LINE-1 RETROTRANSPOSON ACTIVATION IN MOUSE FETAL OOCYTES

by

Marla Tharp

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

April 2020

Abstract

After her pivotal discovery of transposable elements or “jumping genes”, Barbara McClintock left future generations of scientists with the challenge of understanding how such elements control gene expression and development by rearranging the genome in response to genetic and environmental stress. In this work, we make strides toward understanding the impact of transposable element activity in mouse female germ cell development. Germ cells face an evolutionary conundrum: the remarkable responsibility of transmitting high quality genetic information to the subsequent generation while simultaneously providing windows of opportunity for the genome to diversify and adapt that are advantageous for the species fitness. This is an ideal task for transposons. Indeed, early in development oocytes encounter a gauntlet of stressors that together create a natural genomic shock and permit the release of active transposons. Specifically, DNA methylation, an elaborate epigenetic system that normally represses transposons, is surrendered for epigenetic reprogramming of germ cells while concurrent chromosome breakage and rearrangement are ongoing for meiotic prophase I, creating a permissive environment for transposon activity.

The active L1 retrotransposon is expressed in this window of genome stress during development. This L1 burst correlates with death of up to 80% of fetal oocytes, a decades old observation and paradoxical phenomenon as female reproductive success relies on the size and quality of a finite ovarian reserve. A role for L1 in fetal oocyte attrition (FOA) has been characterized. L1-encoded protein ORF1p is heterogeneously expressed between fetal oocytes, and those with elevated ORF1p levels are preferentially killed. Using these findings as a foundation, I aimed to understand the mechanisms by which L1 activity triggers FOA and the biological significance of L1 expression that may outweigh the massive consequences to the oocyte supply. Here, I characterize the mechanisms of L1-mediated FOA that stem from two catalytic activities of L1 ORF2p required for retrotransposition: reverse transcriptase and endonuclease activities. I then successfully block FOA by inhibiting these mechanisms, specifically, by combined inhibition of L1 reverse transcriptase activity with the antiretroviral AZT and the DNA

damage checkpoint activation through mutation of checkpoint kinase 2 (CHK2). Now, for the first time, we study the entire fetal oocyte population and understand the mechanisms of L1-mediated FOA, its impact on oogenesis and fertility, and the origin of L1 heterogeneity among oocytes that determines their fates. Surprisingly, I find that while FOA initially serves as quality control, when bypassed, oocytes are able to reduce genotoxic threats of L1 and differentiate, resulting in a maximized ovarian reserve at postnatal ages without compromising fertility.

Thesis Advisor, Primary Reader: Alex Bortvin, MD, PhD

Secondary Reader: Yumi Kim, PhD

Acknowledgements

I would like to acknowledge and thank the mentors, colleagues, and friends whose efforts have made this thesis work possible. I sincerely thank my PhD mentor, Dr. Alex Bortvin, for the opportunity to conduct research in his lab, and for years of valuable advice with my experiments, presentations, writing, and career. The dedication and passion that Alex shows toward his research, in addition to his ability to think critically on any subject matter, has been incredibly inspiring to me. His high expectations have motivated me to grow as a research scientist.

Many thanks to past and present members of the Bortvin lab. To Dr. Safia Malki, for her mentorship from my first day in the Bortvin lab. I was incredibly fortunate to have her guidance in designing and executing experiments and critical advice on presentations and writing that continues to this day. To Dr. Chiara De Luca for her scientific discussions, boundless support, and cherished friendship. To graduate student alumni, Drs. Valeriya Gaysinskaya and Pavol Genzor, for their scientific discussions and assistance with computational analyses. To Rejeanne Juste for every-day lab management and genotyping of mice. To previous lab members, Casey Hussey, Karen Griffin, and Dr. Xinyi Wang for their contributions to various experiments presented in this thesis.

I thank my colleagues at the Carnegie Institution for Science, Department of Embryology. To Drs. Allison Pinder and Fred Tan, for their expertise with next generation sequencing experiments and computational analyses. To Dr. Mahmud Siddiqui, for his assistance with confocal microscopy and FACS sorting. To Eugenia Dikovsky, for mouse care and generation of transgenic mice. To Dr. Zhao Zhang and the ZZ lab members, for years of valuable advice at lab meetings. To Drs. Ethan Greenblatt and Matt Sieber, for their generous advice with my presentations for conferences and job interviews. To the funding sources that made this thesis work possible, including Carnegie's endowment and NIH grants R21 HD090514 to A.B. and F31 HD088053 to M.E.T.

I thank my thesis committee members, Drs. Yumi Kim, Kathy Burns, and Mark Van Doren, for their critical reading of my manuscript and thesis, as well as advice in committee meetings. I am grateful

to the faculty and administrative staff of the Johns Hopkins Cell, Molecular, Developmental Biology and Biophysics (CMDB) graduate program, especially to my laboratory rotation mentors, Drs. Allan Spradling, Samer Hattar, and Vince Hilser, for the opportunity to study diverse research questions and experimental techniques. Also, many thanks to my classmates in the CMDB program, especially to Sunny Malhotra and Caitlin Pozmanter for their support and friendship.

I thank my mentors from the Woods Hole Embryology Course, including Drs. Rich Schneider, David Sherwood, and numerous other faculty, for sharing their expertise and providing the unique opportunity for me to foster my creativity and passion for science without the pressures of academia. I also thank my mentors and role models from the University of Illinois at Urbana-Champaign, Drs. Phil Newmark and James J. Collins, III. Their endless generosity and support in training me as a researcher have inspired me to pursue this PhD degree and a career in science.

I thank my family, Elizabeth Barca, David Tharp, and Voula Papadopolous, for teaching me the values of education, hard work, and independence. Finally, countless thanks to my dearest friend, Grace Redeker, for twenty years of friendship and her tremendous positivity.

Table of Contents

List of Tables

List of Figures

List of Appendices

CHAPTER 1

Introduction

1.1 The paradox of genetic instability in germ cells

We all began as a single-celled fusion of male and female gametes that gives rise to every cell type in our body. Therefore, the most critical job of gametes is to maintain high quality genetic and epigenetic material during their development that will ultimately be transmitted to the cells of future generations. The real challenge though, is that while preserving genome integrity for the benefit of the individual, gamete genomes must simultaneously retain the potential to adapt to genetic or environmental stress that is advantageous for survival of the species. Gametes are the product of germ cells, and during germ cell development, a number of windows of opportunity remain open to test new arrangements of the genome for such diversification. These developmental windows will be introduced in detail below, including the specification and reprogramming of the epigenome in primordial germ cells that sets the stage for oocyte-specific differentiation, activation and defense against transposable elements, meiotic events including genetic recombination, and oocyte quality control reflected as fetal oocyte attrition. My thesis work explores the fine line between life and death that female germ cells walk during fetal development as they test the limits of tolerable genome instability.

1.2 Reviving the immortal germline

Germ cells are considered totipotent and genetically immortal because they can give rise to all cells of an entire organism that itself produces germ cells to pass to the next generation for an endless series of generations. Therefore, germ cells are the stem cells of a species. How are germ cells able to possess such remarkable plasticity to become some of the most differentiation cell types, spermatozoa and egg, but be also revive the totipotency to make all cell types in the next generation? A considerable part of the story lies in the epigenome reprogramming in the primordial germ cells (PGCs), the founder cells of the germline.

In the mouse embryo, PGCs are specified at embryonic day (E) 6.25 out of the 20-day gestation period[1]. In most organisms including flies, worms and frogs, PGCs are specified much earlier in development by inheriting cytoplasmic germ cell determinants or germ plasm. In *Drosophila*, germ plasm is formed from maternally deposited mitochondria and RNAs driven by the localization of *Oskar* mRNA to the posterior embryo[2]. Germ plasm is necessary and sufficient to specify germ cells in the embryo, demonstrated by transplantation of the normally posteriorly localized germ plasm to the anterior embryo that lead to generation of ectopic PGCs in the anterior[3]. In contrast, PGCs are inductively specified in mice, and likely all mammals, in response to signals from extraembryonic tissues[1]. Mouse PGCs arise from a small number of proximal epiblast cells that become competent to take on PGC fate in response to cues from the extraembryonic ectoderm and visceral ectoderm such as BMPs 4, 8b, and 2 that act to inhibit somatic differentiation programs[4]. Select epiblast cells express *fragilis* in response to BMP-driven phosphorylation of SMAD proteins, making them competent for PGC fate[5]. Germ cell enriched genes such as *blimp1* are specifically expressed in PGCs to repress the somatic program and ultimately result a population of 40 founder PGCs by E7.25 that will go on to migrate to the embryonic gonad and expand[4].

PGCs do not originate with the potency to generate all cell types of the body. This naïve pluripotency is restored by reprogramming the PGC epigenome through genome-wide erasure of DNA methylation and rewriting of covalent modifications on histone N-terminal tails[6]. Reorganization of the

epigenome at this large of scale occurs in only one other context aside from PGCs, that is pre-implantation embryos. However, in preimplantation embryos, re-methylation occurs shortly thereafter during implantation, including in the epiblast cells that become PGCs[6]. For this reason, PGCs initially possess high levels of DNA methylation after specification that become passively eliminated from the genome in a replication dependent manner during PGC migration into the gonad and PGC expansion. Based on whole-genome bisulfite sequencing, two sequential waves of demethylation occur and are defined by their specificity to different genomic regions and timing (Fig. 1-1)[7]. Between the time of PGC specification and migration, ~E6.5 to E9.5, the bulk of demethylation is carried out globally throughout the genome (Fig. 1-1). However, during colonization of the gonad and germ cell expansion, between E9.5 and E13.5, demethylation becomes locus-specific, to imprinting control regions, meiotic and gamete differentiation genes, and the X chromosome(Fig. 1-1)[7]. The timing of demethylation in these waves is critical. Particularly, premature expression of second wave genes that are normally protected from demethylation by DNMT1 during the first wave results in precocious germ cell differentiation and meiotic entry due to misexpression of meiotic licensing genes, subsequent DNA breaks and ultimately germ cell death[8].

Misexpression of other loci characteristic to the second wave of DNA demethylation during epigenetic reprogramming are also critical, especially imprinted genes and transposable elements. Several dozens of genes in mammals have expression in only one of two parental chromosomes, this choice installed by the epigenome[9]. Developmental defects and disease can occur upon aberrant expression of imprinted genes. For example, a number of imprinted genes are expressed in the placenta and when disrupted, lead to defects in growth of the fetus[9]. Additionally, in mammals, DNA methylation also plays a primary role in genome defense against transposable elements by targeting promoters of active elements, particularly the Long Interspersed Nuclear Element (LINE-1 or L1) retrotransposons, to suppress their transcription. In contrast, Intracisternal A Particle (IAP) retrotransposons are largely resistant to reprogramming[7]. L1 is demethylated and expressed during epigenetic reprogramming, leading

to deleterious complications with germ cell genome integrity and survival discussed at length in later sections. Interestingly, L1 elements are readily capable of retrotransposition in imprinted gene regions and are often found not far from each other in the genome[10]. Perhaps a link between the reprogramming of imprinted genes and L1 exists due to their proximity and L1 expression is an indirect consequence of demethylation at these imprinted regions.

The dynamics of re-methylation and restoration of imprints differ between males and females and is relevant to L1 expression. Post sex-determination male germ cells enter a cell cycle arrest, now termed prospermatogonia, and re-establish DNA methylation *de novo* between E16.5 and P3 using DNA methyltransferase 3 family enzymes, rather than DNMT1 that prefers hemimethylated substrates for maintenance methylation[11]. This involves rearrangement of chromatin landscape and rewriting of histone modifications as well as generation of piRNA to guide DNA methylation to specific loci, especially L1 sequences[12,13]. Failure to re-methylate the genome in mutants such as Dnmt3a or Dnmt3l results in hypomethylated imprinting control regions and derepression of transposable elements and failed spermatogenesis[14,15]. In contrast, post sex-determination, female germ cells enter meiosis and remain demethylated until later stages during oocyte growth[16]. This creates a unique and dangerous window of opportunity for transposable elements to be expressed without restraint in female germ cells.

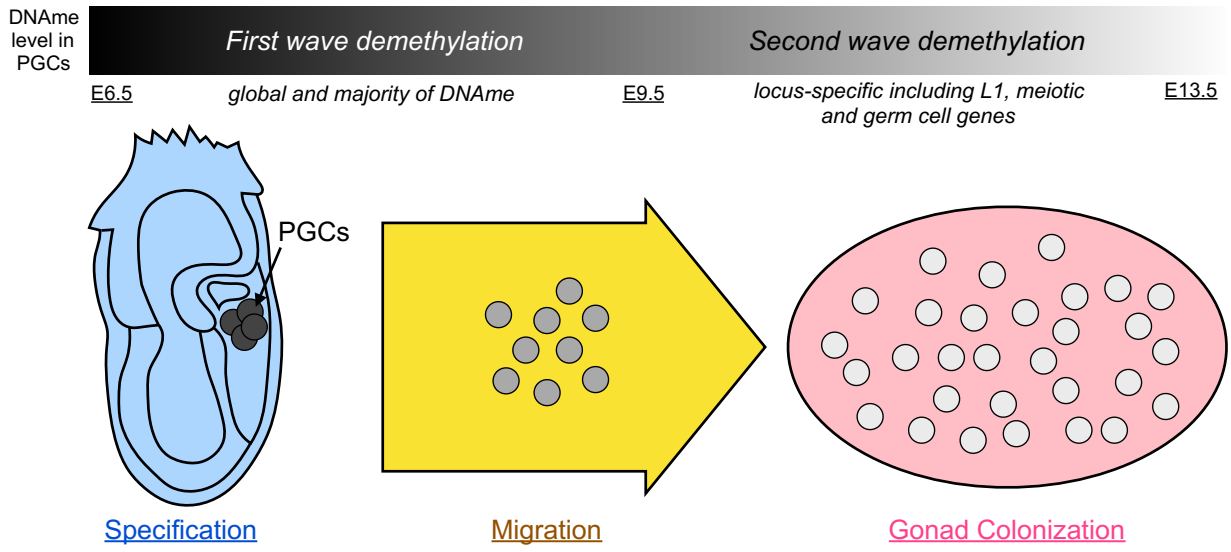


Figure 1-1: Primordial germ cells (PGCs) set the stage for oogenesis. Diagram of PGC specification, migration, and colonization of the gonad in the context of the two waves of DNA methylation (DNAme) erasure during epigenetic reprogramming. PGC specification begins at ~E6.5, resulting in PGCs that are still highly methylated as shown based on the dark gray color on the gradient scale of black (high DNAme) to white (low DNAme). PGCs begin to migrate to the gonad while they complete the first wave of DNAme erasure that is global and accounts for the majority of demethylation during epigenetic reprogramming (E6.5 – E9.5). This is shown by PGCs of a gray color towards the middle of the gradient scale. Subsequently, PGCs colonize and expand their population in the gonad between concurrent with the second wave of demethylation (E9.5 – E13.5). This wave is locus-specific and removes DNA methylation from meiotic and germ cell development genes as well as active LINE-1 (L1) retrotransposons. Resulting PGCs have extremely low levels of DNA methylation remaining as shown by the light gray color.

1.3 Germ cells battle endogenous parasites

Perhaps the most striking information from the human genome project was that less than 5% of our genomes are comprised on coding sequences while at least 50% is accounted for by repeat sequences[17]. This repetitive DNA includes transposons. Transposons and genetic transposition were discovered by the geneticist Barbara McClintock in the mid-1940s and awarded the Nobel prize in 1983 based on experiments surrounding a pair of genetic loci in maize named Dissociation and Activator that together triggered spontaneous mutations[18]. Chromosome breakage at the Dissociation locus that was enhanced by Activator expression were linked to pigment phenotypes in kernels[19]. At the time of her groundbreaking discovery, McClintock proposed that transposons are a feature of the genome to rearrange itself in response to threats in the environment and believed that transposition was a crucial aspect of gene regulation and developmental programs of cells.

Today, we understand that transposons are truly successful and influential mutagens as supported by their existence in the genomes of a diverse range of organisms from bacteria to human and comprising a large percentage of the genome. When domesticated by the host, transposons can regulate gene expression by relocating transcription factor binding sites or pieces of adjacent exons to entirely new locations of the genome and as a consequence, rewire gene regulatory networks or create novel gene products at that location[20,21]. These rearrangements will for the most part be neutral, but in some cases, manifest as a phenotypic change by providing an adaptive benefit. Indeed, transposons have been linked to numerous mutations leading to evolution of new traits in Nature. For example, industrial melanism, a phenomenon in moths that selects for pigment phenotypes advantageous for camouflage to soot-covered trees after the industrial revolution, has been mapped to a Type II DNA transposon insertion in the *cortex* gene[22]. Another striking example is a L2 LINE retrotransposon insertion in the stickleback *GDF6* enhancer that is associated with loss of skeletal armor plates in freshwater populations compared to marine founder populations[23]. Transposons have also been involved in shaping genome architecture. For example, the allotetraploid *Xenopus laevis* has duplicated its genome compared to the diploid *Xenopus*

tropicalis, and chromosomes of the *Xenopus laevis* subgenome are covered with Tc1/Mariner DNA transposons[24]. Instances such as the above described represent genetic scars of transposon mobilization that occurred millions of years ago and their presence repeatedly selected for. In fact, the majority of transposons exist as molecular fossils that have lost the capability to mobilize, as host cells are constantly inventing ways to mutate and eliminate elements from genomic regions. However, specific classes of transposons remain potentially active and these transposons differ among species[25].

LINE-1 retrotransposons are active in humans and mice

Broadly, transposons come in two different flavors with distinct biological features such as structure and life cycle. The DNA transposons, or class II transposons physically excise from their original genomic locus and use a double-stranded or single-stranded DNA intermediate for transposition. The canonical "cut and paste" mechanism uses a double-stranded DNA intermediate, while others use more distinct mechanisms such as rolling circle replication used by *Helitrons* and a still ambiguous mechanism used by *Mavericks* that use a single-stranded DNA intermediate in a "copy and paste" fashion[25]. DNA transposons are active in lower organisms including bacteria, but inactive in humans[26]. However, DNA transposon sequences are still found in human genomes and have important biological roles even though they no longer have the capacity to jump.

In contrast, class I retrotransposons mobilize via a "copy and paste" mechanism using an RNA intermediate. In this way, the original DNA sequence remains intact and a new cDNA created using reverse transcriptase activity encoded by the retrotransposon for insertion at a new genomic locus. Retrotransposons can be further divided into two classes. First, the endogenous retroviruses containing long-terminal repeats (LTR) and analogous sequences to the retroviral *gag*, *pol*, and *env* that create virus-like particles and retrotransposition machinery[26]. Human endogenous retroviruses (HERVs) are inactive, but in mice, LTR retrotransposons are highly abundant and active, including the Intracisternal A Particle

(IAP) retrotransposons. Indeed, mosaic activity of an IAP element due to methylation state at the agouti locus is responsible for the range of phenotypes in those mice[27].

Non-LTR retrotransposons that contain a poly(A) include the autonomous Long Interspersed and non-autonomous Short Interspersed Nuclear Elements (LINE and SINE). LINE-1 (L1) retrotransposons are the only remaining active autonomous transposons in humans. Only about 100 full-length copies of L1 are retrotranspositionally competent; however, full-length and truncated L1 elements combined account for about 20% of human genome [26,28,29]. This is due to negative selection against maintaining active and potentially deleterious L1 elements in the genome[30]. Interestingly, mice are threatened by almost 3,000 active elements. The L1 family in mice can be subdivided based on promoter type, with the majority of full-length elements belonging to the Md_A and Md_TF sub-families (Fig. 1-2A)[31,32]. A generic full-length human L1 element is 7 kb long with two open reading frames, *ORF1* and *ORF2* that encode retrotransposition machinery (Figure 1-2B) [33-35]. *ORF1* encodes an RNA-binding protein that functions as a major component of the L1 ribonucleoprotein particle (RNP) and a nucleic acid chaperone during L1 insertion [36-38]. ORF1p preferentially binds the L1 RNA from which it was derived in cis, but is also capable of binding other L1 RNAs and mRNAs[39]. ORF2p has endonuclease and reverse transcriptase activities that are crucial for L1 retrotransposition [40,41]. A generalized mechanism of L1 retrotransposition involves single-stranded DNA nicking by ORF2p, annealing of the L1 poly(A)-tail to the released DNA strand, which also provides a 3'-OH end to prime L1 reverse transcription, second strand synthesis and ligation of the final product to genomic DNA [40,42,43] (Figure 1-2C). A few missing pieces of this mechanism known as target-primed reverse transcription include: (1) how the RNA strand is removed from the RNA:DNA hybrid, (2) whether this requires host factors, and (3) how the second DNA break is made to complete retrotransposition.

Thus far, I have described a beneficial role for transposons in promoting genetic diversity and adaptation, but active transposons can also be tremendously harmful genomic parasites with their expression serving as a proxy for pathology. The first evidence of active transposition leading to disease

was identified in patients with Hemophilia where a truncated L1 insertion in exon 14 of factor VIII gene was observed in two different cases that mapped back to a maternal full-length L1 sequence[44]. Growing evidence for L1 overexpression in cancer tissues and with autoimmune diseases have been reported, but whether L1 serves as a passenger or driver mutagen for the diseases is still an open question[26].

Host cells go to great lengths to silence L1

The pathology of L1 is not only through successful mutagenic insertions, but also through the genome instability created during L1 overexpression and retrotransposition attempts. L1 overexpression in mammalian cultured cells causes accumulation of DNA damage, activation of DNA damage checkpoints and cell death [41,45]. Interestingly, both reverse transcriptase and endonuclease activities of ORF2p individually have detrimental effects to cell viability [45]. To protect genomes from the detrimental L1 activity, multiple epigenetic and post-transcriptional mechanisms of transposon silencing have evolved [35,46-48]. Often, these mechanisms work together and compensate for each other, especially in germ cells, where controlling L1 activity is especially critical as any resulting mutations will be heritable.

Chromatin structure is a robust primary epigenetic defense mechanism against L1 by modifying histones and DNA. DNA methylation has been historically involved in transcriptional repression and transposon silencing. Unless there is destabilization of the epigenome either during development or due to environmental stress, DNA methylation robustly represses L1 expression. In Section 1.1, the impact of epigenetic reprogramming in early embryonic development and PGCs on derepression of L1 was described in detail. In such cases, repressive histone modifications and their necessary chromatin-modifying enzymes step in to repress L1 expression[49]. Particularly, histone H3 lysine 9 methylation that is characteristic of constitutive heterochromatin and repression of repetitive elements. Histone H3K9me3 is enriched on active LINE elements compared to truncated elements. Histone H3K9me2 represses L1 in PGCs, between E7.5 and E12.5, concurrent and likely compensating for the erasure of DNA methylation on L1 during epigenetic reprogramming in PGCs[50]. Another compensatory mechanism to repress L1

during epigenetic reprogramming in PGCs involves histone H2A/H4R3me2 that inhibits RNA polymerase II binding and written by the chromatin-modifying enzyme PRMT5 when localized to the nucleus[51]. Conditional mutation of PRMT5 in PGCs results in ectopic L1 expression and ultimately loss of germ cells[51].

Post-transcriptional silencing of L1 using small-RNA based systems are active in germ cells, predominantly through PIWI interacting RNAs (piRNAs). 25-32 nucleotides-long piRNAs are generated from transposon transcripts processed by PIWI proteins that belong to the Argonaut family of proteins[52-54]. In mice, there are three PIWI proteins named MILI (PIWIL2), MIWI (PIWIL1), and MIWI2 (PIWIL4) that are involved in piRNA biogenesis along with other factors such as the Tudor-domain-containing proteins (TDRD), but MILI and MIWI alone have RNA slicing ability. Long precursor transposon transcripts are processed into primary piRNAs using PLD6[55]. Secondary piRNAs are derived from transposons transcripts via the ping-pong cycle that involves MILI-bound primary piRNAs as a guide and slicer complex[56]. Secondary piRNAs can interact with either MILI or MIWI2. Ping-pong amplification of piRNAs and transposon cleavage occurs in the cytoplasm. However, piRNAs can play a role for transposon silencing in the nucleus as well by targeting *de novo* DNA methylation machinery to transposon genomic loci[11,12].

A striking sexual dimorphism in piRNA phenotypes exists in mice. In males, piRNAs are essential for germ cell development and fertility, with loss of PIWI proteins or piRNAs leading to upregulation of transposon expression, DNA damage and cell death[56-59]. In females, however, piRNAs are non-essential, with mutation of a number of combinations of PIWI proteins and piRNA biogenesis factors resulting in normal fertility[60,61]. Interestingly, humans have a fourth PIWI protein that is important in oocytes[62]. The rational for the lack of piRNA significance in oocytes is ambiguous, but necessitates a role for other L1 repression mechanisms. One hypothesis is that quality control mechanisms in oocytes are very robust and may rapidly kill oocytes with high L1 expression, negating the need for piRNAs and allowing the functional relevance to be dampened throughout evolution.

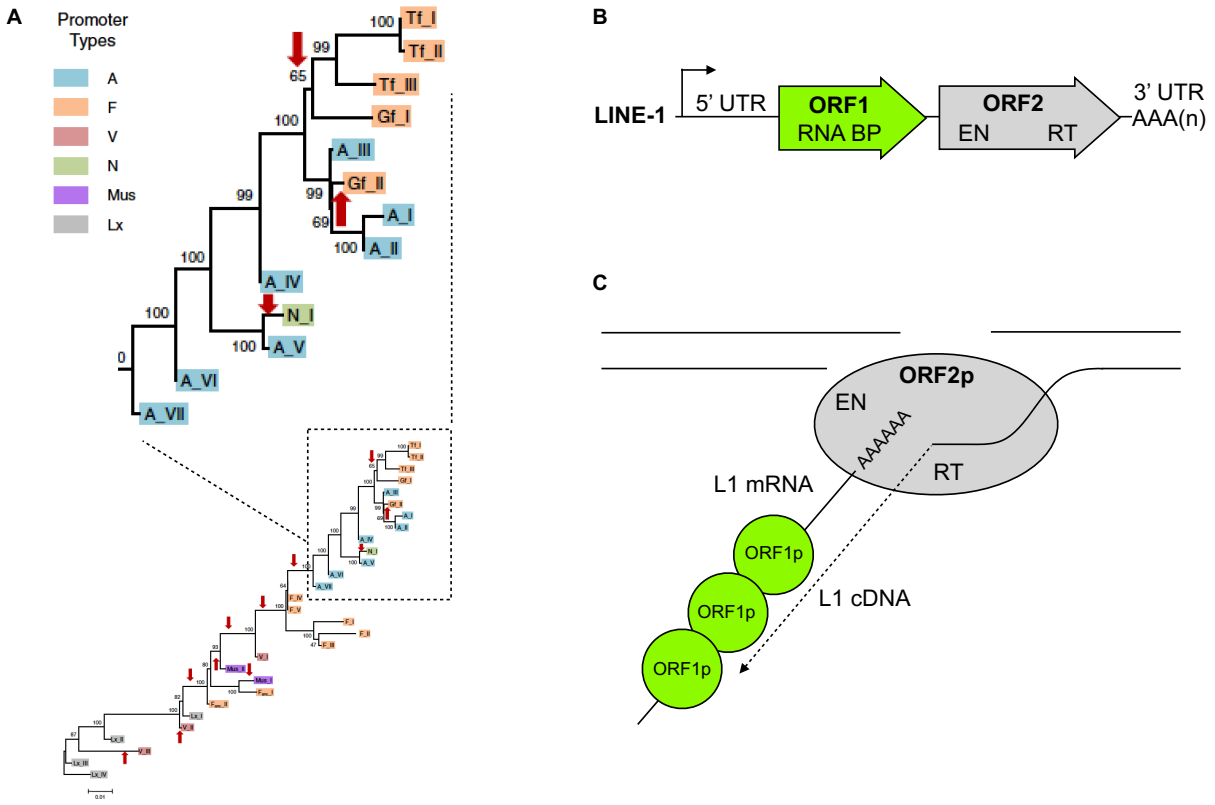


Figure 1-2: Retrotransposon LINE-1 (L1) structure, mechanism of action, and evolution. **A)** Phylogenetic tree of mouse L1 families based on ORF2 sequence (longest non-recombining part)[32]. Shown in the tree are families characterized by their promoter monomer types A, F (encompassing Tf and Gf), V, N, Mus, or Lx. Zoomed in section shows L1 promoter types Tf and A are the youngest, most active elements in the tree. Red arrows show the gain of a new 5' UTR for that family. **B)** Structure of full-length, bicistronic L1 element including promoter (arrow), 5'UTR, two open reading frames ORF1 and ORF2, 3'UTR with poly(A). **C)** Diagram of L1 mechanism of retrotransposition called target-primed reverse transcription (TPRT). L1-encoded ORF2p possesses catalytic activities for TPRT, endonuclease (EN) to make a single-stranded break at the site of attempted insertion and reverse transcriptase (RT) activity that generates a new cDNA copy for insertion from a L1 RNA template initially associated with L1 encoded

ORF1p. The L1 RNA from which ORF1p and ORF2p are derived preferentially binds in cis with its own L1 ORF1p.

1.4 Meiotic prophase I necessitates genome quality control

Germ cells undergo a unique cell division called meiosis during their differentiation into male and female gametes. The purpose of meiosis is to reduce chromosome content from diploid to haploid by two subsequent cell divisions with only one round of replication. In this way, each gamete results with only one copy of each chromosome and after fusion of egg and sperm, the zygote maintains diploid state in the next generation.

Entry and progression through meiotic prophase I (MPI)

After sex-determination, female germ cells enter meiosis while male germ cells undergo a mitotic arrest and do not enter meiosis until postnatal day 9. Female meiotic entry occurs asynchronously over a span of 4 days between E12.5 and E16.5 in an anterior to posterior wave across the gonad[63]. This wave is driven by retinoic acid signaling and the Stimulated by Retinoic Acid gene *Stra8* expression in oocytes. Corresponding expression of meiotic genes such as *Dmc1* followed in this pattern. The wave of meiotic initiation is independent of germ cell number, but dependent on germ cell position and proximity to the meiosis initiation factor in the ovary[64]. Female gonocytes subsequently enter MPI, a prolonged phase of meiosis involving dynamic chromosome rearrangements such as pairing, synapsis, and recombination to promote genetic diversity by creating new combinations of existing alleles. MPI consists of pre-leptotene, leptotene, zygotene, pachytene, and diplotene stages, followed by arrest of oocytes in the dictyate stage. These stages can be identified based on chromosome morphology (using markers such as SYCP3) and the degree of DNA break repair (using markers such as histone gH2AX that marks DNA damage) (Fig 1-3A).

In pre-leptotene, leptotene and zygotene stages, homologous chromosomes must pair and synapse, a process that relies on the formation of programmed DNA double stranded breaks (DSBs) that

are also critical for genetic recombination[65]. A major player in these processes is Spo11, a topoisomerase that catalyzes the DSBs at specific recombination "hot spots" in the genome[66,67]. Interestingly, Spo11-independent DSBs are also observed that are likely due to L1 activity during MPI as loss of piRNA pathway protein Maelstrom leads to derepression of L1 and an increase in Spo11-independent DSBs in *Mael*$^{-/-}$*;Spo11*$^{-/-}$ spermatocytes[59,68,69]. Loss of Spo11 results in depletion of oocytes by the first wave of folliculogenesis[70]. Genetic recombination occurs during the pachytene stage of MPI and is completed in mid-pachynema. Interestingly, due to the repeat nature of L1, it is often used as a recombination substrate for DNA repair. A DNA damage checkpoint is employed post-recombination to ensure the efficient repair of programmed DSBs. If able to pass the checkpoint, MPI cells enter the diplotene stage where homologs begin to desynapse, but remain physically connected via chiasmata that provide stability during the first meiotic division. Oocytes then arrest in the dictyate stage, but with persisting transcriptional activity to accumulate a stockpile of RNA for later use during early embryogenesis. Only after ovulation induced by gonadotropin and steroid hormone signaling do oocytes exit this arrested stage and undergo meiosis I. This is followed by meiosis II upon fertilization. In both of the meiotic divisions, only one cell receives the majority of the cytoplasm to grow into the mature oocyte while the other cells are reduced to polar bodies.

The DNA damage checkpoint safeguards oocyte genomes

The primary job of germ cells is to transmit high quality genetic (and epigenetic) information to the next generation. Transmission of defective genomes can impact the development and viability of offspring. Meiotic defects are a major contributor to genome instability and can result in large chromosomal rearrangements, loss of genetic material or aneuploidy. DSBs are a particularly dangerous genomic legion and are normally limited to only ten DSBs per cell per day[71]. However, the genomes of meiotic cells are ravaged by DSBs for the programmed breaking and repair of DNA during MPI and for this reason, DSB repair is meticulously regulated. The canonical DNA damage checkpoint is activated in

the mid-pachytene stage of MPI once recombination is completed to detect unrepaired DNA breaks or asynapsis[72,73]. This checkpoint uses a cascade of kinases including ATR and Checkpoint kinase 2 (CHK2) to relay information regarding unrepaired DNA breaks to apoptotic machinery, p53 and p63, through phosphorylation to eliminate damaged oocytes (Fig. 1-3B)[74]. The importance of the DNA damage checkpoint can be demonstrated using CHK2 mutant mice, which ignore the presence of lethal amounts of DNA damage, but are otherwise viable and fertile. For example, upon irradiation, oocytes are normally killed due to the copious amounts of DNA breaks. However, in *Chk2*$^{-/-}$, these oocytes are rescued. This effect is also observed by introducing *Chk2*$^{-/-}$ into *Dmc1*$^{-/-}$ oocytes that are incapable of repairing meiotic DSBs[74]. Therefore, such checkpoints are truly critical for establishing the quality and quantity of oocytes in the ovarian reserve. Other reported mechanisms to safeguard oocyte genomes include a meiotic silencing model that depends on the level of gene expression to detect and eliminate oocytes with abnormal numbers of chromosomes[75].

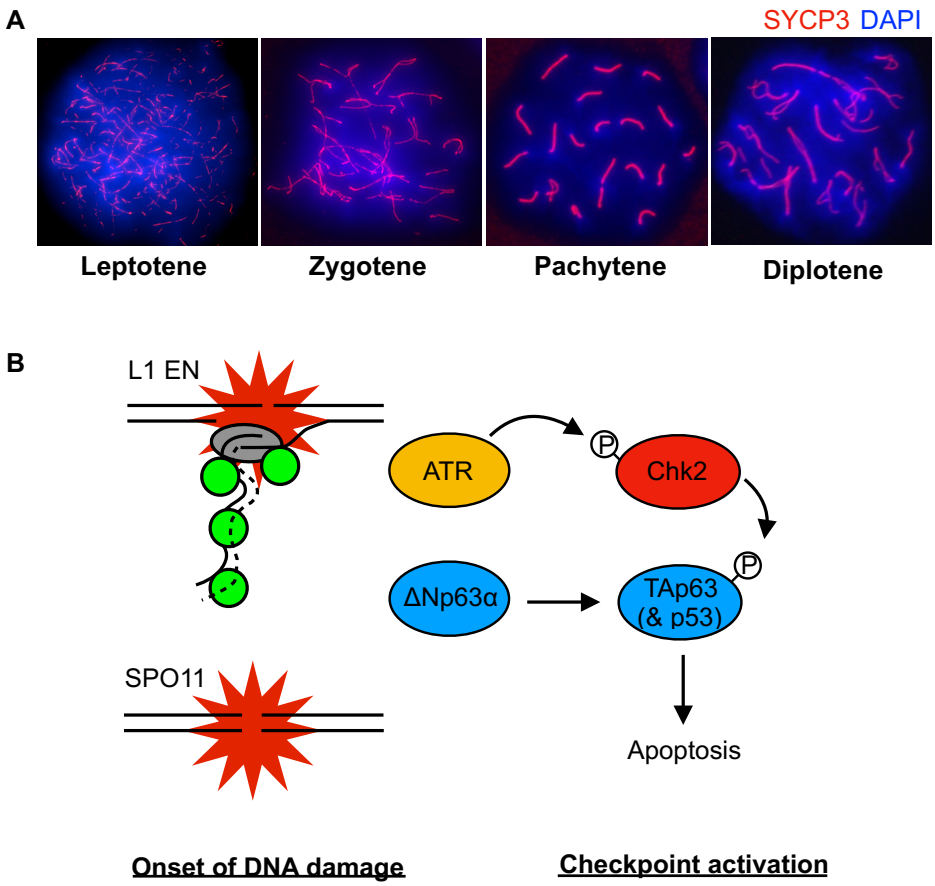


Figure 1-3: Meiotic prophase I (MPI) substages and the DNA damage checkpoint. **A)** Representative images of oocytes in Leptotene, Zygotene, Pachytene, and Diplotene stages of MPI based on synaptonemal complex morphology. Synaptonemal complex labeled by SYCP3 and DNA in DAPI. In Leptotene, chromosomes begin to pair, in zygotene homologous chromosomes synapsis, in Pachytene, chromosomes are fully synapsed and perform genetic recombination, in diplotene, homologous chromosomes begin to desynapse, but do not fully separate. After, oocytes arrest in Dictyate until the first meiotic division. **B)** Molecular players in the canonical DNA damage checkpoint in oocytes. The genome is exposed to the onset of DNA damage from programmed meiotic breaks by Spo11 or L1 endonuclease activity (EN) in Leptotene and Zygotene stages. After meiotic recombination in the Pachytene stage, ATR detects unrepaired DNA breaks and in response, phosphorylates the downstream kinase CHK2. CHK2 in

turn phosphorylates apoptotic machinery p53 and the active isoform of Tap63 that switches from the dominant negative isoform in the Pachytene stage.

1.5 Fetal oocyte attrition shapes the ovarian reserve

Oogenesis programs are highly complex and vary across metazoan species to accommodate the diverse reproductive strategies observed in nature. Some species owe their success to producing large numbers of offspring, especially invertebrates like *D. melanogaster* and *C. elegans*, in hopes that some offspring will survive. This strategy relies on robust germline stem cell-based oogenesis programs to continuously renew the egg supply throughout reproductive life [76]. In contrast, mammalian females are devoid of germline stem cells, and instead, are endowed with a non-renewable ovarian reserve to ultimately produce few offspring [77-79]. Generation of competent mammalian oocyte for fertilization takes considerable energy and a lifetime of preparation and quality control.

Female fertility and reproductive lifespan in mammals critically depend on the size and quality of the ovarian reserve of primordial follicles, a supply of arrested oocytes and associated somatic cells established by birth [80-83]. Paradoxically, the ovarian reserve of about 2 million primordial follicles at birth reflects only a smaller share (~20%) of all oocytes initially specified in the fetal ovary, and is further depleted throughout the woman's reproductive lifespan [84,85] (Fig. 1-4A). The majority of fetal oocytes generated are lost by evolutionarily conserved fetal oocyte attrition (FOA)[84,86-89]. Primordial germ cells expand their population slowly during migration to the gonad and more rapidly once the gonad is reached, where they are subsequently determined as female oocytes prior to FOA. Understanding why females create only to eliminate this large proportion of oocytes is of great significance.

Programmed cell death is a conserved feature of oogenesis. Since the discovery of FOA in the 1960s, however, the mechanisms and underlying developmental or physiological rationale of programmed oocyte death in mammals remains debated. Over the ensuing decades, inadequate oocyte support (death by neglect), altruistic behavior of sister oocytes to support the survival of few (death by

self-sacrifice), and meiotic DNA damage (death by defect) featured prominently in models attempting to explain this phenomenon [90-92].

Death by self-sacrifice

The death by self-sacrifice hypothesis suggests that within a cyst of connected oocytes arising by germ cell proliferation with incomplete cytokinesis, one is fated to be the survivor, while the remaining altruistically donate their cytoplasmic components through intercellular bridges to support the health and growth of the survivor (Fig. 1-4B). This hypothesis is based on the nurse cell dumping process in *Drosophila melanogaster* oogenesis. During early oocyte development in *Drosophila*, connected cysts comprised of 16 germ cells are formed; one of which acquires oocyte fate, while the remaining 15 become nurse cells. Nurse cells undergo programmed cell death and donate their cytoplasmic components through intercellular bridges to the surviving oocyte, a process that is essential for *Drosophila* fertility[93].

Mouse germ cells also undergo proliferation with incomplete cytokinesis to form interconnected oocyte cysts[94]. Oocytes within each cyst are connected by intercellular bridges that are maintained by the protein TEX14[95]. Whether oocyte intercellular bridges in mice function similarly to those *Drosophila*, or are vestigial, remains an open question. By electron microscopy, intercellular bridges between mouse oocytes decrease in diameter between E14.5 and E17.5 as cysts break down into individual primordial follicles[96]. Organelles are observed in these bridges, but there is no evidence supporting the requirement for transport of cytoplasmic materials for oocyte viability, as *Tex14* mutant female mice are fertile[97]. In contrast, male *Tex14* mutants are infertile[95]. Currently, the leading hypothesis with respect to death by self-sacrifice in mouse oocytes is that transport of organelles to surviving oocytes and corresponding growth is mediated through phagocytosis, not intercellular bridges[96].

Death by neglect

The death by neglect hypothesis suggests that somatic cell-derived oocyte survival factors are limited, and therefore, oocytes must compete with each other for these factors or suffer apoptosis (Fig. 1-4C)[90]. This type of developmental program has been observed during formation of the central nervous system, where a large number of neurons are killed as a consequence of limited growth factors to support their survival[98]. In oocytes, death by neglect is supported by changes in the oocyte endowment upon manipulation of the response to proinflammatory cytokine germ cell survival factors. Specifically, mutation of *caspase-11*, a mediator of the inflammatory response, results in a decreased oocyte endowment at birth[99]. Mutation of *caspase-2*, an executor of apoptosis in response to cytokine insufficiency, can rescue oocyte death in *caspase-11* mutants[99]. Complementary studies using *ex vivo* culture of wild-type ovaries show that when insufficient amounts of growth factors or cytokines in the media are provided such as stem cell growth factor (SCF) and leukemia inhibitory factor (LIF), increased germ cell apoptosis is observed[100]. Therefore, amounts of such cues are important for fetal oocyte survival.

Death by defect

The death by defect hypothesis is related to oocyte quality control where only the most competent oocytes will be permitted to survive and undergo folliculogenesis (Fig. 1-4D). While a larger oocyte endowment at birth may increase the reproductive lifespan of a woman, the retention of poor quality of follicles can lead to offspring with disease-causing mutations. As described in Section 1.3, oocyte quality from a genetic perspective is most vulnerable during MPI, where complex chromosome rearrangements and programmed DSBs ravage the genome. Aside from the canonical DNA damage checkpoint through CHK2 as described, other cell death pathways have been implicated in the quality control of oocytes with persisting meiotic defects. First, Caspase 9 involved in the mitochondrial apoptotic pathway has been shown to be required for FOA. In *Casp9*$^{-/-}$ mutant ovaries, oocyte number is significantly increased by E19.5, which corresponds with increased gH2AX levels in surviving oocytes[101,102]. Additionally,

inhibition of X-linked inhibitor of apoptosis protein (XIAP) that prevents cell death in the presence of CASP9 has an opposing negative effect on oocyte survival. However, these effects are only temporary, and by E23.5, oocyte number returns to the wild-type level. Lastly, RNF212 is an essential protein that is required for crossover formation, but also impedes DSB repair and acts to promote the elimination of oocytes with meiotic defects. Mutation of RNF212 results in an increased oocyte supply that contain oocytes with less DSBs[103].

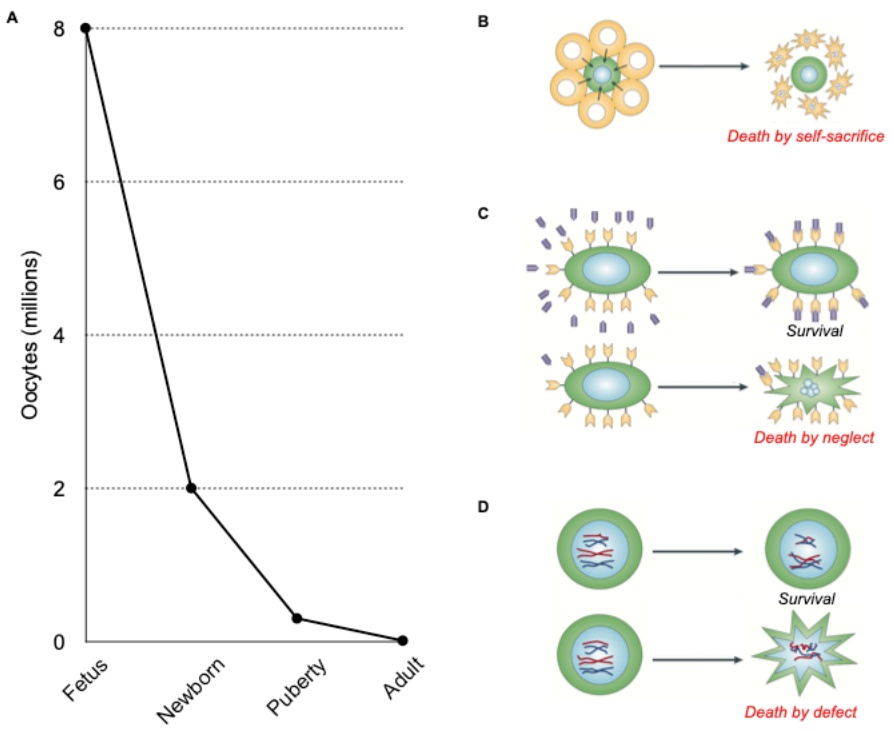


Figure 1-4: Hypotheses surrounding mechanisms and rationale for fetal oocyte attrition. **A)** Chart depicting the depletion of oocytes in humans over lifespan at fetus, newborn, puberty, and adult ages. The steepest decline in oocyte number is between fetus and newborn stages, the time of FOA. **B-D)** Tested hypotheses for mechanisms of FOA[90]. **B)** Death by self-sacrifice occurs when one oocyte in a cyst of connected oocytes is fated to be the survivor, and the remaining sister oocytes donate their cytoplasmic materials to support the survivor while they perish. Orange cells are those that die and green cell is surviving oocyte. **C)** Death by neglect suggests that a limited amount of growth factors or hormones are produced to support oocytes. Purple arrows are growth factors that are limited and orange arrows are receptors. Oocytes that receive enough of the growth factor, while oocytes that die do not have an adequate amount. **D)** Death by defect suggests that FOA eliminates oocytes with excess DNA damage. Shown is a surviving oocyte with intact chromosomes compared to a dying oocyte with broken chromosomes.

1.6 A role for L1 in FOA

Another hypothesis of FOA related to death by defect ties together the above described oogenesis programs from 1.1, 1.2, and 1.3 to implicate retrotransposon LINE-1 (L1) in FOA [69] (Fig. 1-5A-B). In prior work from our lab, a striking correlation of L1 with DNA damage and fate of fetal oocytes was reported based on expression of L1 ORF1p [69] (Fig. 1-5C). The role of L1 in FOA was demonstrated using mouse models that overexpressed L1, including induction of Tet-ORFeus, a synthetic L1 element and mutating the RNA binding protein Maelstrom involved in piRNA biogenesis in males[69,104]. Both models showed an increase in L1 ORF1p expression in fetal oocytes and an increased occurrence of FOA. The role for L1 in FOA was corroborated by experiments to inhibit L1 activity using the reverse transcriptase inhibitor azidothymidine (AZT), that has been shown to effectively block reverse transcriptase activity of L1 ORF2p in vitro [105,106]. A remarkable protective effect on oocyte survival was observed with treatment of pregnant females with AZT on fetal oocytes; however, this effect was temporary and limited to the early stages of MPI.

Cumulatively, these observations introduced L1 as a strong influence for death by defect FOA, suggesting that Nature has evolved a way to produce enough excess oocytes to compensate for this window of L1 activity. Further, it suggests that heterogeneity in L1 expression levels among the fetal oocyte population exists and dictates oocyte fate and quality (Figure. 1-5D). Many outstanding questions surrounding L1-driven FOA remain. First, by what mechanisms does L1 cause FOA? The transience of FOA prevention by AZT indicates the presence of other pathways killing oocytes. Second, how does L1 cause cytotoxicity? In AZT-treated oocytes that did not undergo FOA, increased incidence of common meiotic defects such as DNA damage and asynapsis were observed compared to controls, suggesting that L1 activity was causing damage or disrupting meiotic events by attempting to insert into meiotic breaks[69]. However, the effect of AZT on reverse transcriptase suggests the possibility of reverse transcription intermediates such as RNA:DNA hybrids or single-stranded cDNA after RNA hydrolysis could lead to genome instability and FOA.

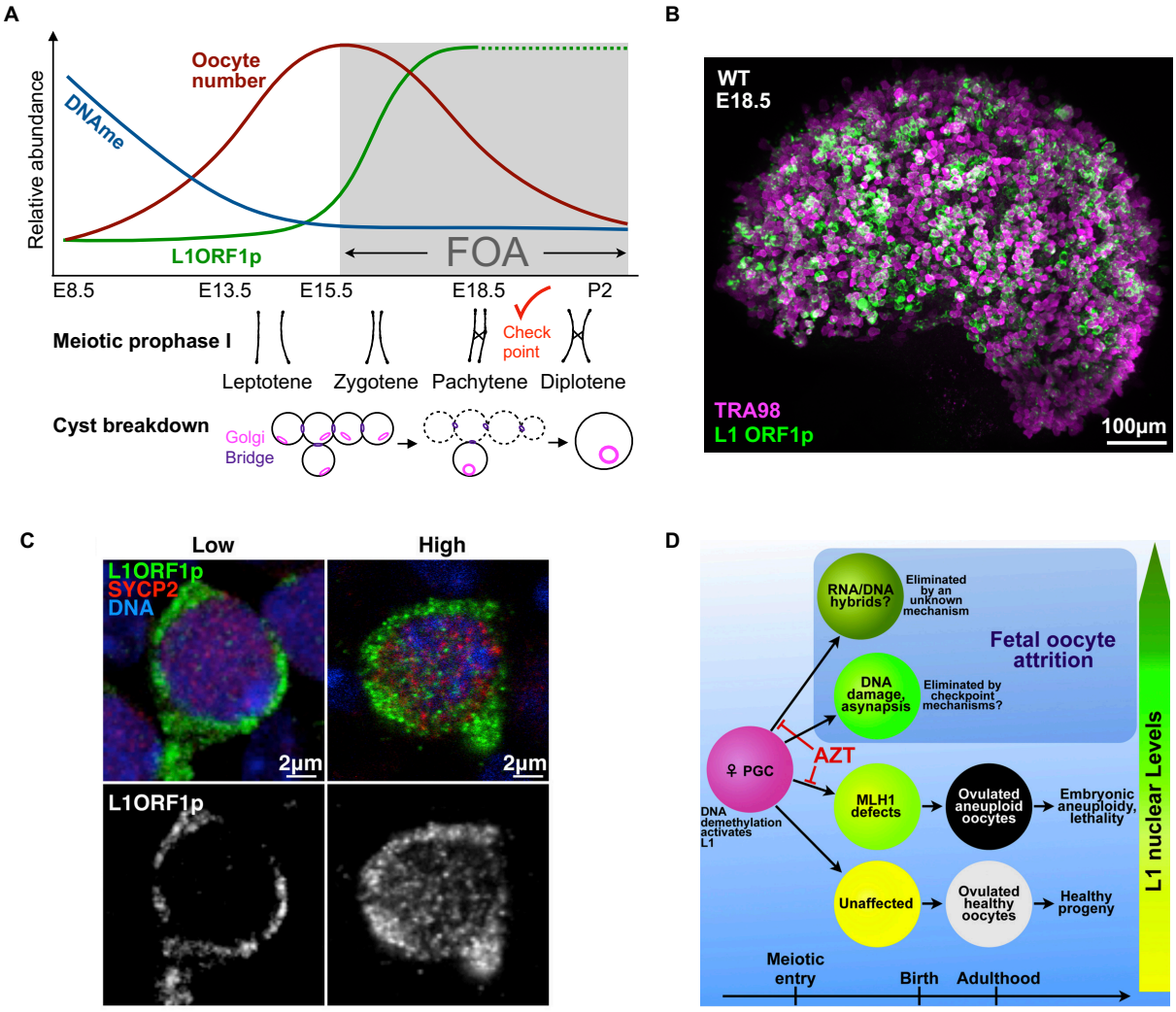


Figure 1-5: L1 expression in fetal oocytes determines their fate. **A)** Model depicting developmental events concurrent with FOA. Genome-wide DNA demethylation (blue line) for epigenetic reprogramming occurs between E8.5 and E13.5 that includes L1 elements. L1 expression is permitted as a consequence of demethylation, with L1 ORF1p observed by E15.5 and increasing in abundance by E18.5. By P2, the average L1 ORF1p level in the oocyte population is decreased, but it is unclear whether this is due to loss of oocytes expressing high levels of L1 ORF1p, or another silencing mechanism. L1 ORF1p expression is anti-correlated with the oocyte supply size (dark red line). Fetal oocytes enter meiotic prophase I, subdivided into Leptotene, Zygotene, Pachytene, and Diplotene stages and involving a strict DNA

damage checkpoint at the mid-pachytene stage to eliminate oocytes with unrepaired DNA breaks (red checkmark). Finally, oocyte cyst breakdown is concurrent with FOA, involving closure of intracellular bridges between oocytes and accumulation of organelles such as Golgi in surviving oocytes. **B)** Whole-mount immunofluorescence labeling of germ cell marker TRA98 and L1 ORF1p in E18.5 ovaries. Scalebar:100μm. **C)** Examples of individual oocytes with low and high nuclear L1 ORF1p expression[69]. **D)** Model depicting conclusions and future directions for FOA as a quality control response to the level of L1 expression[69]. Oocytes with highest levels of L1 (dark green and lime green) undergo FOA due to genome instability caused by DNA damage or intermediates of reverse transcription given the sensitivity to AZT. Oocytes with intermediate levels of L1 that survive may result in defective progeny such as those with crossover defects in meiosis leading chromosome abnormalities (MLH1 defects). Oocytes with lowest levels of L1 (yellow) go on to form healthy progeny.

1.7 Thesis aims

The work of this thesis aims to broadly understand the mechanisms and biological significance of FOA in light of the role for L1 activity. Importantly, I aim to understand whether preventing FOA, if possible, would benefit or perturb fertility of the female upon reproductive age. Further, understanding the developmental basis behind heterogeneity in L1 expression in fetal oocytes is addressed, as level of L1 in an individual oocyte determines its fate during FOA. A pivotal prerequisite to achieving these aims was to develop a system to prevent FOA in order to study the population of oocytes normally eliminated.

This thesis is divided into four parts. Chapter 2 aims to identify the mechanisms of FOA, their triggers related to L1 activity and downstream cell death pathways. Chapter 3 investigates the biological relevance of FOA and therefore a physiological role for L1 derepression in oocytes. Chapter 4 goes back to the beginning to understand how differential levels of L1 are established that dictate fetal oocyte fate. Last, discussion of the future directions and implications of this work will follow.

CHAPTER 2

Mechanisms of fetal oocyte attrition

From "Maximizing the ovarian reserve by evading LINE-1 genotoxicity" (Fig. 1 and 2).

Introduction

In this chapter, I investigated the molecular triggers and mechanisms underlying L1-driven FOA. The deleterious effects of L1 activity on fetal oocyte genome integrity and cell viability are likely attributed to the two catalytic activities of L1 ORF2 protein that are required for its autonomous retrotransposition, endonuclease and reverse transcriptase activities, and how they disrupt events during MPI. Indeed, both endonuclease and reverse transcriptase activities can independently reduce cellular viability in vitro by apoptosis or by inducing a senescence-like state[45].

The ideal approach to determine the molecular mechanisms of L1-driven FOA would be to genetically manipulate L1 in oocytes. However, in mice, such models are not yet feasible and major caveats have been encountered in attempts disrupt L1 activity. For example, mutating L1 elements using CRISPR-Cas9 creates lethal amounts of DNA breaks given that L1 is a repeated sequence and highly abundance in the mouse genome. Strategies using cytosine or adenine base-editors overcome this problem while still perturbing L1 activity, but contain many off-target mutations. Additionally, knockdown of L1 elements by RNA interference result in off-target effects as L1 elements can be found in certain genes. For these reasons, I chose to block L1 activity using a small molecule inhibitor of reverse transcriptase. Reverse transcriptase inhibitors have proven effective to prevent L1 retrotransposition *in vitro*[107,108]. Further, AZT can be transmitted between the placenta barrier and has been shown to positively influence the course of retroviral disease in neonates when administered to the pregnant mother[109]. Administration of AZT to pregnant mothers can also prevent FOA, consistent with the hypothesis that AZT inhibits the reverse transcriptase activity of L1 in oocytes[69]. However, the mechanism of action and specificity to L1 reverse transcriptase remain unknown. Further, this positive effect of AZT on FOA was temporary, with oocyte number returning to that of untreated mice shortly after birth.

L1 reverse transcriptase activity and FOA

The mechanistic link between L1 reverse transcriptase activity and oocyte death is critical for

understanding the role of L1 in FOA. The protective effect of AZT treatment in fetal oocytes suggests that the process of reverse transcription is cytotoxic, either by facilitating complete L1 insertions or by creating reverse transcription intermediates. Complete L1 retrotransposition events are rare, and therefore unlikely to drive massive amounts of FOA. When measured, only one in every eight embryos examined possessed a *de novo* L1 insertion[110]. On the other hand, RNA:DNA hybrids and ssDNA can be cytotoxic upon accumulation, and are linked to the host immune response due to their resemblance to viral nucleic acids. For this reason, I chose to interrogate the role for L1 reverse transcription intermediates and the innate immune system as the mechanism of AZT-sensitive FOA.

To test whether L1 reverse transcription intermediates and the innate immune system are the mechanistic link between L1 reverse transcriptase activity and FOA, I took both candidate and unbiased approaches. First, I interrogated candidate host factors involved in nucleic acid processing with the goal of manipulating the amount of L1 intermediates in oocytes. L1 ORF2p does not possess the catalytic activities necessary to metabolize reverse transcription intermediates, so host enzymes may be hijacked for this process. RNA:DNA hybrids and ssDNA are generated in the cell during a number of different processes aside from L1 reverse transcription, including transcription, DNA repair, and lagging strand DNA synthesis, and the cell has acquired enzymes to process them[111]. Candidate enzymes tested were RNase H2, which hydrolyzes the RNA strand of RNA:DNA hybrids, and three-prime repair exonuclease 1 (Trex1), which cleaves ssDNA flap structures. I also took an unbiased gene expression profiling approach to uncover cell death pathways active during FOA that are reduced upon AZT treatment, especially related to innate immune responses. I focused on the innate immunity because accumulation of RNA:DNA hybrids and ssDNA, such as that seen as a consequence of mutations in *RNase H2* and *Trex1,* induce a type I interferon response through the cGAS/STING pathway[112-115]. These mutations are linked to immune disorders in humans such as Aicardi-Goutières syndrome (AGS), a neuroinflammatory disorder[112]. More specific to L1, a type I interferon response leading to neural toxicity was observed in *Trex1* mutant pluripotent stem cells upon their differentiation into neurons, due to accumulation of L1

ssDNA in the cytoplasm[113]. Interestingly, this immunogenic response in neurons was prevented by treatment with reverse transcriptase inhibitors[113]. Additionally, immune cell lineages such as macrophages occupy the female and male gonad, and play a role in reproductive development[116,117].

L1 endonuclease activity and FOA

Given its direct impact to DNA integrity, I hypothesized that the endonuclease activity of L1 ORF2p also contributes to FOA. The observation that positive effects of reverse transcriptase inhibitor AZT treatment on FOA are temporary suggests that another mechanism governs FOA. Additionally, the endonuclease activity of L1 ORF2p is not impacted by AZT, and DNA damage will still be induced. Indeed, the timing of FOA in the presence of AZT correlates with the activation of the DNA damage checkpoint in the mid-pachytene stage of MPI. Oocytes are tolerant of DNA DSBs prior to the checkpoint activation, because programmed DNA breaks for meiotic recombination are abundant and not yet repaired.

To test the role of DNA damage through L1 endonuclease activity, either by creating more breaks or inhibiting repair of programmed meiotic breaks, I chose to perturb the DNA damage checkpoint through mutation of CHK2. Inhibition of L1 endonuclease activity using a drug would be too blunt of a tool, as, unlike reverse transcriptase, essential cellular processes utilize endonuclease activity and would also be affected. CHK2 is a crucial player in culling oocytes that contain excess unrepaired DNA DSBs by way of the DNA damage checkpoint[118]. *Chk2* mutant mice are viable, fertile, and, unlike in wild-type mice, unrepaired DSBs are ignored and do not cause oocyte death [74,119]. This was shown by irradiation and DNA repair mutants that normally result in infertility, but can be rescued by adding mutation of CHK2. However, whether CHK2 plays a role in FOA under physiological conditions is unknown.

Cumulatively, I identified two phases of FOA associated with the two activities of L1 ORF2p, endonuclease and reverse transcriptase. I uncovered the triggers, mechanisms, and timing with respect to MPI events that underlie these mechanisms to gain a comprehensive picture of L1-driven FOA.

2.1 Intermediates of L1 reverse transcription are reduced upon AZT treatment

To investigate whether accumulation of L1 reverse transcription intermediates triggers FOA by way of the innate immune response, I isolated and measured L1 ssDNA in untreated and AZT-treated oocytes and ovarian somatic cells (a negative control that does not express L1). To isolate ssDNA, I first treated total DNA with RNase A to remove residual mRNA and RNA within RNA:DNA hybrids. Subsequently, RNase A-treated DNA (input) was treated with double-stranded (ds) DNase to remove genomic DNA, leaving ssDNA (naked or once part of an RNA:DNA hybrid) (Fig. 2-1A). To measure L1 ssDNA, quantitative PCR was performed to determine the abundance of L1 *ORF1* DNA in dsDNase-treated samples relative to input samples. The amount of L1 ssDNA is minimal relative to L1 input DNA considering that L1 is a repeat and comprises ~20% of the mouse genome. Indeed, I reproducibly observed a reduction of L1 *ORF1* ssDNA in AZT-treated oocytes compared to untreated oocytes (Fig. 2-1B). Although this decrease was slight, the samples we compared are untreated oocytes containing sub-lethal levels of L1 activity and AZT-treated oocytes where L1 activity is suppressed, so extremely high levels of L1 ssDNA are not expected in either case. Importantly though, a large reduction in L1 *ORF1* ssDNA was observed in untreated ovarian somatic cells that do not express L1 compared to untreated oocytes that was not seen with corresponding AZT-treated samples (Fig. 2-1B). A number of other important controls were included: oocytes treated with RNase A followed by both dsDNase and ssDNase (P1) to eliminate all DNA as a negative controls and oocytes treated with RNase A and P1 alone to understand the proportion of L1 genomic DNA to L1 ssDNA in the samples (Figure 2-1A, B; Table 2-1). Together, our experiments validate that AZT treatment reduces the presence of L1 reverse transcription intermediates and supports the hypothesis that FOA is triggered by accumulation of L1 reverse transcription intermediates above a lethal threshold.

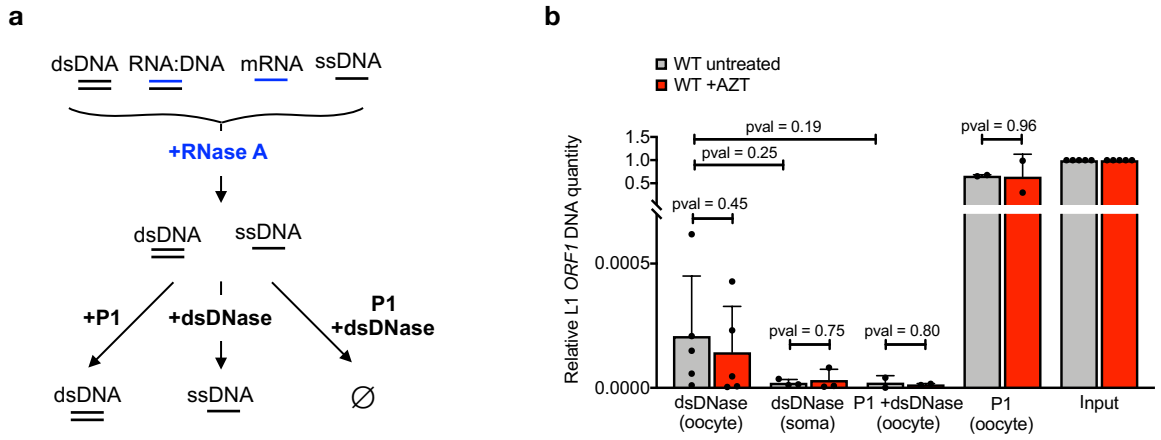


Figure 2-1: Isolation and quantification of L1 single-stranded DNA. **A)** Schematic for isolating L1 reverse transcription intermediates. Total DNA isolated from untreated and AZT-treated sorted oocytes and somatic cells after RNase A treatment to remove mRNA and RNA within RNA:DNA hybrids. ≥6 pairs of WT CD1 E16.5-E17.5 ovaries from single litter per sample. Subsequently, RNase A-treated DNA (input) was treated with double-stranded (ds) DNase to isolate single-stranded (ss) DNA (n=5 for oocytes and n=3 for somatic cells). Input DNA was also treated with the ssDNase P1 to control for dsDNA relative to ssDNA (n=2). Input DNA was also treated with P1 and dsDNase to control for residual dsDNA after dsDNase treatment (n=2). **B)** Relative quantity of L1 *ORF1* DNA by quantitative PCR. Samples first normalized to the single copy gene *Ifnb1,* then to L1 *ORF1* of input DNA. Dots indicate individual biological replicates; data are mean +SD. Stats for comparison of untreated to AZT-treated samples by two-tailed paired Student's t-test, ns $p>0.05$; stats for comparison of WT untreated and AZT-treated, dsDNase-treated oocytes to WT untreated and AZT-treated negative controls by Mann-Whitney test, ns $p>0.05$.

q-PCR measurements of L1 *ORF1* DNA after dsDNase and P1 treatments of untreated (UT) and AZT-treated oocyte and somatic cell DNA

Mean Cq *Ifnb1*	Oocyte UT 1	Oocyte UT 2	Oocyte UT 3	Oocyte UT 4
dsDNase	0	36.658144	0	38.11229272
dsDNase + P1	NA	NA	NA	0
P1	NA	NA	NA	27.02460683
Input	25.3581305	25.5554647	24.9007628	26.47881674
Mean Cq* L1 *ORF1	Oocyte UT 1	Oocyte UT 2	Oocyte UT 3	Oocyte UT 4
dsDNase	25.1636675	24.8221068	26.1474619	24.70875693
dsDNase + P1	NA	NA	NA	28.63429386
P1	NA	NA	NA	14.67333475
Input	12.4547193	12.5928568	12.06382714	14.04951693
Relative L1 *ORF1* DNA expression to *Ifnb1* input DNA	Oocyte UT 1	Oocyte UT 2	Oocyte UT 3	Oocyte UT 4
dsDNase	1.15	1.67	0.42	3.41
dsDNase + P1	NA	NA	NA	0.2253
P1	NA	NA	NA	3590.58
Input	7696.57	8023.41	7332.05	5518.27
Relative L1 *ORF1* DNA expression to L1 *ORF1* input DNA	Oocyte UT 1	Oocyte UT 2	Oocyte UT 3	Oocyte UT 4
dsDNase	0.00014942	0.00020814	5.72828E-05	0.000618038
dsDNase + P1	NA	NA	NA	4.0828E-05
P1	NA	NA	NA	0.650671315
Input	1	1	1	1
Statistical Analysis comparing untreated vs. AZT-treated	Mean Oocyte UT +dsDNase vs. Mean Oocyte +AZT +dsDNase			
p-value	0.45			
t	0.8253			
Signficance	ns			
Statistical Analysis comparing to negative controls	Mean Oocyte UT +dsDNase vs Mean Soma UT +dsDNase			
p-value	0.25			
Signficance	ns			

*Stats determined from two-tailed paired Student's t-test for comparing untreated to AZT-treated samples
Normalized relative L1 ORF1 DNA expression to L1 ORF1 input DNA values

*Stats determined from Mann-Whitney test for comparing dsDNase-treated oocyte samples to negative controls
Normalized relative L1 ORF1 DNA expression to L1 ORF1 input DNA

Oocyte UT 5	Oocyte +AZT 1	Oocyte +AZT 2	Oocyte +AZT 3	Oocyte +AZT 4	Oocyte +AZT 5	Soma UT 1	Soma UT 2	Soma UT 3
0	0	0	0	0	0	36.37573	0	0
38.72167131	NA	NA	NA	0	0	NA	NA	NA
26.56758027	NA	NA	NA	27.86391791	26.28745665	NA	NA	NA
25.90036861	27.45139755	25.81506696	25.11749696	25.64620693	26.15622847	25.0688	26.18734	27.74895
Oocyte UT 5	Oocyte +AZT 1	Oocyte +AZT 2	Oocyte +AZT 3	Oocyte +AZT 4	Oocyte +AZT 5	Soma UT 1	Soma UT 2	Soma UT 3
29.72704883	32.99998544	30.31926382	26.80741356	24.14275765	25.56312291	26.96255	29.7138	31.57888
33.17776771	NA	NA	NA	29.40485717	29.37645405	NA	NA	NA
13.71579239	NA	NA	NA	14.68846581	13.49797163	NA	NA	NA
13.15562074	14.88633583	13.11311596	12.40940579	12.95380836	13.48224657	12.19189	13.5716	15.13507
Oocyte UT 5	Oocyte +AZT 1	Oocyte +AZT 2	Oocyte +AZT 3	Oocyte +AZT 4	Oocyte +AZT 5	Soma UT 1	Soma UT 2	Soma UT 3
0.0703	0.021	0.044	0.31	2.848	1.52	0.27	0.0872	0.0703
0.0064	NA	NA	NA	0.0743	0.107	NA	NA	NA
4640.29	NA	NA	NA	1992	6472.02	NA	NA	NA
6841.04	6038.61	6700.25	6700.25	6653.97	6562.36	7538.18	6295.04	6251.56
Oocyte UT 5	Oocyte +AZT 1	Oocyte +AZT 2	Oocyte +AZT 3	Oocyte +AZT 4	Oocyte +AZT 5	Soma UT 1	Soma UT 2	Soma UT 3
1.02762E-05	3.47762E-06	6.56692E-06	4.62669E-05	0.000428015	0.000231624	3.58E-05	1.39E-05	1.12E-05
9.3553E-07	NA	NA	NA	1.11663E-05	1.63051E-05	NA	NA	NA
0.678301837	NA	NA	NA	0.29937015	0.986233611	NA	NA	NA
1	1	1	1	1	1	1	1	1
Mean Soma UT +dsDNase vs. Mean Soma +AZT +dsDNase				Mean Oocyte UT +dsDNase +P1 vs. Mean Oocyte +AZT +dsDNase +P1				
0.7504				0.8039				
0.3645				0.3181				
ns				ns				
Mean Oocyte UT +dsDNase vs Mean Oocyte UT +dsDNase + P1				Mean Oocyte +AZT +dsDNase vs Mean Soma +AZT +dsDNase				
0.1905				0.7857				
ns				ns				

Soma +AZT 1	Soma +AZT 2	Soma +AZT 3
36.4377452	0	36.7581074
NA	NA	NA
NA	NA	NA
26.1123197	24.8438762	24.6956256
Soma +AZT 1	Soma +AZT 2	Soma +AZT 3
31.2095411	29.0224249	25.640481
NA	NA	NA
NA	NA	NA
13.4820753	12.1055202	12.0623121
Soma +AZT 1	Soma +AZT 2	Soma +AZT 3
0.029	0.0552	0.521
NA	NA	NA
NA	NA	NA
6338.83	6793.79	6382.9
Soma +AZT 1	Soma +AZT 2	Soma +AZT 3
4.575E-06	8.1251E-06	8.1624E-05
NA	NA	NA
NA	NA	NA
1	1	1
Mean Oocyte UT +P1 vs. Mean Oocyte +AZT +P1		
0.9582		
0.06579		
ns		
Mean Oocyte +AZT +dsDNase vs Mean Oocyte +AZT +dsDNase + P1		
0.8571		
ns		

Table 2-1: Relative quantities for untreated and AZT-treated oocyte and soma samples for ssDNA isolation and controls.

2.2 A candidate approach to identify mechanisms of FOA

Host enzymes that function to resolve RNA:DNA hybrids and ssDNA can be hijacked by L1 for retrotransposition. For example, RNase H2 has been found to function in both promoting and suppressing L1 retrotransposition based on different in vitro assays[120,121]. Transcriptional data generated from E13.5, E15.5, and E18.5 oocytes described in further detail below reveals that expression of RNase H2 and other genes encoding nucleotide-processing enzymes decreased at E15.5 compared to E13.5 or E18.5 stages (Fig. 2-2A). This correlates with the onset of AZT-sensitive FOA. Based on the timing of these transcriptional changes, I hypothesized that perturbing the functions of nucleotide-processing host enzymes in attempt to increase L1 reverse transcription intermediate abundance may increase FOA. In contrast, overexpression of such enzymes during FOA may lessen oocyte death by removing L1 reverse transcription intermediates and phenocopy the effect of AZT.

RNase H2

RNase H2 resolves RNA:DNA hybrids in a number of physiological circumstance in the cell and may be involved in processing reverse transcription intermediates during L1 retrotransposition since L1 ORF2p does not possess autonomous RNase H activity. RNase H2 is comprised of three subunits: A, B, and C, with A possessing the catalytic activity. To determine whether RNase H2 activity is directly involved in FOA, presumably through a role in metabolizing L1 RNA:DNA hybrids, I studied the AGS mouse model containing a point mutation in RNase H2A (G37S) rendering it catalytically inactive[112]. RNase H2A (G37S) mutant embryos are 20% smaller than control littermates, die at birth, and have increased L1 DNA and interferon responsive gene expression[112]. In these mutants, I quantified the number of oocytes per ovary at E15.5, when the maximum number of oocytes is present, E18.5, after the bulk of FOA has occurred, and E13.5 during oocyte proliferation to understand defects unrelated to FOA that may impact oocyte number. I observed that oocyte number per ovary was indeed decreased between the wild type and *RNase H2A (G37S)* mutant at all stages tested (Fig. 2-2B). This result corroborates the reported phenotype that mutant embryos are approximately 20% smaller than the wild type, and suggests

that RNase H2 activity is important for early oocyte development, perhaps in resolving RNA:DNA hybrids in R-loops or removing Okazaki fragment RNA primers during replication of the lagging strand, but not specific to L1 activity or FOA[112].

Since the *RNaseH2 (G37S)* mutant mouse model is a blunt tool that constitutively expresses the catalytically inactive RNase H2 enzyme in all tissues, I sought to generate a model more fit to study RNase H2 activity in FOA. To investigate the role of RNase H2 specifically during FOA, we generated a mouse model to induce expression of all three RNase H2 subunits and a GFP reporter using tetracycline/doxycycline inducible promoter (Fig. 2-2C). I grew and electroporated embryonic stem cells with plasmid containing the RNase H2ACB-GFP transgene (Tg) and a hygromycin resistance gene. After treating ESCs with hygromycin and selecting resistant colonies that took up the plasmid, I treated the RNase H2 Tg ESCs with doxycycline (DOX) to observe GFP reporter expression. GFP positive cell lines were then used for injection into mouse blastocysts to generate RNase H2 Tg mice. I administered DOX to pregnant RNase H2 Tg mice at E12.5 and observed GFP-positive embryos (Fig. 2-2D). In these embryos, I confirmed overexpression of RNase H2A by Western blot (Fig. 2-2E). Finally, I quantified oocytes at E18.5 when half of the oocytes are killed in normal conditions and rescued upon AZT treatment. However, I observed that oocyte number was comparable at all stages between RNase H2 Tg and wild type ovaries (Fig. 2-2F). Cumulatively, our results suggest that RNase H2 may not be involved in processing L1 reverse transcription intermediates.

Trex1

Trex1 metabolizes both dsDNA and ssDNA in a 3' to 5' manner, and has been extensively characterized for its role as an anti-viral enzyme. During viral infection, Trex1 is employed to cleave viral ssDNA present in the cytoplasm to avoid its accumulation and the mounting of an innate immune response[113]. Mutations in *Trex1* are associated with disease of the immune system[113]. In attempt to determine whether L1 ssDNA intermediates utilize Trex1 for their processing in a similar fashion to viral

ssDNA, and at the same time, determine a potential role for Trex1 in FOA, we perturbed the function of Trex1 using the CRISPR/Cas9 system. The Trex1 mutation was generated by electroporating Cas9/sgRNA RNP into CD1 zygotes that were cultured to 2-cell stage and oviduct transferred into recipient CD1 females. The resulting mutation was a deletion of 8 nucleotides and insertion of 2 nucleotides that resulted in a frame shift at codon 119 and immediate termination. I quantified oocytes at E15.5, E18.5, and P2 in *Trex1*$^{-/-}$ compared to control *Trex1*$^{+/+}$ and *Trex1*$^{+/-}$ littermates. Mutant embryos and pups were viable with no significant difference in oocyte number compared to control littermates at any stages tested (Fig. 2-2G). Interestingly, in the CD1 genetic background of these mice, FOA as a whole was insignificant between E15.5 and P2 in both control and mutant conditions. Considering the negligible expression of *Trex1* mRNA in fetal oocytes and the insignificant effect on oocyte survival upon *Trex1*$^{-/-}$, I conclude that Trex1 is not involved in FOA, at least in the CD1 genetic background (Fig. 2-2A, G).

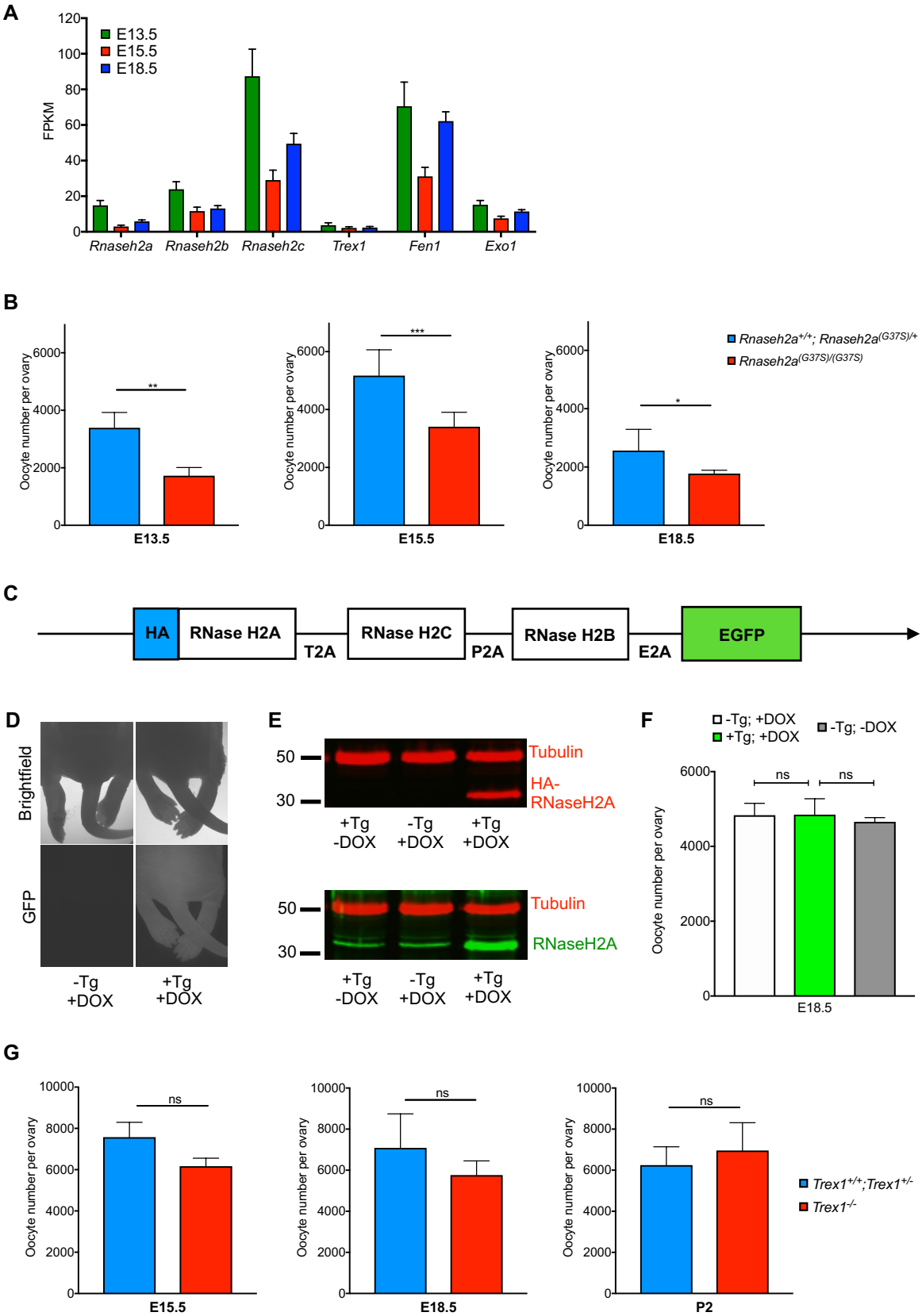

A
E13.5
E15.5
E18.5
FPKM
Rnaseh2a
Rnaseh2b
Rnaseh2c
Trex1
Fen1
Exo1
B
Oocyte number per ovary
E13.5
E15.5
E18.5
Rnaseh2a+/+; Rnaseh2a(G37S)/+
Rnaseh2a(G37S)/(G37S)
C
HA
RNase H2A
T2A
RNase H2C
P2A
RNase H2B
E2A
EGFP
D
Brightfield
GFP
-Tg +DOX
+Tg +DOX
E
Tubulin
HA-RNaseH2A
RNaseH2A
+Tg -DOX
-Tg +DOX
+Tg +DOX
F
-Tg; +DOX
+Tg; +DOX
-Tg; -DOX
ns
E18.5
G
Trex1+/+;Trex1+/-
Trex1-/-
P2

Figure 2-2: Characterizing a role for RNase H2 and Trex1 in FOA. **A)** Expression of RNase H2 A, B, and C mRNA in E13.5, E15.5, and E18.5 oocytes. **B)** Quantification of oocytes per ovary at E13.5, E15.5 and E18.5 in RNase H2A (G37S) mutant and wild-type embryos. **C)** Model of RNase H2ACB-GFP transgene (Tg) containing an HA tag on RNase H2A and peptide cleavage sites between RNase H2A and H2B, H2C and H2B, and H2B and GFP. **D)** DOX-treated RNase H2 +Tg and -Tg E18.5 embryos. **E)** Western blot detection of HA-RNase H2A and endogenous RNase H2A in DOX treated and untreated, +Tg and -Tg E18.5 embryo carcass lysate. **F)** Quantification of oocytes per ovary at E18.5 in DOX treated and untreated, +Tg and -Tg embryos. **G)** Quantification of oocyte per ovary at E15.5, E18.5 and P2 in *Trex1* mutant and control embryos.

2.3 An unbiased approach to identify mechanisms of FOA

To better understand how fetal oocytes are killed as a consequence of L1 reverse transcriptase activity, I took an unbiased RNA sequencing approach to compare gene expression profiles of the untreated wild-type and AZT-treated oocytes and ovaries at the onset (E15.5) and peak (E18.5) of oocyte death, when AZT has an effect. I also obtained transcriptional data for E13.5 ovaries and oocytes (data analysis not shown). At E15.5, untreated and AZT-treated samples contain similar oocyte populations because this stage is prior to any significant amount of FOA. At E18.5, untreated samples undergo FOA and contain only surviving oocytes while AZT-treated samples contain surviving oocytes and an additional population of rescued oocytes. I hypothesized that gene expression differences related to FOA mechanisms, cell autonomous or non-autonomous, may be revealed by such enrichment of rescued oocytes.

Initial analysis of biological replicates was performed by mapping to the mouse genome using Top Hat aligner that considers splice variants followed by Cuffdiff that performs pairwise differential gene expression analysis[122,123]. These results were then imported into R and analyzed using cummeRbund program[124]. By multidimensional scaling (MDS), samples were clustered based on developmental stage and tissue/cell of origin. Four main clusters formed that represented E15.5 and E18.5 ovaries and E15.5 and E18.5 oocytes (Figure 2-3A). Separation of untreated and AZT-treated samples in these respective clusters was less distinguished. Therefore, I conclude that AZT treatment does not have global effects on transcription, highlighting its specificity to reverse transcriptase.

Next, I viewed differential gene expression between pairs of untreated and AZT-treated samples by volcano plot. At E15.5, prior to FOA, the gene expression changes between untreated and AZT-treated ovary or oocyte samples were less significant than those observed between untreated and AZT-treated ovary and oocyte samples at E18.5 (Figure 2-3B). As these gene expression changes may give insight into FOA mechanisms, I performed gene ontology (GO) analysis to understand pathways that may be changing between untreated and AZT-treated ovary and oocyte samples at E18.5[125,126]. Interestingly, GO

terms associated with genes enriched in untreated samples that undergo FOA normally were related to immunity, apoptosis, stress response, and development (Figure 2-3C; Table 2-2 - 2-5). Such terms were exciting to us as they seem to be associated with cell death that were absent upon FOA evasion in AZT treated ovaries and oocytes. Last, I analyzed individual genes that were differentially expressed between untreated and AZT-treated samples related to these pathways described below including the antiviral immune response, complement system, and apoptosis (Figure 2-4D).

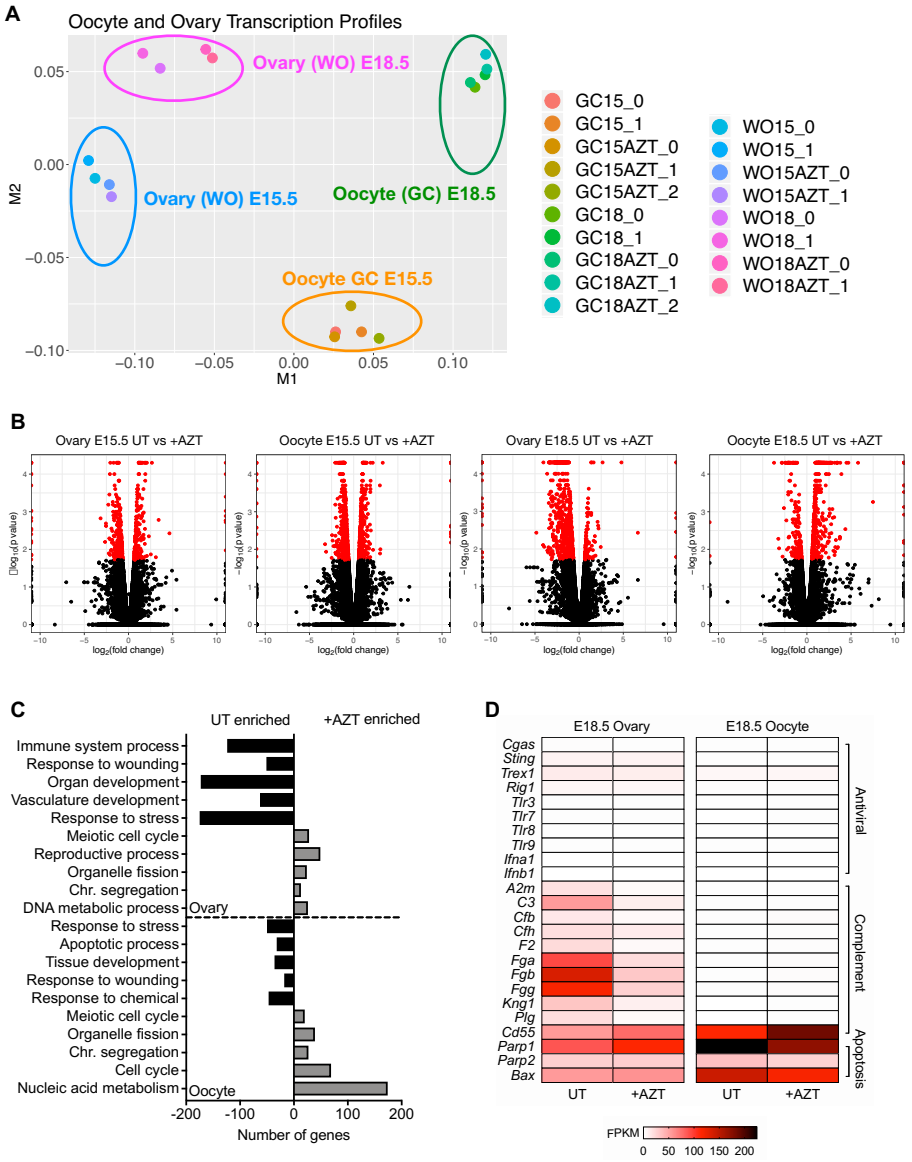


Figure 2-3: Transcriptional analysis of FOA. **A)** MDS plot of E15.5 and E18.5, ovary and oocyte, untreated and AZT-treated samples. **B)** Volcano plots displaying pairwise comparison of differential gene expression between E15.5 untreated and E15.5 +AZT ovaries and oocytes, E18.5 untreated and E18.5 +AZT ovaries and oocytes. Red dots indicate fold change difference is significant and black dots indicate not significant. **C)** Gene ontology enrichment with AZT in E18.5 ovaries and oocytes. **D)** Immunity gene expression in untreated (UT) and AZT-treated E18.5 ovaries and oocytes.

GO terms associated with E18.5 AZT-treated ovaries

Term	Count	P-Value	Fold Enrichment
meiotic cell cycle	28	7.40E-20	10.7
meiotic nuclear division	23	4.70E-17	11.6
meiotic cell cycle process	23	1.80E-16	10.9
meiosis I	16	7.50E-13	13.5
spermatogenesis	30	7.80E-12	4.7
male gamete generation	30	8.20E-12	4.7
sexual reproduction	37	1.40E-10	3.4
reproductive process	49	1.80E-10	2.7
reproduction	49	1.90E-10	2.7
single organism reproductive process	46	3.40E-10	2.8
male meiosis	10	1.00E-09	20.5
meiotic chromosome segregation	12	3.00E-09	12.4
gamete generation	31	9.90E-09	3.4
multi-organism reproductive process	37	1.10E-08	2.9
chromosome organization involved in meiotic cell cycle	10	5.30E-08	13.4
nuclear division	24	6.80E-08	3.8
germ cell development	18	1.20E-07	5
organelle fission	24	2.30E-07	3.6
multicellular organismal reproductive process	32	2.60E-07	2.8
multicellular organism reproduction	32	3.10E-07	2.8
cellular process involved in reproduction in multicellular organism	18	1.10E-06	4.3
nuclear chromosome segregation	13	1.00E-05	5.1
cell cycle	35	1.10E-05	2.2
DNA metabolic process	26	1.20E-05	2.7
ncRNA metabolic process	17	2.10E-05	3.6
chromosome segregation	14	2.30E-05	4.3
cell cycle process	29	2.30E-05	2.4
DNA recombination	12	3.60E-05	4.9
cellular nitrogen compound metabolic process	94	6.90E-05	1.4
chromosome organization	27	9.40E-05	2.3
nucleic acid metabolic process	77	9.80E-05	1.5

spermatid development	11	1.40E-04	4.6
gene silencing	12	3.00E-04	3.8
nitrogen compound metabolic process	95	3.50E-04	1.3
heterocycle metabolic process	83	3.50E-04	1.4
spermatid differentiation	11	3.60E-04	4.1
organic cyclic compound metabolic process	85	5.20E-04	1.4
cellular aromatic compound metabolic process	83	5.20E-04	1.4
developmental process involved in reproduction	21	6.60E-04	2.3
DNA repair	14	7.70E-04	3
nucleobase-containing compound metabolic process	80	8.70E-04	1.4
cellular metabolic process	128	1.20E-03	1.2
metabolic process	139	1.30E-03	1.2
cellular macromolecule metabolic process	109	1.30E-03	1.3
macromolecule metabolic process	118	1.40E-03	1.2
multi-organism process	42	1.60E-03	1.6
primary metabolic process	127	2.50E-03	1.2
organic substance metabolic process	133	2.70E-03	1.2
macromolecule methylation	10	2.80E-03	3.4
organelle organization	56	3.60E-03	1.4
RNA processing	18	3.90E-03	2.2
methylation	11	4.20E-03	3
gene expression	72	1.30E-02	1.3
cellular response to DNA damage stimulus	15	1.50E-02	2.1
RNA metabolic process	60	2.50E-02	1.3
translation	14	4.60E-02	1.8
cellular amide metabolic process	18	5.50E-02	1.6
peptide biosynthetic process	14	5.50E-02	1.8

Table 2-2: GO enrichment with genes increased in E18.5, AZT-treated wild-type ovaries. Top 60 terms shown with number of genes associated (Count), P-Value, and Fold enrichment. Determined using DAVID bioinformatics resources 6.8. Bold letters indicate terms used in Fig. 2-3C.

GO terms associated with E18.5 untreated ovaries

Term	Count	P-Value	Fold Enrichment
single-multicellular organism process	261	1.70E-19	1.6
system development	211	2.50E-18	1.7
response to stress	175	4.30E-18	1.9
vasculature development	63	1.40E-17	3.5
cardiovascular system development	81	1.90E-17	2.8
circulatory system development	81	1.90E-17	2.8
animal organ development	173	2.60E-17	1.8
multicellular organism development	226	2.90E-17	1.6
blood vessel development	59	3.90E-16	3.4
anatomical structure development	241	5.20E-16	1.5
regulation of multicellular organismal process	144	2.30E-15	1.9
response to wounding	51	4.10E-15	3.7
negative regulation of cellular process	196	5.60E-15	1.7
developmental process	243	6.90E-15	1.5
response to inorganic substance	52	1.80E-14	3.5
blood vessel morphogenesis	51	1.90E-14	3.5
immune system process	124	5.70E-14	2
single-organism developmental process	237	5.80E-14	1.5
wound healing	43	6.50E-14	3.9
negative regulation of biological process	203	7.80E-14	1.6
protein complex subunit organization	87	4.20E-13	2.3
macromolecular complex subunit organization	117	7.10E-13	2
negative regulation of multicellular organismal process	76	7.90E-13	2.4
response to oxygen-containing compound	95	2.50E-12	2.1
response to metal ion	36	5.00E-12	4
response to external stimulus	117	5.80E-12	1.9
regulation of response to external stimulus	57	1.00E-11	2.8
negative regulation of hydrolase activity	37	1.40E-11	3.8
tissue development	101	1.60E-11	2
response to organic substance	138	2.60E-11	1.7
negative regulation of peptidase activity	28	3.00E-11	4.8

regulation of cell proliferation	91	3.50E-11	2.1
regulation of multicellular organismal development	100	3.60E-11	2
nucleosome assembly	21	3.90E-11	6.7
regulation of biological quality	151	5.30E-11	1.7
negative regulation of proteolysis	32	5.50E-11	4.1
chromatin silencing	23	6.10E-11	5.8
defense response	84	7.90E-11	2.1
angiogenesis	40	8.20E-11	3.4
protein activation cascade	18	8.70E-11	7.9
protein complex assembly	76	9.00E-11	2.2
protein complex biogenesis	76	9.40E-11	2.2
positive regulation of biological process	209	1.00E-10	1.5
extracellular matrix organization	28	1.10E-10	4.6
negative regulation of hemostasis	14	1.10E-10	11.6
negative regulation of blood coagulation	14	1.10E-10	11.6
extracellular structure organization	28	1.20E-10	4.5
negative regulation of gene expression, epigenetic	23	1.40E-10	5.6
regulation of hydrolase activity	60	1.60E-10	2.5
regulation of response to stress	72	1.60E-10	2.2
negative regulation of coagulation	14	1.90E-10	11.1
nucleosome organization	22	2.40E-10	5.7
myeloid cell differentiation	36	3.40E-10	3.5
regulation of peptidase activity	34	5.40E-10	3.6
regulation of developmental process	111	5.80E-10	1.8
homeostatic process	90	6.00E-10	2
macromolecular complex assembly	84	6.10E-10	2
cell proliferation	96	6.10E-10	1.9

Table 2-3: GO enrichment with genes increased in E18.5, untreated wild-type ovaries. Top 60 terms shown with number of genes associated (Count), P-Value, and Fold enrichment. Determined using DAVID bioinformatics resources 6.8. Bold letters indicate terms used in Fig. 2-3C.

GO terms associated with E18.5 AZT-treated oocytes

Term	Count	P-Value	Fold Enrichment
nucleic acid metabolic process	174	7.50E-17	1.7
cellular nitrogen compound metabolic process	206	2.10E-16	1.6
cellular macromolecule metabolic process	245	2.50E-16	1.5
heterocycle metabolic process	189	4.70E-16	1.7
nitrogen compound metabolic process	212	5.90E-16	1.6
nucleobase-containing compound metabolic process	186	6.00E-16	1.7
cellular aromatic compound metabolic process	189	1.50E-15	1.6
organic cyclic compound metabolic process	192	3.30E-15	1.6
primary metabolic process	275	3.50E-14	1.3
cellular metabolic process	273	7.30E-14	1.3
organic substance biosynthetic process	190	1.00E-13	1.6
macromolecule metabolic process	254	2.10E-13	1.4
macromolecule biosynthetic process	168	2.30E-13	1.6
cellular macromolecule biosynthetic process	165	2.80E-13	1.7
cellular biosynthetic process	186	4.10E-13	1.6
gene expression	174	5.20E-13	1.6
biosynthetic process	190	5.40E-13	1.5
organic substance metabolic process	282	1.10E-12	1.3
regulation of gene expression	148	6.20E-12	1.7
regulation of nitrogen compound metabolic process	145	2.50E-11	1.7
metabolic process	285	5.90E-11	1.3
RNA metabolic process	146	8.60E-11	1.6
cell cycle	69	9.00E-11	2.3
regulation of macromolecule biosynthetic process	137	9.40E-11	1.7
regulation of cellular macromolecule biosynthetic process	134	1.60E-10	1.7
regulation of nucleobase-containing compound metabolic process	135	1.90E-10	1.7
nuclear division	39	3.00E-10	3.3
cellular nitrogen compound biosynthetic process	153	4.30E-10	1.6
regulation of biosynthetic process	140	6.20E-10	1.6
chromosome segregation	27	7.50E-10	4.4
nuclear chromosome segregation	24	8.10E-10	4.9

regulation of cellular biosynthetic process	138	9.60E-10	1.6
regulation of macromolecule metabolic process	175	1.20E-09	1.5
regulation of primary metabolic process	172	1.60E-09	1.5
heterocycle biosynthetic process	134	1.70E-09	1.6
organelle fission	39	2.00E-09	3.1
regulation of RNA metabolic process	123	2.20E-09	1.7
aromatic compound biosynthetic process	134	2.20E-09	1.6
nucleobase-containing compound biosynthetic process	132	2.60E-09	1.6
regulation of cellular metabolic process	170	1.10E-08	1.4
organic cyclic compound biosynthetic process	134	1.20E-08	1.6
regulation of metabolic process	179	1.50E-08	1.4
nucleic acid-templated transcription	119	2.10E-08	1.6
chromosome organization	52	2.50E-08	2.3
cell cycle process	53	2.60E-08	2.3
RNA biosynthetic process	119	2.80E-08	1.6
organelle organization	120	3.50E-08	1.6
sister chromatid segregation	18	5.00E-08	5.3
regulation of transcription, DNA-templated	115	6.00E-08	1.6
regulation of nucleic acid-templated transcription	115	6.30E-08	1.6
regulation of RNA biosynthetic process	115	6.90E-08	1.6
cell division	32	5.60E-07	2.8
meiotic cell cycle process	18	6.60E-07	4.5
mitotic nuclear division	27	6.70E-07	3.1
meiotic cell cycle	20	7.40E-07	4
sister chromatid cohesion	10	8.90E-07	9.5
DNA conformation change	18	1.90E-06	4.1
mRNA transport	13	2.20E-06	5.9

Table 2-4: GO enrichment with genes increased in E18.5, AZT-treated wild-type oocytes. Top 60 terms shown with number of genes associated (Count), P-Value, and Fold enrichment. Determined using DAVID bioinformatics resources 6.8. Bold letters indicate terms used in Fig. 2-3C.

GO terms associated with E18.5 untreated oocytes

Term	Count	P-Value	Fold Enrichment
cellular response to chemical stimulus	47	5.40E-08	2.3
response to wounding	18	7.60E-07	4.4
response to organic substance	48	9.90E-07	2
response to nitrogen compound	25	1.10E-06	3.1
response to chemical	59	1.20E-06	1.8
tissue development	36	1.20E-06	2.4
cellular response to organic substance	38	1.90E-06	2.3
response to external stimulus	40	2.30E-06	2.2
response to organonitrogen compound	22	5.60E-06	3.1
response to endogenous stimulus	31	7.20E-06	2.4
negative regulation of multicellular organismal process	25	1.30E-05	2.7
response to oxygen-containing compound	31	1.50E-05	2.3
negative regulation of biological process	62	1.60E-05	1.6
positive regulation of macromolecule metabolic process	43	1.70E-05	1.9
anatomical structure development	71	1.90E-05	1.5
apoptotic process	32	1.90E-05	2.3
positive regulation of metabolic process	45	1.90E-05	1.9
cellular amide metabolic process	23	2.10E-05	2.8
response to stress	50	2.10E-05	1.8
response to cytokine	18	2.40E-05	3.4
developmental process	72	2.90E-05	1.5
cell death	34	3.10E-05	2.1
cellular response to oxygen-containing compound	23	3.10E-05	2.7
positive regulation of cellular metabolic process	42	3.30E-05	1.9
response to growth factor	17	3.40E-05	3.4
single-multicellular organism process	73	4.20E-05	1.5
animal organ development	49	4.60E-05	1.7
programmed cell death	32	5.00E-05	2.1
cellular response to endogenous stimulus	24	5.80E-05	2.5
peptide metabolic process	20	5.80E-05	2.9
single-organism developmental process	70	6.30E-05	1.5

negative regulation of response to external stimulus	11	6.80E-05	5
gene expression	64	7.70E-05	1.5
cellular response to growth factor stimulus	16	8.20E-05	3.4
cellular response to nitrogen compound	16	9.40E-05	3.3
response to mechanical stimulus	10	9.70E-05	5.4
regulation of multicellular organismal process	41	9.80E-05	1.8
response to oxidative stress	13	1.20E-04	3.9
response to organic cyclic compound	21	1.30E-04	2.6
response to inorganic substance	15	1.40E-04	3.4
negative regulation of protein metabolic process	21	1.40E-04	2.6
anatomical structure formation involved in morphogenesis	24	1.40E-04	2.4
cellular component biogenesis	39	1.40E-04	1.8
blood vessel development	16	1.70E-04	3.1
nitrogen compound metabolic process	74	1.80E-04	1.4
negative regulation of cellular protein metabolic process	20	1.90E-04	2.6
cellular nitrogen compound metabolic process	71	2.00E-04	1.4
regulation of macromolecule metabolic process	67	2.10E-04	1.5
positive regulation of nitrogen compound metabolic process	29	2.10E-04	2.1
cellular component assembly	36	2.30E-04	1.9
multicellular organism development	62	2.50E-04	1.5
positive regulation of macromolecule biosynthetic process	27	2.50E-04	2.1
cellular response to organonitrogen compound	14	2.70E-04	3.4
response to hormone	18	2.70E-04	2.7
anatomical structure morphogenesis	39	2.90E-04	1.8
cellular response to stress	28	3.10E-04	2.1
vasculature development	16	3.10E-04	3
regulation of cellular amide metabolic process	12	3.20E-04	3.8

Table 2-5: GO enrichment with genes increased in E18.5, untreated wild-type oocytes. Top 60 terms shown with number of genes associated (Count), P-Value, and Fold enrichment. Determined using DAVID bioinformatics resources 6.8. Bold letters indicate terms used in Fig. 2-3C.

Complement system and immune cell recruitment

Differential gene expression and gene ontology enrichment analyses in untreated compared to AZT-treated samples at E18.5 revealed a number of candidate cell death mechanisms overrepresented in the untreated samples, and thus related to oocytes undergoing FOA (Figure 2-3D; Table 2-6). These were genes related to immunity and apoptosis. Retrotransposon activity has been associated with activation of the innate immune system as its intermediates resemble those generated by retrovirus activity. Endogenous retroviruses have even shaped the transcriptional network underlying proinflammatory interferon response[127]. One complication is that antiviral immune responses occur in the cytoplasm, whereas L1 retrotransposition occurs at the nuclear genome. L1 overexpression in brain and aged tissues has been linked to antiviral innate immune signatures such as the cGAS-STING pathway that result in expression of inflammatory genes. However, how these intermediates are transported to the cytoplasm, if at all, remains unknown[113,115]. For these reasons, I focused on the immunity signatures identified by RNA-seq that correlated with FOA.

I first analyzed genes reported in studies of immune response to L1 ssDNA. Interestingly, expression of these mRNAs corresponding to canonical antiviral innate immunity genes including the cGAS/STING pathway and type-I interferon response were negligible in fetal oocytes and whole ovaries (Figure 2-3D). This suggests that oocytes do not permit L1 retrotransposition intermediates in the cytoplasm to carry out these types of responses, especially since fetal oocytes are in meiosis, a non-dividing state. However, from less biased analysis of the transcriptomic data, I found gene expression signatures related to an alternative form of immunity that correlated with FOA, being upregulated in untreated compared to AZT-treated ovaries. These were genes encoding components of the complement system, an essential host defense that involves phagocytic clearance of infected cells by immune cells such as macrophages (Figure 2-3D). The complement system is activated in response to infection by molecules such as C3 convertase, which were significantly upregulated in E18.5 untreated ovaries. In contrast, E18.5 AZT-treated oocytes that evade FOA showed increased expression of the *Cd55* or *Daf*

(Decay-accelerating factor) gene compared to untreated oocytes that protects cells from autologous complement attack [128]. The correlation of complement system activation and FOA at early stages was validated by reduced occurrence of macrophages in ovaries that evaded FOA at E18.5 compared to ovaries that undergo FOA normally (Figure 2-4A, B). Roles for the complement system in follicle development and fertility in males and females have been reported, where mutations leading to increased C3 show reduced breeding success and reduction in ovarian quality[129,130].

It is unclear whether the complement system directly detects oocytes with excess L1 ssDNA or if it indirectly detects dead oocytes that were killed by way of another mechanism in response to L1 activity. Other genes that were differentially expressed between E18.5 untreated and AZT-treated oocytes were those involved in apoptosis. These genes included PARP1 and BAX, both of which show increased expression in untreated compared to AZT-treated oocytes. PARP1 is normally involved in DNA repair during MPI and identified in germ cell cysts. When cleaved, cPARP1 becomes a pro-apoptotic protein and has been implicated in oocyte death[131]. Interestingly, cPARP1 activity is more prone to ssDNA breaks, a feature that may allow this pathway to be active during meiotic recombination without eliminating cells with programmed DSBs. It is a possibility that cPARP1-mediated FOA kills oocytes with accumulated L1 reverse transcription intermediates, and these oocytes are subsequently removed by the complement system in an indirect manner.

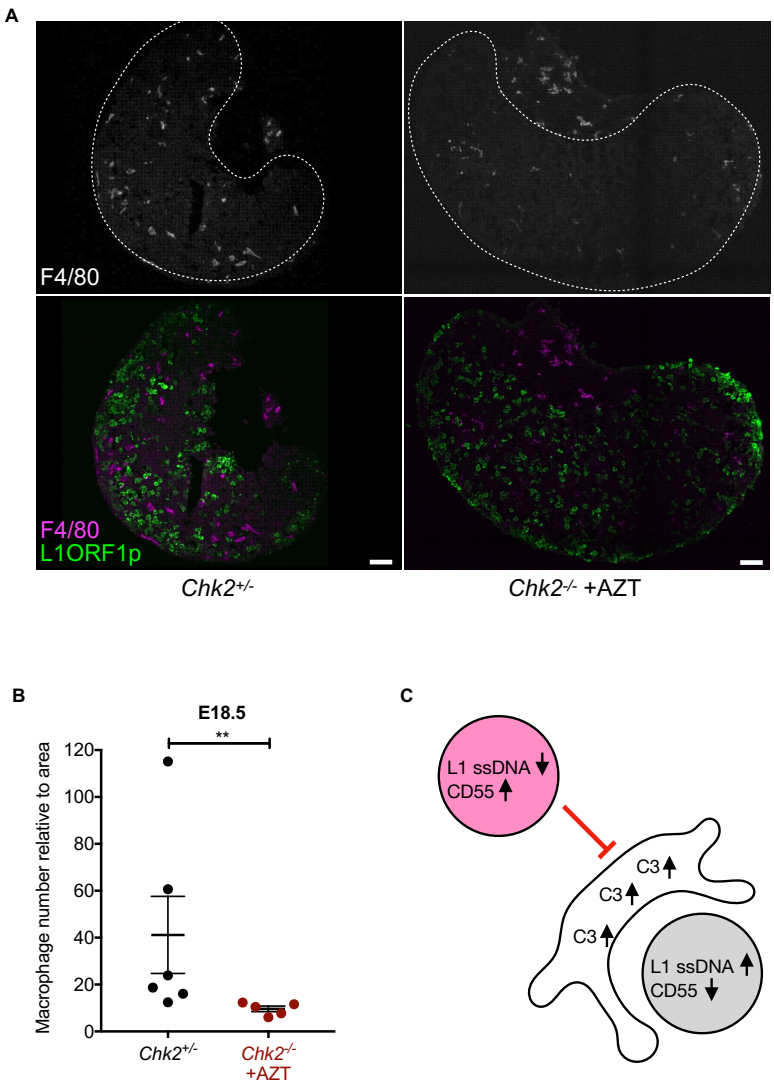


Figure 2-4: Immune cell recruitment with FOA. **A)** Immunofluorescence labeling with macrophage marker F4/80 in E18.5 ovaries of untreated (*Chk2*$^{+/-}$) and AZT-treated (*Chk2*$^{-/-}$ +AZT) conditions. Ovary separated from soma with a dotted white line border determined by L1 ORF1p specifically marking oocytes. Scalebars:50µm. **B)** Quantification of macrophage number per ovary section area in untreated (*Chk2*$^{+/-}$) (n=6) and AZT-treated (*Chk2*$^{-/-}$ +AZT) (n=5) conditions. Dots indicate independent ovary sections from 3 ovary samples per condition; data are mean ± s.e.m. Stats by Mann-Whitney test, **p<0.01. **C)** Models for direct or indirect targeting of oocytes for FOA by the complement system.

Differential gene expression between E18.5 untreated and AZT-treated ovaries (WO) and oocytes (GC) related to innate immunity								
Gene ID	**WO UT FPKM**	**WO UT SD**	**WO +AZT FPKM**	**WO +AZT SD**	**GC UT FPKM**	**GC UT SD**	**GC +AZT FPKM**	**GC +AZT SD**
A2m	16.43	3.12725	2.63	0.63582	0.04	0.0376221	0.11	0.0806205
C3	52.60	10.6102	9.55	1.977145	0.10	0.0649075	0.86	0.333063
Cfb	12.47	2.58755	3.49	0.94033	1.09	0.383805	1.48	0.47659
Cfh	17.37	3.28355	10.67	2.0786	0.67	0.234973	0.51	0.193949
F2	19.67	4.0612	3.47	1.006075	0.69	0.3103075	1.03	0.44478
Fga	91.01	18.41365	18.67	4.1168	0.01	0.02539769	1.37	0.648415
Fgb	133.05	26.9265	29.43	6.50225	0.05	0.06369655	1.50	0.747715
Fgg	115.31	22.822	25.97	5.8551	0.17	0.138305	1.40	0.730395
Kng1	31.26	6.585	8.57	2.14355	0.00	0	0.74	0.383123
Plg	18.40	3.7837	2.55	0.75252	0.00	0	0.50	0.2955345
Cd55	53.73	10.27045	75.63	14.2048	110.56	21.5785	182.67	37.856
Rig1	5.19496	1.095925	4.27313	0.924515	1.00612	0.30282	1.08949	0.32031
Cgas	0.990735	0.3153725	0.539244	0.205667	0.394996	0.166022	0.369805	0.161995
Sting	5.86239	1.38373	6.17467	1.459995	0.710732	0.317509	0.432963	0.2366685
Trex1	10.8487	2.68575	9.69629	2.492505	4.25683	1.37535	4.66047	1.465855
Ifna1	0	0	0	0	0	0	0	0
Ifna2	0	0	0	0	0	0	0	0
Ifna4	0	0	0	0	0	0	0	0
Ifna5	0	0	0	0	0	0	0	0
Ifna6	0	0	0	0	0	0	0	0
Ifna7	0	0	0	0	0	0	0	0
Ifna9	0	0	0	0	0	0	0	0
Ifna11	0	0	0	0	0	0	0	0
Ifna12	0	0	0	0	0	0	0	0
Ifna13	0	0	0.194454	0.1857495	0.406853	0.3236085	0.541643	0.4133385
Ifna14	0	0	0	0	0	0	0	0
Ifna15	0	0	0	0	0	0	0	0
Ifna16	0	0	0	0	0	0	0	0
Ifnab	0	0	0	0	0	0	0.101765	0.144745
Ifnb1	0	0	0	0	0	0	0	0
Ifne	0	0	0	0	0.226699	0.2195755	0	0
Ifng	0	0	0.11892	0.120405	0.187143	0.168207	0.0476681	0.06644895
Ifnk	0.442195	0.6562625	0.32979	0.512295	0	0	0.120507	0.187584
Ifnl2	0	0	0	0	0	0	0	0
Ifnl3	0	0	0	0	0	0	0	0
Ifnz	0	0	0	0	0	0	0	0
Tlr1	0.587301	0.2445245	0.321135	0.184907	0.0988623	0.08137685	0.0284408	0.0412921
Tlr2	1.45452	0.462175	1.04174	0.365745	0.320391	0.1727635	0.0889026	0.0734442
Tlr3	0.787298	0.276791	0.518213	0.197626	0.0504267	0.05127765	0.0780164	0.0624228
Tlr4	1.65274	0.475705	0.899736	0.301022	0.0512179	0.04363105	0.0927232	0.0735979
Tlr5	3.52569	0.87283	3.22705	0.81112	0.110977	0.0870935	0.0993224	0.0781308
Tlr6	0.82516	0.319635	0.421512	0.2037005	0	0	0.045466	0.0555985
Tlr7	0.211276	0.122234	0.150545	0.101795	0	0	0.0172116	0.0379109
Tlr8	0.116264	0.09045	0.118574	0.083358	0.0632321	0.05380745	0.34794	0.1684435
Tlr9	0.0223925	0.02981535	0.0418509	0.04722405	0.0445282	0.0493404	0.0577135	0.05102025
Tlr11	0.00812128	0.01617601	0	0	0	0	0	0
Tlr12	0.156021	0.110487	0.0860098	0.0681956	0.0279781	0.03384715	0.0638025	0.05612925
Tlr13	0.527435	0.209942	0.43631	0.1784085	0.0218295	0.02870515	0.116882	0.0746045
Nod1	5.58116	1.19171	6.20503	1.321175	0.39751	0.167425	0.464113	0.189863
Nod2	0.329846	0.14394	0.321246	0.139259	1.09407	0.3271	1.20718	0.36079

Table 2-6: List of mean FPKM and standard deviation (SD) for innate immunity pathway genes in untreated and AZT-treated E18.5 wild-type ovaries and oocytes.

2.4 The DNA damage checkpoint as a mechanism of FOA

To test our hypothesis that FOA is a result of genotoxic stress from both L1 reverse transcriptase activity and L1 endonuclease-instigated activation of the DNA damage checkpoint, I examined the role of CHK2, a main player in the DNA damage checkpoint. We first compared the oocyte number between *Chk2*$^{-/-}$ and control *Chk2*$^{+/-}$ ovaries. *Chk2*$^{+/-}$ ovaries were used as the control after observing differences in oocyte number due to mixed genetic background when comparing *Chk2*$^{-/-}$ to WT C57Bl/6J (B6) mice (Figure 6-1A-C).

Given that CHK2 has known roles in other cell-cycle related processes, I first analyzed the role of CHK2 in MPI progression by quantifying the percentage of oocytes in preleptotene, leptotene, zygotene, pachytene and diplotene stages based on morphology of the synaptonemal complex as described in the introduction. I quantified percent oocytes in MPI stages at E15.5, prior to FOA, and at E18.5 during FOA from *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ mice (Figure 2-5A, B). We found no significant difference between *Chk2*$^{-/-}$ and *Chk2*$^{+/-}$ ovaries at E15.5 that contain predominantly pre-leptotene, leptotene and zygotene stages (Figure 2-5A). However, at E18.5 when the majority of oocytes have reached the pachytene and diplotene stages, an over-representation of diplotene oocytes were observed in *Chk2*$^{-/-}$ mice (Figure 2-5B; Table 2-7). I conclude that CHK2 does not impact meiotic progression in oocytes, but the enrichment for diplotene oocytes at E18.5 reflects the presence of oocytes spared in the absence of DNA damage checkpoint activation that are normally killed.

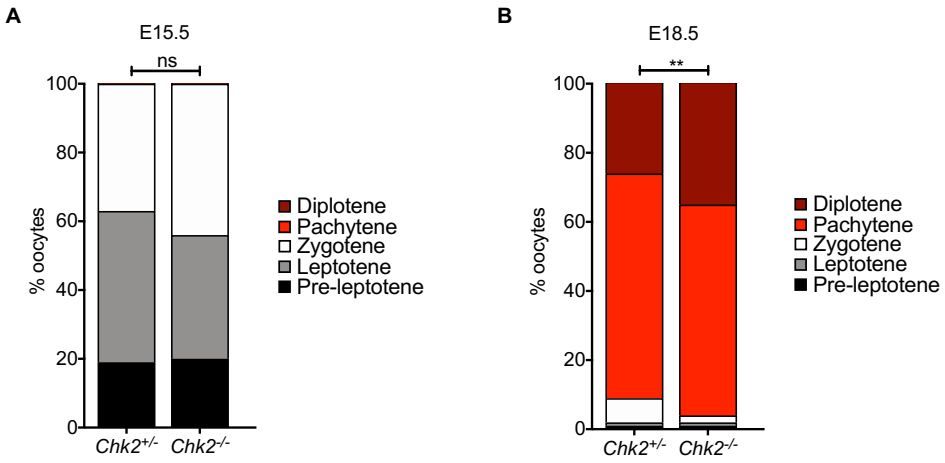


Figure 2-5: Meiotic prophase I progression in absence of CHK2. **A-B)** Comparison of meiotic progression between *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ ovaries at E15.5 (**A)** and E18.5 (**B)** by quantifying percent of oocyte in pre-leptotene, leptotene, zygotene, pachytene and diplotene stages based on synaptonemal complex morphology and the presence of meiotic double-stranded breaks using γH2AX staining. Stats by Chi-square test, ns $p>0.05$, $^{**}p<0.01$.

MPI stage distribution of Chk2+/- and Chk2-/- oocytes

Samples	Pre-leptotene %	Leptotene %	Zygotene %	Pachytene %	Diplotene %	n oocytes	n litters
E15.5 *Chk2+/-*	18.7	43.8	36.7	0.8	0	128	1
E15.5 *Chk2-/-*	19.7	36.3	43.5	0.5	0	2030	3
E18.5 *Chk2+/-*	0	0.9	6.8	65.4	26.8	221	2
E18.5 *Chk2-/-*	0	0.8	1.6	61.1	36.5	303	3

Statistical analysis of MPI stage distribution of Chk2+/- and Chk2-/- oocytes

Sample	Chi-square	P[b]	Significance
E15.5 *Chk2+/- vs* **E1**	3.1	0.5413	ns
E18.5 *Chk2+/- vs* **E**	15.01	0.0047	**

(ns) - P>0.05; (**) - P<0.01

[b]analyzed by Chi-square test

Table 2-7: Quantification and statistical analysis of meiotic progression in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ ovaries at E15.5 and E18.5. Stages quantified by morphology of synaptonemal complex and quantified as percent of given stage out of total oocytes calculated. Statistical significance determined by Chi-square test.

To determine whether CHK2 is involved in FOA, I quantified oocyte number at three different timepoints. First, I analyzed oocyte number at E15.5, when the maximum number of oocytes are present in C57Bl/6J (B6) genetic background and when FOA begins. Until this point, germ cells are proliferating and asynchronously entering meiotic prophase where proliferation stops. To rule out a role for CHK2 in proliferation of germ cells, I quantified oocytes at E15.5 and observed no significant difference in oocyte number between *Chk2*$^{-/-}$ and *Chk2*$^{+/-}$ ovaries, suggesting that CHK2 does not impact germ cell proliferation (Figure 2-6A; Table 2-8, 2-9). Next, I compared oocyte number per ovary in *Chk2*$^{-/-}$ and *Chk2*$^{+/-}$ at E18.5, when approximately half of the oocytes are eliminated in wild-type mice due to FOA, and again at postnatal day 2 (P2), the endpoint of FOA after an additional smaller population of oocytes are eliminated and after DNA damage checkpoint activation [69]. At E18.5, I observed that the oocyte number is comparable between *Chk2*$^{-/-}$ and *Chk2*$^{+/-}$ ovaries, both losing a major portion of the maximum oocyte supply between E15.5 and E18.5 during early MP I stages (Figure 2-6B). However, by P2, *Chk2*$^{-/-}$ ovaries contained significantly more oocytes than P2 *Chk2*$^{+/-}$ ovaries, comparable to the number in *Chk2*$^{-/-}$ ovaries at E18.5 (Fig. 2-6C). These results suggest a role for CHK2 in FOA beginning at E18.5, which coincides with the onset of DNA damage checkpoint activation. This transition to DNA damage checkpoint-dependent FOA is supported by a comparable number of oocytes between *Chk2*$^{-/-}$ and *Chk2*$^{+/-}$ at E17.5, prior to checkpoint activation, and upregulation of the CHK2 target TAp63 to trigger apoptosis at E18.5, but not at E15.5 where the dominant negative isoform is expressed (Fig. 2-6D, E; Fig. 6-2A, B) [132]. Importantly, CHK2-dependent FOA was distinct from AZT-sensitive FOA occurring in early MPI stages.

Since the onset of CHK2 function in FOA at E18.5 coincided with the endpoint of the protective effect of AZT [69], I wanted to test if AZT treatment of *Chk2* mutant mice prevented FOA completely. I first validated that daily administration of AZT to pregnant females starting at E13.5 until P2 prevents FOA between E15.5 and E18.5, but fails to maintain this effect until P2 in control *Chk2*$^{+/-}$ mice (Figure 2-6A-C) [69]. In contrast, AZT treatment of *Chk2*$^{-/-}$ animals preserved more oocytes by P2 than either

condition alone (Figure 2-6C). I included additional biological replicates for all P2 stages using an alternative ovary section-based quantitation method to emphasize reproducibility (Fig. 2-6F). In some cases, all fetal oocytes initially generated at E15.5 persisted to P2 in *Chk2*$^{-/-}$ +AZT ovaries, resulting in a maximized oocyte supply (Fig. A-C, F, G). Notably, ending AZT treatment of *Chk2*$^{-/-}$ mice at E18.5 (rather than at P2) did not increase the oocyte supply at P2 (Figure 2-6C, H, I; Table 2-8, 2-9). These observations identify two distinct mechanisms of FOA. First, a mechanism driven by L1 reverse transcriptase activity that is sensitive to AZT is active throughout MPI, but naturally predominant in early fetal stages. In later fetal stages, a second mechanism dependent on CHK2 in the mid-pachytene stage of MPI is driven by DNA damage and meiotic defects associated with L1 endonuclease activity. Importantly, by understanding these two mechanisms, we can prevent FOA beyond birth and, for the first time, have a system that presents the opportunity to study fetal oocytes that are otherwise eliminated.

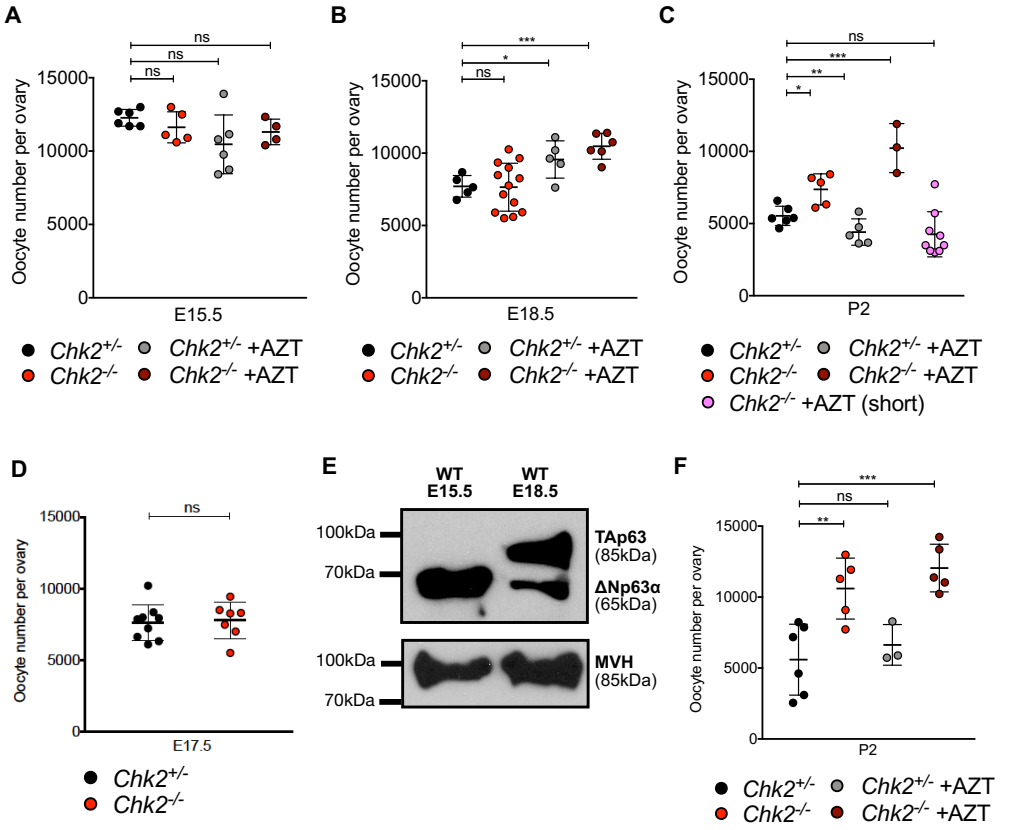

A
B
C
Oocyte number per ovary
15000
10000
5000
0
ns
*
***
**
E15.5
E18.5
P2
Chk2+/-
Chk2+/- +AZT
Chk2-/-
Chk2-/- +AZT
Chk2-/- +AZT (short)
D
E17.5
E
WT
E15.5
WT
E18.5
100kDa
70kDa
TAp63
(85kDa)
ΔNp63α
(65kDa)
MVH
(85kDa)
F
G
n oocytes = 5,200
n oocytes = 8,159
n oocytes = 4,178
n oocytes = 11,900
P2
100µm
Chk2+/-
Chk2-/-
Chk2+/- +AZT
Chk2-/- +AZT
H
Continuous AZT treatment
Short AZT treatment
E13.5 E14.5 E15.5 E16.5 E17.5 E18.5 P0 P1 P2
I
% oocyte population
100
50
0
E15.5
E18.5
P2

Figure 2-6: AZT-treated *Chk2*$^{-/-}$ ovaries evade FOA**. A-C)** Oocyte number per ovary at E15.5 **(A)**, E18.5 **(B)** and P2 **(C)** in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ untreated and AZT-treated ovaries. Dots indicate independent ovary samples; data are mean ± SD; n≥3 ovaries per sample. Stats by two-tailed unpaired Student's t-test, ns p>0.05; *p<0.05; **p<0.01; ***p<0.001. **D)** Oocyte number per ovary at E17.5 in *Chk2*$^{-/-}$ vs *Chk2*$^{+/-}$ embryos. **E)** Western blot detection of p63 in E15.5 and E18.5 whole ovaries. Mouse vasa homolog (MVH) used as loading control. A shift from the dominant negative isoform of p63 present at E15.5 to the active Tap63 isoform at E18.5 is observed in correlation with activation of the DNA damage checkpoint and rescue of oocytes in CHK2 mutants beginning at E18.5. **F)** Oocyte number per ovary at P2 in untreated and AZT-treated *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ conditions estimated by counting oocytes in every 5th section throughout an entire ovary. The trend observed in FOA dynamics reproduces that from whole-mount oocyte quantification method in Fig. 2-6C. Raw oocyte numbers slightly differ due to the estimation method in sections vs. counting each oocyte in the whole-mount method. Stats by two-tailed unpaired Student's t-test, ns p>0.05;**p<0.01; ***p<0.001. Dots indicate independent ovary samples; data are mean ± SD; n≥3 ovaries per sample from at least 3 different embryos of a single litter. **G)** TRA98 labeling of P2 *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ untreated and AZT-treated ovaries. Scalebars:100μm. **H)** Timeline of short and continuous AZT treatments. **I)** Summary of FOA dynamics in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ untreated and AZT-treated ovaries, including short and continuous AZT treatments at P2.

Oocyte number per ovary in Chk2+/- and Chk2-/- untreated and AZT-treated mice

Sample	Mean oocyte number	Std Dev	% of E15.5 oocytes (max)	n ovaries	n embryos	n litters
E15.5 *Chk2*+/- (N1)	12268.2	569.1	100	6	3	2
E15.5 *Chk2*+/- +AZT	10461.5	2002.3	100	6	3	2
E15.5 *Chk2*-/- (N0)	11620	1061.6	100	5	3	2
E15.5 *Chk2*-/- +AZT	11306.8	871.2	100	4	3	1
E18.5 *Chk2*+/-	7707.2	745.4	62.8	5	3	2
E18.5 *Chk2*+/- +AZT	9560.6	1288.9	91.4	5	3	2
E18.5 *Chk2*-/-	7645.9	1656.6	65.8	13	7	3
E18.5 *Chk2*-/- +AZT	10474.17	895.5	92.6	6	3	2
P2 *Chk2*+/-	5536.3	667.6	45.1	6	3	2
P2 *Chk2*+/- +AZT	4411.2	911	42.2	5	3	2
P2 *Chk2*-/-	7369.4	1075.1	63.4	5	3	2
P2 *Chk2*-/- +AZT	10225.7	1700.6	90.4	3	3	2
P2 *Chk2*-/- +AZT (to E18.5)	4258	1563	37.7	9	5	3
P2 *Chk2*+/- (sections)	5589	2495	36.4	6	3	1
P2 *Chk2*+/- +AZT (sections)	6630	1431	NA	3	3	1
P2 *Chk2*-/- (sections)	10598	2147	NA	5	3	1
P2 *Chk2*-/- +AZT (sections)	12045	1680	69.07	5	3	1
Wild-type B6	4708	1163	100	12	6	3
***Chk2*$^{+/+}$ (N1F1)**	11335	1567	100	8	5	4
***Chk2*$^{+/-}$ (N1F1)**	11182	895.4	100	8	5	4
***Chk2*$^{-/-}$ (N1F1)**	10152	1663	100	9	6	5
***Chk2*$^{+/-}$ (N5F1)**	5593	1098	100	5	3	2

Table 2-8: Oocyte number per ovary in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ untreated and AZT treated conditions. Ovaries analyzed at E15.5, E18.5, and P2 stages as well as quantified by whole-mount and sectioning methods. If not specified, whole-mount method was used for quantification.

Statistical analysis of oocyte number per ovary in Chk2+/- and Chk2-/- untreated and AZT treated ovaries

Sample	t-test	p[a]	Significance
E15.5 *Chk2*+/- untreated *vs* E15.5 *Chk2*+/- +AZT	2.126	0.0594	ns
E15.5 *Chk2*+/- untreated *vs* E15.5 *Chk2*-/- untreated	1.297	0.2268	ns
E15.5 *Chk2*+/- untreated *vs* E15.5 *Chk2*-/- +AZT	2.134	0.0654	ns
E18.5 *Chk2*+/- untreated *vs* E18.5 *Chk2*+/- +AZT	2.784	0.0238	*
E18.5 *Chk2*+/- untreated *vs* E18.5 *Chk2*-/- untreated	0.07856	0.9384	ns
E18.5 *Chk2*+/- untreated *vs* E18.5 *Chk2*-/- +AZT	5.491	0.0004	***
P2 *Chk2*+/- untreated *vs* P2 *Chk2*+/- +AZT	2.367	0.0421	*
P2 *Chk2*+/- untreated *vs* P2 *Chk2*-/- untreated	3.469	0.0071	**
P2 *Chk2*+/- untreated *vs* P2 *Chk2*-/- +AZT	6.199	0.0004	***
P2 *Chk2*+/- untreated *vs* P2 *Chk2*-/- +AZT (to E18.5)	1.875	0.0835	ns
P2 *Chk2*+/- +AZT *vs* P2 *Chk2*-/- +AZT (to E18.5)	0.1993	0.8454	ns
P2 *Chk2*-/- untreated *vs* P2 *Chk2*-/- +AZT (to E18.5)	3.931	0.002	**
P2 *Chk2*-/- +AZT *vs* P2 *Chk2*-/- +AZT (to E18.5)	5.626	0.0002	***
P2 *Chk2*+/- untreated *vs* P2 *Chk2*+/- +AZT (sections)	0.6562	0.5326	ns
P2 *Chk2*+/- untreated *vs* P2 *Chk2*-/- untreated (sections)	3.525	0.0065	**
P2 *Chk2*+/- untreated *vs* P2 *Chk2*-/- +AZT (sections)	4.911	0.0008	***
E15.5 WT B6 *vs* E15.5 *Chk2*-/- (N0)	11.43	<0.0001	***
E15.5 WT B6 *vs* E15.5 *Chk2*+/- (N1)	14.9	<0.0001	***
E15.5 WT B6 *vs* E15.5 *Chk2*+/+ (N1F1)	10.88	<0.0001	***
E15.5 WT B6 *vs* E15.5 *Chk2*+/- (N1F1)	13.3	<0.0001	***
E15.5 WT B6 *vs* E15.5 *Chk2*-/- (N1F1)	8.848	<0.0001	***
E15.5 WT B6 *vs* E15.5 *Chk2*-/- (N5F1)	1.451	0.1673	ns

P>0.05; (ns) - P>0.05; (*) - P<0.05; (**) - P<0.01; (***) - P<0.001 [a]analyzed by unpaired t-test

Table 2-9: Statistical analysis of oocyte number per ovary in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ untreated and AZT treated conditions. Statistical significance determine using unpaired Student's t-test.

Conclusions

Based on our observations described above, along with previous studies[69], I propose the following model to explain the molecular triggers and mechanisms of L1-driven FOA (Figure 2-7). Most evident in the early stages of FOA (E15.5-E18.5), but not exclusive to them, L1 reverse transcriptase activity generates intermediates such as L1 ssDNA that promote genotoxic stress alleviated by AZT treatment. Oocytes exceeding a lethal threshold of L1 reverse transcription intermediates reduce *CD55* expression, resulting in activation of the complement system of immunity and macrophage recruitment to clear affected oocytes. Oocytes that do not reach this threshold continue to express *CD55* at high levels to prevent complement attack. Still, an open question remains whether complement activation is a direct response to L1 reverse transcription intermediates, or indirect via activation of proapoptotic pathways. Support for the latter situation comes from decreased expression of pro-apoptotic Poly ADP-ribose polymerase 1 (PARP-1) and apoptosis regulator BAX in AZT-treated oocytes that avoid FOA compared to untreated oocytes. In the late stages of FOA (E18.5-P2), persisting DNA damage and common meiotic defects promoted by L1 endonuclease activity activate DNA damage checkpoint-mediated apoptosis that can be prevented by mutation of CHK2 (Figure 2-7). I report that the combined inhibition of retrotransposon L1 reverse transcriptase activity and the CHK2-dependent DNA damage checkpoint prevents FOA, thus preserving potentially the entire fetal oocyte population.

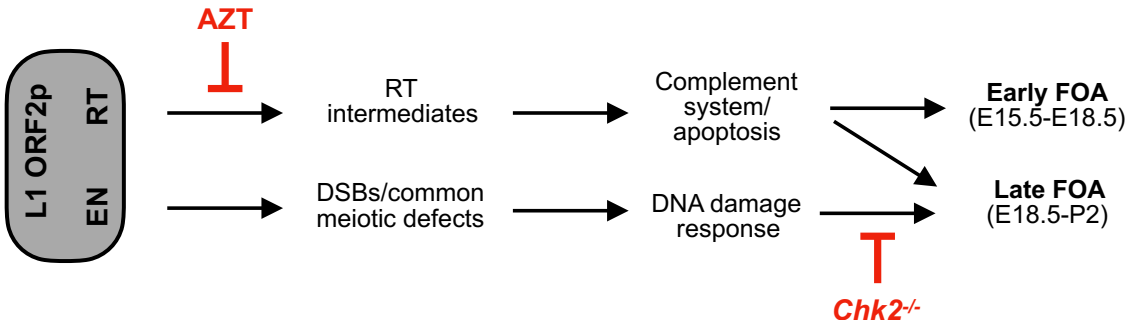


Figure 2-7: Model of triggers and mechanisms of FOA related to L1 ORF2p reverse transcriptase (RT) and endonuclease (EN) activities. In early stages of FOA, L1 RT intermediates whose formation are inhibited by AZT treatment lead to downstream activation of the complement system of immunity and proapoptotic pathways. In late stages of FOA, DNA damage and meiotic defects as a consequence of L1 EN activity throughout meiotic prophase I activates the DNA damage checkpoint through CHK2. Mechanisms in the early stages are also active in late stages based on the observation that FOA persists in late stages when AZT is administered to *Chk2*$^{-/-}$ mice in early stages exclusively.

CHAPTER 3

Biological relevance of fetal oocyte attrition

From "Maximizing the ovarian reserve by evading LINE-1 genotoxicity" (Fig. 3 and 4)

Introduction

Having established a system to prevent FOA into postnatal stages, we now can ask questions about the biological relevance of FOA and L1 expression for reproduction. My overarching question is whether prevention of FOA is beneficial for reproduction by increasing the ovarian reserve, or detrimental for reproduction by contaminating the oocyte supply with poor quality eggs that may lead to mutations in offspring? Here, I have investigated this question in two parts:

I. FOA as oocyte quality control

In Part I of this chapter, I investigated whether manipulating oocytes to evade FOA by our system results in a poorer quality ovarian reserve by enriching the oocyte supply with those containing highest levels of L1 and DNA damage or common meiotic defects that would normally be eliminated. Oocytes preserved by AZT treatment have been reported to contain highest amounts of genomic defects[69]. To investigate how oocytes handle this excess genome stress, I have followed the fate of these rescued oocytes further into development by combining AZT treatment and *Chk2*$^{-/-}$. I have determined the significance of the piRNA pathway, particularly, whether the non-essentiality of piRNAs in female germ cells compared to male germ cells is due to FOA already eliminating high L1-expressing oocytes that would benefit from piRNA-mediated L1 silencing. This is an outstanding question in the transposon and oogenesis field that is addressed using our system to evade FOA.

II. FOA in oocyte differentiation and physiology

Part II of this chapter investigates postnatal oogenesis and fertility in the absence of FOA. It is unknown whether FOA is an essential process for oocyte differentiation or even whether oocytes lost to FOA are of the same cellular identity as oocytes that normally survive. An outstanding question in the field is whether FOA in mice is analogous to nurse cell dumping in *Drosophila* that sacrifices nurse cells

to support the oocyte fated for survival. In this case, oocytes lost to FOA would represent nurse cells, and these presumptive nurse cells may now be observed in our system that prevents FOA. Here I have employed genomic, histological, and functional assays to assess the oocyte population upon FOA evasion. I have also considered the potential implications for FOA evasion by L1 and CHK2 inhibition in treatment of pre-mature ovarian failure in mammals as a novel therapeutic. Further, I have attempted to understand how L1 activity is important in this window of fetal oogenesis, perhaps for maintaining the adaptive potential in oocytes.

3.1 FOA as oocyte quality control

To understand consequences of preserving oocytes with substantial genotoxic stress in the ovarian reserve quality, I first confirmed that AZT-treated oocytes contain high levels of L1 ORF1p and meiotic defects by measuring the nuclear abundance of L1 ORF1p and γH2AX (DNA breaks) at E18.5, just prior to DNA damage checkpoint activation. These oocytes with high levels of defects are likely eliminated by the DNA damage checkpoint. Using *Chk2*$^{-/-}$ mice, we can now follow rescued oocytes longer to determine whether such defects are also observed at later stages.

Upon FOA evasion, oocytes ultimately reduce excessive genotoxic stress

I first quantified L1 expression and meiotic defects in our system by measuring fluorescence intensity of L1 ORF1p and γH2AX in individual E18.5 oocytes. I performed co-immunofluorescence with the germ cell marker TRA98, followed by DAPI counter-staining on ovarian sections, and took the relative mean nuclear (RMN) fluorescence intensity of the marker of interest in DAPI-positive regions of TRA98-positive cells. I then took the average of these values from a pool of approximately 200 oocytes for each genotype or treatment measured. Consistent with previous findings, a population of AZT-treated E18.5 *Chk2*$^{+/-}$ or *Chk2*$^{-/-}$ oocytes that prevent FOA in early MPI stages showed higher L1 ORF1p abundance compared to untreated E18.5 *Chk2*$^{+/-}$ or *Chk2*$^{-/-}$ controls (Figure 3-1A, C; Table 3-1). Of note, in my hands, these effects were observed only with the L1 ORF1p antibody against the full-length protein, but not with the one against the C-terminus only (data not shown). AZT-sensitive oocytes also contained highest nuclear levels of γH2AX in E18.5 *Chk2*$^{+/-}$ or *Chk2*$^{-/-}$ oocytes (Fig. 3-1A; Table 3-2). Therefore, FOA indeed serves as a quality control mechanism in early MPI stages between E15.5 and E18.5, prior to DNA damage checkpoint activation.

I have now extended our studies to observe the persistence of genotoxic stress in oocytes evading FOA beyond the time of birth to P2. I repeated experiments to measure nuclear L1 ORF1p and γH2AX in P2 oocytes. Interestingly, oocytes did not continue to accumulate defects. In fact, in all conditions, L1

ORF1p expression was significantly reduced, suggesting that even oocytes with highest levels of L1 that evade FOA can reduce L1 by a mechanism independent of cell death (Figure 3-1B,C; Table 3-1). γH2AX levels were also reduced at P2 compared to E18.5 in all conditions, agreeing with previous studies that show significant DNA repair capacity in *Chk2*$^{-/-}$ mutant oocytes (Fig. 3-1B; Table 3-2) [74]. This cell death-independent reduction of L1 and γH2AX levels indicates that while FOA acts as quality control for genome integrity, if given a chance, oocytes can ultimately reduce genotoxicity. Therefore, evasion of FOA can increase the ovarian reserve with oocytes of comparable quality to those naturally surviving.

To test whether this reduction of L1 occurs at the level of protein reflected by L1 ORF1p or L1 RNA, I isolated total RNA from untreated and AZT-treated oocytes from ovaries between E14.5 and E18.5, performed reverse transcription reactions to generate cDNA, and performed quantitative RT-PCR to measure abundance of L1 *ORF1*. I used L1 *ORF1* region for the analysis because it better reflects full-length L1 elements rather than those truncated at the 3' end during retrotransposition. At E15.5, E16.5, E17.5, and E18.5, L1 *ORF1* RNA was increased in AZT-treated oocytes compared to untreated, but it was statistically significant only at E17.5 (Figure 3-1D). At E14.5, L1 *ORF1* was the unchanged with AZT treatment. The difference between untreated and AZT-treated oocytes at E15.5, but not E14.5, is likely due to a slight increase in the speed of oocyte developmental timing and thus onset of FOA in CD1 mice, the strain used for this experiment, compared to C57Bl/6J (B6) strain. L1 *ORF1* expression was increased dramatically between E14.5 and E15.5, but was subsequently decreased by E18.5 (Figure 3-1D; Table 3-3). This reduction was observed in both untreated and AZT-treated oocytes, suggesting that the FOA-independent reduction of L1 occurs at the post-transcriptional rather than post-translational level.

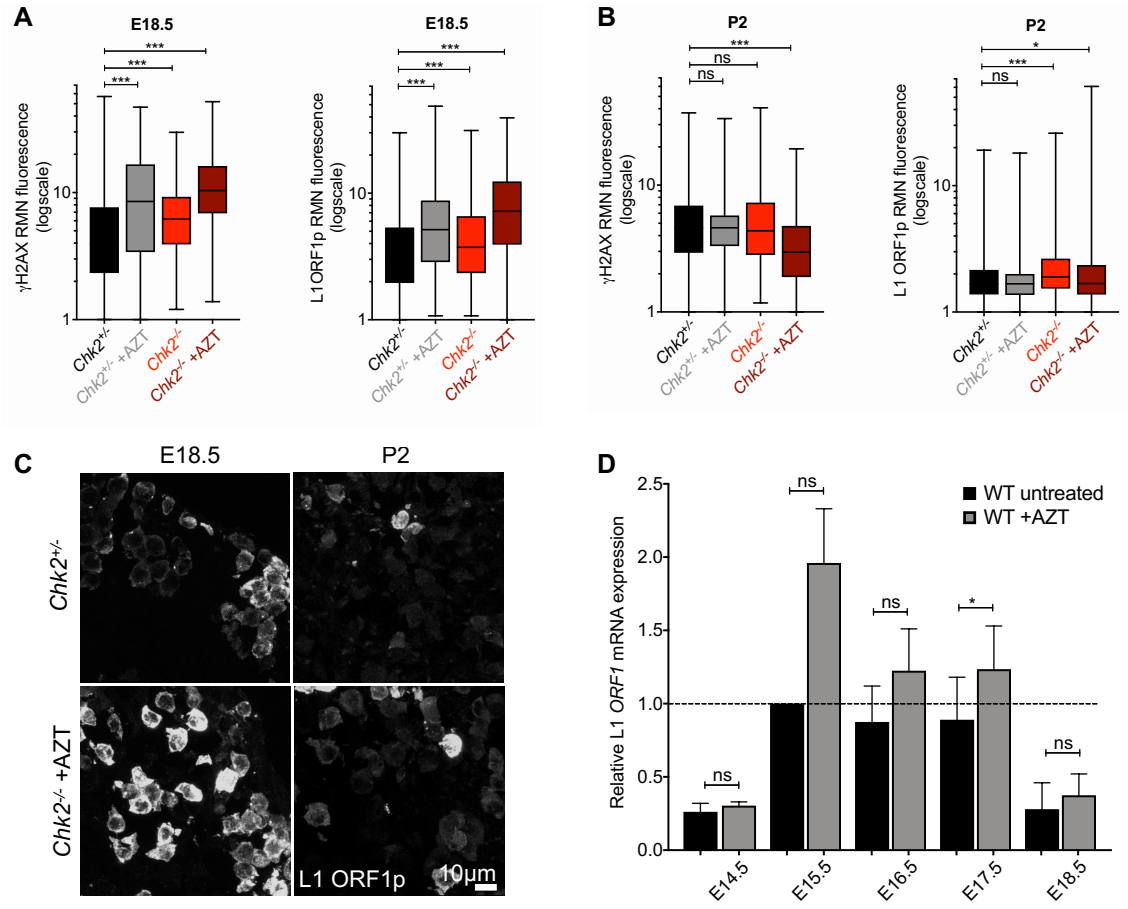


Figure 3-1: Reduction of genotoxicity upon FOA evasion. **A-B)** RMN fluorescence γH2AX and L1 ORF1p levels in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ untreated and AZT-treated oocytes at E18.5 **(A)** and P2 **(B).** Box plots indicate center line at median value, box limits at upper and lower quartile values, and whiskers at max and min values; n≥118 oocytes per sample. Stats by Kolmogorov-Smirnov test, ns p> 0.05; *p<0.05; ***p<0.001. **C)** L1 ORF1p immunofluorescence in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ +AZT ovary sections at E18.5 and P2. Scalebars:10μm. **D)** L1 *ORF1* mRNA expression in UT and AZT-treated oocytes by qRT-PCR. Normalized to E15.5 UT represented by dotted horizontal line. *Actb* gene used as endogenous control. Each bar represents two biological replicates. Stats by paired t-test, ns p>0.05; *p<0.05.

L1ORF1p relative mean nuclear levels in Chk2+/- and Chk2-/- untreated and AZT-treated

Samples	n embryos	n litters	n oocytes	Min	1st Quartile	Median	3rd Quartile	Max
E18.5 *Chk2+/-*	3	3	666	1	1.952	3.154	5.358	30.04
E18.5 *Chk2+/-* +AZT	3	2	366	1.074	2.854	5.157	8.755	48.73
E18.5 *Chk2-/-*	3	3	509	1.074	2.334	3.741	6.582	31.28
E18.5 *Chk2-/-* +AZT	3	3	439	1	3.919	7.193	12.38	39.4
P2 *Chk2+/-*	3	3	712	1	1.378	1.66	2.156	19.08
P2 *Chk2+/-* +AZT	3	2	150	1	1.363	1.672	2.002	18.12
P2 *Chk2-/-*	2	2	270	1	1.533	1.892	2.646	26
P2 *Chk2-/-* +AZT	3	3	505	1	1.376	1.682	2.356	61.01

Statistical analysis of L1ORF1p relative mean nuclear levels in Chk2+/- and Chk2-/- untreated and AZT-treated

Sample	D	p[c]	Significance
E18.5 *Chk2+/-* UT *vs* E18.5 *Chk2+/-* +AZT	0.2471	<0.0001	***
E18.5 *Chk2+/-* UT *vs* E18.5 *Chk2-/-* UT	0.1244	0.0003	***
E18.5 *Chk2+/-* UT *vs* E18.5 *Chk2-/-* +AZT	0.3944	<0.0001	***
P2 *Chk2+/-* UT *vs* P2 *Chk2+/-* +AZT	0.07811	0.4363	ns
P2 *Chk2+/-* UT *vs* P2 *Chk2-/-* UT	0.2035	<0.0001	***
P2 *Chk2+/-* UT *vs* P2 *Chk2-/-* +AZT	0.09459	0.0101	*

(ns) - P>0.05; (*) - P<0.05; (***) - P<0.001

[c]analyzed by Kolmogorov-Smirnov test

Table 3-1: RMN values and statistical analysis for L1 ORF1p in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ untreated and AZT-treated oocytes.

γH2AX relative mean nuclear levels in Chk2+/- and Chk2-/- untreated and AZT-treated

Samples	n embryos	n litters	n oocytes	Min	1st Quartile	Median	3rd Quartile	Max
E18.5 *Chk2+/-*	2	2	261	1	2.322	3.954	7.648	56.96
E18.5 *Chk2+/-* +AZT	2	2	222	1	3.399	8.517	16.67	47.01
E18.5 *Chk2-/-*	2	1	221	1.204	3.905	6.172	9.256	29.79
E18.5 *Chk2-/-* +AZT	2	2	347	1.385	6.864	10.38	16.19	51.94
P2 *Chk2+/-*	2	2	205	1	2.926	4.559	6.898	37.12
P2 *Chk2+/-* +AZT	2	1	118	1	3.305	4.608	5.761	33.44
P2 *Chk2-/-*	2	2	315	1.178	2.814	4.363	7.276	40.7
P2 *Chk2-/-* +AZT	2	2	335	1	1.884	2.97	4.768	19.31

Statistical analysis of γH2AX relative mean nuclear levels in Chk2+/- and Chk2-/- untreated and AZT-treated

Sample	D	p[c]	Significance
E18.5 *Chk2+/-* UT *vs* E18.5 *Chk2+/-* +AZT	0.312	<0.0001	***
E18.5 *Chk2+/-* UT *vs* E18.5 *Chk2-/-* UT	0.2645	<0.0001	***
E18.5 *Chk2+/-* UT *vs* E18.5 *Chk2-/-* +AZT	0.4889	<0.0001	***
P2 *Chk2+/-* UT *vs* P2 *Chk2+/-* +AZT	0.1044	0.3877	ns
P2 *Chk2+/-* UT *vs* P2 *Chk2-/-* UT	0.07123	0.5542	ns
P2 *Chk2+/-* UT *vs* P2 *Chk2-/-* +AZT	0.3016	<0.0001	***

(ns) - P>0.05; (***) - P<0.001

[c]analyzed by Kolmogorov-Smirnov test

Table 3-2: RMN values and statistical analysis for histone γH2AX in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ untreated and AZT-treated oocytes.

Quantitative RT-PCR for L1 *ORF1* in untreated (UT) and AZT-treated fetal oocytes (E14.5-E18.5)				
Replicate 1	**E14.5 UT**	**E14.5 +AZT**	**E15.5 UT**	**E15.5 +AZT**
Mean Cq *Actb*	23.78	24.17	25.20	25.20
Mean Cq L1 *ORF1*	19.69	20.04	21.70	21.03
Replicate 2	**E14.5 UT**	**E14.5 +AZT**	**E15.5 UT**	**E15.5 +AZT**
Mean Cq *Actb*	23.87	23.79	23.75	24.36
Mean Cq L1 *ORF1*	20.44	19.91	18.02	17.42
Relative L1 *ORF1* expression (normalized to E15.5 and *Actb*)	**E14.5 UT**	**E14.5 +AZT**	**E15.5 UT**	**E15.5 +AZT**
Replicate 1	0.32	0.33	1	1.59
Replicate 2	0.2	0.28	1	2.31
Statistical Analysis comparing untreated vs. AZT-treated	**E14.5 UT vs. E14.5 +AZT**		**E15.5 UT vs. E15.5 +AZT**	
p-value	0.4148		0.2306	
t	1.311		2.639	
Significance	ns		ns	

*Both replicates 1 and 2 for E14.5 and E14.5 +AZT were performed in same batch as and normalized to replicate 2 for E15.5.
*All other replicate 1 samples normalized to E15.5 replicate 1 and all replicate 2 samples normalized to E15.5 replicate 2.
*Stats determined from two-tailed paired Student's t-test using Relative L1 *ORF1* expression values.

E16.5 UT	**E16.5 +AZT**	**E17.5 UT**	**E17.5 +AZT**	**E18.5 UT**	**E18.5 +AZT**
25.53	25.52	24.39	25.15	21.94	22.32
22.70	21.43	21.62	21.74	21.79	20.92
E16.5 UT	**E16.5 +AZT**	**E17.5 UT**	**E17.5 +AZT**	**E18.5 UT**	**E18.5 +AZT**
23.59	22.51	23.33	24.17	22.74	22.64
17.70	16.87	17.36	17.83	18.13	17.85
E16.5 UT	**E16.5 +AZT**	**E17.5 UT**	**E17.5 +AZT**	**E18.5 UT**	**E18.5 +AZT**
0.63	1.51	0.6	0.94	0.1	0.23
1.12	0.94	1.18	1.53	0.46	0.52
E16.5 UT vs. E16.5 +AZT		**E17.5 UT vs. E17.5 +AZT**		**E18.5 UT vs. E18.5 +AZT**	
0.6284		0.0546		0.2247	
0.6604		11.63		2.714	
ns		*		ns	

Table 3-3: Relative quantities of mRNA in E14.5 through E18.5 untreated and AZT-treated oocytes

L1-targeting piRNA biogenesis during FOA

The reduction of L1 mRNA and ORF1p levels prompted me to examine the involvement of transposon defense mechanisms in active silencing of L1 during FOA. Given existing erasure of DNA methylation from epigenetic reprogramming, I chose to focus on the piRNA pathway, as robust downregulation of L1 RNA between E15.5 and E18.5 coincided with increased expression of genes encoding piRNA pathway machinery (Figure 3-2A). Further, Mili, the predominant Piwi family protein expressed in fetal oocyte was upregulated at the RNA and protein levels between E15.5 and P2 (Figure 3-2A, B)[60].

To determine whether the induction of the piRNA pathway corresponds with production of piRNAs, I isolated and sequenced small RNAs from wild-type CD1 ovaries at E15.5, E18.5, and P2. Sequencing was analyzed using piPipes program[133]. First, analysis of read length distribution identified two major species, one of 22-23 nucleotides in length (characteristic of endosiRNAs) and the other of 26-27 nucleotides in length (characteristic of piRNAs). Consistent with piRNA pathway activation at E18.5, RNAs of 26-27 nucleotides in length become highly abundant at E18.5 and P2 compared to E15.5, where RNAs of 22-23 nucleotides dominated the profile (Figure 3-2C) [134,135]. These piRNAs were produced in a Mili-dependent manner, because at P2, *Mili* mutant ovaries did not contain RNAs of 26-27 nucleotides, while wild-type ovaries at P2 showed a similar peak to that in CD1 P2 ovaries (Figure 3-2D). Interestingly, generation of piRNAs was not dependent on Maelstrom, an RNA-binding protein involved in the piRNA pathway[59]. *Mael* mutant ovaries showed similar profiles of small RNAs at P2, compared to *Mael* heterozygous*, Mili* wild-type, and CD1 wild-type P2 ovaries.

Next, I aligned small RNA reads to repetitive elements in the mouse genome to determine whether they target L1 elements. Indeed, a massive increase was observed in antisense small RNAs targeting evolutionarily young L1MdA and L1MdT elements at E18.5 and P2, compared to E15.5 (Figure 3-2E,F; Table 3-4). Again, generation of these piRNAs is Mili-dependent, as *Mili* mutant P2 ovaries had negligible amounts of 26-27 nucleotides-long small RNAs targeting any repeat element (Figure 3-2G).

Lastly, antisense piRNAs targeting IAP elements were also observed and were highly abundant (data not shown). Our small RNA analysis shows that the onset of L1 activation in fetal oocytes strongly correlates with generation of L1-targeting piRNAs.

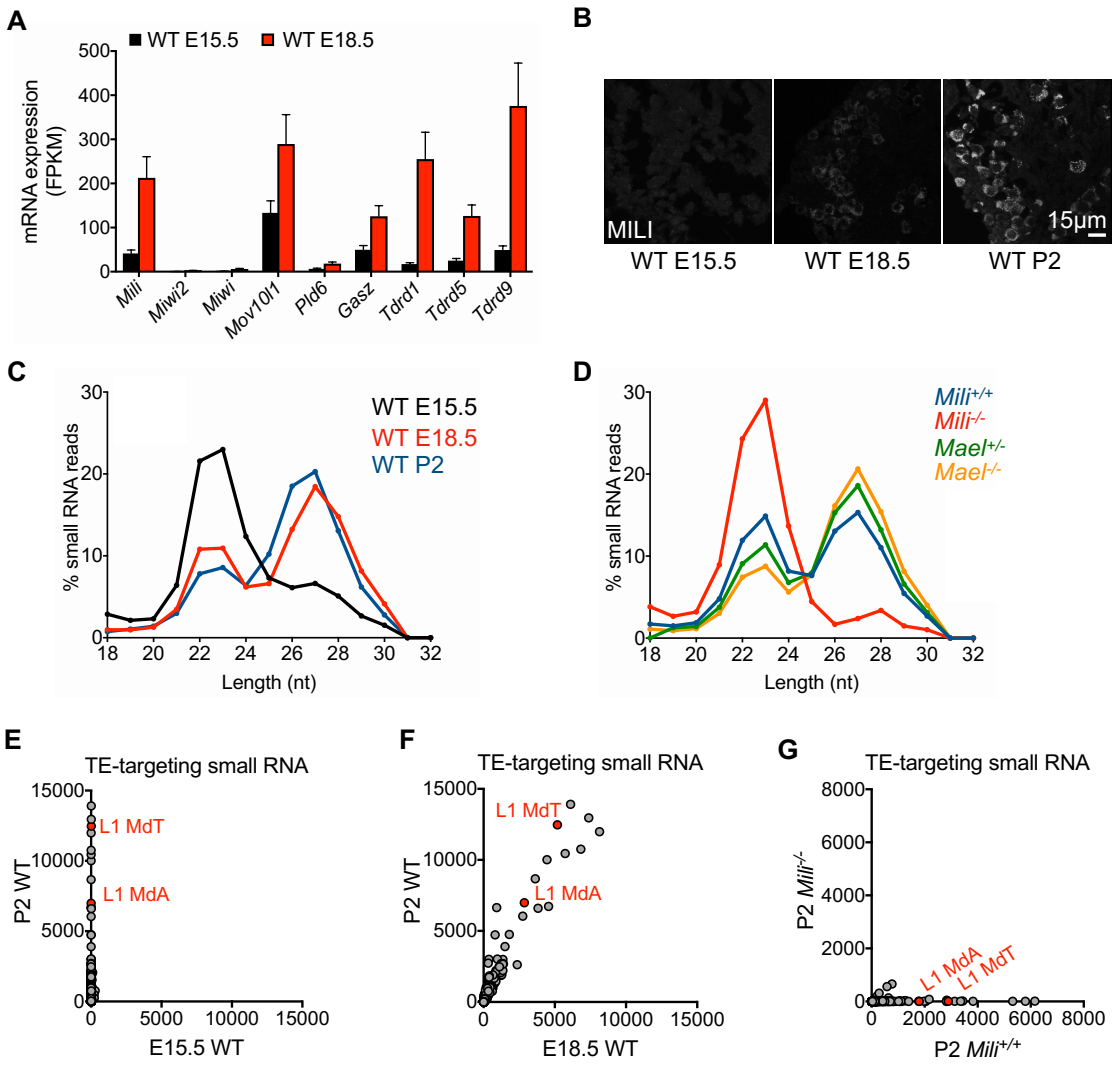


Figure 3-2: piRNA biogenesis during FOA. **A)** piRNA machinery gene expression in E15.5 and E18.5 oocytes. **B)** MILI expression in E15.5, E18.5, and P2 ovary sections. Scalebars:15μm. **C)** Length distribution of 18-32nt small RNA reads from E15.5, E18.5, and P2 WT ovaries. **D)** Length distribution of 28-32nt small RNA reads from P2 *Mili*$^{+/+}$, *Mili*$^{-/-}$, *Mael*$^{+/-}$, and *Mael* $^{-/-}$ ovaries. **E-G)** Antisense transposon-targeting small RNAs in E15.5vsP2 WT **(E)**, E18.5vsP2 WT **(F)**, and P2 *Mili*$^{+/+}$vsP2 *Mili*$^{-/-}$ **(G)** ovaries.

Abundance of small RNAs aligning to repetitive elements in E15.5 and P2 ovaries

Repeat Name	E15.5 sense	P2 sense	E15.5 antisense	P2 antisense
L1Md_A	11.68	6132.95	5.31	6987.44
L1Md_F	1.27	779.55	2.62	1112.59
L1Md_F2	119.61	13115.89	12.37	13914.63
L1Md_F3	3.01	2398.02	4.54	2992.09
L1Md_Gf	3.54	1398.88	1.11	457.4
L1Md_T	53.71	22740.71	16.83	12468.4

Abundance of small RNAs aligning to repetitive elements in E18.5 and P2 ovaries

Repeat Name	E18.5 sense	P2 sense	E18.5 antisense	P2 antisense
L1Md_A	6877.15	6132.95	2857.98	6987.44
L1Md_F	768.96	779.55	553.7	1112.59
L1Md_F2	9516.59	13115.89	6110.6	13914.63
L1Md_F3	1859.88	2398.02	1356.12	2992.09
L1Md_Gf	1991.62	1398.88	214.34	457.4
L1Md_T	32240.31	22740.71	5184.09	12468.4

Abundance of small RNAs aligning to repetitive elements in P2 Mili+/+ and Mili-/- ovaries

Repeat Name	Mili+/+ sense	Mili-/- sense	Mili+/+ antisense	Mili-/- antisense
L1Md_A	7830.27	27.83	1776.42	10.72
L1Md_F	383.93	2.75	282.78	1.46
L1Md_F2	6123.57	387.2	3437.91	19.38
L1Md_F3	1181.35	6.97	705.68	4.63
L1Md_Gf	867	5.84	107.16	4.7
L1Md_T	13730.71	87.52	2834.84	44.74

Table 3-4: Evolutionarily young L1 elements aligning to small RNA reads from E15.5, E18.5, P2, wild-type and P2 *Mili*$^{-/-}$ ovaries.

piRNAs are non-essential for silencing L1 upon FOA evasion

A persisting question is the functional relevance of piRNAs for female germ cells, as a strong sexual dimorphism in piRNA phenotypes exists resulting in male piRNA pathway mutant mice sterile while females can produce viable offspring. We wondered whether FOA masked a functional role for the piRNA pathway in females by eliminating high L1-expressing oocytes, negating the necessity for piRNAs. To answer this question, we turned back to our system to prevent FOA using *Chk2*$^{-/-}$+AZT. To this system, we added a *Mili* mutation to also remove piRNA production. We predicted that in oocytes evading FOA with highest L1 expression, piRNAs may be important to reduce this expression for survival. In normal conditions, this would be supplementary to FOA, perhaps to prune non-lethal levels of L1 existing in surviving oocytes. We quantified oocyte number in *Mili*$^{-/-}$*;Chk2*$^{-/-}$ +AZT ovaries and found that it is still significantly increased compared to that of *Mili*$^{+/-}$*;Chk2*$^{+/-}$ and *Mili*$^{-/-}$*;Chk2*$^{+/-}$ controls (Figure 3-3A; Table 3-5). Next, we quantified L1 ORF1p expression in *Mili*$^{-/-}$*;Chk2*$^{-/-}$ +AZT ovaries. *Mili*$^{-/-}$*;Chk2*$^{-/-}$ +AZT oocytes continue to reduce L1 ORF1p by P2, similar to controls (Figure 3-3B; Table 3-6).

Based on these experiments, generation of L1-targeting piRNAs observed is non-essential in normal conditions or in conditions with FOA evasion. Therefore, alternative or redundant mechanisms to accomplish L1 reduction remain uncovered. One possibility is the RNAi pathway, as endo-siRNA-length small RNAs were observed and a specific ovarian isoform of Dicer (DicerO) that processes siRNAs exists. However, *DicerO* mutant phenotypes manifest at later stages of oogenesis and appear unrelated to FOA [136]. Another attractive possibility is the extensive family of KRAB-ZFP proteins known to silence L1 and endogenous retroviruses in absence of other mechanisms [137,138]. A number of KRAB-ZFPs are expressed in fetal oocytes based on our RNA sequencing analysis.

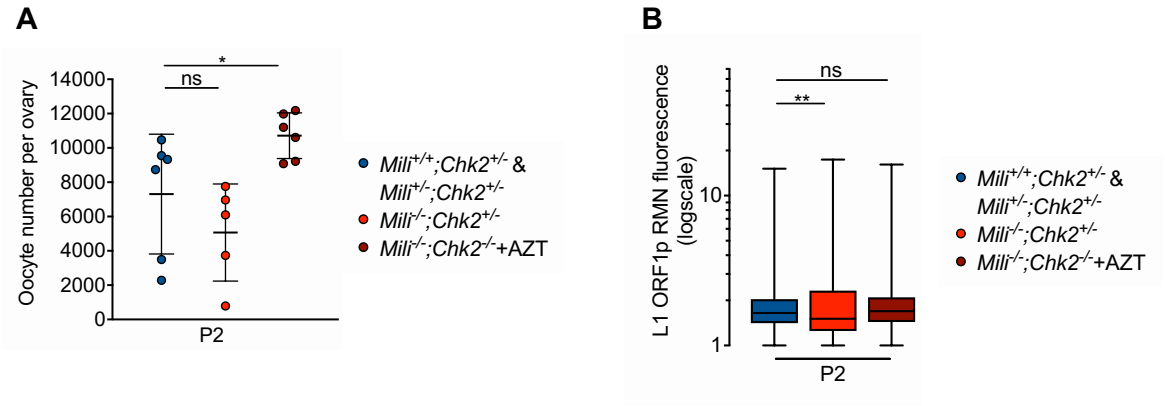


Figure 3-3: L1 ORF1p reduction is independent of FOA and piRNAs**. A)** Oocyte number in *Mili;Chk2* double mutant and control P2 ovaries. Dots indicate independent ovary samples; data are mean ± SD.; n≥3 ovaries per condition. Stats by two-tailed unpaired Student's t-test, ns $p>0.05$; *$p<0.05$. **B)** RMN fluorescence L1ORF1p levels in *Mili;Chk2* double mutant and control ovaries. Box plots indicate center line at median value, box limits at upper and lower quartile values, and whiskers at max and min values; n≥127 oocytes from at least 3 individual embryos per condition. Stats by Kolmogorov-Smirnov test, ns $p>0.05$; **$p<0.01$.

Oocyte number per ovary in Chk2;Mili ovaries at P2

Sample	Mean oocyte number	Std Dev	n ovaries	n embryos	n litters
Mili$^{+/+}$*;Chk2*$^{+/-}$ *(Chk2+/+)*	7306.3	3489.9	6	3	2
Mili$^{-/-}$*;Chk2*$^{+/-}$ *(Chk2+/+)*	5067	2828.2	5	3	2
Mili$^{-/-}$*;Chk2*$^{-/-}$ +AZT	10711.7	1332.1	6	3	1

Statistical analysis of oocyte number per ovary in Mili;Chk2 ovaries at P2

Sample	t-test	P[a]	Significance
Mili +/+ ;Chk2 +/- (Chk2+/+)	1.151	0.2793	ns
Mili +/+ ;Chk2 +/- (Chk2+/+)	2.233	0.0496	*

(ns) - P>0.05; (*) - P<0.05

[a]analyzed by unpaired t-test

Table 3-5: Quantification of oocyte number per ovary and statistical analysis of P2 untreated and AZT-treated *Mili*$^{-/-}$*;Chk2*$^{-/-}$ and control mice.

L1ORF1p relative mean nuclear fluorescence in Mili;Chk2 oocytes at P2

Samples	n embryos	n litters	n oocytes	Min	1st Quartile	Median	3rd Quartile	Max
Mili$^{+/+}$*;Chk2*$^{+/-}$ *(Chk2+/+)*	2	2	199	1	1.41	1.647	2.038	15.13
Mili$^{-/-}$*;Chk2*$^{+/-}$ *(Chk2+/+)*	2	2	150	1	1.249	1.507	2.323	17.41
Mili$^{-/-}$*;Chk2*$^{-/-}$ +AZT	2	1	127	1	1.435	1.695	2.094	16.13

Statistical analysis of L1ORF1p relative nuclear mean fluorescence in Mili;Chk2 oocytes at P2

Sample	D	P[c]	Significance
Mili +/+ ;Chk2 +/- (Chk2+/+)	0.1932	0.0034	**
Mili +/+ ;Chk2 +/- (Chk2+/+)	0.1324	0.1322	ns

(ns) - P>0.05; (**) - P<0.01

[c]analyzed by Kolmogorov-Smirnov test

Table 3-6: Values for L1 ORF1p RMN and statistical analysis in P2 untreated and AZT-treated *Mili*$^{-/-}$*;Chk2*$^{-/-}$ and control mice.

3.2 FOA in oocyte differentiation and physiology

The striking reduction of L1 expression and repair of DNA breaks in *Chk2*$^{-/-}$ +AZT oocytes prompted me to test their differentiation and developmental capacity into mature, functional follicles. If these processes are uninhibited, the ovarian reserve is effectively increased by evading FOA. Oocyte differentiation into a mature follicle involves dramatic changes in gene expression, growth, and interactions with surrounding somatic granulosa cells. I use Part II of this chapter to explore these aspects of oocyte development and physiology in the presence and absence of FOA.

Gene expression profiles of individual oocytes in the absence of FOA

I performed single-cell RNA sequencing to understand the diversity of oocyte populations present in normal conditions and upon FOA evasion. First, I dissociated whole ovaries from E18.5 wild-type mice and E18.5 mice treated with AZT. I hypothesized that if oocytes lost during FOA were of a fundamentally different identity or fate, this would be enriched in the AZT-treated sample and reflected upon clustering of oocyte transcriptomes. First, I subset oocytes from ovarian somatic cells from untreated and AZT-treated samples based on expression of oocyte-specific genes *Ddx4, Dazl,* and *Mael,* in addition to the lack of *Xist* expression that was restricted to somatic cells (Fig. 3-4A). I then integrated the untreated and AZT-treated oocyte datasets and found that the two datasets overlay each other without any deviating populations specific to a particular sample. Therefore, I conclude that oocytes eliminated during FOA do not have fundamentally different identities from those that survive at the transcriptional level (Fig. 3-4B).

Next, I performed a clustering T-SNE analysis on the integrated oocyte datasets. Oocytes were separated into eleven clusters; however, these clusters remained in contact with each other, suggesting that oocyte populations were not different types of cells, but existing in a trajectory of different states (Fig. 3-4C). I took a closer look at this trajectory by mapping gene expression to the clusters. Early oocyte marker gene *Ccnb3* was mainly expressed in clusters 0 and 1, which were the earliest oocytes

developmentally (Fig. 3-4D, G; Table 3-7). Late oocyte marker gene *Gdf9* that is expressed in primordial follicles and throughout follicle growth was found in clusters 9 and 10, which contain the latest oocytes developmentally (Fig. 3-4F, G; Table 3-7). Intermediate oocyte marker *Dppa3* was observed in all clusters, but expression was increased towards the later stage clusters (Fig. 3-4E, G; Table 3-7). While the same overall developmental trajectory was observed in untreated and AZT-treated oocytes, I wanted to determine whether either condition had an enrichment of oocytes in any particular cluster or part of the trajectory. Upon comparing percent of total untreated or AZT-treated oocytes belonging to each cluster, I found that early clusters are enriched for AZT-treated oocytes while late clusters are enriched for untreated oocytes (Fig. 3-H; Table 3-8). I hypothesized that this shift is due to prevention of FOA in early MPI stages by AZT-treatment, between E15.5 and E18.5. Therefore, sampling of oocytes for sequencing would enrich for developmentally earlier oocytes in the AZT-treated population. Further support of this shift is observed when comparing single-oocyte to bulk oocyte RNA sequencing data as described previously in Chapter 2. Bulk oocyte data confirms an increase in the expression of genes expressed in early MPI stages in AZT-treated oocytes compared to untreated at E18.5 (Fig. 3-4I). However, bulk oocyte RNA sequencing data did not reveal significant differential gene expression for *Ccnb3, Dppa3* and *Gdf9,* which were used as marker genes for the developmental trajectory in the single-oocyte analysis (Fig. 3-4I). These differences may only be revealed with looking at single-cell level.

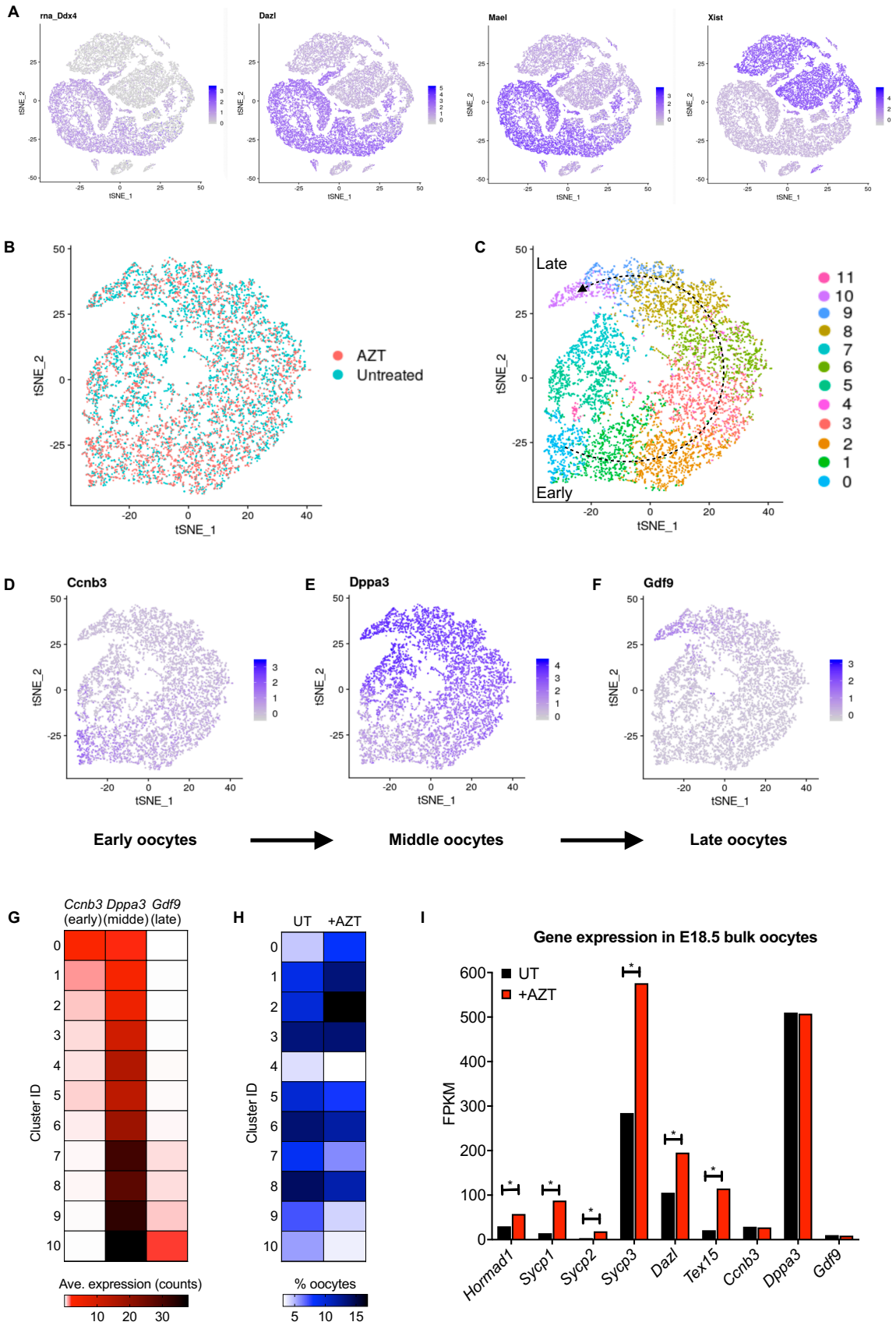

A
rna_Ddx4
Dazl
Mael
Xist
tSNE_1
tSNE_2
B
AZT
Untreated
C
Late
Early
11
10
9
8
7
6
5
4
3
2
1
0
D
Ccnb3
E
Dppa3
F
Gdf9
Early oocytes
Middle oocytes
Late oocytes
G
Ccnb3 Dppa3 Gdf9
(early) (midde) (late)
Cluster ID
Ave. expression (counts)
10 20 30
H
UT +AZT
% oocytes
5 10 15
I
Gene expression in E18.5 bulk oocytes
UT
+AZT
FPKM
Hormad1
Sycp1
Sycp2
Sycp3
Dazl
Tex15
Ccnb3
Dppa3
Gdf9

Figure 3-4: Oocyte populations that do not undergo FOA are slightly enriched for early MPI oocytes. **A)** T-SNE plot of integrated untreated and AZT-treated individual ovarian cells showing *Ddx4, Dazl, Mael*, and *Xist* expression used for sub-setting oocytes from somatic cells. Cells with average log fold change >0.5 for *Ddx4*, *Dazl*, and *Mael* and average log fold change expression <0.5 for *Xist* that is restricted to somatic cells were considered oocytes. **B)** T-SNE plot displaying integrated E18.5 UT (blue) and AZT-treated (orange) oocytes colored by original identities. **C)** T-SNE plot displaying cluster analysis of integrated E18.5 UT and AZT-treated oocytes. 11 clusters formed, cluster 11 removed from downstream analyses because top genes were mitochondrial, indicative of poor quality. Developmental trajectory based on known marker genes is indicated by arrows. **D-F)** Expression of early, middle, and late oocyte markers in integrated E18.5 untreated and AZT-treated sub-set oocytes. (**D)** *Ccnb3* is expressed in oocytes in early meiotic prophase I stages, (**E)** *Dppa3* is expressed in most oocytes and is used as a marker of oocytes in the middle of the developmental trajectory, and (**F)** *Gdf9* is expressed in late stage oocytes or primordial follicles. **G)** Heatmap of average expression of early (*Ccnb3*), middle (*Dppa3*), and late (*Gdf9*) marker genes within individual clusters. Clusters ordered from early to late developmental timepoints. **H)** Percent of oocytes in untreated or AZT-treated samples belonging to each cluster, ordered from early to late developmental timepoints. ~15,000 single-cells and ~50,000 reads/cell collected each from WT CD1 E18.5 untreated and AZT-treated ovaries from one litter of at least 6 pairs of ovaries each. Experiment was repeated one time for both untreated and AZT-treated samples, but with ~5000 oocytes per sample and ~7,000-10,000 reads/cell (data not shown). **I)** Differential gene expression (FPKM) of meiotic genes between E18.5 untreated and AZT-treated wild-type oocytes. ns $p>0.05$; *$p < 0.05$, stats determined by the Cuffdiff program in Cufflinks.

Top significant genes per oocyte cluster (ave_logFC > 0.2, p_val > 0.01)

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Ccnb3	Hbb-bt	Slc20a1	Prdm16	Hsf2	Hba-a1	Tbc1d24	Tuba1c
Gm424181	Hba-a1	Id1	Cd55	Marf11	Hbb-bs1	Id2	Tmsb101
Usp26	Hbb-bs	Marcksl1	Lhx8	Slc39a1	Tsc22d32	Cd55	Ldhb1
1700013H16Rik1	Tob11	Pou3f1		Slc24a5	Pcsk1n1	D10Wsu102e	**Dppa3**
Maoa1	Malat11	Id3		Insig1		Snhg111	Tmem14c1
Top2a	RP24-160N8.1	Crip2		Paip11		Ccno1	Skp1a
Lepr	Gapdh	Cdkn1c		Foxo3		Zbed31	Acat11
Scml2	Dmrtc21	Tob1		Actb3		Gdpd11	Hmgb1
Mtl5	Ftl11	Ftl1		Cbx3		Id11	Ooep1
Meioc1	Gm42418	Dmrtc2		Lrrc8a		Kit1	Cox7b1
1700080O16Rik	Maoa	1700031P21Rik		Bhlhe41		Atp2b11	Stk32a1
4933416C03Rik1	Pet2	Malat1		Myef2		Ninj11	Tuba1b
Agl	Hba-a2	Prdm16		Nmt22		Pcsk1n	Uchl1
Pet21	Slbp	Tsc22d3		Ric1		Ogt1	Sec61g
Exosc21	4933416C03Rik			Ptp4a23		Hes1	Rps4x
H2-D1	Jund			Alkbh52		Ybx21	Ndufb2
Tsc22d33	Tex19.2			Lyrm4		Sdc11	Ube2c1
Hist1h2ba	Tsc22d31			Set2		Taf101	Taldo11
Hspa2	Ly6k			Lsm14b3			Actb
Malat12	Slc20a11			C330006A16Rik			Ift271
H2bfm	Gadd45g			Id21			Tpd521
Plk4	Cdkn1c1			2810474O19Rik			Gtsf11
Gm7120	Myt1l			Ddah1			Rbakdn1
Ly6k1	Nxf2						Gbas
Tob12	Ccnb3						Mphosph61
Esx1	1700013H16Rik						Mat2b1
Topbp1	Tcea3						Kdm1b1
H1f01	Tex19.1						Gdpd12
Iqcb1	Meioc						2700094K13Rik1
Spdya	Cotl1						Crlf11
Rad51ap2	Irf2bpl						Tcl11

Cluster 8	Cluster 9	Cluster 10	*Cluster 11*
Zbed3	Gm15389	Ube2c3	*mt-Nd1*
Gdpd1	Tmsb102	Ldhb3	*mt-Cytb*
Dppa3	Ube2c2	Cd164l23	*mt-Nd4*
Tcl1	**Dppa3**	Gm153893	*mt-Atp6*
Dlgap4	Actb1	Stk32a3	*mt-Co3*
G3bp2	Ldhb2	Mphosph63	*mt-Nd2*
Crlf1	Mrps361	**Dppa3**	*mt-Co1*
Rnaset2a	Stk32a2	**Gdf9**	*mt-Nd3*
Bpgm	Eif4ebp12	Tmsb103	*Loxl2*
Prc1	Nos11	Unc13c1	*Glmp*
Tuba1c	Sdhaf4	2700094K13Rik3	*Lepr1*
Stat3	Tbc1d242	Ooep3	*Malat13*
Tmsb10	Hmgb32	Tubb2b	*Cd92*
Cd164l2	2700094K13Rik2	Mrps362	*Kit4*
Rdx	Dlgap41	4930562C15Rik1	*BC005561*
Ldhb	Rdx2	Sfr13	*Tma73*
Tmem14c	Gtsf12	Zp3	*Zp21*
Kit	Mphosph62	Rdx3	*Abi3bp3*
Ndufaf4	Tcl12	Nlrp4f3	*Ccnb1*
Arpc5l	Tuba1c2	Tuba1c3	*Rpl361*
Tpd52	Gbas1	Gtsf13	*Cox8a4*
Rbakdn	Rnaset2a2	Bfsp1	*Commd7*
Ppfibp2	Kit2	Dpf3	
Got2	Zfyve211	Zbed33	
0610012G03Rik	Gdpd13	Deptor1	
Lsm14b	Ndufaf41	Iqca1	
Sdc1	Ift432	Bpgm3	
Ninj1	Ptpn181	G3bp22	
Stk32a	Tmem14c2	Hsd17b102	
Pdzk1	Abi3bp1	Usp21	
Nos1	Eif1ax1	Actb2	

Table 3-7: Top 30 genes in single-cell clusters from integrated untreated and AZT-treated E18.5 oocyte sub-sets. Genes in bold letters indicate those used to determine developmental trajectory.

Number of oocytes per cluster by original sample identity

Sample	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Total
UT	194	396	426	580	64	170	430	592	371	618	306	235	4382
AZT	284	467	593	473	35	109	272	413	204	401	146	123	3520

Table 3-8: Number of oocytes per cluster belonging to untreated vs. AZT-treated original identity.

FOA is non-essential for Balbiani body formation during cyst breakdown

A prominent hallmark of oocyte differentiation is breakdown of germ cell cysts and the simultaneous growth of the Balbiani body in surviving oocytes. The Balbiani body is a conserved oocyte cytoplasmic aggregate of organelles such as Golgi, endoplasmic reticulum, and mitochondria, and RNAs that are important for supporting oocyte survival and early embryo development (Fig. 3-5A)[139]. Evidence from electron and fluorescence microscopy experiments to observe the Balbiani body in mouse ovaries suggests that its formation may be accomplished through transport of organelles from dying sister oocytes within a germ cell cyst[96,140]. I hypothesized that, if this was the case, the growth of the Balbiani body may be disturbed in the absence of FOA since sister oocytes are no longer dying.

To assess differentiation potential of oocytes evading FOA, I characterized the formation and growth of the Balbiani body as a proxy [96,140,141]. Balbiani body size based on Golgi content observed by immunofluorescence labeling with GM130 increased from 40 μm^2 at E18.5 to 70 μm^2 at P2 in *Chk2*$^{+/-}$ untreated oocytes as noted in other studies[96]. A similar increase in Golgi content between E18.5 and P2 was observed in untreated as well as AZT-treated *Chk2*$^{-/-}$ and *Chk2*$^{+/-}$ oocytes (Figure 3-5B, C; Table 3-9). Therefore, oocytes allowed to reach P2 upon FOA evasion, regardless of initial L1 expression level, have the potential to differentiate (Figure 3-5B, C).

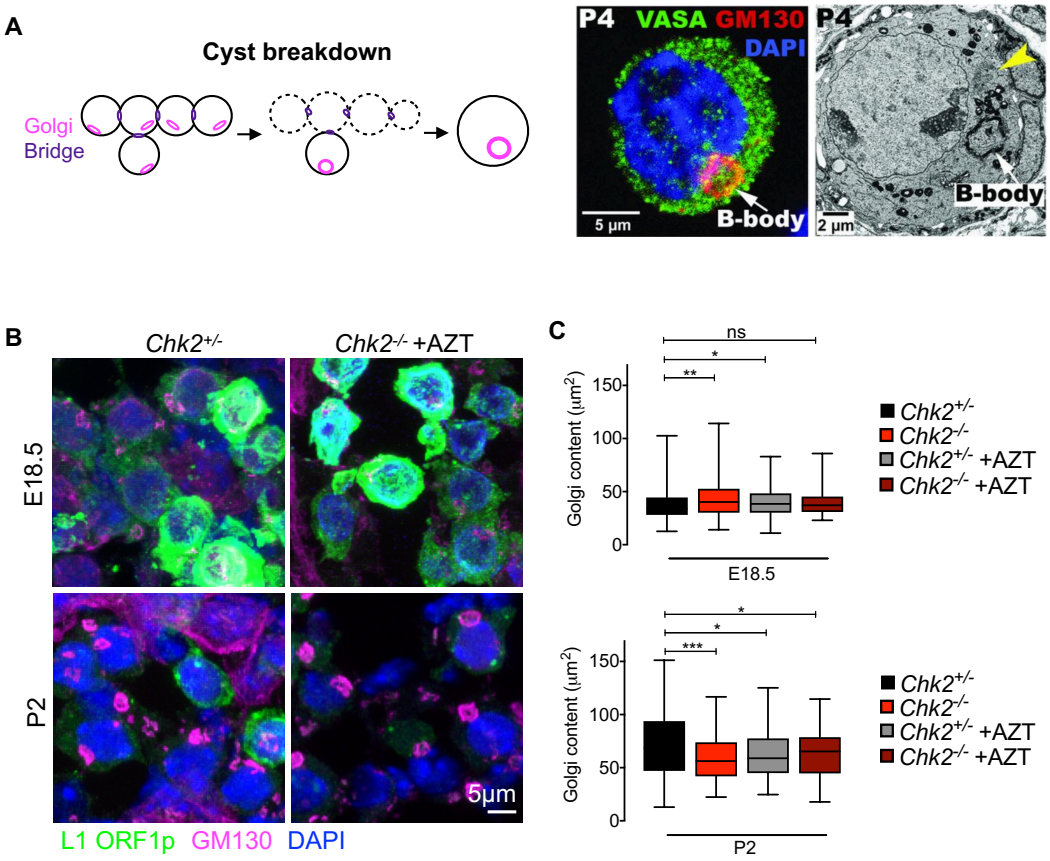


Figure 3-5: Balbiani body formation and growth does not require FOA. **A)** Diagram of events during cyst breakdown. Golgi elements become larger and part of an organelle cloud called the Balbiani body (pink). Intercellular bridges breakdown (purple)[142]. To the right are images of Balbiani body structure by immunofluorescence using GM130 (Golgi marker) and by electron microscopy[96]. **B)** GM130 and L1 ORF1p expression in representative images of *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ +AZT ovary sections at E18.5 and P2. Scalebars: 5µm. **C)** Golgi area (µm^2) in *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ untreated and AZT-treated E18.5, and P2 ovaries. Box plots indicate center line at median value, box limits at upper and lower quartile values, and whiskers at max and min values; n≥70 oocytes per sample. Stats by Kolmogorov-Smirnov test, ns $p>0.05$; *$p<0.05$; **$p<0.01$; ***$p<0.001$.

GM130 area in Chk2+/- and Chk2-/- untreated and AZT-treated

Samples	n embryos	n litters	n oocytes	Min	1st Quartile	Median	3rd Quartile	Max
E18.5 *Chk2+/-*	2	1	205	12.6	28.07	34.03	44.52	102.6
E18.5 *Chk2+/-* +AZT	2	1	141	10.91	30.18	38.47	48.32	83.01
E18.5 *Chk2-/-*	2	1	132	14.08	30.15	40.23	52.7	114.2
E18.5 *Chk2-/-* +AZT	2	2	74	22.95	30.84	37.17	45.31	85.88
P2 *Chk2+/-*	2	2	195	12.91	46.99	68.36	93.79	151.1
P2 *Chk2+/-* +AZT	2	1	130	24.74	44.93	58.86	77.62	125.2
P2 *Chk2-/-*	2	1	168	22.34	41.89	56.25	73.87	116.7
P2 *Chk2-/-* +AZT	2	2	70	17.82	44.6	65.35	78.72	114.6

Statistical analysis of GM130 area in Chk2+/- and Chk2-/- untreated and AZT-treated

Sample	D	p[c]	Significance
E18.5 *Chk2+/-* UT *vs* E18.5 *Chk2+/-* +AZT	0.162	0.0249	*
E18.5 *Chk2+/-* UT *vs* E18.5 *Chk2-/-* UT	0.1895	0.0063	**
E18.5 *Chk2+/-* UT *vs* E18.5 *Chk2-/-* +AZT	0.1546	0.1487	ns
P2 *Chk2+/-* UT *vs* P2 *Chk2+/-* +AZT	0.1641	0.03	*
P2 *Chk2+/-* UT *vs* P2 *Chk2-/-* UT	0.2227	0.0003	***
P2 *Chk2+/-* UT *vs* P2 *Chk2-/-* +AZT	0.1905	0.0476	*

(ns) - P>0.05; (*) - P<0.05; (**) - P<0.01; (***) - P<0.001

[c]analyzed by Kolmogorov-Smirnov test

Table 3-9: Values for Golgi element area and statistical analysis in untreated and AZT-treated *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ ovaries.

FOA is non-essential for folliculogenesis and fertility

Ovarian follicle development involves coordinated processes of oocyte maturation and proliferation of neighboring somatic cells that provide nutrients and signals such as Kit and steroid hormones to support the oocyte[143]. Follicle assembly begins with diplotene-arrested oocytes being surrounded by a flattened layer of somatic granulosa cells to form primordial follicles[144]. Subsequently, oocytes enlarge and granulosa cells convert from a flat to cuboidal shape to generate primary follicles[144]. Upon further proliferation and accumulation of granulosa cell layers, primary follicles transition into secondary (pre-antral) follicles[144]. Finally, a fluid-filled cavity forms, marking the transition from pre-antral to antral follicles, many of which undergo atresia[144]. An ovulated oocyte is released from its antral follicle, leaving behind a mass of somatic cells called a corpus luteum[144].

Given the importance of oocyte-soma communication for folliculogenesis, I hypothesized that an increased ovarian reserve upon FOA evasion might impede folliculogenesis by skewing the ratio of oocytes to granulosa cells produced. To test this hypothesis, I used immunohistochemistry to visualize follicle types and to quantify follicle number in normal conditions and upon FOA evasion. First, I characterized follicle types present at two postnatal stages in control ovaries and those that have evaded FOA: 1) at P4, during the primordial to primary follicle transition, and 2) at P19, when all follicle types can be observed, but prior to ovulation. I observed all characteristic growing follicle types of P4 and P19 ovaries in *Chk2*$^{+/-}$, *Chk2*$^{-/-}$, and *Chk2*$^{-/-}$ +AZT females, suggesting that FOA is not essential for folliculogenesis, nor does its evasion impede folliculogenesis (Figure 3-6A). I next quantified the number of follicles per ovary at P4 and P19, and classified them as primordial or non-primordial (primary, secondary or antral) follicles. Importantly, quantification of primordial and non-primordial follicles revealed an increase in follicles in *Chk2*$^{-/-}$ +AZT upon FOA evasion in both P4 and P19 stages (Figure 3-6A, B; Table 3-10). Therefore, increasing the oocyte supply by preventing FOA can increase the ovarian reserve into juvenile aged mice. Interestingly, the major difference between follicle number between *Chk2*$^{+/-}$ ovaries that experienced FOA and *Chk2*$^{-/-}$ +AZT ovaries that avoided FOA was in the primordial

follicle population. This increase in follicle number observed in ovaries that have evaded FOA compared to normal conditions was diminished in ovaries of ten-month-old females that had been continuously producing litters during this time span (Fig. 3-6C, D).

Finally, to test the effect of increased follicle number on fertility, I crossed females born to *Chk2*$^{-/-}$ +AZT-treated females that have reached reproductive maturity to *Chk2*$^{+/-}$ males and monitored the number of litters and litter size compared to untreated *Chk2*$^{-/-}$ and *Chk2*$^{+/-}$ females. Both litter number per female and pups per litter over 10 months was comparable among the three conditions, suggesting that FOA is not obligatory for fertility, nor is it detrimental to have an increased number of follicles in the ovarian reserve (Figure 3-6E, F; Table 3-11).

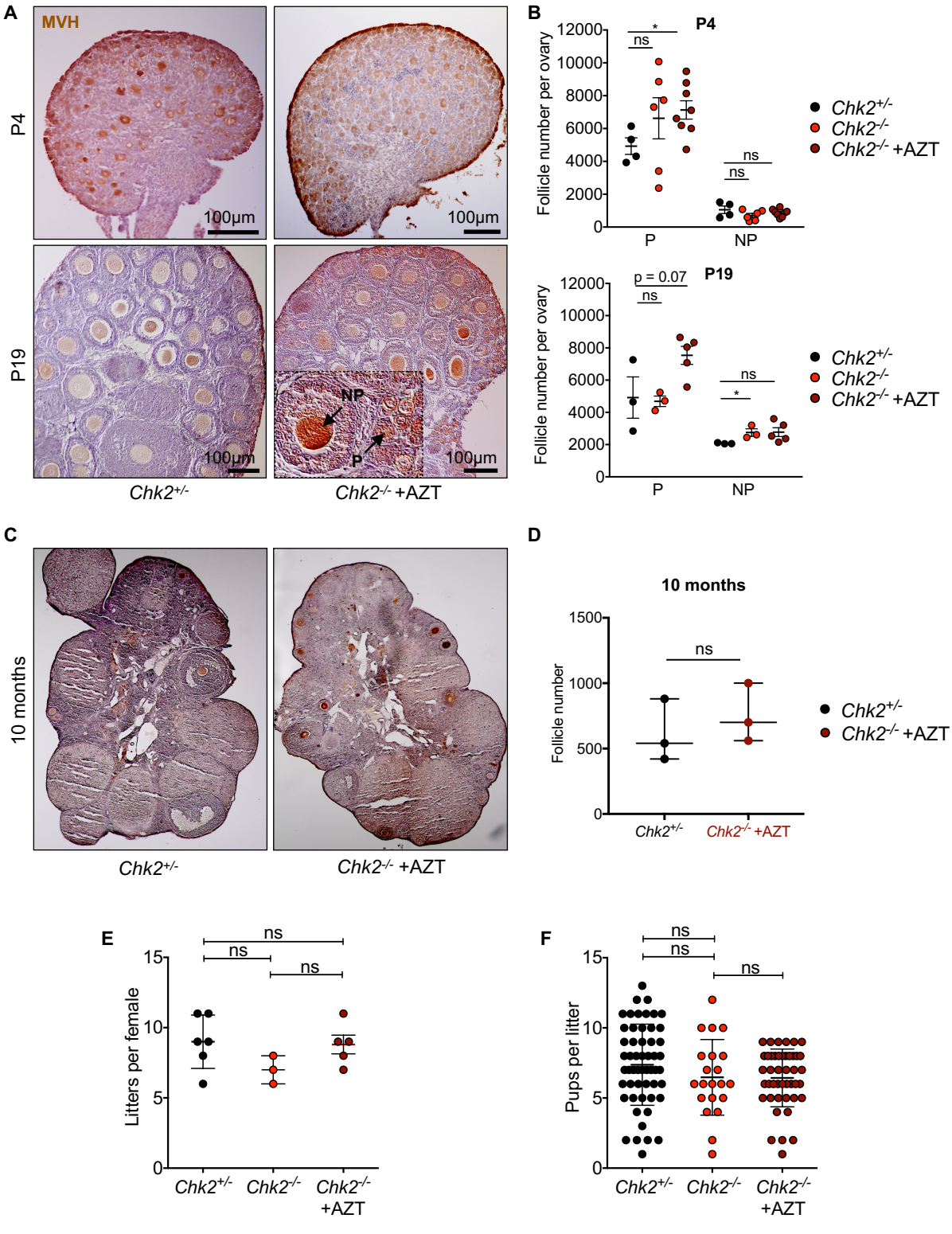

A
MVH
P4
P19
100µm
NP
P
Chk2+/-
Chk2-/- +AZT
B
P4
P19
Follicle number per ovary
ns
p = 0.07
*
P
NP
Chk2+/-
Chk2-/-
Chk2-/- +AZT
C
10 months
D
10 months
Follicle number
E
Litters per female
F
Pups per litter

Figure 3-6: Folliculogenesis and fertility upon FOA evasion. **A)** Mouse vasa homolog (MVH) and hematoxylin staining of representative sections of P4 and P19 *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ +AZT ovaries. Scalebars:100μm. Inset in P19 *Chk2*$^{-/-}$ +AZT shows an example of a primordial (P) vs. non-primordial follicle (NP). **B)** Quantification of P and NP follicles in P4 and P19 ovaries. Dots indicate independent ovary samples; data are mean ± s.e.m.; n≥3 ovaries per sample. Stats by two-tailed unpaired Student's t-test, ns p>0.05; *p≤0.05. **C)** MVH and hematoxylin staining of representative section of 10 month old *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ +AZT ovaries. **D)** Quantification of total follicles in 10 month old ovaries. Dots indicate independent ovary samples; data are mean ± s.e.m.; n=3 ovaries per sample, each ovary from an independent female. Stats by two-tailed unpaired Student's t-test, ns p>0.05. **E)** Number of litters per female over 10 months from *Chk2*$^{+/-}$, *Chk2*$^{-/-}$, and *Chk2*$^{-/-}$ +AZT females crossed to *Chk2*$^{+/-}$ males. Dots indicate independent females; data are mean ± SD; n≥3 females per condition. Stats by two-tailed unpaired Student's t-test, ns p>0.05. **F)** Pups per litter from *Chk2*$^{+/-}$, *Chk2*$^{-/-}$, and *Chk2*$^{-/-}$ +AZT females crossed to *Chk2*$^{+/-}$ males. Dots indicate independent litters; data are mean ± SD; n≥21 litters per condition. Stats by two-tailed unpaired Student's t-test, ns p>0.05.

Follicle number per ovary in Chk2+/- untreated and Chk2-/- untreated and AZT-treated mice

Sample	P follicles	SEM P	NP follicles	SEM NP	Total	SEM total	n ovaries	n mice
P4 Chk2+/-	4929	501.7	1053	231.3	5981	520.5	4	3
P4 Chk2-/-	6623	1252	705	125.5	7328	1303	6	3
P4 Chk2-/- +AZT	7130	558.6	863.8	82.38	7994	591.7	8	4
P19 Chk2+/-	4917	1283	2077	21.86	6993	1264	3	3
P19 Chk2-/-	4687	326.4	2750	230.3	7430	209.5	3	3
P19 Chk2-/- +AZT	7537	557.5	2771	272	10308	464.8	5	3

Statistical analysis for follicle number per ovary in Chk2+/- and Chk2-/- untreated and AZT-treated ovaries

Sample	t-test	P[a]	Significance
P4 Chk2+/- vs P4 Chk2-/- P follicles	1.049	0.325	ns
P4 Chk2+/- vs P4 Chk2-/- +AZT P follicles	2.511	0.031	*
P4 Chk2+/- vs P4 Chk2-/- NP follicles	1.442	0.187	ns
P4 Chk2+/- vs P4 Chk2-/- +AZT NP follicles	0.964	0.358	ns
P4 Chk2+/- vs P4 Chk2-/- total follicles	0.801	0.446	ns
P4 Chk2+/- vs P4 Chk2-/- +AZT total follicles	2.174	0.054	*
P19 Chk2+/- vs P19 Chk2-/- P follicles	0.174	0.871	ns
P19 Chk2+/- vs P19 Chk2-/- +AZT P follicles	2.191	0.071	ns
P19 Chk2+/- vs P19 Chk2-/- NP follicles	2.911	0.044	*
P19 Chk2+/- vs P19 Chk2-/- +AZT NP follicles	1.912	0.191	ns
P19 Chk2+/- vs P19 Chk2-/- total follicles	0.341	0.75	ns
P19 Chk2+/- vs P19 Chk2-/- +AZT total follicles	2.983	0.025	*

(ns) - P>0.05; (*) - P<0.05

[a]analyzed by unpaired t-test

Table 3-10: Quantification and statistical analysis of primordial (P), non-primordial (NP) and total follicles per ovary in untreated and AZT-treated *Chk2*$^{+/-}$ and *Chk2*$^{-/-}$ conditions.

Number of litters for Chk2+/- untreated and Chk2-/- untreated and AZT-treated females

Genotype	Mean number of litters	Std. Dev.	n females
Chk2+/-	9	1.897	6
Chk2-/-	7	1	3
***Chk2-/-* +AZT**	8.8	1.483	5

Statistical analysis for number of litters for Chk2+/- untreated and Chk2-/- untreated and AZT-treated females

Sample	t-test	P[a]	Significance
Chk2+/- vs Chk2-/-	1.673	0.1382	ns
Chk2+/- vs Chk2-/- +AZT	0.1914	0.8525	ns
Chk2-/- vs Chk2-/- +AZT	1.837	0.1158	ns

(ns) - P>0.05

[a]analyzed by unpaired t-test

Pups per litter for Chk2+/- untreated and Chk2-/- untreated and AZT-treated females

Sample	Mean pups per litter	Std. Dev.	n females	n litters
Chk2+/-	7.37	2.89	6	54
Chk2-/-	6.391	2.658	4	21
Chk2-/- +AZT	6.432	2.062	5	44

Statistical analysis for pups per litter for Chk2+/- untreated and Chk2-/- untreated and AZT-treated females

Sample	t-test	P[a]	Significance
Chk2+/- vs Chk2-/-	1.392	0.1679	ns
Chk2+/- vs Chk2-/- +AZT	1.811	0.0733	ns
Chk2-/- vs Chk2-/- +AZT	0.069	0.9452	ns

(ns) - P>0.05

[a]analyzed by unpaired t-test

Table 3-11: Fertility assays and statistical analyses including quantification of number of litters per female and number of pups per litter for *Chk2*$^{+/-}$, *Chk2*$^{-/-}$ and *Chk2*$^{-/-}$ +AZT females crossed to *Chk2*$^{+/-}$ males after ten months.

Conclusion

Cumulatively, our evidence supports the hypothesis that FOA acts as quality control for the genome to prevent transmission of damaged genetic material at the hands of the predominant genotoxic threat, L1 [145]. This leads to my first major finding that in the absence of repressive DNA methylation, FOA, and piRNAs, oocytes can still extinguish L1 expression and common meiotic defects. Given that cells go to great lengths to protect themselves from transposons and have evolved diverse and numerous silencing mechanisms, which mechanisms fetal oocytes are using is a fascinating open question. It has been described previously that piRNAs are non-essential for fertility in oocytes, and thus FOA was used as an explanation, since oocytes with dangerous levels of L1 are likely to be killed[60]. Here I disproved this hypothesis, showing that L1 expression continues to decrease in *Mili*$^{-/-}$*;Chk2*$^{-/-}$ +AZT oocytes that do not generate piRNAs nor undergo FOA. Since piRNAs are involved in *de novo* methylation in male gonocytes, perhaps significant piRNA production in oocytes would lead to premature re-methylation and is therefore suppressed. I speculate that the RNAi pathway may be involved in downregulating L1, as endo-siRNA-length small RNAs were observed. However, the Dicer mutant phenotypes manifest at later stages of oogenesis[136]. Another attractive possibility is that some members of the extensive family of KRAB-ZFP proteins repress L1 in absence of other mechanisms[137,138].

My second major finding was that folliculogenesis and fertility were not impacted by FOA prevention. Even up to 19 days old, the ovarian reserve remained significantly increased when comparing total follicle number between AZT-treated *Chk2*$^{-/-}$ females compared to controls. Therefore, our data support FOA as a non-obligatory program for oogenesis and that all oocytes generated have the potential to differentiate and are not fundamentally different in identity or developmental fate. This excludes hypotheses that require germ cell apoptosis for proper oogenesis, a developmental program exemplified in organisms such as *Drosophila* and *C. elegans*[146]. However, more precise methods allowing specific labeling of rescued vs normally surviving oocytes and eventually offspring rather than enriching for them are necessary for this conclusion. Further support that FOA is non-essential comes from ex-vivo studies

of ESC or iPSC-derived germ cells combined with appropriate somatic cells that can form primordial follicle-like structures without any opportunity for FOA[147].

CHAPTER 4

The origin of differential L1 expression in fetal oocytes

Introduction

Heterogeneity of L1 expression between individual oocytes determines their fate. A critical question remains as to the developmental basis of this heterogeneity. This chapter presents a collection of work that was a collaborative effort with Dr. Safia Malki, exploring three different oogenesis programs with a potential for underlying the establishment of differential L1:

1. *Epigenetic reprogramming*
2. *Asynchronous meiotic entry*
3. *Distribution of cytoplasmic contents in germ cell cysts*

4.1 Epigenetic reprogramming

Given the intimate relationship between DNA demethylation during epigenetic reprogramming and L1 derepression, we hypothesized that differential levels of DNA methylation may underlie differential L1 expression among fetal oocytes. Indeed, the timing and extent of DNA methylation erasure during epigenetic reprogramming is critical for germ cell development. This is exemplified using a conditional mutation of the DNA methyltransferase 1 (DNMT1) that results in premature expression of locus-specific genes that are normally demethylated during the second wave of demethylation[8]. Premature expression of such genes resulted in precocious germ cell differentiation and meiotic entry and ultimately, death. Interestingly, any link to L1 expression was not observed, despite L1 demethylation occurring in the second wave of demethylation as well.

We tested the connection between DNA methylation and L1 expression in two parallel ways: 1) in FACS-sorted, live, untreated and AZT-treated oocytes at E18.5 where the AZT-treated population is enriched for high L1 ORF1p-expressing oocytes that are normally killed, and 2) in FACS-sorted, fixed, immunostained oocytes that were isolated based on high and low levels of L1 ORF1p at E18.5. For all samples, genomic DNA was extracted and bisulfite-converted so that unmethylated cytosine becomes uracil, while methylated cytosine remains as cytosine. A specific amplicon covering 9 CpGs in the 5' UTR of an active L1 TF element was PCR amplified and sequenced by Wide-Seq performed by the Purdue Genomics Core (Fig. 4-1A). Wide-Seq produced between 1000-4000 reads per sample, and the average % methylation was determined for each of these samples by aligning reads to the genome and quantifying methylated vs unmethylated CpGs using Bismark. In untreated and AZT-treated E18.5 oocytes populations, there was a decrease in average methylation level in the AZT-treated oocytes. E15.5 oocytes showed a similar level of methylation as the AZT-treated E18.5 oocytes, suggesting that oocytes rescued from FOA by AZT treatment with highest levels of L1 are those containing the lowest levels of DNA methylation (Fig. 4-1B). We confirmed that AZT-treatment did not impact *de novo* DNA methylation by determining average % methylation in male untreated and AZT-treated prospermatogonia,

which, unlike oocytes, restore methylation prior to birth. *De novo* remethylation occurred normally in both untreated and AZT-treated male prospermatogonia between E13.5 and E18.5 (Fig. 4-1B).

The anti-correlation between L1 expression and DNA methylation levels in oocytes was further established by sequencing the L1 TF amplicon in oocytes containing high versus low levels of L1 ORF1p at E18.5 (Fig. 4-1C). Indeed, oocytes isolated based on having highest levels of L1 ORF1p contained the lowest levels of DNA methylation and vice versa (Fig. 4-1D, E). Samples containing a mixture of high and low L1 ORF1p-expressing oocytes possessed intermediate levels of DNA methylation.

To determine whether the relationship between amount of DNA methylation and L1 expression level extended to expression levels of other genes, whole-genome bisulfite sequencing will be performed on similarly FACS-sorted untreated and AZT-treated oocytes. Ovarian somatic cell populations will be included as controls as they should be highly methylated. I hypothesize that given the importance of timing and efficiency of DNA methylation erasure during epigenetic reprogramming, L1 may serve as a biomarker for oocytes that do not reprogram well[8].

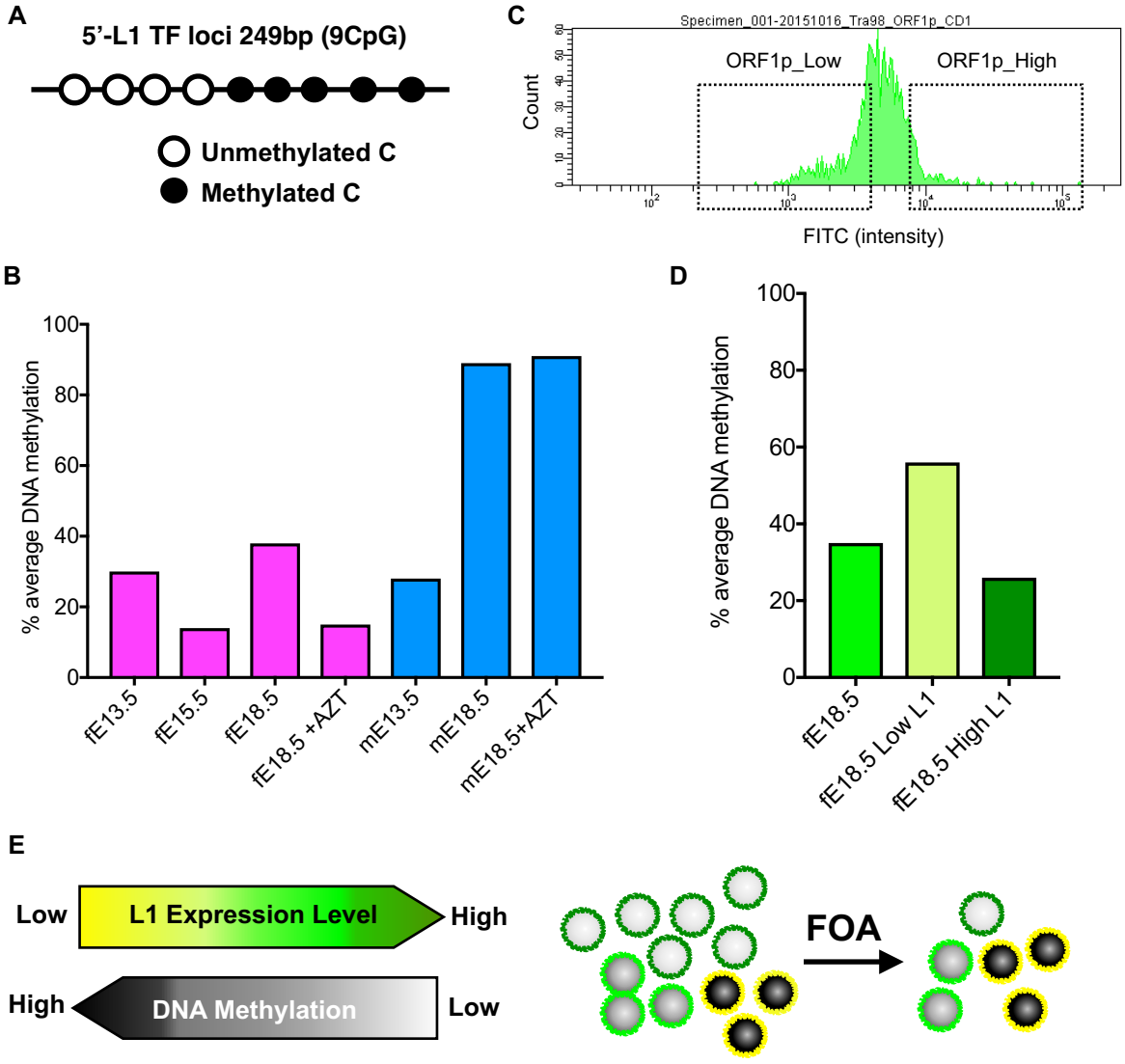


Fig. 4-1: L1 DNA methylation in fetal oocytes. **A)** Diagram of 9 CpGs in L1 TF element interrogated in bisulfite PCR experiments. **B)** Average methylation of 9 CpGs in L1 TF element determined by bisulfite PCR and wide-sequencing in sorted female germ cells from E13.5, E15.5, E18.5, and E18.5 +AZT and male germ cells at E13.5, E18.5 and E18.5 +AZT. **C)** Scheme to isolate oocytes containing high and low levels of L1 ORF1p by FACS. FITC intensity shown on the X axis and ORF1p low population taken from first one-third of FITC intensity and ORF1p high population from last one-third of FITC intensity. **D)** Average methylation of 9 CpGs in L1 TF element determined by bisulfite PCR and wide-sequencing in

sorted female germ cells at E18.5 by high and low L1 ORF1p expression and combined. **E)** Model reflecting the above data that DNA methylation levels are anti-correlated with L1 expression levels in oocytes, and thus oocytes with lowest methylation of L1 elements and highest L1 expression undergo FOA. *Data for Fig. 4-1 obtained by Safia Malki.*

4.2 Meiotic entry

Meiotic entry is an orchestrated process that involves complex interactions with germ cells and their somatic gonad environment. As described in the introduction, meiotic initiation occurs in an anterior to posterior wave throughout the gonad between E12.5 – E16.5 (Fig. 4-2A)[63]. Therefore, the spatial relationship of germ cells within the gonad, in other words, the proximity to the meiosis-inducing factor such as retinoic acid, is critical for meiotic initiation. This process is independent of germ cell number as well as intrinsic germ cell differentiation programs[64,148]. We wondered whether the timing of meiotic entry was involved in differential L1 methylation and expression. L1 is passively demethylated during PGC proliferation and diluted with each replication. Perhaps, oocytes that enter meiosis last do additional rounds of replication and dilution of DNA methylation resulting in their preferential FOA.

I developed a method to label oocytes that enter meiosis last in the posterior ovary by injecting 5-ethynyl-2'-deoxyuridine (EdU) to label these oocytes while they are still replicating. Oocytes that enter meiosis first in the anterior ovary will not take up the EdU as they are no longer replicating. After testing a number of different injection times, I found that injection at E14.0 was the best to label the posterior half of the ovary at E14.5 (Fig. 4-2B). Next, I wanted to test whether posterior oocytes possessed highest levels of L1 expression as a result of potentially further diluting DNA methylation. However, upon tracing EdU-positive cells to E18.5, I observed EdU-positive, TRA98-positive oocytes with both high and low levels of L1 ORF1p as well as EdU-negative, TRA98-positive oocytes with both high and low levels of L1 ORF1p (Fig. 4-2C). These experiments suggest that L1 expression level is independent of meiotic entry. To support this claim, I performed quantitative RT-PCR on anterior and posterior regions of E15.5 ovaries (anterior and posterior one-third of the ovary in attempt to lessen contamination). Similarly, I observed no difference in L1 *ORF1* mRNA nor *Sycp3* mRNA expression between anterior and posterior regions while *Stra8* expression was significantly increased in the posterior as expected at E15.5 (Fig. 4-2D).

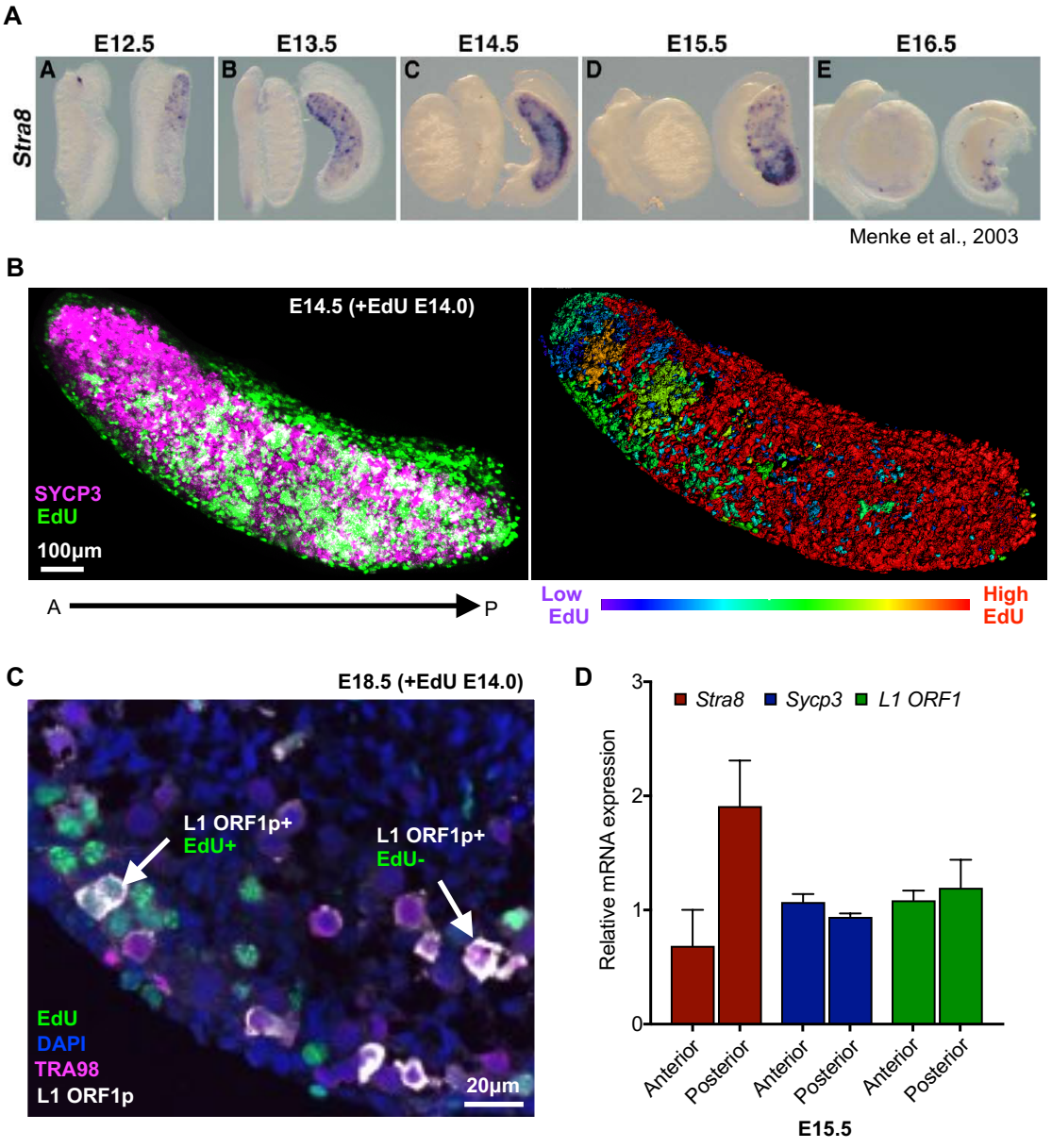


Fig. 4-2: Tracing oocytes based on timing of meiotic entry. **A)** Figure from Menke et al., 2003 showing that meiotic entry occurs in an anterior to posterior wave between E12.5 and E16.5 across the embryonic ovary that is reflected by *Stra8* expression in oocytes. **B)** 12 hour pulse-chase experiment showing E14.5 ovary that had been injected with EdU at E14.0. On the left, oocytes labeled with SYCP3 and EdU labels only posterior half of the ovary that was still replicating at E14.0 while the anterior portion had already exited mitosis and entered meiosis. Scalebar:100µm. On the right shows intensity of EdU labeling with

red reflecting highest EdU abundance and purple reflecting lowest EdU abundance. **C)** Section of E18.5 ovary that had been injected with EdU at E14.0 to label posterior oocytes. Oocytes are labeled with germ cell marker TRA98 and L1 ORF1p. DAPI stains all nuclei. Scalebar:20μm. **D)** Quantitative RT-PCR detection of *Stra8*, *Sycp3*, and L1 *ORF1* from RNA isolated from anterior and posterior segments of E15.5 wild-type CD1 ovaries.

4.3 Distribution of L1 via intercellular bridges within oocyte cysts

We wanted to test a role for oocyte cyst architecture in the establishment of differential L1 expression. First, we observed the relationship between intercellular bridges that connect oocytes within a cyst relative to L1 ORF1p expression within individual connected oocytes to determine variation of L1 ORF1p within and between cysts. Using TEX14, the protein that maintains intercellular bridges and prevents cytokinesis, as a marker to distinguish connected oocytes, we were able to map cysts within whole-mount ovaries of wild-type CD1 mice (Fig. 4-3A). In connected oocytes, we observed and quantified expression of L1 ORF1p in wild-type conditions at E15.5, prior to FOA (Fig. 4-3B, C). After quantifying L1 ORF1p in individual oocytes comprising three independent cysts, I found that the greatest difference in L1 ORF1p was between cysts, not within cysts. Slight differences between oocytes within cysts were also observed. These results suggest that oocytes within a cyst try to distribute L1 amongst each other, perhaps to prevent accumulation of lethal amounts within any individual oocytes. However, some cysts are endowed with greater amounts of L1 ORF1p to distribute (Fig. 4-3D)

To test the functional significance for intercellular bridges in L1 distribution and FOA, we analyzed L1 ORF1p expression in oocytes of TEX14 knockout mice. *Tex14*$^{-/-}$ females lack intercellular bridges based on electron microscopy and immunofluorescence labeling with other intercellular bridge proteins and are fertile[97]. We hypothesized that if intercellular bridges play a role in differential L1 expression, L1 ORF1p distribution and FOA dynamics may be altered in *Tex14*$^{-/-}$. First, we quantified oocytes in *Tex14*$^{-/-}$ compared to control combined *Tex14*$^{+/+}$ and *Tex14*$^{+/-}$ at our FOA timepoints of interest, E15.5, E18.5 and P2. Interestingly, at E15.5, there were no differences in oocyte number per ovary upon loss of TEX14. However, at E18.5 and P2, there was a decrease in oocytes in *Tex14*$^{-/-}$ compared to controls (Fig. 4-3E). Next, we quantified L1 ORF1p relative mean nuclear fluorescence in individual oocytes of *Tex14*$^{-/-}$ and *Tex14*$^{+/-}$ controls. We observed increased heterogeneity in L1 ORF1p expression in *Tex14*$^{-/-}$ with more oocytes having extraordinarily high levels of L1 ORF1p that were not observed in controls (Fig. 4-3F). One conclusion is that intercellular bridges allow better distribution of L1 between

oocytes to prevent accumulation within a subpopulation that will ultimately surpass the lethal threshold of L1 genotoxicity.

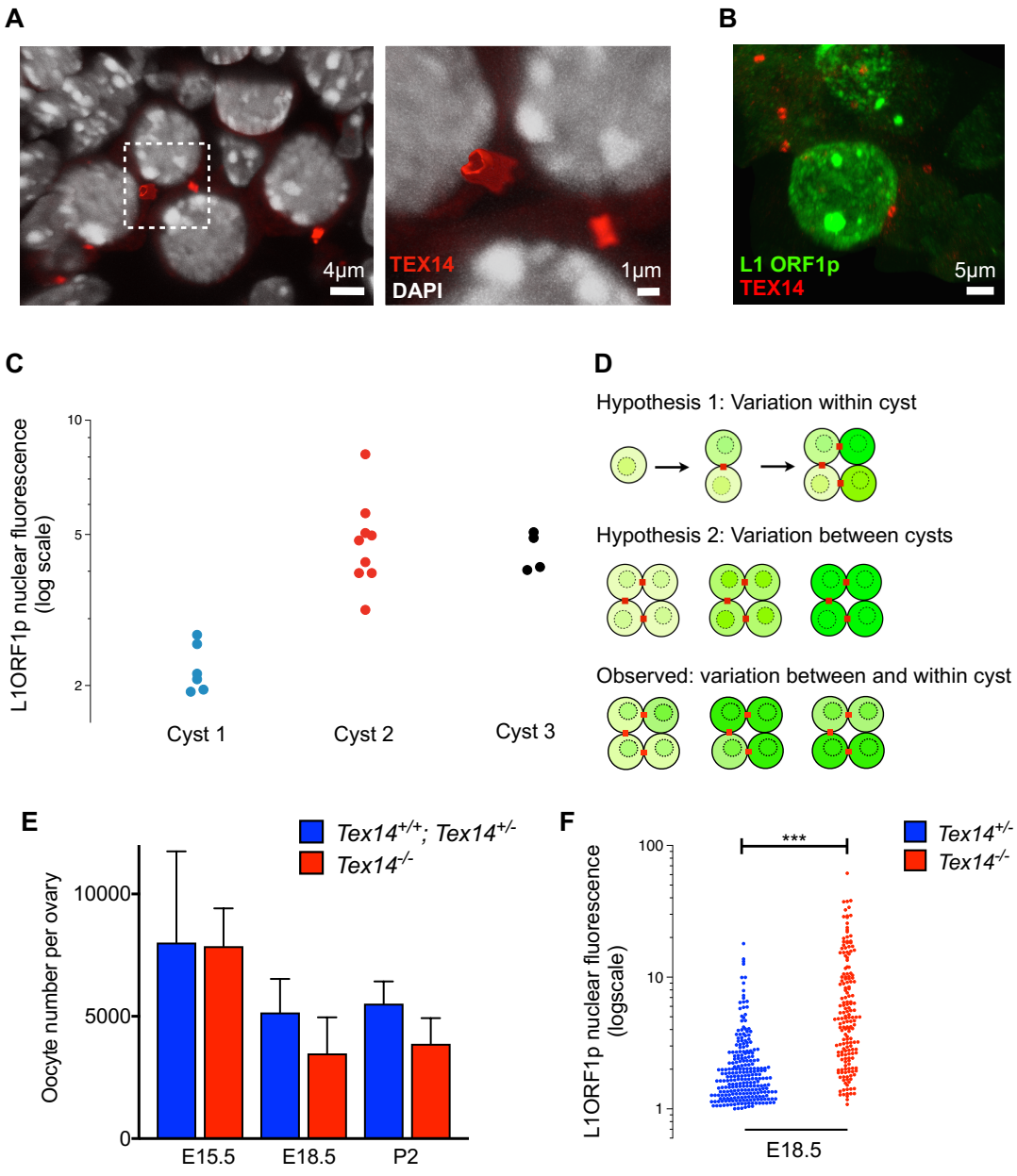


Fig. 4-3: L1 distribution within oocyte cysts. **(A)** Cluster of E15.5 WT CD1 oocytes connected by intercellular bridges labeled with the antibody TEX14. DAPI labels DNA. Inset surrounded by dotted white line zooms in on single TEX14-positive bridge with surface rendering to highlight tube structure. Scalebars:4µm for original and 1µm for inset. **B)** Cluster of E18.5 WT CD1 oocytes connected by intercellular bridges labeled with TEX14 and L1 ORF1p. Scalebar: 5µm. **C)** L1 ORF1p relative mean

nuclear fluorescence in connected oocytes within three individual cysts in E15.5 WT CD1 ovaries. **D)** Hypotheses and observed results for variation of L1 ORF1p distribution within cysts of WT mice. **E)** Oocyte number per ovary at E15.5, E18.5, and P2 in *Tex14*$^{-/-}$ compared to control (combined *Tex14*$^{+/+}$ and *Tex14*$^{+/-}$). **F)** L1 ORF1p relative mean nuclear fluorescence in *Tex14*$^{-/-}$ compared to control *Tex14*$^{+/-}$ oocytes at E18.5. Stats by Kolmogorov-Smirnov test, ***p<0.001. *Data obtained by Safia Malki for Fig. 4-3F.*

Conclusion

We explored three different fetal oogenesis programs with the potential to create differential L1 expression between the fetal oocyte population: epigenetic reprogramming, meiotic entry, and distribution within oocyte cysts. Overall, our results elucidate that regulation of L1 expression at the epigenetic level plays an important role in the origin of differential L1 expression among the fetal oocyte population based on DNA methylation. L1 expression level in oocytes appears independent of the timing of meiotic entry from the perspective of the meiotic initiation factor wave across the gonad. Last, transport of L1 within oocyte cysts may play a complementary role to determine the level of L1 in individual oocytes. Perhaps the regulation of differential L1 expression is a combination of mechanisms, originating at epigenetic differences between individual oocytes that prime the transcriptional output of L1 and amount of L1 protein made. Following its production, L1 may be distributed within individual cysts when they are still open around E15.5. Altogether, these mechanisms set up oocyte fate during FOA.

CHAPTER 5

Going forward

Discussion and future directions

The work presented in this thesis has deepened our understanding of the relationship between L1 and oocyte hosts, especially concerning programmed cell death in the fetal ovary. This work also supports a biological significance for FOA as quality control of the ovary and that at least in the first generation, FOA is not necessary for fertility. Going forward, the studies presented in this thesis, particularly the establishment of the *Chk2*$^{-/-}$ +AZT model that prevents FOA into postnatal stages by preserving oocytes that initially contained highest levels of L1 genotoxicity, has opened up a number of exciting areas of future study in the fields of transposons, oogenesis, and evolution.

5.1 Locus-specific expression of L1 elements

Until now, I have used a combination of broad experimental readouts for L1 expression and activity that include: L1 ORF1 mRNA expression by quantitative RT-PCR, L1 ORF1p expression by immunofluorescence, and the effect of reverse transcriptase inhibitor AZT. These approaches provide a big picture of L1 expression dynamics; however, it is a low resolution picture with some caveats. For example, L1 is expressed as a bicistronic transcript, and can possess an intact ORF1 region while its ORF2 region is degenerate. In this case, L1 ORF1p would be expressed, but the element does not possess the catalytic functions for retrotransposition provided by ORF2. Greater precision can be introduced into L1 expression analysis by identifying specific genomic loci in fetal oocytes where L1 is expressed, and characterizing these sequences. Approximately 3000 L1 elements, belonging to the youngest L1Md_T, L1Md_A and L1Md_Gf subfamilies, remain intact and active in the mouse genome[149]. It is important to distinguish whether expressed L1 elements in oocytes are intact, and therefore, relevant to FOA pathology.

In attempt to identify the specific genomic loci of L1 elements expressed in fetal oocytes, I used short-read mRNA sequencing data from sorted oocytes at E13.5, E15.5, E15.5 treated with AZT, E18.5 and E18.5 treated with AZT described in Chapter 2. Proper alignment and quantification of transposable

elements using short-read mRNA sequencing is a complex analysis due to the repeat nature of transposable elements and their high copy number in the genome. Various computational pipelines have been published with the goal of improving these types of analyses by addressing issues such as how to classify multi-mapping reads[149,150]. I began by comparing three strategies for alignment and quantification of transposable elements. First, I compared alignment strategies that discarded vs allowed multi-mapping reads. Normally, reads mapping multiple times in the genome are discarded, but in the case of transposable elements, much quantitative information is lost this way since they are highly repetitive sequences. To test the difference between unique and multi-mapping alignment methods, I used the STAR program with an annotation file to all mouse repetitive elements (mm10 RepeatMasker)[151,152]. I used the STAR option --outFilterMultimapNmax to keep reads that map to multiple locations of the genome, and compared these results to the default that keeps only uniquely mapped reads. In the unique mapping case, 16,752 different L1Md_T, L1Md_A and L1Md_Gf loci (from now on referred to as "young" L1) were identified as having >0 counts, while in the multi-mapping case, 25,646 loci were identified. The young L1 loci in the multi-mapped case encompassed all uniquely mapped young L1 loci identified. The locus assigned to ambiguous multi-mapped reads is based on the alignment score that is either the best of multiple or randomly chosen from multiple of the same score. To better resolve this uncertainty in multi-mapping read assignment, I used the computer program Telescope that reassigns multi-mapped reads using expectation maximization algorithms based on where the unique reads mapped[150]. After running Telescope, 7,035 young L1 loci were identified, 6,695 of which were in common with multi-mapped loci before Telescope and 340 new loci after using Telescope.

I decided to perform the remainder of the analysis using multimapping in STAR followed by Telescope to reassign the multi-mapped reads. Top hits for young L1 elements expressed in fetal oocytes were identified by setting a counts threshold of approximately 150, and normalizing to total number of reads against protein coding genes for each sample. Of the top hits, I manually checked for intact ORF1 and ORF2 by obtaining the sequence from mouse mm10 UCSC genome browser using coordinates from

the mouse RepeatMasker annotation file[152]. Then, I searched for intact ORFs in the sequence corresponding to the number of amino acids for ORF1 (371 amino acids) and ORF2 (1281 amino acids) using MacVector software and validated these ORFs using the NCBI Basic Local Alignment Search Tool for protein sequences and the L1Base2 (Fig. 5-1A)[153]. The remaining 64 young L1 sequences fell into four categories: containing intact ORF1 and ORF2, intact ORF1 only, intact ORF2 only and degenerate ORF1 and ORF2. Positions of expressed L1 sequences with intact or degenerate ORFs did not appear biased toward any particular chromosome (Fig. 5-1B). Slightly over 50% of the young L1 elements contained both ORFs intact, and were expressed in fetal oocytes at all stages except E13.5 (Fig. 5-1B, C). Interestingly though, the L1 sequences with highest expression were those that contained degenerate ORFs, suggesting that the cell figured out a way to silence these more dangerous elements (Fig. 5-1C). The expression pattern dynamics of the intact L1 mRNA sequences correlated with that observed by quantitative PCR of L1 ORF1 mRNA, where E13.5 oocytes showed minimal L1 expression, E15.5 oocytes possessed highest levels of L1 expression, E18.5 oocytes had reduced L1 expression, and E18.5 +AZT oocytes had high L1 expression similar to E15.5 (Fig. 3-1D and 5-1D). These results also confirm that AZT prevents death of oocytes containing high levels of L1 RNA, as suggested by L1 ORF1p experiments[69].

Looking at individual sequences of intact L1 elements, a few different expression pattern dynamics emerged, including the dominant pattern described above, but also one where E15.5 has the highest L1 expression, and AZT-treated E18.5 oocytes do not contain higher amounts of L1 (Fig 5-1E). Understanding the landscape of specific genomic loci for L1 elements expressed in fetal oocytes will allow us to better understand the contributing elements to FOA and provides a list of target sequences to mutate in attempt to directly implicate L1 in FOA. Also, understanding these specific L1 loci will be critical in analysis of the origin of differential L1 expression due to differential DNA demethylation described in Chapter 4 by correlating L1 loci to differentially methylated regions during epigenetic reprogramming.

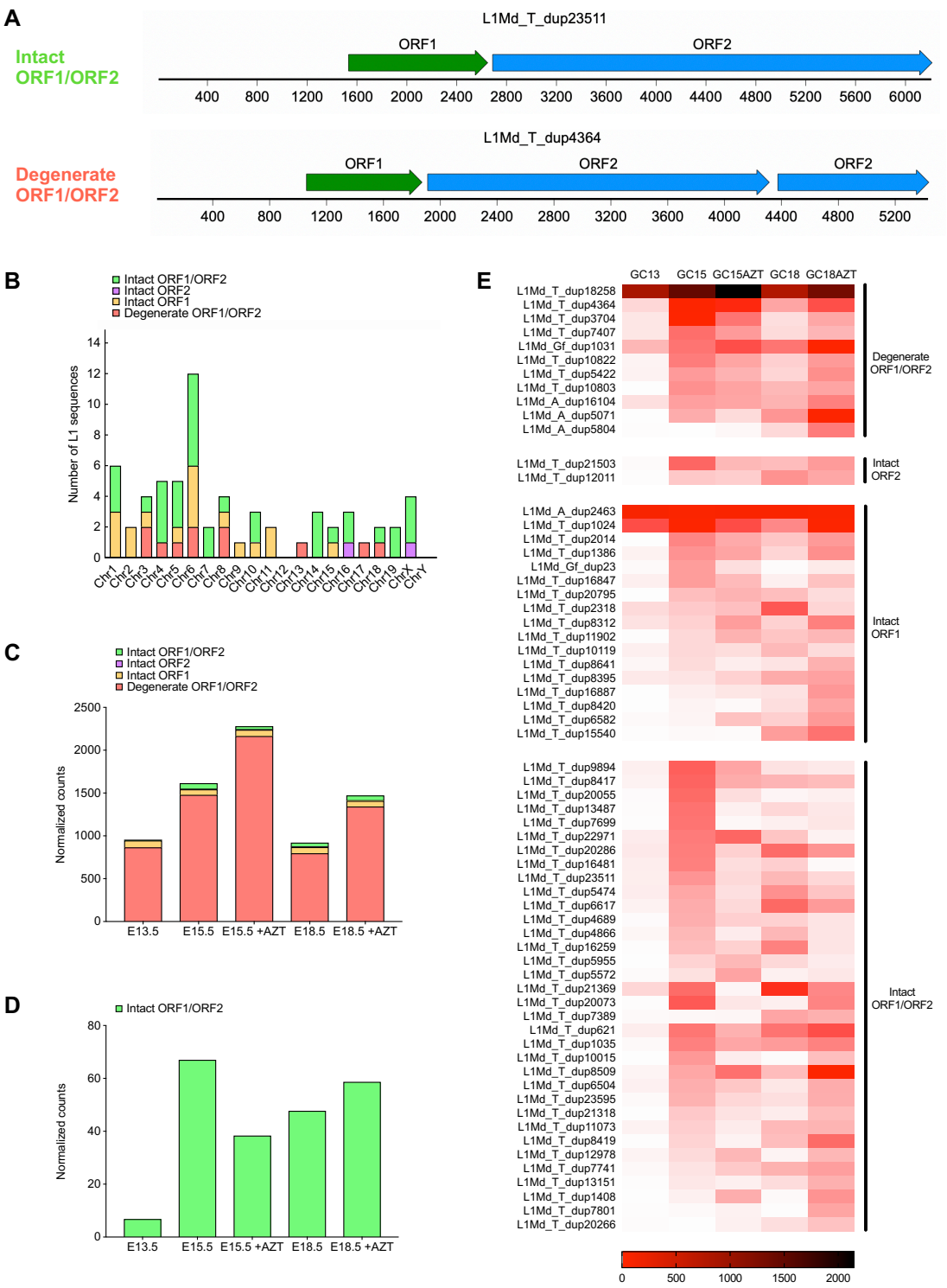

A
Intact ORF1/ORF2
L1Md_T_dup23511
ORF1
ORF2
400 800 1200 1600 2000 2400 2800 3200 3600 4000 4400 4800 5200 5600 6000
Degenerate ORF1/ORF2
L1Md_T_dup4364
ORF1
ORF2
ORF2
400 800 1200 1600 2000 2400 2800 3200 3600 4000 4400 4800 5200
B
Intact ORF1/ORF2
Intact ORF2
Intact ORF1
Degenerate ORF1/ORF2
Number of L1 sequences
0 2 4 6 8 10 12 14
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16 Chr17 Chr18 Chr19 ChrX ChrY
C
Intact ORF1/ORF2
Intact ORF2
Intact ORF1
Degenerate ORF1/ORF2
Normalized counts
0 500 1000 1500 2000 2500
E13.5 E15.5 E15.5 +AZT E18.5 E18.5 +AZT
D
Intact ORF1/ORF2
Normalized counts
0 20 40 60 80
E13.5 E15.5 E15.5 +AZT E18.5 E18.5 +AZT
E
GC13 GC15 GC15AZT GC18 GC18AZT
L1Md_T_dup18258
L1Md_T_dup4364
L1Md_T_dup3704
L1Md_T_dup7407
L1Md_Gf_dup1031
L1Md_T_dup10822
L1Md_T_dup5422
L1Md_T_dup10803
L1Md_A_dup16104
L1Md_A_dup5071
L1Md_A_dup5804
Degenerate ORF1/ORF2
L1Md_T_dup21503
L1Md_T_dup12011
Intact ORF2
L1Md_A_dup2463
L1Md_T_dup1024
L1Md_T_dup2014
L1Md_T_dup1386
L1Md_Gf_dup23
L1Md_T_dup16847
L1Md_T_dup20795
L1Md_T_dup2318
L1Md_T_dup8312
L1Md_T_dup11902
L1Md_T_dup10119
L1Md_T_dup8641
L1Md_T_dup8395
L1Md_T_dup16887
L1Md_T_dup8420
L1Md_T_dup6582
L1Md_T_dup15540
Intact ORF1
L1Md_T_dup9894
L1Md_T_dup8417
L1Md_T_dup20055
L1Md_T_dup13487
L1Md_T_dup7699
L1Md_T_dup22971
L1Md_T_dup20286
L1Md_T_dup16481
L1Md_T_dup23511
L1Md_T_dup5474
L1Md_T_dup6617
L1Md_T_dup4689
L1Md_T_dup4866
L1Md_T_dup16259
L1Md_T_dup5955
L1Md_T_dup5572
L1Md_T_dup21369
L1Md_T_dup20073
L1Md_T_dup7389
L1Md_T_dup621
L1Md_T_dup1035
L1Md_T_dup10015
L1Md_T_dup8509
L1Md_T_dup6504
L1Md_T_dup23595
L1Md_T_dup21318
L1Md_T_dup11073
L1Md_T_dup8419
L1Md_T_dup12978
L1Md_T_dup7741
L1Md_T_dup13151
L1Md_T_dup1408
L1Md_T_dup7801
L1Md_T_dup20266
Intact ORF1/ORF2
0 500 1000 1500 2000

Figure 5-1: Characterization of locus-specific L1 sequences expressed during FOA. **A)** Example L1 sequences with "intact" vs "degenerate" ORF1 and ORF2. **B)** Total number of young L1 sequences expressed in E13.5, E15.5, E18.5 and E18.5 +AZT oocytes. Sequences are further categorized by whether they contain intact ORF1 and ORF2 (green), intact ORF2 only (purple), intact ORF1 only (orange) or degenerate ORF1 and ORF2 (red). **C)** Normalized counts for young L1 sequences expressed in E13.5, E15.5, E15.5 +AZT, E18.5 and E18.5 +AZT oocytes and categorized by intact or degenerate ORFs. **D)** Normalized counts for intact young L1 sequences expressed in E13.5, E15.5, E15.5 +AZT, E18.5 and E18.5 +AZT oocytes. **E)** Heatmap displaying expression of 64 individual, locus-specific young L1 sequences in E13.5, E15.5, E15.5 +AZT, E18.5 and E18.5 +AZT oocytes and categorized by intact or degenerate ORFs.

L1 ID	ORFs Intact	Chr	Start (exon)	Stop (exon)	L1Base2 cat./ID
L1Md_T_dup18258	None	13	9832021	9838665	FLnI-L1/#7705
L1Md_T_dup4364	None	3	102866185	102871622	FLnI-L1/#6157
L1Md_T_dup3704	None	3	41763853	41770371	FLnI-L1/#6198
L1Md_T_dup7407	None	5	107578919	107584270	FLnI-L1/#5281
L1Md_Gf_dup1031	None	17	30500005	30504719	FLnI-L1/#1207
L1Md_T_dup10822	None	8	67035359	67041544	FLnI-L1/#8148
L1Md_T_dup5422	None	4	49460400	49467367	FLnI-L1/#13131
L1Md_T_dup10803	None	8	65704113	65712009	FLnI-L1/#8210
L1Md_A_dup16104	None	18	5725794	5730924	FLnI-L1/#5004
L1Md_A_dup5071	None	6	95809857	95814880	FLnI-L1/#2922
L1Md_A_dup5804	None	6	107249034	107255170	FLnI-L1/#2901
L1Md_T_dup21503	ORF2	16	57586095	57593165	ORF2-L1/#9801
L1Md_T_dup12011	ORF2	X	73268096	73275199	ORF2-L1/#239
L1Md_A_dup2463	ORF1	3	33618969	33624989	FLI-L1/#6685
L1Md_T_dup1024	ORF1	1	107035321	107041388	FLI-L1/#9028
L1Md_T_dup2014	ORF1	2	17124176	17131427	FLI-L1/#186
L1Md_T_dup1386	ORF1	1	138305467	138312353	FLI-L1/#9012
L1Md_Gf_dup23	ORF1	1	51433415	51438958	FLI-L1/#9684
L1Md_T_dup16847	ORF1	11	37752442	37758534	FLI-L1/#13866
L1Md_T_dup20795	ORF1	15	52837587	52845318	FLI-L1/#1933
L1Md_T_dup2318	ORF1	2	56792768	56801310	FLI-L1/#286
L1Md_T_dup8312	ORF1	6	78032420	78033415	FLI-L1/#2958
L1Md_T_dup11902	ORF1	9	100073828	100079677	FLI-L1/#1365
L1Md_T_dup10119	ORF1	8	3907166	3914646	FLI-L1/#8149
L1Md_T_dup8641	ORF1	6	116877123	116884013	FLI-L1/#2561
L1Md_T_dup8395	ORF1	6	89527391	89534263	FLI-L1/#2576
L1Md_T_dup16887	ORF1	11	41105275	41111909	FLI-L1/#13900
L1Md_T_dup8420	ORF1	6	95834702	95841316	FLI-L1/#2400
L1Md_T_dup6582	ORF1	5	26379422	26385697	FLI-L1/#5287
L1Md_T_dup15540	ORF1	10	6281128	6286747	FLI-L1/#10305
L1Md_T_dup9894	ORF1 and ORF2	7	104027601	104034674	FLI-L1/#2450
L1Md_T_dup8417	ORF1 and ORF2	6	95658065	95663747	FLI-L1/#2245
L1Md_T_dup20055	ORF1 and ORF2	14	104704915	104712436	FLI-L1/#1958
L1Md_T_dup13487	ORF1 and ORF2	X	100859053	100865917	FLI-L1/#1179
L1Md_T_dup7699	ORF1 and ORF2	6	19145469	19151801	FLI-L1/#2195
L1Md_T_dup22971	ORF1 and ORF2	18	45016906	45024783	FLI-L1/#590
L1Md_T_dup20286	ORF1 and ORF2	15	103410636	103417544	FLI-L1/#904
L1Md_T_dup16481	ORF1 and ORF2	10	112167315	112174741	FLI-L1/#2647
L1Md_T_dup23511	ORF1 and ORF2	19	35904002	35910216	FLI-L1/#1000
L1Md_T_dup5474	ORF1 and ORF2	4	53668404	53674903	FLI-L1/#1446
L1Md_T_dup6617	ORF1 and ORF2	5	29349521	29356677	FLI-L1/#373
L1Md_T_dup4689	ORF1 and ORF2	3	134880938	134887427	FLI-L1/#703
L1Md_T_dup4866	ORF1 and ORF2	4	18606847	18612728	FLI-L1/#1428
L1Md_T_dup16259	ORF1 and ORF2	10	86234064	86240153	FLI-L1/#2758

L1Md_T_dup5955	ORF1 and ORF2	4	91654618	91660721	FLI-L1/#1471
L1Md_T_dup5572	ORF1 and ORF2	4	66355025	66361105	FLI-L1/#1450
L1Md_T_dup21369	ORF1 and ORF2	16	46642577	46649283	FLI-L1/#409
L1Md_T_dup20073	ORF1 and ORF2	14	106963963	106970297	FLI-L1/#1925
L1Md_T_dup7389	ORF1 and ORF2	5	104887885	104894355	FLI-L1/#311
L1Md_T_dup621	ORF1 and ORF2	1	61491787	61499623	FLI-L1/#193
L1Md_T_dup1035	ORF1 and ORF2	1	107822787	107828917	FLI-L1/#72
L1Md_T_dup10015	ORF1 and ORF2	7	120506016	120512065	FLI-L1/#2438
L1Md_T_dup8509	ORF1 and ORF2	6	104257151	104264421	FLI-L1/#2109
L1Md_T_dup6504	ORF1 and ORF2	5	17238271	17247254	FLI-L1/#381
L1Md_T_dup23595	ORF1 and ORF2	19	51949137	51955753	FLI-L1/#1007
L1Md_T_dup21318	ORF1 and ORF2	16	41054276	41060891	FLI-L1/#499
L1Md_T_dup11073	ORF1 and ORF2	8	99466277	99472357	FLI-L1/#1797
L1Md_T_dup8419	ORF1 and ORF2	6	95794367	95800789	FLI-L1/#2115
L1Md_T_dup12978	ORF1 and ORF2	X	69930908	69938522	FLI-L1/#1212
L1Md_T_dup7741	ORF1 and ORF2	6	21393957	21399849	FLI-L1/#2196
L1Md_T_dup13151	ORF1 and ORF2	X	79737866	79744900	FLI-L1/#1202
L1Md_T_dup1408	ORF1 and ORF2	1	139716332	139722200	FLI-L1/#89
L1Md_T_dup7801	ORF1 and ORF2	6	25550943	25557814	FLI-L1/#2160
L1Md_T_dup20266	ORF1 and ORF2	14	124654232	124660284	FLI-L1/#1940

E13.5 Counts	E15.5 Counts	E15.5 +AZT Counts	E18.5 Counts	E18.5 +AZT Counts
861.345361	1447.355983	2141.931184	781.0313501	1311.09589
0.951391683	7.750176064	4.803902962	2.120466397	3.95140996
0.593530476	4.976279407	3.195174498	0.767035397	2.070797943
0.591381011	3.307334567	2.355590366	0.849540169	1.750624901
1.736340603	3.186163959	3.958506903	3.18036837	4.9849033
0.252981734	3.019862591	2.140938452	1.148058462	2.386506186
0.26381078	2.357966249	1.808372456	0.99963531	2.625895948
0	2.578093041	2.181065584	1.708898678	2.15610165
0.761164713	2.246593349	2.117813926	1.797268759	2.947935764
0	1.964905019	0.848783095	2.484294018	5.685595178
0.031354034	0.016781733	0.16175241	0.939448744	3.037708624
0.07892245	3.485288774	1.696154896	1.585234967	2.346026224
0.029192894	1.297225031	1.193914413	2.541163306	2.059432291
72.37310034	41.34413	51.2724921	49.64537885	32.86272895
3.989508743	5.105719967	4.00814139	2.761396848	5.318649603
0.009730965	2.810309174	1.996413509	1.344435305	2.383889688
0.2854105	2.53798205	2.156075357	0.988994326	2.614066175
0.095136833	2.307500134	0.775155122	0.042061681	0.487266326
0.344871021	2.301061736	1.641707339	0.382668355	1.475792157
0	1.545361683	1.792345648	1.546410467	0.818668908
0.87569343	1.266171448	1.504609461	3.755567333	0.998151138
0.704905041	1.069459461	2.279299329	1.07470958	2.962410957
0.029192894	0.942723457	1.833784261	1.385781137	1.725475144
0.371902776	0.933034875	0.809053556	1.671163813	0.842085601
0.183790248	0.865144064	0.433272714	0.673013598	1.849572377
0.501641189	0.645017272	1.078307023	1.996033439	2.216898686
0.011892104	0.386525496	0.607014358	0.69289057	2.432533271
0	0.245735893	0.085493072	0.767938139	2.135625636
0.090814554	0.228299408	1.478856774	1.153790277	2.404435684
0.029192894	0.10973602	0.0797845	2.306042058	3.231721434
0.378409543	3.610508812	2.113360197	0.696950395	0.501139016
0.27569121	3.508824493	1.978887747	1.763380128	1.603814387
0.059460521	3.391231445	0.798068264	0.217738908	0.350687208
0.045407277	3.240098508	0.374198871	0.967418668	0.563457772
0.038923859	3.188638475	0.174283302	0.358758257	0.608656205
0.458430074	3.030011253	3.480486133	1.397324864	0.277166039
0.518930305	2.994047027	1.132649108	3.361724014	2.503044279
0.417321727	2.956967969	0.80262883	1.166129949	0.129826278
0.409740228	2.49474179	0.921487586	1.705661497	0.957646925
0.438933122	2.071246362	0.736424954	2.577386376	1.409151866
0.281088221	1.970446836	0.917886196	3.428625233	2.366731684
0.059460521	1.93352406	1.146680318	1.08710265	0.64073922
0.209735596	1.699561921	0.455709919	1.723893177	0.651165759
0.038923859	1.557451024	1.122220725	2.920727975	0.6439474

0	1.040261011	1.719301302	0.935362219	0.455859675
0.074600171	0.727179933	2.174706301	0.332380226	0.558294257
0.975164217	3.357496884	0.21944973	4.618163023	2.773131423
0.035676313	3.746750515	0.583472311	0.376140594	2.829637032
0.130813146	0.167817334	0.191493808	2.110754853	1.881027055
0.416223646	3.152491366	1.847984546	3.210117078	3.961984512
0.095136833	2.925295002	2.082092615	2.348183839	2.891655302
0	2.413198514	0.437859484	0.080010138	1.546864817
0.235704293	2.072033818	3.275874428	1.584572516	5.929145429
0.321086813	1.92238148	1.340914784	0.56914676	1.959503269
0.09622324	1.818756483	0.963370848	0.759631595	1.937345571
0	1.285123901	0.69715313	0.445429651	1.596990167
0.278938756	1.168743021	0.464286766	1.665245105	1.777850481
0.023784208	1.044516901	0.266164727	1.553921066	3.36917066
0.249734188	0.900005246	1.732803922	0.248256863	1.73515861
0.154597354	0.670626374	1.278458651	1.836622215	2.269529993
0.021623069	0.557398341	0.357380359	0.817430323	1.888586274
0.011892104	0.3229997	1.949006343	0.080832783	2.525362022
0.029192894	0.092408659	0.312261837	0.0313406	2.138510636
0.048654824	0	0.326155261	0.78030451	1.468270595

Table 5-1: Top hits for locus-specific L1 sequences. Listed are L1 sequence ID, ORFs intact, chromosome number, start and stop coordinates from Repeat Masker annotation, category and L1 sequence ID from L1Base2, and normalized counts for E13.5, E15.5, E15.5 +AZT, E18.5, and E18.5 +AZT sorted oocyte samples.

5.2 Interplay between MPI events and L1 expression

Germ cells are exceptionally vulnerable to L1 expression during MPI. It is during MPI that L1 expression is most commonly observed and where phenotypes related to genome instability and cell death as a consequence of L1 expression manifest. For example, piRNA mutant spermatocytes that cannot silence L1 expression fail during MPI[59]. Additionally, wild-type oocytes that express L1 at high levels fail during MPI[69]. The complex interplay between L1 regulation and MPI events remains to be explored in mechanistic detail to understand what creates this incompatibility in germ cells.

Programmed DSBs are generated during MPI by the Spo11 enzyme for genetic recombination[66]. One outstanding question is whether L1 can take advantage of Spo11-generated DSBs and opportunistically insert at these sites, or instead, preferentially create their own breaks using the endonuclease activity of ORF2p. By utilizing meiotic DSBs, L1 may interfere with their processing and repair, contributing to the genetic instability observed. In contrast, L1 may assist in repair of meiotic DSBs by serving as molecular glue through their retrotransposition. This has been shown in cell culture in mutants for non-homologous end-joining, and suggests that L1 does not require micro-homology for insertion[154]. To test the contribution of meiotic DSBs vs. DNA breaks generated by L1 endonuclease activity to genome instability during MPI, I used the *Spo11* mutant mouse model that does not create meiotic DSBs[70]. Importantly, non-meiotic DNA breaks are observed in *Spo11*$^{-/-}$ germ cells based on immunofluorescence labeling of RAD51, a marker of DSB repair[69]. Further, in *Spo11*$^{-/-}$*;Mael*$^{-/-}$ spermatocytes that cannot silence L1, an increase in non-meiotic breaks is observed and can be attributed to L1 rather than other sources of DNA damage[59]. To analyze a role for meiotic vs. non-meiotic DNA breaks in oocyte death reflected as FOA, I quantified oocyte number in untreated and AZT-treated *Spo11*$^{-/-}$, *Spo11*$^{+/-}$ and *Spo11*$^{+/+}$ in E18.5 ovaries. I hypothesized that if meiotic DSBs predominantly contributed to FOA, we observe less FOA in *Spo11*$^{-/-}$. However, at E18.5, we observed normal FOA in *Spo11*$^{-/-}$, *Spo11*$^{+/-}$ and *Spo11*$^{+/+}$ ovaries, agreeing with our *Chk2*$^{-/-}$ oocyte data that FOA prior to DNA damage checkpoint activation is independent of meiotic DSBs (Fig. 5-2A). Interestingly, the ability to rescue

oocytes with AZT was less effective in *Spo11*$^{-/-}$ than controls, a result that requires further exploration (Fig. 5-2A). An important future direction will be to analyze the number of oocytes in *Spo11*$^{-/-}$ vs control P2 ovaries, as the role of meiotic vs non-meiotic breaks may not manifest until after activation of the DNA damage checkpoint. This idea is supported by prior analysis of *Spo11*$^{-/-}$ vs wild-type postnatal ovaries between P0 and P60, where oocyte number per ovary becomes progressively decreased in *Spo11*$^{-/-}$ ovaries over time compared to wildtype[155]. However, a number of other phenotypes have been reported in *Spo11*$^{-/-}$ meiotic cells such as delayed meiotic progression and asynapsis due to lack of meiotic breaks that must be considered when interpreting results[70]. I also observed a significant decrease in L1 ORF1p expression in oocytes at E16.5 in AZT-treated *Spo11*$^{-/-}$ compared to AZT-treated *Spo11*$^{+/-}$ ovaries (AZT treatment is insignificant at E16.5 with respect to oocyte number as oocyte loss is still minimal), reflective of a potential delay in MPI progression (Fig. 5-2B). Therefore, the question of whether L1 has a preference to retrotranspose into existing programmed meiotic breaks versus creating *de novo* breaks using endonuclease activity of L1 ORF2p may be better addressed by chromatin immunoprecipitation experiments to determine if L1 proteins and meiotic DSBs co-localize, and at which genomic loci in wild-type and *Spo11*$^{-/-}$ oocytes.

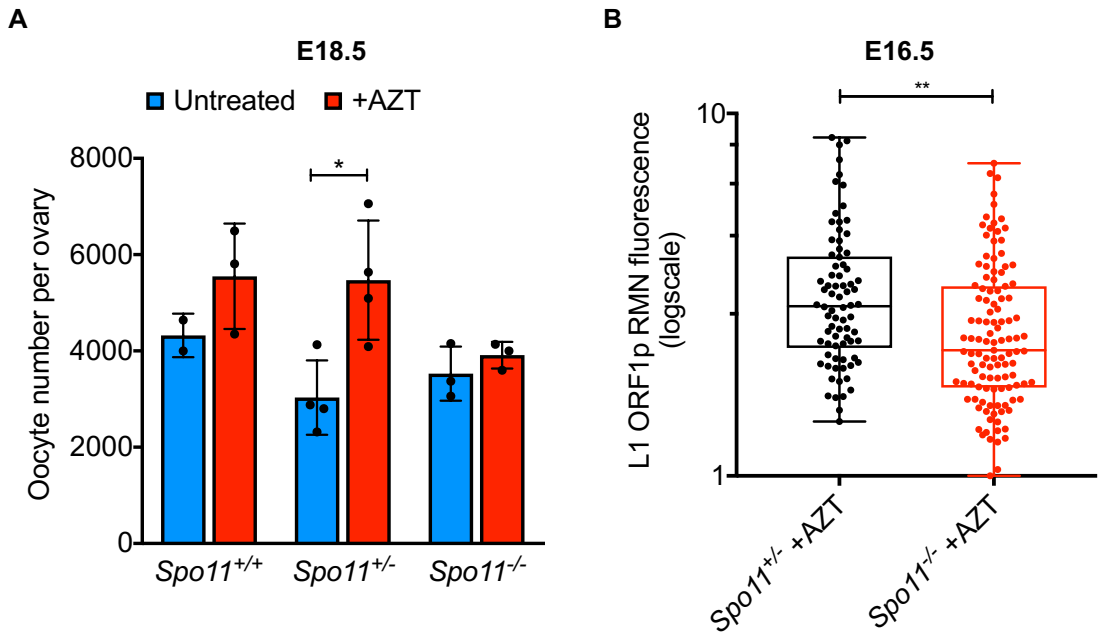


Figure 5-2: Analysis of Spo11 mutant oocytes. **A)** Oocyte number in untreated and AZT-treated *Spo11*$^{+/+}$, *Spo11*$^{+/-}$ and *Spo11*$^{-/-}$ ovaries at E18.5. B) Relative mean nuclear fluorescence intensity of L1 ORF1p in AZT-treated *Spo11*$^{+/-}$ and *Spo11*$^{-/-}$ oocyte at E16.5.

5.3 Human health and adaptation relevance for FOA prevention

Due to the conservation of FOA and L1 expression in human oocytes, an interesting question to entertain is whether the conservation of FOA and L1 expression in human oocytes may lead our work to some biomedical relevance for female fertility. By blocking FOA with the goal of increasing ovarian reserve, inhibition of L1 and CHK2 activity may be a novel avenue for therapeutics for improvement of pre mature ovarian failure, premature menopause in women undergoing cancer treatment, or even normal age-related decline in female fertility. However, this type of treatment would be purely preventative, as the drug administration impacts the fetal ovary rather than adult. Important considerations are the differences in the types and abundance of L1 elements in humans vs mice. Also, the contribution of L1 to FOA in humans is not explored, and there are a number of differences between L1 abundance and regulation between mice and humans.

We still do not understand how FOA prevention impacts the future generations of offspring. In light of evolution, fetal oogenesis is a critical window of opportunity for L1 activity in combination with meiotic recombination to promote genetic diversity and adaptation to environmental stress, an influence that may outweigh the short-term consequences of FOA to the ovarian reserve size of an individual. Perhaps mammals have evolved to produce oocytes in excess to account for this window. Future studies analyzing retrotransposition frequency and mutations in a larger pedigree from *Chk2*$^{-/-}$ +AZT females that did not experience FOA will answer these big questions of whether lack of quality control in fetal oogenesis manifests in the disease susceptibility or adaptive potential of future generations. We consider experiments to lineage label oocytes based on L1 expression and trace individual fetal oocytes into postnatal timepoints.

CHAPTER 6

Materials and methods

6.1 Mice

For this study, *Chk2*$^{-/-}$ mice in a mixed C57Bl/6 and 129X1/Sv genetic background were used [119]. *Chk2*$^{-/-}$ mice were backcrossed one time to C57Bl/6 to generate *Chk2*$^{+/-}$ controls. We chose *Chk2*$^{+/-}$ as a control to account for the mixed genetic background resulting in a significant increase in oocyte number compared to wild-type mice of pure C57Bl/6 background (Fig. 6-1A-C). To determine how genetic background contributed to differences in oocyte number between *Chk2*$^{-/-}$ and C57Bl/6 wild-type mice, we backcrossed *Chk2*$^{-/-}$ for 4 generations. This reduced the percent genome containing homozygous 129X1/Sv SNPs from approximately 4% to 0.03% (SNP genotyping by DartMouse, Dartmouth School of Medicine). *Mili*$^{+/+}$, *Mili*$^{+/-}$, and *Mili*$^{-/-}$ mice used were in C57Bl/6 genetic background. *Mili*$^{-/-}$;*Chk2*$^{-/-}$ mice and control littermates were generated by crossing *Chk2*$^{-/-}$ and *Mili*$^{+/-}$ animals. Wild-type mice of CD1 (Charles River Laboratories) genetic background were used for all quantitative PCR, mRNA (bulk ovary and oocyte and single-cell) and small RNA sequencing experiments unless otherwise noted. *Tex14* +/+, *Tex14*+/- and *Tex14-/-* mice were used in C57Bl/6 as well as mixed C57Bl/6 and 129 genetic backgrounds[95]. All experimental procedures were performed in compliance with ethical regulations and approved by the IACUC of the Carnegie Institution for Science.

6.2 AZT treatment

50mg/kg/day AZT was administered daily by gavage to ~8-month-old pregnant female from E13.5 until experiment end point [69]. AZT (Sigma Aldrich, Cat# A2169) was diluted to 5mg/ml in nuclease free water, aliquoted and stored at -20°C. In other words, 10µl of prepared stock solution (5mg/ml) was administered per gram of mouse body weight per day.

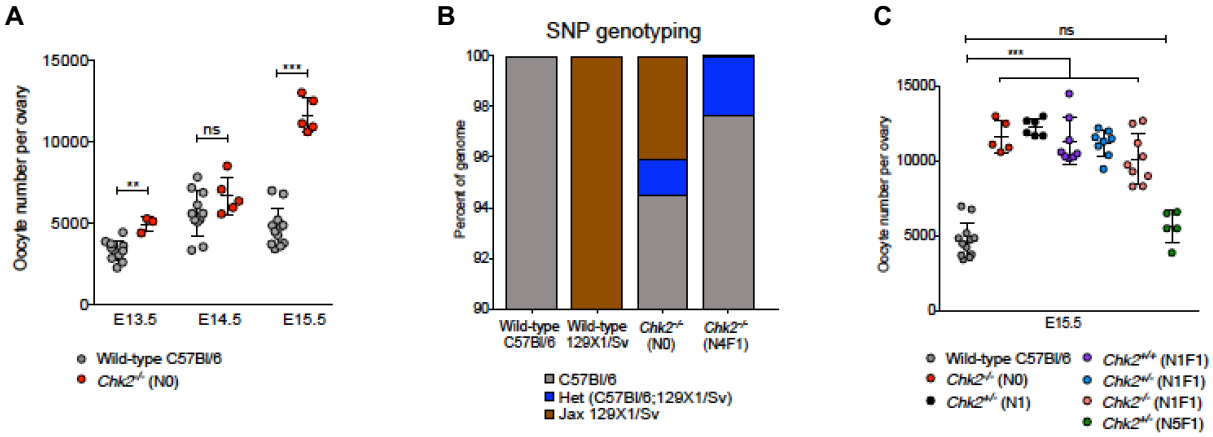


Figure 6-1: Maximum oocyte number per ovary depends on genetic background. Genetic background influences maximum oocyte number. **A)** We observed a significant difference in maximum oocyte number at E15.5 and earlier stages at E13.5 and E14.5 between ovaries of *Chk2*$^{-/-}$ (N0) mice used for all other experiments and wild-type C57Bl/6 mice. **B)** Although appearing C57Bl/6, *Chk2*$^{-/-}$ (N0) mice were of mixed C57Bl/6 and 129X1/Sv genetic backgrounds. We performed SNP genotyping (DartMouse, Dartmouth School of Medicine) to assess the amount of 129X1/Sv specific polymorphisms remaining in the *Chk2*$^{-/-}$ (N0) genome to determine whether genetic background contributed to the difference in maximum oocyte number observed, or whether this difference was related to CHK2 function. Shown is percent genome of *Chk2*$^{-/-}$ (N0) mice belonging to C57Bl/6 or Jax 129X1/Sv origin, alongside wild-type C57Bl/6 and wild-type 129X1/Sv genomes that contain 100% of SNPs from respective origins. Approximately 4% of the *Chk2*$^{-/-}$ (N0) genome was homozygous 129X1/Sv. We repeated SNP genotyping after backcrossing *Chk2*$^{-/-}$ (N0) to wild-type C57Bl/6 four times (N4F1), finding that homozygous 129X1/Sv SNPs are eliminated. Scale is from 90 to 100% on the Y-axis. **C)** Maximum oocyte number per E15.5 ovary from respective genotypes and genetic backgrounds: wild-type C57Bl/6, *Chk2*$^{-/-}$ (N0), *Chk2*$^{+/-}$ (N1), *Chk2*$^{+/+}$ (N1F1), *Chk2*$^{+/-}$ (N1F1), and *Chk2*$^{-/-}$ (N1F1), and *Chk2*$^{+/-}$ (N5F1). *Chk2*$^{-/-}$ (N0) and *Chk2*$^{+/-}$ (N1) data are repeated from Fig. 2-6A and Table 2-8 – 2-9. Dots indicate independent ovary samples; data are mean $\pm$SD; n$\geq$5 ovaries. Stats by two-tailed unpaired Student's t-test, ns

$p>0.05$;***$p<0.001$. Only in *Chk2*$^{+/-}$ (N5F1), after five backcrosses to C57Bl/6, is the maximum oocyte number comparable to that of wild-type C57Bl/6, indicating that maximum oocyte number is dependent on genetic background (Table 2-8 – 2-9). Similar oocyte numbers observed between *Chk2*$^{+/+}$ (N1F1), *Chk2*$^{+/-}$ (N1F1), and *Chk2*$^{-/-}$ (N1F1), indicating that the difference in maximum oocyte number is not due to CHK2 function. For all other experiments, we compared *Chk2*$^{-/-}$ (N0) to *Chk2*$^{+/-}$ (N1) as a control rather than wild-type C57Bl/6 to avoid differences in oocyte number related to genetic background while still being able to obtain 100% progeny of desired genotype using the following crosses: WT B6 x *Chk2*$^{-/-}$ (N0) for the *Chk2*$^{+/-}$ control group and *Chk2*$^{-/-}$ (N0) x *Chk2*$^{-/-}$ (N0) for the *Chk2*$^{-/-}$ experimental group.

6.3 Immunostaining

For whole-mount ovaries, ovarian cryo-sections, and meiotic spreads, immunofluorescence staining was performed according to previously described protocols[69,156].

Paraffin sections. Colorimetric immunostaining using DAB substrate was performed. Ovaries were fixed in Bouin's overnight at 4°C, transferred to 70% ethanol overnight, and embedded in paraffin. Ovaries were sectioned into 10µm slices. Sections were deparaffinized by washing slides in Citrisolv (Decon Labs) 3x 15 minutes, re-hydrated through graded ethanol washes, blocked with hydrogen peroxide for 10 minutes, avidin block for 15 minutes (Vector Laboratories), biotin block (Vector Laboratories), and goat serum block (Vector Labs cat # PK-4001). Samples were incubated overnight at 4°C with anti-DDX4/MVH and for 30 minutes at room temperature with biotinylated goat anti-rabbit IgG secondary antibody (Vector Labs cat # PK-4001). Samples were incubated 30 minutes at room temperature with Vectastain ABC reagent (Vector Labs cat # PK-4001) followed by DAB detection. Slides were dipped in hematoxylin, rehydrated in ethanol, dipped in Citrisolv and mounted.

Cell suspension. Immunofluorescence staining was performed in a gel matrix. Ovarian cells were dissociated using dissociation buffer containing 0.025% trypsin, 2.5 mg/mL collagenase, and 0.1mg/mL DNase I. Ovaries incubated at 37˚C for 30 minutes, pipetting cells every 10 minutes to dissociate. 10% fetal bovine serum was added to inactivate trypsin. Cells were pelleted for 5 min at 500rpm and resuspended in 10% FBS in GBSS. Suspension filtered in 40µm filter and pelleted again. Resuspend in 2% PFA and fix 10 minutes on ice. Pellet cells and resuspend in Collagen solution (mix on ice: 20µl 10xPBS, 2.7µl NaOH, 56µl water, 121.22µl collagen for total of 200µl). Incubate at 37 degrees for 30 minutes until collagen solidifies. Perform immunofluorescence steps on collagen embedded cells. Wash 2x in 1x PBS, permeabilize in 0.1% Tx100-PBS for 5 minutes, and incubate in primary antibody

overnight at 4 degrees or 1 hour at room temperature. Wash once with 0.1% Tx100-PBS and 2x with 0.05% Tx100-PBS. Block in 10% serum. Incubate in secondary antibody 2 hours at room temperature in 3% serum. Repeat wash sequence. To release cells from collagen, first replace wash with warm GBSS and crush pellet with blue pestle. Remove GBSS and replace with 25mg collagenase in 1ml warm GBSS. Incubate at 37 degrees for 30 minutes with agitation and pipetting until dissolved. Filter cells once gel matrix is gone in 40μm filter. Pellet cells and resuspend in PBS for FACS or other downstream analysis.

Antibodies. Primary antibodies: Anti-Germ cell-specific antigen antibody [TRA98], rat monoclonal, abcam cat. #ab82527, diluted to 1:500 for immunofluorescence; Anti-L1 ORF1p (full length protein), rabbit polyclonal, diluted to 1:500 for immunofluorescence [69]; Anti-GM130, mouse monoclonal, BD Biosciences cat. # 610822, diluted to 1:200 for immunofluorescence; Anti-phospho-Histone H2A.X (Ser139) clone JBW301, mouse monoclonal, Millipore Sigma cat. # 05-636, diluted to 1:1000 for immunofluorescence; Anti-SYCP3, rabbit polyclonal, abcam cat. # ab15093, diluted to 1:500 for immunofluorescence. Anti-PIWIL2, rabbit polyclonal, abcam cat. # ab181340, diluted to 1:50 for immunofluorescence; Anti-F4/80 antibody [CI:A3-1], rat monoclonal, abcam cat. # ab6640, diluted to 1:100 for immunofluorescence; Anti-DDX4/MVH, rabbit polyclonal, abcam cat. # ab13840, diluted to 1:200 for immunostaining on paraffin sections and 1:1000 for western blot; Anti-p63(4A4), mouse monoclonal, Santa Cruz Biotechnology cat. # sc-8431, diluted to 1:500 for western blot. Secondary antibodies: Alexa donkey ant- rabbit 488 (Invitrogen, cat # A-21206**)** diluted 1:1000 for immunofluorescence; Alexa donkey anti-rabbit 568 (Invitrogen, cat # A10042) diluted 1:1000 for immunofluorescence; Alexa donkey anti-mouse 488 (Invitrogen, cat # A-21202) diluted 1:1000 for immunofluorescence; Alexa donkey anti-mouse 594 (Invitrogen, cat # A-21203) diluted 1:1000 for immunofluorescence; Alexa donkey anti-rat 647 (Invitrogen, cat # 150155**)** diluted 1:1000 for immunofluorescence; Goat anti-mouse IgG (H+L)-HRP Conjugate (BioRad cat # 1721011) diluted 1:2000 for Western blot; Goat anti-rabbit IgG (H+L)-HRP Conjugate (BioRad cat # 1721019) diluted

1:2000 for Western blot.

6.4 Quantification of oocyte number

Whole-mount ovaries. We quantified oocytes per ovary at E15.5, E18.5, and P2 using a whole-mount immunofluorescence and tissue clearing method previously described [156]. Ovaries were dissected and labeled with the germ cell-specific antibody TRA98 [157]. After immunofluorescence, samples were treated with ScaleA2 clearing reagent for 7 days, changing solution each day. Confocal imaging through the entire tissue using SP5 confocal microscope (Leica) followed by 3D reconstruction of Z-stacked images using Imaris software (Bitplane) were performed. The Imaris spot algorithm was then used for oocyte counting. Specific parameters used for oocyte counting by spot analysis were 8μm for estimated diameter and quality score above 30. For statistical analysis, average number of oocytes per ovary was counted in embryos from at least 3 different litters with few exceptions noted (Table 2-9). Variability between numbers due to timing of plug during the day as well as natural variation between embryos that is more apparent at earlier stages between litters and embryos. Statistical significance determined using two-tailed unpaired Student's t-test (Table 2-9).

Ovary sections. We used a complementary oocyte quantification method based on ovarian cryo-sections rather than whole-mount ovaries in cases with limited sample availability as they could be used for analysis of other markers in parallel such as L1 ORF1p. In these cases, such as for additional replicates of *Chk2*$^{-/-}$ and *Chk2*$^{+/-}$ untreated and AZT-treated ovaries and *Mili*$^{-/-}$;*Chk2*$^{-/-}$ +AZT ovaries and associated controls, we quantified oocytes by labeling with germ cell-specific antibody TRA98 and counterstained nuclei with DAPI. We then scored oocyte number in every 5th section through the entire ovary, and estimated total number per ovary. Statistical significance was determined using two-tailed unpaired Student's t-test.

We quantified primordial and non-primordial (primary, secondary, antral) follicles per ovary at P4 and P19 using ovarian paraffin sections. We performed immunohistochemistry and DAB staining on paraffin-embedded and sectioned ovaries and quantified follicles in every 5th section through the entire ovary, then estimated total number per ovary. Sections were labeled with the cytoplasmic germ cell marker MVH and nuclei counterstained with hematoxylin. MVH-positive follicles were quantified and categorized as primordial or non-primordial based on the number of somatic cell layers surrounding the oocyte. At least three ovaries from three different females were quantified for each experimental group and two-tailed unpaired Student's t-test used to determine statistical significance (Table 3-10).

6.5 Analysis of L1ORF1p and γH2AX relative mean nuclear fluorescence

Protocol to determine relative mean nuclear fluorescence (RMN) executed as previously described[69]. Ovary sections of 8μm thickness were stained with DAPI, germ cell marker TRA98, and L1 ORF1p. Confocal stacks were taken through the section. Imaris bitplane surface algorithm was used to generate a surface around each DAPI-positive nucleus in TRA98-positive germ cells. Then, RMN fluorescence was calculated for the channel containing L1 ORF1p signal within the surface. Specific parameter used for surface algorithm to outline nuclei was surface grain size of 0.283μm. This procedure was used in the same manner to calculate γH2AX RMN fluorescence. Each germ cell RMN value was then divided by the average of three RMN values from TRA98-negative somatic cell nuclei that should not contain L1 ORF1p nor γH2AX to normalize for background fluorescence. On average, about 200 oocytes were quantified per experimental group. Oocytes come from at least 3 different ovaries and 2 different litters unless noted otherwise (Table 3-1 – 3-2). Statistical significance was determined using two-tailed Kolmogorov-Smirnov test.

6.6 Analysis of Golgi element area

Ovary sections of 8μm thickness were stained with DAPI, germ cell marker TRA98 and GM130 [96]. Confocal stacks were taken through the section. Imaris bitplane was used to generate a surface around each GM130-positive Balbiani body region in a TRA98-positive cell. Area of surface generated for GM130 channel was calculated. Each bar represents 70 to 200 individual oocytes measured. Oocytes come from at least 3 different ovaries and 2 different litters unless noted otherwise (Table 3-9). Statistical significance was determined using two-tailed Kolmogorov-Smirnov test.

6.7 EdU labeling

For EdU labeling of embryonic ovaries, pregnant females were injected with 2.5μl/g body weight of 5mg/ml stock EdU. For posterior ovary labeling, inject at approximately E14.0. Dissection, fixation, blocking, antibody incubation and tissue clearing performed as described for whole-mount immunofluorescence. Ovaries can also be cryosectioned and used for immunofluorescence protocol on sections. EdU detection using Molecular probes by life technologies Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay kit follows secondary antibody incubation and washes. Incubate 30 minutes at room temperature for detection followed by 3 x 30 minutes washes in 1x PBS + 0.05% Tx100 and incubation with ScaleA2 clearing reagent if necessary.

6.8 Meiotic chromatin spread preparation

Ovaries were dissociated into a single-cell suspension using dissociation buffer containing 0.025% trypsin, 2.5 mg/mL collagenase, and 0.1mg/mL DNase I. One volume of Hypotonic buffer (30mM Tris-HCl, pH 8.2, 50mM sucrose, 17mM sodium citrate) was added to the cell suspension and set on nutator for 30 minutes. Cells were pelleted and supernatant replaced with 100μM sucrose, pH 8.2 solution. Approximately 600μl sucrose solution per ovary pair. Slides were dipped in fixative (1% PFA, 0.15% Triton X-100, pH 9.2) and 20μL resuspended cells pipetted along bottom edge. Cells were slowly spread

around slide by tilting the slide gently. Slides were dried in humid chamber for 2 hours, then treated with 0.08% Photo-Flo (Kodak). Slides used immediately for immunostaining or stored at -80°C.

6.9 Microscopy

Imaging of whole-mount ovaries and ovary sections was performed using TCS-SP5 laser-scanning confocal microscope (Leica, Buffalo Grove, IL), histological sections using Nikon Eclipse E800 microscope equipped with a Diagnostic Instruments model 2.3.1 digital camera, and meiotic spreads using Olympus BX61 microscope equipped with a Hamamatsu C4742-95 digital Camera as previously described [69,156]. Image analysis was completed using Imaris (Bitplane) and ImageJ.

6.10 Western blot

6-12 whole ovaries were lysed in RIPA buffer containing 50mM Tris-HCl pH=8, 150mM NaCl, 1%NP40, 1%SDS, 1mMEDTA, and 10% glycerol. 1mM PMSF and 1mM Halt protease and phosphatase inhibitor cocktail were added to buffer just before lysis. Ovaries were homogenized using RNase-free pestle and protein quantified using BCA. Lysates were run on 12% polyacrylamide running, 4% polyacrylamide stacking gel. Proteins transferred overnight at 4°C to PVDF membrane that had been activated for 15 seconds in 100% methanol followed by 2 minutes water and 15 min in transfer buffer. Membrane rinsed with PBS + 0.05% Tween-20 and blocked with 5% nonfat milk in PBS + 0.05% Tween-20 for 1 hour at room temperature. Primary antibodies were incubated overnight at 4°C in blocking buffer. Secondary antibodies used at 1:2000 and incubated 1 hour at room temperature. Detection by ECL was performed. Original western blots shown below (Fig. 6-2).

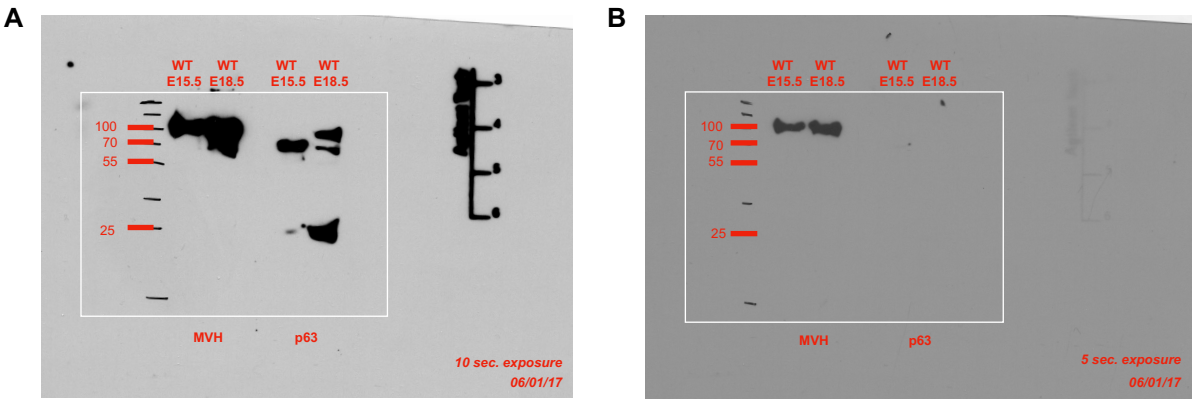


Figure 6-2: Original Western blots. **A)** Labeling of p63 and mouse vasa homolog (MVH) in wild-type E15.5 and E18.5 ovary lysates with 10 seconds exposure. **B)** Labeling of p63 and mouse vasa homolog (MVH) in wild-type E15.5 and E18.5 ovary lysates with 5 seconds exposure.

6.11 FACS

Isolation of oocytes (live). Ovaries were dissociated into a single-cell suspension using dissociation buffer containing 0.025% trypsin, 2.5 mg/mL collagenase, and 0.1mg/mL DNase I. Ovaries incubated at 37˚C for 30 minutes, pipetting cells every 10 minutes to dissociate. 10% fetal bovine serum was added to inactivate trypsin. Cell suspensions were filtered using 40μm filter, centrifuged at 1000 rpm, resuspended in PBS containing 3% fetal bovine serum, and filtered again. Cells were stained with propidium iodide for 10 minutes. Gating for negative red fluorescence was used to eliminate dead cells from the sort. Approximately 96% viability on average detected. Oocytes were FACS-sorted from remaining ovarian somatic cells into PBS containing 3% fetal bovine serum based on size and complexity using forward and side scatter parameters (Fig. 6-3A) [158]. BD FACS Aria III sorter was used for data collection. FACS Diva software was used for data analysis. Purity assessment of oocyte samples were determined by immunofluorescence detection of germ cell marker TRA98 on fixed sorted cells as well as by quantitative RT-PCR detection of germ cell gene mouse vasa homolog (MVH) (Fig. 6-3B, C).

Isolation of high and low L1 ORF1p-expressing oocytes (fixed). To isolate E15.5 and E18.5 high and low L1 ORF1p-expressing fetal oocytes by FACS, we fixed and double-immunostained ovarian single cell suspensions for Tra98 and L1 ORF1p as described in the Immunostaining cell suspension method. Only Tra98-positive cells were collected to ensure oocyte purity. Tra98-positive cells in the lower one-third of the L1 ORF1p intensity range and higher 1/3 L1 ORF1p intensity range collected for low and high L1 oocyte populations respectively (Fig 4-1C). All Tra98-positivie cells collected for E13.5 as L1 ORF1p is not yet expressed.

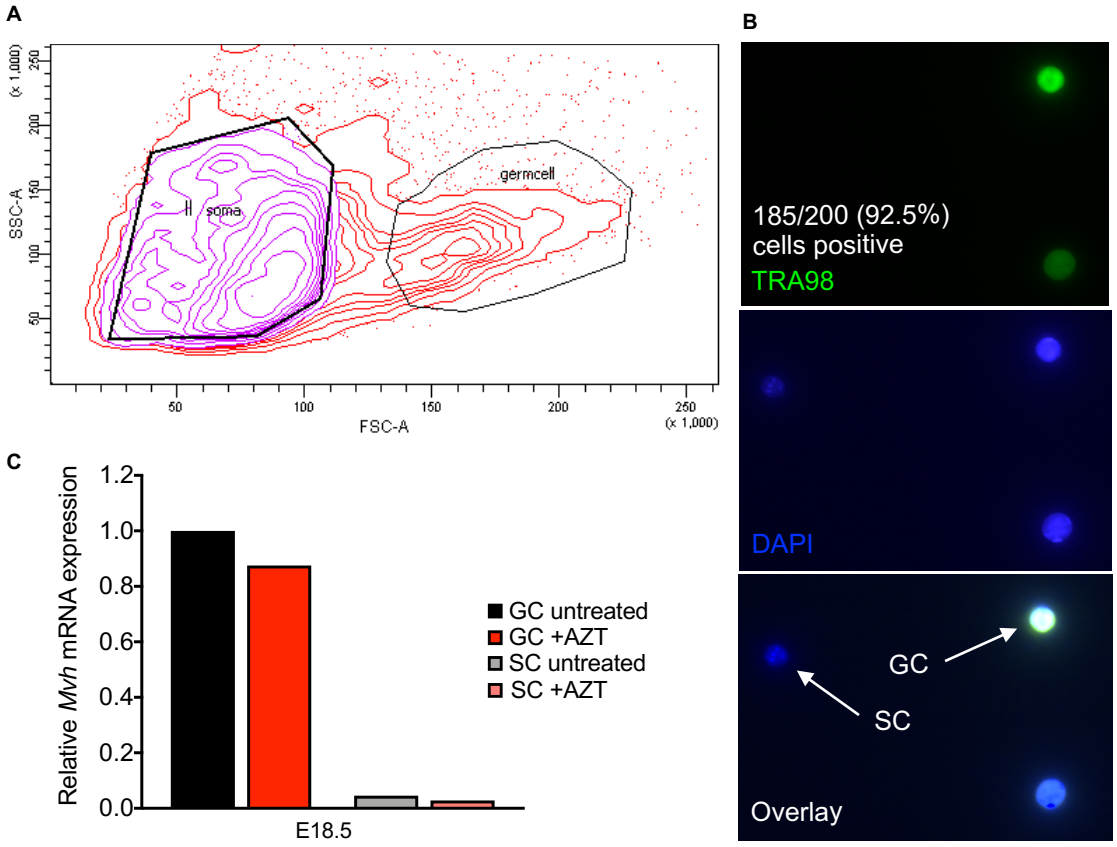


Figure 6-3: Oocyte FACS profile and purity assessment**. A)** FACS profile during isolation of oocytes from ovarian somatic cells based on side and forward scatter parameters. **B)** Analysis of oocyte purity in sorted oocyte sample based on TRA98 and DAPI staining. **C)** Analysis of oocyte purity in untreated and AZT-treated sorted oocytes and ovarian somatic cells using quantitative RT-PCR detection of germ cell-specific gene *Mvh* normalized to untreated oocyte sample with *Actb* as an endogenous control. One biological replicate per sample, each sample contains oocytes from at least 6 embryos.

6.12 Bulk ovary and oocyte mRNA-Seq

RNA preparation. Wild-type CD1 whole ovaries or sorted oocytes were obtained. At least 3 pairs of ovaries from a single litter or oocytes sorted from at least 6 ovary pairs from a single litter were used per biological replicate. 2 biological replicates were used for each whole ovary sample (E15.5 and E18.5 untreated and AZT-treated samples), 2 biological replicates were used for each sorted oocyte untreated sample (E15.5 and E18.5), and 3 biological replicates used for each sorted oocyte AZT-treated sample (E15.5 +AZT and E18.5 +AZT). RNA was extracted from samples using Trizol reagent (Invitrogen), DNaseI-treated using TURBO DNA-free kit (Ambion), and libraries generated using ribo-zero kit. 75bp unpaired, single-end reads were sequenced on Illumina Next-seq 500 system.

mRNA-Seq computational analysis. Any remaining rRNA sequences were removed computationally using Bowtie by aligning reads to mm10 rRNA genome. Non-rRNA reads were subsequently mapped to mm10 genome using Tophat splice aligner [65]. To determine differential gene expression, cuffdiff was used followed by cummeRbund in R to obtain FPKM values and generate plots. See computational pipelines 1 and 2 in Appendix for details on processing mRNA-Seq reads in terminal and differential gene expression analysis using cummeRbund in R. GO pathway enrichment analyses was performed using DAVID Bioinformatics Resources 6.8[66, 67]. See computational pipeline 6 in Appendix for details on locus-specific L1 analysis.

6.13 Small RNA-Seq

Small RNA preparation. Protocol adapted from ZZ Lab Illumina TruSeq Small RNA Cloning Protocol (Sep., 2015). Total RNA was extracted from at least three pairs of whole ovaries using mirVana miRNA isolation kit. RNA was run on 15% urea gel and 18-35 nucleotide region excised based on ladder markers for 18-nt and 35nt custom RNA oligos. Small RNAs were eluded from gel slice with 0.3M NaCl overnight at room temperature. 3' and 5' adapters were ligated. Samples were run on a second 10% urea polyacrylamide gel and 60nt to 100nt size region extracted and RNAs precipitated overnight in 0.3M NaCl. Reverse transcription reaction performed to generate cDNA. Libraries amplified by PCR, run on agarose gel and extracted using QIAquick Gel Extract kit. Libraries were run on the bioanalyzer for high sensitivity DNA to determine concentration and quality. Sometimes multiple bands are observed on the bioanalyzer because adapters are larger than small RNAs in between and may create different hybridization species. 75 or 150 base pair reads sequenced on Illumina Next-Seq 500 system. At least 3 pairs of ovaries from a single litter were used per biological replicate. Data shown represents one biological replicate per condition. Experiment repeated for WT E18.5 and WT P2 conditions (replicate data not shown). WT samples were of CD1 genetic background. *Mili*$^{+/+}$ and *Mili*$^{-/-}$ were of B6 genetic background.

Small RNA-Seq computational analysis. piPipes small RNA analysis pipeline was used to determine small RNA length distribution and to align reads to repeats[68]. First, we computationally removed adapter sequences and piPipes removes reads mapping to rRNA. Following this, piPipes aligns remaining reads to miRNA hairpins described in miRbase. These reads were removed, and used for normalization after aligning remaining non-rRNA, non-miRNA reads to the provided genome (mm10) using Bowtie. See computational pipeline 3 in Appendix for details on workflow used for processing and analysis of small RNA-Seq reads using piPipes in terminal.

6.14 Single-cell RNA-Seq

Single-cell RNA Preparation. At least six whole ovaries from a single litter were dissociated into a single-cell suspension using dissociation buffer containing 0.025% trypsin, 2.5 mg/mL collagenase, and 0.1mg/mL DNase I. Cells were pelleted and washed with PBS 3 times. Viability and cell count were determined using trypan blue staining and Countess II automated cell counter. Samples with greater than 90% viability and 1 million cells per milliliter were used for sequencing. GEM generation, barcoding, and library construction were performed with 10x genomics Chromium Genome Reagent Kit (v2 Chemistry). Libraries were sequenced on Illumina Next-Seq 500 system. For the untreated sample, 16,448 cells with 48,924 mean reads per cell were sequenced. For the AZT-treated sample, 15,551 cells with 48,450 mean reads per cell were sequenced. Experiment replicated for both UT and AZT-treated ovaries using fewer cells (~5000 cells per sample) and fewer reads (7,000-10,000 reads per cell). Replicate data not shown.

Single-cell RNA-Seq computational analysis. Differential gene expression and clustering analysis was performed using Cell Ranger v3.0 and Seurat v3.0 packages[69, 70]. Untreated and AZT-treated oocyte are subset from total ovarian cells based on expression of *Ddx4, Dazl,* and *Maelstrom* and lack of expression of *Xist.* Oocyte datasets are integrated and cluster analysis performed. Oocyte clusters are ordered from those containing early, middle, and late stage oocytes based on marker gene expression (*Ccnb3* = early, *Dppa3* = middle, and *Gdf9* = late, expression values represented as Ave log FC). Percent of oocytes belonging to untreated and AZT-treated samples were calculated for each cluster. See computation pipeline 4 in Appendix for details on clustering and differential gene expression analysis of single-cell RNA-Seq reads using Seurat in R.

6.15 Bisulfite PCR-Seq

Bisulfite-converted DNA preparation. Bisulfite conversation and purification performed using Epigentek kit. Nested PCR was performed to amplify target 5' L1 region containing 9 CpGs. PCR product run on agarose gel and gel extracted. DNA sent for Wide-Seq by Purdue Genomics Core.

Nested PCR primer sequences. F1-GTT AGA GAA TTT GAT AGT TTT TGG AAT AGG; R1-CCA AAA CAA AAC CTT TCT CAA ACA CTA TAT; F2-TAG GAA ATT AGT TTG AAT AGG TGA GAG GT; R2-TCA AAC ACT ATA TTA CTT TAA CAA TTC CCA[159].

Bisulfite PCR-Seq computational analysis. See computational pipeline 5 in Appendix for details on processing and analysis of bisulfite-converted wide-Seq reads using Bismark.

6.16 L1 reverse transcription intermediate isolation and quantification

Sorted oocytes or somatic cells from ≥6 pairs of E16.5-E17.5 WT CD1 untreated or AZT-treated ovaries from a single litter per sample were collected and treated with lysis buffer (100mM Tris pH8.5, 50mM NaCl, 5mM EDTA, 0.2% SDS) containing Proteinase K for 2 hours at 55°C. Lysate was treated with RNase A for 30 minutes at 37°C to remove mRNA and RNA from RNA:DNA hybrids, and followed by DNA extraction using Phenol:Chloroform pH=8 and salt/isopropanol precipitation. DNA abundance was measured using qBIT hsDNA system. Approximately 25ng used as input for downstream reactions. RNase A-treated DNA (input) was exposed to either dsDNase to isolate ssDNA, ssDNase (P1) + dsDNase to eliminate all DNA, or P1 alone to isolate dsDNA. P1 nuclease (NEB, cat# M0660S) reactions were performed in 10µL volume, incubated at 37°C for 30 minutes followed by inactivation at 75°C for 10 minutes. dsDNase (Thermofisher, cat# EN0771) reactions were performed in 20 µL volume, incubated at 37°C for 10 minutes followed by inactivation at 55°C for 5 minutes with addition of 10mM DTT. Resulting samples diluted equally and used for q-PCR detection of L1 *ORF1* and single-copy gene *Ifnb1*. Relative quantities from q-PCR are normalized to *Ifnb1* input relative quantity to account for total DNA concentration across samples. Then, resulting relative quantities are normalized to L1 *ORF1* input for respective sample. Five biological replicates used for untreated and AZT-treated, dsDNase-treated oocyte DNA, three biological replicates for untreated and AZT-treated, dsDNase-treated somatic cells, two biological replicates for untreated and AZT-treated, P1-treated oocyte DNA and P1 + dsDNase-treated oocyte DNA. DNase I-treated oocyte DNA and no template controls were also performed (data not shown). Statistical significance determined using two-tailed paired Student's t-test for comparison of untreated to AZT-treated samples and Mann-Whitney test for comparison of WT untreated and AZT-treated, dsDNase-treated oocytes to WT untreated and AZT-treated negative controls.

qPCR primer sequences. F-L1ORF1: ATG GCG AAA GGT AAA CGG AG; R-L1ORF1: AGT CCT

TCT TGA TGT CCT CT; F-ifnb1: CTG CGT TCC TGC TGT GCT TCT CCA; R-ifnb1: TTC TCC GTC ATC TCC ATA GGG ATC.

6.17 Quantitative RT-PCR

We isolated RNA from samples such as FACS-sorted oocytes and somatic cells or anterior and posterior thirds of whole ovaries using TRIZOL reagent (Invitrogen). RNA was DNase treated using TURBO DNA-free kit (Ambion). cDNA synthesis reactions were performed using oligo dT and Superscript III First-Strand Synthesis System for RT-PCR (Invitrogen). cDNA was diluted equally and added to the quantitative (q) RT-PCR reactions containing SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). qRT-PCR was performed on CFX96 Touch Real-Time PCR Detection System to detect SYBR Green. Relative quantities were analyzed using ΔΔCt methods with Actb as the housekeeping control gene.

qRT-PCR primer sequences. F-actb: CGG TTC CGA TGC CCT GAG GCT CTT; R-actb: CGT CAC ACT TCA TGA TGG AAT TGA; F-mvh: TGG CAG AGC GAT TTC TTT TT; R-mvh: CGC TGT ATT CAA CGT GTG CT; F-L1ORF1: ATG GCG AAA GGT AAA CGG AG; R-L1ORF1: AGT CCT TCT TGA TGT CCT CT; F-Stra8: GTA TCG CCG TAA CTC CCA GA; R-Stra8: GCA GAT GAC CCT CAC ACA AG; F-Sycp3: GTC AAG GGC TAA AGC AGT CG; R-Sycp3: TTG GCC GTG ACC TTT AAT TC.

6.18 Transgenic mice generation

RNase H2ACB-GFP construct. Transgenes were cloned into pBS31 using Gibson assembly. pBS31 sequence was analyzed by addgene online tool using EcoRI and MluI restriction sites to generate Gibson assembly homology sequences that were 30 nucleotides in length. The inserts were cloned according to the plasmid map and confirmed by Sanger sequencing. Inserts are color-coded and primer sequences underlined. The construct used for electroporation after confirmation of intact inserts and plasmid backbone was WT RNase H2 #1. The G37S mutation in the RNase H2A(G37S) mutant mouse is highlighted in yellow.

RNase H2ACB-GFP plasmid sequence.

GACCGATCCAGCCTCCGCGGCCCCGAATTCATGTACCCATACGATGTTCCAGATTACGCTGA
TCTCAGCGAGCTGGAGAGGGACAATACGGGTCGTTGTCGTCTGAGTTCTCCTGTACCTGCTG
TGTGTCTCAAGGAGCCGTGCGTTCTGGGCGTGGATGAAGCGGGCCGAGGCCCTGTGCTTGGT
CCCATGGTCTACGCCATCTGTTACTGCCCCCTGTCTCGCTTGGCAGATCTGGAGGCCCTGAAA
GTGGCAGACTCTAAGACCTTGACAGAGAACGAGCGGGAGAGGCTCTTTGCGAAAATGGAGG
AGGATGGAGACTTTGTGGGTTGGGCTTTGGACGTCCTGTCTCCAAACCTGATCTCTACCAGC
ATGCTTGGGCGAGTCAAGTACAACCTCAACTCCCTGTCACACGATACAGCTGCGGGGCTGAT
ACAGTACGCACTGGACCAGAATGTGAATGTCACTCAGGTATTTGTGGACACTGTAGGAATGC
CAGAGACATACCAGGCTCGATTACAACAGCACTTTCCCGGGATAGAGGTGACAGTCAAGGC
CAAAGCTGACTCCCTGTTCCCTGTGGTCAGTGCTGCCAGCATCTTTGCCAAGGTGGCCCGAG
ACAAGGCTGTGAAGAACTGGCAGTTTGTGGAAAATTTACAGGATCTGGACTCCGATTATGGC
TCAGGCTATCCCAATGATCCCAAGACCAAAGCCTGGCTGAGGAAACATGTGGACCCTGTGTT
TGGCTTTCCCCAGTTTGTACGGTTCAGTTGGAGCACAGCCCAGGCCATCCTGGAGAAGGAGG
CAGAAGATGTCATTTGGGAGGACTCAGAAGCTGAGGAGGATCCTGAAAGACCGGGAAAAAT
CACCTCCTACTTCAGCCAGGGCCCGCAGACCTGCCGCCCTCAGGCCCCCCACAGATACTTCC

AGGAGCGAGGCCTGGAGGCAGCCAGCAGCCTCGGCTCCGGCGAGGGCAGGGGAAGTCTTCT

AACATGCGGGGACGTGGAGGAAAATCCCGGCCCAATGAAGAACCCGGAGGAAGCGGCAGA

CGGGAAACAGCGTATTCACCTGCGCCCTGGCTCGCTGCGTGGTGCCGCACCGGCTAAGCTGC

ACCTCCTGCCCTGCGACGTTCTAGTCAGCCGGCCGGCCCCGGTAGATCGCTTCTTCACGCCC

GCCGTCCGCCACGATGCAGACGGGCTACAGGCGTCGTTTCGCGGTCGCGGCCTGCGGGGCG

AGGAGGTAGCTGTGCCGCCAGGGTTTGCGGGATTCGTGATGGTGACGGAGGAGAAGGGAGA

GGGGTTGATAGGGAAACTGAACTTCTCCGGGGACGCGGAGGACAAAGCGGACGAGGCGCA

GGAGCCGCTGGAGCGGGACTTCGACCGCCTTATCGGGGCCACCGGCAGCTTCAGCCATTTCA

CCTTGTGGGGTCTGGAAACGGTCCCGGGTCCAGATGCCAAAGTGCATAGGGCCCTAGGTTGG

CCCAGCCTCGCAGCAGCGATTCACGCCCAGGTCCCTGAGGACGGCAGCGGCGCCACAAACT

TCTCTCTGCTAAAGCAAGCAGGTGATGTTGAAGAAAACCCCGGGCCTATGGCCGGAGGTCG

GGACCGCGGGGACTTGGCGGCCAGGCAGCTAGTGTTCCTACTTCCAGAACATTTAAAAGATG

CCTCGAAGAAGAAGAAGAAGAGCAGCCTCCTGTTCGTAAAGCTGGCCAACCCGCACTCAGG

GGAAGGAGCCACATACTTAATTGATATGTGTCTTCAACAGCTGTTTGAAATAAAAGTTTTCA

AGGAAAAACACCATTCTTGGTTTATAAATCAATCAGTTCAATCAGGGGGCCTTCTCCACTTT

GCCACACCCATGGATCCATTGTTCCTGCTCCTTCACTATCTCCTAAAGGCTGGCAAAGAGGG

GAAGTATCAGCCCTTGGACCAAGTCGTGGTCGATGACACGTTTCCAGATTGCACCTTGCTGC

TGAGATTTCCTGAGCTTGAAAAGTCACTTCGGCATGTGACAGAGGAAAAAGAAGTGAACAG

CAAGAAGTACTATAAGTACAGCTCAGAGAAGACATTGAAGTGGCTGGAGAAGAAGGTCAAC

CAAACTGTGGTGGCACTAAAAGCTAATAATGTCAATGTTGGAGCCCGGGTTCAGTCATCTGC

ATATTTCTCTGGTGGTCAGGTTTCCAGGGACAAGGAAGAGGATTATGTTCGCTATGCCCATG

GTCTGATCTCTGATTACATCCCTAAAGAACTGAGTGATGATTTATCCAAGTTCTTGAAGCTTC

CAGAACCTCCAGCTTCATTGACCAACCCTCCATCAAAGAAACTAAAGTTATCAGATGAGCCT

GTAGAAGCCAAAGAAGATTACACTAAGTTTAACACTAAAGACTTGAAGACCGGCAAGAAAA

ATAGCAAAATGACTGCAGCTCAGAAGGCTTTGGCTAAAGTTGACAAAAGTGGAATGAAAAG

TATCGATGCCTTTTTTGGTGCAAAAAATAAAAAAACTGGAAAGATTGGCTCGGGCCAGTGTA

CTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAGAGCAACCCAGGTCCCATGGTGAGC

AAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAA

ACGGCCACAAGTTCAGCGTGTCTGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGAC

CCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCC

TGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTC

AAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA

ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCT

GAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTAC

AACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCA

AGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC

CCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCC

TGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC

CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAACGCGTTGAGAACTTCAGGGTGAG

TTTGGG

HA Tag – RNase H2A

Gibson assembly homology: GACCGATCCAGCCTCCGCGGCCCCGAATTC

ATG HA tag: ATGTACCCATACGATGTTCCAGATTACGCT

RNase H2A – T2A

RNase H2A left primer, no START codon: GATCTCAGCGAGCTGGAG

RNase H2A right primer, no STOP codon: GAGGCTGCTGGCTGCCTC

RNase H2A right primer, no STOP codon, reverse complement: GAGGCAGCCAGCAGCCTC

T2A:

GGCTCCGGCGAGGGCAGGGGAAGTCTTCTAACATGCGGGGACGTGGAGGAAAATCCCGGCC

CA

RNase H2C – P2A

RNase H2C left primer: ATGAAGAACCCGGAGGAAG

RNase H2C right primer, no STOP codon: GTCCTCAGGGACCTG

RNase H2C right primer, no STOP codon (reverse complement): CAGGTCCCTGAGGAC

P2A:

GGCAGCGGCGCCACAAACTTCTCTCTGCTAAAGCAAGCAGGTGATGTTGAAGAAAACCCCG

GGCCT

RNase H2B – E2A

RNase H2B left primer: ATGGCCGGAGGTCGGGAC

RNase H2B right primer, no STOP codon: AATCTTTCCAGTTTTTTTA

RNase H2B right primer, no STOP codon (reverse complement): TAAAAAAACTGGAAAGATT

E2A:

GGCTCGGGCCAGTGTACTAATTATGCTCTCTTGAAATTGGCTGGAGATGTTGAGAGCAACCC

AGGTCCC

GFP (c3GIC9 addgene EGFP)

GFP left primer: ATGGTGAGCAAGGGCGAGGA

GFP right primer: TTACTTGTACAGCTCGTCCAT

GFP right primer (reverse complement): ATGGACGAGCTGTACAAGTAA

Gibson assembly homology: ACGCGTTGAGAACTTCAGGGTGAGTTTGGG

Mutation of embryonic stem cells (ESCs). KH2 ESCs were grown on 0.1% gelatin coated plates containing DR4 MEF feeder cells. Feeder cell media contains 5ml Pen/Strep (100X), 5ml Glutamax and 50ml FBS to 500ml with DMEM+glucose. ESC media contains 5ml Pen/Strep 100X, 5ml Glutamax, 75ml FBS, 0.5ml 2-mercaptoethanol, 5ml NEAA, 125µl GSK3a/b inhibitor, 75µl Src inhibitor and 50µl LIF to 500ml with KO DMEM+glucose. ESCs grown to 20 million cells over 2 10cm petri dishes. Harvest, ESCs with 0.05% trypsin and count. Centrifuge and resuspend in volume to obtain 15 million cells in 750µl PBS. Add to cuvette with 30µg flip-in plasmid (containing transgene and hygromycin resistance gene) + 15µg FLPe plasmid in 30µl total volume. Pipette up and down to mix well. Set electroporation to 0.5kV (500V) and 25µF capacitance. Pulse twice for a few seconds each. Let cuvette and cells stand in hood for a while to recover, then add cells drop by drop to petri dish containing feeder cells and ESC media. Leave overnight. Next day add ESC media with 140µg/ml hygromycin added to begin the selection. After 9 days, pick surviving colonies and transfer to 96-well plate. Use for genotyping, expanding and freezing cell lines. Transgene and plasmid backbone sequences were verified using Wide-Seq prior to electroporation. See in computational pipeline 6 in Appendix for details on processing and analysis of Wide-Seq reads and visualization in IGV.

Preparation of ESCs for blastocyst injection. Prior to blastocyst injection, genotype to ensure presence of transgene, karyotype to ensure number of chromosomes is correct and test induction of GFP reporter upon DOX administration (1µg/ml DOX) for each ESC clone of interest (Fig. 6-4A, B). For injection, expand ESC clone in 6-well plate. Harvest cells with 0.05% trypsin, stop reaction with ESC media and pipette gently to dissociate. Let sit in 20ml ESC media divided between 2x 15ml conicals for 10-15 min to sediment feeder cells. Remove top 5ml from each conical, combine into single conical, spin.

Resuspend pellet in 0.5 ml ESC injection media. Injection media contains 200µl of 1M HEPES to 10ml ESC media. Keep cells on ice until injection.

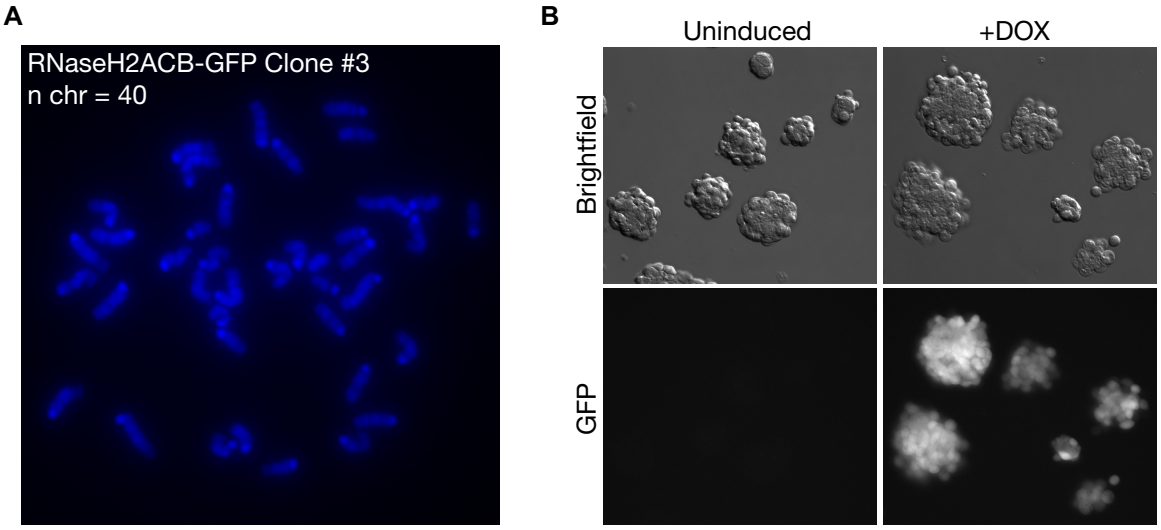


Figure 6-4: ESC preparation for blastocyst injection. **A)** Karyotype of representative RNaseH2ACB-GFP clone #3 ESC used for injection. 40 chromosomes counted. **B)** Expression of GFP reporter gene in ESC colonies that have been induced with DOX or uninduced. Brightfield and GFP channels shown.

CRISPR-Cas9 mutation of Trex1. *Trex1* mutation was generated by electroporating Cas9/sgRNA RNP into CD1 zygotes that were cultured to 2-cell stage and oviduct transferred into recipient CD1 females. The guide RNA sequence used was: TTGTGTGCCACAAGGCAGCA. Resulting mutation was interpreted as deletion of 8 nucleotides and insertion of 2 nucleotides that resulted in a frame shift at codon 119 and immediate termination (PC_ATQR*).

Mutation (founder #8554-1):

>TREX1_MOUSE Q91XB0 Three-prime repair exonuclease 1, ECO:0000250|UniProtKB:Q9NSU2 (3.1.11.2) (3'-5' exonuclease TREX1, ECO:0000303|PubMed:10391904).

MGSQTLPHGHMQTLIFLDLEATGLPSSRPEVTELCLLAVHRRALENTSISQGHPPPVPRP PRVVDKLSLCIAPGKACSPGASEITGLSKAELEVQGRQRFDDNLAILLRAFLQRQPQ**P(118)CC L**VAHNGDRYDFPLLQTELARLSTPSPLDGTFCVDSIAALKALEQASSPSGNGSRKSYSLG SIYTRLYWQAPTDSHTAEGDVLTLLSICQWKPQALLQWVDEHARPFSTVKPMYGTPATTG TTNLRPHAATATTPLATANGSPSNGRSRRPKSPPPEKVPEAPSQEGLLAPLSLLTLLTLA IATLYGLFLASPGQ

6.19 Fertility Assays

Chk2$^{-/-}$ females that were treated with AZT during their fetal development and raised to adults were crossed to *Chk2*$^{+/-}$ males. *Chk2*$^{-/-}$ females that were untreated as well as *Chk2*$^{+/-}$ females that were untreated were crossed to *Chk2*$^{+/-}$ males for controls. The number of live pups per litter at the day of birth from 6 *Chk2*$^{-/-}$ +AZT females, 3 *Chk2*$^{-/-}$ females, and 6 *Chk2*$^{+/-}$ females were monitored for at least 10 months. Number of litters over 10 months were reported in females that survived the duration of the assay. Statistical significance was determined using unpaired Student's t-test.

6.20 Statistical information

Two-tailed unpaired student's t-test used for oocyte quantification and fertility assays. Two-tailed paired student's t-test used for quantitative PCR comparing L1 *ORF1* DNA between WT untreated and WT +AZT samples and quantitative RT-PCR. Mann-Whitney test used for macrophage analysis and for quantitative PCR comparing L1 *ORF1* DNA between WT untreated dsDNase (oocyte) and WT untreated negative controls. Two-tailed Kolmogorov-Smirnov test used for all RMN fluorescence experiments. Chi-square test used for meiotic progression analysis. Statistics calculated using GraphPad Prism 7 Software.

6.21 Data and materials availability

The NCBI Sequence Read Archive project number for all high-throughput sequencing data is: PRJNA543598. All relevant data are available from the authors.

APPENDIX

Computational pipeline 1: Processing mRNA-Seq reads

mRNA-Seq pipeline

#samples
GC15-1, GC15-2, GC15AZT-1, GC15AZT-2, GC15AZT-3
GC18-1, GC18-2, GC18AZT-1, GC18AZT-2, GC18AZT-3
WO15-1, WO15-2, WO15AZT-1, WO15AZT-2
WO18-1, WO18-2, WO18AZT-1, WO18AZT-2

```
#concatenate and unzip
$ zcat *fastq.gz > GC15-1.fastq

#run fastqc
$ fastqc GC15-1.fastq

#open html, want mean to be in green region, OK for some failures

#removing rRNA
$ bowtie -p16 --un filtered_GC15-1.fastq
/mnt/sequence/genomes/mouse/mm10-ribosomal GC15-1.fastq

#use filtered_GC15-1.fastq going forward as rRNA is removed

#align reads to mm10 genome using tophat
$ nohup tophat -p16 -o Mapping_GC15-1 /mnt/sequence/genomes/mouse/mm10
/mnt/sequence/tharp/GC_RNAseq_CD1/GC15-1/filtered_GC15-1.fastq

#output is bam file

#calculate differential gene expression using cuffdiff
$ cuffdiff -p15 -o
/mnt/sequence/tharp/GC_RNAseq_CD1/CuffDiff_analysis/Diff_GC15-1
/mnt/sequence/tharp/iGenomes/Mus_musculus/UCSC/mm10/Annotation/Genes/g
enes.gtf -L GC15,WO15 GC15-1_accepted_hits.bam,GC15-
2_accepted_hits.bam WO15-1_accepted_hits.bam,WO15-2_accepted_hits.bam

#biological replicates bam files are separated by comma

#import into RStudio for cummerBund analysis to make graphs
#see cummeRbund scripts
```

cummeRbund

Marla Tharp

November 3, 2019

#load cummeRbund

```
.libPaths( "/mnt/sequence/R/3.6.1/cummeRbund" )
library( "cummeRbund" )
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: parallel
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter,
##     Find, get, grep, grepl, intersect, is.unsorted, lapply, Map,
##     mapply, match, mget, order, paste, pmax, pmax.int, pmin,
##     pmin.int, Position, rank, rbind, Reduce, rownames, sapply,
##     setdiff, sort, table, tapply, union, unique, unsplit, which,
##     which.max, which.min
```

```
## Loading required package: RSQLite
```

```
## Loading required package: ggplot2
```

```
## Loading required package: reshape2
```

```
## Loading required package: fastcluster
```

```
##
## Attaching package: 'fastcluster'
```

```
## The following object is masked from 'package:stats':
##
##     hclust
```

```
## Loading required package: rtracklayer
```

```
## Loading required package: GenomicRanges
```

```
## Loading required package: stats4
```

```
## Loading required package: S4Vectors
```

```
##
## Attaching package: 'S4Vectors'

## The following object is masked from 'package:base':
##
##     expand.grid

## Loading required package: IRanges

## Loading required package: GenomeInfoDb

## Loading required package: Gviz

## Loading required package: grid

##
## Attaching package: 'cummeRbund'

## The following object is masked from 'package:GenomicRanges':
##
##     promoters

## The following object is masked from 'package:IRanges':
##
##     promoters

## The following object is masked from 'package:BiocGenerics':
##
##     conditions
```

#import cuffdiff data

```
cuff<-readCufflinks("/mnt/sequence/tharp/Marla_sequencing/bam/DGE_paper1518AZT/")
cuff
```

```
## CuffSet instance with:
##   8 samples
##   24347 genes
##   34535 isoforms
##   27685 TSS
##   26886 CDS
##   679784 promoters
##   775180 splicing
##   574448 relCDS
```

#Find fpkm and stdev of gene set

```
myGeneIds<-c("Sycp3","Hormad1","Dazl")
myGeneIds
```

```
## [1] "Sycp3"   "Hormad1" "Dazl"
```

```
myGenes<-getGenes(cuff, myGeneIds)
```

```
## Warning: Closing open result set, pending rows

## Warning: Closing open result set, pending rows
```

```
myGenes
```

```
## CuffGeneSet instance for  3  genes
##
## Slots:
```

```
##   annotation
##   fpkm
##   repFpkm
##   diff
##   count
##   isoforms    CuffFeatureSet instance of size 7
##   TSS         CuffFeatureSet instance of size 3
##   CDS         CuffFeatureSet instance of size 5
##   promoters       CuffFeatureSet instance of size 3
##   splicing        CuffFeatureSet instance of size 3
##   relCDS      CuffFeatureSet instance of size 3
```

```
fpkm(myGenes)
```

```
##    gene_id sample_name      fpkm   conf_hi  conf_lo quant_status     stdev
## 1     Dazl        GC15  231.5150  333.1590 129.8710           OK  50.82200
## 2     Dazl     GC15AZT  311.1150  469.0980 153.1310           OK  78.99150
## 3     Dazl        GC18  181.9330  255.7970 108.0700           OK  36.93200
## 4     Dazl     GC18AZT  395.2700  590.7280 199.8110           OK  97.72900
## 5     Dazl        WO15   80.7366  113.9200  47.5533           OK  16.59170
## 6     Dazl     WO15AZT  157.3010  229.5210  85.0802           OK  36.11000
## 7     Dazl        WO18  116.3910  164.0660  68.7158           OK  23.83750
## 8     Dazl     WO18AZT  118.8780  166.0240  71.7327           OK  23.57300
## 9  Hormad1        GC15  102.4460  143.6590  61.2327           OK  20.60650
## 10 Hormad1     GC15AZT  156.5960  223.3410  89.8513           OK  33.37250
## 11 Hormad1        GC18   51.5597   71.1062  32.0131           OK   9.77325
## 12 Hormad1     GC18AZT  116.5650  161.3050  71.8255           OK  22.37000
## 13 Hormad1        WO15   55.5476   77.6794  33.4157           OK  11.06590
## 14 Hormad1     WO15AZT   70.4881   98.7260  42.2501           OK  14.11895
## 15 Hormad1        WO18   29.2370   40.7732  17.7007           OK   5.76810
## 16 Hormad1     WO18AZT   44.1129   60.7463  27.4795           OK   8.31670
## 17   Sycp3        GC15  521.0810  759.3750 282.7870           OK 119.14700
## 18   Sycp3     GC15AZT  711.4150 1090.1200 332.7140           OK 189.35250
## 19   Sycp3        GC18  489.1940  693.2310 285.1570           OK 102.01850
## 20   Sycp3     GC18AZT 1164.2000 1764.5100 563.9020           OK 300.15500
## 21   Sycp3        WO15  249.6380  367.7520 131.5230           OK  59.05700
## 22   Sycp3     WO15AZT  366.9830  544.6850 189.2800           OK  88.85100
## 23   Sycp3        WO18  298.1300  434.6140 161.6450           OK  68.24200
## 24   Sycp3     WO18AZT  428.8980  623.3300 234.4650           OK  97.21600
```

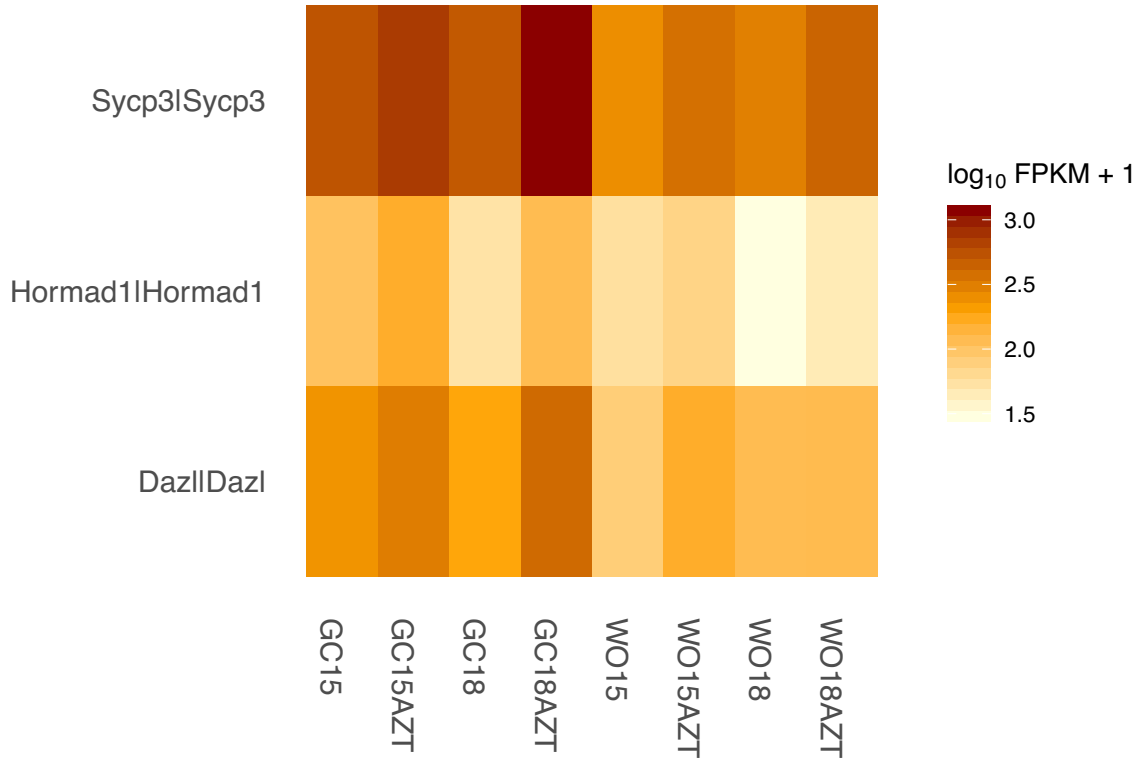
#Generate heatmap to visualize gene set

```
hm <-csHeatmap(myGenes)
```

```
## Using tracking_id, sample_name as id variables
```

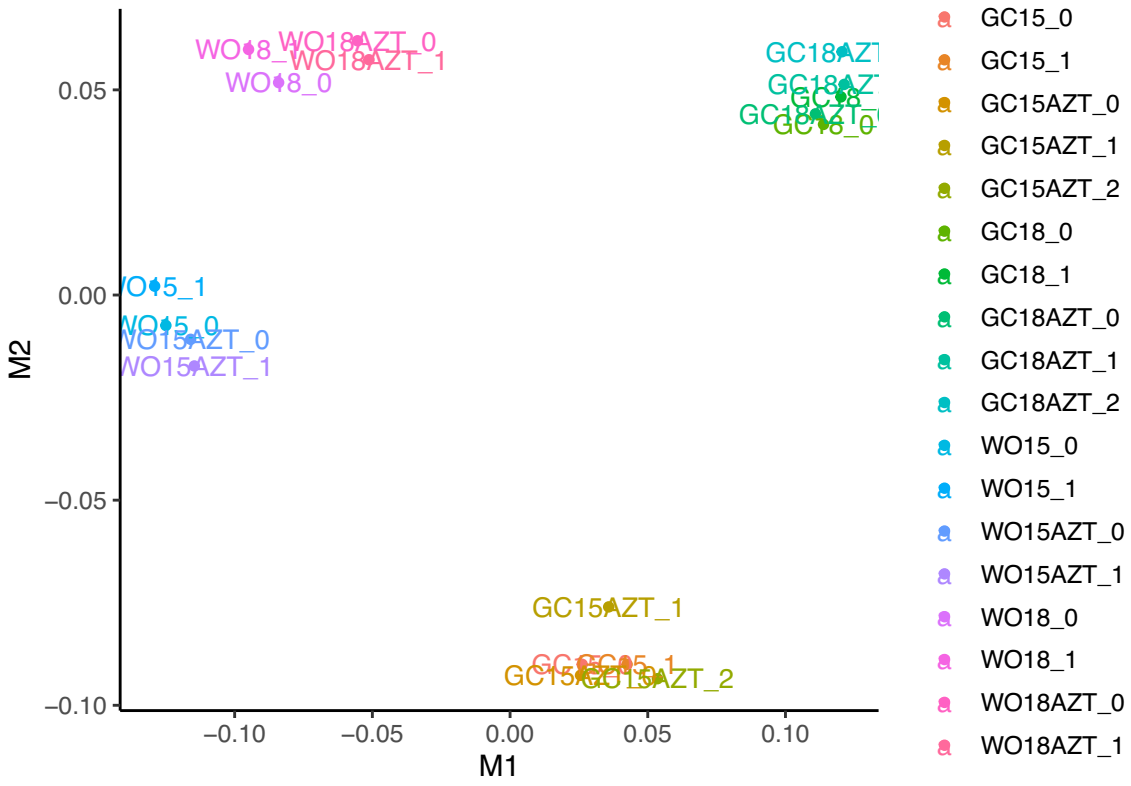
```
## No id variables; using all as measure variables
```

```
hm <- hm + theme(axis.text.x=element_text(size=12),axis.text.y=element_text(size=12))
hm
```

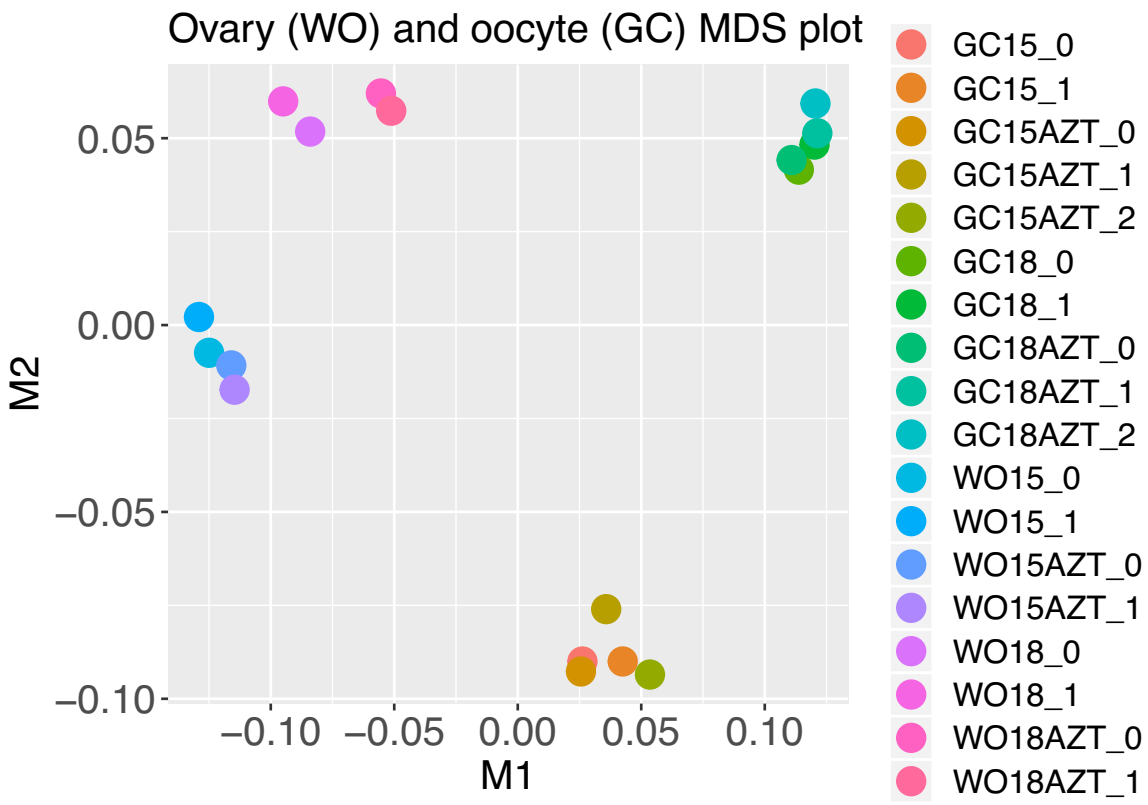


#Generate MDS plot

```
p <- MDSplot( genes(cuff), replicates=TRUE )
p + theme_classic( base_size=12 )
```

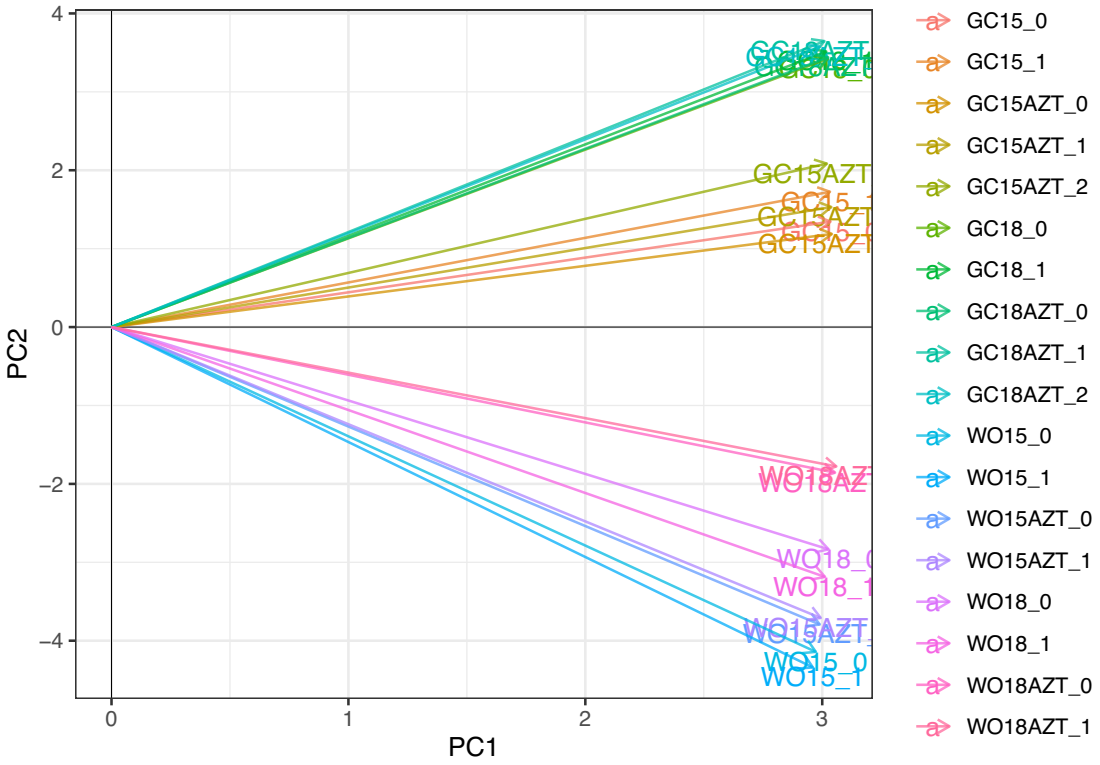



```
b <- ggplot_build( p )
ggplot( data.frame(b$data[2]) ) +
  aes(x=M1, y=M2) +
  geom_point( aes( x, y, color=label ), size=5) +
  theme(axis.text = element_text(size=16), text = element_text(size=16)) +
  labs(title = "Ovary (WO) and oocyte (GC) MDS plot" ) +
  theme(plot.title = element_text(size=16))
```



#Generate PCA plot

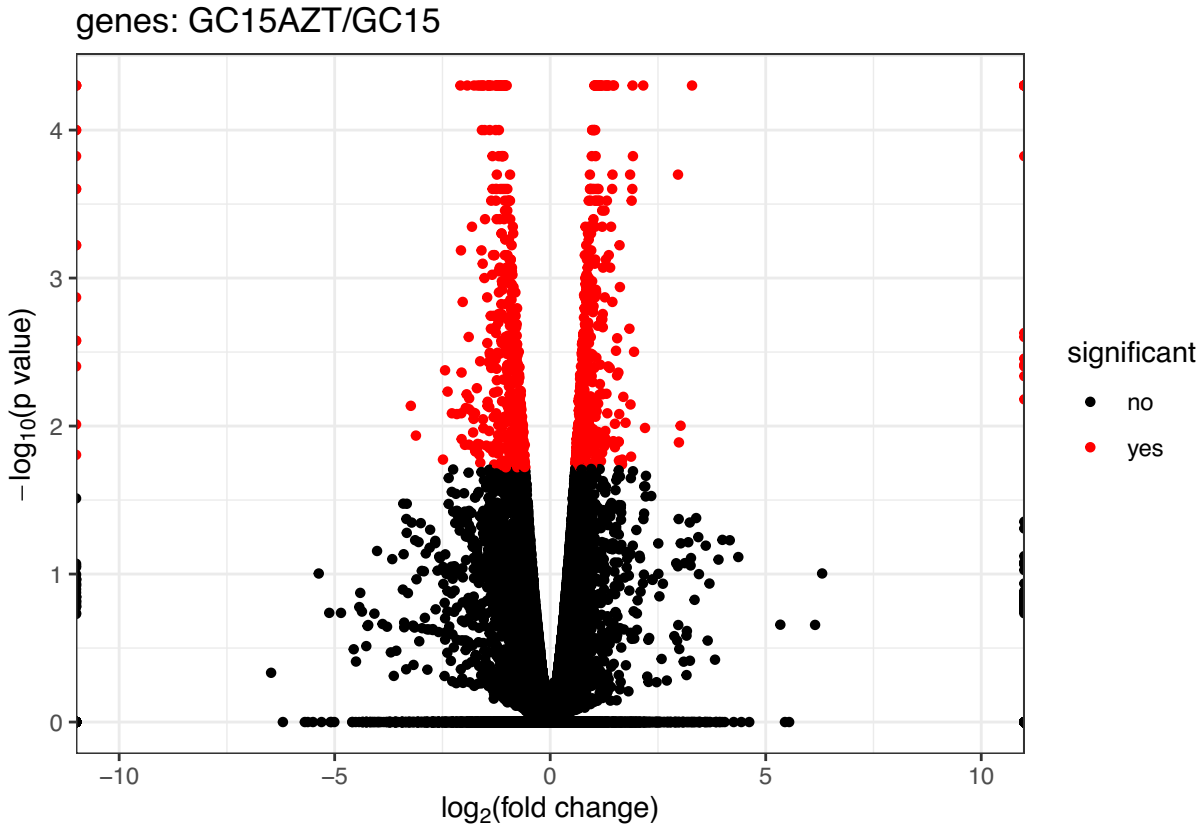
```
q<- PCAplot(genes(cuff), x="PC1", y="PC2", replicates = T, showPoints = F, pseudocount = T)
q
```



#Generate volcano plots

```
v<-csVolcano(genes(cuff), 'GC15', 'GC15AZT', alpha=0.05, showSignificant=TRUE, xlim = c(-10,10))
v
```

```
## Warning: Removed 11 rows containing missing values (geom_point).
```


genes: GC15AZT/GC15
4
3
2
1
0
−log10(p value)
−10
−5
0
5
10
log2(fold change)
significant
no
yes

Computational pipeline 3: Processing small RNA-Seq reads and analysis

Small RNA-Seq pipeline

#samples
WT E15.5 ovary, E18.5 ovary, P2 ovary
Mili+/+ P2 ovary, *Mili-/-* P2 ovary
Mael+/- P2 ovary, *Mael-/-* P2 ovary

#create directory to contain analysis
```
$ mkdir 180319-piPipes
$ cd 180319-piPipes
```

#create copy of raw .fastq files
```
$ zcat /mnt/sequence/FastQ/180315_NB501387_0215_AHKVFNBGX5/FastQ/AB-marlatharp/smallRNA_ATCACG/*.fastq.gz > smallRNA.fastq
```

#remove adapters using code from ZZ (for output: "smallreads" has no adapters, "shortreads" has adapters)
```
$ perl /mnt/sequence/zhang/bin/rm_3linker.pl smallRNA.fastq
```

#load piPipes
```
$ source /mnt/sequence/linux_3/piPipes/perl5/perlbrew/etc/bashrc
$ source /mnt/sequence/linux_3/piPipes/env/bin/activate
$ module load R/3.3.3
$ module load piPipes/git
```

#align reads and quantitate a single condition
```
$ nohup piPipes small -i smallRNA.fastq -g mm10 -c 8 -o smallRNA_out > nohup.smallRNA.out &
```

#compare two conditions and normalize to miRNA for unoxidized samples
```
$ nohup piPipes small2 -a smallRNAMilipl_out -b smallRNAMilimin_out -g mm10 -c 8 -N miRNA
```

#details for understanding output files and commands at https://github.com/bowhan/piPipes/wiki/smallRNA-seq

Seurat analysis

#connecting to server and start Seurat

```
.libPaths( "/mnt/sequence/R/3.6.1/Seurat/" )
.libPaths()
```

```
reticulate::use_virtualenv( "/mnt/sequence/R/3.6.1/Seurat/env/" )
reticulate::py_config()
```

```
library( "Seurat" )
```

```
## Registered S3 method overwritten by 'R.oo':
##   method        from
##   throw.default R.methodsS3
```

#import first dataset to Seurat object (Untreated)

```
MT1.data <- Read10X( "/mnt/sequence/10x/MT1_count/outs/filtered_feature_bc_matrix/" )
MT1 <- CreateSeuratObject( MT1.data, project="Untreated" )
MT1
```

```
## An object of class Seurat
## 27998 features across 16446 samples within 1 assay
## Active assay: RNA (27998 features)
```

#Normalization

```
MT1 <- NormalizeData( MT1 )
```

```
MT1 <- FindVariableFeatures( MT1 )
```

```
MT1 <- ScaleData( MT1 )
```

```
## Centering and scaling data matrix
```

#Run PCA

```
MT1 <- RunPCA( MT1 )
```

```
MT1 <- FindNeighbors( MT1 )
```

```
## Computing nearest neighbor graph
```

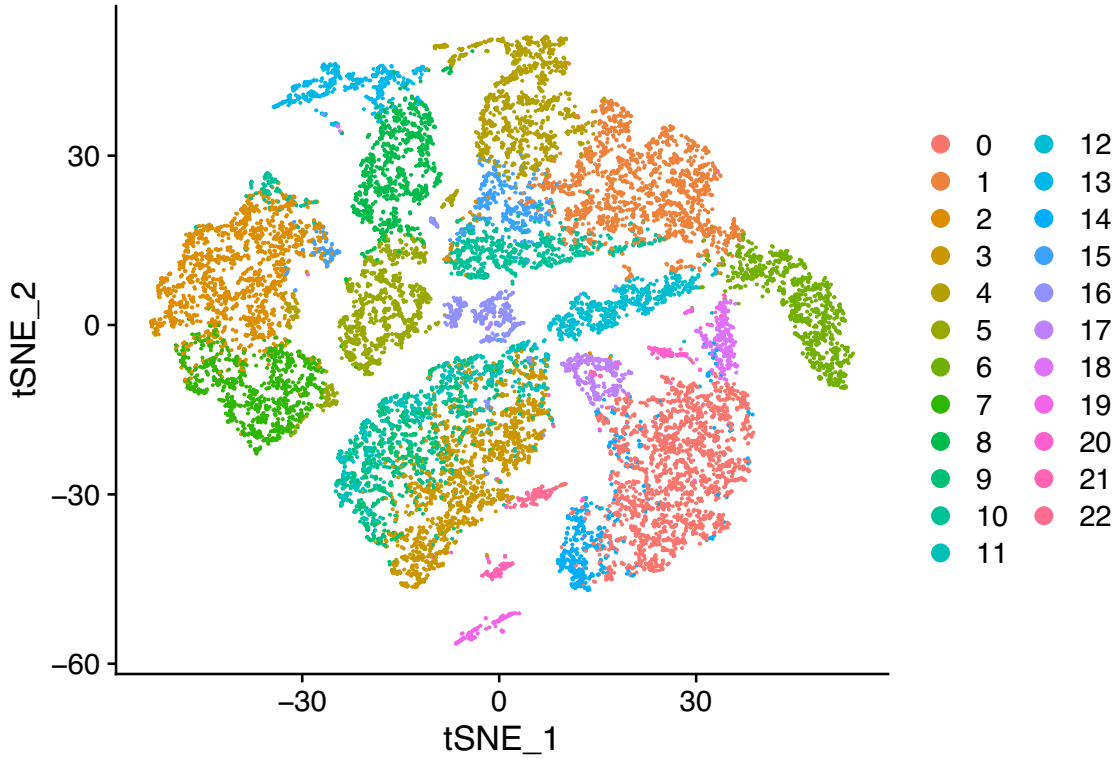
```
## Computing SNN
```

```
MT1 <- FindClusters( MT1 )
```

#Run tSNE

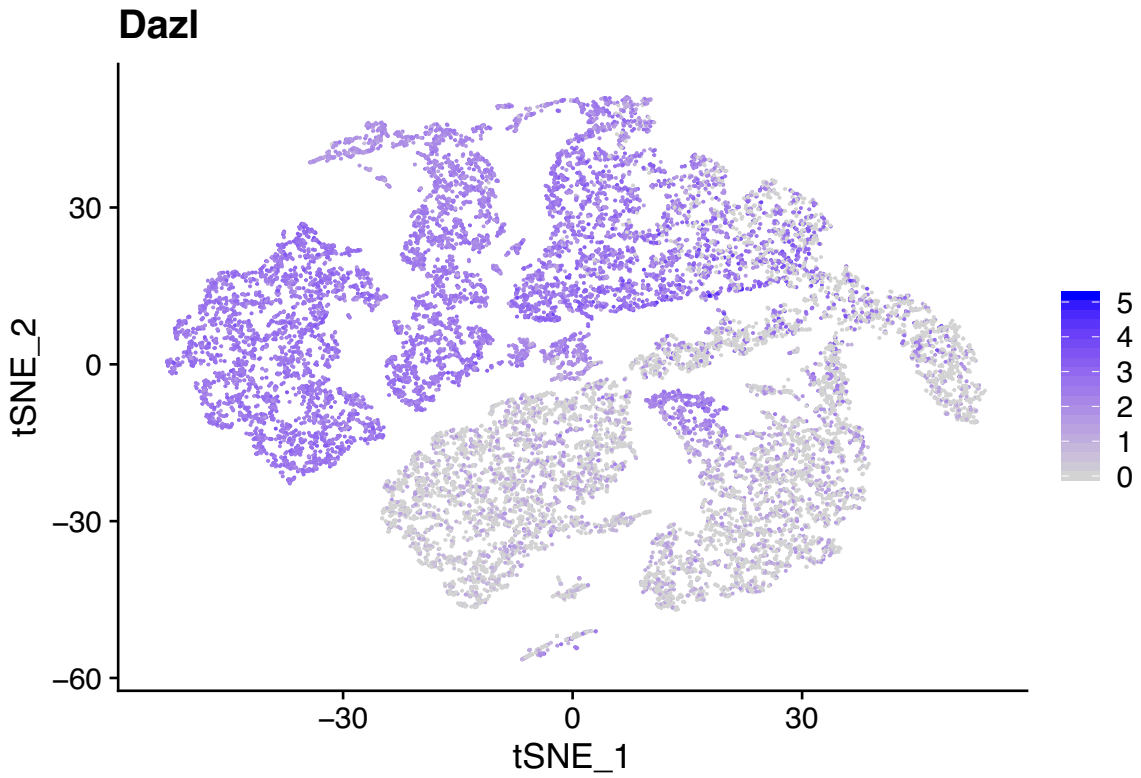
```
MT1 <- RunTSNE( MT1 )
```

```
DimPlot(object = MT1, reduction = "tsne")
```



#Feature plot showing expression of gene of interest

```
FeaturePlot(MT1, "Dazl")
```



#Subset oocytes based on marker genes

```
MT1S <- subset(x=MT1, subset = Ddx4 > 0.5 & Dazl > 0.5 & Mael > 0.5 & Xist < 0.5)
MT1S
```

#import second dataset to Seurat object (+AZT), normalize and subset oocytes

```
MT2.data <- Read10X( "/mnt/sequence/10x/MT2_count/outs/filtered_feature_bc_matrix/" )
MT2 <- CreateSeuratObject( MT2.data, project="AZT" )
MT2
```

```
## An object of class Seurat
## 27998 features across 15551 samples within 1 assay
## Active assay: RNA (27998 features)
```

```
MT2 <- NormalizeData( MT2 )
```

```
MT2 <- FindVariableFeatures( MT2 )
```

```
MT2 <- ScaleData( MT2 )
```

```
## Centering and scaling data matrix
```

```
MT2 <- RunPCA( MT2 )
```

```
MT2 <- FindNeighbors( MT2 )
```

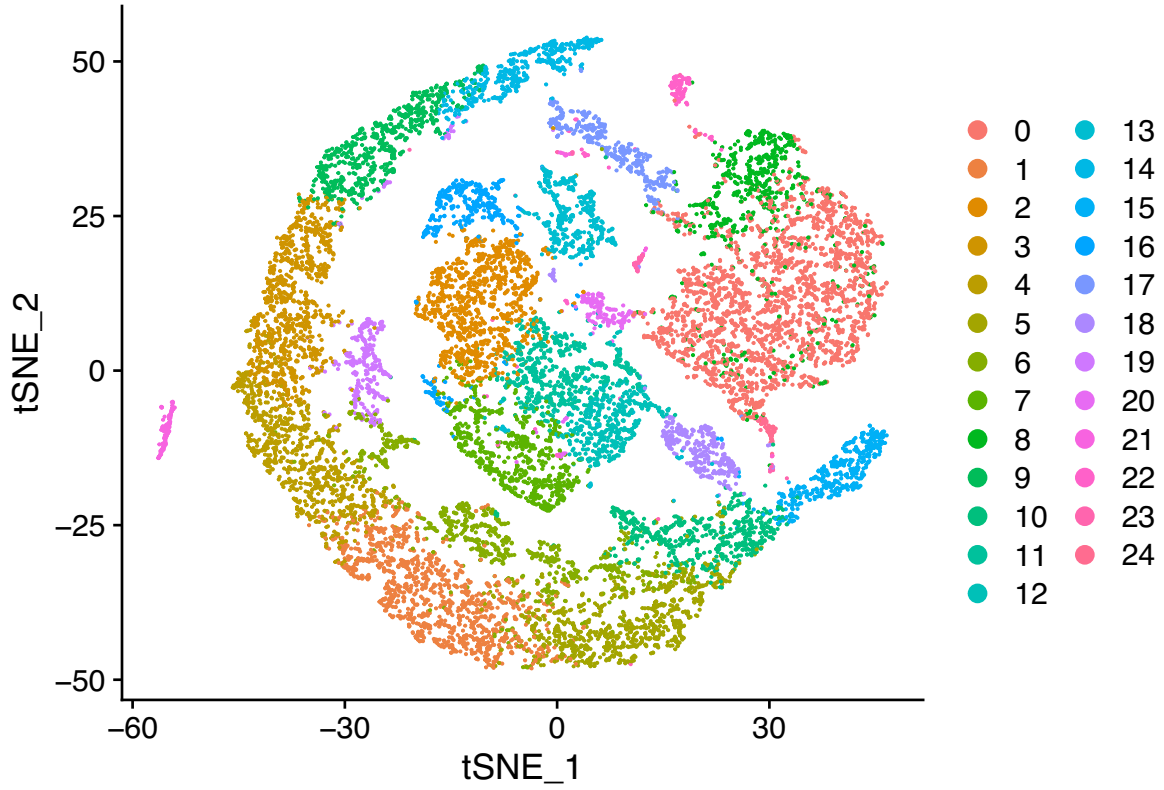
```
## Computing nearest neighbor graph
```

```
## Computing SNN
```

```
MT2 <- FindClusters( MT2 )
```

```
MT2 <- RunTSNE( MT2 )
```

```
DimPlot(object = MT2, reduction = "tsne")
```




```
MT2S <- subset(x=MT2, subset = Ddx4 > 0.5 & Dazl > 0.5 & Mael > 0.5 & Xist < 0.5)
MT2S
```

```
## An object of class Seurat
## 27998 features across 3520 samples within 1 assay
## Active assay: RNA (27998 features)
##  2 dimensional reductions calculated: pca, tsne
```

#Integrate oocyte subset datasets

```
MT12S <- merge( MT1S, MT2S, add.cell.ids=c("Untreated","AZT") )
MT12S
```

```
## An object of class Seurat
## 27998 features across 7902 samples within 1 assay
## Active assay: RNA (27998 features)
```

#Adjust for batch effect differences

```
MT12S.list <- SplitObject(MT12S, split.by = "orig.ident")
```

#preprocessing, log-normalization,identify variable features

```
for (i in 1:length(MT12S.list)) {
    MT12S.list[[i]] <- NormalizeData(MT12S.list[[i]], verbose = FALSE)
    MT12S.list[[i]] <- FindVariableFeatures(MT12S.list[[i]], selection.method = "vst",
        nfeatures = 2000, verbose = FALSE)}
```

```
reference.list <- MT12S.list[c("Untreated","AZT")]
MT12S.anchors <- FindIntegrationAnchors(object.list = reference.list, dims = 1:30)
```

```
## Computing 2000 integration features
## Scaling features for provided objects
## Finding all pairwise anchors
## Running CCA
## Merging objects
## Finding neighborhoods
## Finding anchors
##  Found 11838 anchors
## Filtering anchors
##  Retained 4906 anchors
## Extracting within-dataset neighbors
```

```
MT12S.integrated <- IntegrateData(anchorset = MT12S.anchors, dims = 1:30)
```

```
## Merging dataset 2 into 1
## Extracting anchors for merged samples
## Finding integration vectors
## Finding integration vector weights
## Integrating data
```

```
library(ggplot2)
library(cowplot)
```

```
##
## ************************************************************
## Note: As of version 1.0.0, cowplot does not change the
##   default ggplot2 theme anymore. To recover the previous
##   behavior, execute:
##   theme_set(theme_cowplot())
## ************************************************************
```

#Normalizations with integrated data

```
DefaultAssay(MT12S.integrated) <- "integrated"
```

```
MT12S.integrated <- ScaleData( MT12S.integrated, verbose = FALSE )
```

```
MT12S.integrated <- RunPCA( MT12S.integrated)
```

```
MT12S.integrated <- FindNeighbors( MT12S.integrated )
```

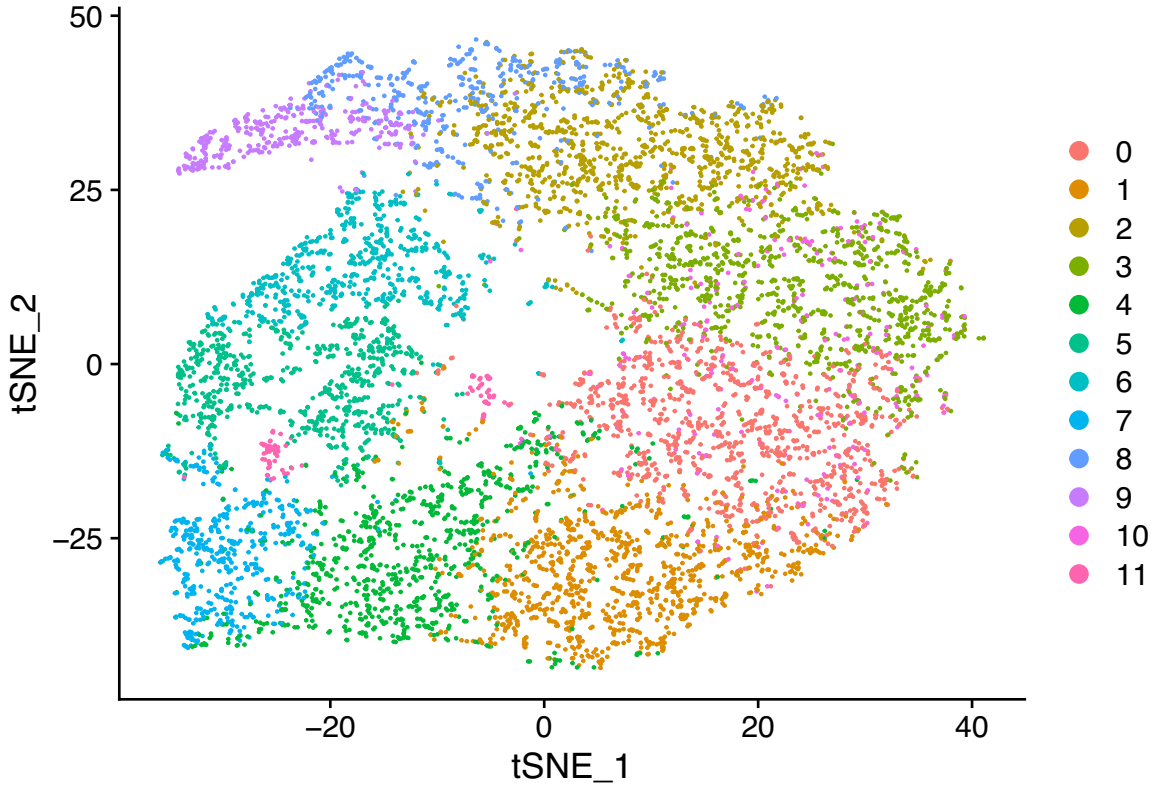
```
## Computing nearest neighbor graph
## Computing SNN
```

```
MT12S.integrated <- FindClusters( MT12S.integrated )
```

#tSNE with integrated data

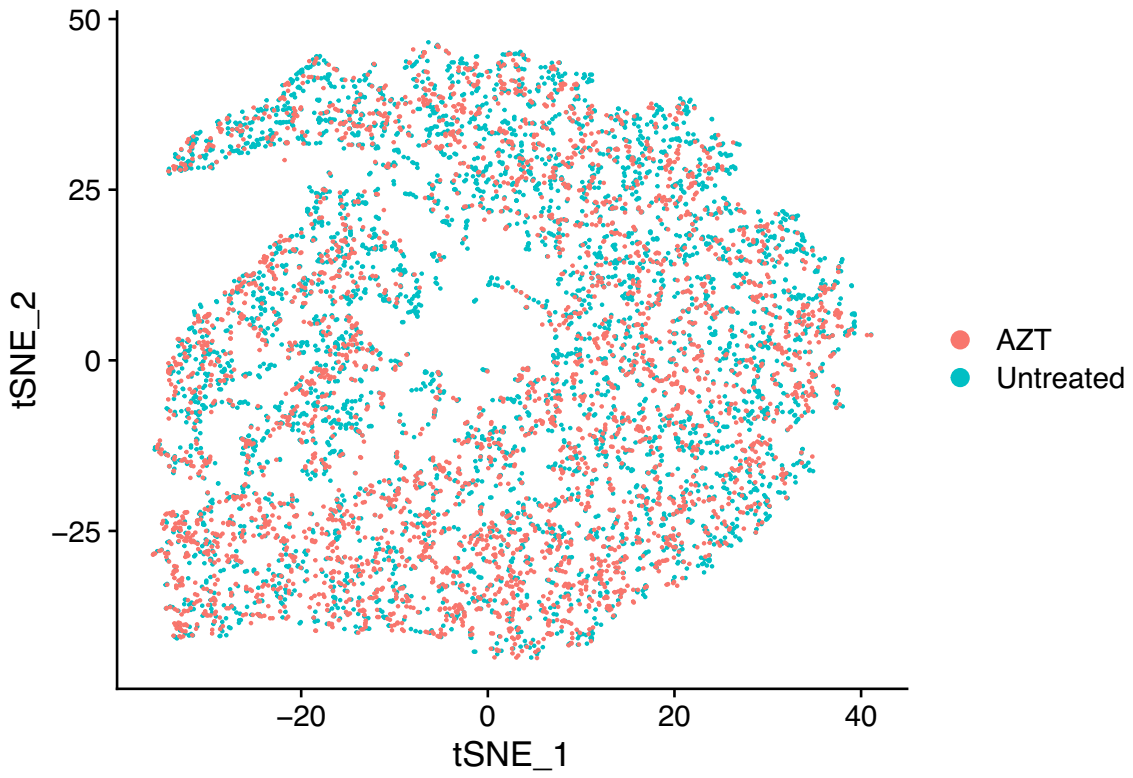
```
MT12S.integrated <- RunTSNE( MT12S.integrated)
```

```
DimPlot(object = MT12S.integrated, reduction = "tsne")
```



#tSNE with original identities

```
DimPlot(object = MT12S.integrated, reduction = "tsne", group.by = "orig.ident")
```




```
table(MT12S.integrated$orig.ident, MT12S.integrated$integrated_snn_res.0.8)
```

```
##
##                0   1   2   3   4   5   6   7   8   9  10  11
##   AZT        473 593 401 413 467 272 204 284 146 123 109  35
##   Untreated 580 426 618 592 396 430 371 194 306 235 170  64
```

#determine gene expression in each cluster

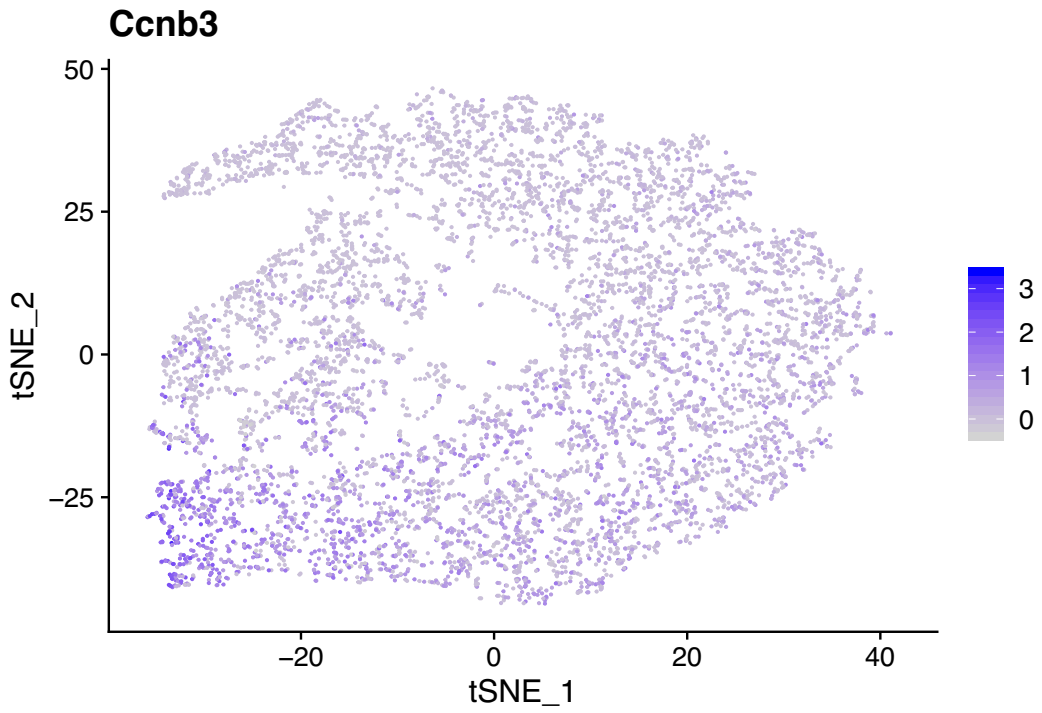
```
a <- FindAllMarkers( MT12S.integrated, assay="RNA" )
```

```
## Calculating cluster 0
## Calculating cluster 1
## Calculating cluster 2
## Calculating cluster 3
## Calculating cluster 4
## Calculating cluster 5
## Calculating cluster 6
## Calculating cluster 7
## Calculating cluster 8
## Calculating cluster 9
## Calculating cluster 10
```

```
## Calculating cluster 11
```

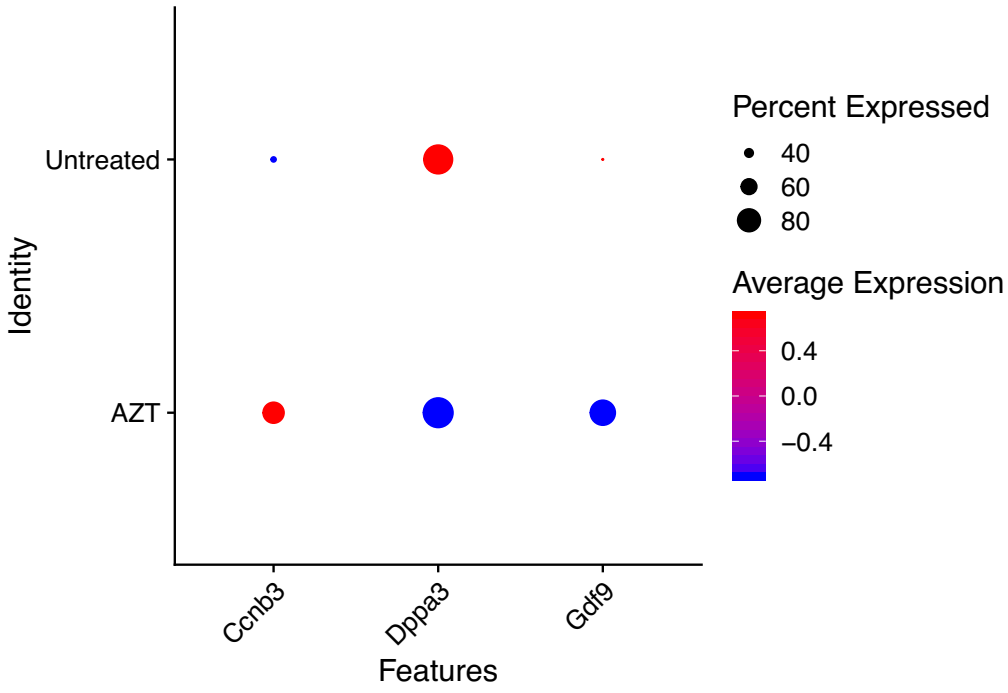
```
a
```

```
FeaturePlot(MT12S.integrated, "Ccnb3")
```



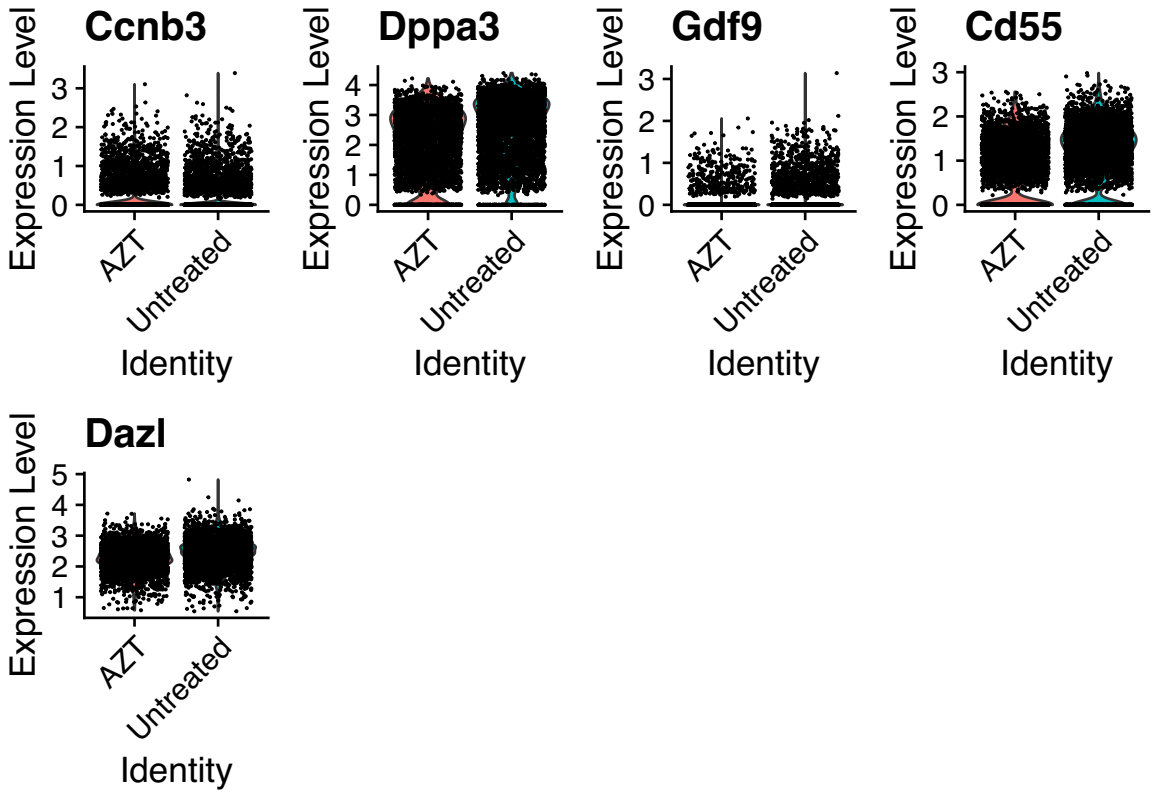
#Plot expression of genes and % cells that express gene per sample

```
Idents(MT12S.integrated) <- factor(Idents(MT12S.integrated))
markers.to.plot <- c("Ccnb3","Dppa3","Gdf9")
DotPlot(MT12S.integrated, features = rev(markers.to.plot), cols = c("blue","red"), group.by = "orig.ide
```

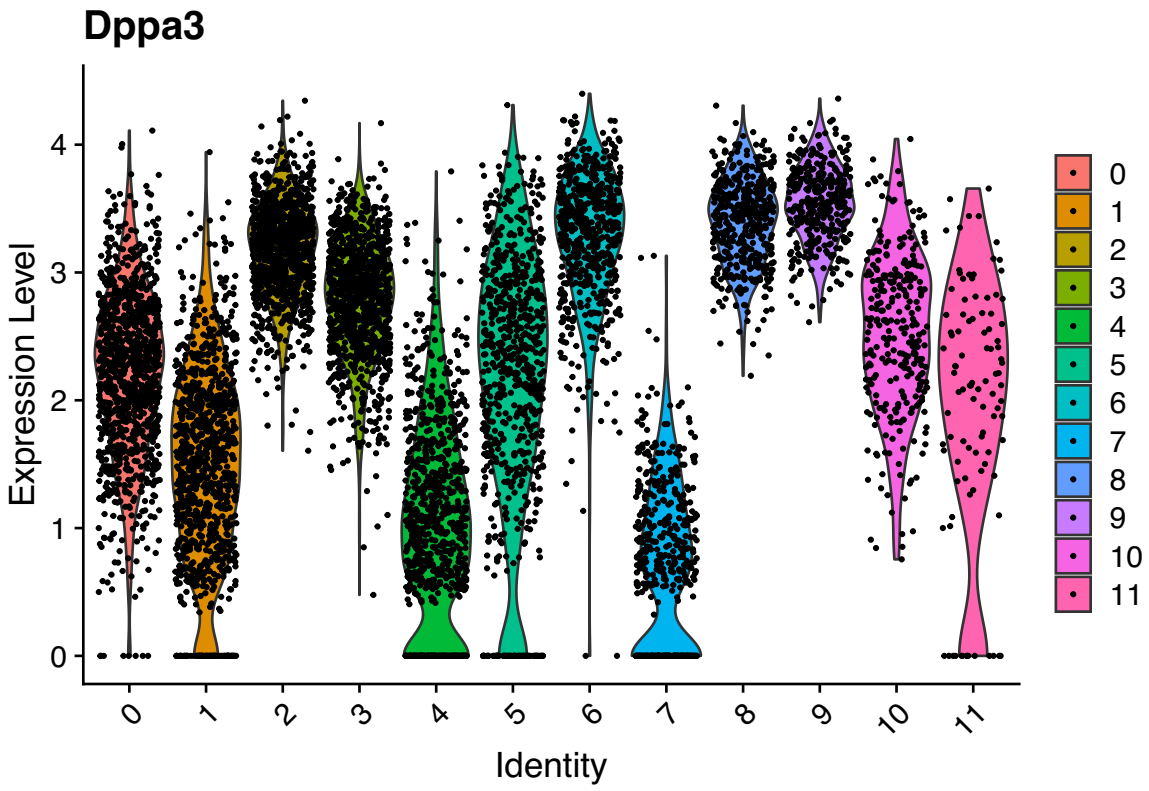


#generate violin plots of gene expression

```
VlnPlot(MT12S.integrated, features = c("Ccnb3","Dppa3", "Gdf9", "Cd55", "Dazl"), pt.size = 0.1, ncol = 
```




```
VlnPlot(MT12S.integrated, features = c("Dppa3"), pt.size = 0.5, assay = "RNA")
```



#find expression of all genes, use RNA assay for most accurate expression value rather than integrated dataset

```
AverageExpression(MT12S.integrated, assays = "RNA")
```

```
## Finished averaging RNA for cluster 0
## Finished averaging RNA for cluster 1
## Finished averaging RNA for cluster 2
## Finished averaging RNA for cluster 3
## Finished averaging RNA for cluster 4
## Finished averaging RNA for cluster 5
## Finished averaging RNA for cluster 6
## Finished averaging RNA for cluster 7
## Finished averaging RNA for cluster 8
## Finished averaging RNA for cluster 9
## Finished averaging RNA for cluster 10
## Finished averaging RNA for cluster 11
```

#find expression of candidate genes, use RNA assay

```
AverageExpression(MT12S.integrated, assays = "RNA", features = c("Gdf9", "Ccnb3", "Dppa3"))
```

```
## Finished averaging RNA for cluster 0
```

```
## Finished averaging RNA for cluster 1
## Finished averaging RNA for cluster 2
## Finished averaging RNA for cluster 3
## Finished averaging RNA for cluster 4
## Finished averaging RNA for cluster 5
## Finished averaging RNA for cluster 6
## Finished averaging RNA for cluster 7
## Finished averaging RNA for cluster 8
## Finished averaging RNA for cluster 9
## Finished averaging RNA for cluster 10
## Finished averaging RNA for cluster 11
## $RNA
##                  0          1          2           3          4           5
## Gdf9   0.02144741 0.01734374  0.2823828  0.06897318 0.02161643  0.04306164
## Ccnb3  0.33259766 0.58740259  0.1018547  0.19077568 1.06764310  0.46483246
## Dppa3 10.07834591 4.82399513 24.7112920 16.88619235 2.91609397 11.62177218
##                6          7           8           9          10         11
## Gdf9   0.2824178 0.02256228  0.49356602  1.56289612  0.05283977 0.5022479
## Ccnb3  0.1039557 2.55202290  0.05687947  0.04138699  0.32502153 0.4764757
## Dppa3 28.3193210 1.52665219 30.33739537 34.88336458 13.93866028 9.4098044
```

Computational pipeline 5: Processing wide-Seq reads

Analysis of transgene and plasmid backbone sequences

#build reference genome from .fasta file
$ bowtie2-build PBS31_G37S_RnaseH2ACB_EGFP.fasta PBS_G37S_RNaseH2ACB_EGFP

#align reads to reference genome
$ bowtie2 -p 48 -x ../bowtie2/genomes/PBS_G37S_RNaseH2ACB_EGFP -1 ../bowtie2/rawdata/003547_63-RNaseH2_S28_L001_R1_001.fastq.gz -2 ../bowtie2/rawdata/003547_63-RNaseH2_S28_L001_R2_001.fastq.gz -S G37S_63.sam

#create sorted.bam file
$ samtools view -Sb -F 4 G37S_63.sam > G37S_63.bam
$ samtools sort G37S_63.bam -o G37S_63.sorted.bam
$ samtools index G37S_63.sorted.bam.bam

#load sorted.bam into IGV with corresponding reference genome

Analysis of L1 DNA methylation amplicon

#build reference genome from .fasta file (genome folder contains L1 unconverted amplicon.fa)
$ bismark_genome_preparation –bowtie2 /mnt/sequence/tharp/L1Amplicon/

#align reads to reference genome
$ bismark --bowtie2 --bam -p 4 --score_min L,0,-0.4 /mnt/sequence/tharp/L1Amplicon/ --se reads.fq

#deduplication
$ deduplicate_bismark -s --bam reads_bismark_bt2.bam

#M-bias on deduplicated samples
$ bismark_methylation_extractor --multicore 4 --mbias_only --single-end reads_bismark_bt2.deduplicated.bam

#Report for html showing m-bias to determine parameters for methylation extractor
$ bismark2report reads_bismark_bt2.deduplicated.M-bias.txt

#make directory for output called "methextract," but do not go inside
$ mkdir methextract

#methylation extract
$ bismark_methylation_extractor --multicore 2 --gzip --single-end --no_overlap --report --ignore 6 --ignore_3prime 1 -o methextract --comprehensive reads_bismark_bt2.deduplicated.bam > nohupextract

Computational pipeline 6: Locus-specific L1 mRNA-Seq analysis

#samples
GC15-1, GC15-2, GC15AZT-1, GC15AZT-2, GC15AZT-3
GC18-1, GC18-2, GC18AZT-1, GC18AZT-2, GC18AZT-3

#align reads to mm10 using STAR (unique)
$ STAR --runMode alignReads --genomeDir /mnt/sequence/genomes/mouse/mm10/STAR/ --sjdbGTFfile /mnt/sequence/genomes/mouse/mm10_rmsk_TE.gtf --runThreadN 16 --quantMode TranscriptomeSAM GeneCounts --outSAMtype BAM SortedByCoordinate --outFileNamePrefix GC18-1_mm10_rmsk --readFilesIn /mnt/sequence/tharp/Marla_sequencing/mRNA_oocyte/rawdata_-rRNA_fastq/filtered_GC18-1.fastq

$ samtools sort Aligned.toTranscriptome.out.bam -o output.sorted.bam
$ samtools index output.sorted.bam
$ samtools idxstats output.sorted.bam > output.txt

#align reads to mm10 using STAR (multi-map)
$ STAR --runMode alignReads --genomeDir /mnt/sequence/genomes/mouse/mm10/STAR/ --sjdbGTFfile /mnt/sequence/genomes/mouse/mm10_rmsk_TE.gtf --runThreadN 16 --quantMode TranscriptomeSAM GeneCounts --outSAMtype BAM SortedByCoordinate --outFileNamePrefix GC18-1_mm10_rmsk --readFilesIn /mnt/sequence/tharp/Marla_sequencing/mRNA_oocyte/rawdata_-rRNA_fastq/filtered_GC18-1.fastq --outFilterMultimapNmax 100

$ samtools sort Aligned.toTranscriptome.out.bam -o output.sorted.bam
$ samtools index output.sorted.bam
$ samtools idxstats output.sorted.bam > output.txt

#align reads to mm10 using STAR (multi-map with downstream re-assignment)
$ STAR --runMode alignReads --genomeDir /mnt/sequence/genomes/mouse/mm10/STAR/ --runThreadN 16 --outSAMtype BAM SortedByCoordinate --readFilesIn /mnt/sequence/tharp/Marla_sequencing/mRNA_oocyte/rawdata_-rRNA_fastq/filtered_GC18-1.fastq --outFilterMultimapNmax 100

#re-assignment of multi-mapped reads using Telescope
$ telescope assign /mnt/sequence/tharp/L1_mm10_telescope/GC18-1/Aligned.sortedByCoord.out.bam /mnt/sequence/genomes/mouse/mm10_rmsk_TE.gtf --attribute transcript_id --updated_sam

#normalize to total number of reads after rRNA removal

References

1 Ohinata, Y. *et al.* Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207-213, doi:10.1038/nature03813 (2005).

2 Lehmann, R. Germ Plasm Biogenesis--An Oskar-Centric Perspective. *Current topics in developmental biology* **116**, 679-707, doi:10.1016/bs.ctdb.2015.11.024 (2016).

3 Illmensee, K. & Mahowald, A. P. 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, doi:10.1073/pnas.71.4.1016 (1974).

4 Hayashi, K., de Sousa Lopes, S. M. & Surani, M. A. Germ cell specification in mice. *Science (New York, N.Y.)* **316**, 394-396, doi:10.1126/science.1137545 (2007).

5 Saitou, M., Barton, S. C. & Surani, M. A. A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293-300, doi:10.1038/nature00927 (2002).

6 Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* **3**, 662-673, doi:10.1038/nrg887 (2002).

7 Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Molecular cell* **48**, 849-862, doi:10.1016/j.molcel.2012.11.001 (2012).

8 Hargan-Calvopina, J. *et al.* Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. *Developmental cell* **39**, 75-86, doi:10.1016/j.devcel.2016.07.019 (2016).

9 Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet* **2**, 21-32, doi:10.1038/35047554 (2001).

10 Greally, J. M. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 327-332, doi:10.1073/pnas.012539199 (2002).

11 Molaro, A. *et al.* Two waves of de novo methylation during mouse germ cell development. *Genes & development* **28**, 1544-1549, doi:10.1101/gad.244350.114 (2014).

12 Aravin, A. A. & Bourc'his, D. Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes & development* **22**, 970-975, doi:10.1101/gad.1669408 (2008).

13 Pezic, D., Manakov, S. A., Sachidanandam, R. & Aravin, A. A. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes & development* **28**, 1410-1428, doi:10.1101/gad.240895.114 (2014).

14 Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96-99, doi:10.1038/nature02886 (2004).

15 Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-903, doi:10.1038/nature02633 (2004).

16 Gahurova, L. *et al.* Transcription and chromatin determinants of de novo DNA methylation timing in oocytes. *Epigenetics & chromatin* **10**, 25, doi:10.1186/s13072-017-0133-5 (2017).

17 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).

18 Comfort, N. C. From controlling elements to transposons: Barbara McClintock and the Nobel Prize. *Trends in genetics : TIG* **17**, 475-478, doi:10.1016/s0168-9525(01)02383-6 (2001).

19 Fedoroff, N. V. McClintock's challenge in the 21st century. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20200-20203, doi:10.1073/pnas.1215482109 (2012).

20 Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature genetics* **42**, 631-634, doi:10.1038/ng.600 (2010).

21 Moran, J. V., DeBerardinis, R. J. & Kazazian, H. H., Jr. Exon shuffling by L1 retrotransposition. *Science (New York, N.Y.)* **283**, 1530-1534, doi:10.1126/science.283.5407.1530 (1999).

22 Van't Hof, A. E. *et al.* The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**, 102-105, doi:10.1038/nature17951 (2016).

23 Indjeian, V. B. *et al.* Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. *Cell* **164**, 45-56, doi:10.1016/j.cell.2015.12.007 (2016).

24 Session, A. M. *et al.* Genome evolution in the allotetraploid frog Xenopus laevis. *Nature* **538**, 336-343, doi:10.1038/nature19840 (2016).

25 Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annual review of genetics* **41**, 331-368, doi:10.1146/annurev.genet.40.110405.090448 (2007).

26 Kazazian, H. H., Jr. & Moran, J. V. Mobile DNA in Health and Disease. *The New England journal of medicine* **377**, 361-370, doi:10.1056/NEJMra1510092 (2017).

27 Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nature genetics* **23**, 314-318, doi:10.1038/15490 (1999).

28 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040

291/5507/1304 [pii] (2001).

29 Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562, doi:10.1038/nature01262 (2002).

30 Boissinot, S., Entezam, A. & Furano, A. V. Selection against deleterious LINE-1-containing loci in the human lineage. *Molecular biology and evolution* **18**, 926-935, doi:10.1093/oxfordjournals.molbev.a003893 (2001).

31 Naas, T. P. *et al.* An actively retrotransposing, novel subfamily of mouse L1 elements. *The EMBO journal* **17**, 590-597, doi:10.1093/emboj/17.2.590 (1998).

32 Sookdeo, A., Hepp, C. M., McClure, M. A. & Boissinot, S. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mobile DNA* **4**, 3, doi:10.1186/1759-8753-4-3 (2013).

33 Jachowicz, J. W. & Torres-Padilla, M. E. LINEs in mice: features, families, and potential roles in early development. *Chromosoma* **125**, 29-39, doi:10.1007/s00412-015-0520-2 (2016).

34 Babushok, D. V. & Kazazian, H. H., Jr. Progress in understanding the biology of the human mutagen LINE-1. *Hum Mutat* **28**, 527-539, doi:10.1002/humu.20486 (2007).

35 Goodier, J. L. & Kazazian, H. H., Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* **135**, 23-35, doi:10.1016/j.cell.2008.09.022 (2008).

36 Martin, S. L. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *J Biomed Biotechnol* **2006**, 45621, doi:S1110724306456217 [pii] 10.1155/JBB/2006/45621 (2006).

37 Martin, S. L. Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA biology* **7**, 706-711 (2010).

38 Khazina, E. *et al.* Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nature structural & molecular biology* **18**, 1006-1014, doi:10.1038/nsmb.2097 (2011).

39 Wei, W. *et al.* Human L1 retrotransposition: cis preference versus trans complementation. *Molecular and cellular biology* **21**, 1429-1439, doi:10.1128/mcb.21.4.1429-1439.2001 (2001).

40 Feng, Q., Moran, J. V., Kazazian, H. H., Jr. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905-916 (1996).

41 Gasior, S. L., Wakeman, T. P., Xu, B. & Deininger, P. L. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* **357**, 1383-1393, doi:S0022-2836(06)00142-2 [pii] 10.1016/j.jmb.2006.01.089 (2006).

42 Cost, G. J., Feng, Q., Jacquier, A. & Boeke, J. D. Human L1 element target-primed reverse transcription in vitro. *The EMBO journal* **21**, 5899-5910, doi:10.1093/emboj/cdf592 (2002).

43 Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595-605, doi:10.1016/0092-8674(93)90078-5 (1993).

44 Kazazian, H. H., Jr. *et al.* Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164-166, doi:10.1038/332164a0 (1988).

45 Wallace, N. A., Belancio, V. P. & Deininger, P. L. L1 mobile element expression causes multiple types of toxicity. *Gene* **419**, 75-81, doi:10.1016/j.gene.2008.04.013 (2008).

46 Goodier, J. L. Restricting retrotransposons: a review. *Mobile DNA* **7**, 16, doi:10.1186/s13100-016-0070-z (2016).

47 Deniz, O., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat Rev Genet*, doi:10.1038/s41576-019-0106-6 (2019).

48 Di Giacomo, M. *et al.* Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Molecular cell* **50**, 601-608, doi:10.1016/j.molcel.2013.04.026 (2013).

49 Bulut-Karslioglu, A. *et al.* Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Molecular cell* **55**, 277-290, doi:10.1016/j.molcel.2014.05.029 (2014).

50 Zamudio, N. *et al.* DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes & development* **29**, 1256-1270, doi:10.1101/gad.257840.114 (2015).

51 Kim, S. *et al.* PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Molecular cell* **56**, 564-579, doi:10.1016/j.molcel.2014.10.003 (2014).

52 Bortvin, A. PIWI-interacting RNAs (piRNAs) - a mouse testis perspective. *Biochemistry. Biokhimiia* **78**, 592-602, doi:10.1134/s0006297913060059 (2013).

53 Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science (New York, N.Y.)* **313**, 320-324, doi:10.1126/science.1129333 (2006).

54 Houwing, S. *et al.* A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**, 69-82, doi:10.1016/j.cell.2007.03.026 (2007).

55 Watanabe, T. *et al.* MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Developmental cell* **20**, 364-375, doi:10.1016/j.devcel.2011.01.005 (2011).

56 Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science (New York, N.Y.)* **316**, 744-747, doi:10.1126/science.1142612 (2007).

57 Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & development* **22**, 908-917, doi:10.1101/gad.1640708 (2008).

58 Castaneda, J. *et al.* Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. *The EMBO journal* **33**, 1999-2019, doi:10.15252/embj.201386855 (2014).

59 Soper, S. F. *et al.* Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Developmental cell* **15**, 285-297, doi:10.1016/j.devcel.2008.05.015 (2008).

60 Lim, A. K. *et al.* The nuage mediates retrotransposon silencing in mouse primordial ovarian follicles. *Development (Cambridge, England)* **140**, 3819-3825, doi:10.1242/dev.099184 (2013).

61 Kabayama, Y. *et al.* Roles of MIWI, MILI and PLD6 in small RNA regulation in mouse growing oocytes. *Nucleic acids research* **45**, 5387-5398, doi:10.1093/nar/gkx027 (2017).

62 Williams, Z. *et al.* Discovery and Characterization of piRNAs in the Human Fetal Ovary. *Cell reports* **13**, 854-863, doi:10.1016/j.celrep.2015.09.030 (2015).

63 Menke, D. B., Koubova, J. & Page, D. C. Sexual differentiation of germ cells in XX mouse gonads occurs in an anterior-to-posterior wave. *Developmental biology* **262**, 303-312, doi:10.1016/s0012-1606(03)00391-9 (2003).

64 Arora, R. *et al.* Meiotic onset is reliant on spatial distribution but independent of germ cell number in the mouse ovary. *Journal of cell science* **129**, 2493-2499, doi:10.1242/jcs.189910 (2016).

65 Page, S. L. & Hawley, R. S. The genetics and molecular biology of the synaptonemal complex. *Annual review of cell and developmental biology* **20**, 525-558, doi:10.1146/annurev.cellbio.19.111301.155141 (2004).

66 Keeney, S., Giroux, C. N. & Kleckner, N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375-384, doi:10.1016/s0092-8674(00)81876-0 (1997).

67 Keeney, S. Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome dynamics and stability* **2**, 81-123, doi:10.1007/7050_2007_026 (2008).

68 Carofiglio, F. *et al.* SPO11-independent DNA repair foci and their role in meiotic silencing. *PLoS genetics* **9**, e1003538, doi:10.1371/journal.pgen.1003538 (2013).

69 Malki, S., van der Heijden, G. W., O'Donnell, K. A., Martin, S. L. & Bortvin, A. A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. *Developmental cell* **29**, 521-533, doi:10.1016/j.devcel.2014.04.027 (2014).

70 Baudat, F., Manova, K., Yuen, J. P., Jasin, M. & Keeney, S. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Molecular cell* **6**, 989-998 (2000).

71 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).

72 Rinaldi, V. D., Bolcun-Filas, E., Kogo, H., Kurahashi, H. & Schimenti, J. C. The DNA Damage Checkpoint Eliminates Mouse Oocytes with Chromosome Synapsis Failure. *Molecular cell* **67**, 1026-1036.e1022, doi:10.1016/j.molcel.2017.07.027 (2017).

73 Carroll, J. & Marangos, P. The DNA damage response in mammalian oocytes. *Frontiers in genetics* **4**, 117, doi:10.3389/fgene.2013.00117 (2013).

74 Bolcun-Filas, E., Rinaldi, V. D., White, M. E. & Schimenti, J. C. Reversal of female infertility by Chk2 ablation reveals the oocyte DNA damage checkpoint pathway. *Science (New York, N.Y.)* **343**, 533-536, doi:10.1126/science.1247671 (2014).

75 Cloutier, J. M. *et al.* Histone H2AFX Links Meiotic Chromosome Asynapsis to Prophase I Oocyte Loss in Mammals. *PLoS genetics* **11**, e1005462, doi:10.1371/journal.pgen.1005462 (2015).

76 Greenspan, L. J., de Cuevas, M. & Matunis, E. Genetics of gonadal stem cell renewal. *Annual review of cell and developmental biology* **31**, 291-315, doi:10.1146/annurev-cellbio-100913-013344 (2015).

77 Lei, L. & Spradling, A. C. Female mice lack adult germ-line stem cells but sustain oogenesis using stable primordial follicles. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8585-8590, doi:10.1073/pnas.1306189110 (2013).

78 Zhang, H. *et al.* Experimental evidence showing that no mitotically active female germline progenitors exist in postnatal mouse ovaries. *Proc Natl Acad Sci U S A* **109**, 12580-12585, doi:10.1073/pnas.1206600109 (2012).

79 Hanna, C. B. & Hennebold, J. D. Ovarian germline stem cells: an unlimited source of oocytes? *Fertil Steril* **101**, 20-30, doi:10.1016/j.fertnstert.2013.11.009 (2014).

80 Pelosi, E., Simonsick, E., Forabosco, A., Garcia-Ortiz, J. E. & Schlessinger, D. Dynamics of the ovarian reserve and impact of genetic and epidemiological factors on age of menopause. *Biol Reprod* **92**, 130, doi:10.1095/biolreprod.114.127381 (2015).

81 Grive, K. J. & Freiman, R. N. The developmental origins of the mammalian ovarian reserve. *Development (Cambridge, England)* **142**, 2554-2563, doi:10.1242/dev.125211 (2015).

82 Pelosi, E., Forabosco, A. & Schlessinger, D. Genetics of the ovarian reserve. *Frontiers in genetics* **6**, 308, doi:10.3389/fgene.2015.00308 (2015).

83 Findlay, J. K., Hutt, K. J., Hickey, M. & Anderson, R. A. What is the "ovarian reserve"? *Fertility and sterility*, doi:10.1016/j.fertnstert.2014.10.037 (2014).

84 Baker, T. G. A Quantitative and Cytological Study of Germ Cells in Human Ovaries. *Proc R Soc Lond B Biol Sci* **158**, 417-433 (1963).

85 Kurilo, L. F. Oogenesis in antenatal development in man. *Human genetics* **57**, 86-92 (1981).

86 Baker, T. G. A quantitative and cytological study of oogenesis in the rhesus monkey. *Journal of anatomy* **100**, 761-776 (1966).

87 Beaumont, H. M. & Mandl, A. M. A quantitative and cytological study of oognonia and oocytes in the foetal and nonatal rat. *Proc R Soc Lond B Biol Sci* **155**, 557-579 (1962).

88 Burgoyne, P. S. & Baker, T. G. Perinatal oocyte loss in XO mice and its implications for the aetiology of gonadal dysgenesis in XO women. *J Reprod Fertil* **75**, 633-645 (1985).

89 Ioannou, J. M. Ooegenesis in the Guinea-Pig. *Journal of embryology and experimental morphology* **12**, 673-691 (1964).

90 Tilly, J. L. Commuting the death sentence: how oocytes strive to survive. *Nature reviews. Molecular cell biology* **2**, 838-848, doi:10.1038/35099086 (2001).

91 Hartshorne, G. M., Lyrakou, S., Hamoda, H., Oloto, E. & Ghafari, F. Oogenesis and cell death in human prenatal ovaries: what are the criteria for oocyte selection? *Molecular human reproduction* **15**, 805-819, doi:10.1093/molehr/gap055 (2009).

92 Pepling, M. E. & Spradling, A. C. Mouse ovarian germ cell cysts undergo programmed breakdown to form primordial follicles. *Developmental biology* **234**, 339-351, doi:10.1006/dbio.2001.0269 (2001).

93 Cox, R. T. & Spradling, A. C. A Balbiani body and the fusome mediate mitochondrial inheritance during Drosophila oogenesis. *Development (Cambridge, England)* **130**, 1579-1590, doi:10.1242/dev.00365 (2003).

94 Lei, L. & Spradling, A. C. Mouse primordial germ cells produce cysts that partially fragment prior to meiosis. *Development (Cambridge, England)* **140**, 2075-2081, doi:10.1242/dev.093864 (2013).

95 Greenbaum, M. P. *et al.* TEX14 is essential for intercellular bridges and fertility in male mice. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4982-4987, doi:10.1073/pnas.0505123103 (2006).

96 Lei, L. & Spradling, A. C. Mouse oocytes differentiate through organelle enrichment from sister cyst germ cells. *Science (New York, N.Y.)* **352**, 95-99, doi:10.1126/science.aad2156 (2016).

97 Greenbaum, M. P., Iwamori, N., Agno, J. E. & Matzuk, M. M. Mouse TEX14 is required for embryonic germ cell intercellular bridges but not female fertility. *Biology of reproduction* **80**, 449-457, doi:10.1095/biolreprod.108.070649 (2009).

98 Giehl, K. M. Trophic dependencies of rodent corticospinal neurons. *Reviews in the neurosciences* **12**, 79-94, doi:10.1515/revneuro.2001.12.1.79 (2001).

99 Morita, Y. *et al.* Caspase-2 deficiency prevents programmed germ cell death resulting from cytokine insufficiency but not meiotic defects caused by loss of ataxia telangiectasia-mutated (Atm) gene function. *Cell death and differentiation* **8**, 614-620, doi:10.1038/sj.cdd.4400845 (2001).

100 Morita, Y. *et al.* Requirement for phosphatidylinositol-3'-kinase in cytokine-mediated germ cell survival during fetal oogenesis in the mouse. *Endocrinology* **140**, 941-949, doi:10.1210/endo.140.2.6539 (1999).

101 Ene, A. C., Park, S., Edelmann, W. & Taketo, T. Caspase 9 is constitutively activated in mouse oocytes and plays a key role in oocyte elimination during meiotic prophase progression. *Developmental biology* **377**, 213-223, doi:10.1016/j.ydbio.2013.01.027 (2013).

102 Liu, X., Castle, V. & Taketo, T. Interplay between Caspase 9 and X-linked Inhibitor of Apoptosis Protein (XIAP) in the oocyte elimination during fetal mouse development. *Cell Death & Disease* **10**, 790, doi:10.1038/s41419-019-2019-x (2019).

103 Qiao, H. *et al.* Impeding DNA Break Repair Enables Oocyte Quality Control. *Molecular cell* **72**, 211-221.e213, doi:10.1016/j.molcel.2018.08.031 (2018).

104 O'Donnell, K. A., An, W., Schrum, C. T., Wheelan, S. J. & Boeke, J. D. Controlled insertional mutagenesis using a LINE-1 (ORFeus) gene-trap mouse model. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E2706-2713, doi:10.1073/pnas.1302504110 (2013).

105 Dai, L., Huang, Q. & Boeke, J. D. Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. *BMC Biochem* **12**, 18, doi:10.1186/1471-2091-12-18 (2011).

106 Jones, R. B. *et al.* Nucleoside analogue reverse transcriptase inhibitors differentially inhibit human LINE-1 retrotransposition. *PloS one* **3**, e1547, doi:10.1371/journal.pone.0001547 (2008).

107 Dai, L., Huang, Q. & Boeke, J. D. Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. *BMC biochemistry* **12**, 18, doi:10.1186/1471-2091-12-18 (2011).

108 Banuelos-Sanchez, G. *et al.* Synthesis and Characterization of Specific Reverse Transcriptase Inhibitors for Mammalian LINE-1 Retrotransposons. *Cell chemical biology*, doi:10.1016/j.chembiol.2019.04.010 (2019).

109 Sharpe, A. H., Jaenisch, R. & Ruprecht, R. M. Retroviruses and mouse embryos: a rapid model for neurovirulence and transplacental antiviral therapy. *Science (New York, N.Y.)* **236**, 1671-1674, doi:10.1126/science.3037694 (1987).

110 Richardson, S. R. *et al.* Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome research* **27**, 1395-1405, doi:10.1101/gr.219022.116 (2017).

111 Crossley, M. P., Bocek, M. & Cimprich, K. A. R-Loops as Cellular Regulators and Genomic Threats. *Molecular cell* **73**, 398-411, doi:10.1016/j.molcel.2019.01.024 (2019).

112 Pokatayev, V. *et al.* RNase H2 catalytic core Aicardi-Goutieres syndrome-related mutant invokes cGAS-STING innate immune-sensing pathway in mice. *The Journal of experimental medicine* **213**, 329-336, doi:10.1084/jem.20151464 (2016).

113 Thomas, C. A. *et al.* Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation. *Cell stem cell* **21**, 319-331.e318, doi:10.1016/j.stem.2017.07.009 (2017).

114 De Cecco, M. *et al.* L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73-78, doi:10.1038/s41586-018-0784-9 (2019).

115 Simon, M. *et al.* LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. *Cell metabolism* **29**, 871-885.e875, doi:10.1016/j.cmet.2019.02.014 (2019).

116 DeFalco, T., Bhattacharya, I., Williams, A. V., Sams, D. M. & Capel, B. Yolk-sac-derived macrophages regulate fetal testis vascularization and morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E2384-2393, doi:10.1073/pnas.1400057111 (2014).

117 Ono, Y. *et al.* CD11c+ M1-like macrophages (MPhis) but not CD206+ M2-like MPhi are involved in folliculogenesis in mice ovary. *Scientific reports* **8**, 8171, doi:10.1038/s41598-018-25837-3 (2018).

118 Zannini, L., Delia, D. & Buscemi, G. CHK2 kinase in the DNA damage response and beyond. *Journal of molecular cell biology* **6**, 442-457, doi:10.1093/jmcb/mju045 (2014).

119 Hirao, A. *et al.* Chk2 is a tumor suppressor that regulates apoptosis in both an ataxia telangiectasia mutated (ATM)-dependent and an ATM-independent manner. *Molecular and cellular biology* **22**, 6521-6532, doi:10.1128/mcb.22.18.6521-6532.2002 (2002).

120 Choi, J., Hwang, S. Y. & Ahn, K. Interplay between RNASEH2 and MOV10 controls LINE-1 retrotransposition. *Nucleic acids research* **46**, 1912-1926, doi:10.1093/nar/gkx1312 (2018).

121 Benitez-Guijarro, M. *et al.* RNase H2, mutated in Aicardi-Goutieres syndrome, promotes LINE-1 retrotransposition. *The EMBO journal* **37**, doi:10.15252/embj.201798506 (2018).

122 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

123 Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* **31**, 46-53, doi:10.1038/nbt.2450 (2013).

124 cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. v. 2.29.0 (2019).

125 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

126 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).

127 Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science (New York, N.Y.)* **351**, 1083-1087, doi:10.1126/science.aad5497 (2016).

128 Sun, X. *et al.* Role of decay-accelerating factor in regulating complement activation on the erythrocyte surface as revealed by gene targeting. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 628-633, doi:10.1073/pnas.96.2.628 (1999).

129 Lee, A. S. *et al.* Rare mutations in the complement regulatory gene CSMD1 are associated with male and female infertility. *Nature communications* **10**, 4626, doi:10.1038/s41467-019-12522-w (2019).

130 Fan, X. *et al.* Single-cell reconstruction of follicular remodeling in the human adult ovary. *Nature communications* **10**, 3164, doi:10.1038/s41467-019-11036-9 (2019).

131 Ghafari, F., Gutierrez, C. G. & Hartshorne, G. M. Apoptosis in mouse fetal and neonatal oocytes during meiotic prophase one. *BMC developmental biology* **7**, 87, doi:10.1186/1471-213x-7-87 (2007).

132 Livera, G. *et al.* p63 null mutation protects mouse oocytes from radio-induced apoptosis. *Reproduction (Cambridge, England)* **135**, 3-12, doi:10.1530/rep-07-0054 (2008).

133 Han, B. W., Wang, W., Zamore, P. D. & Weng, Z. piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics (Oxford, England)* **31**, 593-595, doi:10.1093/bioinformatics/btu647 (2015).

134 Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203-207, doi:nature04916 [pii]

10.1038/nature04916 (2006).

135 Aravin, A. A. & Hannon, G. J. Small RNA silencing pathways in germ and stem cells. *Cold Spring Harb Symp Quant Biol* **73**, 283-290, doi:sqb.2008.73.058 [pii]

10.1101/sqb.2008.73.058 (2008).

136 Flemr, M. *et al.* A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. *Cell* **155**, 807-816, doi:10.1016/j.cell.2013.10.001 (2013).

137 Wolf, G. *et al.* The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes & development* **29**, 538-554, doi:10.1101/gad.252767.114 (2015).

138 Yang, P., Wang, Y. & Macfarlan, T. S. The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. *Trends Genet* **33**, 871-881, doi:10.1016/j.tig.2017.08.006 (2017).

139 Jamieson-Lucy, A. & Mullins, M. C. The vertebrate Balbiani body, germ plasm, and oocyte polarity. *Current topics in developmental biology* **135**, 1-34, doi:10.1016/bs.ctdb.2019.04.003 (2019).

140 Pepling, M. E., Wilhelm, J. E., O'Hara, A. L., Gephardt, G. W. & Spradling, A. C. Mouse oocytes within germ cell cysts and primordial follicles contain a Balbiani body. *Proc Natl Acad Sci U S A* **104**, 187-192, doi:0609923104 [pii]

10.1073/pnas.0609923104 (2007).

141 Kloc, M., Bilinski, S. & Etkin, L. D. The Balbiani body and germ cell determinants: 150 years later. *Curr Top Dev Biol* **59**, 1-36, doi:10.1016/S0070-2153(04)59001-4 (2004).

142 Chao, J. T. & Niwa, M. Nursing the Follicles. *Developmental cell* **37**, 7-8, doi:10.1016/j.devcel.2016.03.019 (2016).

143 Thomas, F. H. & Vanderhyden, B. C. Oocyte-granulosa cell interactions during mouse follicular development: regulation of kit ligand expression and its role in oocyte growth. *Reproductive biology and endocrinology : RB&E* **4**, 19, doi:10.1186/1477-7827-4-19 (2006).

144 Barnett, K. R., Schilling, C., Greenfeld, C. R., Tomic, D. & Flaws, J. A. Ovarian follicle development and transgenic mouse models. *Human reproduction update* **12**, 537-555, doi:10.1093/humupd/dml022 (2006).

145 Hunter, N. Oocyte Quality Control: Causes, Mechanisms, and Consequences. *Cold Spring Harb Symp Quant Biol* **82**, 235-247, doi:10.1101/sqb.2017.82.035394 (2017).

146 Matova, N. & Cooley, L. Comparative aspects of animal oogenesis. *Developmental biology* **231**, 291-320, doi:10.1006/dbio.2000.0120 (2001).

147 Hikabe, O. *et al.* Reconstitution in vitro of the entire cycle of the mouse female germ line. *Nature* **539**, 299-303, doi:10.1038/nature20104 (2016).

148 Dokshin, G. A., Baltus, A. E., Eppig, J. J. & Page, D. C. Oocyte differentiation is genetically dissociable from meiosis in mice. *Nature genetics* **45**, 877-883, doi:10.1038/ng.2672 (2013).

149 Teissandier, A., Servant, N., Barillot, E. & Bourc'his, D. Tools and best practices for retrotransposon analysis using high-throughput sequencing data. *Mobile DNA* **10**, 52, doi:10.1186/s13100-019-0192-1 (2019).

150 Bendall, M. L. *et al.* Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS computational biology* **15**, e1006453, doi:10.1371/journal.pcbi.1006453 (2019).

151 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

152 Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics (Oxford, England)* **31**, 3593-3599, doi:10.1093/bioinformatics/btv422 (2015).

153 Penzkofer, T. *et al.* L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic acids research* **45**, D68-d73, doi:10.1093/nar/gkw925 (2017).

154 Morrish, T. A. *et al.* Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* **446**, 208-212, doi:10.1038/nature05560 (2007).

155 Di Giacomo, M. *et al.* Distinct DNA-damage-dependent and -independent responses drive the loss of oocytes in recombination-defective mouse mutants. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 737-742, doi:10.1073/pnas.0406212102 (2005).

156 Malki, S., Tharp, M. E. & Bortvin, A. A Whole-Mount Approach for Accurate Quantitative and Spatial Assessment of Fetal Oocyte Dynamics in Mice. *Biology of reproduction* **93**, 113, doi:10.1095/biolreprod.115.132118 (2015).

157 Tanaka, H. *et al.* A germ cell-specific nuclear antigen recognized by a monoclonal antibody raised against mouse testicular germ cells. *Int J Androl* **20**, 361-366 (1997).

158 Wojtasz, L., Daniel, K. & Toth, A. Fluorescence activated cell sorting of live female germ cells and somatic cells of the mouse fetal gonad based on forward and side scattering. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **75**, 547-553, doi:10.1002/cyto.a.20729 (2009).

159 Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis (New York, N.Y. : 2000)* **35**, 88-93, doi:10.1002/gene.10168 (2003).

BIOGRAPHICAL SKETCH

NAME OF APPLICANT: Marla Tharp

POSITION TITLE: Predoctoral Fellow

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	YEARS	FIELD OF STUDY
University of Illinois at Urbana-Champaign, Urbana, IL	B.S.	2010-2014	Molecular and Cellular Biology
Johns Hopkins University/Carnegie Institution for Science, Baltimore, MD	Ph.D.	2014-2020	Cell, Molecular, and Developmental biology and Biophysics

A. Personal Statement

As a whole, my research experiences have investigated questions of developmental genetics and evolution in a number of model systems, especially related to germ cell quality, embryogenesis, and regeneration. Undergraduate research in the Newmark Laboratory at the University of Illinois at Urbana-Champaign provided me with a strong foundation in cell and developmental biology, with a focus on stem cells and reproduction, and experimental techniques in molecular and cellular biology. There, I designed and executed projects in two different flatworm species, culminating in data presented at meetings and in multiple research papers, including my own first author manuscript **(1, 2)**. I went on to do my graduate studies in the Johns Hopkins University CMDB program, which fostered my long-term career goal of becoming an academic researcher by providing opportunities to mentor students in the classroom and the lab, develop background knowledge in diverse areas of biology through courses, weekly seminars, and four laboratory rotations prior to choosing a thesis lab **(3)**. Additionally, during one summer of my graduate studies, I had the opportunity to participate in the Embryology Course at the Marine Biological Labs at Woods Hole that provided an unparalleled experience to learn and explore a large number of experimental model organisms and techniques in developmental biology **(4)**. My thesis work was conducted in the lab of Dr. Alex Bortvin at the Carnegie Institution for Science, who has pioneered studies on the role of retrotransposons in germ cell quality and survival, generating many of the tools and techniques that are utilized in my proposed project. In the Bortvin lab, I investigated how transposable element activity can lead to massive oocyte demise during fetal development in mice **(5)**. I found that retrotransposon LINE-1 cytotoxicity in fetal oocytes is conferred by two different mechanisms: the first involving persistent DNA damage and defects in meiosis, and the second related to accumulation of reverse transcription intermediates and the innate immune system. My findings have important implications for understanding the establishment of the finite ovarian reserve in humans, and how germ cell quality control is critical for subsequent generations. My expertise includes computational skills to analyze big data projects such as R programming, techniques of mouse genetics and ovary histological analysis, confocal imaging, fluorescence activated cell sorting, and sequencing experiments that include bulk and single-cell mRNA-Seq, small RNA-Seq, and bisulfite-Seq.

B. Positions and Honors

Positions and Employment

2014-2020	Predoc fellow. Johns Hopkins University/Carnegie Institution for Science. Lab of Dr. Alex Bortvin.
Summer 2018	Student researcher, Embryology Course. Marine Biological Labs at Woods Hole, Massachusetts.
2011-2014	Undergraduate researcher. University of Illinois at Urbana-Champaign. Lab of Dr. Phil Newmark.

Academic and Professional Honors

2016-2020	Ruth L. Kirschstein National Pre-Doctoral Research Service Award, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH
2019	Junior Investigator Award, FASEB Mobile DNA Conference
2018	Burroughs Wellcome Fund and Surdna Foundation Scholarships for Embryology: Concepts and Techniques in Modern Developmental Biology Course, Marine Biological Laboratory in Woods Hole
2016	Cell, Molecular, Developmental Biology and Biophysics Departmental Retreat Poster Award, Johns Hopkins University
2014	Thomas Hunt Morgan Fellowship, Incoming Graduate Student Award, Johns Hopkins University
2014	Senior Thesis with Highest Distinction, University of Illinois at Urbana-Champaign
2014	Outstanding Undergraduate Research Award, University of Illinois at Urbana-Champaign
2013	Procter and Gamble Award for Undergraduate Research, University of Illinois at Urbana-Champaign

---

C. Contributions to Science

1. During my undergraduate studies, I conducted four years of research in Dr. Phil Newmark's Lab at the University of Illinois at Urbana-Champaign. My earliest contribution focused on stem cell biology of the parasitic flatworm *Schistosoma mansoni*. These parasites cause schistosomaisis, a neglected tropical disease affection over 200 million people worldwide. My role in the project was optimizing RNA interference techniques to disrupt genes of interest in these parasites. We performed an RNAi screen of genes enriched in schistosome stem cells and found that EdU positive dividing cells are lost upon treatment. These experiments have great implications for treating schistosomaisis because by eliminating their reservoir of stem cells, these parasites may die much faster in the host.

> Collins JJ 3rd, Wang B, Lambrus BG, **Tharp ME,** Iyer H, Newmark PA. Adult somatic stem cells in the human parasite Schistosoma mansoni. Nature. 2013 Feb 28;494(7438):476-9. doi: 10.1038/nature11924. Epub 2013 Feb 20. PubMed PMID: 23426263; PubMed Central PMCID: PMC3586782.

2. Also in Dr. Newmark's Lab, I became interested in organ regeneration, stem cell, and germ cell biology using the planarian *Schmidtea mediterranea* as a model organism. My primary contribution was investigating systemic hormone signaling in planarian sexual development. *S. mediterranea* exists in sexual and asexual strains, prompting an evolutionary question as to how this sexuality is acquired or lost.

I hypothesized that steroid hormone signaling may be an answer. In analyzing RNA sequencing data of genes enriched in sexual versus asexual strains of *S. mediterranea*, I chose a list of sexual-specific steroid hormone receptors to knockdown using RNA interference. Disruption of one of these genes, *nhr-1,* showed a complete loss of somatic reproductive organs as well as differentiated germ cells. Interestingly, this receptor was expressed in the somatic reproductive organs alone, suggesting that germ cells require cell non-autonomous hormone signals from these organs for their development. This work identified *nhr-1* as a key component for planarian reproductive development and a potential sexualizing agent.

> **Tharp ME,** Collins JJ 3rd, Newmark PA. A lophotrochozoan-specific nuclear hormone receptor is required for reproductive system development in the planarian. Dev Biol. 2014 Dec 1;396(1):150-7. doi: 10.1016/j.ydbio.2014.09.024. Epub 2014 Sep 30. PubMed PMID: 25278423; PubMed Central PMCID: PMC4253560.

3. My studies of nuclear hormone receptor signaling in germ cell development led me to do a graduate school rotation project to study the biophysical interactions of the glucocortiocoid receptor in Dr. Vince Hilser's lab. I contributed to a study that determined the role of co-factor TSG101 and DNA binding interactions in glucocorticoid receptor function.

> White JT, Rives J, **Tharp ME,** Wrabl JO, Thompson EB, Hilser VJ. Tumor susceptibility gene-101 regulates glucocorticoid receptor through disorder-mediated allostery. BioRxiv. 2020. doi: 10.1101/2020.02.02.931485.

4. During my graduate research, I was selected to attend the Embryology Course at Marine Biological Labs in Woods Hole. There, I conducted exploratory experiments in a vast number of model organisms using both classic and cutting edge developmental biology techniques. One experiment I performed with colleagues from Dr. Bob Goldstein's lab optimized a protocol for staining and fluorescence confocal imaging of live tardigrade *Hypsibius exemplaris*, or water bear, embryos. We collected and soaked tardigrade embryos in the lipophilic fluorescent dye, FM 4-64, which stains cell membranes and vesicles. This allowed visualization and confocal live imaging of the early cleavages of tardigrade embryonic development.

> McGreevy KM, Heikes KL, Kult S, **Tharp ME,** Goldstein B. Fluorescent Cell Staining Methods for Living *Hypsibius exemplaris* Embryos. Cold Spring Harb Protoc. 2018 Nov 1;2018(11). doi: 10.1101/pdb.prot106021. PubMed PMID: 30385676.

5. Due to my interest in developmental biology, germ cells, and evolution, I joined Dr. Alex Bortvin's lab at the Carnegie Institution for Science for my graduate research. My research investigated fetal oocyte attrition (FOA) in mice, a process that selectively eliminates up to 80% of developing oocytes before birth in mammals, and thus significantly influences the quality and quantity of oocytes in the finite ovarian reserve. Specifically, I focused on the role of the retrotransposon LINE1 (L1), and how activation of L1 as a consequence of primordial germ cell epigenetic reprogramming triggers FOA. My contributions include elucidating two mechanisms of FOA that are attributed to L1 activity. These include elimination by way of the DNA damage checkpoint mediated by checkpoint kinase 2 (CHK2) that is activated in late meiotic prophase I to promote apoptosis of oocytes with excess DNA damage and meiotic defects. Second, the complement system of innate immunity that is responsive to L1 reverse transcriptase activity and attenuated using the reverse transcriptase inhibitor AZT. I also established a method to prevent FOA by treating CHK2 mutant mice with AZT that allowed me to determine that FOA was non-essential for oogenesis and fertility using genome-wide sequencing and histological approaches.

a. Malki S, **Tharp ME,** Bortvin A. A Whole-Mount Approach for Accurate Quantitative and Spatial Assessment of Fetal Oocyte Dynamics in Mice. Biol Reprod. 2015 Nov;93(5):113. doi: 10.1095/biolreprod.115.132118. Epub 2015 Sep 30. PubMed PMID: 26423126
b. **Tharp ME,** Bortvin A. DjPiwiB: A Rich Nuclear Inheritance for Descendants of Planarian Stem Cells. Dev Cell. 2016 May 9;37(3):204-6. doi: 10.1016/j.devcel.2016.04.022. PubMed PMID: 27165550.
c. **Tharp ME,** Bortvin A. De novo DNA Methylation: Who's Your DADdy?. Trends Genet. 2019 Nov;35(11):785-787. doi: 10.1016/j.tig.2019.09.001. Epub 2019 Oct 6. PubMed PMID: 31597610.
d. **Tharp ME,** Malki S, Bortvin A. Maximizing the ovarian reserve in mice by evading LINE-1 genotoxicity. Nat Commun. 2020 Jan 16;11(1):330. doi: 10.1038/s41467-019-14055-8. PubMed PMID: 31949138; PubMed Central PMCID: PMC6965193.

Complete List of Published Work in MyBibliography:

https://www.ncbi.nlm.nih.gov/myncbi/marla.tharp.1/bibliography/public/