteins interact transiently on chromatin in oocytes. ZFP-1 is able to localize to chromatin in the absence of CSR-1 (Figure 4), likely through a PHD-finger-based interaction as described in Chapter 2. This suggests that if any interaction at the chromatin level exists between ZFP-1 and CSR-1, ZFP-1 is likely to be recruiting CSR-1 rather than the reverse. It is difficult to speculate what role CSR-1 with ZFP-1 has in the oocytes, given CSR-1's involvement with histone mRNA processing. Further experiments are needed to understand the connection between ZFP-1 and RNAi.

Future considerations for ZFP-1 and RNAi

zfp-1(RNAi) has been shown to affect both exogenous RNAi efficiency (Dudley et al, 2002; Kim et al, 2005) and transcriptional gene silencing of a transgene (Grishok et al, 2005). Our studies here have helped clarify the function of the PHD fingers of ZFP-1, but have not clearly brought us to the biological function of those PHD fingers or *zfp-1* in RNAi in general.

Some preliminary Western blot experiments (data not shown) suggested that the histone depletion phenotype seen in knockdown of the RNAi pathway components may also be present in *zfp-1(RNAi)* knockdown. This was difficult to reproduce. However, histone genes are overrepresented among ZFP-1 targets (26 out of 70 histone genes, see Appendix I Table 1), and it may have a role in their expression. Knockdown of *zfp-1* by RNAi did not result in the histone mRNA misprocessing defect seen in the RNAi pathway mutants in Chapter 3, but *zfp-1* may regulate histones in a different way. Insufficient amount of histones may explain the lethality phenotypes seen in *zfp-1*/nDf17 (Chapter 2, Figure 6), although it is not possible to carry out analysis of these worms to determine whether that is the case. A decrease in histone abundance may have severe effects on development and on the efficiency of RNAi. It will be interesting to see what future experiments will tell us about the connection between RNAi and chromatin.

Part 3:

CSR-1 has a complex job description

We have shown that CSR-1 is involved in histone mRNA processing in the germline, which explains the published phenotypes of sterility and chromosome segregation defects of the *csr-1* and other RNAi pathway mutants. We have also shown that CSR-1's effect on histone production is relevant in somatic nuclei in the intestinal cells. These results do not explain the role of the majority of siRNAs bound to CSR-1 that target protein-coding genes throughout the genome. However complex the many roles of CSR-1, it seems that this complexity is likely to be specific to *C. elegans.* This is quite conceivable since the worm has the largest number of Argonaute proteins than any other species known (Boisvert & Simard, 2008), suggesting that these proteins are likely to have functions carried out by non-Argonaute-mediated mechanisms in other organisms.

In Chapter 3, we proposed that since the U7 snRNA is missing in *C. elegans*, small RNAs may assume its role in directing the cleavage of the 3′UTR of histone mRNA transcripts. We showed that CSR-1 interacts with histone mRNA in an *ego-1*-dependent manner, consistent with CSR-1 targeting the histone mRNA transcripts. We also showed that overexpression of transgenic histone proteins rescues the lethality caused by RNAi of *csr-1* or *ego-1*. Although it is clear that loss of histone protein is responsible for the phenotypes of the RNAi mutants, we do not yet know the precise mechanism of histone mRNA processing in *C. elegans*. One possibility discussed in Chapter 3 is that CSR-1 mediates the cleavage of the 3′ ends of histone mRNAs by its known endonucleolytic capabilities (Aoki et al, 2007). Alternatively, CSR-1 may act to recruit other mRNA-specific cleavage factors that have yet to be identified. One important question regarding the scenario where CSR-1 cleaves the histone mRNA transcript is how does

CSR-1 know which targets to cleave? As was reported, the 3′ end the cleavage of histone mRNA is very specific to directly downstream of the conserved stem-loop, which may result either from direct CSR-1 cleavage or in conjunction with 3′ exonuclease trimming of the message. However, CSR-1-bound siRNAs are complementary to entire histone transcripts, similarly to many other genes, therefore there must be a regulatory mechanism used by CSR-1 and its slicer activity. Nevertheless, CSR-1 does have the ability to cleave, thus it is probable that a mechanism exists to promote CSR-1 cleavage activity only in specific scenarios.

Further experiments are needed in order to understand the precise mechanism of both histone mRNA processing in *C. elegans* and the role of the 22G RNAi pathway in the germline, One possible experiment would be to create a GFP reporter construct containing the promoter and 3′ UTR regulatory sequences of a histone gene. Mutations could then be made to the regulatory elements and the resulting transgenic lines could be treated with *csr-1(RNAi)*. The changes in GFP expression would then act as a readout for the requirement of CSR-1, as opposed to scoring lethality. This reporter may allow identification of additional factors working with CSR-1 in histone mRNA processing.

It is clear from the work presented here that there are complex regulatory mechanisms associated with histone post-translational modifications and the production of histone proteins. This work has advanced our understanding of the proteins essential to the eukaryotic cell and the complex function of histones. Understanding how histone proteins are regulated at every possible level has proven essential to the biology of the cell time and time again.

Figure 1: Alignment of ZFP-1 PHD1 with various PHD fingers. The top panel shows alignment with PHD fingers known to bind to H3K4me0, with the residues that interact with the lysine side chain highlighted in blue. The bottom panel shows ZFP-1 PHD1 aligned with PHD fingers known to interact with H3K4me3, with residues known to form an aromatic cage around H3K4me3 highlighted in orange. In both panels, any residues known to differentiate between the methyl states of H3R2 are also highlighted in green. Conserved cysteine and histidine residues that represent the PHD finger domain are highlighted in bold.

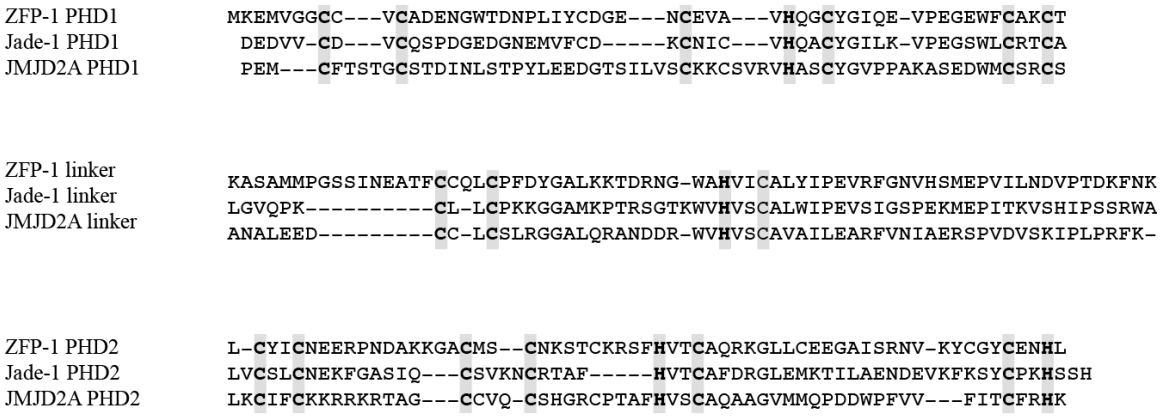


Figure 2: ZFP-1 PHD1-PHD2 primary sequence alignment with Jade-1 PHD1-PHD2 (Zhou et al, 2004) and JMJD2A PHD1-PHD2 (Gray, 2005). The bold highlights indicate conserved histidine and cysteine residues that are predicted to coordinate zinc ions. In the case of the defined PHD fingers, these residues constitute the domain-defining motif.

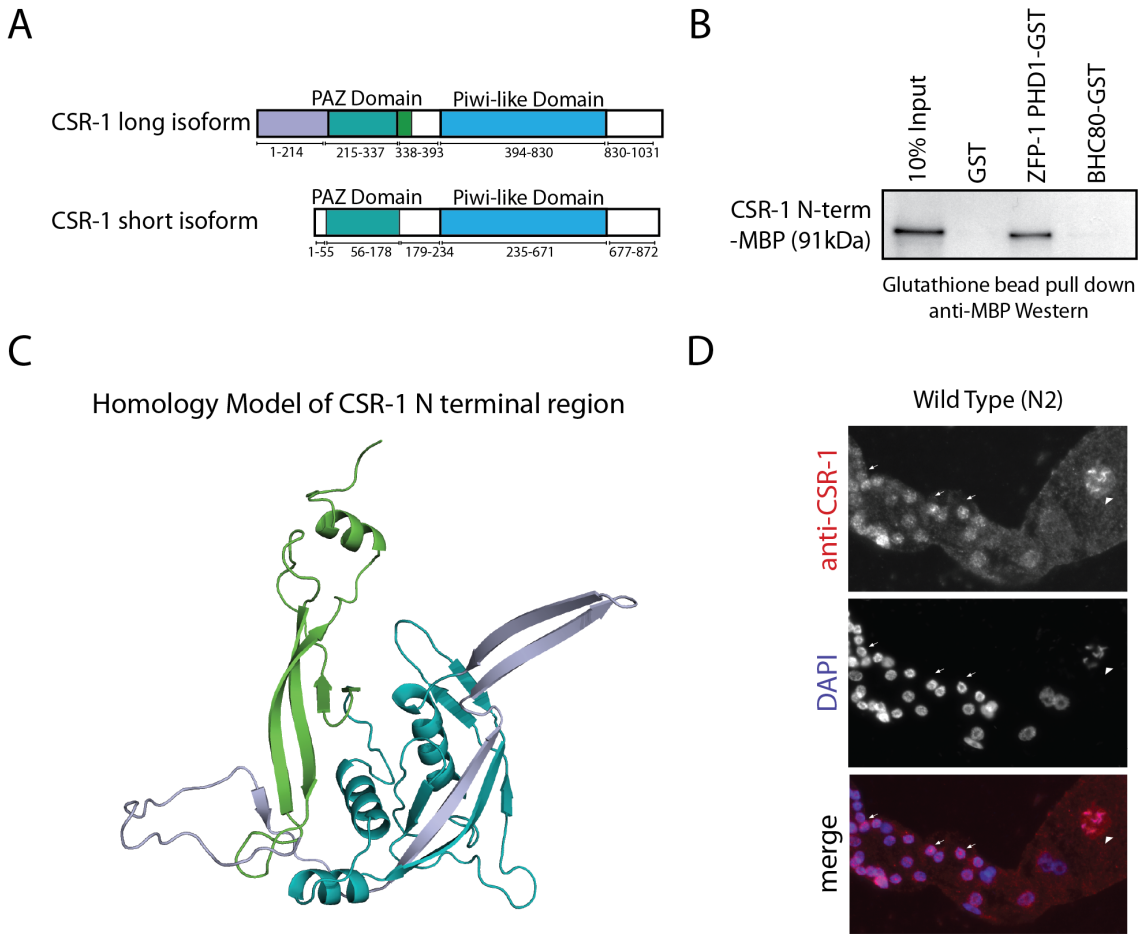


Figure 3: CSR-1 N-terminus interacts with ZFP-1 PHD1 and localizes to chromosomes in oocytes. (A) Schematic representing the two isoforms of CSR-1. Both isoforms contain the traditional Argonaute domains PAZ (teal) and PIWI-like (light blue), but the long isoform contains an additional N-terminal region not present in the short isoform. (B) Pull-down experiment with CSR-1(N-term)-MBP incubated with various GST-fusion proteins using glutathione beads followed by Western blot analysis with anti-MBP antibodies. The CSR-1 N-terminal region (shown in the homology model in (C)) is pulled down by ZFP-1 PHD1-GST, but not by GST alone or BHC80-GST. (C) Homology model of the N-terminal region of CSR-1 based on the crystal structure of the human Argonaute2 (Schirle & Macrae, 2012). The light

purple region represents the N-terminus of CSR-1 present only in the long isoform of the protein, the teal region represents the PAZ domain and the green region represents the beginning of mid-region of the protein, corresponding to the boxes in (A). (D) Immunostaining of dissected adult gonads from wild-type worms with anti-CSR-1 antibodies (Aoki et al, 2007)(pink) shown with DNA (DAPI, blue). Arrows indicate CSR-1 present in P-granules in the more distal germline, arrowhead indicates a maturing oocyte that shows CSR-1 on the chromosomes. Image was taken at 200x zoom. (NOTE: These images were taken using the freeze-crack method as compared to a different immunostaining protocol used in Chapters 2 and 3. See Materials and Methods for details.)

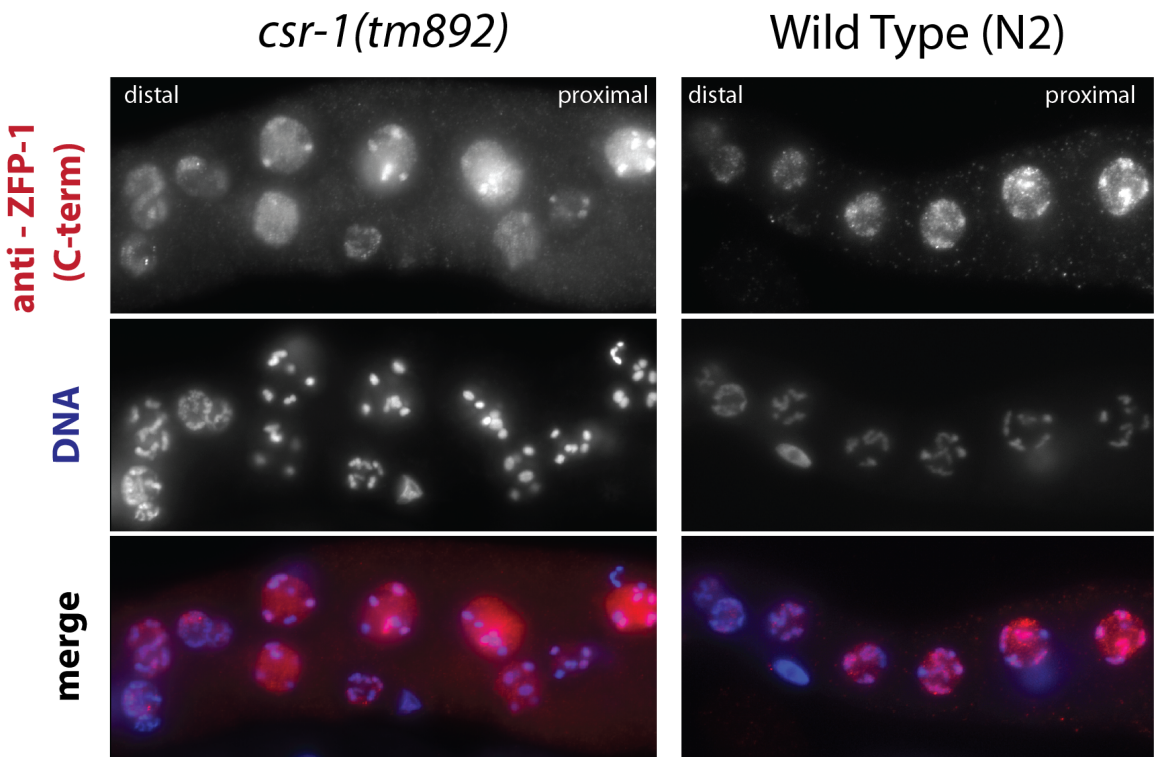


Figure 4: ZFP-1 localizes to chromosomes in wild-type and *csr-1(tm892)* sterile mutants. Immunostaining of dissected wild-type and *csr-1(tm892)* adult worm gonads with anti-ZFP-1 C-terminal antibodies (Mansisidor et al 2011) (pink) and DNA (DAPI, blue). ZFP-1 localizes to chromosomes in wild type (right panel), as seen in Chapter 2, Figures 2 and 3. In *csr-1(tm892)* sterile worms, ZFP-1 still localizes to chromosomes in oocytes. The malformed nuclei in the *csr-1(tm892)* gonad indicate the sterility associated with the loss of *csr-1*. (NOTE: These images were taken using the same immunostaining protocol as in Chapters 2 and 3, see Materials and Methods for details.)

Materials and Methods

Standard immunostaining as used in Chapters 2 and 3

Adult worms were hand picked and placed in PBS (137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4) with 0.01% Tween and 25mM sodium azide in a glass petri dish. Worms were dissected using a scalpel and fixed with 2% paraformaldehyde for 1 hour at room temperature and post-fixed in ice-cold methanol for 5 minutes. Worms were transferred to glass tubes and blocked with 3% (w/v) BSA in PBS. Blocked worms were then incubated overnight at 4°C with anti-ZFP-1 antibodies (Mansisidor et al, 2011) diluted 1:8000, washed 3 times with PBST and incubated for 1 hour at room temperature with anti-rabbit Alexa Fluor 555 (Molecular Probes) diluted 1:200. Worms were then mounted on polylysine slides (LabScientific, Inc) using Mounting solution with DAPI (Vector Laboratories) and all images were taken on Zeiss AxioImager Z1. Results shown are representative of at least five biological replicates.

Immunostaining using freeze-crack method (used in Figure 3)

Adult worms were washed off an agarose plate with PBS into a siliconized glass tube and subsequently washed 2 more times. Approximately 10µl of worms were then pipetted onto a polylysine slide and left to settle. A coverslip was then placed on top of the slide at a right angle so that the edge of the coverslip extended over the edge of the slide. Mild pressure was applied to burst the adult worms. The slide was then carefully placed on dry ice and left to freeze. The coverslip was popped off and immediately placed into pre-chilled 100% methanol and placed at -20°C for 15 minutes. The slide was then transferred to pre-chilled 100% acetone and placed at -20°C for 15 minutes. The slides were then washed with 100µl of PBST three times and blocked with 3% BSA for 30 minutes in a humid chamber. Blocked slides were then incubated for 1 hour

in a humid chamber with anti-CSR-1 (Aoki et al, 2007) antibodies diluted 1:1000, washed 3 times with PBST and incubated for 1 hour at room temperature with anti-rabbit Alexa Fluor 555 (Molecular Probes) diluted 1:200. Mounting solution with DAPI (Vector Laboratories) was used and all images were taken on Zeiss AxioImager Z1. Results shown are representative of at least five biological replicates.

References

Aoki K, Moriguchi H, Yoshioka T, Okawa K, Tabara H (2007) In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. *EMBO J* **26:** 5007-5019

Boisvert ME, Simard MJ (2008) RNAi pathway in C. elegans: the argonautes and collaborators. *Current topics in microbiology and immunology* **320:** 21-36

Claycomb JM, Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, Conte D, Mello CC (2009) The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* **139:** 123-134

Di Lorenzo A, Bedford MT (2011) Histone arginine methylation. *FEBS Lett* **585:** 2024-2031

Dudley NR, Labbé J-C, Goldstein B (2002) Using RNA interference to identify genes required for RNA interference. *Proc Natl Acad Sci USA* **99:** 4191-4196

Eissenberg JC, Shilatifard A (2010) Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Dev Biol* **339:** 240-249

Fiedler M, Sánchez-Barrena MJ, Nekrasov M, Mieszczanek J, Rybin V, Müller J, Evans P, Bienz M (2008) Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. *Mol Cell* **30:** 507-518

Gray SG (2005) Functional Characterization of JMJD2A, a Histone Deacetylase- and Retinoblastoma-binding Protein. *Journal of Biological Chemistry* **280:** 28507-28518

Gregory GD, Vakoc CR, Rozovskaia T, Zheng X, Patel S, Nakamura T, Canaani E, Blobel GA (2007) Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. *Mol Cell Biol* **27:** 8466-8479

Grishok A, Sinskey JL, Sharp PA (2005) Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. *Genes Dev* **19:** 683-696

Iberg AN, Espejo A, Cheng D, Kim D, Michaud-Levesque J, Richard S, Bedford MT (2007) Arginine Methylation of the Histone H3 Tail Impedes Effector Binding. *Journal of Biological Chemistry* **283:** 3006-3010

Ishizaka A, Mizutani T, Kobayashi K, Tando T, Sakurai K, Fujiwara T, Iba H (2012) Double plant homeodomain (PHD) finger proteins DPF3a and -3b are required as transcriptional co-activators in SWI/SNF complex-dependent activation of NF-kappaB RelA/p50 heterodimer. *J Biol Chem* **287:** 11924-11933

Kim JK, Gabel HW, Kamath RS, Tewari M, Pasquinelli A, Rual J-F, Kennedy S, Dybbs M, Bertin N, Kaplan JM, Vidal M, Ruvkun G (2005) Functional genomic analysis of RNA interference in C. elegans. *Science* **308:** 1164-1167

Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*

Miller TC, Rutherford TJ, Johnson CM, Fiedler M, Bienz M (2010) Allosteric remodelling of the histone H3 binding pocket in the Pygo2 PHD finger triggered by its binding to the B9L/BCL9 co-factor. *J Mol Biol* **401:** 969-984

Mohan M, Lin C, Guest E, Shilatifard A (2010) Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nature reviews Cancer* **10:** 721-728

Muramoto T, Muller I, Thomas G, Melvin A, Chubb JR (2010) Methylation of H3K4 Is required for inheritance of active transcriptional states. *Curr Biol* **20:** 397-406

Musselman CA, Kutateladze TG (2011) Handpicking epigenetic marks with PHD fingers. *Nucleic Acids Res* **39:** 9061-9071

Ptashne M (2007) On the use of the word 'epigenetic'. *Curr Biol* **17:** R233-236

Rando OJ, Chang HY (2009) Genome-wide views of chromatin structure. *Annu Rev Biochem* **78:** 245-271

Rando OJ, Verstrepen KJ (2007) Timescales of genetic and epigenetic inheritance. *Cell* **128:** 655-668

Sanchez R, Zhou MM (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36:** 364-372

Schaner C (2006) Germline chromatin. *WormBook*

Schaner CE, Deshpande G, Schedl PD, Kelly WG (2003) A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. *Dev Cell* **5:** 747-757

Schirle NT, Macrae IJ (2012) The Crystal Structure of Human Argonaute2. *Science*

Smith E, Shilatifard A (2010) The Chromatin Signaling Pathway: Diverse Mechanisms of Recruitment of Histone-Modifying Enzymes and Varied Biological Outcomes. *Molecular Cell* **40:** 689-701

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14:** 1025-1040

van Steensel B (2011) Chromatin: constructing the big picture. *EMBO J* **30:** 1885-1895

Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A, Liu XS, Rinn J, Schier AF (2010) Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464:** 922-926

Zhou MI, Wang H, Foy RL, Ross JJ, Cohen HT (2004) Tumor suppressor von Hippel-Lindau (VHL) stabilization of Jade-1 protein occurs through plant homeodomains and is VHL mutation dependent. *Cancer Research* **64:** 1278-1286

Appendix I: Gene lists and primers

The following list of genes was compiled by overlapping the total ZFP-1-bound target genes from Mansisidor et al, 2011, with published datasets for H3K4me2/3 targets from Gu & Fire, 2009 and the germline specific mountain from topomap studies in Kim et al, 2001. Those genes with annotated names are given in Table 1 and the following information for those tested includes any information given on Wormbase.org. The primers used for real time PCR are indicated for each gene tested. Table 2 lists the ZFP-1 target genes that correspond to histone genes. The gene names and class of histone are given.

Gu SG, Fire A (2009) Partitioning the C. elegans genome by nucleosome modification, occupancy, and positioning. *Chromosoma* **119:** 73-87

Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS (2001) A gene expression map for Caenorhabditis elegans. *Science* **293:** 2087-2092

Mansisidor AR, Cecere G, Hoersch S, Jensen MB, Kawli T, Kennedy LM, Chavez V, Tan M-W, Lieb JD, Grishok A (2011) A Conserved PHD Finger Protein and Endogenous RNAi Modulate Insulin Signaling in Caenorhabditis elegans. *PLoS Genet* **7:** e1002299

Table 1: ZFP-1 germline-specific targets with H3K4me2/3 peaks

ZFP-1 targeted germline genes	Genes also with H3K4me2/3 peak	Annotated name
C04H5.6	C07H6.4	
C05D2.5	C07H6.5	cgh-1
C07H6.4	C08F8.2	pfd-1
C07H6.5	C16A3.7	tag-182
C08F8.2	C24F3.2	
C16A3.7	C26D10.2	hel-1
C24F3.2	C27A2.3	ify-1
C26D10.2	C30C11.2	rpn-1
C27A2.3	C30C11.4	msi3p
C30C11.2	C32F10.2	lin-35
C30C11.4	C34G6.5	
C32F10.2	C37A2.4	cye-1
C34G6.5	C37C3.9	
C37A2.4	C47E8.4	daf-21
C37C3.9	C50C3.8	bath-42
C47E8.4	C54G10.2	rfc-1
C50C3.8	C56C10.1	
C54G10.2	D1014.8	spr-1
C56C10.1	D1046.2	
D1014.8	F22D3.1	ceh-38
D1046.2	F26B1.2	
D2013.5	F30F8.8	taf-5
F07A11.2	F31D4.3	fkb-6
F12F6.5	F31E3.3	rfc-4
F17A9.2	F32B6.3	nhr-4
F22D3.1	F32E10.1	
F26B1.2	F43G6.1	dna-2
F26H9.4	F49E11.1	mbk-2
F30F8.8	F52H3.2	
F31D4.3	F53A3.2	polh-1
F31E3.3	F54D5.9	
F32B6.3	F56F3.1	pqn-45
F32E10.1	H21P03.2	
F33G12.2	K01G5.5	
F43G6.1	K05C4.7	
F49E11.1	K06A5.4	knl-2
F52H3.2	K07A1.12	lin-53
F53A3.2	K07H8.10	
F54D5.9	K08B4.1	
F55A11.6	K08E3.6	cyk-4
F56F3.1	K08E7.1	
H21P03.2	K08F4.2	
K01G5.5	K08F4.3	

K02B12.8	K12D12.1	top-2
K05C4.7	M03C11.4	tag-235
K06A5.4	M04F3.1	rpa-2
K07A1.12	M7.2	klc-1
K07H8.10	R05D3.2	
K08B4.1	T01C3.7	fib-1
K08E3.6	T02E1.2	
K08E7.1	T05H10.5	ufd-2
K08E7.3	T07A9.1	
K08F4.2	T09A5.10	lin-5
K08F4.3	T10B5.3	
K12D12.1	T10B5.6	knl-3
M03C11.4	T20G5.11	rde-4
M04F3.1	T22D1.10	ruvb-2
M7.2	T22F3.4	rpl-11
R05D11.5	T23H2.1	npp-12
R05D3.2	T24G10.2	
R12E2.10	T26A5.5	
T01C3.7	T26A5.7	set-1
T02E1.2	T26A8.4	
T05E11.4	W01B11.3	nol-5
T05H10.5	W02D9.1	
T07A9.1	W02D9.3	
T09A5.10	W03C9.7	mex-1
T10B5.3	W08D2.7	mtr-4
T10B5.6	Y11D7A.12	flh-1
T20G5.11	Y43F4B.4	npp-18
T22D1.10	ZC168.4	cyb-1
T22F3.4	ZK1010.3	
T23H2.1	ZK742.2	
T24G10.2		
T26A5.5		
T26A5.7		
T26A8.4		
W01B11.3		
W02D9.1		
W02D9.3		
W03C9.7		
W08D2.7		
Y11D7A.12		
Y43F4B.4		
ZC168.4		
ZK1010.3		
ZK742.2		
ZK829.9		

Primers for mRNA quantification by qRT-PCR in *zfp-1*(*RNAi*) (5' to 3')

All targets have peaks of ZFP-1, H3K4me2/3 and are germline enriched

All ZFP-1 peaks have been confirmed to be strong in embryos

1. *bath-42* – The C50C3.8 gene encodes a protein with a meprin-associated Traf homology (MATH) domain that may be involved in apoptosis.

Forward Primer

TCAAGAATGTTTAGTGTCCCTG

Reverse Primer

TCCTCTCCTCTGGCTAATCC

2. *cgh-1* - *cgh-1* encodes a putative DEAD-box RNA helicase, orthologous to budding yeast Dhh1p, fission yeast Ste13p, Drosophila ME31B, and human DDX6; CGH-1 inhibits physiological apoptosis in oocytes, keeping it down to a normal level of ~50% in hermaphrodite gonads; independently of apoptosis, CGH-1 is also required for sperm function, oocyte fertilization, and early embryonic cytokinesis;

Forward Primer

ACTTCCAGGGAATTCTCGAC

Reverse Primer

ACACCAAGAAGAGTAAGCTCCT

3. *cyb-1* – *cyb-1* encodes a cyclin B homolog that appears to be required both for embryogenesis (via maternally encoded CYB-1) and for postembryonic development and fertility (via

zygotically encoded CYB-1); removing maternal product in cyb-1(RNAi) embryos causes multiple nuclei, and affected embryos arrest development at an early embryonic stage.

Forward Primer

ATTTCAGCTATGTTCGTTGCC

Reverse Primer

GAGGAAGACAAGCGAAGATGG

4. *cye-1* - *cye-1* encodes a homolog of the G1 cell cycle regulator cyclin E; CYE-1 is required for progression through the G1 phase of the cell cycle and for endoreduplication in intestinal cells, and is required maternally both for embryonic development and for the proper execution of several post-embryonic lineages (e.g. vulval development); CYE-1 is provided maternally and expressed ubiquitously in nuclei during embryonic development; postembryonic expression is restricted to actively proliferating blast cells, including the germline.

Forward Primer

CGAAGTACTAAGGAAGGTCCA

Reverse Primer

CATAGCTGGAAGTGGTTGTC

5. *cyk-4* - *cyk-4* encodes a Rho GAP (Rho guanosine triphosphatase (GTPase) activating protein); during development, cyk-4 is required for two different processes: 1) paternally provided CYK-4 regulates anterior-posterior polarity in the one-cell embryo by controlling the actomyosin cytoskeletal network, and 2) together with ZEN-4, CYK-4 is sufficient to promote formation of the large bundles of microtubules that form the central spindle during anaphase and polarize the foregut epithelium

Forward Primer

TATAGATGAAGAGCCGAATGAGGG

Reverse Primer

GTGGTAGTGGTGGTAGTTGTC

6. *daf-21* - *daf-21* encodes a member of the Hsp90 family of molecular chaperones; DAF-21 activity is required for larval development, negative regulation of dauer larva formation, and a number of specific chemosensory behaviors, such as the response to chemicals and odorants detected by the ASE and AWC sensory neurons

Forward Primer

GGAGAAGAAGGAGGGAGAGG

Reverse Primer

TGGACAAGCTCTTGTAGAACTC

7. *ify-1* - Interactor of FizzY protein

Forward Primer

CAACAACACTTCAGTGCAAAGC

Reverse Primer

TTTGACGAGTTTACAACCAGCC

8. *klc-1* - *klc-1* encodes one of two *C. elegans* kinesin light chains; by homology, KLC-1 is predicted to function, along with kinesin-like heavy chains, as part of a multimeric kinesin complex involved in intracellular transport; consistent with this, KLC-1 has been shown to interact with UNC-116/kinesin, KCA-1/kinesin cargo adaptor, and the ARX-2/Arp2/3 complex component in yeast two-hybrid assays; RNAi experiments indicate that klc-1 activity is required for normal embryonic development.

Forward Primer

CTCTCGAAAGTTGCTATCACC

Reverse Primer

GACATCACATTGCTATCCATCAC

9. *knl-3* - *knl-3* encodes a novel protein; KNL-3 activity is essential for formation of a functional kinetochore and thus, for proper chromosome segregation and spindle pole separation; KNL-3 localizes to kinetochores throughout mitosis and this localization requires the activity of HCP-4 (CENP-C), a conserved kinetochore component that likely directs kinetochore assembly

Forward Primer

TCAAAGAGATGAGCCAATACGA

Reverse Primer

GCGATTTCAATAGCTGACTTGAG

10. *nol-5* – nucleolar protein

Forward Primer

GGAGGAGAAGGTCAAAGAGGA

Reverse Primer

GAAGAGTTGAGCACGGTAGG

11. *lin-35* - *lin-35* encodes the C. *elegans* retinoblastoma protein (Rb) ortholog; lin-35 was first identified in screens for synthetic multivulva (synMuv) genes and as a class B synMuv gene, functions redundantly with class A genes to antagonize Ras signaling and negatively regulate vulval development; in addition, lin-35 activity is required redundantly with: 1) pha-1 and ubc-18 for early steps in pharyngeal morphogenesis, 2) fzr-1 for normal patterns of postembryonic proliferation, 3) xnp-1 for somatic gonad development, and 4) psa-1 for fertility and embryonic and larval development; on its own, lin-35 is also required for wild-type levels of fertility

Forward Primer

AAACTAATACACCGCCACCA

Reverse Primer

CCGAACCCTCAAGAGAAACC

12. *lin-5* - *lin-5* encodes a novel protein with a centrally located alpha-helical coiled-coil domain and ten potential proline-directed kinase phosphorylation sites; LIN-5 activity is required for mitosis and cytokinesis at all stages of development and specifically, for spindle positioning, chromosome alignment and segregation, and perhaps also for blocking exit from mitosis when chromosome segregation fails

Forward Primer

AAGAGAACGAAACAGTCACC

Reverse Primer

GATTGAAGGAACAGCTCTTTCTC

13. *mbk-2* - *mbk-2* encodes one of two *C. elegans* members of the DYRK (dual-specificity Yak1-related kinase) family of proteins that includes S. cerevisiae Yak1 and the Drosophila minibrain and DYRK2 kinases; MBK-2 activity is required maternally for the egg-to-embryo transition that occurs during the earliest stages of embryonic development; specifically, MBK-2 is required for: 1) posterior localization of the germ plasm components PIE-1, POS-1, and PGL-1, and 2) post-fertilization degradation of a subset of maternal proteins including the MEI-1 and MEI-2 meiosis-specific katanin subunits, the OMA-1 oocyte maturation factor, and residual PIE-1 that remains anteriorly localized after its normal posterior segregation

Forward Primer

GATTGGTGGTGCGAATAATGGT

Reverse Primer

AATGCCTTGATTACCTGTCCA

14. *mex-1* - *mex-1* encodes a CCCH-type zinc-finger protein that is required maternally for segregation of P granules, germ cell formation, and somatic cell differentiation in the early embryo; MEX-1 is expressed cytoplasmically in germ line blastomeres, is a component of P granules, and is required for restricting PIE-1 expression and function to these cells

Forward Primer

TGGATAAGCTAATGTTGGAGGA

Reverse Primer

TGTTGTTGAATGAGTCATCGG

15. *msi3p* - C30C11.4 encodes a member of the Hsp70 family of heat shock proteins; loss of C30C11.4 activity via large-scale RNAi screens results in a variety of defects including embryonic and larval lethality, slow growth rates, and locomotory and morphological defects; in addition loss of C30C11.4 has been reported to result in a small but reproducible reduction in adult lifespan in otherwise long-lived animals.

Forward Primer

GGTTCTTGGATTCGACATCGG

Reverse Primer

AAAGATACGCAAGCTGGAGTG

16. *npp-12* - *npp-12* encodes Gp210, an evolutionarily conserved membrane protein of the nuclear pore complex (NPC); NPP-12 is required for embryonic viability and normal nuclear membrane structures.

Forward Primer

TTCACCGAAATGGACTACTTCTG

Reverse Primer

GTAACTTGGCCTTCTTTCTTTCC

17. *pfd-1* - *pfd-1* encodes a putative prefoldin subunit 1, orthologous to human PFDN1, that is required for normal microtubule nucleation and growth, embryonic viability, distal tip cell (DTC) migration, and locomotion; PFD-1 is expressed cytoplasmically in most, if not all, tissues; pfd-1(RNAi) animals have sterile progeny and are uncoordinated, and pfd-1(RNAi) embryos show a reduced microtubule growth rate

Forward Primer

CCAATGAAACACGAATGCGG

Reverse Primer

TCATCCAAATCCACCAAGTTCT

18. *pqn-1* - *abu-1* encodes a transmembrane protein with a predicted signal sequence, a glutamine/asparagine-rich domain and multiple cysteine-rich repeats (DUF139); abu-1 expression is induced by blockage of the unfolded-protein response in the endoplasmic reticulum, and ABU-1 may help protect the organism from damage by improperly folded nascent protein.

Forward Primer

GATTTCAATGGCGAAGATGGT

Reverse Primer

GTGGAGACAAACCCGTATTT

19. *rde-4* – *rde-4* encodes a dsRNA binding protein (dsRBP); during the initiation phase of RNA interference (RNAi), RDE-4 appears to function in recognition and subsequent cleavage of long-

trigger dsRNA molecules into small, interfering RNAs (siRNAs); RDE-4 exists in solution as a homodimer and binds long dsRNA with high affinity in a sequence-non-specific manner

Forward Primer

TTCCTACAAGTGCTGAATTTCC

Reverse Primer

TTTGATTGCACATCGTGCAT

20. *set-1* - Predicted methyltransferase (contains a SET domain)

Forward Primer

AATGGTGTATCGATGCGACAAA

Reverse Primer

TCGATTTCCACCACTTTCGTTT

21. *spr-1* - *spr-1* encodes the *C. elegans* ortholog of the human corepressor CoREST that functions in HDAC-containing complexes to mediate transcriptional repression; in *C. elegans*, spr-1 was first identified in screens for genetic suppressors of the egg-laying defective phenotype of sel-12/presenilin mutants; subsequent genetic analyses showed that spr-1 likely functions as a negative regulator of LIN-12/Notch signaling in multiple cells, including those of the somatic gonad, vulva, and early embryo

Forward Primer

TACAAAGCGAGTCTGATGGTTC

Reverse Primer

CAGCCAGAACTTCAATCTCTTCC

Non-germline enriched, zfp-1 targets with H3K4me:

22. *xnp-1* *(*B0041.7) - *xnp-1* encodes an ATP-dependent DNA helicase of the SNF2 family that is orthologous to human XNP/ATR-X, which is associated with a number of X-linked mental retardation syndromes; in *C. elegans*, xnp-1 activity is required at high temperatures for embryogenesis, somatic gonad development, fertility, and vulval morphogenesis

Forward Primer

GAATCCGAATCCGAATCTGAAGAC

Reverse Primer

TTGAGCGTTTCCGTTTCTCTG

23. *pdk-1* – kinase in insulin signaling pathway

Forward Primer

GTGCTCATCCAGAAAGACGG

Reverse Primer

AGTGTTCTCTTCCTCCTCATCC

24. *spo-11* - *spo-11* encodes an ortholog of Spo11p in S. cerevisiae that is required for meiotic DNA recombination, and (in conjunction with CHK-2 and MRE-11) for the localization of RAD-51 to foci in late zygotene/early pachytene nuclei

Forward Primer

ACGTGGGATTCTAGCAACTTC

Reverse Primer

ATTTCAATACCATGCGGATCAG

Germline enriched, *zfp-1* targets, no H3K4me

25. *let-99* - *let-99* encodes a protein with a DEP domain (Domain found in Dishevelled, Egl-10, and Pleckstrin); LET-99 has no obvious homologs in non-nematodes, but has a truncated paralog (LRG-1) in *C. elegans*; LET-99 is required for the mitotic spindle to be properly oriented with respect to the axis of cellular polarity, both during anterior migration and rotation of the nuclear-centrosome complex and during anaphase

Forward Primer

CATCTGTTCGACGAAATTCTATGG

Reverse Primer

GGAACATTGTTAGGGATGATTCGT

26. *srgp-1* - Cdc42-interacting protein CIP4

Forward Primer

AACTTCATCAACATCAGCAGACAG

Reverse Primer

CGTGCGGAACTCTTTAGCAG

Non *zfp-1* targets, germline enriched with H3K4me:

27. B0336.5

Forward Primer

TCGAAACAAAGCCATCATCTC

Reverse Primer

CGGTTGGTGCTAATAATATCCAC

28. *npp-15* (C29E4.4)

Forward Primer

GGACTTGAAGAGAACTCCAC

Reverse Primer

TTAGCACTCATCCAGACCA

29. *zfp-1* mRNA (to check that RNAi works) on exon 2 and the 3-4 junction, RNAi clone targets C-terminal end to get both long and short

Forward Primer

GGAACGTACATAGTATGGAACCAG

Reverse Primer

GAAAGCTCCTCTTGCACGTC

30. actin (*act-3*) exon-exon junction

Forward Primer

GAAGTGCGACATTGATATCCGT

Reverse Primer

CTTGATCTTCATGGTTGATGGG

Previous targets: (besides *pdk-1*)

daf-16 – FOXO transcription factor

Forward Primer

TTCAAGCCAATGCCACTACC

Reverse Primer

ACACGATTGAATTCCAGGCAG

***egl-30* -** *egl-30* encodes an ortholog of the heterotrimeric G protein alpha subunit Gq (Gq/G11 class) that affects viability, locomotion, egg laying, synaptic transmission, and pharyngeal pumping

Forward Primer

GAGGTCAGCGATCAGAAAGG

Reverse Primer

TTCCATTCGGTTCTCGTTGTC

B0041.2a - B0041.2 encodes, by alternative splicing, three isoforms of an unfamiliar protein paralogous to AIN-1, and homologous to Brugia malayi 14748.m00068, 14052.m00191, and 14963.m01790; B0041.2 protein binds GEI-4 and GEI-16 in yeast two-hybrid experiments

Forward Primer

CGTAATCCATTCTTTGCTCAACAC

Reverse Primer

TTGAGGATCCATCCATGACAG

***ncl-1* (ZK112.2)** - *ncl-1* encodes a cytoplasmic B-box zinc finger protein that negatively regulates rRNA and 5S RNA transcription, nucleolus size, and body size

Forward Primer

GCCAAGCAGATCCAACTCAC

Reverse Primer

GCTGTTTGAATACCATAACCTCC

Non-target: *lys- 7* - *lys-7* encodes an enzyme homologous to an antimicrobial lysozyme encoded by the LYS4 gene of the protozoan parasite Entamoeba histolytica

Forward Primer

GTCAGAGCAATCTGGATTCAGG

Reverse Primer

GTTAATCCGGATTGTCTGGCTC

Table 2: Histone genes targeted by ZFP-1

Histone Genes	Gene name	Class	Histone Genes targeted by ZFP-1	Gene name	Class
B0035.10	his-45	H3	C50F4.5	his-41	H2B
B0035.7	his-47	H2A	C50F4.7	his-37	H4
B0035.8	his-48	H2B	F08G2.3	his-42	H3
B0035.9	his-46	H4	F17E9.10	his-31	H4
C50F4.13.2	his-35	H2A	F17E9.12	his-32	H3
C50F4.5	his-41	H2B	F22B3.1	his-63	H3
C50F4.7	his-37	H4	F22B3.2	his-64	H4
E03A3.3	his-69	H3	F45F2.12	his-8	H2B
E03A3.4	his-70	H3	F45F2.13	his-6	H3
F07B7.10	his-51	H2A	F45F2.2	his-39	H2B
F07B7.11	his-54	H2B	F45F2.3	his-5	H4
F07B7.3	his-53	H2A	F45F2.4	his-7	H2A
F07B7.4	his-52	H2B	F55G1.10	his-60	H4
F07B7.5	his-49	H3	F55G1.11	his-61	H2A
F07B7.9	his-50	H4	F55G1.2	his-59	H3
F08G2.1	his-44	H2B	F55G1.3	his-62	H2B
F08G2.2	his-43	H2A	K03A1.6	his-38	H4
F08G2.3	his-42	H3	K06C4.10	his-18	H4
F17E9.10	his-32	H3	K06C4.5	his-17	H3
F17E9.12	his-31	H4	M163.3	his-24	H1
F17E9.13	his-33	H2A	T10C6.11	his-4	H2B
F17E9.9	his-34	H2B	T10C6.12	his-3	H2A
F22B3.1	his-64	H4	T10C6.14	his-1	H4
F22B3.2	his-63	H3	T23D8.5	his-67	H4
F35H10.1	his-30	H2A	T23D8.6	his-68	H2A
F35H10.11	his-29	H2B			
F45E1.6.2	his-71	H3			
F45F2.12	his-8	H2B			
F45F2.13	his-6	H3			
F45F2.2	his-39	H2B			
F45F2.3	his-5	H4			
F45F2.4	his-7	H2A			
F54E12.1	his-55	H3			
F54E12.3	his-56	H4			
F54E12.4	his-58	H2B			
F54E12.5	his-57	H2A			
F55G1.10	his-61	H2A			
F55G1.11	his-60	H4			

F55G1.2	his-59	H3			
F55G1.3	his-62	H2B			
H02I12.6	his-66	H2B			
H02I12.7	his-65	H2A			
K03A1.6	his-38	H4			
K06C4.10	his-18	H4			
K06C4.11	his-19	H2A			
K06C4.12	his-22	H2B			
K06C4.13	his-27	H3			
K06C4.2	his-28	H4			
K06C4.3	his-21	H2A			
K06C4.4	his-20	H2B			
K06C4.5	his-17	H3			
M163.3	his-24	H1			
T10C6.11	his-4	H2B			
T10C6.12	his-3	H2A			
T10C6.13	his-2	H3			
T10C6.14	his-1	H4			
T23D8.5	his-67	H4			
T23D8.6	his-68	H2A			
W05B10.1	his-74	H3			
Y49E10.6.2	his-72	H3			
ZK131.1	his-26	H4			
ZK131.10	his-16	H2A			
ZK131.2	his-25	H3			
ZK131.3	his-9	H3			
ZK131.4	his-10	H4			
ZK131.5	his-11	H2B			
ZK131.6	his-12	H2A			
ZK131.7	his-13	H3			
ZK131.8	his-14	H4			
ZK131.9	his-15	H2B			

Appendix II

A side note on histone specific antibodies

As we showed in Chapter 3, the antibodies specific for H4 vary greatly in their ability to detect differences in the levels of histone protein. These antibodies tend to be alarmingly hypersensitive and therefore reach saturation very easily. While this provides for a good loading control, it also makes detecting small differences in histone proteins challenging. During one of my committee meetings, Dr. Mann raised a question about the sensitivity of H3K4me2 specific antibodies. This prompted me to test the sensitivity of my antibodies by using synthetic biotinylated histone tail peptides used for binding assays. I found that antibodies specific to H3K4me2, were also very strongly bound to H3K4me1. Antibodies specific to H3K4me3 were more selective, thus I used only anti-H3K4me3 for my ChIP experiments **(Figure 1)**.

It is possible that some residual binding to H3K4me1 would not greatly affect the ChIP experiment selecting for H3K4me2. However, having attended several conferences either specific to chromatin or with large sections devoted to the study of chromatin marks, there is invariably a poster remarking on the abilities of antibodies specific to one histone mark to bind another even more strongly, in some cases to an opposing mark as well (e.g. H3K9me vs H3K36me). For my own work, only the methylation states of H3K4 seemed relevant, so I did not test my antibodies on other histone tail PTMs. However, both by word-of-mouth at conferences and from my own experience, I have learned that some antibodies specific to one particular methylation state or histone tail can also bind others quite strongly. This is important because antibodies that can bind marks that are correlated to completely opposing states of transcription could lead to very different datasets from a given *in vivo* pull-down experiment. It seems that although most rigorous scientists who work with these antibodies are aware of their

variation in sensitivities, at this point in time no real catalogue exists to document which antibodies are in fact, specific. There are groups working towards producing such a catalogue, but I urge readers of this thesis to be aware of this variability when using histone specific antibodies.

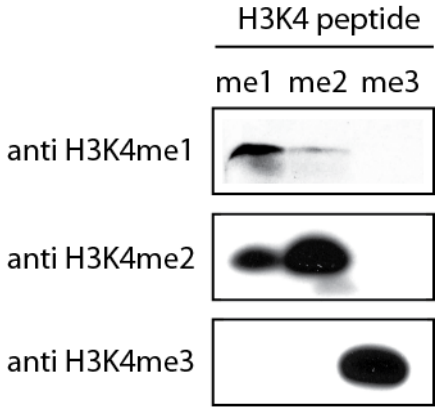


Figure 1: Histone tail peptides with three methylation states of H3K4 were used to test the specificity of anti H3K4 methylation specific antibodies. Antibodies specific for H3K4me1 also show reactivity with H3K4me2; antibodies specific for H3K4me2 also strongly bind to H3K4me1; antibodies specific for H3K4me3 select only the H3K4me3 state in this experiment.