UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title

Investigating how the MES Chromatin Regulators Protect Germline Immortality in C. elegans

Permalink

Author

Cockrum, Chad

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**Investigating how the MES chromatin regulators
protect germline immortality in *C. elegans***

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

MOLECULAR, CELL
AND
DEVELOPMENTAL BIOLOGY

by

Chad Cockrum

December 2021

The Dissertation of Chad Cockrum is
approved:

_____
Professor Susan Strome, chair

_____
Professor John Tamkun

_____
Professor Grant Hartzog

_____
Professor Jeremy Sanford

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

iv

ABSTRACT

Investigating how the MES chromatin regulators

protect germline immortality in *C. elegans*

Chad Cockrum

The germline has the amazing power and the tremendous responsibility to create an entirely new organism and control the information that gets transmitted to offspring. 2 unique properties grant the germline this power: totipotency and immortality. How the germline acquires and maintains these properties are burning questions. The primordial germ cells (PGCs) must launch and maintain a germline-appropriate gene expression program to acquire and protect germ fate. Chromatin regulation or DNA packaging is one mechanism that can tackle this challenge by allowing PGCs to selectively 'turn on' genes that encode germline-appropriate RNAs. How chromatin regulation impacts gene expression in PGCs is poorly understood. In *C. elegans*, the MES chromatin regulators are required maternally for germline development in offspring. MES-2, MES-3, and MES-6 form the worm version of the H3K27 histone methyltransferase (HMT) Polycomb Repressive Complex 2 (PRC2), and MES-4 is an H3K36 HMT. The essential role of PRC2 in ensuring germline survival and immortality in *C. elegans* is to repress genes on the X chromosome. MES-4's role has not yet been determined and is far more puzzling. An attractive model is that MES-4 instructs PGCs to express germline genes by transmitting an epigenetic 'memory of germline' from parent germlines to offspring PGCs. The focus of my thesis was testing this model using a combination of genetics, transcriptomics, and microscopy. I

found that neither MES-4 or its specific role in transmitting a memory of germline are required for PGCs to launch a germline program and develop into a healthy germline. Instead, MES-4's critical role in protecting germline survival and immortality is repressing X genes, similar to PRC2's role. This encouraged me to test whether MES-4 and PRC2 cooperate to repress the same X genes in PGCs. I indeed found this to be the case. I also identified the THAP transcription factor LIN-15B as a major driver of X mis-expression and germline death in *mes-4* mutants. Together my findings answered how the MES chromatin regulators protect germline immortality, and raised interesting follow-up questions, such as how do MES-4 and LIN-15B balance levels of X-chromosome expression in PGCs? My thesis work is one of the many testaments to the power of transcriptomics to study gene regulation at the whole-genome scale. However, many researchers who seek to embrace transcriptomics face the hurdle or 'activation energy' to get started. I led and published a major effort by the Strome lab to provide such researchers with a launchpad into transcriptomics; I discussed foundations, best practices, and commonly used strategies, and I summarized available resources.

ACKNOWLEDGEMENTS

CHAPTER 1

INTRODUCTION

The germline is totipotent and immortal

Germ cells (i.e. gametes and their precursors) have 2 unique properties that enable them to form each new generation of organism. First, they are 'totipotent', meaning that they can generate every cell type, including somatic cells and new germ cells. Second, they are 'immortal'; while somatic lineages die off in each generation, the germline survives and connects generations (Figure 1-1). How germline immortality is protected to ensure the survival of a species is a key question in developmental biology. Germ cells must acquire and protect their germ fate by launching and maintaining a germline-appropriate gene expression program. Germ cells must 1) express genes that encode products needed for germline development and function ('germline genes') and 2) repress genes that encode products that are toxic to the germline, such as those that drive somatic development ('somatic genes'). Failure to repress somatic genes can cause germ cells to reprogram into somatic cells and lose their immortal property. Therefore, identifying mechanisms that define and shape a germ cell's gene expression program is an important goal in the field and will help us understand how germline immortality is protected. In my thesis, I used the worm *Caenorhabditis elegans* as a powerful model organism to investigate how the primordial germ cells (PGCs) regulate their gene expression program to maintain their germ fate and ensure their survival.

Germline development in *C. elegans*

       C. elegans is a well-established and premier model for studying germline development because of its genetic tractability, short life cycle, and large brood size, and because its germline can be easily visualized and isolated. Moreover, engineering new mutants using CRISPR/Cas9 genome editing has become astonishingly efficient in *C. elegans* (Arribere et al. 2014; Dickinson et al. 2015; Paix et al. 2017; Dockshin et al. 2018). *C. elegans* has 2 sexes: hermaphrodite and male. The hermaphrodite's germline proliferates inside 2 gonad arms and produces both oocytes and sperm (Figure 1-2). Hermaphrodites can fertilize their oocytes by using their own sperm to create 'self' progeny or by using sperm from a male to create 'outcross' progeny. The male's germline proliferates inside 1 gonad arm and only produces sperm. In both sexes, the germline is organized as an 'assembly line'. At the distal end of the gonad, the somatic distal tip cell (DTC) signals to Notch/GLP-1 in the germline, to maintain a pool of mitotic germline stem cells (Kimble and Simpson, 1997; Seydoux and Schedl, 2001; Crittenden et al. 2003). As germ cells migrate away from the DTC (i.e. move proximally), the strength of Notch signaling decreases, causing those cells to enter meiosis. This first occurs during the L3 larval stage (Hubbard and Greenstein, 2005). Hermaphrodites first produce sperm during the L4 larval stage and store them in the 2 spermathecae during the L4 larval stage. They then switch completely to producing oocytes during adulthood. In contrast, males produce sperm continuously throughout adulthood.

In the embryo following fertilization, the germline or 'P' lineage is set apart

from somatic lineages by a series of 4 asymmetric cell divisions (Figure 1-2). Each

asymmetric division generates 1 somatic blastomere and 1 P blastomere. The final

asymmetric cell division generates the somatic blastomere D and the germline

blastomere and founder cell P4 (Strome and Wood, 1983). P4 is committed to

germline fate and gives rise to all germ cells. P4 divides symmetrically to generate

the 2 sister PGCs Z2 and Z3, which arrest in the G2 phase of the cell cycle and

remain physically connected due to incomplete cytokinesis for the remainder of

embryogenesis (Fukuyama et al. 2006; Bauer et al. 2021). Once the L1 larva hatches

and begins to feed, the PGCs launch their gene expression program and re-enter the

cell cycle (Strome and Updike, 2015). Amazingly, the PGCs proliferate into over a

thousand germ cells in adults (Kimble and White, 1981).

Specifying germline fate

While mechanisms of germline specification differ across species, there

appear to be 2 general modes: preformation and induction (Strome and Updike, 2015)

(Figure 1-3). In organisms that have a preformed germline, such *as Drosophila*

melanogaster, *Xenopus laevus*, and *Danio rerio*, germline identity is continuous; it is

transmitted from the mother's oocyte to the offspring's PGCs via 'germ plasm'. Germ

plasm specifies the germline in *Drosophila* and *Xenopus* embryos. In both systems,

transplantation of germ plasm to ectopic sites is sufficient to create functional germ

cells at those sites (Illmensee and Mahowald, 1974; Tada et al. 2012). Well-known

3

and rigorously studied components of germ plasm are small membrane-less cytoplasmic condensates called germ granules, or P granules in worms (Strome and Wood, 1982; Updike and Strome, 2010). 2 lines of evidence from *Drosophila* suggest that germ granules are germline determinants. First, mutants that lack germ granules do not make a germline (Nusslein-Volhard et al. 1987). Second, engineered mislocalization of the germ-plasm mRNA *oskar* to the wrong region of the embryo causes germ granules and germ cells to form at that region (Ephrussi and Lehmann, 1992; Smith et al. 1992). Organisms that specify their germline by induction, such as mouse and human, do not inherit maternal germ plasm. Instead, germline identity must be newly established in a subset of embryonic cells via inductive signaling in each generation (Strome and Updike et al. 2015). For example, in mouse post-implantation embryos, BMP4 signaling from extra-embryonic tissue induces WNT3-primed epiblast cells to activate a suite of transcription factors that collectively specify germline fate (Saitou M., 2002; Aramaki et al. 2013; Magnusdottir et al. 2013).

C. elegans embryos inherit and segregate germ granules to the PGCs, suggesting that their germlines are preformed (Strome and Wood et al. 1982). However, several observations suggest that germ granules are not necessary or sufficient to specify the germline in *C. elegans*. First, *pptr-1* mutant embryos that fail to partition maternally loaded germ granules to the germline develop into fertile adults (Gallo et al. 2010). Second, elimination of germ granules by simultaneous depletion of 4 essential germ-granule components, GLH-1, GLH-4, PGL-1, and PGL-

3, by RNAi does not impair germline specification (Updike et al. 2014; Knutson and Egelhofer et al. 2017). Third, mis-segregation of germ granules to somatic cells in *mes-1* mutant embryos is not sufficient to convert those somatic cells into germ cells (Strome et al. 1995). Instead, germ granules in *C. elegans* protect germ cells from mis-expressing a somatic program and from developing as somatic cells after the germline has been specified (Updike et al. 2014; Knutson and Egelhofer et al. 2017). This raises the question: how is the germline specified in *C. elegans*? In my thesis, I tested a strong candidate, the chromatin regulator MES-4, for a role in specifying germline.

Global inhibition of transcription in early germ cells

Global inhibition of transcription elongation in early germ cells is a conserved strategy to prevent toxic accumulation of somatic transcripts and acquisition of somatic fates (Seydoux et al. 1996; Nakamura et al. 1996; Martinho et al. 2004, Seydoux and Braun, 2006; Nakamura and Seydoux, 2008). In the *C. elegans* P blastomeres, transcription elongation is globally inhibited by PIE-1 (Pharynx Intestine in Excess-1). PIE-1 binds to positive transcription elongation factor b (P-TEFb), thereby sequestering and preventing it from phosphorylating the C-terminal domain of RNA Polymerase II (Seydoux et al. 1996). Loss of PIE-1 causes the P blastomeres to turn on their transcriptome and develop into pharyngeal muscle and intestine like the somatic blastomere EMS (sister of the P2 germline blastomere) (Mello et al. 1992). *Drosophila* uses a similar strategy to block transcription elongation in early

germ cells (Nakamura et al. 1996; Martinho et al. 2004): Pgc (polar granule

component) binds to and prevents P-TEFb from phosphorylating the CTD of Pol II

(Hanyu-Nakamura et al. 2008). In mouse, transcription is not globally repressed in

early germ cells. Instead, the transcription factor BLIMP1 specifically represses genes

that are required for mesodermal fate (Ohinata et al. 2005; Kurimoto et al. 2008;

Aramaki et al. 2013; Magnusdottir et al. 2013). Thus, while preventing toxic

accumulation of somatic transcripts is essential in early germ cells, different species

have adopted different strategies to do so.

The *C. elegans* germline must continue to globally repress transcription once

the PGCs are born. PIE-1 is degraded in the newly formed PGCs (~100-cell stage

embryo), coinciding with a brief turn-on of transcription that is likely caused by

relieving P-TEFb sequestration (Seydoux and Dunn, 1997). Quick re-establishment

and maintenance of global repression relies on the actions of several chromatin

regulators (Schaner et al. 2003; Schaner and Kelly, 2006). One is the H3K4me2

demethylase SPR-5 (homolog of mammalian Lsd1), which removes the active histone

modification H3K4me2 in PGCs. *spr-5* mutants accumulate H3K4me2 in PGCs over

generations, coinciding with a progressive loss of fertility (mortal germline phenotype

or 'Mrt'). This suggests that a failure to remove H3K4me2 in PGCs damages their

health and immortality (Katz et al. 2009; Nottke et al. 2011; Greer et al. 2014). In

further support of this, loss of the H3K9me2 histone methyltransferase MET-2 also

causes a Mrt phenotype and an accumulation of H3K4me2 in PGCs over generations

(Andersen and Horvitz, 2007). Interestingly, *spr-5; met-2* double mutants have a

synthetic sterile phenotype that manifests within 1 generation (Kerr et al. 2014; Greer et al. 2014; Carpenter et al. 2021). The histone modification H3K36me has also been implicated in maintaining global transcriptional repression in PGCs. Loss of the H3K36 HMT MES-4 or the candidate H3K36me reader MRG-1 causes PGCs to prematurely launch transcription of their genome (Furuhashi et al. 2010; Miwa et al. 2019).

Several chromatin regulators in worms, including MES-4, MRG-1, Polycomb Repressive Complex 2 (PRC2), and the histone chaperones LIN-53 and FACT, prevent the reprogramming of germ cells into somatic cells when a master regulator of somatic development is ectopically expressed in the germline. (Tursun et al. 2011); Patel et al. 2012; Seelk et al. 2016; Kolundzic et al. 2018; Hajduskova et al. 2019; UI Fatima and Tursun, 2020). Loss of any of those chromatin regulators allows germ cells to mis-express some neuronal transcripts and convert toward neurons when the transcription factor CHE-1 is ectopically expressed in the germline. Recently, loss of PRC2 activity was shown to cause upregulation of some neuronal transcripts and to push germ cells toward a neuronal fate even without ectopically expressing CHE-1 (Kaneshiro et al. 2019).

Germ granules

Although germ granules do not specify the germline in *C. elegans*, they are necessary for proper germline development (Kawasaki et al. 1998, Kawasaki et al. 2004, Spike et al. 2008; Updike et al. 2014; Gallo et al. 2010). Loss of germ granules

causes the germline to accumulate somatic transcripts and at low frequency reprogram toward neuronal or muscle cells (Updike et al. 2014; Knutson and Egelhofer et al. 2017). Germ granules are enriched for RNA-binding proteins, such as Vasa (GLH), Nanos (NOS), and several Argonautes that are involved in small RNA processes (Sheth et al, 2010; Updike and Strome, 2010); Updike et al. 2011; Gao and Arkov, 2012). Germ granules also extend the nuclear pore complex by covering the cytoplasmic side of about 75% of nuclear pores. This positions germ granules to intercept mRNAs as they are exported from the nucleus (Updike and Strome, 2011). An attractive model is that germ granules serve as transcriptome-surveillance hubs that distinguish 'self' (germline) transcripts destined for translation from 'non-self' (somatic) transcripts that should not be translated (Knutson and Egelhofer et al. 2017). Key players in distinguishing 'self' vs 'non-self' appear to be the germ granule-localized argonautes CSR-1 and PRG-1 and their associated 22G and 21U small RNAs, respectively (Claycomb et al. 2009; Shirayama et al. 2012; Wedeles et al. 2013; Seth et al. 2013).

The MES chromatin regulators

A forward genetic screen identified the *mes* genes as being maternally required for offspring germline survival and development (Capowski et al. 1991) (Figure 1-4). The name *mes* refers to their mutant phenotype: 'Maternal-effect sterile'. *mes* M+Z- (Maternal MES positive and Zygotic MES negative) mutants are fertile because they inherited MES protein from their mother (M+). Since those

mutants cannot synthesize zygotic MES product (Z-), they produce all sterile *mes* M-Z- offspring that have few or no germ cells. The germline in *mes* M-Z- mutants dies during early larval development. Loss of *ced-3* or *ced-4*, 2 genes required for apoptosis, does not suppress germline death in *mes-3* M-Z- mutants, suggesting that germline death is caused by an apoptosis-independent mechanism, probably by necrosis (Paulsen et al. 1995). Removal of just maternal MES protein renders *mes* M-Z+ mutants sterile, underscoring the critical role of maternal MES protein.

The MES proteins form 2 chromatin regulators called histone methyltransferases (HMTs), which catalyze methylation of histones (Bannister and Kouzarides, 2011). MES-2, MES-3, and MES-6 physically interact to form the worm version of the H3K27 HMT Polycomb Repressive Complex 2 (PRC2), which was first identified in *Drosophila* (Lewis, 1978; Duncan, 1982; Xu et al. 2001; Muller et al, 2002; Fong et al. 2002; Bender et al. 2004). MES-4 is an H3K36 HMT (Bender et al. 2006; Rechtsteiner et al. 2010; Furuhashi et al. 2010; Kreher et al. 2018). How the MES chromatin regulators protect germline survival and immortality is the major question of my thesis work. I hypothesized that the MES chromatin regulators ensure appropriate gene expression patterns in PGCs. In chapter 2, I describe how I tested this hypothesis and my findings.

An overview of chromatin regulation

Multicellular organisms need to develop distinct and specialized tissues. This requires genetically identical cells to launch separate tissue-appropriate gene

expression programs that support tissue form and function. One conserved strategy to launch and maintain different gene expression programs from the same genome is the differential packaging of DNA into chromatin. The basic unit of chromatin is the nucleosome, comprised of a histone octamer (2 of each histone H2A, H2B, H3, and H4) that wraps ~147 bp of DNA and results in 5- to 10-fold DNA compaction (Kornberg, 1974; Felsenfeld and Groudine, 2003) (Figure 1-5). Nucleosomes are repeating units on chromatin and are connected by ~10-80 bp of linker DNA to form 'beads on a string' (Olins and Olins, 1974; Oudet et al. 1975), which is then folded into a 30 nm chromatin fiber for a total compaction of ~50-fold. This fiber is stabilized by histone H1 binding to each nucleosome and the adjacent linker DNA (Noll and Kornberg, 1977; Thoma and Koller, 1977; Robinson and Rhodes, 2006). Several higher orders of chromatin compaction exist beyond the 30 nm fiber (Felsenfeld and Groudine, 2003). By modulating the level of DNA packaging, genetic elements can be made more or less accessible to transcription machinery and a constellation of factors that impact transcription. In support of this, active genes were found to be more sensitive to DNAse I digestion than inactive genes (Weintraub and Groudine et al. 1976). A key question is how DNA packaging is regulated to control transcription.

While extensive regulation likely takes place at all levels of DNA packaging, many in the field have focused on regulation at the level of histones (Felsenfeld and Groudine, 2003; Kouzarides et al. 2007). Histones are conserved basic proteins that each contain a globular domain within the nucleosome and an N-terminal tail that

10

protrudes outside of the nucleosome. Residues in the N-terminal tails are frequently

targeted for post-translational modification; however, some residues within the

globular domain are targeted as well (Bannister and Kouzarides, 2011). Several types

of histone modifications have been identified: methylation, acetylation,

phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deimination, and

proline isomerization (Kouzarides, 2007). I will focus on histone acetylation and

methylation. Histone acetyltransferases (HATs) catalyze histone acetylation using the

substrate acetyl coA (Bannister and Kouzarides, 2011). Histone acetylation 'relaxes'

chromatin by neutralizing positively charged lysine residues in histones, which

weakens the stabilizing electrostatic interactions between histones and negatively

charged DNA. Thus, histone acetylation makes genes more permissible to

transcription. Histone methyltransferases (HMTs) catalyze histone methylation using

the substrate S-adenosylmethionine (Bannister and Kouzarides, 2011). In contrast to

histone acetylation, histone methylation does not neutralize positively charged

residues; instead, it regulates DNA packaging by recruiting effector proteins. One

class of effector proteins is ATP-dependent chromatin remodelers, which use energy

from ATP hydrolysis to manipulate nucleosomes (e.g. promote nucleosome sliding or

histone exchange) and regulate chromatin accessibility (Allis and Jenuwein, 2016).

The 'histone code' hypothesis proposes that the combination of histone modifications

on a gene determines that gene's transcriptional status (Jenuwein and Allis, 2001).

This hypothesis is attractive but has been contested by numerous studies showing that

histone modifications can have context-dependent roles in transcription and can

recruit different chromatin reader or effector complexes. For example, trimethylation of Lysine 36 in the tail of histone 3 (H3K36me3) has both repressive and activating roles; it represses spurious transcription in coding regions of the yeast genome and increases transcription of X-linked genes in the *Drosophila* male (Carroza et al. 2005; Bell et al. 2008; Venkatesh and Workman, 2013). My thesis work focused on 2 histone modifications, H3K36me3 and H3K27me3 (Figure 1-6).

The marks of MES: H3K36me3 and H3K27me3

H3K27me3 causes transcriptional repression and is found in facultative heterochromatin on developmentally regulated genes (Azuara et al. 2006; Boyer et al. 2006; Jorgensen et al. 2006; Lee et al. 2006; Trojer and Reinberg, 2007; Margueron and Reinberg, 2011). H3K27me3 is catalyzed by the conserved Polycomb Repressive Complex 2 (PRC2). One model for how H3K27me3 causes repression is by recruitment of Polycomb Repressive Complex 1 (PRC1). PRC1 binds to H3K27me3 through its chromodomain and is thought to repress transcription by catalyzing ubiquitylation of H2AK119 and by regulating higher-order chromatin structure independently from its role in histone modification (de Napoles et al. 2004; Wang et al. 2004; Eskeland et al. 2010; Margueron and Reinberg, 2011). MES-2, MES-3, and MES-6 form the worm version of Polycomb Repressive Complex (PRC2) (Xu et al. 2001; Ketel et al. 2005) (Figure 1-6). The worm complex catalyzes both H3K27me2 and H3K27me3 (Fong et al. 2002; Bender et al. 2004). MES-2 is the HMT subunit of PRC2 that contains a SET domain and is homologous to

Drosophila E(Z) (Enhancer of Zeste) and human EZH2 (Holdeman et al. 1998; Cao et al. 2002; Kuzmichev et al. 2002; Bender et al. 2004). MES-6 is homologous to *Drosophila* ESC (Extra Sex Combs) and mammalian EED (Korf et al. 1998). MES-3 is a *C. elegans*-specific component of PRC2 (Paulsen et al. 1995). In *C. elegans*, loss of PRC2 eliminates H3K27me3 in most regions of the germline but not in the soma, suggesting the existence of an additional H3K27me3 HMT (Bender et al. 2004).

H3K36me3 is associated with actively transcribed genes and has context-dependent roles in transcriptional regulation (Bannister et al. 2005; Rechtsteiner et al. 2010). In yeast, the HMT Set2 catalyzes all levels of H3K36 methylation (me1, me2, and me3) co-transcriptionally by binding (via its SRI domain) to the hyperphosphorylated C-terminal domain (CTD) of elongating RNA Polymerase II (Strahl et al. 2002; Li et al. 2003; Kizer et al. 2005; Venkatesh and Workman, 2013). H3K36me3 deposition in the wake of elongating RNA Pol II recruits Rpd3S (Reduced potassium deficiency 3 small), a histone deacetylase complex that deacetylates nucleosomes, to close chromatin and prevent spurious transcription from cryptic promoters in gene bodies (Keogh et al. 2005; Carrozza et al. 2005). On the other hand, H3K36me3 has an activating role in *Drosophila* X-chromosome dosage compensation (Bell et al. 2008). In XY male flies, the HMT Hypb catalyzes H3K36me3 on X-linked loci, which recruits the MSL (male-specific lethal) complex to acetylate nucleosomes via its HAT subunit, MOF (Males absent On the First) (Buscaino et al, 2003; Bell et al. 2008; Wagner and Carpenter, 2012). This mechanism increases X-chromosome transcription by 2-fold in XY males to match

the level of autosome transcription and the level of X-chromosome transcription in XX females (Bell et al. 2008; Lucchesi and Kuroda, 2015). The mammalian NSD H3K36 HMTs have also been shown to promote or repress transcription depending on context (Huang and Zhu, 2018; Nimura et al. 2009; Lucio-Eterovic et al. 2010; Wagner and Carpenter, 2012). MES-4 is 1 of the 2 H3K36 HMTs in *C. elegans* (Figure 1-6)*.* It catalyzes all H3K36me2 and some H3K36me3 (Bender et al. 2006; Rechtsteiner et al. 2010; Furuhashi et al. 2010) (Figure 1-6). MES-4 is homologous to *Drosophila* dMES-4 and mammalian NSD1, NSD2, and NSD3, and is 1 of the few H3K36 HMTs that can catalyze H3K36me3 independently from transcription elongation (Bender et al. 2006; Wagner and Carpenter, 2012). Mammalian absent, small, or homeotic discs 1-like (Ash1l) is another H3K36 HMT with this special ability (Miyazaki et al. 2013). The other H3K36me3 HMT in worms, MET-1, is homologous to yeast Set2 and like Set2 catalyzes H3K36me3 co-transcriptionally. MET-1 has no essential function in *C. elegans*, but I found that its loss does cause low levels of sterility at high temperatures (Kreher et al. 2018). H3K36me is distributed across approximately 20% of the genome and is enriched in gene bodies, in line with it being a mark associated with actively transcribed genes (Rechtsteiner et al. 2010). The role of MES-4 and H3K36me in protecting germline immortality in *C. elegans* is particularly puzzling and poorly understood.

The role of PRC2 and MES-4 in X-chromosome repression

In sexually reproducing organisms, different sexes have the same number of autosomes but often have a different number of X chromosomes. This creates an X-dosage problem that must be resolved by some form of dosage compensation (Straub and Decker, 2007). Intriguingly, different species use widely different dosage compensation strategies. In humans, females randomly inactivate 1 of their 2 Xs to match the transcription level from genes on the male's single X. In flies, transcription of genes on the male's single X is increased 2-fold to match the female's 2 Xs. In somatic cells of worms, a Dosage Compensation Complex (DCC) decreases transcription of genes on each of the hermaphrodite's 2 Xs by 2-fold to match the male's single X (Meyer 2005; Strome et al. 2014).

The vast majority of genes on the X chromosome are repressed during most stages of germline development except oogenesis (Kelly et al. 2002; Arico et al. 2011; Tzur et al. 2018). During meiosis in males, a process called meiotic silencing of unsynapsed chromosomes (MSUC) causes the single unpaired X to be coated and silenced by H3K9me2, a mark of constitutive heterochromatin (Bean et al. 2004; Checci and Engebrecht, 2011; Strome et al. 2014). Silencing of unsynapsed chromatin is a conserved feature of the germline and is thought to act as a surveillance system for diploid genomes, targeting foreign single-copy DNA sequences like transposons (Bean et al. 2004). Due to X-chromosome silencing in the male germline, germline genes that need to be expressed during germline stages shared by both sexes (e.g. proliferation and early meiosis) were selected against being X-linked; however, some remained X-linked, and therefore require some form of

dosage compensation to equalize their expression between the sexes. The *C. elegans* germline does not express the DCC and instead is thought to rely on the MES system for X dosage compensation (Strome et al. 2014). Indeed, X-chromosome repression is a well-documented and major focus of PRC2 and MES-4 in germ cells (Fong et al. 2002; Kelly et al. 2002; Bender et al. 2006; Gaydos et al. 2012, Gaydos et al. 2014; Lee et al. 2017). This raises the hypothesis that PRC2 and MES-4 protect germline immortality by repressing the X chromosomes, perhaps as the germline's version of X dosage compensation.

It was demonstrated that PRC2 protects germline immortality by repressing genes on the X chromosome (Gaydos et al. 2014). The major impact to gene expression in fertile *mes-2* M+Z- mutant adults is mis-expression of hundreds of X-linked genes (Gaydos et al. 2012). This is not surprising since H3K27me3 is concentrated on the X chromosomes in germ cells. Fertility analyses of *mes-3* M-Z-mutants that inherit different endowments of X chromosomes demonstrated that X chromosome repression is PRC2's essential role in the germline (Gaydos et al. 2014). In those analyses, all *mes-3* M-Z- mutant XX hermaphrodites were sterile, as were XO mutant males that inherited their single X from the oocyte and therefore with a history of expression. Intriguingly, most XO mutant males that inherited their single X from the sperm and therefore with a history of repression were fertile. Thus, *mes-3* M-Z- mutants can be fertile if they maintain repression of their X chromosomes. It was further shown that enrichment of repressive H3K9me2 on the sperm-inherited X

was the reason why most *mes-3* M-Z- XO males were fertile and why their germlines were able to maintain X repression (Gaydos et al. 2014).

Like PRC2, MES-4 may protect germline immortality by repressing genes on the X chromosome. Germlines from fertile *mes-4* M+Z- mutant hermaphrodite adults mis-express X genes, many of which are also mis-expressed in germlines from fertile *mes-2* M+Z- mutant hermaphrodite adults (Gaydos et al. 2012). Moreover, *mes-4* M-Z- mutant males that inherited their X from the sperm can be fertile (Garvin et al. 1998). In early embryos and the germline, MES-4 and H3K36me3 are concentrated on the autosomes and the left tip of the X chromosome but are largely absent from the remainder of the X chromosome (Rechtsteiner et al. 2010). This suggests that MES-4 represses X genes indirectly. One model proposes that MES-4-catalyzed H3K36me3 on autosomes repels PRC2 and helps concentrate its repressive activity on the X chromosomes. In support of this model, H3K36me3 is known to antagonize PRC2 activity in vitro and in vivo (Yuan et al. 2011, Schmitges et al. 2011; Gaydos et al. 2012, Evans et al. 2015). Indeed, loss of MES-4 causes acquisition of H3K27me3 in H3K36me3-depleted regions of the genome and reduction of H3K27me3 on the X chromosomes (Gaydos et al. 2012). One observation that challenges this model is that *mes-4* M-Z- mutants have a more severe germline phenotype compared to *mes-3* M-Z- mutants; *mes-4* M-Z- mutant adults completely lack a germline, whereas some *mes-3* M-Z- mutant adults have partially formed germlines (Figure 3-14). Another possible model for how MES-4 indirectly represses the X chromosome is that MES-4-catalyzed H3K36me3 sequesters a transcriptional activator (e.g. a HAT) on

17

autosomes and away from the X chromosome. Recent investigations in yeast and worms support a sequestration role for H3K36me3 (Cabianca et al. 2019; Georgescu et al. 2020).

MES-4 epigenetically transmits a memory of germline

Genome-wide patterns of histone marking can be epigenetically transmitted from the parent germline to offspring and propagated during offspring development. Recently, intergenerational epigenetic transmission of histone marking was demonstrated to impact offspring transcription and health in *C. elegans* (Kaneshiro et al. 2019). A major obstacle to propagating patterns of histone marking during development is DNA replication. Amazingly, during DNA replication modified histones can be passed from the parent chromosome to the 2 daughter chromatids by histone chaperones, providing a mechanism to transmit 'memories' of gene expression (Margueron and Reinberg, 2010). DNA replication dilutes levels of histone marking, since each daughter chromatid receives only a portion of parental histones. To prevent loss of parental patterns of histone marking over the course of several rounds of cell division, HMTs must restore full levels of marking to the daughter chromatids. One model suggests that some HMTs bind to the same mark they catalyze via a chromatin reader domain (e.g. chromodomain), enabling them to 1) target previously marked parental histones that were incorporated into daughter chromatids and 2) catalyze new marking on adjacent unmarked histones (Margueron

et al. 2009; Margueron and Reinberg, 2010; Sankaran et al. 2016; Poepsel et al. 2018).

MES-4 propagates gamete-inherited patterns of H3K36me3 through cell divisions during embryogenesis (Rechtsteiner et al. 2010; Furuhashi et al. 2010; Kreher et al. 2018). MES-4 may propagate its mark using the mechanism described above. MES-4 may bind chromatin via its 3 PHD fingers, which are domains known to bind methylated lysines (Sanchez and Zhou, 2011). In support of this, immunostaining experiments showed that 2 point mutations in the first PHD finger of *mes-4* completely dissociate MES-4 from chromosomes (Bender et al. 2006). However, it is unknown whether MES-4 directly binds H3K36me. 3 lines of evidence from immunostaining experiments show that MES-4 cannot bind to or act on chromosomes that lack pre-existing H3K36me3. First, MES-4 does not catalyze detectable levels of de novo methylation on chromosomes that lack pre-existing H3K36me3 (Furuhashi et al. 2010). Second, in the 1-cell embryo, maternal MES-4 protein associates with wild-type sperm-inherited chromosomes but is not detectable on sperm-inherited chromosomes that lack H3K36me (*met-1; mes-4* double mutant father) (Kreher et al. 2018). Third, an undergraduate in the Strome lab and I generated 'hybrid' embryos that inherited some H3K36me3(+) chromosomes and some H3K36me3(-) chromosomes and found that MES-4 maintained the hybrid pattern of marking through many rounds of cell division (Figure 4-3). Thus, MES-4 seems to play a truly epigenetic role: it maintains gamete-inherited patterns of histone marking through development.

MES-4 transmits an epigenetic 'memory of germline' from parents to offspring. A fascinating observation from chromatin immunoprecipitation (ChIP)-chip data in early embryos was that MES-4 catalyzes H3K36me3 on genes that were previously transcribed in the maternal germline. This suggested a tantalizing model for how MES-4 protects germline immortality (Figure 1-7): First, H3K36me3 is catalyzed on genes expressed in the maternal germline to create a memory of germline, which is then transmitted to the offspring via the oocyte. Second, maternally loaded MES-4 maintains the memory through cell divisions and delivers it to the PGCs. Finally, the PGCs 'deploy' the memory to launch a germline-appropriate gene expression program that specifies germ fate. This deployment may depend on downstream effectors, like the H3K36me3 reader MRG-1. In support of this, loss of MRG-1 causes maternal-effect sterility, like loss of MES-4 and PRC2 (Fujita et al. 2002; Takasaki et al. 2007). MRG-1's homologs (yeast Eaf3, fly MSL3, and mammalian MRG15) are subunits in both histone acetyltransferase (HATs) complexes and histone deacetylase complexes (HDACs), and therefore may activate and/or repress transcription (Joshi and Struhl, 2005; Chen et al. 2009). Loss of MRG-1 also causes maternal-effect sterility. In chapter 3 of my thesis, I tested whether MES-4's transmission of an epigenetic memory of germline is essential for offspring PGCs to launch a germline-appropriate gene expression program.
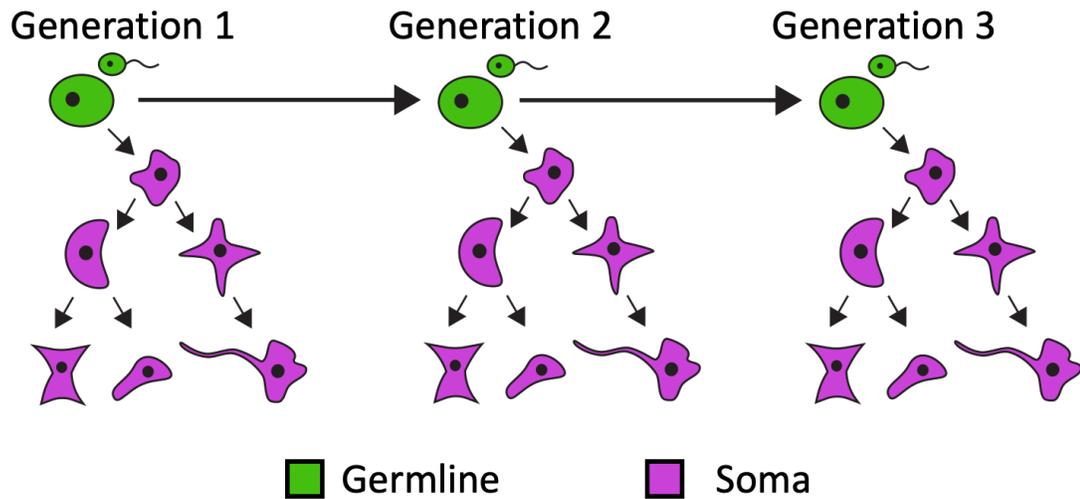
Figure 1-1. Germline totipotency and immortality. The union of an oocyte and sperm (green gametes in the top row) generates a master stem cell (the zygote) that can form every cell type, including somatic cells (purple) and new germ cells (green). This property is called 'totipotency' and is unique to the germline. The germline is also immortal. Somatic lineages die off in each generation, but the germline survives and connects generations. The totipotency and immortality properties of the germline allow it to form each new generation of organism and ensure the survival of a species.
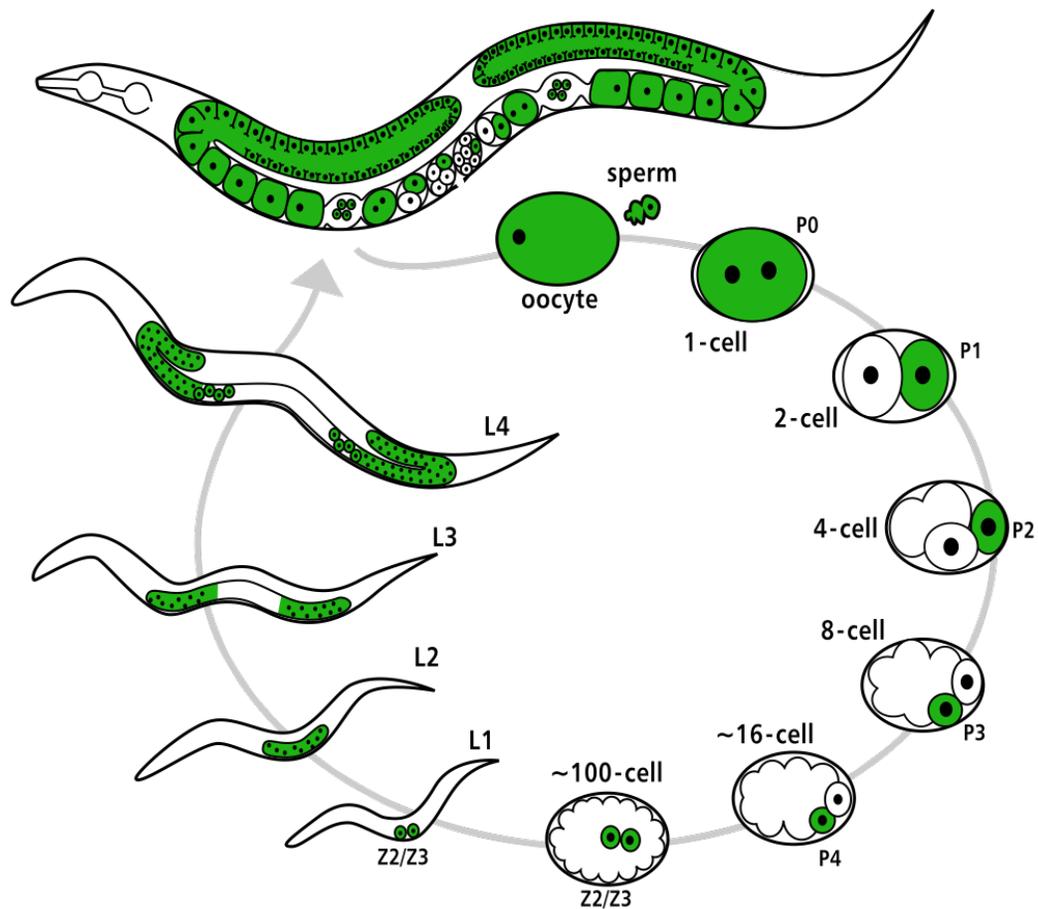
Figure 1-2. *C. elegans* germline development. The *C. elegans* germline (green) is
continuous, meaning that germline identity is present at every developmental stage in
the life cycle. *C. elegans* hermaphrodites (top) have 2 gonad arms, which encase their
germline and generate both oocytes and sperm to make 'self-progeny'. In the embryo
after fertilization, the germline is set apart from somatic lineages by a series of 4
asymmetric cell divisions that each generate 1 somatic blastomere (white) and 1
germline or 'P' blastomere. P4 is the germline founder cell and first primordial germ
cell (PGC). P4 divides symmetrically to form the 2 sister PGCs, Z2 and Z3. The
PGCs remain quiescent until the larva (L1) hatches and feeds, after which the PGCs
re-enter the cell cycle and begin to proliferate. The PGCs generate over 1000 germ
cells through larval development and adulthood.
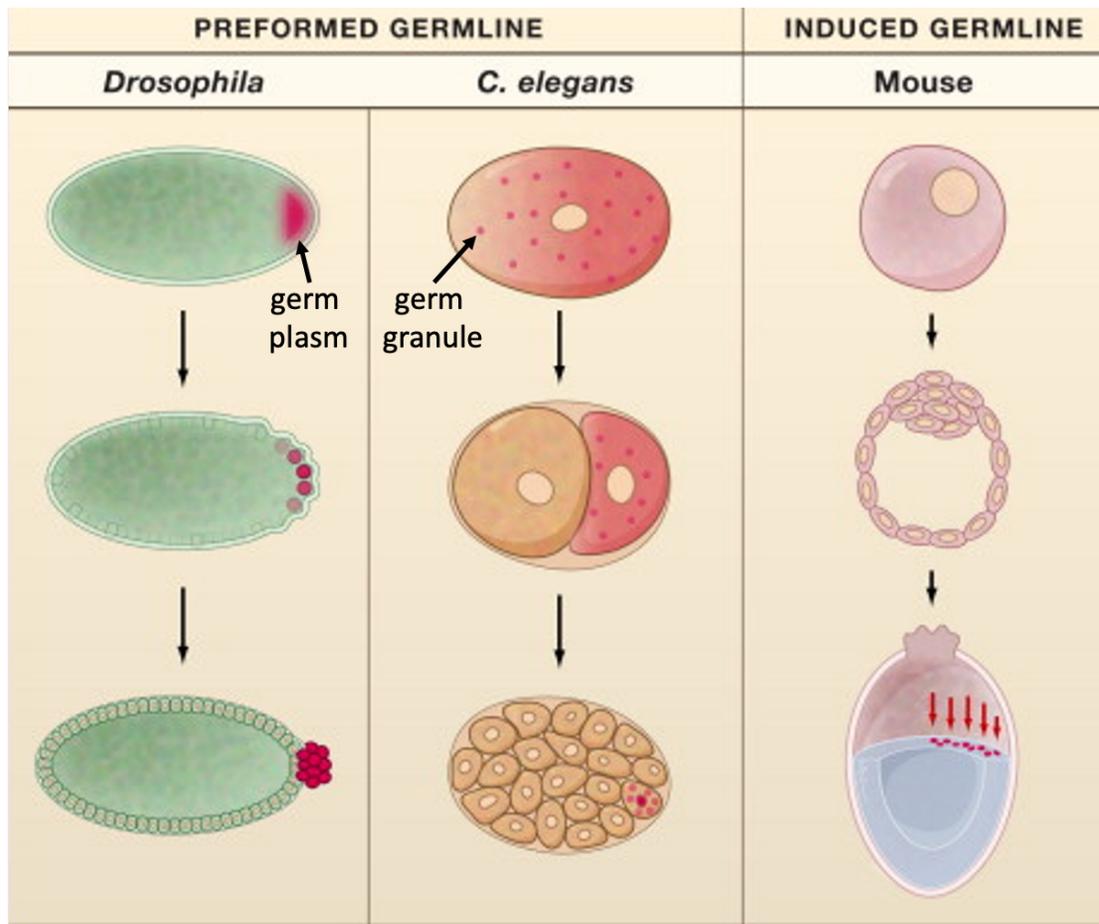
Figure 1-3. Modes of germline specification. Preformation and induction are the 2 modes of germline specification. The germline in *Drosophila melanogaster* and *Caenorhabditis elegans* is preformed by the partitioning of maternally provided germ plasm during embryogenesis. Germ granules are components of germ plasm that are the premier candidates for being germline determinants. They indeed appear to determine the germline in *Drosophila*. However, several observations have demonstrated that they do not determine the germline in *C. elegans* (Gallo et al. 2010; Updike et al. 2014; Knuston and Egelhofer et al. 2017). Mice, like other mammals, must newly establish germline identity in a subset of embryonic cells by induction from extraembryonic tissue. Figure adapted from Seydoux and Braun, 2006.
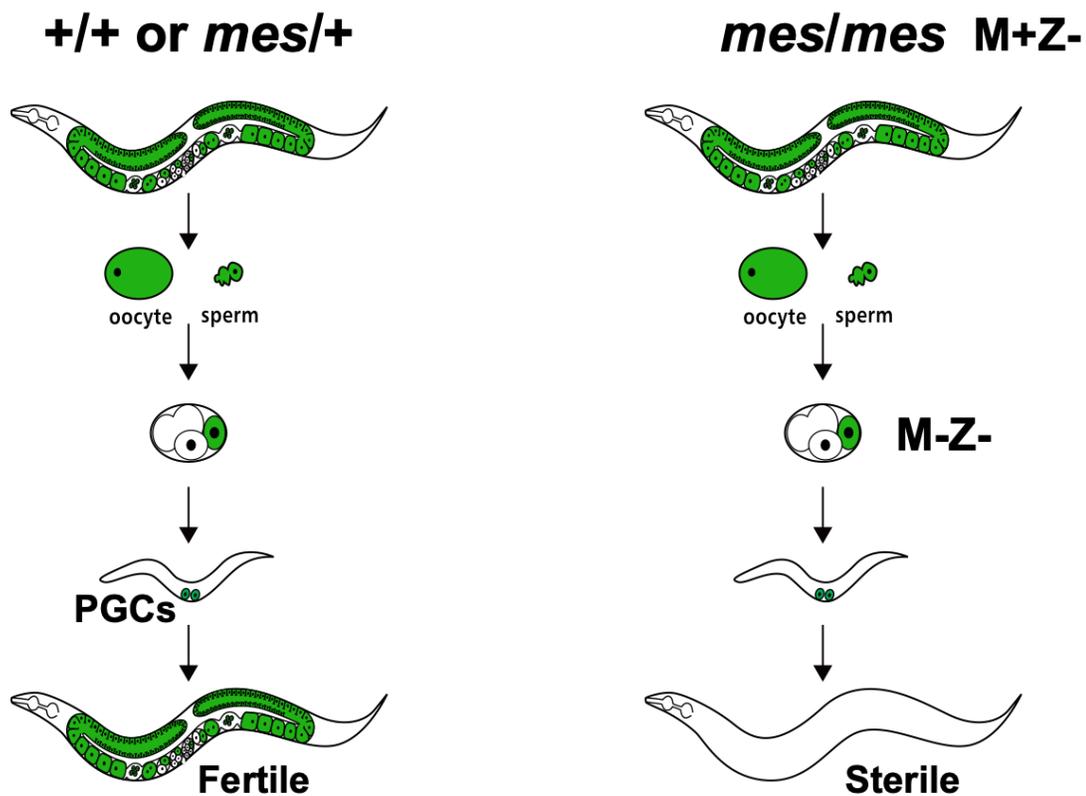
Figure 1-3. Mutation in *mes* genes causes maternal-effect sterility. The *mes* genes were identified by the Strome lab in a forward genetic screen as genes required maternally for germline development in offspring. Wild-type (+/+) or *mes/+* heterozygous mothers synthesize MES protein and deliver it to their offspring. As a result, all offspring, even *mes/mes* homozygous mutant offspring, develop into fertile adults.. Fertile *mes/mes* (or just '*mes*') animals inherited maternal MES protein but cannot synthesize new or zygotic MES protein, and are therefore called M+Z- for Maternal *mes(+)* and Zygotic *mes(-)*. Offspring of *mes* M+Z- mothers are *mes* M-Z- because they do not inherit maternal MES protein or synthesize zygotic MES protein. As a result, all *mes* M-Z- offspring develop into sterile adults that lack a germline. Their PGCs are born and begin to proliferate, but the nascent germline dies during early larval development.
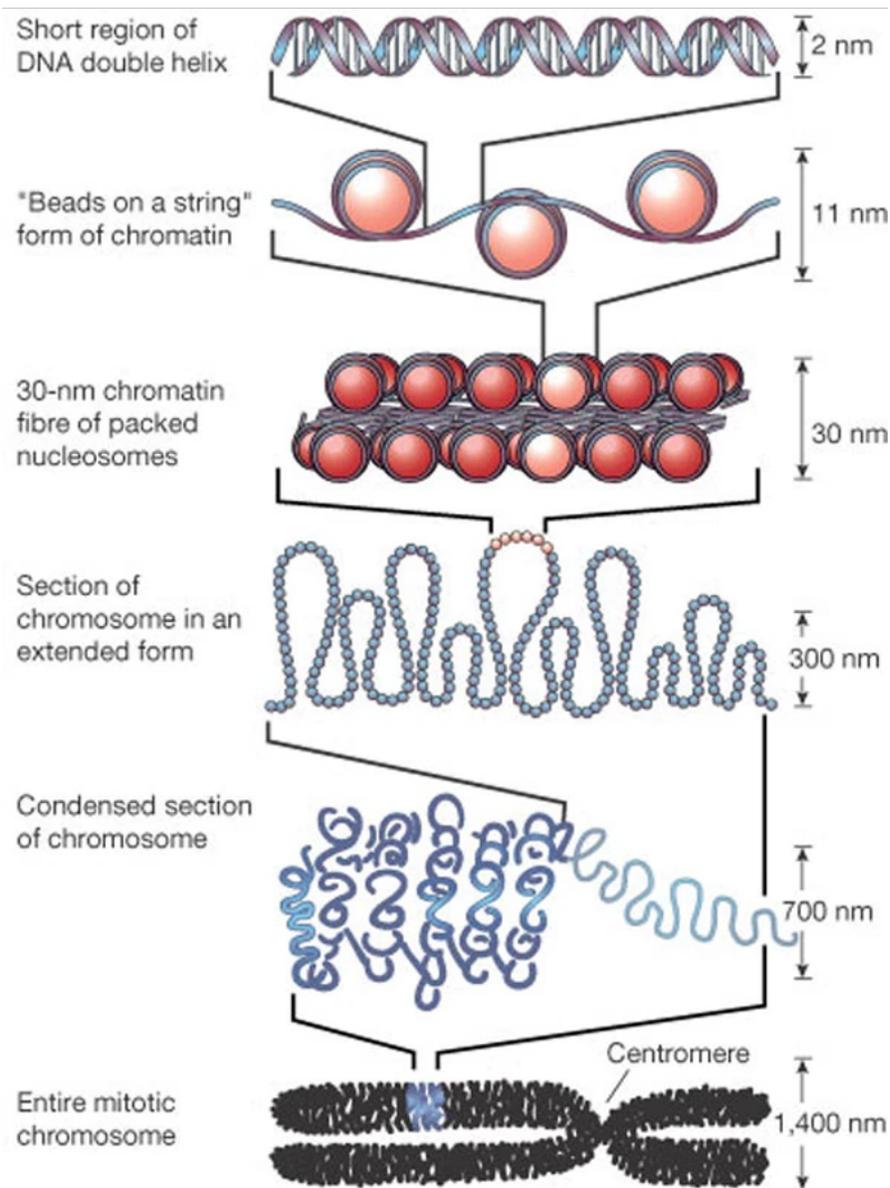
Figure 1-5. Packaging of DNA into chromatin. The basic and repeating unit of chromatin is the nucleosome (spheres) that wraps DNA approximately 2 times. DNA is packaged into progressively higher orders of compaction. Figure adapted from Felsenfeld and Groudine, 2003.
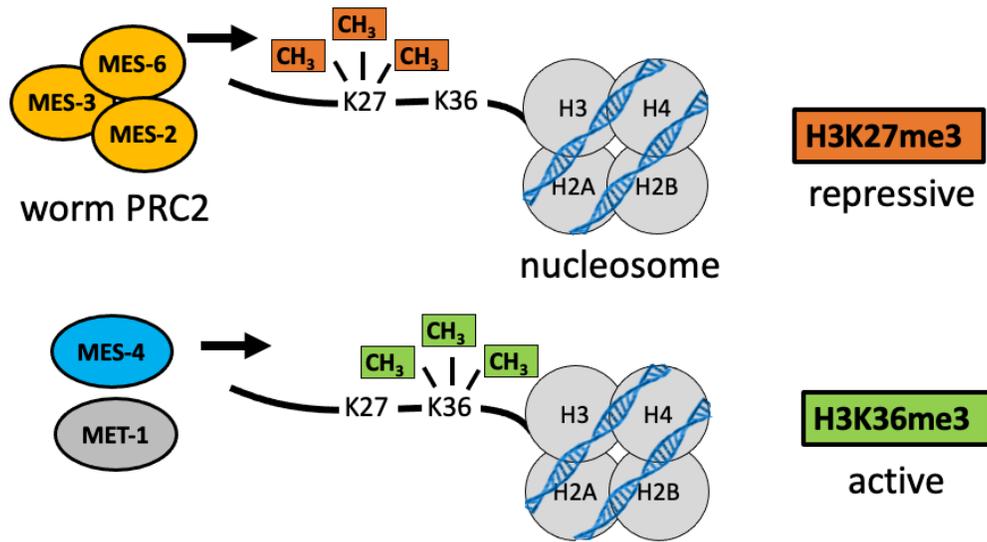
Figure 1-6. The MES proteins are histone methyltransferases. A nucleosome contains an octamer of histones (2 H2A, 2 H2B, 2 H3, and 2 H4). Histones have a globular domain (gray circles) and an N-terminal tail that protrudes from the nucleosome. Those tails are frequently targeted for covalent modification. 2 histone modifications are trimethylation of lysine 27 and trimethylation of lysine 36 in the tail of histone 3, called H3K27me3 and H3K36me3, respectively. H3K27me3 causes transcriptional repression and is catalyzed by the worm version of Polycomb Repressive Complex 2 (PRC2), which is composed of MES-2, MES-3, and MES-6. MES-2 contains a SET domain and is the catalytic subunit. H3K36me3 is associated with active transcription and is catalyzed by MES-4 and MET-1.
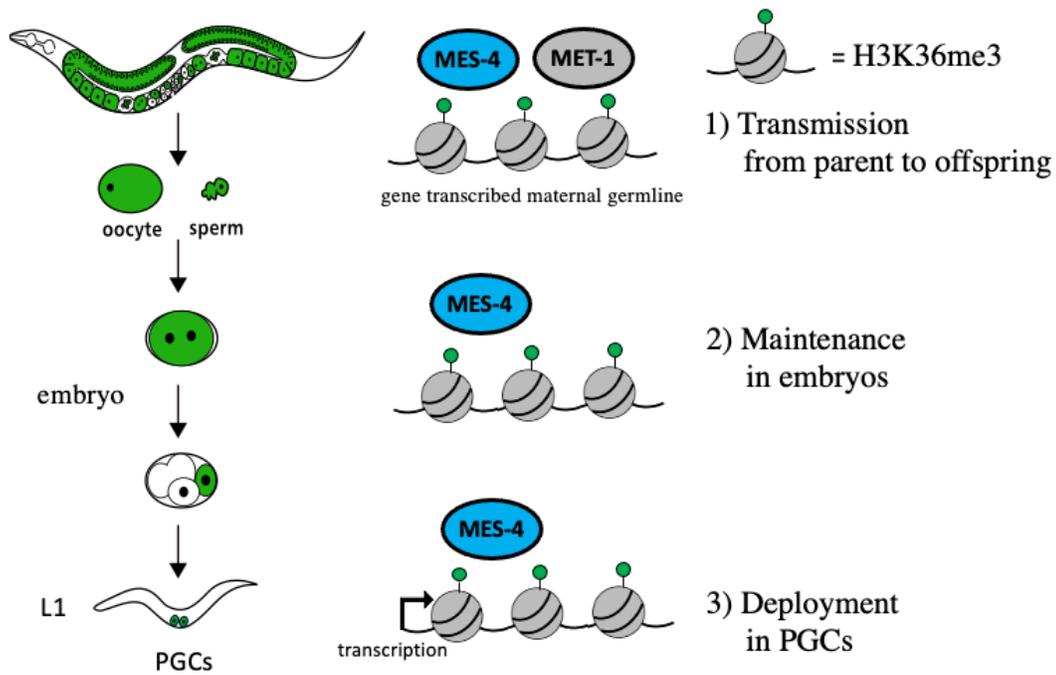
Figure 1-7. MES-4 transmits an epigenetic memory of germline. Chromatin immunoprecipitation (ChIP)-chip experiments in early-embryo extracts found that MES-4 maintains H3K36me3 on genes that were previously transcribed in the maternal germline (Rechtsteiner et al. 2010; Gaydos et al. 2012). Transmission of this epigenetic 'memory of germline' from parents to offspring may be critical for PGCs to acquire germline fate and launch a germline-appropriate gene expression program. Our model is the following: 1) the H3K36 HMTs, MES-4 and MET-1, establish a memory of germline in the maternal germline by marking transcribed genes with H3K36me3 (green circles). This memory is transmitted to offspring via the oocyte. 2) Maternally loaded MES-4 maintains the memory in embryos by restoring high levels of marking after each round of cell division, eventually delivering the memory to the PGCs. 3) The memory is 'deployed' in PGCs to ensure that they transcribe the correct genes for germline fate.

CHAPTER 2

A Primer for Generating and Using Transcriptome Data and Gene Sets

The text and figures in this chapter are excerpted from the previously published paper: Cockrum, C., Kaneshiro, K. R., Rechtsteiner A., Tabuchi, T. M., and Strome, S. (2020). A primer for generating and using transcriptome data and gene sets. *Development (Cambridge, England)* 147, dev193854.

ABSTRACT

Transcriptomic approaches have provided a growing set of powerful tools with which to study genome-wide patterns of gene expression. Rapidly evolving technologies enable analysis of transcript abundance data from particular tissues and even single cells. This Primer discusses methods that can be used to collect and profile RNAs from specific tissues or cells, process and analyze high-throughput RNA-sequencing data, and define sets of genes that accurately represent a category, such as tissue-enriched or tissue-specific gene expression.

INTRODUCTION

Analysis of gene expression patterns is a common and important component of many modern biological research projects. Such analysis can provide insights into how gene expression patterns drive cell fate and function, and how mutations, drugs, diseases, physiological stimuli, and stress impact gene expression programs. The first level of analysis of gene expression patterns is quantifying levels of gene transcripts. Such 'transcriptome' analysis often involves determining which RNAs are characteristic of certain cells, tissues, or stages of development. This requires the availability of high-confidence lists of RNAs in those samples. Investigators face the

challenges of deciding how to profile RNA populations, comparing their RNA profiles with already-published profiles and determining which transcripts are characteristic of particular cells, tissues, stages, mutants or interventions.

This Primer – intended as an overview for those new to transcriptome analysis – describes widely used methods for isolating cells and tissues and preparing samples for transcript profiling, and discusses considerations in processing RNA-sequencing (RNA-seq) data and generating lists of genes or 'gene sets' expressed in particular cell types. We use examples from the nematode *Caenorhabditis elegans*, but stress that the lessons and considerations extend across systems. In the last few years, single-cell RNA-seq has become increasingly popular. Because processing and analyzing such data involve many unique considerations and have been reviewed elsewhere, we do not delve deeply into single-cell RNA-seq but instead refer readers to reviews devoted to this subject for further details (Hwang et al. 2018; Luecken and Theis, 2019).

STRATEGIES FOR ISOLATING TISSUES AND CELLS FOR TRANSCRIPT PROFILING

Transcriptome data from whole animals or embryos can provide important stage information, such as how a single mutation impacts an organism's transcriptome through its life cycle. Furthermore, it is typically straightforward to obtain large amounts of RNA for profiling (depending on the organism). More commonly, however, researchers are likely to want transcriptome data from particular tissues or

cell types. Below, we describe approaches to isolate tissues and cell types from which RNAs can be extracted (see Figure 2-1 and examples in Table 2-1).

Comparison of animals that have versus lack a tissue or cell of interest

Comparing the transcript profiles of animals that contain (e.g. wild-type animals) versus lack a tissue or cell type of interest (e.g. mutant animals) can identify transcripts enriched in that tissue/cell type (Figure 2-1A). For example, Reinke et al. identified genes that are expressed more highly in *C. elegans* germline tissue than in somatic tissues (i.e. germline enriched) by comparing levels of transcripts in wild-type adults (germline plus soma) and *glp-4* mutant animals that lack a germline (just soma) (Reinke et al. 2000, 2004). One advantage of this technique is that animals can be harvested in large quantities to yield sufficient RNAs for sensitive profiling. Although powerful, this genetic strategy assumes that all differential expression between animals is due only to the presence versus absence of the tissue/cell type, which may not always be the case; there may, for example, be compensatory changes in gene expression in other tissues. Furthermore, mutants that cleanly remove a tissue/cell type from living animals may not be available.

Dissection of tissues

One common way to isolate tissues or cells from animals for RNA profiling is by hand dissection (Figure 2-1B). This can be faster (and potentially less disruptive to the tissue) than other techniques that require several time-consuming steps, such as fluorescence-activated cell sorting (see below), but can be labor-intensive and require

specialized dissection skills to collect enough material for RNA-seq (typically many thousands of cells). Some new technologies allow profiling from one or a few samples of tissue or cells, which can significantly reduce the hands-on time required to isolate enough material by dissection. Two important considerations are whether a particular tissue can be cleanly separated from surrounding tissues (for example, the germline is encased by the somatic gonad, making them difficult to separate) and the degree of cell-type complexity of the tissue of interest. Because the presence of a mixture of tissues or cell types may confound the specificity of a gene set, transcriptome data should be interpreted with caution. In some cases (particularly for small organisms or early developmental stages), hand dissection of a particular tissue may not be feasible.

Fluorescence-activated cell sorting (FACS)

To isolate specific cell types that are difficult to isolate by dissection or to obtain sufficient quantities of cells for sensitive transcript profiling, cell sorting approaches can be used, provided the tissue of interest can be efficiently dissociated into single cells (Figure 2-1C). Cell sorting by FACS requires either cell type-specific expression of a fluorescent protein (e.g. GFP) or application of a fluorescent antibody that binds to a cell type-specific cell surface marker. In either case, cells can be dissociated from large quantities of animals, after which a FACS machine separates fluorescent cells from non-fluorescent cells. One important consideration for FACS is the cell purity after sorting. Some variables that can influence cell purity are the level

and specificity of fluorescent protein or surface marker expression, the stringency of

the fluorescence threshold, and sorting time. FACS can take several hours, during

which time cells may deteriorate, lose fluorescence intensity, and sustain alterations

to gene expression (e.g. misexpression of genes involved in stress responses). It is

important to ensure that the fluorescent protein (or cell surface marker) is highly

expressed and specific to the cell type of interest and that cell isolation and FACS

sorting time are limited. CRISPR gene-editing technology has made it straightforward

to engineer cell type-specific expression of fluorescent proteins in many organisms

(e.g. Dickinson and Goldstein, 2016; Paix et al. 2015; Wang et al. 2016; Dokshin et

al. 2018).

mRNA 'tagging' in cells or tissues of interest

Instead of profiling mRNAs from isolated cells or tissues, an alternative

strategy is to 'tag' mRNAs in a cell type of interest and then selectively purify those

tagged mRNAs (Figure 2-1D). For instance, Roy et al. genetically engineered *C.*

elegans to express FLAG-tagged PAB-1 [poly(A) binding protein-1] only in specific

cells; mRNAs bound to FLAG::PAB-1 in those cells were then co-

immunoprecipitated using anti-FLAG antibodies (Roy et al. 2002). Unlike FACS,

mRNA tagging does not require dissociation of cells from tissues. Therefore, in cases

for which cell dissociation is difficult, mRNA tagging may be a preferable choice. A

few considerations are the need to select a promoter that will drive expression of

tissue-specific FLAG::PAB-1, the assumption that FLAG::PAB-1 expression will not

impact tissue health, and the need to formaldehyde cross-link transcripts to

FLAG::PAB-1 to facilitate robust co-immunoprecipitation of mRNAs, which can

introduce background signal. Notably, an improved mRNA tagging method called

'PAT-seq' [poly(A) tagging sequencing] has been developed (Blazie et al. 2015).

Improvements to this method include enhancing the efficiency of FLAG

immunoprecipitation by driving expression of a PAB-1 transgene with three instead

of one FLAG tag and driving more reliable PAB-1 expression from a single-copy

transgene inserted into the genome (instead of relying on variable expression from an

extrachromosomal array).

Isolation of single cells

In recent years, it has become feasible to profile transcripts from single cells

(Hwang et al. 2018; Luecken and Theis, 2019) (Figure 2-1E). A huge advantage of

this strategy is that it provides cell type-specific transcriptome data. In some cases, it

may be possible to identify cells of interest by size and/or shape and isolate them by

hand dissection; an example is *C. elegans* early embryos (Osborne Nishimura et al.

2015). Alternatively, cells can be identified by cell type-specific expression of a

fluorescent protein (as used for FACS) and then manually isolated. In some cases,

researchers may instead aim to compare single-cell transcriptomes between hundreds

to thousands of different cell types (e.g. to determine cell-type heterogeneity within a

tumor). For such a high-throughput project, a common strategy uses microfluidics to

separate a suspension of cells into thousands of individual tiny droplets, each of

which will contain a single cell from the suspension and a 'cocktail' of ingredients to

prepare an RNA-sequencing library (Xu et al. 2019). Although the dissociation of

cells from animals and tissues is straightforward and inexpensive, the delivery of

single cells to individual droplets and subsequent preparation of many sequencing

libraries is expensive. Moreover, rare transcripts are often not detected or

inconsistently detected among biological replicates (termed 'dropouts');

consequently, transcriptome information may be incomplete and/or variable. Owing

to the high variance in single-cell RNA-seq experiments, they typically require many

more replicates compared with conventional RNA-seq experiments. As single-cell

technologies continue to improve, such issues are likely to become less of a barrier to

the use of these technologies.

GENERATING AND PROCESSING RNA-SEQ DATA

After isolating RNAs from a biological sample, the next goal is to measure

how many RNAs in the sample were produced from each gene (Figure 2-2A).

Although many studies use the term 'gene expression' to describe a gene's

transcriptional activity, most RNA-profiling techniques measure RNA abundance,

which is impacted not only by transcription but also by transcript processing and

degradation. Popular genome-wide RNA-profiling techniques include high-

throughput sequencing of cDNA or RNA libraries, hybridization to DNA

microarrays, and serial analysis of gene expression (SAGE). These approaches are

reviewed elsewhere (Schulze and Downward, 2001; Yamamoto et al. 2001; Stark et

al. 2019). The current 'go-to' method is next-generation sequencing (NGS) of cDNA

libraries (e.g. Illumina, PacBio, Ion Torrent, and Oxford Nanopore technologies). In this section, we discuss some considerations and best practices for preparing cDNA libraries for high-throughput sequencing and for analyzing and visualizing sequencing data (see Figure 2-2).

cDNA library preparation

First, researchers must decide whether to generate cDNA libraries from polyA(+)-selected mRNAs (mature transcripts with a polyA tail) or from RNA depleted of ribosomal RNAs. One key difference between these methods is the library sequencing depth needed to detect transcripts of interest, typically transcripts produced from protein-coding genes. Libraries made from polyA(+)-selected mRNA samples are highly enriched for protein-coding transcripts, and thus usually require less sequencing depth to detect transcripts of interest. Although libraries made from rRNA-depleted RNA samples should be enriched for protein-coding transcripts, they may still have a large representation of rRNA, requiring deeper sequencing to detect protein-coding RNAs. However, rRNA-depleted RNA samples do have the advantages of detecting non-polyadenylated transcripts (e.g. some histone transcripts, pre-mRNAs and non-coding RNAs) and avoiding the biases of polyA(+)-selection toward enrichment of transcripts with longer polyA tails and sequencing coverage skewed toward the 3′ end of genes.

Deciding on the appropriate sequencing depth largely depends on how much sampling of a library is needed to detect transcripts of interest. More depth is required to detect rare transcripts and when the diversity or 'complexity' of transcripts is high

(e.g. many thousands of genes are expressed, or many transcripts are alternatively spliced). Because experimental design and goals vary widely, we recommend choosing a sequencing depth based on that reported in previous studies with a similar experimental design, or on the results of a pilot experiment. One way to identify an optimal sequencing depth is a saturation analysis, which helps identify a minimum depth needed to detect the majority of transcripts or to have sufficient statistical power for differential expression testing (Wang et al. 2012; Robinson and Storey, 2014).

Researchers must also decide how many biological replicates to prepare for sequencing. Biological replicates are cDNA libraries prepared from different collections of the same type of biological sample (e.g. the same tissue). Biological replicates are necessary to test an experiment's reproducibility, to determine biological variation and to perform statistical tests such as differential expression. It is best practice to sequence a minimum of three biological replicates of cDNA libraries; however, if high variation between replicates is expected, then more than three should be sequenced. To estimate the number of replicates needed for differential expression analysis, a power calculation can be performed using predicted values for sample variance and effect size (Conesa et al. 2016).

Quality control of sequencing data

The first step in processing high-throughput sequencing data is quality control (QC) analysis of the raw 'reads'. QC analysis of reads can detect technical problems that occurred during library preparation or sequencing, such as the presence of PCR artifacts or sample contamination. Some common and user-friendly tools for QC analysis are FastQC and fastp (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; Chen et al. 2018), which analyze the sequence quality and GC content of reads, the presence of adapters and duplicated reads, and more. Acceptable QC metrics depend on experimental design (for general recommendations and best practices, see Conesa et al. 2016 and Koch et al. 2018).

Mapping of reads to a genome

To identify the genomic locus from which each read originated, reads are mapped to the organism's reference transcriptome or genome. Some popular mapping tools are STAR and HISAT2 (Dobin et al. 2013; Kim et al. 2019). For a high-quality sample, more than 70% of reads are expected to map to the genome. A low percentage of mapped reads may be a symptom of sample contamination with RNA from a different organism or excess primers or library adapters in the sequenced library. Another useful analysis for checking sample quality is visualization of read coverage over genes and exons in a genome browser (e.g. the UCSC genome browser) or metagene profile. Non-uniform coverage may reveal issues that occurred during sample and/or library preparation, such as RNA degradation or inefficient reverse transcription. Commonly used tools to perform QC of the mapping step

include Picard and Qualimap (http://broadinstitute.github.io/picard/; Okonechnikov et al. 2016).

Estimation of transcript abundance

To estimate transcript abundance, the numbers of mapped reads per transcript can be counted with tools such as HTSeq or featureCounts (Anders et al. 2015; Liao et al. 2014). Samples are then normalized to account for biases, such as sequencing depth and transcript length. The choice of normalization strategy can have a large impact on data interpretation and depends on the research goal.

If the research goal is to compare gene expression within a sample, then RPKM (reads per kilobase of transcript per million reads), FPKM (fragments per kilobase of transcript per million mapped reads) and TPM (transcripts per million) are frequently used to normalize for transcript length and library-size effects and to report transcript abundance (Mortazavi et al. 2008; Li et al. 2010). For RPKM/FPKM, the first step corrects for library size by dividing each gene's reads or fragments by the total reads/fragments in millions. The second step corrects for transcript length bias by dividing each gene's library size-adjusted counts (values after the first step) by the respective transcript length in kilobases. TPM is related to RPKM/FPKM except that the order of normalization steps is reversed. First, a gene's read or fragment count is divided by transcript length in kilobases to obtain 'transcripts' per gene. Then, each gene's number of transcripts is divided by the library's total transcripts in millions. An advantage of using TPM is that genes with the same read/fragment coverage will

contribute equally to the total number of transcripts in the library, regardless of gene length. In contrast, with RPKM/FPKM, longer genes tend to contribute more to the total number of reads in the library. Variation in the degree of this bias can cause RPKM/FPKM values to differ between sample libraries, making it challenging to compare a gene's RPKM/FPKM value between samples. For this reason, TPM is now widely preferred over RPKM/FPKM. For a more detailed explanation and comparison of RPKM, FPKM and TPM, see https://rna-seqblog.com/rpkm-fpkm-and-tpm-clearly-explained/. If the research goal is to compare transcript abundance between two samples, it is more appropriate to use statistical analysis of differential expression (see below), which requires other normalization methods that account for transcript composition biases (Wagner et al. 2012; Zhao et al. 2020).

After normalization, it is good practice to perform another round of QC by comparing transcript profiles of all samples in the data set, usually by clustering analyses. Principal component analysis (PCA; Figure 2-2B) is a common strategy to find and visualize a few dimensions that explain most of the variance between transcript profiles (called 'dimensionality reduction'), which helps to quickly identify transcript profiles that are similar or different (Wall et al. 2003; https://rna-seqblog.com/statquest-pca-clearly-explained/). A heatmap of hierarchical clustering (Figure 2-2C) is another common visualization strategy. Biological replicates should have similar transcript profiles, and so should cluster together and away from samples that are expected to have different profiles (e.g. samples from different conditions). PCA and hierarchical clustering can also detect batch effects (differences between

samples that are due to their preparation and/or sequencing in separate batches) and other technical biases (for descriptions of clustering analysis, batch effects, and best practices, see Conesa et al. 2016 and Koch et al. 2018).

Differential expression (DE) analysis

Identifying differentially expressed genes between two conditions is a common goal in transcriptome analyses. Several widely used and free tools, such as edgeR and DESeq2, can identify and assign statistical significance to differentially expressed genes from raw count data (Robinson et al. 2010; Love et al. 2014). edgeR and DESeq2 have detailed and user-friendly explanations of the software, to help users understand and choose appropriate parameter settings for their analysis. Both packages provide methods that normalize for differences in library depth based on mapped read counts and are appropriate for comparisons between samples. Another strategy considers only reads mapped to a set of 'negative' or 'control' transcripts. For example, 'spike-in' transcripts added in equal amounts to all samples during library preparation are assumed to have equal abundances in the transcript profiles (Jiang et al. 2011). In a typical RNA-seq experiment, thousands of genes are separately tested for DE, creating a multiple hypothesis testing burden that increases the number of false positives: many genes that are truly not differentially expressed have low P-values and are incorrectly deemed differentially expressed. edgeR and DESeq2 address this problem by adjusting the P-value for multiple hypothesis

testing. Commonly, a gene is called differentially expressed if its adjusted *P*-value (also called q-value or false discovery rate; FDR) is below 0.05 or 0.01.

Two popular ways to visualize DE analysis are MA plots (Figure 2-2D) and volcano plots (Figure 2-2E). An MA plot shows, for each transcript, the log fold change in transcript abundance between samples (M) versus the log of the average transcript abundance (A), giving insight into the expression levels of mis-regulated genes. A volcano plot shows, for each transcript, the statistical significance (usually the log of the adjusted *P*-value or FDR) versus the log fold change in transcript abundance between samples, giving insight into the statistical significance of mis-regulated genes.

GENE SETS

Gene sets commonly classify gene products (transcripts) as either 'enriched/depleted' in, 'expressed' in, or 'specific' to a particular tissue, cell type, sex, stage, or other variable. Here, we describe methods to define and assign meaning to gene sets in the context of a tissue of interest. We recommend defining gene sets using statistical criteria (e.g. *P*-values from DE analysis), but note that gene sets defined by non-statistical criteria can also yield biological insights. For cases in which non-statistical criteria are chosen, we urge researchers to consider and test the impact of those choices on the results of downstream analyses (e.g. gene set enrichment analysis).

Defining 'enriched/depleted' gene sets

DE analysis is often used to classify genes as either tissue enriched or tissue depleted by identifying genes that are expressed at higher or lower levels in one sample relative to another (Figure 2-2D,E). For example, a muscle-enriched gene can be defined as one that encodes a transcript found at a statistically significantly higher abundance level in a sample of muscle tissue relative to a sample of all tissues. DE analysis is the best practice to classify tissue-enriched genes, because the classification is based on statistical criteria (e.g. genes with an FDR below a specific threshold) and avoids setting arbitrary thresholds of transcript abundance (see below). As transcripts can be expressed but not enriched in a tissue, DE analysis alone usually does not define a complete tissue-expressed gene set.

Defining 'expressed' gene sets

To classify a gene as tissue expressed, researchers typically set an abundance threshold in an RNA profile. Genes that produce higher RNA levels than the threshold are called 'expressed', whereas those that produce lower RNA levels are called 'not expressed'. For this strategy, the choice of threshold is usually arbitrarily set somewhere within the 1-5 TPM or RPKM/FPKM range. One way to choose a TPM threshold is by analyzing a histogram of log-transformed TPM values in a sample. Most histograms show a bimodal distribution (Figure 2-3A); the mode at high TPM values reflects highly expressed genes, whereas the mode at low TPM values reflects lowly expressed genes or noise. A reasonable TPM threshold can be set to a value between these two modes. Because transcript profiles and goals differ

between experiments, there is no gold-standard abundance threshold. Therefore, an appropriate threshold must be chosen and validated for each experiment, realizing that there is a trade-off between sensitivity (inclusion of genes) and specificity (inclusion of only genes that are truly expressed). Previously defined gene sets can help guide selection of a threshold (Figure 2-3B). It is therefore useful to compare different thresholds: a higher-TPM threshold to generate a 'stringent' gene set and a lower-TPM threshold to generate a 'relaxed' gene set (Figure 2-3A). A stringent threshold provides high specificity by omitting truly non-expressed genes, but at the cost of incorrectly excluding true lowly expressed genes (false negatives) (Figure 2-3B). A relaxed threshold provides high sensitivity to detect low-abundance transcripts, but at the cost of incorrectly classifying some non-expressed genes as expressed (false positives). Because defining a gene set this way uses arbitrary criteria, we recommend evaluating its accuracy using existing gene sets and/or other types of gene expression data, such as images of fluorescent reporters or fluorescence *in situ* hybridization.

Another method is to define a 'background' level of TPM by measuring TPM levels in intergenic regions where no transcription occurs (Ramsköld et al. 2009; https://github.com/BgeeDB/BgeeCall). Genes that produce transcripts at a level above this background TPM value are considered to be expressed. Background TPM values can be used to calculate the FDR and a false negative rate (FNR) for the number of genes detected as expressed at different TPM thresholds. An appropriate

TPM threshold can be set to the one that best balances high specificity (low FDR) and high sensitivity (low FNR) in detecting expressed genes.

Defining 'specific' gene sets

A tissue-specific gene set defines genes that are exclusively expressed in one tissue compared with every other tissue. As such, they are highly desired gene sets by researchers who want to understand key genes that drive the development or function of a tissue type. To define a tissue-specific gene set, we recommend identifying transcripts that are tissue enriched by DE analysis and then identifying which of those transcripts are detected only in that tissue (e.g. by strict TPM thresholds). It is important to realize that a gene may be inappropriately classified as tissue specific if the transcript profiling data used for classification failed to detect the gene's transcripts in every other tissue. Some reasons for this might be low sequencing depth, low transcript abundance, stage-specific expression, and few cells in a tissue that express the gene (e.g. rare neuron subtypes).

Evaluating and comparing gene sets

Researchers may want to evaluate the quality of their gene sets by comparing them with already-published gene sets. Commonly, two or more tissue-expressed gene sets that define the same tissue's transcriptome are compared to identify similarities and differences (Figure 2-3B). Such tissue-expressed gene sets are expected to share a majority of genes. However, some differences in membership are

expected due to differences in sample preparation, RNA-profiling technology, sequencing depth, and/or bioinformatic methods used to define gene sets. To reduce sources of variation, it is best to compare gene sets that have been processed and defined using identical methods (if possible); often, this requires one to re-process published transcriptome data. We recommend that researchers re-evaluate their gene sets as new transcriptome data become available.

Gene set enrichment analysis

After defining a gene set of interest, a common goal is to assign biological meaning to the set using gene set enrichment analysis. This type of analysis tests whether a gene set is enriched for genes that share a common biological feature (e.g. protein function, co-regulation, signaling pathway or epigenetic environment). Therefore, it is important to use accurate and high-confidence sets of genes that define those features. Large efforts by several groups, such as the Gene Ontology Consortium, maintain and update publicly available databases of gene sets classified by biological feature (Ashburner et al. 2000; The Gene Ontology Consortium, 2019). Popular gene set enrichment analysis tools, such as GSEA, DAVID and g:Profiler, use these large databases and various statistical methods to comprehensively test many biological features for enrichment (Mootha et al. 2003; Subramanian et al. 2005; Huang et al. 2009; Raudvere et al. 2019; for a review and tutorial on best practices and various approaches to gene set enrichment analysis, see Reimand et al. 2019 and Maleki et al. 2020). Gene set enrichment analysis is an excellent method

for researchers to gain insights into biology from their RNA-seq experiments and generate new and exciting hypotheses.

USEFUL DATA SETS AND RESOURCES

Some C. elegans transcriptome data

Table 2-1 presents some *C. elegans* transcriptome data and gene sets that exemplify diverse methodologies used by the community. For each study, we describe the biological samples used for transcript profiling, how they were isolated, how transcripts were profiled, and which gene sets were defined. Our goal in this Primer was to present a broad view of available data sets, not to judge or evaluate them. We recommend that researchers conduct their own evaluations of data sets they intend to use for a research goal (see the 'Footnotes' in Table 2-1 for issues to consider). An extended set of references is provided in Table 2-3.

Some widely used resources

Table 2-2 provides links to some valuable resources that are heavily used by researchers studying human tissues and popular model organisms. WormBase, FlyBase, HumanBase and other organism-focused resource hubs collect and curate transcriptome data and provide useful tools for visualization and analysis of those data. The ENCODE and modENCODE projects produced and processed a massive amount of transcriptome data from human, mouse, *C. elegans*, and *Drosophila* (Celniker et al. 2009; Hillier et al. 2009; Gerstein et al. 2010, 2014). The Expression

Atlas provides transcriptome data from more than 40 species. NEXTDB is a collection of *C. elegans in situ* hybridization data. A relatively new effort, the Chan Zuckerberg Cell Atlas Initiative (https://www.czbiohub.org/projects/cell-atlas), will include transcriptome data in its ambitious goal to map every cell type in the human body.

CONCLUSIONS

Transcriptomic approaches are now a powerful component of biologists' standard toolbox. A current exciting direction is defining the RNA population in individual cells in an organism, which is providing unprecedented insights into cells' specialized gene expression patterns and functions. It is crucial that researchers new to transcriptomics learn its basics and best practices. For example, gene sets are commonly used resources, but if not defined accurately, they can lead to misinterpretation of data. Therefore, we encourage researchers to re-evaluate the reliability of gene sets as new transcriptome data and analyses become available. Our aim in this Primer was to provide a 'launch pad' into transcriptomics by introducing some commonly used methods and important considerations for designing transcriptomics experiments and analyzing the data. However, research goals should shape the designs and analyses. Researchers should carefully evaluate different approaches to identify which are best for their specific research goals. We hope that this Primer provides a strong foundation for scientists seeking to embrace the power of transcriptomics.

ACKNOWLEDGEMENTS
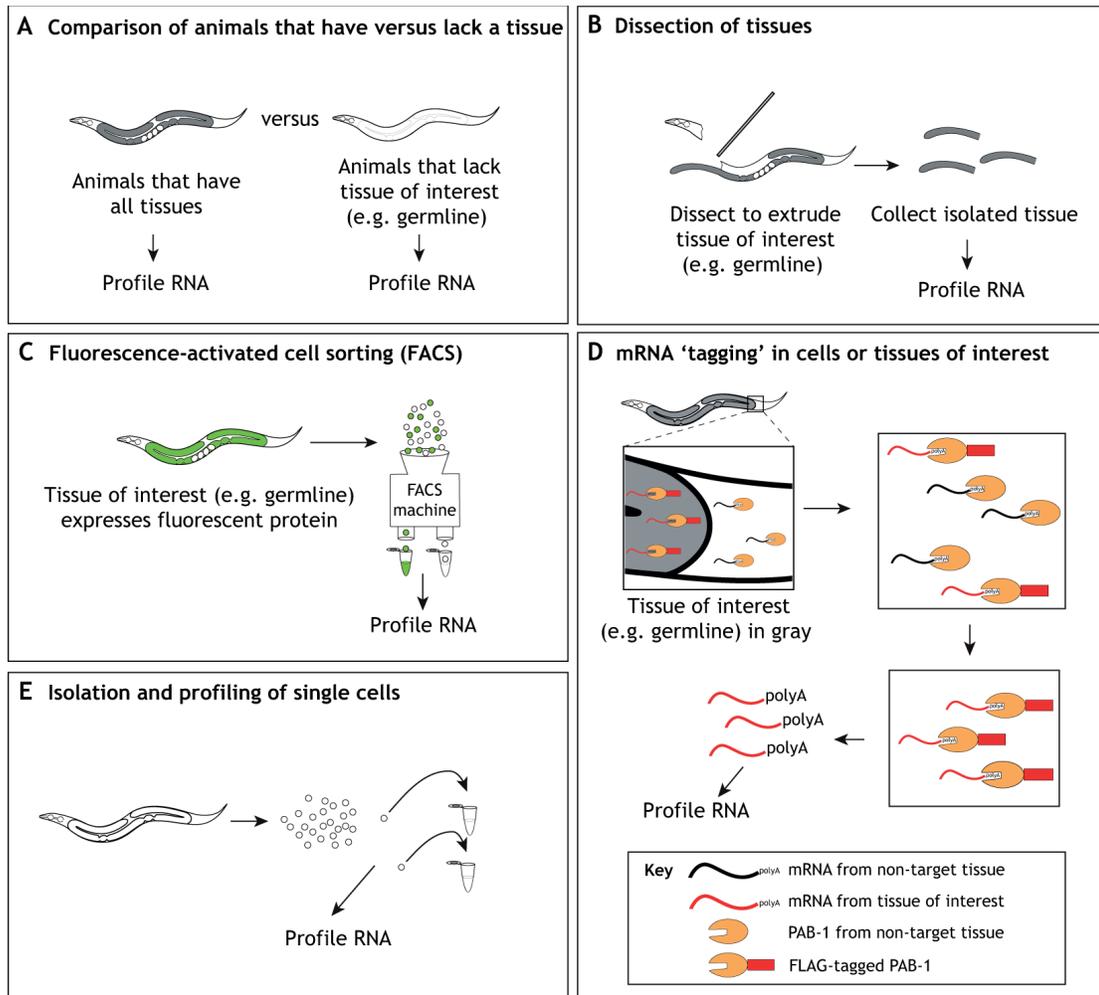
Figure 2-1: Strategies for isolating tissues and cells for transcript profiling. (A-E) Strategies for isolating tissues and cells for transcript profiling. Schematics representing different approaches for isolating samples for RNA profiling. Advantages and disadvantages of each method are described in the main text. See Table 2-1 for a reference to an example of each.
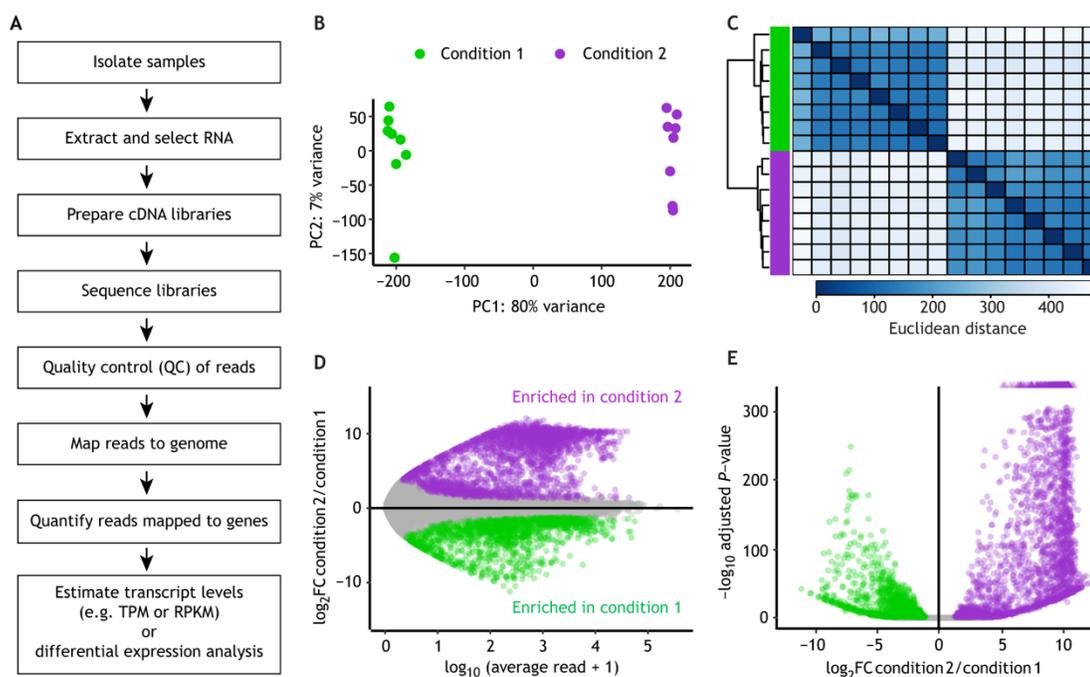
Figure 2-2. Flow chart of steps to prepare libraries for RNA-seq and process and visualize sequencing data. (A) Flow chart of steps. (B-E) Common ways to visualize RNA-seq data and differential expression (DE) analysis, using data from 16 libraries across two sample conditions (Ortiz et al. 2014) and the R package DESeq2 for DE analysis. Green represents condition 1, and purple represents condition 2. (B,C) Visualization of sample clustering using log-transformed counts. (B) Principal component analysis (PCA) showing distances between samples (colored dots) along the first two principal components, which together capture most of the variance between samples. (C) Heatmap showing hierarchical clustering of samples by Euclidean distance. The order of samples across columns is identical to the order down rows. Each tile represents one sample-sample comparison. The dendrogram shows the hierarchical clustering. Darker blue tiles indicate a smaller distance (larger similarity) between samples than lighter blue tiles. In both plots, samples cluster by condition, suggesting that the condition variable explains most of the variation between samples. The clean separation is a good indication of high reproducibility among biological replicates. (D,E) Visualization of DE data. Colored circles and triangles are differentially expressed protein-coding genes, defined as having at least a 2-fold difference in transcript abundance between conditions and having a P-value (adjusted for multiple hypothesis testing) ≤0.05. Differentially expressed genes with negative fold changes define an 'enriched in condition 1' gene set, and genes with positive fold changes define an 'enriched in condition 2' gene set. (D) MA plot showing a transcript's average abundance (read count) versus its fold change in condition 2 compared with condition 1. (E) Volcano plot showing a transcript's fold change versus its adjusted P-value. Triangles are genes that have a significance value that exceeds the maximum value of the y-axis.

Figure 2-3: Selecting an mRNA abundance threshold to define 'expressed' genes in a transcript profile. (A) Histogram plot showing a distribution of transcript abundance values for each protein-coding gene (20,258) expressed as $\log_{10}$(TPM+a pseudo-count of 1) in a transcript profile made from germlines dissected from adult *C. elegans* (Tabuchi et al. 2018). The maximum value of the *y*-axis was artificially set to a lower value (cutting off many genes with 0 TPM) in order to show the shape of the distribution across all TPM values. The dashed vertical lines represent two choices for minimum TPM thresholds to define 'expressed' genes in the transcript profiles. The blue line indicates a 'relaxed' minimum threshold of $\log_{10}$(TPM+1) ≈ 0.30 (or TPM=1). The red line indicates a 'stringent' minimum threshold of $\log_{10}$(TPM+1) ≈ 1.04 (or TPM=10). (B) Venn diagrams comparing the two gene sets defined above using a relaxed threshold (left, blue circle) or stringent threshold (right, red circle) to sets of previously defined 'germline-expressed' genes (gold circles) and 'germline-silent' genes (gray circles) (Wang et al. 2009; Kolasinska-Zwierz et al. 2009). Intersections indicate genes that are found in both compared gene sets. This analysis demonstrates the trade-off of sensitivity and specificity when choosing thresholds: although the relaxed threshold is more inclusive of lowly expressed genes previously defined as germline expressed, it is also more inclusive of genes previously defined as germline silent. Despite being more inclusive, the relaxed threshold still does not include 267 genes (~6%) previously defined as germline expressed. This may be due to differences in transcript profiling techniques used to generate the sets (RNA-seq versus SAGE) and/or the TPM threshold being too stringent.
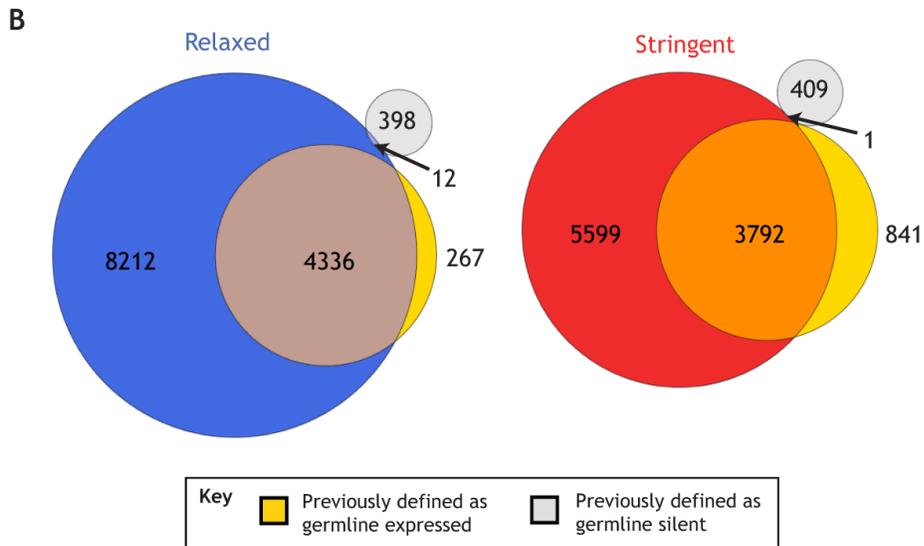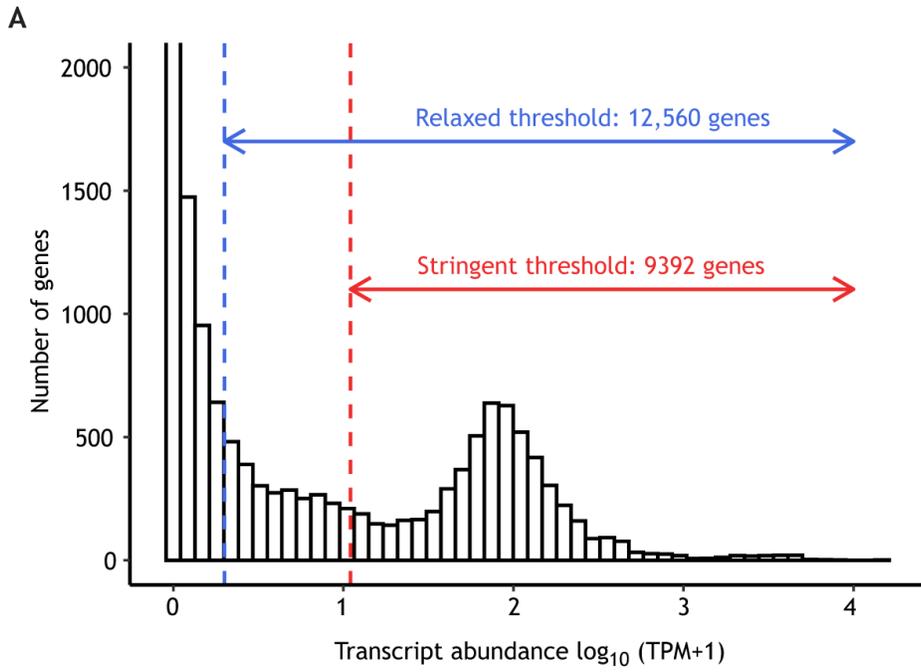
Published paper	Sample preparation	RNA preparation	Platform	Gene sets (number of genes)
Reinke et al., 2004*,¶,**,‡‡	Young adult hermaphrodites with and without a germline (wild type and *glp-4* mutants). Adult males with and without a germline.	polyA(+)	Microarray	Germline enriched in hermaphrodites (3144) Germline enriched in males (1092) Hermaphrodite biased (1935) Male biased (1260) Hermaphrodite adult soma enriched (460) Male adult soma enriched (430)
Ortiz et al., 2014**,‡‡	Oogenic and spermatogenic gonads dissected from sexually transformed mutant animals.	polyA(+)	RNA-seq	Spermatogenesis enriched (2748) Oogenesis enriched (1732) Gender neutral (6274)
Kaletsky et al., 2018¶,‡‡	FACS-sorted hypodermal cells (P*Y37A1B.5*), intestinal cells (P*ges-1*), muscle cells (P*myo-3*) and neurons (P*unc-119*) from day 1 adult hermaphrodites.	Total RNA	RNA-seq	Tissue (expressed | enriched | unique): Hypodermis (7191 | 584 | 248) Intestine (9604 | 519 | 264) Muscle (7691 | 426 | 190) Neurons (8437 | 867 | 616) Ubiquitously expressed (expressed in all tissues) (5360)
Blazie et al., 2015¶,‡‡	mRNA IPed from transgenic mixed-stage worms expressing 3xFLAG-tagged PAB-1 in intestine (P*ges-1*), pharynx (P*myo-2*) and body muscle (P*myo-3*).	mRNA IPed with tagged PAB-1	RNA-seq	Intestine expressed (7355) Pharyngeal muscle expressed (3094) Body muscle expressed (2604) Intestine specific (4091) Pharyngeal muscle specific (310) Body muscle specific (329)
Hashimshony et al., 2012‡,§	Single cells (AB, ABa, ABp, P1, EMS, P2, C, P3) isolated from dissected early embryos.	polyA(+)	RNA-seq	Differentially expressed between sister blastomeres Newly transcribed in blastomeres
Cao et al., 2017§	Applied single-cell RNA-seq combinatorial indexing of cells or nuclei (sci-RNA-seq) to obtain single cell RNA-seq data from all cells of L2 worms.	polyA(+)	RNA-seq	Expression levels for all genes in seven tissues and 27 cell types in L2 larvae Expression from ten neuronal cell types divided further into 40 fine-grained neuronal cell type clusters Enriched in a tissue (3925) Enriched in a cell type (1939)

*Microarrays have a limited dynamic range, assay only annotated genes and may ignore different isoforms.

‡Only assays 3′ ends of polyA(+) transcripts and ignores splicing.

§Noise in single-cell/low-input profiling.

¶Possible mixture of stages/tissues or cell populations after FACS sorting, pull-down, or sample preparation.

**Subset of genes in a particular category were verified by an independent method.

‡‡Report significant overlap with previously defined gene set(s).

IPed, immunoprecipitated; P*gene*, promoter of gene used to drive expression of fluorescent protein or other transgene.

Table 2-1. Examples of different sample preparation, transcript profiling, and gene sets in *C. elegans*.

Type of resource	Resource name	Species	Website
Resource hub that curates gene expression data and offers tools for analysis	WormBase	*Caenorhabditis elegans*	https://wormbase.org
	FlyBase	*Drosophila melanogaster*	https://flybase.org
	ZFIN (Zebrafish Information Network)	*Danio rerio*	https://zfin.org
	SGD (Saccharomyces Genome Database)	*Saccharomyces cerevisiae*	https://www.yeastgenome.org
	TAIR (The Arabidopsis Information Resource)	*Arabidopsis thaliana*	https://www.arabidopsis.org
	MGI (Mouse Genome Informatics)	*Mus musculus*	http://www.informatics.jax.org
	RGD (Rat Genome Database)	*Rattus norvegicus*	https://rgd.mcw.edu
	HumanBase	*Homo sapiens*	https://hb.flatironinstitute.org
Large-scale generation and analysis of gene expression data	modENCODE	*Caenorhabditis elegans*, *Drosophila melanogaster*	http://www.modencode.org
	ENCODE	*Mus musculus*, *Homo sapiens*	https://www.encodeproject.org
	FlyAtlas 2	*Drosophila melanogaster*	http://flyatlas.gla.ac.uk/ FlyAtlas2/index.html
	The Genome BC Gene Expression Consortium	*Caenorhabditis elegans*	http://elegans.bcgsc.ca/home/ ge_consortium.html
	NEXTDB (The Nematode Expression Pattern Database)	*Caenorhabditis elegans*	https://nematode.nig.ac.jp
Large collection and analysis of gene expression data	Expression Atlas	40+ species	https://www.ebi.ac.uk/gxa/home
	GExplore	*Caenorhabditis elegans*	http://genome.sfu.ca/gexplore/

Table 2-2. Some resources for transcriptome data and analyses for widely studied organisms.

CHAPTER 3

The MES chromatin regulators protect germline immortality by repressing the X chromosome via regulation of the transcription factor LIN-15B

ABSTRACT

Maternally synthesized products play critical roles in development of offspring. A premier example is the *C. elegans* H3K36 methyltransferase MES-4, which is essential for germline survival and development in offspring. How maternal MES-4 protects germline immortality is not well-understood, but its role in H3K36 methylation hinted that it may regulate gene expression in Primordial Germ Cells (PGCs). We tested this hypothesis by profiling transcripts from single pairs of PGCs dissected from wild-type and *mes-4* M-Z- (no Maternal or Zygotic MES-4) newly hatched larvae. We found that *mes-4* M-Z- PGCs display normal turn-on of most germline genes and normal repression of somatic genes, but dramatically up-regulate hundreds of genes on the X chromosome. Analysis of *mes-4* M-Z- mutants that inherited different endowments of X chromosome(s) from gametes demonstrated that X mis-expression is the cause of germline death. Intriguingly, removal of the THAP transcription factor LIN-15B from *mes-4* mutants reduced X mis-expression and prevented germline death. *lin-15B* is X-linked and mis-expressed in *mes-4* M-Z- newborn germlines, identifying it as a critical target for MES-4 repression. The above findings extend to MES-2/3/6, the *C. elegans* version of Polycomb Repressive Complex 2. We propose that maternal MES-4 and PRC2 cooperate to protect

55

germline survival by preventing germline-toxic expression of genes on the X chromosome, including the gene that encodes the key transcription factor LIN-15B.

INTRODUCTION

Many critical events during early development are orchestrated by maternally synthesized gene products. Mutations in genes that encode such products in the mother can cause 'maternal-effect' phenotypes in offspring. These phenotypes are usually severe developmental defects. Maternal-effect lethal genes, which cause maternal-effect lethality in offspring, encode products that guide crucial events in early embryo development, such as pattern formation and embryonic genome activation (e.g. the PAR proteins in *C. elegans*, BICOID in *Drosophila*, and Mater in mouse) (Frohnhofer and Nusslein-Volhard, 1987; Tong et al. 2000; Goldstein and Macara, 2007). Maternal-effect sterile genes encode products needed for fertility of the offspring. A few genes in this category encode proteins in germ granules (e.g. PGL-1 in *C. elegans* and VASA in Drosophila) (Rongo and Lehnmann, 1996; Kawasaki et al. 1998). Another fascinating set of genes in this category encode chromatin regulators, which are the focus of this paper.

The *C. elegans* MES proteins were identified in genetic screens for maternal-effect sterile mutants, hence their name (MES for Maternal-Effect Sterile) (Capowski et al. 1991). MES-2, MES-3, and MES-6 assemble into a trimeric complex that is the *C. elegans* version of Polycomb Repressive Complex 2 (PRC2) (Xu et al. 2001; Bender et al. 2004). PRC2 is a histone methyltransferase (HMT) that methylates Lys

27 on histone H3 (H3K27me) to repress genes that are packaged by those methylated nucleosomes (Margueron and Reinberg, 2011). MES-4 is an HMT that methylates Lys 36 on H3 (H3K36me), which marks actively transcribing genes and has context-dependent roles in transcriptional regulation (Bender et al. 2006; Rechtsteiner et al. 2010; Furuhashi et al. 2010; Kreher et al. 2018). Although PRC2 and MES-4 catalyze opposing flavors of histone marking, the loss of either causes nearly identical mutant phenotypes (Capowski et al. 1991). Worms that inherit a maternal load of gene product but cannot synthesize zygotic product (referred to as *mes* M+Z- mutants) are fertile. Worms that do not inherit a maternal load or produce zygotic gene product (*mes* M-Z- mutants) are sterile due to death of nascent germ cells in early- to mid-stage larvae. In *mes* M-Z+ mutants, zygotically synthesized product does not rescue fertility, highlighting the critical importance of maternal product. PRC2's roles in transcriptional repression and development have been intensively studied and are well defined across species, including roles in *C. elegans* germline development (Bender et al. 2004, Gaydos et al. 2012; Gaydos et al. 2014, Patel et al. 2012, Kaneshiro et al. 2019, Delaney et al. 2019). In contrast, how MES-4 ensures the survival of nascent germ cells is unknown and particularly puzzling.

One possibility for MES-4 function is that maternal MES-4 promotes expression of genes required for germline development. Support for this comes from analyses of mutants that ectopically express germline genes in their soma (e.g. *lin-15B, lin-35,* and *spr-5; met-2* mutants), and as a result, have developmental defects (Unhavaithya et al. 2002; Wang et al. 2005; Cui et al. 2006; Curran et al. 2009;

Petrella et al. 2011; Wu et al. 2012; Goetsch et al. 2017; Carpenter et al. 2021).

Concomitant loss of MES-4 from these mutants prevents ectopic expression of

germline genes and restores worm health. In wild-type early embryos, MES-4 and

methylated H3K36 associate with genes that were expressed in the maternal germline,

regardless of whether they are transcribed in embryos (Furuhashi et al. 2010;

Rechtsteiner et al. 2010). Focusing on H3K36me3, genetic tests showed that MES-4

strictly maintains pre-existing patterns of H3K36me3 and is unable to catalyze *de*

novo H3K36me3 marking of genes (Furuhashi et al. 2010); the other H3K36 HMT in

C. elegans, MET-1, like H3K36 HMTs in other systems, catalyzes *de novo*

H3K36me3 on genes in response to transcriptional turn-on (Kizer et al. 2005;

Furuhashi et al. 2010; Kreher et al. 2018). Taken together, these findings suggested

the appealing model that in embryos maternal MES-4 maintains H3K36me3 marking

of germline-expressed genes and in that way transmits an epigenetic "memory of

germline", a developmental blueprint, to the primordial germ cells (PGCs) of

offspring.

Two findings challenge the model that MES-4 somehow promotes expression

of germline genes. First, among *mes-4* M-Z- mutants, hermaphrodites (with 2 X

chromosomes) are always sterile, while males (with 1 X chromosome) can be fertile

(Garvin et al. 1998). This suggested that the dosage of X-linked genes matters for the

Mes-4 mutant phenotype. Second, profiling transcripts in the gonads of fertile *mes-4*

M+Z- mutant hermaphrodites revealed that the most dramatic effect of losing zygotic

MES-4 is up-regulation of genes on the X (Bender et al. 2006; Gaydos et al. 2012).

Notably, the X chromosomes are normally kept globally repressed during all stages of germline development in males and during most stages of germline development in hermaphrodites (Kelly et al. 2002; Reinke et al. 2004; Wang et al. 2009, Arico et al. 2011; Tzur et al. 2018), and likely as a consequence, most germline-expressed genes are located on the autosomes. These findings focused attention on the X chromosome and raised the question – what role does maternal MES-4 serve to ensure that PGCs survive and develop into a full and healthy germline?

To investigate the role of MES-4 in PGCs, the cells that critically rely on maternal MES-4 for survival, and to formally test the model that MES-4 promotes expression of germline genes, we performed transcript profiling in dissected single pairs of PGCs from wild-type versus *mes-4* M-Z- mutant larvae. We asked if absence of maternal MES-4 causes PGCs to 1) fail to turn on germline genes, 2) inappropriately turn on somatic genes, and/or 3) inappropriately turn on X-linked genes. We found that in *mes-4* PGCs most germline genes were turned on normally, thus disproving the model that the major role of MES-4 is to promote expression of germline genes in PGCs. Most somatic genes were kept off, arguing that MES-4 does not protect the germline by opposing somatic development. The most dramatic impact to the transcriptome in *mes-4* PGCs was up-regulation of hundreds of X-linked genes, many of which are part of an oogenesis program. Our genetic analysis of *mes-4* mutants with different X-chromosome endowments from the oocyte and sperm demonstrated that up-regulation of X-linked genes is the cause of death of nascent germ cells in *mes-4* M-Z- mutant larvae. A surprising finding was that maternal MES-

4 promotes germline health independently of transmitting H3K36me3 marking of parental chromosomes. We identified the transcription factor LIN-15B, an X-linked gene up-regulated in *mes-4* mutants, as a major cause of X mis-expression and germline death in *mes-4* mutants. Performing similar tests of PRC2 (*mes-3*) M-Z-mutant larvae revealed that their PGCs up-regulate many X-linked genes in common with *mes-4* PGCs, and that removal of LIN-15B restores the health of their germline, as it does for *mes-4* mutants. This study revealed that maternal MES-4 and PRC2 cooperate to ensure germline survival and health in offspring by preventing mis-expression of genes on the X chromosome, and that both operate through a key X-encoded transcription factor.

RESULTS

MES-4 is not required for PGCs to launch a germline program

MES-4 propagates an epigenetic 'memory' of a germline gene expression program during embryogenesis by maintaining H3K36me3 on genes that were previously expressed in parental germlines (Rechtsteiner et al. 2010; Furuhashi et al. 2010; Kreher et al. 2018). Our model predicts that delivery of this memory to offspring PGCs instructs them to launch a gene expression program that promotes germline proliferation and development. To test this model, we performed RNA-sequencing to determine whether PGCs from *mes-4* M-Z- (Maternal MES-4 minus and Zygotic MES-4 minus) mutant larvae, which completely lack MES-4, fail to turn

on a germline program (Figure 3-1A). We developed a hand-dissection strategy that enables us to isolate in <30 minutes single sets of sister PGCs (2 cells), marked by a specifically and highly expressed germline marker (GLH-1::GFP), from wild-type or *mes-4* M-Z- mutant larvae for RNA-seq library preparation. We performed differential expression analysis to identify genes that are significantly down-regulated (DOWN) or up-regulated (UP) in *mes-4* PGCs compared to wild-type (wt) PGCs. Our analysis identified 176 DOWN genes and 450 UP genes (Figure 3-1B-D).

To determine whether the DOWN genes include germline genes that fail to turn on normally in *mes-4* mutant PGCs, we analyzed transcript levels and fold changes (*mes-4* vs. wt) for genes that are members of 2 'germline' gene sets: 1) a 'germline-specific' set containing 168 genes that are expressed in germline tissue and not expressed in somatic tissues, and 2) a 'germline-enriched' set containing 2176 genes that are expressed at higher levels in adults with a germline compared to adults that lack a germline. We found that the majority of germline-specific genes (162 of 168 genes or 96%) and germline-enriched genes (2111 of 2176 genes or 97%) are not significantly DOWN in *mes-4* PGCs (Figure 3-1B). Some germline-specific genes and germline-enriched genes are DOWN, but those numbers are not more than expected by chance (Figure 3-1E). Since some gene expression defects may not manifest until after *mes-4* PGCs have started dividing, we used our hand-dissection strategy to isolate sets of 2 PGC-descendants, which we call 'early' germ cells (EGCs), from wt and *mes-4* mutant L2 larvae and profiled their transcripts. We found that *mes-4* EGCs down-regulate more germline-specific and more germline-enriched

genes than *mes-4* PGCs. However, like *mes-4* PGCs, *mes-4* EGCs turn on most germline-specific and germline-enriched genes normally (Figure 3-6).

As an independent test of differential expression in PGCs, we selected 3 genes and performed smFISH to measure and compare their transcript levels between *mes-4* and wt PGCs. 2 of the genes we tested, *cpg-2* and *pgl-3,* are members of our germline-specific set and were determined by our RNA-seq analysis to be DOWN or not DOWN respectively, in *mes-4* PGCs (Figure 3-2A, Figure 3-8). Corroborating our RNA-seq analysis, smFISH analysis showed that the average transcript abundance of *cpg-2* is significantly lower in *mes-4* vs wt PGCs, while the average transcript abundance of *pgl-3* is not significantly different (Figure 3-2B-C). The other gene we tested by smFISH, *chs-1*, is a member of our germline-enriched set and is consistently not mis-expressed in our RNA-seq or smFISH analysis (Figure 3-2C, Figure 3-8). Together, our RNA-seq and smFISH analysis showed that *mes-4* PGCs turn on most germline genes normally.

MES-4 is not required to keep somatic genes off in PGCs

Chromatin regulators can protect tissue-appropriate transcription patterns by serving as a 'barrier' to promiscuous transcription factor activity. Loss of MES-4 and PRC2 have both been shown to allow mis-expression of neuronal-target genes upon ectopic expression of the transcription factor CHE-1 in the germline (Patel et al. 2012, Seelk et al. 2016). Moreover, loss of PRC2 activity in the *C. elegans* germline was recently linked to mis-expression of some neuronal genes and conversion to

neuronal fate (Kaneshiro et al. 2019). Maternal MES-4 may promote offspring

germline development by preventing their germlines from turning on a somatic gene

expression program. To test this possibility, we examined whether UP genes in *mes-4*

vs wt PGCs are members of a 'soma-specific' gene set that defines 861 genes

expressed in soma but not germline. We found that only 19 UP genes are soma-

specific, which is not a higher number than expected by chance (Figure 3-1C,F).

Therefore, *mes-4* PGCs do not mis-express a soma-specific program.

MES-4 represses genes on the X chromosome including many oogenesis genes in PGCs

Repression of the X chromosomes in the *C. elegans* germline is essential for

germline health (Gaydos et al. 2014). We found that 311 of the 2808 (11%) protein-

coding X genes are UP in *mes-4* vs wt PGCs. Strikingly, more than half of all UP

genes are on the X chromosome (311 out of 450 genes), and this number is

significantly higher than expected by chance (Figure 3-1F). We found that *mes-4*

EGCs mis-express almost all of the X genes that *mes-4* PGCs do and additionally

mis-express another 273 X genes; together, they mis-express 584 X genes (Figure 3-

6). These data show that *mes-4* PGCs mis-express many genes on the X chromosome

and that X mis-expression becomes more severe in their descendant EGCs.

As an independent test of differential expression, we selected 4 X-linked UP

genes and performed smFISH to compare their transcript levels between *mes-4* vs wt

PGCs (Figure 3-8). smFISH analysis showed that all 4 X genes have higher transcript

abundance in *mes-4* PGCs than in wt PGCs (Figure 3-2D-F), corroborating our transcriptome analysis. These data reveal that the X chromosome is the primary focus of MES-4 regulation in PGCs.

While most X-linked genes are repressed during germline proliferation and spermatogenesis, some are normally turned on during oogenesis (Kelly et al. 2002; Arico et al. 2011; Tzur et al. 2018, Figure 3-7). We examined whether X-linked UP genes are those that are normally turned on during oogenesis by comparing our set of X-linked UP genes to a set of 'oogenesis' genes, defined as 470 X-linked and 1201 autosome-linked genes that are expressed at higher levels in dissected adult oogenic germlines than in dissected spermatogenic germlines (Ortiz et al. 2014). We found that 209 of the 311 (67%) X-linked UP genes and 25 of the 149 (17%) autosomal UP genes are in the oogenesis set, which are both higher numbers than expected by chance (Figure 3-1F). The enrichment for oogenesis X-linked genes was especially high. Strikingly, almost all oogenesis UP genes in *mes-4* PGCs (209 of 234, or 89%) are X-linked. No other germline gene set that we tested (germline-specific, germline-enriched, and spermatogenesis) was enriched in the set of UP genes in *mes-4* PGCs (Figure 3-1F). Our gene ontology (GO) analysis found that the set of X-linked UP genes in *mes-4* PGCs or EGCs are enriched for biological process terms that characterize roles of oogenesis: 'reproduction' and 'embryo development ending in birth or hatching' (Figure 3-7). We conclude that *mes-4* PGCs mis-express an oogenesis program involving many X-linked genes, which may interfere with mutant PGCs' ability to proliferate.

Mis-expression of genes on the X chromosome(s) causes germline death in *mes-4* mutants

Since X mis-expression is the largest defect to the transcriptome in *mes-4* PGCs, we hypothesized that mis-expression of the 2 X chromosomes in germlines of *mes-4* mutant hermaphrodites causes germline death. To test our hypothesis, we asked whether *mes-4* mutant males, which inherit only a single X chromosome (typically from the oocyte), have healthy germlines. We live imaged wild-type and *mes-4* mutant hermaphrodites and males that express a germline-specific GFP reporter (GLH-1::GFP), and scored their germline health qualitatively as either 'absent/tiny' germline, 'partial' germline, or 'full' germline. All live-imaged *mes-4* mutant hermaphrodites lacked a germline (Figure 3-3A). In contrast, some *mes-4* mutant males that inherited their single X from an oocyte (X^{oo} males) had either partial or full germlines (21% and 4%, respectively). Since X chromosomes turn on during oogenesis (Kelly et al. 2002; Arico et al. 2011; Tzur et al. 2018), X^{oo} males inherited an X with a history of expression. Using a *him-8* mutant, we generated wild-type and *mes-4* mutant males that instead inherited their X from a sperm (X^{sp} males), which has a history of repression because it was not turned on previously during spermatogenesis. We tested whether *him-8; mes-4* mutant X^{sp} males that inherited a single X with a history of repression have healthier germlines than *mes-4* mutant X^{oo} males that inherited a single X with a history of expression. Strikingly, most *him-8; mes-4* mutant X^{sp} males made full germlines (67%), compared to only 4% *mes-4* mutant X^{oo} males.

A new and powerful genetic tool uses *gpr-1* over-expression to generate hermaphrodite worms that form a germline entirely composed of 2 genomes inherited from the sperm, or rarely 2 genomes inherited from the oocyte (Bessling and Bringmann 2016; Artilles et al. 2019) (Figure 3-9). Using this tool, we tested whether *mes-4* mutant X^{sp}/X^{sp} hermaphrodites that inherited 2 X chromosomes with a history of repression from a sperm have larger germlines compared to *mes-4* mutant X^{oo}/X^{oo} hermaphrodites that inherited 2 X chromosomes with a history of expression from an oocyte. While all *mes-4* mutant X^{oo}/X^{oo} hermaphrodites lacked a germline, some *mes-4* mutant X^{sp}/X^{sp} hermaphrodites had partial or full germlines (32% and 18%, respectively) (Figure 3-3A). Our combined genetic analysis demonstrates that mis-expression of the X chromosome(s) causes germline death in *mes-4* M-Z- mutants.

MES-4 promotes germline health independently from its role in transmitting H3K36me3 patterns across generations

Transmission of epigenetic information across generations can impact the health of offspring. We hypothesized that maternally loaded MES-4's role in transmitting H3K36me3 patterns from parents to offspring is essential for offspring germline development. If so, then an offspring's failure to inherit parental patterns of H3K36me3 should cause their germline to die even if they receive maternal MES-4. To test our hypothesis, we used the *gpr-1* genetic tool and the GLH-1::GFP germline marker to generate F1 adult offspring whose PGCs inherited 2 H3K36me3(-) genomes from the sperm and either did or did not inherit maternal MES-4.

Importantly, the germline in both types of F1 offspring had the same genotype (*met-1; mes-4*); the only difference between the F1s was the presence or absence of maternally loaded MES-4. We found that over half (57%) of F1 adult offspring had a full germline if their PGCs inherited 2 H3K36me3(-) genomes and maternal MES-4 (Figure 3-3B). In contrast, 0% of F1 adult offspring had a full germline if their PGCs inherited 2 H3K36me3(-) genomes from the sperm and did not inherit maternal MES-4 (Figure 3-3B). This result shows that maternally loaded MES-4 is critical for offspring germline development but that its critical role is not to transmit H3K36me3 patterns from parents to offspring.

Presence of maternal MES-4 allows many F1 offspring whose PGCs inherited 2 H3K36me3(-) genomes from the sperm to make a full germline. One possibility is that maternally loaded H3K36 HMTs can re-establish sufficient levels of H3K36me3 marking to H3K36me3(-) chromosomes in PGCs for germline development. Since MES-4 cannot catalyze de novo H3K36me3 marking on H3K36me3(-) chromosomes (Furuhashi et al. 2010, Kreher et al. 2018), re-establishment of H3K36me3 levels would require the other H3K36 HMT MET-1, which can catalyze de novo marking in response to transcriptional turn-on, like H3K36 HMTs in other species. We found that removal of maternal MET-1, like removal of maternal MES-4, caused almost all F1 offspring whose PGCs inherited 2 H3K36me3(-) genomes from sperm to lack a germline (Figure 3-3B). These findings show that maternal loads of both H3K36 HMTs are required for F1 offspring whose PGCs inherited 2 H3K36me3(-) genomes from sperm to make a germline, and they support the possibility that newly

established H3K36me3 marking of H3K36me3(-) chromosomes by maternally loaded HMTs can allow PGCs to make a germline.

LIN-15B causes X mis-expression in germlines of *mes-4* M+Z- adults

MES-4 levels are low on the X chromosome(s) during most stages of germline development (Bender et al. 2006; Rechtsteiner et al. 2010; Furuhashi et al. 2010; Gaydos et al. 2012; Kreher et al. 2018). We therefore hypothesized that MES-4 represses X genes indirectly by regulating the expression or activity of 1 or more downstream factor(s). For example, MES-4's activity on autosomes may repress X genes by concentrating a transcriptional repressor onto the X or by sequestering a transcriptional activator away from the X (Gaydos et al. 2012; Cabianca et al. 2019; Georgescu et al. 2020). Several lines of evidence make the THAP transcription factor LIN-15B a strong candidate for causing X mis-expression in germlines that lack MES-4. First, our analysis of publicly available LIN-15B ChIP data from whole embryos and larvae found that LIN-15B targets the promoter of many X genes that are repressed by MES-4 in PGCs and/or EGCs (Figure 3-12). Second, *lin-15B* is X-linked and UP in our smFISH analysis of *mes-4* vs wt PGCs (Figure 3-2E and F and Figure 3-8). Third, LIN-15B has been reported to promote expression of X-linked genes in PGCs and adult germlines (Lee et al. 2017, Robert et al. 2020). Finally, *mes-4* and *lin-15B* genetically interact in somatic cells to control expression of germline genes such as genes encoding P-granule components (Petrella et al. 2011).

We hypothesized that LIN-15B causes mis-expression of X genes in germlines that lack MES-4. To test our hypothesis, we used RNA-seq to determine whether gonads dissected from *mes-4* M+Z-*; lin-15B* M-Z- double mutant adults have reduced levels of X mis-expression compared to gonads dissected from *mes-4* M+Z- single mutant adults. Our differential expression analyses showed that *mes-4* M+Z-*; lin-15B* M-Z- adult gonads up-regulate considerably fewer X genes (112 X genes) than *mes-4* M+Z- adult gonads (367 X genes) (Figure 3-4A and Figure 3-13). Furthermore, the 367 X genes up-regulated in *mes-4* M+Z- adult gonads had closer to normal (wild-type) transcript levels in *mes-4* M+Z-*; lin-15B* M-Z- adult gonads (Figure 3-4B). Additional differential expression analysis revealed that 323 of the 367 X genes (88%) up-regulated in *mes-4* M+Z- vs wt adult gonads had significantly lower transcript levels in *mes-4* M+Z-*; lin-15B* M-Z- (Figure 3-4C). Our results show that LIN-15B is responsible for much of the X mis-expression in *mes-4* M+Z- adult gonads.

LIN-15B causes sterility in *mes-4* M-Z- mutants

Since LIN-15B causes X mis-expression in *mes-4* M+Z- adult gonads, we hypothesized that removal of LIN-15B would allow *mes-4* M-Z- mutants to make healthier germlines. We compared distributions of germline size between *mes-4* M-Z- X^{sp}/X^{sp} mutant hermaphrodite adults and those that lacked either maternal LIN-15B (*lin-15B* M-Z+), zygotically synthesized LIN-15B (*lin-15B* M+Z-*)*, or both (*lin-15B* M-Z-*)*. We found that removal of either maternal LIN-15B, zygotic LIN-15B, or both

caused more *mes-4* M-Z- X^{sp}/X^{sp} mutants to make full-sized germlines (Figure 3-4D).

Notably, loss of maternal LIN-15B caused better recovery of germline health than

loss of zygotic LIN-15B, and loss of both had an additive effect. Strikingly, 88% of

mes-4 M-Z-; *lin-15B* M-Z- X^{sp}/X^{sp} mutant adults made full-sized germlines. We

conclude that both maternal and zygotic sources of the transcription factor LIN-15B

cause germline loss in *mes-4* M-Z- mutants.

Removal of LIN-15B may only allow *mes-4* M-Z- mutant hermaphrodites to

make a full-sized germline if that germline inherited 2 X chromosomes with a history

of repression (from sperm), which by itself improves germline health in *mes-4* M-Z-

mutants (Figure 3-3A, Figure 3-4D). We performed a similar test as above except we

analyzed germline size in X^{oo}/X^{sp} mutants that inherited 1 of their 2 X chromosomes

with a history of expression (from the oocyte). We found that 29% of *mes-4* M-Z-;

lin-15B M-Z- X^{oo}/X^{sp} adult mutant hermaphrodites made full-sized germlines

compared to 0% of *mes-4* M-Z-; *lin-15B* M+Z+ adult mutant hermaphrodites (Figure

3-14). This finding demonstrates that removal of LIN-15B can even allow *mes-4* M-

Z- X^{oo}/X^{sp} mutants to make a full-sized germline.

To investigate whether other factors contribute to sterility in *mes-4* M-Z-

mutants, we identified candidate genes that met 1 or more of 4 criteria: 1) they are X-

linked, 2) they are X-linked and UP in *mes-4* PGCs and/or EGCs, 3) there is evidence

of them binding to the promoter region of at least 25% of X-linked UP genes, and 4)

they target a DNA motif that is enriched in the promoter of X-linked UP genes

(Figure 3-12). We selected 23 top candidates or histone acetyltransferases (HATs),

which are involved in transcriptional activation, and tested whether their depletion by RNAi causes *mes-4* M-Z- mutants to make healthier germlines. RNAi depletion of LIN-15B, but none of the 23 candidates, caused *mes-4* M-Z- mutants to make healthier germlines (Figure 3-12). We conclude that LIN-15B is a major contributor to causing germline death in *mes-4* M-Z- mutants.

MES-4 cooperates with PRC2 to repress X genes

In addition to MES-4, the maternally loaded H3K27 HMT Polycomb Repressive Complex 2 (PRC2), composed of MES-2, MES-3, and MES-6, promotes germline survival and development by repressing genes on the X chromosome (Gaydos et al. 2012; Gaydos et al. 2014). Moreover, as in *mes-4* M-Z- mutants, loss of LIN-15B causes *mes-3* M-Z- mutants to make healthier germlines (Figure 3-14). Together, these findings suggest that MES-4 and PRC2 cooperate to repress X genes and protect germline health. To test this possibility, we compared transcript profiles in wild-type (wt), *mes-3* M-Z-, and *mes-4* M-Z- PGCs and EGCs. In Principal Component Analysis (PCA), the top 2 principal components captured 41% of the variance across all samples and clustered *mes-4* and *mes-3* mutant samples together by germline stage and away from wild-type samples (Figure 3-5A). Using differential expression analysis, we identified 354 X genes UP in *mes-3* vs wt PGCs and 443 X genes UP in *mes-3* vs wt EGCs. We found that stage-matched *mes-3* and *mes-4* samples up-regulate a highly similar set of X genes (Figure 3-5B-C). Next, we compared log$_2$(fold change) (mutant vs wt) of mis-regulated X genes between *mes-4*

and *mes-3* PGCs and between *mes-4* and *mes-3* EGCs. We found a positive, albeit small, correlation (0.22 Spearman's correlation coefficient) between PGCs and a stronger correlation between EGCs (0.44 Spearman's correlation coefficient) (Figure 3-5D-E). We conclude that MES-4 and PRC2 repress the same set of X genes in newborn germlines and therefore cooperate to ensure germline survival and health of newborn germlines in M-Z- mutant larvae.

The chromodomain protein MRG-1 is a candidate reader and effector of H3K36me3 that, like MES-4 and PRC2, promotes germline development by repressing X genes (Fujita et al. 2002; Takasaki et al. 2007). To test if MRG-1 represses the same set of X genes as MES-4 and PRC2, we profiled transcripts from PGCs hand-dissected from *mrg-1* M-Z- L1 larvae. We found that *mrg-1* PGCs mis-express 440 X genes, 225 of which are also mis-expressed in *mes-4* PGCs, *mes-4* EGCs, *mes-3* PGCs, and *mes-3* EGCs (Figure 3-15). These findings add MRG-1 to the team of maternal regulators that ensure PGC survival and health by repressing the X.

DISCUSSION

In this study, we investigated how a maternally supplied chromatin regulator protects germline immortality and promotes germline health. We found that nascent *C. elegans* germlines (PGCs and EGCs) that completely lack maternal MES-4 mis-express over a thousand genes, most of which are on the X chromosome. We further demonstrated that X mis-expression is the cause of germline death in *mes-4* M-Z-

mutants. Removal of a single transcription factor, LIN-15B, reduced X mis-expression in the germline of *mes-4* mutant mothers (*mes-4* M+Z-) and was sufficient to allow the majority of their offspring (*mes-4* M-Z-) to develop full-sized germlines. Intriguingly, *lin-15B* is itself X-linked and mis-expressed in nascent germlines that lack MES-4, highlighting *lin-15B* as a key target for MES-4 repression. We favor a model where maternal MES-4 promotes offspring germline development by preventing LIN-15B from activating a germline-toxic program of gene expression from the X chromosome (Figure 3-5F). This work underscores how maternally supplied factors can guide development of specific tissues in offspring by protecting their transcriptome.

Maternal MES-4 catalyzes methylation of H3K36me2 and H3K36me3 on 20% of the genome in embryos (Rechsteiner et al, 2010), marking thousands of genes that were previously expressed in the maternal germline and that need to be expressed in the offspring germline. Yet surprisingly, lack of MES-4 does not impact the expression of most germline genes in PGCs. Furthermore, our genetic findings show that *mes-4* mutants can develop a full-sized germline if they inherit X chromosomes that have a history of repression, demonstrating that MES-4 is not required for PGCs to launch a germline program. Interestingly, MES-4 is required for the mis-expression of germline genes in somatic tissues of several mutants, such as *spr-5; met-2, lin-15B, lin-35* and mutants of DREAM complex components (Wang et al. 2005; Petrella et al. 2011; Carpenter et al. 2020). This suggests that maternal MES-4 has tissue-dependent roles in gene regulation. Such context-dependent roles may be mediated by different

H3K36me3 'reader' complexes, as observed in other organisms (Yochum and Ayer, 2002; Cai et al. 2003; Chen et al. 2009). Our findings clarify the role of maternal MES-4 in regulating the transcriptome of newborn germlines.

There has been a concerted effort in recent years to determine how epigenetic inheritance impacts offspring health (Heard and Martienssen, 2014; Tabuchi et al. 2018; Kaneshiro et al. 2019). Maternal MES-4's role in propagating gamete-inherited H3K36me3 patterns through embryogenesis is a clear example of epigenetic inheritance (Kreher et al. 2018). By taking advantage of the *gpr-1* genetic tool, we demonstrated that <u>inheritance</u> of H3K36me3 patterns from parents is not required for offspring germline development. However, additional loss of either maternal MES-4 or maternal MET-1 (the two H3K36 HMTs in *C. elegans*) rendered almost all worms sterile, suggesting an important role for H3K36me3 marking during early germline development. We speculate that maternal MES-4 and maternal MET-1 cooperate to restore sufficient levels and proper patterns of H3K36me3 marking to chromosomes inherited lacking H3K36me3 to allow germline development. In this scenario, we envision that maternal MET-1 first catalyzes new H3K36me3 marking on genes co-transcriptionally during the first wave of zygotic genome activation in PGCs (Furuhashi et al. 2010; Kreher et al. 2018), after which maternal MES-4 maintains MET-1-generated patterns of H3K36me3 through early germline development to prevent germline death. The importance of H3K36me3 marking is also highlighted by our finding that loss of the candidate H3K36me3 'reader' MRG-1 (homolog of yeast Eaf3, fly MSL3, and human MRG15) (Gorman et al. 1995; Eisen et al. 2001; Cai et

al. 2003; Bertram and Pereira-Smith, 2001; Joshi and Struhl, 2005) causes PGCs to mis-express the same set of X genes as does loss of MES-4 and also causes death of the nascent germline.

Since MES-4 binding and its HMT activity are very low across almost the entire X chromosome, it is likely that MES-4 regulates expression of X genes indirectly in PGCs. One possible mechanism for indirect regulation is that MES-4 generates H3K36me3 on autosomes to repel and concentrate a transcriptional repressor on the X chromosome(s). An attractive candidate repressor is the H3K27 HMT Polycomb Repressive Complex 2 (PRC2): in germlines, PRC2 activity is concentrated on the X chromosome(s), PRC2 represses the same set of X genes as MES-4 (Gaydos et al. 2012; Lee et al. 2017; this work), and loss of PRC2 causes maternal-effect sterility, like loss of MES-4. An alternative possibility is that H3K36me3 in germlines sequesters a transcriptional activator on autosomes and away from the X chromosome(s) as it does to the histone acetyltransferase (HAT) CBP-1 in *C. elegans* intestinal cells (Cabianca et al. 2019; Georgescu et al. 2020).

We identified the THAP transcription factor LIN-15B as a cause of X mis-expression in fertile germlines that lack MES-4 (*mes-4* M+Z- mutant mothers) and a major driver of germline death in their *mes-4* M-Z- mutant offspring. Interestingly, *lin-15B* is itself an X-linked gene that is mis-expressed in *mes-4* M-Z- mutant PGCs and EGCs. This suggests that although maternal MES-4 regulates expression of thousands of genes in offspring germlines, it only needs to repress *lin-15B* to allow germline survival. However, there are likely additional factors besides LIN-15B that

contribute to X mis-expression and germline death in *mes-4* M-Z- mutants, as removal of LIN-15B does not allow all *mes-4* M-Z- mutants to make germlines. We tested whether depletion of 23 candidate HATs and transcription factors improves germline health in *mes-4* mutants; we found no hits other than the THAP transcription factor *lin-15B*. Recent studies found that upregulation of *lin-15B* also causes sterility in *nanos* mutants and *set-2* (H3K4 HMT) mutants and leads to up-regulation of other X genes (Lee et al. 2017, Robert et al. 2020). Those findings coupled with ours suggest that excessive LIN-15B activity causes germline-toxic levels of X-chromosome expression and that the germline uses multiple protective mechanisms to antagonize LIN-15B.

Why is X mis-expression toxic to the germline? Our transcriptome analysis found that the X genes mis-expressed in *mes-4* M-Z- nascent germlines are highly enriched for genes that are normally turned on during oogenesis. This suggests the intriguing possibility that maternal MES-4 promotes germline development of PGCs by antagonizing an oogenesis program that interferes with the proliferative fate of PGCs. Since the oocyte-inherited X chromosome has a history of expression, it may be prone to turning on in, and thereby causing death of, offspring PGCs that lack MES-4. In support of this, offspring PGCs that lack MES-4 can develop into full-sized germlines if they inherit only X chromosomes that have a history of repression (from the sperm). Our findings support a model where MES-4 prevents activation of the oocyte-inherited X chromosome in PGCs by opposing transcription factors such as LIN-15B.

How LIN-15B causes X mis-expression in germlines that lack MES-4 is

unclear. Several studies have focused on LIN-15B's role as a repressor of germline

genes in somatic tissues (Wang et al. 2005; Petrella et al. 2011; Wu et al. 2012).

Recently, LIN-15B was shown to promote repressive H3K9me2 marking in the

promoter of germline-specific genes in soma (Rechtsteiner et al. 2019). In germlines,

LIN-15B may activate expression of X genes indirectly, e.g. by downregulating or

antagonizing a repressor of X genes. Alternatively, LIN-15B may have context-

dependent roles in gene expression, a well-known feature of many transcription

factors (Fry and Farnham, 1999), and may directly activate expression of X genes. In

support of this model, our analysis of LIN-15B ChIP data from whole embryos and

larvae found that LIN-15B is associated with the promoter of many X genes that are

repressed by MES-4 in newborn germlines. Clarification of LIN-15B's mode of

action in germlines requires analysis of germline-specific chromatin patterns of LIN-

15B binding, biochemical experiments, and identification of LIN-15B's functional

partners.

An outstanding question is what launches the germline program in *C. elegans*

PGCs. MES-4 has been a prime candidate, since it transmits H3K36me marking of

germline genes from parent germ cells to offspring germ cells and so has been

invoked as passing a "memory of germline" across generations. Our findings disprove

that "memory of germline" role for MES-4, since *mes-4* mutant PGCs turn on most

germline genes normally and can undergo normal germline development if they

inherit X chromosomes with a history of repression. Other attractive contenders for

specifying the germline fate of PGCs are germ granules and small RNAs. Work from

our lab and others demonstrated that *C. elegans* germ granules (aka P granules)

protect germline fate but are not needed to specify it (Gallo et al. 2010, Updike et al.

2014; Knutson and Egelhofer et al. 2017). 22G small RNAs (22 nucleotides long and

starting with a G) bound to the argonaute CSR-1 have been shown to target most

germline-expressed genes and promote expression of some (Claycomb et al. 2009,

Wedeles et al. 2013; Cecere et al. 2014). Complete loss of CSR-1 or DRH-3, an RNA

helicase that generates 22G RNAs, causes sterility (Duchaine et al. 2006, Claycomb

et al. 2009; Gu et al. 2009). However, hypomorphic mutations in the helicase domain

of DRH-3 that abolish production of most 22G RNAs do not impact germline

formation, suggesting that 22G RNAs are not needed to specify germline fate (Gu et

al. 2009). Moreover, our tests of the impact of loss of 22G small RNAs in PGCs

dissected from *drh-3* mutants demonstrated that they do not play a large role in

regulating gene expression in PGCs (Appendix 1). We propose the intriguing

possibility that in *C. elegans*, germline fate is the default, which must be protected in

the germline (e.g. by MES proteins and P granules) and opposed in somatic tissues

(e.g. by DREAM and LIN-15B).

METHODS

Worm strains

All worms were maintained at 20˚C on 6-cm plates containing Nematode Growth Medium (NGM) spotted with *E. coli* OP50 *(*Brenner, 1974). Strains used and generated (*) in this study are listed below.

DUP64 - *glh-1(sams24[glh-1::GFP::3xFLAG]) I*

SS1491* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; mes-4(bn73)/tmC12[egl-9(tmIs1194)] V*

SS1293* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; mrg-1(qa6200)/qC1[dpy-19(e1259) glp-1(q339)] III*

SS1492* - *mes-3(bn199)/tmC20 [unc-14(tmIs1219) dpy-5(tm9715)] glh-1(sams24[glh-1::GFP::3xFLAG]) I*

SS1476* - *met-1(bn200)/tmC20[unc-14(tmIs1219) dpy-5(tm9715)] I*

SS1494*- *met-1(bn200)/tmC20[unc-14(tmIs1219) dpy-5(tm9715)] glh-1(sams24[glh-1::GFP::3xFLAG]) I; mes-4(bn73)/tmC12[egl-9(tmIs1197)] V*

SS1497* - *glh-1(sams24[glh-1::GFP::3xFLAG]); oxTi421[eft-3p::mCherry::tbb-2 3'UTR + Cbr-unc-119(+)] X*

SS1514* - *glh-1(sams24[glh-1::GFP::3xFLAG]); mes-4(bn73)/tmC12[egl-9(tmIs1194)] V; oxTi421[eft-3p::mCherry::tbb-2 3'UTR + Cbr-unc-119(+)] X*

SS1503* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; him-8(e1489) IV*

SS1500* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; him-8(e1489) IV*; *mes-4(bn73)/tmC12[egl-9(tmIs1194)] V*

SS1498* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; him-8(e1489) IV; oxTi421[eft-3p::mCherry::tbb-2 3'UTR + Cbr-unc-119(+)] X*

SS1493* - *glh-1(sams24[glh-1::GFP::3xFLAG]); him-8(e1489) IV; mes-4(bn73)/tmC12[egl-9(tmIs1194)] V; oxTi421[eft-3p::mCherry::tbb-2 3'UTR + Cbr-unc-119(+)] X*

SS1515* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; ccTi1594[mex-5p::GFP::gpr-1::smu-1 3'UTR + Cbr-unc-119(+), III: 680195] III; hjSi20[myo-2p::mCherry::unc-54 3'UTR] IV*

SS1516* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; ccTi1594[mex-5p::GFP::gpr-1::smu-1 3'UTR + Cbr-unc-119(+), III: 680195] III; hjSi20[myo-2p::mCherry::unc-54 3'UTR] IV; mes-4(bn73)/tmC12[egl-9(tmIs1194)] V*

SS1517* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; ccTi1594[mex-5p::GFP::gpr-1::smu-1 3'UTR + Cbr-unc-119(+), III: 680195] III; hjSi20[myo-2p::mCherry::unc-54 3'UTR] IV; mes-4(bn73)/tmC12[egl-9(tmIs1194)] V; lin-15B(n744) X*

SS1518* - *met-1(bn200) glh-1(sams24[glh-1::GFP::3xFLAG]) I; ccTi1594[mex-5p::GFP::gpr-1::smu-1 3'UTR + Cbr-unc-119(+), III: 680195] III. hjSi20[myo-2p::mCherry::unc-54 3'UTR] IV*

SS1511* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; mes-4(bn73)/tmC12[egl-9(tmIs1194)] V; lin-15B(n744) X*

SS1512* - *mes-3(bn199)/tmC20[unc-14(tmIs1219) dpy-5(tm9715)] glh-1(sams24[glh-1::GFP::3xFLAG]) I; lin-15B(n744) X*

SS1513* - *glh-1(sams24[glh-1::GFP::3xFLAG]) I; mrg-1(qa6200)/qC1[dpy-19(e1259) glp-1(q339)] III; lin-15B(n744) X*

Creation of *mes-3* and *met-1* null alleles by CRISPR-Cas9

We created the null alleles *mes-3(bn199)* and *met-1(bn200)* linked to glh-1::GFP by inserting TAACTAACTAAAGATCT into the 1st exon of each locus. The resulting genomic edit introduced a TAA stop codon in each reading frame, a frame shift in the coding sequence, and a BglII restriction site (AGATCT) for genotyping. Alt-R crRNA oligos (IDT) were designed using CRISPOR (Concordet and Haeussler, 2018) and the UCSC Genome Browser (ce10) to produce highly efficient and specific Cas9 cleavage in the 1st exons of *mes-3* and *met-1*. Ultramer ssDNA oligos (IDT) containing 50 bp micro-homology arms were used as repair templates. We used a *dpy-10* co-CRISPR strategy (Arribere et al. 2014) to isolate strains carrying our desired mutations. Briefly, 2.0 uL of 100 µM *mes-3* or *met-1* crRNA and 0.5 uL of 100 µM *dpy-10* cRNA were annealed to 2.5 uL of 100 µM tracrRNA (IDT) by incubating at 95˚C for 2 minutes, then room temperature for 5 minutes, to produce sgRNAs. We complexed sgRNAs with 5 uL of 40 µM Cas9 protein at room temperature for 5 minutes, added 1 uL of 40 µM *mes-3* or *met-1* repair template and 1 uL of 40 µM *dpy-10(cn64)* repair template, and centrifuged the mix at 13,000g for 10 minutes. All RNA oligos were resuspended in IDT's duplexing buffer (cat). Mixes were injected into 1 or both gonad arms of ~30 DUP64 adults. Transformant progeny were isolated and back-crossed 4x to DUP64.

Isolation of single sets of 2 sister PGCs or 2 EGCs

To obtain near-synchronous animals for dissection, we hatched larvae within a 30-minute window in the absence of food. We then allowed those newly hatched L1 larvae to feed for 30 minutes to start PGC development. Larvae were partially immobilized in 15 uL drops of egg buffer (25 mM HEPES, pH 7.5, 118 mM NaCl, 48 mM KCl, 2 mM MgCl$_2$, 2 mM CaCl$_2$, adjusted to 340 mOsm) on poly-lysine coated microscope slides and hand-dissected using 30-gauge needles to release their gonad primordium (consisting of 2 connected sister PGCs and 2 somatic gonad precursors). Isolation of single sets of sister PGCs from somatic gonad precursors and into tubes for transcript profiling involved fluorescence microscopy to identify PGCs by germline-specific expression of GLH-1::GFP and mouth pipetting using pulled glass capillaries coated with Sigmacote (Sigma SL2) and 1% BSA in egg buffer. 7.5 mg/mL pronase (Sigma P8811) and 5 mM EDTA were added to reduce the adherence of gonad primordia to the poly-lysine coated slides and to weaken cell-cell interactions. While monitoring by microscopy, gonad primordia were moved to a drop of fresh egg buffer and PGCs were separated from gonad precursors using shearing force generated by mouth pipetting. Finally, we transferred single sets of sister PGCs into 0.5 uL drops of egg buffer placed inside the caps of 0.5 mL low-bind tubes (USA Scientific, #1405-2600). We only selected PGCs for transcript profiling if they maintained bright fluorescence of GFP throughout isolation and were clearly separated from somatic gonad precursors. Isolation of EGCs differed in 3 ways: 1) EGCs were dissected from L2 larvae that were fed for 20 hours after hatching, 2) the

2 EGCs that made up 1 sample may have come from different animals, and 3) the stage of each EGC could not be determined and therefore may have differed between samples. Tubes containing single sets of 2 sister PGCs or 2 EGCs were quickly centrifuged, flash frozen in liquid nitrogen, and then stored at -80˚C. A detailed protocol for isolating PGCs and EGCs from larvae is available upon request. At least 11 samples (replicates) of PGCs or EGCs were isolated for each condition.

Isolation of RNAs from adult germlines

We dissected 1st day hermaphrodite adult worms (approximately 20-24 hours post-mid-L4 stage) with 30-gauge needles in egg buffer (see recipe above, except not adjusted to 340 mOsm) containing 0.1% Tween and 1 mM levamisole to extrude their gonads. Gonads were cut at the narrow 'bend' to separate the gonad region containing mitotic and early meiotic germ cells from the region that contains oocytes and/or sperm; the former was used for RNA profiling. 20-60 gonads were mouth pipetted into 500 uL Trizol reagent (Life Technologies, #15596018), flash-frozen in liquid nitrogen, and stored at -80˚C for up to 1 month before RNA extraction.

To release RNAs from gonads in Trizol, we performed 3 freeze-thaw cycles using liquid nitrogen and a 37˚C water bath, while vortexing vigorously between each cycle. RNAs immersed in Trizol were added to phase-lock heavy gel tubes (need vendor) and mixed with 100 uL of 1-bromo-3-chloropropane (BCP) (Sigma #B9673) by handshaking, followed by room temperature incubation for 10 minutes. Samples were then centrifuged at 13,000g and 4˚C for 15 minutes to separate phases. RNAs in

the aqueous phase were precipitated by mixing well with 0.7x-0.8x volumes of ice-cold isopropanol and 1 uL 20 mg/mL glycogen, followed by incubation at -80˚C for 1-2 hours and centrifugation at 13,000g at 4˚C for 30 minutes. RNA pellets were washed 3x with ice-cold 75% ethanol and then resuspended in 15 uL water. RNA concentration was determined by a Qubit fluorometer, purity was determined by a Nanodrop, and integrity (RIN values) was determined by a bioanalyzer or tapestation.

Generation of cDNA sequencing libraries

PGCs and EGCs: Immediately after thawing PGCs and EGCs on ice, a 1:4,000,000 dilution of ERCC spike-in transcripts (Life Technologies, #4456740) were added to each sample. Double-stranded cDNAs from polyA(+) RNAs were generated using a SMART-seq method that combined parts of the SMART-seq2 (Picelli et al. 2014) and SMART-seq v4 (Takara) protocols. Briefly, PGCs and EGCs were lysed at room temperature for 5 minutes in a lysis buffer containing RNAse inhibitors (Takara, #2313) to release mRNAs into solution. 1.2 µM custom DNA primer (IDT) was annealed to transcripts' polyA tails by incubating samples at 72˚C for 3 minutes and then immediately placing them on ice. Reverse transcription to generate double-stranded cDNA was performed using 200 U SmartScribe, 1x first-strand buffer, 2 mM DTT (Takara, #639537), 1 mM dNTPs (Takara, #639125), 4 mM MgCl$_2$, 1 M betaine (Sigma, #B0300), 20 U RNAse Inhibitor (Takara), and 1.2 µM custom template-switching oligo with a Locked Nucleic Acid analog (Qiagen) at 42˚C for 90 minutes, followed by 70˚C for 15 minutes to heat-inactivate the reverse

transcriptase. cDNAs were PCR-amplified according to Takara's SMART-seq v4 protocol for 20 cycles using SeqAmp DNA Polymerase (Takara, #638504) and a custom PCR primer. Amplified cDNAs were purified by SPRI using 1x Ampure XP beads (Agencourt, #A63881) and quantified using a Qubit fluorometer. All custom oligos contained a biotin group on their 5' end to ameliorate oligo concatemerization. Illumina's Nextera XT kit (Illumina, #FC-131-1096) was used with 350-400 pg cDNA as input to prepare dual-indexed Illumina RNA-sequencing libraries according to the manufacturer's instructions. Libraries were PCR-amplified for 14 cycles and purified by SPRI using 0.6x Ampure XP beads.

Adult germlines: Illumina libraries were prepared from polyA(+) RNAs using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490) and the NEBNext Ultra RNA Library Prep Kit (#E7530) according to the manufacturer's instructions. 100 ng total RNA was used to generate cDNAs, which were then amplified using 15 cycles of PCR.

Library concentration was measured by a Qubit fluorometer, and average fragment size was measured by either a bioanalyzer or a tapestation (Agilent). Libraries were multiplexed and paired-end sequenced with 50 cycles on either an Illumina HiSeq2500 or NovaSeq 6000 SP flowcell at the QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Processing and analysis of RNA-seq data

PGCs and EGCs: Paired-end Illumina sequencing reads were aligned to the ce10 (WS220) genome downloaded from Ensembl using Hisat2 (v2.2.1). Samtools (v1.10) was then used to remove duplicate reads and reads with low mapping quality (MAPQ < 10) from the alignment file. The function featureCounts from the subread package (v2.0.1) (Liao et al. 2014) was used to obtain gene-level read (fragment) counts using a ce10 (WS220) transcriptome annotation file (Ensembl), which additionally contained 92 ERCC spike-in transcripts. 11 low-quality transcript profiles that were likely caused by a failure to capture mRNAs from 2 PGCs or EGCs, a well-known problem in single-cell RNA-sequencing, were identified using the R package scuttle (v1.2.0) (McCarthy et al. 2017). The R package DESeq2 (v1.32.0) (Love et al. 2014) was used to identify differentially expressed genes in mutant vs wild-type samples. P-values were adjusted for multiple hypothesis testing by the Benjamini-Hochberg method to produce q-values. Genes with a q-value < 0.05 were considered differentially expressed. The scaling factors used by DESeq2 to normalize transcript profiles were calculated using the R package scran (v1.20.1), which uses a pooling and deconvolution approach to deal with zero-inflation in low-input RNA-seq data (Lun et al. 2016). Bigwig files containing normalized read coverage over the WS220/ce10 genome were generated by bamCoverage from deepTools (Ramirez et al. 2016) using library scaling factors computed by scran, a bin size of 5, and a smoothing window of 15. Average read coverage across replicates was computed using wiggletools (Zerbino et al. 2014) and visualized on the UCSC Genome Browser. Principal component analysis (PCA) was performed using

DESeq2's plotPCA function and variance-stabilized counts from DESeq2 vst function. All visualizations of RNA-seq data were generated by the R packages 'ggplot2' (v3.3.5) and ggpubr (v0.4.0) or the UCSC Genome Browser.

Adult germlines: For transcript profiles from dissected adult germlines, data were processed as described above except the scaling normalization factors were computed by DESeq2.

Gene sets

The germline-specific, germline-enriched, and soma-specific gene sets were previously defined using microarray and SAGE data. The germline-specific and germline-enriched gene sets were previously defined in Rechtsteiner et al. 2010 and Reinke et al. 2004, respectivelyThe soma-specific set was previously defined in Knutson and Egelhofer 2007 and refined in this study by removing genes from the set that have >5 TPM in RNA-seq data from wild-type dissected adult germlines (this study). The oogenesis and spermatogenesis were previously defined in Ortiz et al. 2014 (they called 'oogenic' and 'spermatogenic'). Oogenesis genes are expressed at higher levels in dissected adult oogenic germlines than dissected adult spermatogenic germlines; spermatogenesis genes are the opposite.

Single-molecule fluorescence in-situ hybridization (smFISH) in L1 larvae

100-200 gravid adult mothers were allowed to lay offspring in drops of S basal overnight. Starvation-synchronized L1 offspring were collected and fed HB101

bacteria in S basal for 5 hours. Fed L1s were washed 3-4 times with S basal to remove bacteria and then used for smFISH using a protocol described in van Oudernaarden et al. 2012 with a few modifications. Briefly, L1s were fixed with 3.7% formaldehyde for 45 minutes at room temperature, followed by 3 washes with PBS-Tween (0.1%). Fixed L1s were incubated in 75% ethanol at 4˚C overnight and up to 3 days. RNAs were hybridized to 25 nM RNA probe sets in hybridization buffer (2x SSC, 10% formamide, 0.1% Tween-20, and 0.1 g/mL dextran sulfate) at 37˚C overnight. Afterwards, larvae were washed 2 times in hybridization buffer at 37˚C for 30 minutes, the second of which included 1 ng/uL of DAPI. Larvae were washed 3 times in PBS-Tween (0.1%) and mounted in an anti-fade medium consisting of n-propyl gallate and Vectashield (Vector Laboratories, H-1000). Mounted samples were immediately imaged on a spinning disk confocal microscope using a 100x oil objective to acquire 3D Z-stacks of PGCs; only Z-slices containing GLH-1::GFP signal were imaged. All RNA probe sets were conjugated to Quasar670 fluorescent dye. We purchased the *lsd-1*, *mbk-1*, *pek-1*, and *lin-15B* probe sets from Stellaris. The *cpg-2, pgl-3,* and *chs-1* probe sets were gifts from Dr. Erin Osbourne Nishimura.

Counting transcripts in 3D smFISH images

We developed and used a custom pipeline written in Fiji (version) and MATLAB (version) to batch process raw smFISH 3D images into transcript abundance measurements in PGCs. We adapted much of the MATLAB code and strategy from Raj et al. 2008 to create our pipeline. Images were removed from

downstream analyses if they were obviously very dim in all channels. Numbers of analyzed PGCs per probe set per genotype are indicated in Figure 3-2. A Laplacian of Gaussian (LoG) filter was used to enhance the signal-to-noise contrast in smFISH images. Those filtered 3D images were thresholded by signal intensity to produce binary images. The 'imregionalmax' function from the Imaging Processing Toolbox in MATLAB was used to find and count regional maxima (transcript foci) in those binary images. To isolate and count regional maxima in PGCs, a 2D binary image mask of the PGCs was generated in Fiji from a maximum-intensity Z-projection of the GLH-1::GFP image channel and applied to all Z-slices in the 3D smFISH images. Segmentation of PGCs to create the 2D mask was done in Fiji by first blurring Z-projections using a large Gaussian kernel, and then detecting edges in the blurred image. Boxplots comparing transcript counts in PGCs between genotypes of larvae were generated by the R packages ggplot2 and ggpubr.

To choose an appropriate signal intensity threshold for a 3D smFISH image, we plotted the number of detected regional maxima across 100 increasingly stringent thresholds applied to the image (Raj et al. 2008). In that plot, a range of thresholds that produced similar numbers of detected regional maxima was identified, and a threshold within that range was selected. Since threshold values were similar for all images within an image set (the collection of images acquired on the same day and for one probe set), an averaged threshold (across 5 images) was calculated and applied to all images in the set. Notably, threshold values were similar across

89

different image sets. Manual counting of dots in a few smFISH images gave similar values as our semi-automated pipeline.

Analysis of germline size in adults

All analyses were performed by live imaging 1st day adults (approximately 20-24 hours post-mid-L4 stage) and evaluating germline size using the germline marker GLH-1::GFP. GLH-1 (Vasa) is a component of germ granules and is specifically and highly expressed in the germline. Adult germlines were classified into 1 of 3 categories: (1) 'Full' if its size was similar to that of a wild-type adult germline, (2) 'Partial' if it had at least ~15 GLH-1::GFP(+) germ cells but wasn't large enough to be classified as 'full', and (3) 'Absent/tiny' if it had < 15 GLH-1::GFP(+) germ cells. In rare ambiguous cases where a germline's size could not be distinguished between 2 categories, the germline was assigned to the category of the smaller size. To classify germline size for hermaphrodites, which have 2 gonad arms, we only considered the size of the larger germline; in most cases, the germlines in both gonad arms were similar in size.

We used confocal microscopy to acquire live images of scored germlines in adults. Live adults were placed in a drop of 1 uL H_2O and 1 uL polystyrene microspheres (Polysciences, Inc. #00876) and then immobilized on 6% agarose pads. Images were acquired in Z-stacks using a 20x air objective and then converted to Z-projections of maximum intensity using Fiji. DIC projections were used to outline the body of worms.

Tracking X-chromosome inheritance patterns

Methods used to track X-chromosome inheritance are diagrammed in Figure 3-9. To identify F1 male offspring that received their single X from the oocyte (X^{oo}) or the sperm (X^{sp}), we performed crosses with 1 parent contributing an X-linked *eft-3p::mCherry* transgene. F1 male offspring that inherited the transgene from one parent were easily distinguished by bright cytoplasmic mCherry fluorescence in their soma. We used a *gpr-1* over-expression allele to generate F1 hermaphrodite offspring whose germlines inherited either 2 genomes from the sperm or 2 genomes from the oocyte (Besseling and Bringmann, 2016; Artilles et al. 2019). Those non-Mendelian offspring were visually identified by patterns of mCherry fluorescence in their pharyngeal muscle cells (*myo-2p::mCherry* X).

RNAi depletion

We performed RNAi by feeding worms *E. coli* HT115 bacteria that carry a gene target's DNA sequence in the L4440 vector (Kamath and Ahringer, 2003). Most RNAi constructs were obtained from the Ahringer RNAi library and sequence confirmed. *lin-15B*, *lsy-2*, *nfya-1*, *eor-1*, and *sma-9* RNAi constructs were generated in this study. RNAi constructs were streaked onto LB agar plates containing 100 ug/mL carbenicillin and 10 ug/mL tetracycline. Single clones were cultured overnight (14-17 hours) at 37˚C in LB and carbenicillin (100 ug/mL). The following day, RNAi cultures were spotted onto 6-cm NGM plates containing 1 mM IPTG and 100 ug/mL

carbenicillin (both added by top spreading), and then left to dry for 2 days at room temperature in the dark. To deplete both maternal load and newly synthesized gene product, we placed L4-stage larval mothers onto RNAi plates, grew them 1 day to reach adulthood, then moved those adults to new RNAi plates to lay offspring. RNAi was done at 20˚C. We scored germline size in 1st day adult offspring as described above.

Statistical analysis

Sample sizes for worms scored for germline size and for images used to count transcripts in smFISH analysis are indicated in the respective figures. In our germline-size analysis, 2-tailed Fisher's exact tests were performed to test whether the proportion of worms with a 'full' germline is higher or lower in 1 sample vs another sample. In our smFISH analysis, Mann-Whitney tests were performed to compare transcript abundance between *mes-4* and wt PGCs. Gene set enrichment analyses were performed using hypergeometric tests in R. The sizes of gene sets used for those tests are noted in the respective figure legends. For Fisher's exact tests, hypergeometric tests, and Mann-Whitney tests, P-value designations are $* < 0.01$, $** < .001$, $*** < 1e-4$, and $**** < 1e-5$. In our RNA-seq analysis, differentially expressed genes were identified using Wald tests in the R package DESeq2 (see above) as those with a q-value < 0.05.

Spinning-disk confocal microscopy

All images were acquired using a spinning-disk confocal microscope

equipped with a Yokogawa CSU-X1 confocal scanner unit, Nikon TE2000-E inverted

stand, Hamamatsu ImageEM X2-CCD camera, solid state 405, 488, 561, 640 nm

laser lines, 460/50, 525/50, 593/40, 700/75 nm (EM/BP) fluorescence filters, DIC,

Nikon Plan Apo VC 20×/0.5 air objective, Nikon Plan Apo 100×/1.40 oil objective,

and Micro-Manager software (v 1.4.20). Image processing for images was done in

Fiji (v 2.1.0/1.53C) and photoshop.

Transcription factor analysis motif analyses

Analyses of ChIP data: Bed files containing transcription factor (TF) binding

sites (ChIP-chip or ChIP-seq 'peaks') across the ce10 genome were downloaded from

the modENCODE (Gerstein et al. 2010) and modERN (Kudron et al. 2018) websites.

Each bed file was loaded into R and converted into a GRanges object using the

GenomicRanges package. TF binding sites were assigned to genes using the package

ChIPSeeker. A TF binding site was assigned to a gene if it overlapped with a gene's

TSS region (500 bp upstream of TSS). If a TF had more than 1 set of binding site data

(e.g. 2 ChIP-seq experiments), we processed each set separately and then merged the

TF-assigned genes. TFs enriched in promoters of X-linked UP genes in *mes-4* PGCs

or EGCs (584 genes) compared to all X-linked genes (2808) were identified using

hypergeometric tests (P-value < 0.05).

Analyses of DNA motifs: We downloaded position-weight matrices for known

C. elegans TF DNA motifs from the CisBP database. X-linked genes that contain a

TF's motif in their promoter (500 bp upstream of TSS) were identified using FIMO

from the Meme suite in R using default parameters. Motifs significantly enriched in

promoters of X-linked UP genes in *mes-4* PGCs or EGCs (584 genes) compared to all

X-linked genes (2808) were identified using hypergeometric tests (P-value < 0.05).

Figure 3-1 (on next page). Transcriptome analysis of *mes-4* M-Z- mutant PGCs.
(A) Cartoon illustrating the Maternal-effect sterile (Mes) phenotype in *mes-4* mutants.
mes-4 M+Z- (M for Maternal, Z for Zygotic) mutants cannot synthesize MES-4 (Z-)
but are fertile because they inherited maternal MES-4 gene product (M+). Removal of
maternal MES-4 renders *mes-4* M-Z- mutant adults sterile. Germline is colored green,
and soma is colored white. We profiled transcripts from single sets of 2 sister
Primordial Germ Cells (PGCs) and from single sets of 2 Early Germ Cells (EGCs)
(red boxes) hand-dissected from *mes-4* M-Z- (hereafter called *mes-4*) mutant and
wild-type (wt) L1 and L2 larvae, respectively. (B,C,D) MA plots showing
$\log_2$(average expression) versus $\log_2$(fold change) of transcript abundance for 20,258
protein-coding genes (circles) between *mes-4* and wt PGCs. Genes belonging to a
specific gene set are colored: (B) 168 germline-specific genes are blue, (C) 861 soma-
specific genes are green, and (D) 2808 X-chromosome genes are red. Differentially
expressed genes in *mes-4* vs wt PGCs were identified using Wald tests in DESeq2
(Love et al. 2014) and by setting a q-value < 0.05 significance threshold. Numbers of
all mis-regulated genes (black) and numbers of those in gene sets (colored) are
indicated in the corners; top is upregulated (UP) and bottom is downregulated
(DOWN) in *mes-4* vs wt. (E,F,G) Bar plots showing the expected number (light gray)
and observed number (dark gray) of mis-regulated genes that are members of gene
sets indicated below the bars. Hypergeometric tests were performed in R to test for
gene-set enrichment. P-value designations are **** < 1e-5. (E) Enrichment analyses
for DOWN genes were restricted to 5,858 protein-coding genes that we defined as
'expressed' (minimum average read count of 1) in wt PGCs. Gene set sizes: germline-
specific (140), germline-enriched genes (1867), ubiquitous (2143). (F,G) Enrichment
analyses for UP genes included all 20,258 protein-coding genes in the transcriptome.
Gene set sizes: soma-specific (861), germline-specific (168), germline-enriched
(2176), spermatogenesis-enriched (2498), gender-neutral (5973), oogenesis-enriched
(1671), oogenesis-enriched on X (470), oogenesis-enriched on Autosomes (1201),
chrI (2888), chrII (3508), chrIII (2670), chrIV (3300), chr V (5084), chr X (2808).
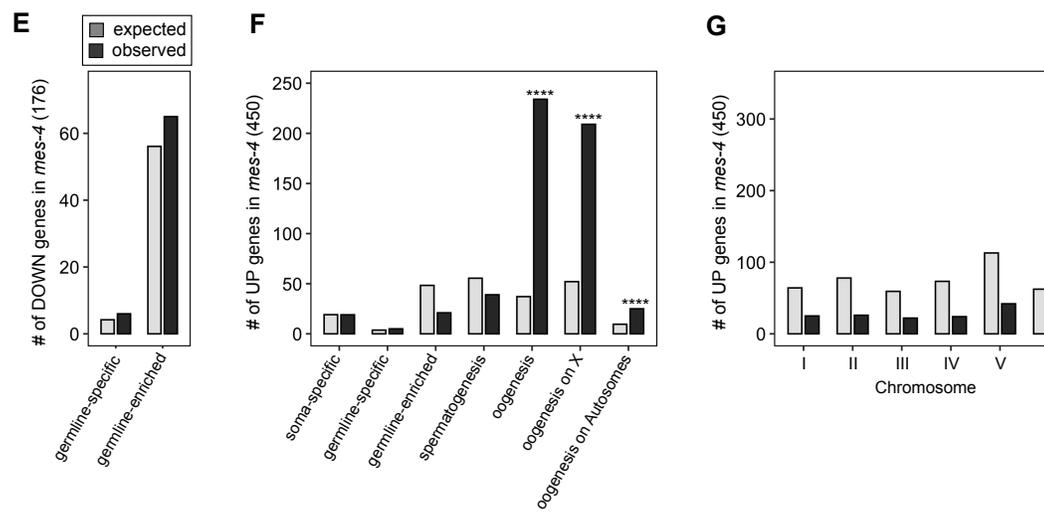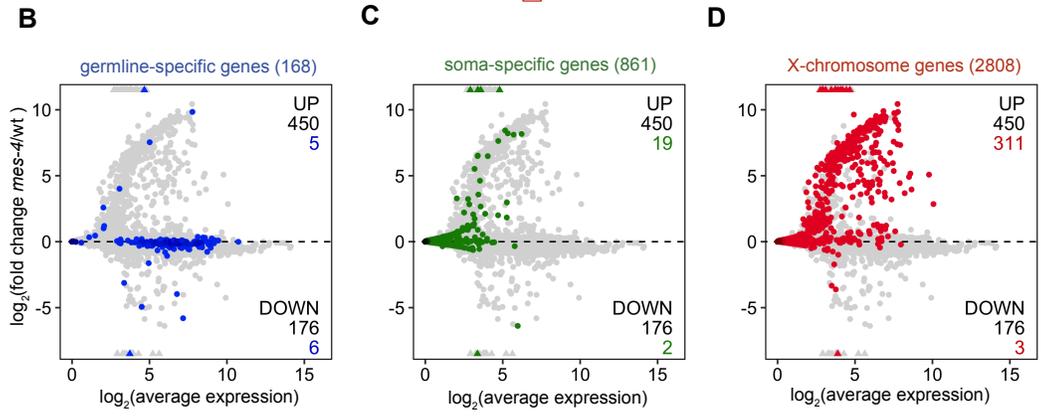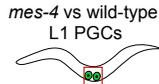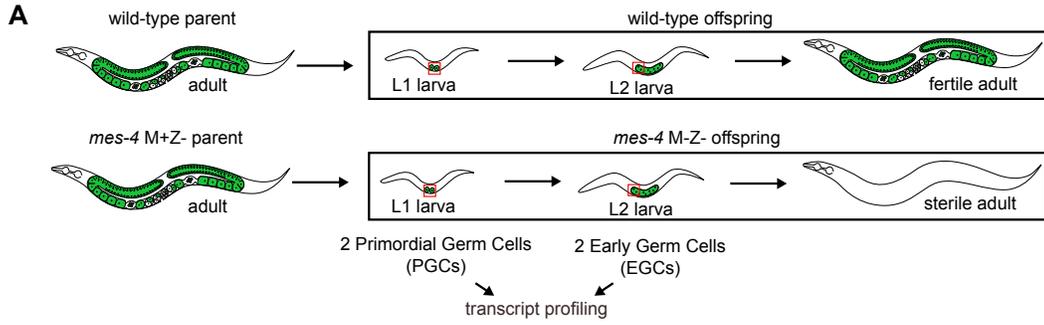
Figure 3-2 (on next page). Transcript quantification in PGCs by single-molecule FISH (smFISH) corroborates the findings from transcript profiling. (A,D) UCSC genome browser images showing average sequencing read coverage over Ensembl gene models for wild-type (wt) PGCs (top track) and *mes-4* PGCs (bottom track). (A) 2 germline-specific genes: *cpg-2* (left) and *pgl-3* (right). *cpg-2* is DOWN and *pgl-3* is not DOWN in transcript profiling of *mes-4* vs wt PGCs. (D) 2 X-linked UP genes in transcript profiling of *mes-4* vs wt PGCs: *lsd-1* (left) and *lin-15B* (right). (B,E) Representative maximum-intensity Z-projection images of smFISH experiments in L1 larvae. DAPI-stained nuclei are in red. GLH-1::GFP is in green. The dashed lines circumscribe PGCs marked by GLH-1::GFP. The 2nd and 3rd images in each set are zoomed insets of the yellow box in the 1st image. Foci in the mRNA channel (3rd image in each set) represent individual transcripts. Scale bars are 10 microns. (B) smFISH RNA probes targeting *cpg-2* (left) and *pgl-3* (right) transcripts. (E) smFISH RNA probes targeting *lsd-1* (left) and *lin-15B* (right) transcripts. (C,F) Transcript quantification in PGCs in smFISH 3D images. Each circle represents 1 quantified image. The number of quantified images for each combination of probe and genotype is indicated. Boxplots show the median, the 25th and 50th percentiles (boxes), and the 2.5th and 97.5th percentiles (whiskers). Mann-Whitney tests were used to compare a gene's transcript counts between *mes-4* and wt PGCs. P-value designations are NS > .01, **** < 1e-5 . (C) Quantification of *cpg-2*, *pgl-3*, and *chs-1* transcripts. *chs-1* is a member of our 'germline-enriched' gene set and is not DOWN in transcript profiling of *mes-4* vs wt PGCs. (F) Quantification of *lsd-1*, *mbk-1*, *pek-1*, and *lin-15B* transcripts, 4 X-linked UP genes in transcript profiling of *mes-4* vs wt PGCs.

mes-4 vs wild-type
L1 PGCs

A

germline-specific DOWN
in *mes-4* PGCs

germline-specific not DOWN
in *mes-4* PGCs

wt

mes-4

cpg-2

wt

mes-4

pgl-3

B

DAPI
GLH-1::GFP   *cpg-2* mRNA

wt

mes-4

DAPI
GLH-1::GFP   *pgl-3* mRNA

wt

mes-4

C

# of mRNA molecules

wt PGCs   *mes-4* PGCs

**** NS NS

n: 20   23   21   27   25   25
probe: *cpg-2*   *pgl-3*   *chs-1*

germline-specific
genes

germline-enriched
gene

D

X UP
in *mes-4* PGCs

X UP
in *mes-4* PGCs

wt

mes-4

lsd-1

wt

mes-4

lin-15B

E

DAPI
GLH-1::GFP   *lsd-1* mRNA

wt

mes-4

DAPI
GLH-1::GFP   *lin-15B* mRNA

wt

mes-4

F

# of mRNA molecules

wt PGCs   *mes-4* PGCs

****   ****   ****   ****

n: 15   10   16   16   15   15   27   17
probe: *lsd-1*   *mbk-1*   *pek-1*   *lin-15B*

X-chromosome
genes

Figure 3-3 (on next page). Maternally loaded MES-4 promotes germline development by repressing the X chromosomes independently from transmitting H3K36me3 across generations. (A,B) Bar plots showing distributions of germline size ('absent/tiny', 'partial', and 'full') in worms with different X-chromosome compositions in their germline. Numbers of scored F1 offspring and the genotype of their germlines are indicated below each bar (*mes-4* indicates *mes-4* M-Z-). 2-cell embryos contain AB (left) and P1 (right) blastomeres. AB generates some somatic tissues, and P1 generates the germline and some somatic tissues. Orange, blue, and green coloring mark the inheritance pattern of X chromosome(s) in F1 offspring: AB and P1 embryo blastomeres and adult germlines colored orange contain only oocyte-inherited X(s), those colored blue contain only sperm-inherited X(s), and those colored green contain 1 oocyte-inherited X and 1 sperm-inherited X. Parent gametes are colored according to the X they transmit to offspring. To generate 'non-Mendelian' F1 hermaphrodite offspring that inherited 2 genomes and therefore 2 Xs from one gamete, we mated fathers to mothers that carry a mutation in *gpr-1* (Besseling and Bringmann, 2016; Artilles et al. 2019). We used 2-tailed Fisher's exact tests to test whether the proportion of F1 adults with a full-sized germline significantly differs between samples. P-value designations are $* < 0.01$, $**** < 1\mathrm{e}\text{-}5$. (A) To generate F1 male offspring that inherited their single X from the sperm, we mated parents that carry the *him-8(e1489)* allele, which causes X chromosome nondisjunction during oogenesis in the hermaphrodite. (B) Presence or absence of sperm-inherited H3K36me3 marking, maternal MES-4, and maternal MET-1 in germlines of F1 offspring is indicated in the schematic of each cross and below each bar.
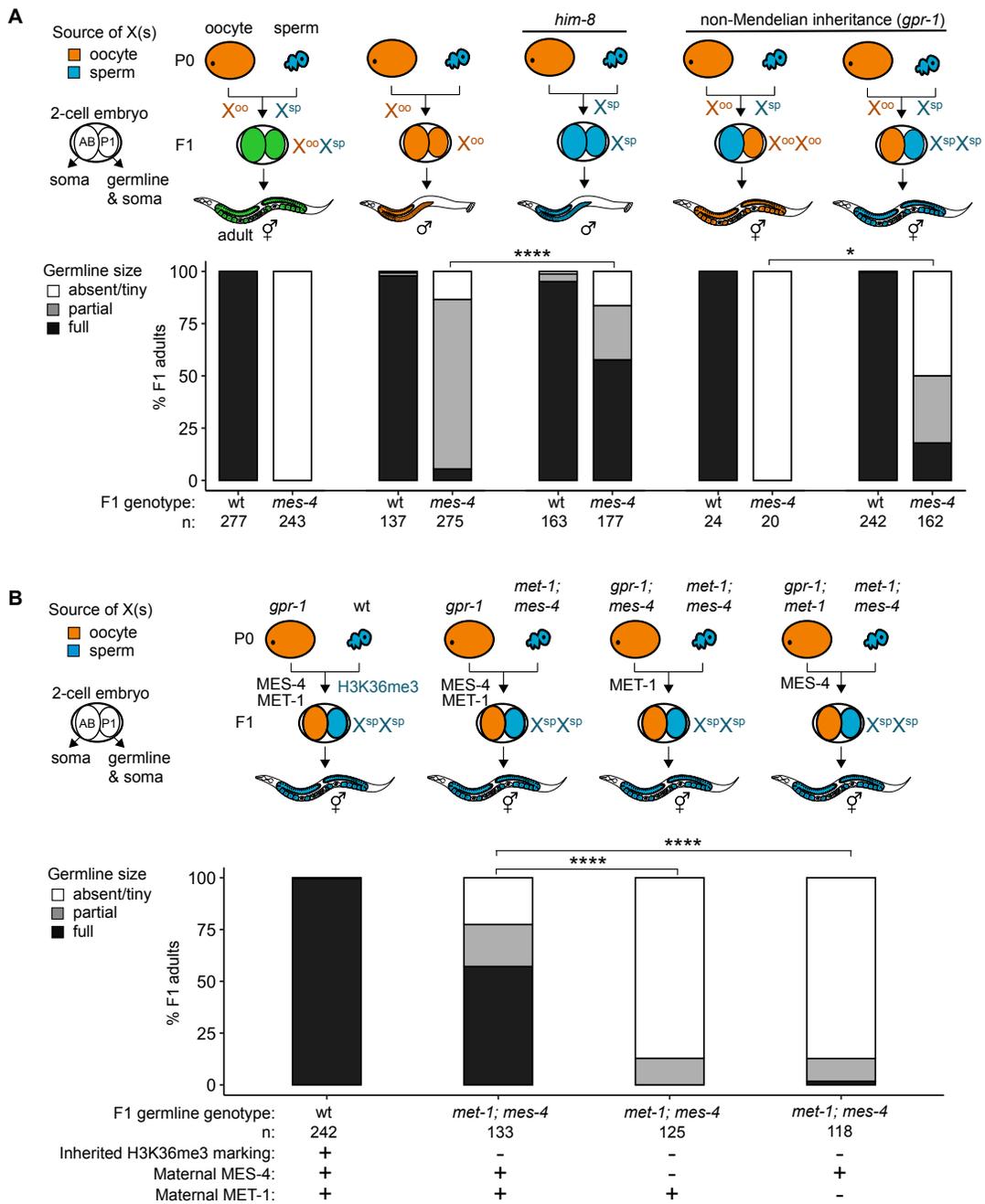
Figure 3-4 (on next page). Loss of LIN-15B reduces X mis-expression in *mes-4* M+Z- adult germlines and suppresses germline death in *mes-4 M-Z-* mutants.
(A) Volcano plot showing $\log_2$(fold change) of transcript abundance and significance [$\log_2$(q-value)] for 20,258 protein-coding genes (circles) between gonads dissected from *mes-4* M+Z- vs wild-type (wt) adults. X-chromosome genes (2808) are colored red. Genes above the horizontal line (q-value of 0.05) are considered significantly mis-regulated. The number of all mis-regulated protein-coding genes (black) and the number of those that are X-linked (red) are indicated in the corners; left is downregulated (DOWN) and right is upregulated (UP) in *mes-4* M+Z- vs wt. (B) Boxplots showing $\log_2$(fold change) in transcript abundance for the set of 303 X-UP genes in *mes-4* M+Z- vs wt gonads between *mes-4* M+Z- vs wt gonads (red) and between *mes-4* M+Z-*; lin-15B* M-Z- vs wt gonads (pink). Boxplots show the median, the 25th and 50th percentiles (boxes), and the 2.5th and 97.5th percentiles (whiskers). Waists around the median indicate 95% confidence intervals. Mann-Whitney tests were used to compare samples. P-value < 1e-5 **** (C) Venn diagram comparing the set of 303 X-UP genes in *mes-4* M+Z- vs wt and the set of 445 X-DOWN genes in *mes-4* M+Z-*; lin-15B* M-Z- vs *mes-4* M+Z- gonads. (D) Bar plot as described in the legend of Figure 3-3A. Genotypes of hermaphrodite and male parents are indicated at the top. All scored F1 offspring are non-Mendelian segregants (caused by the *gpr-1* mutation in mother worms) whose germline inherited 2 genomes and therefore 2 Xs from the sperm. The F1 germline's genotype with respect to *lin-15B* is indicated to the right of the 2-cell embryos; *lin-15B(+/+)* indicates 2 wild-type copies of *lin-15B* and *lin-15B(-/-)* indicates 2 null mutant copies of *lin-15B*. The presence or absence of maternal LIN-15B and zygotic LIN-15B in the germlines of F1 offspring is indicated in the schematic of each cross and below each bar. We used 2-s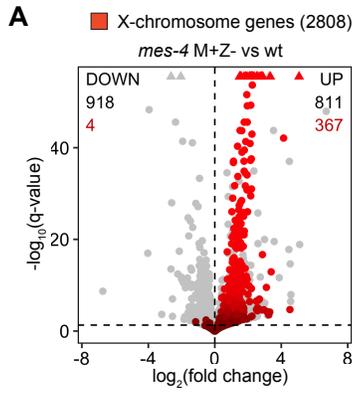ided Fisher's exact tests to test whether the proportion of F1 adults with a full-sized germline significantly differs between samples. P-value designations are ** < 0.001, **** < 1e-5.

mes-4 M+Z- adult hermaphrodite germlines

A X-chromosome genes (2808)

mes-4 M+Z- vs wt

DOWN   UP
918   811
4   367

-log$_{10}$(q-value)

log$_2$(fold change)

B X-UP *mes-4* M+Z- vs wt
(367 genes)

****

log$_2$(fold change mutant/wt)

mes-4 M+Z-   *mes-4* M+Z-; *lin-15B* M-Z-

C X-UP
mes-4 M+Z-
vs wt

44
323
156

X-DOWN
mes-4 M+Z-; *lin-15B* M-Z-
vs mes-4 M+Z-

mes-4 M-Z- adult hermaphrodite germlines

D

Source of X(s)
■ oocyte
■ sperm

2-cell embryo

AB P1

soma   germline
        & soma

F1 germline genotype:	*mes-4* M-Z-; *lin-15B* M+Z+	*mes-4* M-Z-; *lin-15B* M-Z-	*mes-4* M-Z-; *lin-15B* M-Z+	*mes-4* M-Z-; *lin-15B* M+Z-
n:	162	91	102	178
Maternal LIN-15B:	+	−	−	+
Zygotic LIN-15B:	+	−	+	−

Germline size
□ absent/tiny
▨ partial
■ full

% F1 adults

102

Figure 3-5 (on next page). *mes-4* **M-Z- and** *mes-3* **M-Z- nascent germlines mis-express a highly similar set of X genes.** (A) Principal Component Analysis (PCA) including all replicates of wt (blue), *mes-4* M-Z- (green), and *mes-3* M-Z- (red) PGCs (circles) and EGCs (triangles). The percentage of total variance across all samples described by the top 2 principal components is indicated. (B,C) Venn diagrams comparing sets of up-regulated X genes (mutant vs wt) between *mes-4* and *mes-3* PGCs (B) and *mes-4* and *mes-3* EGCs (C). (D,E) Scatterplots comparing $\log_2$(fold change) (mutant vs wt) of transcript abundance for X-UP genes (circles) in either *mes-4* or *mes-3* PGCs (D) or in either *mes-4* or *mes-3* EGCs (E). The Spearman correlation coefficient along with its p-value is indicated at the top of each scatterplot. (F) Cartoon model illustrating how MES-4 protects germline survival by repressing X genes (gray boxes). MES-4 may indirectly repress X genes, including *lin-15B*, by concentrating a repressor (e.g. PRC2) onto the X or by restricting an activator (e.g. histone acetyltransferase or LIN-15B) from the X. Our findings identify LIN-15B as a key player in activating X genes and causing germline death upon loss of MES-4. LIN-15B may activate X genes directly by binding to those genes or indirectly by regulating 1 or more other transcription factors.

A

PC2: 12% variance / PC1: 29% variance

- L1 PGCs
- L2 EGCs

wt
mes-4 M-Z-
mes-3 M-Z-

B X UP L1 PGCs

mes-4
(311)

41 270 84

mes-3
(354)

C X UP L2 EGCs

mes-4
(568)

147 421 22

mes-3
(443)

D X-UP
mes-4 or *mes-3* PGCs
(395 genes)

$\log_2$(fold change *mes-3*/wt) PGCs

$R = 0.22, p = 1.1e\text{-}05$

$\log_2$(fold change *mes-4*/wt) PGCs

E X-UP
mes-4 or *mes-3* EGCs
(590 genes)

$\log_2$(fold change *mes-3*/wt) EGCs

$R = 0.44, p < 2.2e\text{-}16$

$\log_2$(fold change *mes-4*/wt) EGCs

F

Wild-type PGCs

MES-4 ⊣ LIN-15B

concentrate PRC2 on X?
restrict activator from X?

lin-15B

X chromosome
repressed

germline survival

mes-4 M-Z- PGCs

LIN-15B

lin-15B

X chromosome
mis-expressed

germline death

104

Figure 3-6 (on next page). *mes-4* EGCs have more severe transcriptome defects than their progenitor *mes-4* PGCs. MA plots as described in Figure 3-1B-D showing differential expression analysis for *mes-4* vs wt Early Germ Cells (EGCs). (D,E,F) Venn diagrams comparing sets of mis-regulated genes (*mes-4* vs wt) between PGCs and EGCs: (D) DOWN germline-specific genes, (E) UP soma-specific genes, and (F) UP X genes. (G,H,I) Bar plots showing the expected number (light gray) and observed number (dark gray) of mis-regulated genes in either *mes-4* PGCs or EGCs that are members of gene sets indicated below the bars. Hypergeometric tests were performed in R to test for gene-set enrichment. P-value designations are * < 0.01, ** < 0.001, **** < 1e-5. (G) Enrichment analyses for DOWN genes were restricted to 6,682 protein-coding genes that we defined as 'expressed' (minimum average read count of 1) in either wt PGCs or EGCs. Gene set sizes: germline-specific (146), germline-enriched genes (1887). (H,I) Enrichment analyses for UP genes included all 20,258 protein-coding genes in the transcriptome. Gene set sizes (H,I) : soma-specific (169, 629), germline-specific (15,153), germline-enriched (65, 2111), spermatogenesis (171, 2327), oogenesis (470, 1201).

Figure 3-7 (on next page). Analysis of features of mis-regulated genes in *mes-4* PGCs and EGCs. (A,B) MA plots as described in Figure 3-1B-D showing differential expression analysis for (A) *mes-4* vs wt PGCs and (B) *mes-4* vs wt EGCs. Colored circles represent mis-regulated genes (q-value < 0.05); UP genes are in 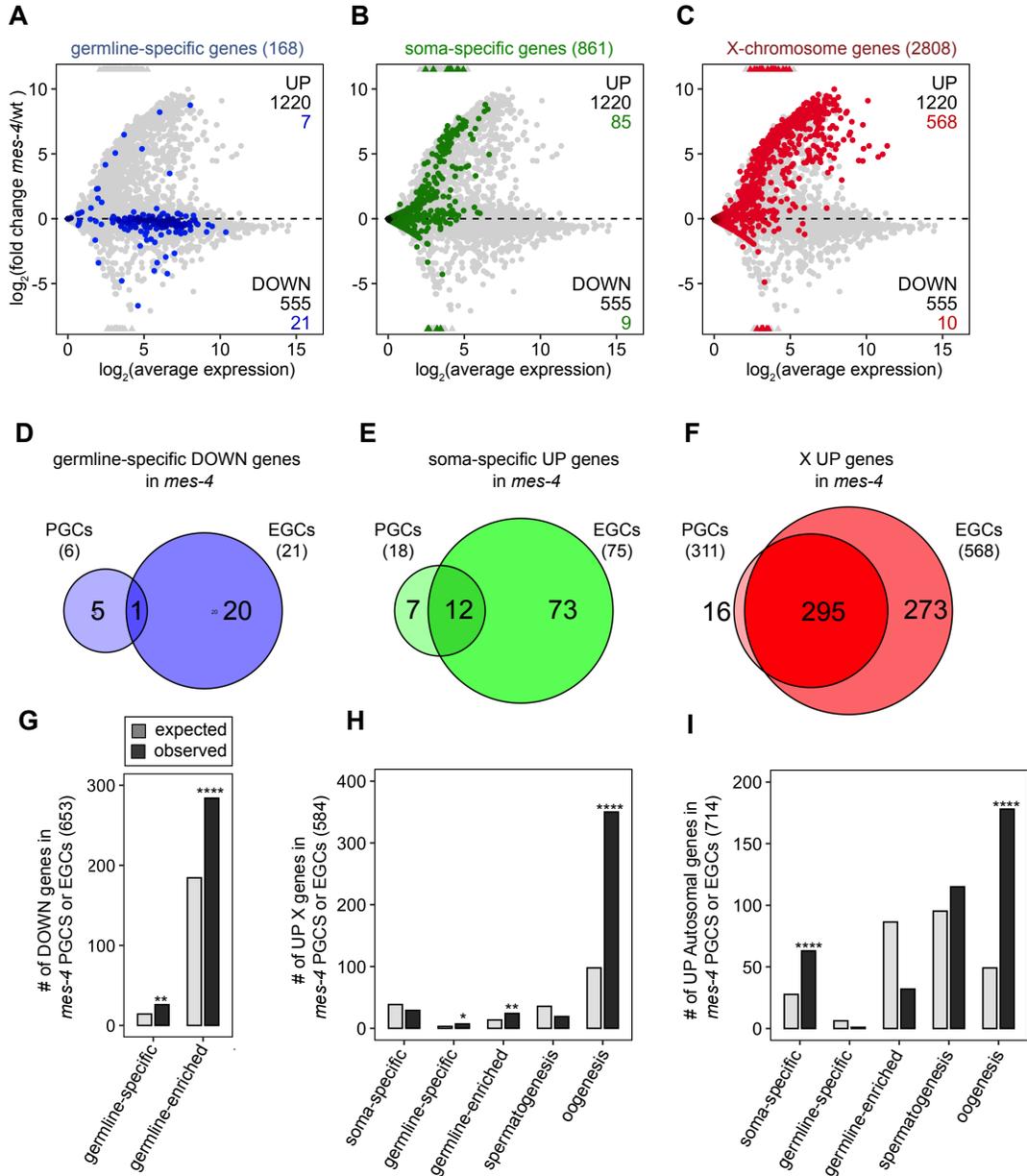pink and DOWN genes are in teal. (C,D) Boxplots showing distributions of transcript abundance levels for autosomal (left) or X- linked (right) expressed genes (gray), UP genes (pink), and DOWN genes (teal) in (C) wt PGCs and (D) wt EGCs. The horizontal line represents the median, the box represents the 25th and 75th quartiles, and the whiskers represent the 5th and 95th quantiles. (E) Line plots showing $\log_2$(average expression) across 10 regions of a wild-type adult hermaphrodite gonad that capture different stages of germline development (data from Tzur et al. 2018, Tables S2A and S10A). We separately analyzed genes on autosomes (left) and on the X (right). Average expression values were calculated in Tzur et al. by normalizing each gene's read counts to the total number of read counts in a sample, multiplying those normalized values by 10^5, and finally averaging across 2 independent samples. Each dot represents the mean expression of a gene set (colors) in 1 of the 10 germline regions. Whiskers correspond to 95% confidence intervals of the mean. Gene set sizes are indicated. We did not analyze the set of DOWN X genes due to its small size (13 genes). (F,G,H) Bar plots showing significantly enriched (Benjamini Hochberg-adjusted P-value < 0.05) gene ontology (GO) biological process terms in sets of mis-regulated genes (*mes-4* PGCs or EGCs): (F) UP X genes, (G) UP autosomal genes, and (H) DOWN genes. Numbers of mis-regulated genes for each GO term are indicated. Analyses of UP genes (F and G) used all 20,258 protein coding genes as a background, while analysis of DOWN genes (H) used our set of 6,345 protein coding genes expressed in wt PGCs or EGCs as a background. All GO analyses were performed using DAVID Bioinformatics Resource 6.8.
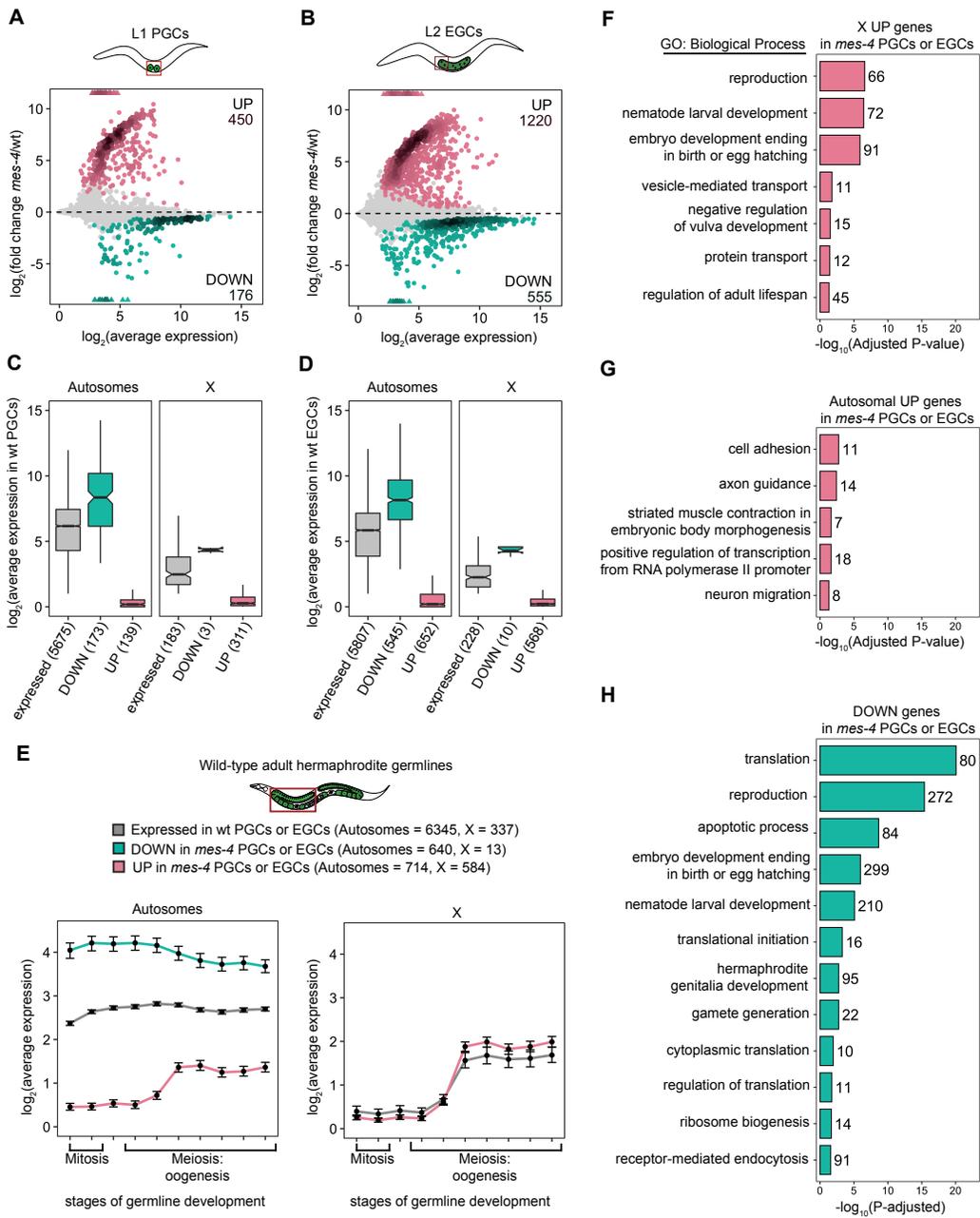
transcript	gene category	smFISH in L1 PGCs		RNA-seq in L1 PGCs	
		$\log_2$(fold change)	P-value Mann Whitney test	$\log_2$(fold change)	Adjusted P-value Wald test
cpg-2	germline-specific	-2.72	1.19e-6	-3.97	1.43e-4
pgl-3	germline-specific	-0.199	0.328	-0.213	0.634
chs-1	germline-enriched	-0.150	0.277	-0.131	0.758
lsd-1	X-linked	4.80	4.57e-5	8.06	1.26e-16
mbk-1	X-linked	2.75	1.96e-6	9.23	3.26e-31
pek-1	X-linked	2.32	3.27e-6	8.55	3.47e-26
lin-15B	X-linked	1.68	4.16e-6	1.22	0.1

Figure 3-8. Comparison of smFISH and transcript profiling data. Table showing differential expression statistics for the 7 genes tested by smFISH. For our smFISH analysis, $\log_2$(fold change) values were calculated by comparing mean transcript counts in *mes-4* PGCs to wt PGCs, and the P-value is from a Mann Whitney test. For our transcript profiling analysis, 'shrunken' $\log2_{(}$fold change) values were calculated using DESeq2 and ashr in R.
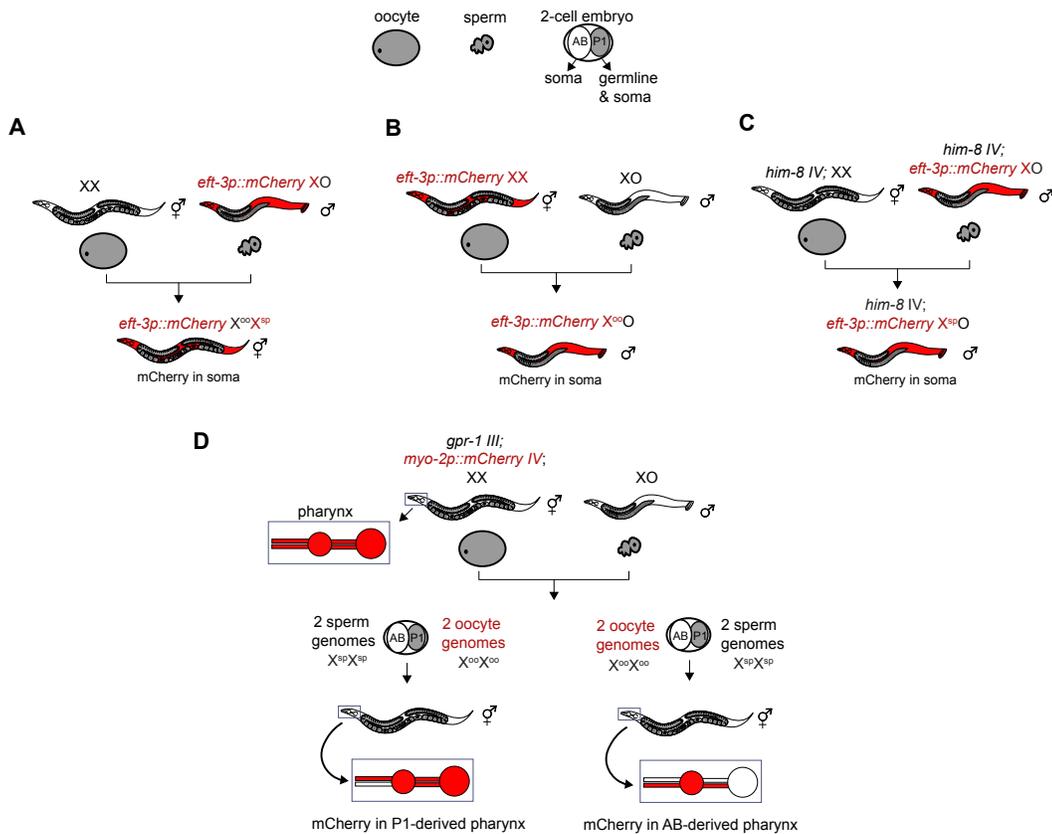
Figure 3-9. Genetic strategies to generate and identify F1 offspring that inherited different X-chromosome endowments from parents. (A-C) An X-linked and soma-expressed mCherry driven by the *eft-3* promoter was used to track X-chromosome inheritance patterns in F1 offspring. (C) To generate males that inherited their single X from the sperm, we used the *him-8(e1489)* allele to cause X chromosome nondisjunction in the maternal germline, which in turn causes some oocytes to lack an X. (D) To generate hermaphrodites whose germline inherits either 2 genomes from the oocyte or 2 genomes from the sperm, we used a *gpr-1* over-expression allele. Expression of a mCherry marker driven by the *myo-2* promoter in AB-derived pharyngeal muscle or P1-derived pharyngeal muscle was used to identify F1 hermaphrodite offspring with non-mendelian inheritance patterns.
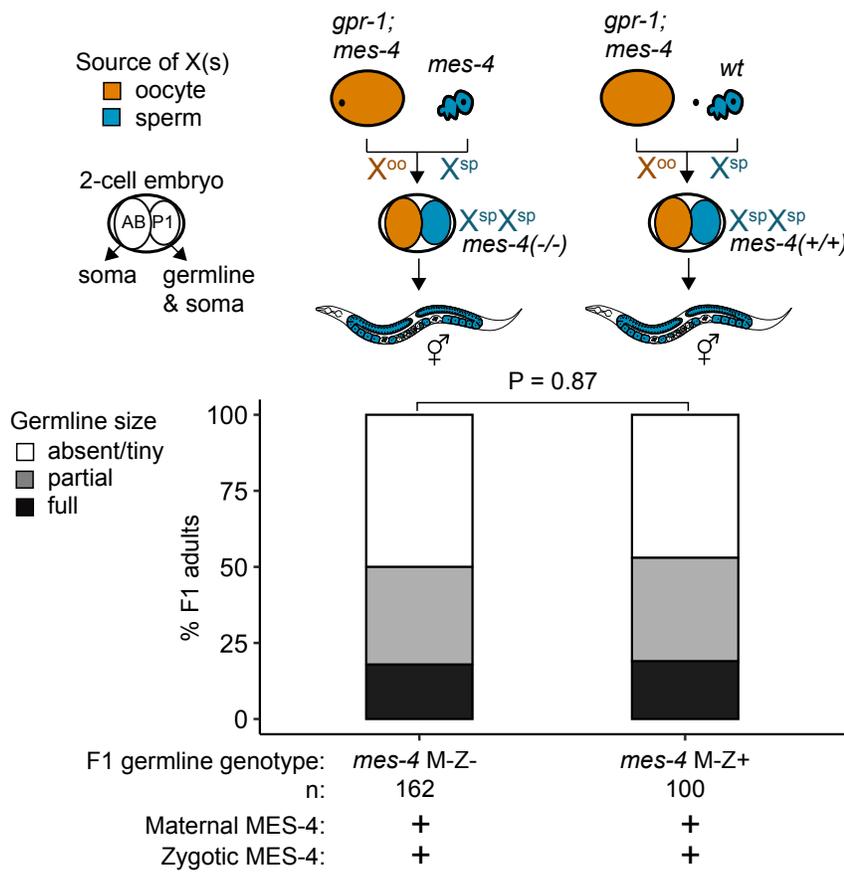
Figure 3-10: *mes-4* M-Z+ X^sp^/X^sp^ mutants do not have healthier germlines than *mes-4* M-Z- X^sp^/X^sp^ mutants. Bar plot (as in Figure 3-3A) showing distributions of germline size ('absent/tiny', 'partial', and 'full') in F1 offspring whose germline inherited 2 X chromosomes from the sperm. Numbers of scored F1 offspring, the genotype of their germlines, and the presence or absence of maternal MES-4 or zygotic MES-4 are indicated below the bars. Notably, the germlines of *mes-4* M-Z+ offspring have 2 wild-type copies of *mes-4* that they inherit from the sperm.
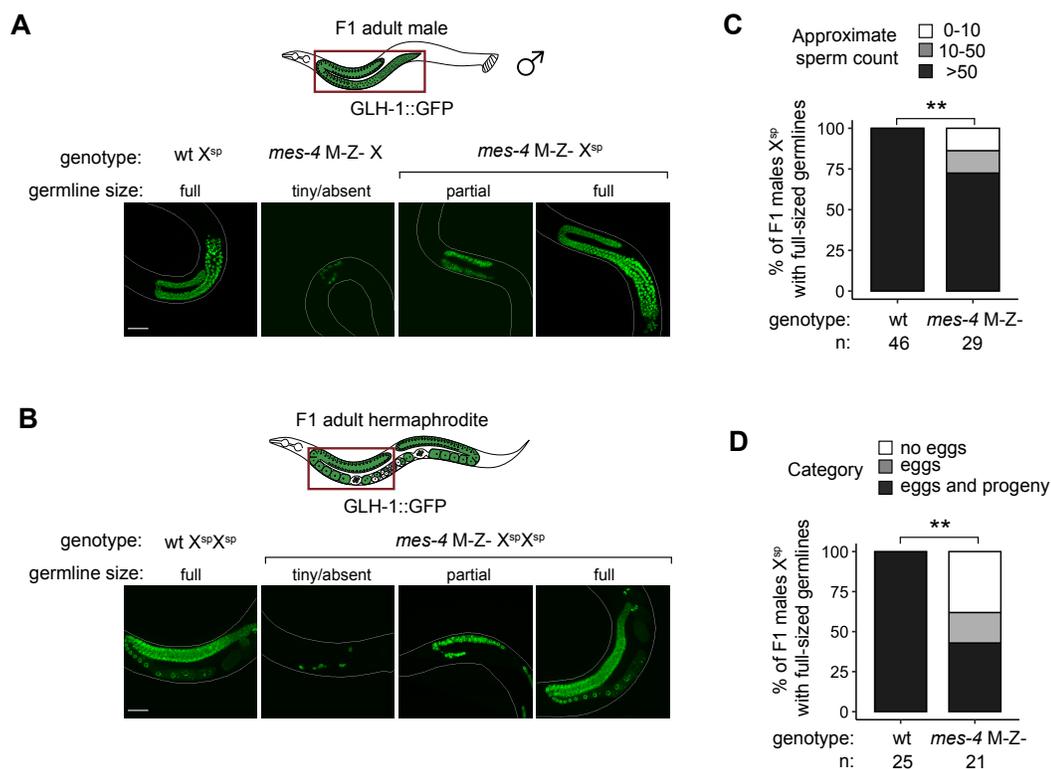
Figure 3-11: Further fertility analyses of *mes-4* mutants that inherit their X's from the sperm. (AB) Representative images of live F1 offspring with a germline scored as either 'tiny/absent', 'partial', or 'full'. Genotypes of imaged worms with respect to the *mes-4* locus and their germline's X-chromosome composition are indicated. Green signal is the germline marker GLH-1::GFP. Worm bodies are outlined in white. Scale bar is 45 µM. (A) F1 male offspring generated using the *him-8(e1489)* allele. All imaged males inherited their single X from the sperm, except the male with a 'tiny/absent' germline; that male inherited its X either from the mother's oocyte or mother's sperm (self-fertilization) (B) F1 hermaphrodite offspring generated using the *gpr-1* allele and form a germline composed of 2 genomes from the sperm. (C) Bar plots showing distributions of sperm count in F1 male offspring across 3 categories: '0-10 sperm', '10-50' sperm, and '>50 sperm'. Only F1 males that were classified as having full germlines were analyzed. We compared proportions of F1 males that have >50 sperm between wt and *mes-4* mutant populations using a 2-sided Fisher's exact test. P-value designation is ** < 0.001. (D) Bar plots showing the proportion of F1 hermaphrodites that have 'no eggs', 'eggs' (but do not produce viable progeny), and 'eggs and progeny'. Only F1 hermaphrodite offspring that were classified as having full germlines were analyzed. We compared proportions of F1 hermaphrodites that have 'eggs and progeny' between wt and *mes-4* mutant populations using a 2-sided Fisher's exact test.

Figure 3-12: Identification and testing of candidate transcription factors for a role in causing Mes-4 sterility. (A) Venn diagram comparing 3 gene sets: X UP genes in *mes-4* PGCs or EGCs, X genes with a LIN-15B ChIP peak in their promoter region (500 bp upstream of TSS), and X genes with a LIN-15B DNA motif in their promoter region. We used LIN-15B ChIP-chip data from embryos (downloaded from modENCODE) and LIN-15B ChIP-seq data from L3 larvae and L4 larvae (downloaded from modERN). (B) Table showing the candidate transcription factors that we tested for a role in causing Mes-4 sterility (columns) and the criteria used to select them (rows). Gray-colored squares indicate a criterion was met, white-colored squares indicate a criterion was not met, and an 'X' indicates a criterion could not be tested (e.g. RNAi depletion caused highly penetrant lethality).

Figure 3-13: Further analysis of how removing LIN-15B impacts the transcriptome of *mes-4* M+Z- dissected adult germlines. (AB) Volcano plots as described in Figure 3-4A showing differential expression analysis for (A) *mes-4* M+Z-; *lin-15B* M-Z- vs wt adult germlines and for (B) *mes-4* M+Z-; *lin-15B* M-Z- vs *mes-4* M+Z- germlines. X genes are colored red. (C) Scatterplot comparing log2(fold change) in *mes-4* M+Z- vs wt adult germlines to $\log_2$(fold change) in *mes-4* M+Z-; *lin-15B* M-Z- vs *mes-4* M+Z- adult germlines. Only genes that are either UP or DOWN (q-value < 0.05) in *mes-4* M+Z- vs wt adult germlines are shown. X genes are colored red. The Spearman's correlation coefficient is indicated for all plotted genes (black) and for plotted X genes (red). (D) Boxplots as described in Figure 3-4B for autosomal UP genes in *mes-4* M+Z- vs wt (left) and for DOWN genes in *mes-4* M+Z- vs wt (right). (E) Venn diagrams as described in Figure 3-4C for the indicated gene sets.
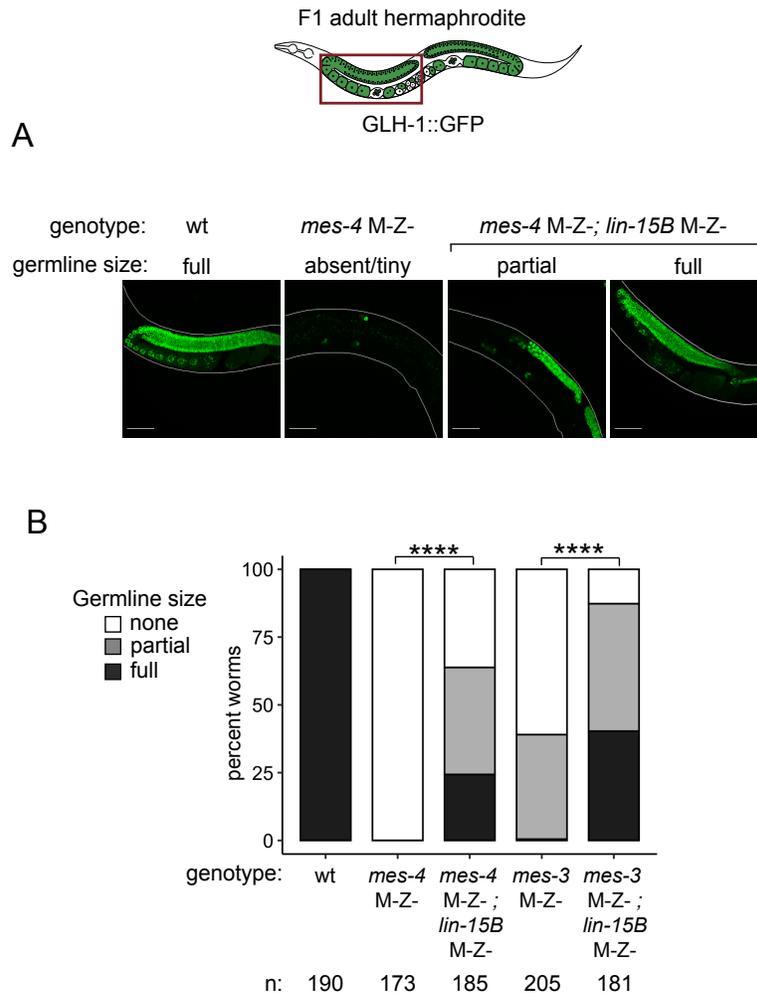
Figure 3-14: Removal of LIN-15B improves germline health in *mes-4* M-Z-X⁰⁰/Xˢᵖ mutant hermaphrodites. (A) Representative images of live F1 adult hermaphrodite offspring with a germline scored as either 'tiny/absent', 'partial', or 'full'. Genotypes of imaged worms with respect to the *mes-4* and *lin-15B* loci are indicated. All F1s inherited 1 X from the oocyte and 1 X from the sperm. Green signal is the germline marker GLH-1::GFP. Worm bodies are outlined in white. Scale bar is 45 µM. (B) Bar plot as described in Figure 3-4D showing distributions of germline size in populations of F1 offspring. Genotypes and sample sizes are indicated below the plot. We compared proportions of F1s with full-sized germlines using 2-sided Fisher's exact tests. P-value designation is **** < 1e-5.
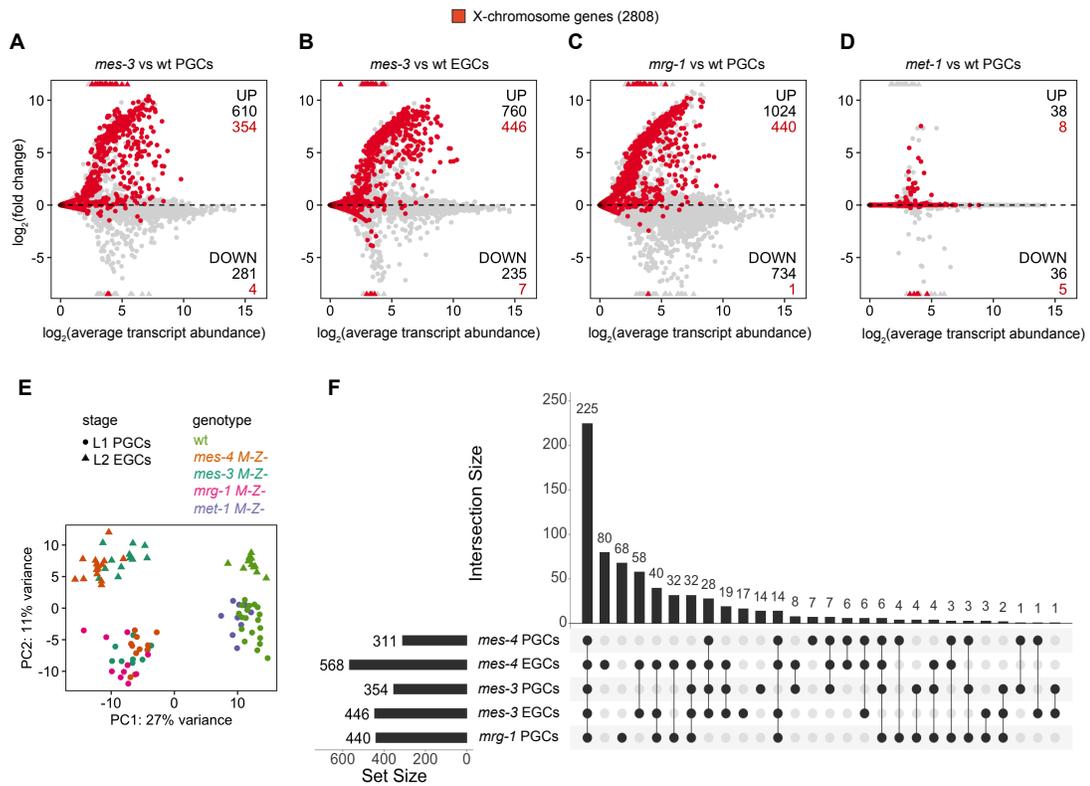
Figure 3-15: Comparison of X mis-expression in PGCs and EGCs dissected from various chromatin regulator mutants. (A-D) MA plots as described in Figure 3-1D showing differential expression analysis for (A) *mes-3* M-Z- vs wt PGCs, (B) *mes-3* M-Z- vs wt EGCs, (C) *mrg-1* M-Z- vs wt PGCs, and (D) *met-1 M-Z-* vs wt PGCs. X genes are colored red. (E) PCA plot showing all analyzed transcript profiles across the first 2 principal components. We used variance-stabilized counts computed by the 'vst' function in DESeq2 to perform PCA. Stages and genotypes of worms are indicated by shape and color, respectively. (F) Upset plot generated using the UpsetR R package comparing sets of UP X (mutant vs wt) genes across *mes-4* PGCs, *mes-4* EGCs, *mes-3* PGCs, *mes-3* EGCs, and *mrg-1* PGCs.

APPENDIX 1

Analysis of the role of 22Gs in regulating

the transcriptome of PGCs

INTRODUCTION

In Chapter 3, I found that the maternally loaded chromatin regulator MES-4 does not specify the germline in *C. elegans*. Other strong candidates to specify the germline were maternally inherited P granules, which segregate exclusively to the germline lineage during embryogenesis. While germ granules are germline determinants in *Drosophila*, they do not specify the germline in *C. elegans* (Gallo et al. 2010; Updike et al. 2014; Knutson et al. 2017). Remaining candidates to specify the germline are CSR-1-associated 22G small RNAs (22 nucleotides long and have a strong bias for a guanosine at the 5' end) that are a type of endo-siRNA and are loaded onto the CSR-1 argonaute (CSR-1:22Gs) (Claycomb et al. 2009).

CSR-1:22Gs have several features that make them promising candidates to turn on a germline program in PGCs. They are abundant in the germline, antisense to most germline-expressed transcripts, maternally loaded into the embryo, and have been shown to promote transcription of some gene targets, such as spermatogenesis genes in males and histone genes (Claycomb et al. 2009; Conine et al. 2010; Avgousti et al. 2012; Seth et al. 2013; Wedeles et al. 2013; Cecere et al. 2014). CSR-1 appears to have an essential role in germline development, since almost all *csr-1* M+Z-mutants are sterile. Rarely, *csr-1* M-Z- mutant embryos are produced but all die due to chromosome segregation defects (Yigit et al. 2006; Claycomb et al. 2009; Gerson-

Gerwitz et al. 2015). CSR-1:22Gs are thought to bind and 'license' genes for transcription by protecting them from piRNA-mediated silencing (Seth et al. 2013; Wedeles et al. 2013). In vivo tethering and genetic experiments support this view: 1) tethering CSR-1 to a gene prevents its silencing by the piRNA pathway (Wedeles et al. 2013), 2) tethering CSR-1 to a previously silent gene can activate its expression (Wedeles et al. 2013), and 3) CSR-1 and presumably its associated 22Gs are required for trans-activation of a previously silent gene (Seth et al. 2013). Thus, CSR-1:22Gs may collectively form an epigenetic memory of germline that is transmitted from parent to offspring and is necessary for offspring PGCs to launch a germline program. However, recent studies have challenged the view that CSR-1 promotes transcription of its gene targets (Gerson-Gerwitz et al. 2015; Campbell and Updike et al. 2015). Here, I aimed to clarify the role of CSR-1 in regulating gene expression in PGCs.

The WAGO (worm-specific argonaute) pathway is the other 22G pathway in the germline. WAGO-associated 22Gs are secondary siRNAs that are produced by primary siRNA triggers (e.g. piRNAs) and silence transcription of germline-inappropriate gene targets (e.g. transposons and other repetitive elements). (Gu et al. 2009; Billi et al. 2014).

To investigate the roles of the CSR-1:22G and WAGO:22G pathways in regulating proper gene expression patterns in PGCs, I profiled mRNAs from PGCs that were depleted of most 22G RNAs. Biogenesis of 22G RNAs in the germline requires the RNA helicase DRH-3 (Gu et al. 2009, Billi et al. 2014). *drh-3* null mutants are sterile (Duchaine et al. 2006) Animals carrying a *drh-3(ne4253)*

hypomorphic mutation are depleted of most 22G RNAs yet are fertile at permissive temperature (20°C) (Gu et al. 2009). For this project, I used CRISPR/Cas9 to generate the *drh-3(bn197)* hypomorphic allele that is linked to the germline-specific marker glh-1::GFP (for transcript profiling) and that produces the same mutant protein as *drh-3(ne4253)* (Figure 4-1A).

RESULTS

Germlines dissected from *drh-3(bn197)* mutant adults are slightly depleted of CSR-1-associated 22Gs and very depleted of WAGO-associated 22Gs

I first examined whether the *drh-3(bn197)* allele that I generated depletes 22G RNAs in the germline. I profiled 16-30 nucleotide-long small RNAs from dissected wild-type (wt) and *drh-3(bn197)* adult germlines. I found that 22G RNA levels were dramatically reduced in *drh-3(bn197)* vs wt germlines (Figure 4-1B). Henceforth, I refer to the *drh-3(bn197)* allele as *drh-3*. To identify gene targets that have significantly reduced antisense 22Gs in *drh-3* vs wt germlines, I performed differential expression analysis using DESeq2 and set a significance threshold of q-value < 0.05. I found >5000 gene targets that have significantly reduced antisense 22G RNAs in *drh-3* germlines ('DOWN' genes), the vast majority of which are protein-coding genes (Figure 4-1C-D). Next, I examined CSR-1:22G gene targets and WAGO:22G gene targets. I found that CSR-1:22G gene targets have slightly lower levels of antisense 22Gs and WAGO:22G gene targets have dramatically lower levels

119

of antisense 22Gs in *drh-3* vs wt germlines (Figure 4-1E). Furthermore, 1563 of 4153 (38%) CSR-1:22G targets and 994 of 1071 (93%) WAGO:22G targets have significantly lower levels of antisense 22Gs in *drh-3* vs wt germlines (Figure 4-1F). I conclude that *drh-3* mutant adult germlines have lower levels of antisense 22G RNAs than wt adult germlines and that WAGO:22G targets are more impacted than CSR-1:22G targets.

PGCs dissected from *drh-3(bn197)* L1 larvae turn on a germline program

The key question of this project is whether CSR-1:22Gs are required for PGCs to launch a germline program. I profiled mRNAs from wild-type and *drh-3* mutant PGCs dissected from recently hatched L1 larvae and performed differential expression analysis to identify mRNAs that had significantly higher or lower abundance in *drh-3* vs wt PGCs. I tested whether *drh-3* PGCs fail to turn on a 'core' germline program by examining 2 germline gene sets: 1) 168 germline-specific genes, which are expressed in the germline and not expressed in soma, and 2) 2111 germline-enriched genes, which are expressed more highly in the germline compared to soma. I found that almost all (166/168) germline-specific genes were turned on normally in *drh-3* PGCs, and that neither germline set had more down-regulated genes than expected by chance (Figure 4-2A and D). Next, I tested whether the 1539 DOWN CSR-1:22G gene targets (from Figure 4-1E) had lower levels of mRNA in *drh-3* vs wt PGCs. I found that 24 of the 1539 DOWN CSR-1:22G gene targets had a lower level of mRNA in *drh-3* vs wt PGCs, but the vast majority were turned on

normally (Figure 4-2B and E). I conclude that partial depletion of CSR-1:22Gs in *drh-3(bn197)* mutant PGCs does not prevent those PGCs from turning on a germline program.

I performed the same analysis for the 945 DOWN WAGO:22G targets and found that most were not mis-regulated. However, I observed that a dramatically higher number of DOWN WAGO:22Gs had higher levels of mRNA in *drh-3* PGCs than expected by chance (Figure 4-2C and E). This suggests that the WAGO:22G pathway is required to repress some of its targets in PGCs. I tested whether *drh-3* PGCs mis-express a gene expression program that is inappropriate for PGCs by examining 3 gene sets: 1) soma-specific genes, 2) oogenesis genes, and 3) spermatogenesis genes (see Methods for details). Interestingly, I observed more up-regulated genes than expected by chance for all 3 gene sets (Figure 4-2E). In chapter 3, I found that MES-4 protects germline immortality by repressing genes on the X chromosome. This prompted me to test whether the mis-regulated genes in *drh-3* PGCs are enriched on the X chromosome. I found that not to be the case (Figure 4-2F). I conclude that depletion of 22Gs in *drh-3(bn197)* mutant PGCs causes some targets of the WAGO:22G pathway and several PGC-inappropriate genes to aberrantly turn on.

CONCLUSION

A burning question is what specifies the germline in *C. elegans*. In chapter 3, I ruled out the MES chromatin regulators. In this study, I provide evidence that

suggests that CSR-1:22Gs don't do it either. Importantly, the *drh-3* hypomorph generated and used in this study does not fully deplete CSR-1-associated 22Gs. Thus, CSR-1:22Gs may be required to turn on more germline genes than identified in this study. Interestingly, loss of WAGO:22Gs correlates with aberrant turn-on of many of their gene targets in *drh-3* PGCs. Moreover, several classes of PGC-inappropriate genes are mis-expressed in *drh-3* PGCs, namely soma-specific genes, oogenesis genes, and spermatogenesis genes, Notably, it was shown that CSR-1 represses spermatogenesis genes in dissected adult hermaphrodite germlines (Campbell and Updike 2015). My findings indicate that CSR-1:22Gs have a similar role in hermaphrodite PGCs. Overall, this work suggests that the primary role of 22G RNAs in regulating the transcriptome of PGCs is to repress genes that would interfere with a PGC program. Perhaps, this is how 22G RNAs ensure proper germline development.

METHODS

Worm strains

All worms were maintained at 20˚C on 6-cm plates containing Nematode Growth Medium (NGM) spotted with *E. coli* OP50 *(*Brenner, 1974). Strains used and generated (*) in this study are listed below.

DUP64 - *glh-1(sams24[glh-1::GFP::3xFLAG]) I*
SS1451* - *glh-1(sams24[glh-1::GFP::3xFLAG]) drh-3(bn197) I*

Creation of a *drh-3* hypomorphic alleles linked to glh-1::GFP by CRISPR-Cas9

See Figure 4-1A for details. The *drh-3(bn197)* allele encodes the same peptide as the *drh-3(ne4253)* allele, which was previously shown to deplete antisense-22G RNAs (Gu et al. 2009). We also made a few silent mutations to create a BseR1 restriction site for genotyping. An Alt-R crRNA oligo (IDT) were designed using CRISPOR (Concordet and Haeussler, 2018) and the UCSC Genome Browser (ce10) to produce highly efficient and specific Cas9 cleavage in *drh-3*'s HELICc domain near the intended edit site. An Ultramer ssDNA oligos (IDT) containing 50 bp micro-homology arms was used as a repair template. We used a *dpy-10* co-CRISPR strategy (Arribere et al. 2014) to isolate strains carrying our desired mutations. Briefly, 2.0 uL of 100 μM *drh-3* 0.5 uL of 100 μM *dpy-10* cRNA were annealed to 2.5 uL of 100 μM tracrRNA (IDT) by incubating at 95˚C for 2 minutes, then room temperature for 5 minutes, to produce sgRNAs. We complexed sgRNAs with 5 uL of 40 μM Cas9 protein at room temperature for 5 minutes, added 1 uL of 40 μM *drh-3* repair template and 1 uL of 40 μM *dpy-10(cn64)* repair template, and centrifuged the mix at 13,000g for 10 minutes. All RNA oligos were resuspended in IDT's duplexing buffer (cat). Mixes were injected into 1 or both gonad arms of ~30 DUP64 adults. Transformant progeny were isolated and back-crossed 4x to DUP64.

mRNA-sequencing from single sets of sister PGCs

To obtain near-synchronous animals for dissection, we hatched larvae within a 30-minute window in the absence of food. We then allowed those newly hatched L1

larvae to feed for 30 minutes to start PGC development. Larvae were partially immobilized in 15 uL drops of egg buffer (25 mM HEPES, pH 7.5, 118 mM NaCl, 48 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, adjusted to 340 mOsm) on poly-lysine coated microscope slides and hand-dissected using 30-gauge needles to release their gonad primordium (consisting of 2 connected sister PGCs and 2 somatic gonad precursors). Isolation of single sets of sister PGCs from somatic gonad precursors and into tubes for transcript profiling involved fluorescence microscopy to identify PGCs by germline-specific expression of GLH-1::GFP and mouth pipetting using pulled glass capillaries coated with Sigmacote (Sigma SL2) and 1% BSA in egg buffer. 7.5 mg/mL pronase (Sigma P8811) and 5 mM EDTA were added to reduce the adherence of gonad primordia to the poly-lysine coated slides and to weaken cell-cell interactions. While monitoring by microscopy, gonad primordia were moved to a drop of fresh egg buffer and PGCs were separated from gonad precursors using shearing force generated by mouth pipetting. Finally, we transferred single sets of sister PGCs into 0.5 uL drops of egg buffer placed inside the caps of 0.5 mL low-bind tubes (USA Scientific, #1405-2600). We only selected PGCs for transcript profiling if they maintained bright fluorescence of GFP throughout isolation and were clearly separated from somatic gonad precursors. Tubes containing single sets of 2 sister PGCs were quickly centrifuged, flash frozen in liquid nitrogen, and then stored at -80˚C. At least 11 samples (replicates) of PGCs were isolated for each condition.

Immediately after thawing PGCs on ice, a 1:4,000,000 dilution of ERCC spike-in transcripts (Life Technologies, #4456740) were added to each sample.

124

Double-stranded cDNAs from polyA(+) RNAs were generated using a SMART-seq

method that combined parts of the SMART-seq2 (Picelli et al. 2014) and SMART-

seq v4 (Takara) protocols. Briefly, PGCs were lysed at room temperature for 5

minutes in a lysis buffer containing RNAse inhibitors (Takara, #2313) to release

mRNAs into solution. 1.2 µM custom DNA primer (IDT) was annealed to transcripts'

polyA tails by incubating samples at 72˚C for 3 minutes and then immediately

placing them on ice. Reverse transcription to generate double-stranded cDNA was

performed using 200 U SmartScribe, 1x first-strand buffer, 2 mM DTT (Takara,

#639537), 1 mM dNTPs (Takara, #639125), 4 mM $MgCl_2$, 1 M betaine (Sigma,

#B0300), 20 U RNAse Inhibitor (Takara), and 1.2 µM custom template-switching

oligo with a Locked Nucleic Acid analog (Qiagen) at 42˚C for 90 minutes, followed

by 70˚C for 15 minutes to heat-inactivate the reverse transcriptase. cDNAs were

PCR-amplified according to Takara's SMART-seq v4 protocol for 20 cycles using

SeqAmp DNA Polymerase (Takara, #638504) and a custom PCR primer. Amplified

cDNAs were purified by SPRI using 1x Ampure XP beads (Agencourt, #A63881)

and quantified using a Qubit fluorometer. All custom oligos contained a biotin group

on their 5' end to ameliorate oligo concatemerization. Illumina's Nextera XT kit

(Illumina, #FC-131-1096) was used with 350-400 pg cDNA as input to prepare dual-

indexed Illumina RNA-sequencing libraries according to the manufacturer's

instructions. Libraries were PCR-amplified for 14 cycles and purified by SPRI using

0.6x Ampure XP beads. Library concentration was measured by a Qubit fluorometer,

and average fragment size was measured by either a bioanalyzer or a tapestation

(Agilent). Libraries were multiplexed and paired-end sequenced with 50 cycles on either an Illumina HiSeq2500 or NovaSeq 6000 SP flowcell at the QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

small RNA-sequencing from dissected adult germlines

We dissected 1st day hermaphrodite (approximately 20-24 hours post-mid-L4 stage) with 30-gauge needles in egg buffer (see recipe above, except not adjusted to 340 mOsm) containing 0.1% Tween and 1 mM levamisole to extrude their gonads. Gonads were cut at the narrow 'bend' to separate the gonad region containing mitotic and early meiotic germ cells from the region that contains oocytes and/or sperm; the former was used for RNA profiling. 200 gonads were mouth pipetted into 500 uL Trizol reagent (Life Technologies, #15596018), flash-frozen in liquid nitrogen, and stored at -80˚C for up to 1 month before RNA extraction.

To release RNAs from gonads in Trizol, we performed 3 freeze-thaw cycles using liquid nitrogen and a 37˚C water bath, while vortexing vigorously between each cycle. RNAs immersed in Trizol were added to phase-lock heavy gel tubes (need vendor) and mixed with 100 uL of 1-bromo-3-chloropropane (BCP) (Sigma #B9673) by handshaking, followed by room temperature incubation for 10 minutes. Samples were then centrifuged at 13,000g and 4˚C for 15 minutes to separate phases. RNAs in the aqueous phase were precipitated by mixing well with 0.7x-0.8x volumes of ice-cold isopropanol and 1 uL 20 mg/mL glycogen, followed by incubation at -80˚C for 1-2 hours and centrifugation at 13,000g at 4˚C for 30 minutes. RNA pellets were

washed 3x with ice-cold 75% ethanol and then resuspended in 15 uL water. RNA concentration was determined by a Qubit fluorometer, purity was determined by a Nanodrop, and integrity (RIN values) was determined by a bioanalyzer or tapestation.

We size-selected 16-30 nucleotide-long small RNAs by PAGE. Since 22G RNAs have a 5' triphosphate group that prevents adapter ligation for library preparation, we treated 1 ug of total RNA with polyphosphatase (Epicentre #RP8092H) according to the manufacturer's instructions. RNAs were boiled in formamide buffer and then run on denaturing 15% Urea/TBE polyacrylamide gels (Biorad, Mini-Protean #4566053). Gels were stained with SYBR gold to visualize RNAs under UV light. Gel slices containing 16-30 nucleotide-long RNAs were excised from gels, then crushed by centrifugation through small pores cut into walls of eppendorf tubes, and then incubated at 4°C overnight in 300 mM NaCl and 1 mM EDTA buffer to extract RNAs (see Silas et al. 2018). RNAs were precipitated using 0.7x-0.8x isopropanol and glycogen as a carrier. RNA pellets were washed 3x with 75% ice-cold ethanol and then resuspended in water.

cDNA libraries were prepared from purified 16-30 nucleotide-long small RNAs using the NEBNext small RNA library prep set for Illumina (#E7330S) according to manufacturer's instructions for 100 ng total RNA input. 16 PCR cycles were used to amplify libraries. Libraries were purified using NEB's Monarch PCR & DNA kit (#T1030), then size-selected by SPRI using 3.7x Ampure XP beads (Agencourt, #A63881), and finally quantified using a Qubit fluorometer and tapestation (Agilent). Libraries were multiplexed and single-end sequenced with 50

127

cycles on an Illumina HiSeq4000 flowcell at the QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Processing and analysis of mRNA-seq data

Paired-end Illumina sequencing reads were aligned to the ce10 (WS220) genome downloaded from Ensembl using Hisat2 (v2.2.1). Samtools (v1.10) was then used to remove duplicate reads and reads with low mapping quality (MAPQ < 10) from the alignment file. The function featureCounts from the subread package (v2.0.1) (Liao et al. 2014) was used to obtain gene-level read (fragment) counts using a ce10 (WS220) transcriptome annotation file (Ensembl), which additionally contained 92 ERCC spike-in transcripts. 11 low-quality transcript profiles that were likely caused by a failure to capture mRNAs from 2 PGCs, a well-known problem in single-cell RNA-sequencing, were identified using the R package scuttle (v1.2.0) (McCarthy et al. 2017). The R package DESeq2 (v1.32.0) (Love et al. 2014) was used to identify differentially expressed genes in mutant vs wild-type samples. P-values were adjusted for multiple hypothesis testing by the Benjamini-Hochberg method to produce q-values. Genes with a q-value < 0.05 were considered differentially expressed. The scaling factors used by DESeq2 to normalize transcript profiles were calculated using the R package scran (v1.20.1), which uses a pooling and deconvolution approach to deal with zero-inflation in low-input RNA-seq data (Lun et al. 2016).

Processing and analysis of small RNA-seq data

Single-end Illumina sequencing reads were trimmed and filtered to retain those that are 16-30 nucleotides long Cutadapt. Reads were then mapped to the ce11 (WBcel235) genome using Bowtie1 and the following parameters: -v 1 -S -M 1 --best --strata --no-unal --tryhard. Samtools (v1.10) was used to sort reads in the alignment file. Sense reads that correspond to structural RNAs (rRNA, snoRNA, tRNA, snRNA) were filtered from the file containing mapped reads using bedtools (Quinlan and Hall, 2010). For the analysis of small RNA lengths in Figure 4-1B, reads were parsed by their length and the identity of their 5'-most base using bioawk. All small RNA reads and the subset of antisense 22G small RNA reads were counted using featureCounts from the subread package (Liao et al. 2014). The R package DESeq2 (v1.32.0) (Love et al. 2014) was first used to calculate scaling factors for sample normalization using gene-level counts from all small RNA reads. Those scaling factors were then applied to differential expression analysis (Wald tests in DESeq2) using gene-level counts from just antisense-22G RNAs. This differential expression analysis identified gene targets that have significantly higher or lower levels of antisense 22G RNAs in *drh-3* vs wt germlines. P-values were adjusted for multiple hypothesis testing by the Benjamini-Hochberg method to produce q-values. Genes with a q-value < 0.05 were considered differentially expressed. All visualizations of RNA-seq data were generated by the R packages 'ggplot2' (v3.3.5) and ggpubr (v0.4.0).

Gene sets

The germline-specific, germline-enriched, and soma-specific gene sets were previously defined based on microarray and SAGE data. The germline-specific and germline-enriched gene sets were previously defined in Rechtsteiner et al. 2010 and Reinke et al. 2004, respectively. The soma-specific set was previously defined in Knutson and Egelhofer et al. 2017 and then refined in chapter 3 of this thesis by removing genes from the set that have >5 TPM in RNA-seq data from wild-type dissected adult germlines. The oogenesis and spermatogenesis were previously defined in Ortiz et al. 2014 (they called 'oogenic' and 'spermatogenic'). Oogenesis genes are expressed at higher levels in dissected adult oogenic germlines than dissected adult spermatogenic germlines; spermatogenesis genes are the opposite. Gene targets of CSR-1:22G RNAs were previously defined in Claycomb et al. 2009 as genes that are antisense to 22G RNAs that co-immunoprecipitated with CSR-1 (Claycomb et al. 2009). Gene targets of WAGO:22G RNAs were previously defined in Gu et al. 2009 as genes that were depleted of 22G RNAs in *rde-3*, *mut-7*, and MAGO12 (1 strain with null alleles for 12 WAGOs) mutants.

Statistical analysis

Gene set enrichment analyses were performed using hypergeometric tests in R. The sizes of gene sets used for those tests are noted in the respective figure legends. P-value designations are * < 0.01, ** < .001, *** < 1e-4, and **** < 1e-5. In our small RNA-seq and mRNA-seq analysis, differentially expressed genes were

identified using Wald tests in the R package DESeq2 (see above) as those with a q-value < 0.05.

Figure 4-1 (on next page). Dissected germlines from *drh-3(bn197)* adult hermaphrodites have reduced abundance of antisense 22Gs for some CSR-1-class and many WAGO-class gene targets. (A) Diagram showing DNA and peptide sequences for the wild-type *drh-3* allele and the mutant hypomorphic *drh-3(bn197)* allele at the site of the *bn197* T834M point mutation in the HELICc (helicase) domain. The *drh-3(bn197)* allele encodes the same peptide as the *drh-3(ne4253)* allele, which was previously shown to deplete antisense-22G RNAs (Gu et al. 2009). The *drh-3(bn197)* allele was generated by CRISPR/Cas9 and contains a BseRI restriction site for genotyping. B) Bar plots showing the distribution of small RNA (sRNA) lengths in wild-type dissected adult gonads (top) and *drh-3(bn197)* dissected adult gonads (bottom). The colors represent the identity of the first nucleotide in the sRNA. C) MA plot showing $\log_2$(average expression) versus $\log_2$(fold change) of antisense-22G RNA abundance for gene transcripts (circles) in *drh-3(bn197)* vs wt gonads. Targets with mis-regulated levels of antisense 22Gs in *drh-3* vs wt germlines (red) were identified using Wald tests in DESeq2 (Love et al. 2014) and by setting a q-value < 0.05 significance threshold. The numbers of mis-regulated targets are indicated in the corners. D) Pie chart showing the fraction of DOWN 22G targets (those with significantly lower abundance of antisense 22Gs in *drh-3* vs wt germlines) that are members of the classes indicated at the right. (EF) only protein-coding genes are included in the analysis. (E) Box plots showing log-transformed fold change (*drh-3* vs mutant) in antisense-22G RNA abundance for CSR-1:22G (teal) and WAGO:22G (gold) protein-coding targets. CSR-1:22G and WAGO:22G targets were previously defined in Claycomb et al. 2009 and Gu et al. 2009, respectively. F) Venn diagram comparing the set of all protein-coding 22G-DOWN targets to sets of CSR-1:22G or WAGO:22G protein-coding targets.
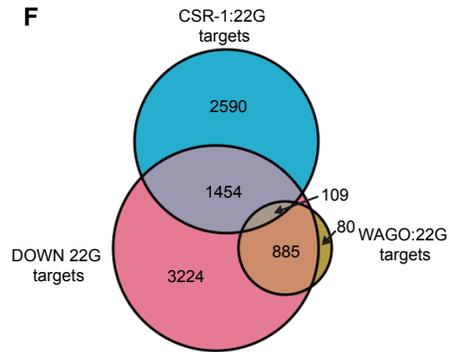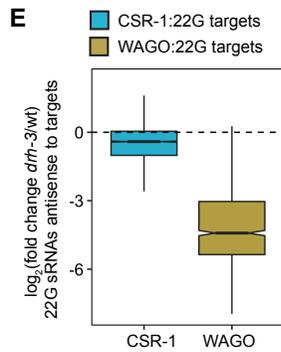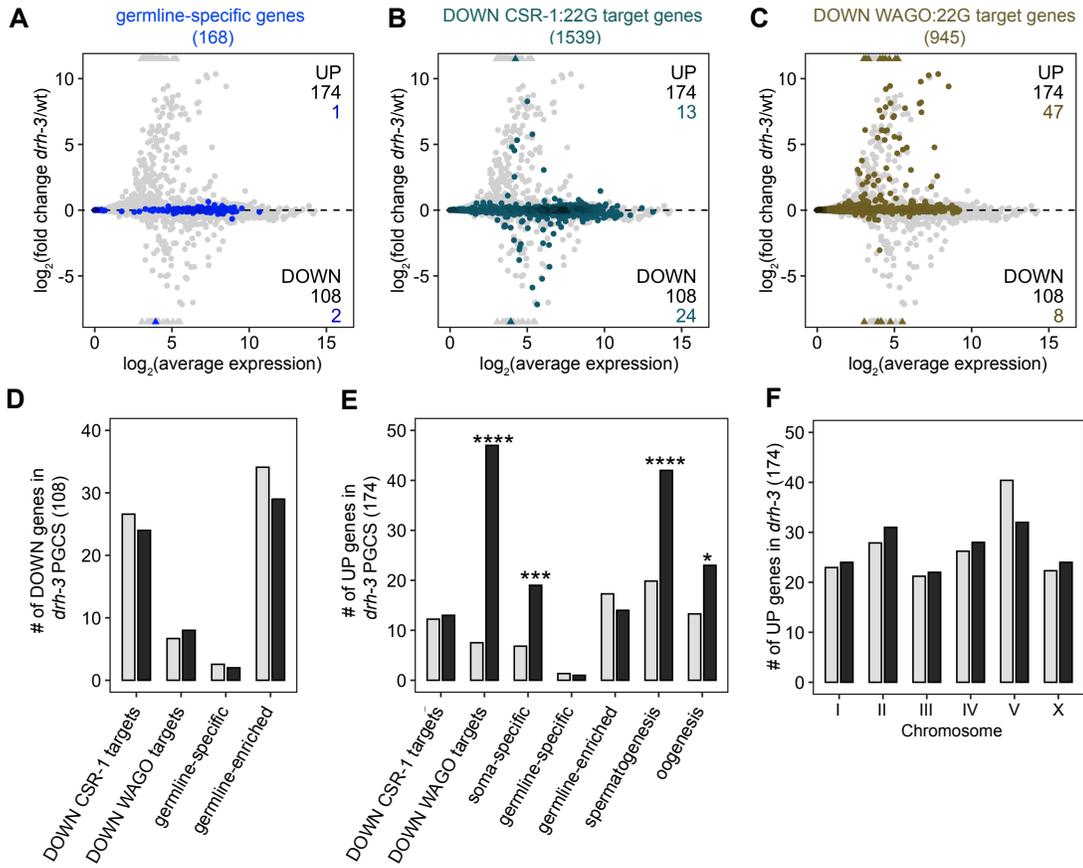
133
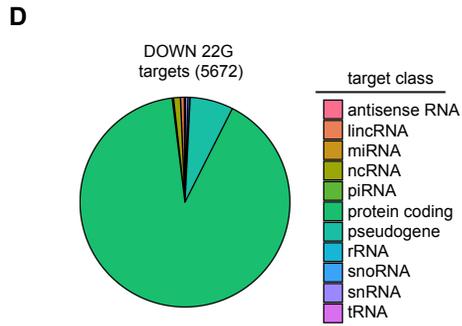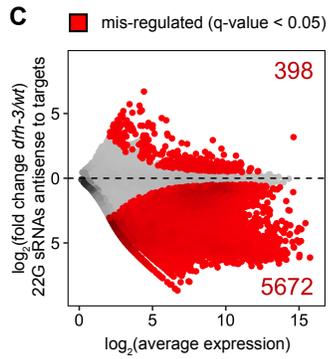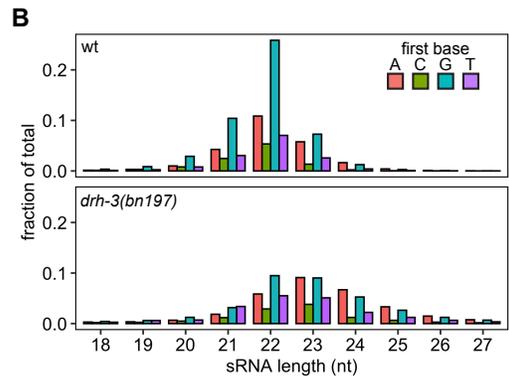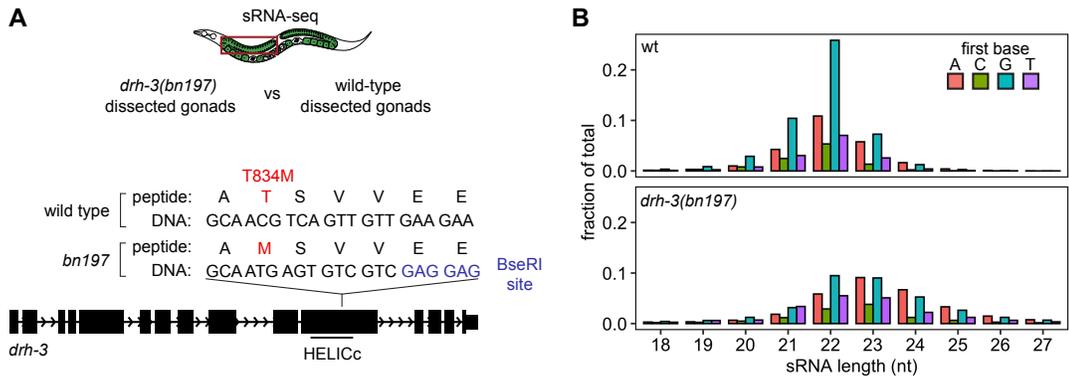
Figure 4-2 (on next page). PGCs dissected from *drh-3(bn197)* L1 larvae turn on a germline program. (A-C) MA plots showing $\log_2$(average expression) versus $\log_2$(fold change) of mRNA abundance for 20,258 protein-coding genes (circles) between *drh-3* and wt PGCs. Genes belonging to a specific gene set are colored: (A) 168 germline-specific genes are blue, (B)1549 DOWN CSR-1:22G target genes are dark teal (from Figure 4-1F), and (C) 945 DOWN WAGO:22G target genes are dark gold Ifrom Figure 4-1F). Differentially expressed genes in *drh-3* vs wt PGCs were identified using Wald tests in DESeq2 (Love et al. 2014) and by setting a q-value < 0.05 significance threshold. Numbers of all mis-regulated genes (black) and numbers of those in gene sets (colored) are indicated in the corners; top is upregulated (UP) and bottom is downregulated (DOWN) in *drh-3* vs wt. (D,E,F) Bar plots showing the expected number (light gray) and observed number (black) of mis-regulated protein-coding genes that are members of gene sets indicated below the bars. Hypergeometric tests were performed in R to test for gene-set enrichment. P-value designations are * < .01, *** < 1e-4, **** < 1e-5. (D) Enrichment analyses for DOWN genes were restricted to 5,858 protein-coding genes that we defined as 'expressed' (minimum average read count of 1) in wt PGCs. Gene set sizes: DOWN CSR-1:22G targets (1456), DOWN WAGO:22G targets (366), germline-specific (140), germline-enriched genes (1867). (E,F) Enrichment analyses for UP genes included all 20,258 protein-coding genes in the transcriptome. Gene set sizes: DOWN CSR-1:22G targets (1549), DOWN WAGO:22G targets (945), soma-specific (861), germline-specific (168), germline-enriched (2176), spermatogenesis (2498), oogenesis (1671), chrI (2888), chrII (3508), chrIII (2670), chrIV (3300), chr V (5084), chr X (2808).

mRNA-seq
drh-3(bn197) vs wild-type
L1 PGCs

A germline-specific genes
(168)

UP
174
1

DOWN
108
2

log$_2$(fold change *drh-3*/wt)
log$_2$(average expression)

B DOWN CSR-1:22G target genes
(1539)

UP
174
13

DOWN
108
24

log$_2$(fold change *drh-3*/wt)
log$_2$(average expression)

C DOWN WAGO:22G target genes
(945)

UP
174
47

DOWN
108
8

log$_2$(fold change *drh-3*/wt)
log$_2$(average expression)

D

# of DOWN genes in *drh-3* PGCS (108)

DOWN CSR-1 targets
DOWN WAGO targets
germline-specific
germline-enriched

E

# of UP genes in *drh-3* PGCS (174)

****
***
****
*

DOWN CSR-1 targets
DOWN WAGO targets
soma-specific
germline-specific
germline-enriched
spermatogenesis
oogenesis

F

# of UP genes in *drh-3* (174)

I   II   III   IV   V   X
Chromosome

135

APPENDIX 2

MES-4 maintains gamete-inherited patterns
of H3K36me3 through development

INTRODUCTION

MES-4 maintains gamete-inherited H3K36me3 patterns during embryogenesis in the absence of transcription (Rechtsteiner et al. 2010; Furuhashi et al. 2010; Kreher et al. 2018). An important question is whether MES-4 functions strictly as a 'maintenance' enzyme or whether it can generate new patterns of marking by catalyzing new (de novo) H3K36me3 onto genomic regions that lack H3K36me3. A couple immunostaining experiments have provided strong evidence that MES-4 is strictly a maintenance enzyme: 1) Kreher et al. showed that MES-4 cannot bind to sperm-inherited chromosomes that lack H3K36me3 [H3K36me3(-)] in the 1-cell embryo, and 2) Furuhashi et al. showed that zygotically expressed MES-4 cannot catalyze detectable levels of H3K36me3 in L1 larval PGCs (Furuhashi et al. 2010; Kreher et al. 2018). A shared limitation of these studies is that they tested MES-4's histone methyltransferase (HMT) role in nascent germ cells before or soon after they launch their germline program. Since transcription can greatly shape the chromatin landscape, it is crucial to examine MES-4's HMT role further into development. An undergraduate in the Strome lab, Audrey Piatt, and I tested whether MES-4 faithfully maintains patterns of gamete-inherited H3K36me3 in the germline through larval development.

136

RESULTS

We generated F1 embryos that inherited H3K36me3(+) chromosomes from feminized *met-1; fem-2* mothers (M+) and H3K36me3(-) chromosomes from *met-1; mes-4* double mutant fathers (P-). Therefore, I will refer to these F1 embryos as H3K36me3 M+P-. Importantly, the H3K36me3 M+P- F1s have MES-4 but lack the other H3K36 HMT in *C. elegans,* MET-1. We allowed the F1 H3K36me3 M+P- embryos to develop into fertile adults and then immunostained H3K36me3 in their oocytes to determine whether MES-4 catalyzed de novo marking on the sperm-inherited (P-) chromosomes. In oocytes, there are 6 sets of paired homologs; the maternally-inherited chromosomes (M+) are adjacent to the corresponding paternally inherited chromosomes (P-). In our imaging analysis of the F1 oocytes, we observed a consistent pattern: 5 'hybrid' homolog sets that were approximately half-stained for H3K36me3 and 1 homolog set that completely lacked H3K36me3 (Figure 4-3). The unstained homolog set is likely the X chromosomes, which are known to be largely devoid of H3K36me3 (Bender et al. 2006; Rechtsteiner et al. 2010; Kreher et al. 2018). We interpret the hybrid staining pattern of the 5 autosomes to indicate that MES-4 faithfully maintained gamete-inherited patterns of H3K36me3 marking through larval development: H3K36me3 was maintained on M+ chromosomes and H3K36me3 was not newly catalyzed on P- chromosomes. Thus, this experiment extends the previous findings mentioned above and provides strong evidence that MES-4 is truly a 'maintenance' HMT that can faithfully transmit epigenetic information across generations and through many rounds of cell division.

METHODS

Strains used in the study

SS1467 - *met-1(bn200)/tmC20[unc-14(tmIs1219) dpy-5(tm9715)] glh-1(sams24[glh-1::GFP::3xFLAG]) I*

SS1494- *met-1(bn200)/tmC20[unc-14(tmIs1219) dpy-5(tm9715)] glh-1(sams24[glh-1::GFP::3xFLAG]) I; mes-4(bn73)/tmC12[egl-9(tmIs1197)] V*

H3K36me3 immunostaining of oocytes

H3K36me3 immunostaining was performed as described in Kreher et al. 2018. Images were acquired using a spinning-disk confocal microscope equipped with a Yokogawa CSU-X1 confocal scanner unit, Nikon TE2000-E inverted stand, Hamamatsu ImageEM X2-CCD camera, solid state 405, 488, 561, 640 nm laser lines, 460/50, 525/50, 593/40, 700/75 nm (EM/BP) fluorescence filters, DIC, Nikon Plan Apo VC 20×/0.5 air objective, Nikon Plan Apo 100×/1.40 oil objective, and Micro-Manager software (v 1.4.20). Image processing for images was done in Fiji (v 2.1.0/1.53C) and photoshop.
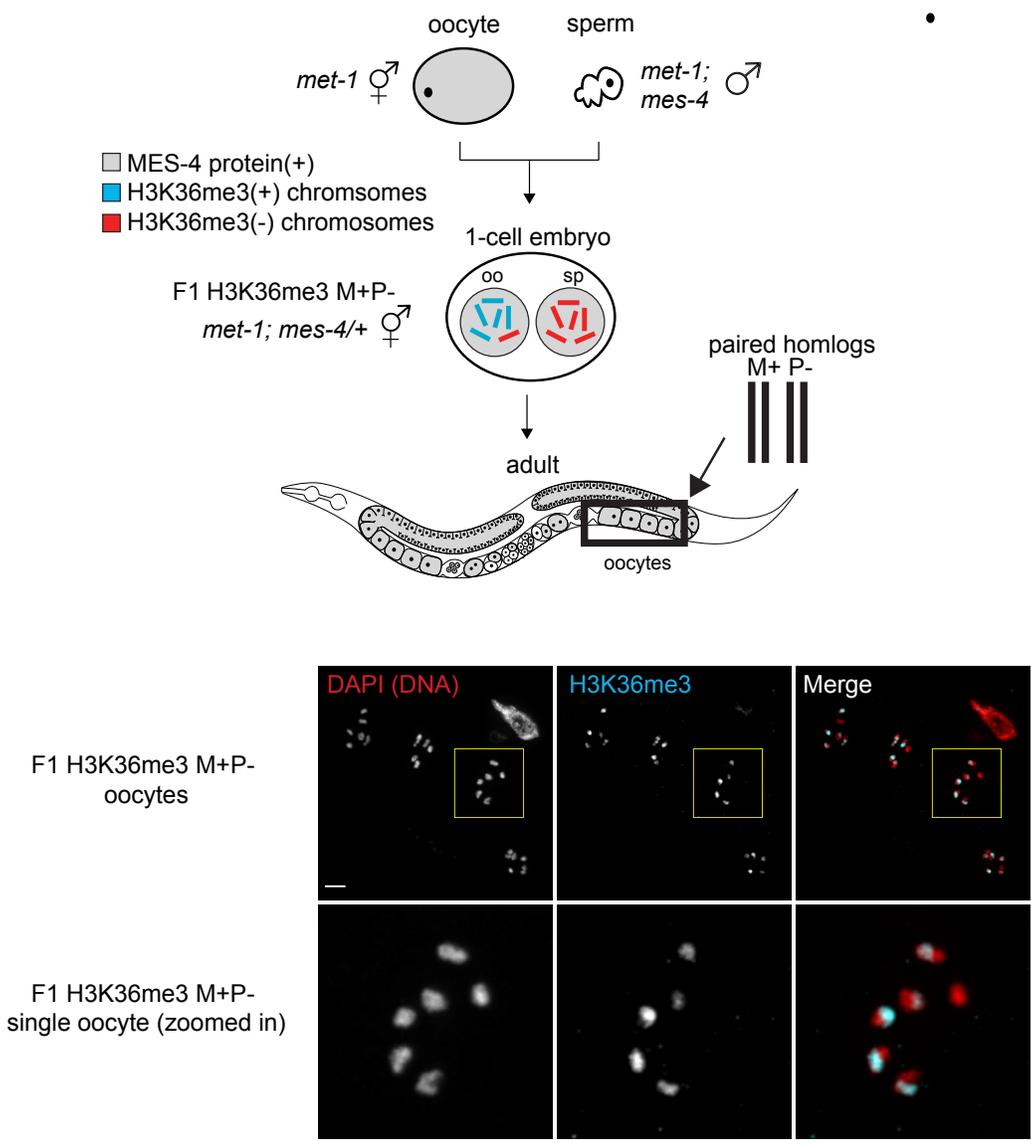
Figure 4-3. MES-4 maintains gamete-inherited patterns of H3K36me3 through development. A schematic diagramming the experiment's strategy (top). We crossed *met-1* mothers to *met-1; mes-4* fathers to generate F1 offspring that inherit H3K36me3(+) chromosomes from the oocyte and H3K36me3(-) chromosomes from the sperm. Thus, we call the F1s 'H3K36me3 M+P-' for Maternal(+) and Paternal(-). Those F1 offspring have the MES-4 H3K36 HMT but lack the MET-1 H3K36 HMT. We allowed the F1s to develop into adults and then immunostained H3K36me3 in their oocytes (bottom). The bottom row are zoomed in images of the yellow-boxed region in the top row. Scale bar is 5 microns. In oocytes, the oocyte-inherited homolog and the sperm-inherited homolog are paired.

REFERENCES

Allis, D. C., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Review Genetics* 17, 487-500.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 31, 166–169.

Andersen, E. C., and Horvitz, H. R. (2007). Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. *Development (Cambridge, England)* 134, 2991–2999.

Aramaki, S., Hayashi, K., Kurimoto, K., Ohta, H., Yabuta, Y. et al. (2013). A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Developmental cell* 27, 516–529.

Arico, J. K., Katz, D. J., van der Vlag J., and Kelly, W. G. (2011). Epigenetic patterns maintained in early Caenorhabditis elegans embryos can be established by gene activity in the parental germ cells. PLoS Genet. 7: e1001391.

Arribere, J. A., Bell, R. T., Fu, B. X. H., Artiles, K. L., Hartman, P. S. et al. (2014). Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics* 198,837-846.

Artiles, K. L., Fire, A. Z., & Frøkjær-Jensen, C. (2019). Assessment and Maintenance of Unigametic Germline Inheritance for C. elegans. *Developmental cell* 48, 827–839

Ashburner, M., Ball C. A., Blake J. A., Botstein D., Butler H., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 25, 25-29.

Avgousti, D. C., Palani, S., Sherman, Y., and Grishok, A. (2012). CSR-1 RNAi pathway positively regulates histone expression in C. elegans. *The EMBO journal* 31, 3821–3832.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., et al. (2006). Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* 8, 532-538.

Bannister, A. J., Schneider R., Myers F. A., Thorne A. W., Crane-Robinson C., et al. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* 280, 17732–17736.

Bannister, A., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395.

Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L., and Hunter, C. P. (2003) Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. *Development*. 130, 889–900.

Bauer, J., Poupart, V., Goupil, E., Nguyen, K., Hall, D. H. et al. (2021). The initial expansion of the C. elegans syncytial germ line is coupled to incomplete primordial germ cell cytokinesis. *Development (Cambridge, England)* 148, dev199633.

Bean, C. J., Schaner, C. E., and Kelly, W. G. (2004). Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. *Nat Genet.* 36, 100-105.

Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A. et al. (2008). Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster. *Molecular and cellular biology* 28, 3401-3409.

Bender, L.B., Cao, R., Zhang, Y., and Strome, S. (2004). The MES-2/MES-3/MES-6 Complex and Regulation of Histone H3 Methylation in C. elegans. *Current Biology* 14, 1639-1643.

Bender, L. B., Suh, J., Carroll, C. R., Fong, Y., Fingerman, I. M., et al. (2006). MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the C. elegans germ line. *Development* 133, 3907-3917.

Bertram, M. J., and Pereira-Smith, O. M. (2001). Conservation of the MORF4 related gene family: identification of a new chromo domain subfamily and novel protein motif. *Gene* 266, 111-121

Besseling, J., and Bringmann, H. (2016). Engineered non-Mendelian inheritance of entire parental genomes in C. elegans. *Nature biotechnology* 34, 982–986.

Blazie, S. M., Babb C., Wilky, H., Rawls, A., Park, J. G. et al. (2015). Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. *BMC Biol.* 13.

Blazie, S. M., Geissel, H. C., Wilky, H., Joshi, R., Newbern, J. et al. (2017). Alternative Polyadenylation Directs Tissue-Specific miRNA Targeting in Caenorhabditis elegans Somatic Tissues. *Genetics* 206, 757–774.

Boeck, M. E., Huynh C., Gevirtzman L., Thompson O. A., Wang G. et al. (2016). The time-resolved transcriptome of C. elegans. *Genome Res.* 26, 1441–1450.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A. et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349-353.

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

Buscaino, A., Köcher, T., Kind, J. H., Holz, H., Taipale, M., et al. (2003). MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex. *Molecular cell*, 11, 1265–1277.

Cabianca, D. S., Muñoz-Jiménez, C., Kalck, V., Gaidatzis, D., Padeken, J., et al. (2019). Active chromatin marks drive spatial sequestration of heterochromatin in C. elegans nuclei. *Nature* 569, 734–739.

Cai, Y., Jin, J. Tomomori-Sato, C., Sato, S., Sorokina, I. et al. (2003). Identification of new subunits of the multiprotein mammalian TRRAP/TIP60-containing histone acetyltransferase complex. *Journal of Biological Chemistry*. 278, 42733-42736.

Campbell, A. C., and Updike, D. L. (2015). CSR-1 and P granules suppress sperm-specific transcription in the C. elegans germline. *Development (Cambridge, England)* 143, 1745–1755.

Cao, J., Packer, J. S., Ramani, V., Cusanovich, D. A., Huynh, C. et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*. 357, 661–667.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H. et al. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298,1039-1043.

Capowski, E. E., Martin, P., Garvin, C., and Strome, S. (1991). Identification of grandchildless loci whose products are required for normal germ-line development in the nematode Caenorhabditis elegans. *Genetics* 129, 1061-1072.

Carpenter, B. S., Lee, T. W., Plott, C. F., Rodriguez, J. D., Brockett, J. S. et al. (2021). Caenorhabditis elegans establishes germline versus soma by balancing inherited histone methylation. *Development* 148, dev196600

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K. et al. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581–592.

Cecere, G., Hoersch, S., O'Keeffe, S., Sachidanandam, R., and Grishok, A. (2014). Global effects of the CSR-1 RNA interference pathway on the transcriptional landscape. *Nature structural & molecular biology* 21, 358–365.

Celniker, S. E., Dillon, L. A. L., Gerstein, M. B., Gunsalus, K. C., Henikoff, S. et al. (2009). Unlocking the secrets of the genome. *Nature* 459, 927–930.

Checchi, P. M., and Engebrecht, J. (2011). Caenorhabditis elegans methyltransferase MET-2 shields the male X chromosome from checkpoint machinery and mediates meiotic sex chromosome inactivation. *PLoS Genetics* 7, e1002267.

Chen, M., Takano-Maruyama, M., Pereira-Smith, O. M., Gaufo, G. O., and Tominaga, K. (2009). MRG15, a component of HAT and HDAC complexes, is essential for proliferation and differentiation of neural precursor cells. *Journal of neuroscience research* 87, 1522–1531.

Chen, S., Zhou, Y., Chen, Y., Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, 884-890.

Chikina, M. D., Huttenhower, C., Murphy, C. T., and Troyanskaya, O. G. (2009). Global prediction of tissue-specific gene expression and context-dependent gene networks in Caenorhabditis elegans. *PLoS Comput. Biol.* 5.

Claycomb, J. M., Batista, P. J., Pang, K. M., Gu, W., Vasale, J. J. et al. (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* 139: 123–134.

Cockrum, C., Kaneshiro, K. R., Rechtsteiner, A., Tabuchi, T. M., and Strome, S. (2020). A primer for generating and using transcriptome data and gene sets. *Development (Cambridge, England)* 147, dev193854.

Concordet, J., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens, *Nucleic Acids Research* 46, W242–W245,

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A. et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*. 17, 1–19.

Conine, C. C., Batista, P. J., Gu, W., Claycomb, J. M., Chaves, D. A. et al. (2010). Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* 107, 3588–3593.

Cui, M., Kim, E. B., and Han, M. (2006). Diverse chromatin remodeling genes antagonize the Rb-involved SynMuv pathways in C. elegans. *PLoS Genetics* 2, e74.

Crittenden, S. L., Eckmann, C. R., Wang, L., Bernstein, D. S., Wickens, M. et al. (2003). Regulation of the mitosis/meiosis decision in the Caenorhabditis elegans germline. *Philosophical transactions of the Royal Society of London. Series B, Biological science* 358, 1359–1362.

Delaney, K., Strobino, M., Wenda, J. M., Pankowski, A., and Steiner, F. A. (2019). H3.3K27M-induced chromatin changes drive ectopic replication through misregulation of the JNK pathway in C. elegans. *Nature communications* 10, 2529.

de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Developmental cell* 7, 663–676.

Diag, A., Schilling, M., Klironomos, F., Ayoub, S., and Rajewsky, N. (2018). Spatiotemporal m(i)RNA Architecture and 3′ UTR Regulation in the C. elegans Germline. *Dev. Cell.* 47, 785-800.e8.

Dickinson, D. J., and Goldstein, B. (2016). CRISPR-based methods for Caenorhabditis elegans genome engineering. *Genetics* 202, 885–901.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C. et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

Dokshin, G. A., Ghanta, K. S., Piscopo, K. M., and Mello, C. C. (2018). Robust Genome Editing with Short Single-Stranded and Long, Partially Single-Stranded DNA Donors in Caenorhabditis elegans. *Genetics* 210, 781-787.

Doyon, Y., Selleck, W., Lane, W. S., Tan, S., Côté, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Mol. Cell Biol.* 24, 1884-1896.

Duchaine, T. F., Wohlschlegel, J. A., Kennedy, S., Bei, Y., Conte, D. Jr., et al. (2006). Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. *Cell* 125, 343–354.

Duncan, I. M. (1982). Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. *Genetics* 102, 49-70.

Ebbing, A., Vértesy, Á., Betist, M. C., Spanjaard, B., Junker, J. P. et al. (2018). Spatial Transcriptomics of C. elegans Males and Hermaphrodites Identifies Sex-Specific Differences in Gene Expression Patterns. *Dev. Cell.* 47, 801-813.e6.

Eisen, A., Utley, R. T., Nourani, A., Allard, S., Schmidt, P., et al. (2001). The yeast NuA4 and *Drosophila* MSL complexes contain homologous subunits important for transcriptional regulation. *J. Biol. Chem.* 276, 3484-3491.

Ephrussi, A., and Lehmann, R. (1992). Induction of germ cell formation by oskar. *Nature* 358, 387-392.

Eskeland, R., Leeb, M., Grimes, G. R., Kress, and C. Boyle, S. (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* 38, 452-464.

Evans, K. J., Huang, N., Stempor, P., Chesney, M. A., Down, T. A., et al. (2016). Stable C. elegans chromatin domains separate broadly expressed and developmentally regulated genes. *PNAS* 113, E7020–E7029

Felsenfeld, G. and Groudine, M. (2003). Controlling the double helix. *Nature* 421, 448-53.

Fong, Y., Bender, L., Wang, W., and Strome, S. (2002). Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of C. elegans. *Science* 296, 2235–2238.

Fox, R. M., Von Stetina, S. E., Barlow, S. J., Shaffer, C., Olszewski, K. L. et al. (2005). A gene expression fingerprint of C. elegans embryonic motor neurons. *BMC Genomics*. 6, 1–23.

Fox, R. M., Watson, J. D., Von Stetina, S. E., McDermott, J., Brodigan, T. M., et al. (2007). The embryonic muscle transcriptome of Caenorhabditis elegans. *Genome Biol.* 8.

Frohnhofer, H.G., Nusslein-Volhard, C. (1987). Maternal genes required for the anterior localization of bicoid activity in the embryo of Drosophila. *Genes Dev.* 1, 880-890.

Fry, C. J., and Farnham, P. J. (1999). Context-dependent transcriptional regulation. *The Journal of biological chemistry 274*, 29583–29586.

Fukuyama, M., Rougvie, A. E., and Rothman, J. H. (2006). C. elegans DAF-18/PTEN mediates nutrient-dependent arrest of cell cycle and growth in the germline. *Curr. Biol.* 16, 773-779.

Fujita, M., Takasaki, T., Nakajima, H., Kawano, T., Shimura, Y. et al. (2002). MRG-1, a mortality factor-related chromodomain protein, is required maternally for primordial germ cells to initiate mitotic proliferation in C. elegans. *Mech. Dev.* 114, 61-69.

Furuhashi, H., Takasaki, R., Rechtsteiner, A., Li T., Kimura H. et al. (2010) Trans-generational epigenetic regulation of C. elegans primordial germ cells. *Epigenetics & Chromatin* 3, 15.

Gallo, C. M., Wang, J. T., Motegi, F., and Seydoux, G. (2010). Cytoplasmic partitioning of P granule components is not required to specify the germline in C. elegans. *Science* 330, 1685-1689.

Gao, M., and Arkov, A. L. (2012). Next generation organelles: structure and role of germ granules in the germline. *Mol. Reprod. Dev.* 80, 610–623.

Gaydos, L. J., Rechtsteiner, A., Egelhofer, T. A., Carroll, C. R., and Strome, S. (2012). Antagonism between MES-4 and Polycomb Repressive Complex 2 Promotes Appropriate Gene Expression in C. elegans Germ Cells. *Cell Rep.* 2, 1169–1177.

Gaydos, L. J., Wang, W., and Strome, S. (2014). Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science (New York, N.Y.)* 345, 1515–1518.

Garvin, C., Holdeman, R., and Strome, S. (1998). The phenotype of mes-2, mes-3, mes-4 and mes-6, maternal-effect genes required for survival of the germline in Caenorhabditis elegans, is sensitive to chromosome dosage. *Genetics* 148, 167–185.

Georgescu, P. R., Capella, M., Fischer-Burkart, S., and Braun, S. (2020). The euchromatic histone mark H3K36me3 preserves heterochromatin through

sequestration of an acetyltransferase complex in fission yeast. *Microbial cell (Graz, Austria)* 7, 80–92.

Gerstein, M., Lu Z., Van Nostrand, E., Cheng, C., Arshinoff, B. et al. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science*. 330, 1775–1787.

Gerstein, M. B., Rozowsky, J., Yan, K. K., Wang, D., Cheng, C. et al. (2014). Comparative analysis of the transcriptome across distant species. *Nature*. 512, 445–448.

Gerson-Gurwitz, A., Wang, S., Sathe, S., Green, R., Yeo, G. W. et al. (2016). A Small RNA-Catalytic Argonaute Pathway Tunes Germline Transcript Levels to Ensure Embryonic Divisions. *Cell* 165, 396–409.

Goetsch, P. D., Garrigues, J. M., and Strome, S. (2017). Loss of the Caenorhabditis elegans pocket protein LIN-35 reveals MuvB's innate function as the repressor of DREAM target genes. *PLoS genetics* 13, e1007088.

Goldstein, B., and Macara, I. G. (2007). The PAR proteins: fundamental players in animal cell polarization. *Developmental cell* 13, 609–622.

Gorman M., Franke A., and Baker B. S. (1995). Molecular characterization of the *male-specific lethal-3* gene and investigations of the regulation of dosage compensation in *Drosophila*. *Development* 121, 463-475.

Greer, E. L., Beese-Sims, S. E., Brookes, E., Spadafora, R., Zhu, Y. et al. (2014). A histone methylation network regulates transgenerational epigenetic memory in C. elegans. *Cell Rep.* 7, 113-126.

Gu, W., Shirayama, M., Conte, D. Jr., Vasale, J., Batista, P. J. et al. (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. *Molecular cell* 36, 231–244.

Haenni, S., Ji, Z., Hoque, M., Rust, N., Sharpe, H., et al. (2012) Analysis of C. elegans intestinal gene expression and polyadenylation by fluorescence-activated nuclei sorting and 3′-end-seq. *Nucleic Acids Res.* 40, 6304–6318.

Hajduskova, M., Baytek, G., Kolundzic, E., Gosdschan, A., Kazmierczak, M. et al. (2019). MRG-1/MRG15 Is a Barrier for Germ Cell to Neuron Reprogramming in Caenorhabditis elegans. *Genetics* 211, 121-139.

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P. and Nakamura, A. (2008). Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature* 451, 730–733.

Harris, T. W., Arnaboldi, V., Cain, S., Chan, J., Chen, W. J. et al. (2019). WormBase: a modern Model Organism Information Resource. *Nucleic Acids Res.* 48, 762–767.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Rep.* 2, 666–673.

Hashimshony, T., Feder, M., Levin, M., Hall, B. K., and Yanai, I. (2015). Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature*. 519, 219–222.

Heard, E., and Martienssen, R. A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* 157, 95–109.

Hillier, L. W., Reinke, V., Green, P., Hirst, M., Marra, M. A. et al. (2009). Massively parallel sequencing of the polyadenylated transcriptome of C. elegans. *Genome Res.* 19, 657–666.

Holdeman, R., Nehrt, S., and Strome, S. (1998). MES-2, a maternal protein essential for viability of the germline in Caenorhabditis elegans, is homologous to a Drosophila Polycomb group protein. *Development* 125, 2457-2467.

Huang, D. W., Sherman, B. T., Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4, 44-57.

Huang, C., and Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophysics reports* 4, 170–177.

Hubbard, E. J. A., and Greenstein, D. (2005). Introduction to the germ line (September 1, 2005), WormBook, ed. The C. elegans Research Community, WormBook, doi/10.1895/wormbook.1.18.1, http://www.wormbook.org.

Hwang, B, Lee, J. H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med.* 50, 96.

Illmensee, K. and Mahowald, A. P. (1974). Transplantation of posterior polar plasm in Drosophila. Induction of germ cells at the anterior pole of the egg.

Proceedings of the National Academy of Sciences of the United States of America 71, 1016–1020.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. *Science* 293, 1074–1080.

Ji, N., and van Oudenaarden, A. (2012). Single molecule fluorescent in situ hybridization (smFISH) of C. elegans worms and embryos. *WormBook : the online review of C. elegans biology*, 1–16.

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R. et al. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* 21, 1543-1551.

Jorgensen, H. F., Giadrossi, S., Casanova, M., Endoh, M., Koseki, H. et al. (2006). Stem cells primed for action: polycomb repressive complexes restrain the expression of lineage-specific regulators in embryonic stem cells. *Cell Cycle* 5, 1411-4.

Joshi, A. A., and Struhl, K. (2005). Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol. Cell.* 20, 971–978.

Kaletsky, R., Yao, V., Williams, A., Runnels, A. M., Tadych, A. et al. (2018). Transcriptome analysis of adult Caenorhabditis elegans cells reveals tissue-specific gene and isoform expression. *PLoS Genet.* 14.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R. et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* 421, 231–237.

Kaneshiro, K. R., Rechtsteiner, A., and Strome S. (2019). Sperm-inherited H3K27me3 impacts offspring transcription and development in C. elegans. *Nat. Commun.* 10, 1271.

Katz, D. J., Edwards, T. M., Reinke, V., and Kelly, W. G. (2009). A C. elegans LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* 137, 308-320.

Kawasaki, I., Shim, Y. H., Kirchner, J., Kaminker, J., Wood W. B. et al. (1998). PGL-1, a predicted RNA-Component of Germ Granules, Is Essential for Fertility in C. elegans. *Cell* 94, 635-45.

Kawasaki, I., Amiri, A., Fan, Y., Meyer, N., Dunkelbarger, S. et al. (2004). The PGL family proteins associate with germ granules and function redundantly in Caenorhabditis elegans germline development. *Genetics* 167, 645–661.

Kelly, W. G., and Fire, A. (1998). Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. *Development* 125, 2451–2456.

Kelly W. G., Schaner, C. E., Dernburg, A. F., Lee, M. H., Kim, S. K. et al. (2002). X-chromosome silencing in the germline of C. elegans. *Development* 129, 479–492.

Keogh. M. C., Kurdistani, S. K., Morris, S. A., Ahn, S. H., Podolny, V. et al. (2005). Cotranscriptional Set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123, 593–605.

Kerr, S. C., Ruppersburg, C. C., Francis, J. W. and Katz, D. J. (2014). SPR-5 and MET-2 function cooperatively to reestablish an epigenetic ground state during passage through the germ line. *Proc. Natl. Acad. Sci. USA* 111, 9509-9514.

Ketel, C. S., Andersen, E. F., Vargas, M. L., Suh, J., Strome, S. et al. (2005). Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. *Mol Cell Biol.* 25, 6857–6868.

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 37, 907-915.

Kimble, J. E., and White, J. G. (1981). On the control of germ cell development in Caenorhabditis elegans. *Dev. Biol.* 81, 208-219.

Kimble, J., and Simpson, P. (1997). The LIN-12/Notch signaling pathway and its regulation. *Annu. Rev. Cell Dev. Biol.* 12, 333-361.

Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L. et al. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol Cell Biol.* 25, 3305–3316.

Koch, C. M., Chiu, S.F., Akbarpour, M., Bharat, A., Ridge, K. M., Barton, E. T., and Winter, D.R. (2018). A Beginner's Guide to Analysis of RNA Sequencing Data. *Am. J. Respir. Cell Mol. Biol.* 59, 145-157.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S. et al. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41, 376–381.

Kolundzic, E., Ofenbauer, A., Bulut, S. I., Uyar, B., Baytek, G. et al. (2018). FACT Sets a Barrier for Cell Fate Reprogramming in Caenorhabditis elegans and Human Cells. *Dev. Cell* 10, 611-626.

Korf, I., Fan, Y., and Strome, S. (1998). The Polycomb group in Caenorhabditis elegans and maternal control of germline development. *Development* 125, 2469-2479.

Kornberg, R. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* 184, 868-871.

Kouzarides, T. (2007). Chromatin Modifications and Their Functions. *Cell* 128, 693-703.

Kreher, J., Takasaki, T., Cockrum, C., Sidoli, S., Garcia, B. A. et al. (2018). Distinct Roles of Two Histone Methyltransferases in Transmitting H3K36me3-Based Epigenetic Memory Across Generations in Caenorhabditis elegans. *Genetics*, 210, 969–982.

Kudron, M. M., Victorsen, A., Gevirtzman, L., Hillier, L. W., Fisher, W. W. et al. (2018). The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of *Drosophila* and *Caenorhabditis elegans* Transcription Factors. *Genetics* 208, 937–949.

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q. et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, 90-97.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K. et al. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.* 22, 1617–1635.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* 16, 2893–2905.

Lee, C. S., Lu, T., and Seydoux, G. (2017). Nanos promotes epigenetic reprograming of the germline by down-regulation of the THAP transcription factor LIN-15B. *eLife* 6, e30201.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine S. S. et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301-13.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* 276, 565-570.

Li, B., Howe, L., Anderson, S., Yates, J. R., 3rd, and Workman, J. L. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem.* 278, 8897–8903

Li B., Ruotti V., Stewart R. M., Thomson J. A., and Dewey C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 26, 493–500.

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923-930.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.

Lucchesi, J. C., and Kuroda, M. I. (2015). Dosage compensation in Drosophila. *Cold Spring Harbor perspectives in biology* 7, a019398.

Lucio-Eterovic, A. K., Singh, M. M., Gardner, J. E., Veerappan, C. S., Rice, J. C. et al. (2010). Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proceedings of the National Academy of Sciences of the United States of America* 107, 16952–16957.

Luecken, M. D. and Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular systems biology* 15.

Lun, A. T., Bach, K., and Marioni, J. C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome biology* 17, 75.

Magnúsdóttir, E., Dietmann, S., Murakami, K., Gunesdogan, U., Tang, F. et al. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell Biol.* 15, 905–915.

Martinho, R. G., Kunwar, P. S., Casanova, J. and Lehmann, R. (2004). A noncoding RNA is required for the repression of RNA polII-dependent transcription in primordial germ cells. *Curr. Biol.* 14, 159–165.

Maleki, F., Ovens, K., Hogan, D. J., and Kusalik, A. J. (2020). Gene Sets Analysis: Challenges, Opportunities, and Future Research. *Front Genet.* 11, 654.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343–349.

Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son J. et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461, 762–767.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. *Nat. Rev. Genet.* 11, 285–296.

McCarthy, D. J., Campbell, K. R., Lun, A. T., and Wills, Q. F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics (Oxford, England)* 33, 1179–1186.

Mello, C. C., Draper, B. W., Krause, M., Weintraub, H. and Priess, J. R. (1992). The pie-1 and mex-1 genes and maternal control of blastomere identity in early C. elegans embryos. *Cell* 70, 163–176.

Meyer ,B. J. (2005). X-chromosome dosage compensation. In WormBook, ed. The C. elegans Research Community, Wormbook, doi/10.1895/wormbook.1.7.1, http://www.wormbook.org

Miwa, T., Inoue, K., and Sakamoto, H. (2019). MRG-1 is required for both chromatin-based transcriptional silencing and genomic integrity of primordial germ cells in Caenorhabditis elegans. *Genes Cells* 24, 377-389.

Miyazaki ,H., Higashimoto, K., Yada, Y, Endo, T. A., Sharif, J. et al. (2013). Ash1 l methylates Lys36 of histone H3 independently of transcriptional elongation to counteract polycomb silencing. *PLoS Genet.* 9, e1003897.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2005). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5, 1–8.

Mootha, V. K., Lindgren, C. M., Eriksson, K., Subramanian, A., Sihag S. et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* 34, 267-273.

Muller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A. et al. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* 111, 197–208.

Nakamura, A., Amikura, R., Mukai, M., Kobayashi, S. and Lasko, P. F. (1996). Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. *Science* 274, 2075–2079.

Nakamura, A., and Seydoux, G. (2008). Less is more: specification of the germline by transcriptional repression. *Development* 135, 3817-3827.

Nimura, K., Ura, K., Shiratori, H., Ikawa, M., Okabe, M. et al. (2009). A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. *Nature* 460, 287–291.

Noll, M., and Kornberg, R. D. (1977). Action of micrococcal nuclease on chromatin and the location of histone H1. *Journal of molecular biology* 109, 393-404.

Nottke, A. C., Beese-Sims, S. E., Pantalena, L. F. Reinke, V., Shi, Y. et al. (2011). SPR-5 is a histone H3K4 demethylase with a role in meiotic double-strand break repair. *Proceedings of the National Academy of Sciences* 108, 12805-12810.

Nusslein-Volhard, C., Frohnhofer, H. G., and Lehmann, R. (1987). Determination of anteroposterior polarity in Drosophila. *Science* 18, 1675-1681.

Ohinata, Y., Payer B., O'Caroll, D., Ancelin, K., Ono, Y. et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* 436, 207-213.

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32, 292–294.

Olins, A.L. and Olins, D.E. (1974). Spheroid chromatin units ([nu] bodies). *Science* 183, 330-332.

Ortiz, M. A., Noble, D., Sorokin, E. P., and Kimble, J. (2014). A new dataset of spermatogenic vs. oogenic transcriptomes in the nematode Caenorhabditis elegans. *G3 Genes, Genomes, Genet.* 4, 1765–1772.

Osborne Nishimura, E., Zhang, J. C., Werts, A. D., Goldstein, B., and Lieb, J. D. (2015). Asymmetric Transcript Discovery by RNA-seq in C. elegans

Blastomeres Identifies neg-1, a Gene Important for Anterior Morphogenesis. *PLoS Genet.* 11, 1–29.

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

Paix, A., Folkmann, A., Rasoloson, D., and Seydoux, G. (2015). High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics* 201, 47–54.

Patel, T., Tursun, B., Rahe, D. P., and Hobert, O. (2012). Removal of Polycomb repressive complex 2 makes C. elegans germ cells susceptible to direct conversion into specific somatic cell types. *Cell Rep.* 29, 1178-1186.

Paulsen, J. E., Capowski, E. E., and Strome, S. (1995). Phenotypic and Molecular Analysis of mes-3, a Maternal-Effect Gene Required for Proliferation and Viability of the Germ Line is C. elegans. *Genetics* 141, 1585-1598.

Picelli, S., Faridani, O. R., Björklund, A. K., Winberg, G., Sagasser, S. et al. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols* 9, 171–181.

Poepsel, S., Kasinath, V., and Nogales, E. (2018). Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nat. Struct. Mol. Biol.* 25, 154–162.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* 26, 841–842.

Raj, A., van den Bogaard, P., Rifkin, S., et al. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* 5, 877–879.

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., et al. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* 44, W160–W165.

Ramsköld, D., Wang, E. T., Burge, C. B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* 5, 1–11.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., et al. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research.* 47, 191-198.

Rechtsteiner, A., Costello, M. E., Egelhofer, T. A., Garrigues, J. M., Strome, S. et al. (2019). Repression of germline genes in caenorhabditis elegans somatic tissues by H3K9 dimethylation of their promoters. *Genetics* 212, 125-140.

Reimand, J., Isser, R., Voisin, V., Kucera, M., Tannus-Lopes, C. et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape, and EnrichmentMap. *Nat Protoc.* 14, 482-517.

Reinke, V., Smith, H. E., Nance, J., Wang, J., Van Doren, C., et al. (2000). A global profile of germline gene expression in C. elegans. *Mol. Cell.* 6, 605–616.

Reinke, V., Gil, I. S., Ward, S., and Kazmer, K. (2004). Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans. *Development* 131, 311–323.

Petrella, L. N., Wang, W., Spike, C. A., Rechtsteiner, A., Reinke, V. et al. (2011). synMuv B proteins antagonize germline fate in the intestine and ensure C. elegans survival. *Development* 138, 1069-1079.

Robert, V. J., Knutson, A. K., Rechtsteiner, A., Garvis, S., Yvert, G., et al. (2020). *Caenorhabditis elegans* SET1/COMPASS Maintains Germline Identity by Preventing Transcriptional Deregulation Across Generations. *Frontiers in cell and developmental biology* 8, 561791.

Robinson, P. J., and Rhodes, D. (2006). Structure of the '30 nm' chromatin fibre: A key role for the linker histone. *Curr. Opin. Struct. Biol.* 16, 336–343.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.

Robinson, D. G. and Storey, J. D. (2014). Determining Appropriate Sequencing Depth Through Efficient Read Subsampling. *Bioinformatics* 30, 3424-3426.

Rongo, C., and Lehmann, R. (1996). Regulated synthesis, transport and assembly of the Drosophila germ plasm. *Trends in genetics* 12, 102–109.

Roy, P. J., Stuart, J. M., Lund, J., and Kim, S. K. (2002). Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. *Nature* 418, 975–979.

Saitou, M., Barton, S. C. and Surani, M. A. (2002). A molecular programme for the specification of germ cell fate in mice. *Nature* 418, 293–300.

Sanchez, R., and Zhou, M. M. (2011). The PHD finger: a versatile epigenome reader. *Trends in biochemical sciences* 36, 364–372.

Sankaran, S. M., Wilkinson, A. W., Elias, J. E., and Gozani, O. (2016). A PWWP Domain of Histone-Lysine N-Methyltransferase NSD2 Binds to Dimethylated Lys-36 of Histone H3 and Regulates NSD2 Function at Chromatin. *The Journal of biological chemistry* 291, 8465–8474.

Schaner, C. E., Deshpande, G., Schedl, P. D., and Kelly, W. G. (2003). A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. *Developmental cell* 5, 747–757.

Schaner, C. E. and Kelly, W. G. (January 24, 2006). Germline chromatin. WormBook, ed. The C. elegans Research Community, WormBook, doi/10.1895/wormbook.1.73.1,

Schulze, A. and Downward, J. (2001). Navigating gene expression using microarrays — a technology review. *Nat Cell Biol.* 3, E190–E195.

Seelk, S., Adrian-Kalchhauser, I., Hargitai, B., Hajduskova, M., Gutnik, S. et al. (2016). Increasing Notch signaling antagonizes PRC2-mediated silencing to promote reprograming of germ cells into neurons. *eLife* 5, e15477

Seydoux, G., Mello, C. C., Pettitt, J., Wood, W. B., Priess, J. R. et al. (1996). Repression of gene expression in the embryonic germ lineage of C. elegans. *Nature* 382, 713–716.

Seydoux, G., and Dunn, M. A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of Caenorhabditis elegans and Drosophila melanogaster. *Development* 124, 2191–2201.

Seydoux, G., and Braun, R. E. (2006). Pathway to Totipotency: Lessons from Germ Cells. *Cell* 127, 891-904.

Seydoux, G., and Schedl, T. (2001). The germline in C. elegans: origins, proliferation, and silencing. *International review of cytology* 203, 139–185.

Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte, D., Jr et al. (2013). The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Developmental cell* 27, 656–663.

Sheth, U., Pitt, J., Dennis, S., Priess, J. R. (2010). Perinuclear P granules are the principal sites of mRNA export in adult C. elegans germ cells. *Development* 137, 1305–14.

Shirayama, M., Seth, M., Lee, Heng-Chi, Gu, W., Ishidate, T. et al. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. *Cell* 150, 65–77.

Silas, S., Jain, N., Stadler, M., Fu, B., Sánchez-Amat, A. et al. (2018). A Small RNA Isolation and Sequencing Protocol and Its Application to Assay CRISPR RNA Biogenesis in Bacteria. *Bio-protocol* 8, e2727.

Smith, J. L., Wilson, J. E., and Macdonald, P. M. (1992). Overexpression of oskar directs ectopic activation of nanos and presumptive pole cell formation in Drosophila embryos. *Cell* 70, 849–859.

Spike, C., Meyer, N., Racen, E., Orsborn, A., Kirchner, J. et al. (2008). Genetic analysis of the Caenorhabditis elegans GLH family of P-granule proteins. *Genetics* 178, 1973–1987

Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z. W., Bone, J. R. et al. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol.* 22, 1298-1306.

Straub, T., and Becker, P. B. (2007). Dosage compensation: the beginning and end of generalization. *Nature reviews genetics* 8, 47–57.

Strome, S., and Wood, W. B. (1982). Immunofluorescence visualization of germ-line-specific cytoplasmic granules in embryos, larvae, and adults of Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U S A.* 79, 1558-1562.

Strome, S., and Wood, W. B. (1983). Generation of asymmetry and segregation of germ-line granules in early C. elegans embryos. *Cell* 35, 15–25.

Strome, S., Martin, P., Schierenberg, E., and Paulsen, J. (1995). Transformation of the germ line into muscle in mes-1 mutant embryos of C. elegans. *Development* 121, 2961–2972.

Strome, S., Kelly, W. G., Ercan, S., and Lieb, J. D. (2014). Regulation of the X chromosomes in Caenorhabditis elegans. *Cold Spring Harbor perspectives in biology* 6, a018366.

Strome, S. and Updike, D. (2015). Specifying and protecting germ cell fate. *Nat Rev Mol Cell Biol* 16, 406–16.

Schmitges, F. W., Prusty, A. B., Faty, M., Stützer, A., Lingaraju, G. M. et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. *Cell* 42, 330–341.

Stoeckius, M., Grün, D., Kirchner, M., Ayoub, S., Torti, F. et al. (2014). Global characterization of the oocyte-to-embryo transition in Caenorhabditis elegans uncovers a novel mRNA clearance mechanism. *EMBO J.* 33, 1751–1766.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L. et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 102, 15545-15550.

Tabuchi, T. M., Rechtsteiner, A., Jeffers, T. E., Egelhofer, T. A., Murphy, C. T. et al. (2018). Caenorhabditis elegans sperm carry a histone-based epigenetic memory of both spermatogenesis and oogenesis. *Nat. Commun.* 9, 1–11.

Takasaki, T., Liu, Z., Habara, Y., Nishiwaki, K., Nakayama, Jun-Ichi et al. (2007). MRG-1, an autosome-associated protein, silences X-linked genes and protects germline immortality in Caenorhabditis elegans. *Development* 134, 757-767.

Thoma, F., and Koller, T. (1977). Influence of histone H1 on chromatin structure. *Cell* 12, 101–107.

The Gene Ontology Consortium (2018). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* 47, D330-D338.

Tintori, S. C., Osborne, Nishimura E., Golden, P., Lieb, J. D., and Goldstein, B. (2016). A Transcriptional Lineage of the Early C. elegans Embryo. *Dev. Cell* 38, 430–444.

Tong, Z. B., Gold, L., Pfeifer, K. E., Dorward, H., Lee, E. et al. (2000). Mater, a maternal effect gene required for early embryonic development in mice. *Nature genetics* 26, 267–268.

Trojer, P. and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* 28, 1–13.

Tursun, B., Patel, T., Kratsios, P., and Hobert, O. (2011). Direct conversion of C. elegans germ cells into specific neuron types. *Science (New York, N.Y.)* 331, 304–308.

UI Fatima, N., and Tursun, B. (2020). Conversion of Germ Cells to Somatic Cell Types in C. elegans. J. *Dev. Biol.* 8, 24.

Updike, D., and Strome, S. (2010). P granule assembly and function in Caenorhabditis elegans germ cells. *J. Androl.* 31, 53-60.

Updike, D. L., Hachey, S. J., Kreher, J. and Strome, S. (2011). P granules extend the nuclear pore complex environment in the C. elegans germ line. *J. Cell Biol.* 192, 939–948.

Updike, D. L., Knutson, A. K., Egelhofer, T. A., Campbell, A. C., and Strome, S. (2014). Germ-granule components prevent somatic development in the C. elegans germline. *Curr. Biol.* 24, 970-975.

Unhavaithaya, Y., Shin, T. H., Miliaras, N., Lee, J., Oyama, T. et al. (2002). MEP-1 and a homolog of the NURD complex component Mi-2 act together to maintain germline-soma distinctions in C. elegans. *Cell* 111, 991–1002.

Venkatesh, S., Workman, J. L. (2013). Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. *Wiley Interdiscip Rev Dev Biol.* 2, 685–700.

Wagner, E. J., and Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* 13,115–126.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples.*Theory Biosci.* 131. 281-285.

Wall, M. E., Rechtsteiner, A., and Rocha, L. M. (2003). Singular Value Decomposition and Principal Component Analysis. In A Practical Approach to Microarray Data Analysis (ed. D. P. Berrar, W. Dubitzky, & M. Granzow, pp. 91–109). Springer US.

Wang, H., Wang, L., Erdjument-Bromage, H. Vidal, M., Tempst, P. et al. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873–878.

Wang, D., Kennedy, S., Conte, D. Jr., Kim, J. K., Gabel, H. W. et al. (2005). Somatic misexpression of germline P granules and enhanced RNA interference in retinoblastoma pathway mutants. *Nature* 436, 593–597.

Wang, X., Zhao, Y., Wong, K., Ehlers, P., Kohara, Y. et al. (2009). Identification of genes expressed in the hermaphrodite germ line of C. elegans using SAGE. *BMC Genomics* 10.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28, 2184–2185.

Wang, H., La, Russa M., and Qi, L. S. (2016). CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.* 85, 227–264.

Wedeles, C. J., Wu, M. Z., and Claycomb, J. M. (2013). Protection of germline gene expression by the C. elegans Argonaute CSR-1. *Developmental cell* 27, 664–671.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. *Science* (New York, N.Y.), 193, 848–856.

Wu, X., Shi, Z., Cui, M., Han, M., and Ruvkun, G. (2012). Repression of germline RNAi pathways in somatic cells by retinoblastoma pathway chromatin complexes. *PLoS genetics* 8, e1002542.

Yamamoto, M., Wakatsuki, T., Hada, A., and Ryo, A. (2001). Use of serial analysis of gene expression (SAGE) technology. *J. Immunol. Methods.* 250, 45–66.

Xu, L., Fong, Y., and Strome, S. (2001). The Caenorhabditis elegans maternal-effect sterile proteins, MES-2, MES-3, and MES-6, are associated in a complex in embryos. *PNAS* 9, 5061-5066.

Xu, X., Wang, J., Wu, L., Guo, J., Song, Y. et al. (2019). Microfluidic Single-Cell Omics Analysis. *Small* 16.

Yigit, E., Batista, P. J., Bei, Y., Pang, K. M., Chen, C. C. et al. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* 127, 747–757.

Yochum, G. S., and Ayer, D. E. (2002). Role for the mortality factors MORF4, MRGX, and MRG15 in transcriptional repression via associations with Pf1, mSin3A, and Transducin-Like Enhancer of Split. *Mol Cell Biol.* 22, 7868-76.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S. et al. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* 286: 7983–7989.

Zerbino, D. R., Johnson, N., Juettemann, T., Wilder, S. P., and Flicek, P. (2014). WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics (Oxford, England)* 30, 1008–1009.

Zhao, S., Ye Z., and Stanton, R. (2020). Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* 26, 903-909.