UNDERSTANDING TRANSCRIPTIONAL RESPONSES TO ENVIRONMENTAL AND DEVELOPMENTAL CUES IN MODEL ORGANISMS USING FULL-LENGTH SINGLE CELL RNA-SEQ

by

Yangqi Su

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Bioinformatics and Computational Biology

Charlotte

2024

Approved by:

_____
Dr. Zhengchang Su

_____
Dr. Jun-tao Guo

_____
Dr. Way Sung

_____
Dr. Shan Yan

ABSTRACT

YANGQI SU. Understanding Transcriptional Responses to Environmental and Developmental Cues in Model Organisms Using Full-Length Single Cell RNA-Seq (Under the direction of DR. ZHENGCHANG SU)

Recent advancements in single-cell RNA sequencing have revolutionized our understanding of gene expression regulation under various biological contexts, providing higher resolution and system-level insights compared to traditional bulk RNA sequencing methods. In this dissertation, we utilize single cell RNA-seq (scRNA-seq) along with various statistical tools to unveil the transcriptomic landscape of four model organisms during development and under different stress conditions. First, we sequenced yeast cells under three stress treatments (hypotonic condition, glucose starvation and amino acid starvation) using a full-length single-cell RNA-Seq method. We found that although single cells from the same treatment showed varying degrees of uniformity, technical noise and batch effects can confound results significantly. However, upon careful selection of samples to reduce technical artifacts and account for batch-effects, we were able to capture distinct transcriptomic signatures for different stress conditions as well as identify putative regulatory relationships between transcription factors and target genes. Our results show that a full-length single-cell based transcriptomic analysis provide a clearer picture of yeast stress response over bulk cell population-based transcriptomic methods.

Second, we present a transcriptomic level analysis of oogenesis in *C. elegans* hermaphrodites. We dissected a hermaphrodite gonad into seven sections corresponding to the mitotic distal region, the pachytene, the diplotene, the early diakinesis region and the 3 most proximal oocytes, and deeply sequenced the transcriptome of each of them along with

that of the fertilized egg using a single-cell RNA-seq protocol. We identified specific gene expression events as well as gene splicing events in finer detail along the oocyte germline and provided novel insights into underlying mechanisms of oogenesis. Through careful review of relevant research literature coupled with patterns observed in our analysis, we attempt to delineate transcripts that may serve functions in the interaction between the germline and cells of the somatic gonad. These results expand our knowledge of the transcriptomic space of the *C. elegans* germline and lay a foundation on which future studies of the germline can be based upon. Lastly, we profiled mature oocytes and 1-cell zygotes of mice and rats to uncover elusive transcriptomic dynamics in the maternal to zygote transition. We confirm the existence of early gene expression in the mouse zygotic while revealing a similar chain of events occurring in the rat zygote. We observe an increase in nascent transcription in both species. Moreover, we find subtle but pervasive signals of differential splicing of genes related to key early zygotic activities occurring in both species. Meanwhile, we find distinct profiles of alternative polyadenylation between zygotes and oocytes in both species, specifically in genes related to major processes within the zygote. Finally, although a more dynamic transcriptomic landscape exists in the mice zygote, the rat zygote also displays similar transcriptomic features, suggesting that minor zygotic activation in rat occurs earlier than originally thought.

ACKNOWLEDGEMENTS

I would sincerely thank my advisor Dr. Zhengchang Su for his professional and patient guidance in my PhD project. He gave many constructional ideas and instructions on the project so this thesis can be realized successfully. I would thank my committee members, Dr. Juntao Guo, Dr. Way Sung and Dr. Shan Yan for their insightful comments and suggestions. I would like to thank the members in my lab for their discussions on my project, as well as Dr. Xiaoxia Cui, Jane Kouranova and Jonathan Shea for providing experimental support and generation of our data. I would thank my parents for their support. Finally, I would thank the Graduate Assistant Support Plan (GASP) for their financial support. I would also like to thank all staff members in the Department of Bioinformatics and Genomics and the international student and scholar office (ISSO) for their supportive help.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

The various characteristics of single cells have been studied extensively through a combination of cell isolation, high throughput multiplexing, amplification, and next generation sequencing, leading to an explosion of techniques to quantify the heterogeneous gene expression, genomic alterations, epigenomic modifications, and proteomic fluctuations of single cells (Buenrostro et al., 2015; Kolodziejczyk, Kim, Svensson, Marioni, & Teichmann, 2015; Shalek et al., 2013). In the forefront of these advancements is single cell RNA sequencing (scRNA-seq). The rapid development of high throughput RNA-seq technology has provided a better platform than microarray to study the transcriptomic profiles of cells, providing researchers the opportunity to quantify gene expression genome-wide in a single assay with higher resolution, better dynamic range, and lower technical variation. However, bulk RNA-seq, which pools together hundreds of thousands of or more cells, masks the hidden variations and intricacies that occur at the single cell level. scRNA-seq overcome these limitations, thereby revealing transcriptional mechanisms of cell differentiation, fate plasticity, and disease (Haque, Engel, Teichmann, & Lonnberg, 2017).

Various scRNA-seq protocols have been developed in the past decade, each with its strength and weaknesses. Roughly, the methodologies can be divided based on sequencing the full length or just the 3'-end of the transcripts (Lieberman et al., 2021). Earlier methods such as SMART-seq and SMART-seq2 utilizes the full transcript and focuses on detecting as many transcripts as possible (Picelli, Bjorklund, et al., 2013; Picelli et al., 2014). However, these protocols rely on relative measurements that might affect the

accuracy and precision of downstream analysis. Furthermore, the manual preparation of samples is limited for scalability and often introduces significant technical variation. Later methods such as Drop-seq and Chromium 10x only sequence the 3'-ends of transcripts and incorporate UMIs to reduce amplification bias and improve cell capture efficiency using microfluidics methods (Hashimshony, Wagner, Sher, & Yanai, 2012; Macosko et al., 2015; Zheng et al., 2017). These newer methods can capture tens to hundreds of thousands of cells at once with low technical variation, providing an opportunity to characterize all the cells of an organism. On the other hand, they suffer from lowered gene detection rates as well as the inability to identify alternative splicing and novel isoforms.

Although there is a prevalence of high sample/cell throughput sequencing in recent single cell studies, full-length based sequencing enables capturing of many more aspects of the transcriptome than 3'end methodologies. In addition to higher reads per sample/cell, which translates to more genes detected per sample/cell, full-length methods allow for the quantification of allelic specific expression, alternative splicing, poly-adenylation, etc. The high degree of read coverage in full-length scRNA-seq data have been shown to resemble bulk RNA-seq data (X. Wang, He, Zhang, Ren, & Zhang, 2021).

In chapter 1 of this dissertation, we used full-length scRNA-seq techniques to better understand transcriptional responses to environmental cues in yeast cells and to developmental cues in C. elegans germline as well as during the formation of zygotes of C. elegans, mice and rats. We show that analysis methodologies developed for bulk RNA-seq data perform well for scRNA-seq data. The high depths of sequencing in most of our

samples along with the usage of full-length protocols allow us to have a detailed and meaningful look into various biological mechanisms.

In chapter 2, we utilize scRNA-seq to study the stress response of yeast. The yeast stress response has been a topic of study in the past, and while there have been a plethora of microarray and proteomics-based research in this area, scRNA-seq provides us with a new way to obtain possible novel insights. Specifically, scRNA-seq has allowed us to shed new light on how yeast responds to glucose starvation, amino acid starvation and hypo-osmotic shock with scRNA-seq.

In chapter 3, we attempt to delineate the mechanism of oogenesis and fertilization in *C. elegans* using scRNA-seq. By dissecting the gonads of adult *C. elegans* hermaphrodites in to predefined sections, we collect and sequence cells along each section of the gonad, allowing us to take a close examination of oocyte maturation until zygote formation. Here, we used scRNA-seq to find new markers and novel transcriptional regulatory mechanisms underlying oogenesis and oocyte to zygote transition in *C. elegans*

In chapter 4, we collect and perform scRNA-seq in mature oocytes and zygotes from mice and rats to determine what transcriptome changes occur during oocyte fertilization. Through differential expression analysis we characterize relevant transcriptional markers of maternal to zygote transition (MZT). Furthermore, we find splicing and polyadenylation changes that may contribute to the transition process. Finally, we identify orthologous genes between rats and mice to find common and different mechanisms that underly MZT.

CHAPTER 2 TRANSCRIPTOMIC CHANGES IN SINGLE YEAST CELLS UNDER VARIOUS STRESS

2.1 Background

The ability to adapt to a changing environment is crucial to survival of individual cells, and this is even more evident for single celled organisms (Causton et al., 2001; Gasch et al., 2000; Price et al., 2001). A rapid change in a cell's surrounding induces stress, which requires the cell to activate complex mechanisms of sensing and signal transduction to adapt (Bahn et al., 2007; Rodriguez, Snoek, De Bono, & Kammenga, 2013; Zaman, Lippman, Zhao, & Broach, 2008). Such mechanisms eventually lead to the expression of genes and proteins, which can often be specific to the type of stress encountered. At the same time, organisms also develop general responses regardless of the type of stress encountered to better deal with a constantly changing environment (Gasch & Werner-Washburne, 2002). The study of stress response in various organisms has led to a better understanding of some of the fundamental aspects of cellular biology (Gasch et al., 2000; Girardot, Monnier, & Tricoire, 2004; Price et al., 2001; Rodriguez et al., 2013).

One such model organism that has been studied extensively is the budding yeast *Saccharomyces cerevisiae*. The budding yeast is a single-cell organism that faces a constantly changing environment when living freely in nature, often having to deal with multiple types of stress at the same time. For many decades, the responses of budding yeast to different environmental perturbations have been studied systematically on different levels, leading to the discovery and understanding of many pathways of the organism(Causton et al., 2001; Gasch et al., 2000; Stefan Hohmann & Mager, 2007;

Morano, Grant, & Moye-Rowley, 2012). However, while yeast may be the most well-studied eukaryotic organism, we are still far from being able to completely model this organism's response to stress. The advent of microarray technology meant simultaneous profiling of thousands of genes of yeast was possible (Lashkari et al., 1997). These developments led to the early analysis of the transcriptomic responses of yeast to different environmental changes, resulting in the discovery of a set of roughly 900 genes, termed the yeast environmental stress response (ESR) genes, which were activated as a general response to multiple types of stress (Gasch et al., 2000). Up-regulated ESR genes were found to be regulated by transcription factors (TFs) Msn2p and Msn4p and related directly to mitigation of stress, while down-regulated ESR genes were found to be involved in ribosomal biogenesis and protein synthesis (Causton et al., 2001; Martinez-Pastor et al., 1996; Schmitt & McEntee, 1996). This general response was thought to be crucial for the survival of yeast cells in preparation for changes in the environment (Gasch et al., 2000).

The yeast response to amino-acid starvation has been shown to be mediated through the general amino acid control (GAAC) pathway (Natarajan et al., 2001), and more specifically through the TF Gcn4p and protein kinase Gcn2p. Gcn2p phosphorylates translation initiation factor Eif2p, thus inhibiting overall translation rates (Dever et al., 1992). Though the abundance of Gcn4p is controlled at the translational level, phosphorylation of Eif2p increases the level of Gcn4p via a mechanism involving delayed ribosomal re-initiation and inhibitory upstream ORFs in the 5' region of the *GCN4* gene (Hinnebusch, 1997), thereby Gcn4p activates the promoters of genes that harbor a GCN response element (Arndt & Fink, 1986). Early transcriptomic studies showed that amino-

acid starvation down-regulated genes related to growth and ribosome biogenesis, while upregulating genes involved in amino-acid biosynthesis, cellular redox reaction, carbohydrate metabolism, cell wall modification, protein folding and degradation, DNA damage repair, fatty acid metabolism, metabolite transport, and autophagy(M. H. Jia et al., 2000).

The yeast has also been studied extensively for responses to changes in carbon sources (16-19). Though yeast prefers glucose as a carbon source during fermentative growth, it also can utilize other carbon sources as alternatives (Rolland, Winderickx, & Thevelein, 2002). During growth in glucose-rich cultures, genes in pathways for utilizing alternative carbon sources, such as galactose, maltose and sucrose, are repressed through a glucose sensitive repressor Mig1p(Klein, Olsson, & Nielsen, 1998). It was found that glucose starvation upregulated genes were involved in oxidative phosphorylation and the TCA cycle, and some of them encode high-affinity glucose transporters (J. Wu, Zhang, Hayes, Panoutsopoulou, & Oliver, 2004). At the same time, glucose starvation results in a drastic reduction in transcription rates and degradation of mRNA as well as almost complete inhibition of translational machinery (Ashe, De Long, & Sachs, 2000; Jona, Choder, & Gileadi, 2000). Another study profiling ribosome of yeast under glucose starvation noted that while overall protein synthesis was reduced, transcription of many stress-response and glucose-repressed genes was increased (Jona et al., 2000; Klein et al., 1998).

As a single-celled organism, yeast in nature may constantly experience sudden changes in surrounding osmolarity. The adaptation of yeast to hyperosmotic stress has been

studied extensively, whereby a sudden increase in osmolarity will cause the yeast cell to shrink and the high osmolarity glycerol response pathway is activated (S. Hohmann, 2009). Less is known about an osmolarity downshift, when there is a rapid influx of water, leading to increase of cell size and turgor pressure, during which the cell wall plays a vital role in preventing the cell from bursting (Smits, Kapteyn, van den Ende, & Klis, 1999). This process initiates the cell integrity pathway (CWI). Glycerol export is mediated via the Fps1p transporter. An influx of calcium ions also occurs (Batiza, Schulz, & Masson, 1996), which results in the activation of the TF Crz1p. The osmolarity sensor Sln1p, the phosphotransferase Ypd1p, and the TF Skn7p form a phosphor-relay system that activates Skn7p by phosphorylation upon cell swelling (S. Li et al., 2002; Tao, Deschenes, & Fassler, 1999). Upon activation, Skn7p activates the transcription of genes related to cell wall biogenesis (S. Li et al., 2002; Tao et al., 1999). Micro-array transcriptomic analysis revealed a reversal of the gene expression response during hyper-osmolarity stress but failed to find a distinct pattern of gene expression during hypotonic stress (Gasch et al., 2000).

In the past decade, the rapid development of high throughput RNA-seq technology has provided a better platform than microarray to study the transcriptomic profiles of organisms, providing the ability to quantify gene expression genome-wide in a single assay with higher resolution, larger dynamic ranges and lower technical variation(Z. Wang, Gerstein, & Snyder, 2009; W. Zhang et al., 2015; Zhao, Fung-Leung, Bittner, Ngo, & Liu, 2014). Moreover, the advent of single-cell RNA-seq (scRNA-seq) methods has provided an unprecedented opportunity to study the response of individual cells of an

organism(Shalek et al., 2014). This is even more evident in the case for the budding yeast, which is itself a single celled organism. Utilizing scRNA-seq, it may now be possible to delve deeper into the intricacies of an individual yeast cell's response to different stress. However, studies applying single-cell transcriptomic studies in micro-organisms such as the budding yeast have been limited in comparison to studies of larger mammalian cells. Recent microfluidics-based methods have allowed high throughput scRNA-seq studies to be possible(Zheng et al., 2017). Most notably, methods such as those developed by 10x Genomics have allowed the simultaneous sequencing of hundreds of thousands of cells ("Genomics X. 1.3 million brain cells from E18 mice," 2017). Though these methods are well optimized for mammalian cells, they are not easily applicable to yeast cells, due to the complexity introduced in sample preparation by having to lyse individual microbial cell walls, as well as the relatively low amounts of mRNA in small microbial cells compared to mammalian cells. Consequently, though studies of yeast stress responses have been extensively carried out throughout the past decade using a variety of techniques, scRNA-seq based studies have been relatively few.

To fill this gap, we adapted a full-length scRNA-seq method to profile transcriptomes of single yeast cells under hypotonic osmolarity, glucose starvation or amino acid starvation. Although our sample size is not on the scale of droplet-based scRNA-seq methods that only sequence the 3'-end of mRNA molecules and are more prone to drop-out effects due to shallow sequencing depths, we sequenced each single-cell transcriptome in full-length to a sufficient depth and are thus less prone to drop-out effects. Furthermore, a single-cell approach allows for more biological repeats than a bulk-based

procedure. Using these data, we not only confirmed many earlier findings based on bulk cell data, but also reveal novel aspects of stress responses in yeast at single-cell level.

2.2 Methods

2.2.1 Cell culture and spheroplasts preparation

A monoclone of the yeast strain S288C (ATCC) was selected using a YPD based agar (10% yeast extract, 20% peptone, 2% glucose and 20% agar) petri plate and stocked at -80 °C until use. To wake up cells, 30 µl thawed yeast stock inoculated in 3 ml YPD medium (1% yeast extract, 2% peptone and 2% glucose) was incubated overnight at 30 °C and 250 rpm. Cells were then expanded at 30 °C and 250 rpm after a 1:50 dilution in the YPD medium until mid-logarithmic phase (OD_{600} between 0.5 and 0.8). Five OD unit (ODU) cells were collected by centrifugation (500 g, 5 min) at room temperature. The cells were resuspended in autoclaved water and collected by centrifugation (500 g, 5 min) at room temperature. The cells were then resuspended in the softening medium (100 mM Hepes-KOH, pH 9.4, 10 mM Dithiothreitol) and incubated in room temperature for 15 min. The cells collected by centrifugation (500 g, 5 min) at room temperature were then resuspended in the Spheroplasts (S) medium (1× YNB, 2% glucose, 1x amino acids, 50 mM Hepes-KOH, pH 7.2, and 1 M sorbitol) (Dunn & Wobbe, 2001) to a concentration of 5 ODU/ml. Zymolyase 100T was added to the spheroplasts suspension to a final concentration of 2 µl/ODU, followed by 60 min incubation at 30°C to remove the cell wall and equilibrate cells to an isotonic, nutrient-rich condition. After two washes in the S medium by centrifugation (500 g, 5 min) at room temperature, spheroplasts were re-

suspended to 5 ODU/ml in the desired treatment solution: amino acids starvation (AAS): S medium (with 1.0 M Sorbitol) without amino acid; glucose starvation (GS): S medium (with 1.0 M Sorbitol) without glucose; hypotonic: S medium without sorbitol; isotonic condition: S medium with 1M sorbitol. Cells were exposed to the treatment for 0.5~2.0 hours before manual harvest.

2.2.2 Single cell harvest

0.5 mL of the spheroplasts were placed on a poly-lysine coated circular cover slip (2 mm diameter) in a petri dish for 5 min at room temperature (23 °C). The cover slip was broken in the center with forceps, and a small piece of cover slip was transferred to a 30 µl perfusion chamber, which was constantly perfused by a desired solution by gravity feeding. The solution change time in the chamber was about 20 sec. Single cells were harvested using a path clamp electrode pipette using a micromanipulator (ROE-200, Sutter) under an inverted microscope (Olympus 1X71) placed on a vibration isolation table (TMC). A cell was harvested in less than 10 nl perfusion solution.

2.2.3 Single cell RNA-seq library preparation

Our method is based on (Tang, Barbacioru, Bao, et al., 2010; Tang, Barbacioru, Nordman, et al., 2010) with modifications to prepare multiplex sequencing libraries using Illumina Nextera XT Kit. Briefly, a harvested cell was quickly transferred using a home-made microinjection system to a 200 µl Eppendorf tube containing 4 µl cell lysis buffer (0.9× PCR Buffer II, 3 mM $MgCl_2$, 0.45% NP40, 4.5 mM DTT, 0.18 U/µl SUPERase-In, 0.36 U/µl RNase Inhibitor, 12.5 nM AUP1 primer, 2 mM dNTP). In most single-cell samples, 0.1 µl (1:2.5x10^4 dilution) ERCC spike-in mRNA (Thermo Fisher) was added.

The cell was lysed at 70 °C for 90 sec, then placed on ice and stored at -80 °C until use. A cell lysate was thawed on ice, and 1 µl reverse transcription mix was added (13.2 U/ µl SuperScript III Reverse transcriptase, 0.4 U/µl Rnase Inhibitor, and 0.07 µg/µl T4 gene 32 protein). The first strand cDNA was synthesized by incubating the tube at 50 °C for 30 min, followed by inactivation of the reverse transcriptase at 70 °C for 10 min, and then the tube was cooled on ice. Free AUP1 primers were removed by adding 1 µl ExoSAP (Affymetrix) to the tube and incubating at 37 °C for 15 min, followed by inactivation of the ExoSAP at 80 °C for 15 min. This step would leave the AUP1 sequences at the 5'-end cDNA intact. A polyA tail was then added to the 3'-end of the first strand cDNA by adding 6 µl TdT mixture (1× PCR Buffer II, 1.5 mM $MgCl_2$, 3 mM dATP, 0.75 U/µl Terminal Transferase and 0.1 U/µl RNase H) and incubating at 37 °C for 15 min, followed by inactivation of the enzyme at 70 °C for 10 min. The resulting products (12 µl) were then divided into two equal portions (each 6 µl), and each was mixed with 19 µl second strand buffer (1× High Fidelity PCR Buffer, 2 mM $MgSO_4$, 0.2 mM each dNTP, 0.3 µM AUP2 primer, and 0.1 U/µl high fidelity Platinum Taq DNA polymerase). The two tubes were subject to one PCR cycle (30 sec at 95 °C, 2 min at 50 °C and 6 min at 72 °C) to synthesize the second-strand cDNA in the form of 5'-AUP2-T24-cDNA-A24-AUP1-3'. Nineteen µl PCR mixture (1× High Fidelity PCR Buffer, 2 mM $MgSO_4$, 0.25 mM each dNTP, 2 µM AUP1 Primer, 2 µM AUP2 Primer, 0.1 U/µl Platium Taq DNA Polymerase High Fidelity) was added to each tube, which brings the volume of each reaction to 44 µl, and cDNA was amplified by 18 PCR cycles (98 °C for 5 sec, 67 °C for 1 min and 72 °C for 6 min). The resulting cDNA from two reactions were combined (total 88 µl) and were further subject

to 12 cycles of PCR with two duplicates, each with 2.4 µl sample and 87.6 µl PCR mixture (1× High Fidelity PCR Buffer, 2 mM $MgSO_4$, 0.375 mM each dNTP, 1 µM AUP1 Primer, 1 µM AUP2 Primer, 0.1 U/µl Platium Taq DNA Polymerase High Fidelity). The products were then combined, and cDNA was revolved on a 1% ager gel (25 µl sample per lane). The band between 300 bases to the loading well was cut and cDNA was purified using a QIA quick gel purification Kit, followed by magnetic beads (GE Health) purification (10:7 sample to beads ratio). After quantification using a Bioanalyzer (Agilent High Sensitivity DNA Kit), the libraries were then prepared using an Illumina Nextera XT or TruSeq DNA Sample Preparation Kit according to the vendor's guide. The libraries were sequenced on an Illumina HiSeq2000 or HiSeq2500 machine (100 base-paired reads). Bulk mRNA was also extracted from population spheroplasts under AAS using a yeast RiboPure™ RNA Purification Kit (Ambion). Different amount of purified bulk mRNA (5pg, 10pg, 20pg, 1,00pg, 1000pg and 10,000pg) were used to construct sequencing libraries in the same way as for single-cell libraries, with the exception that 0.1µl ERCC spike-in mRNA (Thermo Fisher) was added to the lysis buffer, with a concentration of $1:5x10^5$, $1:2.5x10^5$, $1:1.25x10^5$, $1:2.5x10^4$, $1:2.5x10^3$ $1:2.5x10^2$ for the 5pg, 10pg, 20pg, 100pg 1,000pg and 10,000pg input RNA, respectively.

2.2.4 Characterization of single-cell transcriptomes

The raw reads from FASTQ files of different lanes were first combined based on sample ID before being mapped to the *S. cerevisiae* reference genome (SGD R64-2-1) using STAR (version 2.5.2) (Dobin et al., 2013). STAR alignments provided soft clipping of possible adapter sequences at the ends of reads; thus, no prior trimming of reads was

performed. Gene transcription levels were quantified using uniquely mapped reads in TPM (transcript per million mapped reads) by RSEM(v1.2.31) (B. Li & Dewey, 2011) and in raw counts by QoRTs (Hartley & Mullikin, 2015) (default parameters, using the formatted for DESeq2 count files). Principal components analysis (PCA) and uniform manifold approximation and projection (UMAP) of the cells/samples were performed on the $\log_2$(TPM+1) values.

2.2.5 Library Quality Assessment

We adopted the following metrics to quantify the quality of the scRNA-seq libraries from each cell/sample. 1) The library complexity is defined as the number of distinct (unique) read's start positions mapped to the genome (Levin et al., 2010). To directly compare the library complexity of different libraries, we randomly sampled the same number of reads (one million) from each library. 2) The evenness is defined as the averaged coefficient of variation (CV) of the read coverage along each base-pair of the gene body (Levin et al., 2010). Since transcripts of a low copy number are subject to uneven coverages, only the top 50% highly expressed genes are used in the calculation of this measure. 3) Continuity of coverage measures the number of gaps along the exons of a gene, where a gap is defined as a consecutive length of ≥5 bases without any reads mapped. The final gap measure is a weighted average of gaps in all the genes according to each gene's expression in TPM values. 4) Sensitivity, which measures the number of genes detected with at least 5 reads in each sample. And 5) Bio-reads ratio, which measures the proportion of reads uniquely mapped to the organism genome. Sample quality control and filtering were performed slightly differently on the two batches of our libraries due to their different

sequencing depth and quality: samples from batch 1 were retained if the following thresholds were met: detection rate > 500, complexity > 0.1, gap < 0.5, evenness < 1.5; while samples from batch 2 were retained if the following thresholds were met: num uniquely mapped reads > 10000. All quality metrics apart from Sensitivity were calculated for spike-in reads for each sample independently as well. Metrics such as rRNA expression % and Mitochondrial gene expression % were calculated using TPM values, genes used in these calculations are included in Supplementary Table 2-1, Supplementary File 1. An additional metric was calculated for spike-in reads based on the detection-limit procedure described in (Svensson et al., 2017) using TPM expression of cells that had > 2% spike-in rate and at least 8 different spike-ins with non-zero expression. Pearson correlation was calculated between all samples that passed filtering and samples of previous studies(Gasch et al., 2017; Nadal-Ribelles et al., 2019; J. Wang et al., 2022). Normalized expression data was obtained from GEO according to the authors, specifically, (44) and (45) was normalized with the median normalization used in DESeq2(Michael I Love, Huber, & Anders, 2014), and data from (Gasch et al., 2017) was normalized with the SCnorm method (Bacher et al., 2017). In order to select samples similar to the isotonic conditions used in our study, yeast cells aged 2-h were selected from the study of (J. Wang et al., 2022), samples labeled ‘Unstressed’ were selected from (Gasch et al., 2017) and only samples of the BY4741 strain were selected from (44). For each study, we normalized our data according to the specified normalization method and used gene features shared between our data and the selected dataset in the calculation of Pearson correlation.

2.2.6 Differential Expression Analysis and GO/Pathway Enrichment

Analysis of differentially expressed genes was performed using MAST (Finak et al., 2015). For each comparison, only genes expressed in at least 20 percent of samples were used for subsequent analysis. Counts per million (CPM) expression values (calculated with raw counts) were used as the input and all other parameters were left at default. Differential expression results were controlled for gene detection rate. Genes that met the criteria: FDR < 0.05 and absolute fold-change > 1.2 were labeled as differentially expressed. Enrichment analyses for Gene Ontology biological processes(Ashburner et al., 2000; Gene Ontology, 2021), KEGG Pathways(Kanehisa, 2019; Kanehisa, Furumichi, Sato, Kawashima, & Ishiguro-Watanabe, 2022; Kanehisa & Goto, 2000) and Wiki-pathways(Martens et al., 2021) were carried out using clusterProfiler (Yu, Wang, Han, & He, 2012) based on the identified differentially expressed genes.

2.2.7 Promoter and Transcription Factor Binding Sites Analyses

Upstream sequences (maximum length of either 1,000bp or the entire upstream intergenic sequences) of differentially expressed genes (DEGs) were extracted from the *S. cerevisiae* reference genome (SGD R64-2-1). Motifs in the sequences were identified using ProSampler (Y. Li, Ni, Zhang, Li, & Su, 2019), which were then compared with known motifs in the Yeastract database (Monteiro et al., 2020) using TomTom (Gupta, Stamatoyannopoulos, Bailey, & Noble, 2007). This analysis was performed separately for upregulated genes and downregulated genes of each stress condition. TF-gene interactions from Yeastract were extracted for all genes filtering for interactions that were either documented or had gene expression evidence. When comparing our putative TF-gene

relationships with TF-gene relationships from Yeastract, we omitted relationships involving Spt15p and Nhp6ap from our analysis as these TFs are known to bind TATA boxes and are also absent from TF-gene relationships from Yeastract.

Empirical correlation analysis between TF and predicted target DEGs was performed as follows:

Given a table of TF-gene interactions, for each TF:

Step 1: calculate the mean correlation C between TF and its interacting genes.

Step 2: let N be the number of predicted DEGs interacting with the TF, sample N genes from all genes and calculate mean of absolute correlations between the TF and N sampled genes.

Step 3: repeat step M (M >= 10000) times to generate empirical distribution of mean of absolute correlations.

Step 4: p-value = (Number of sampled mean of correlation values > S)/M

2.2.8 Gene Clustering and TF Enrichment

For samples of the same condition and batch, absolute Pearson correlation distance (1-| Pearson correlation|) was calculated between gene TPM expression levels and used for hierarchical clustering using the hclust method in R with method set to 'ward.D2'. Number of clusters was determined with R package NbClust (min.cluster = 2, max.cluster = 200)(Charrad, Ghazzali, Boiteau, & Niknafs, 2014). Each resulting cluster was then enriched for TFs using known Yeastract TF-Gene interactions using a hypergeometric test

from R package clusterProfiler(Yu et al., 2012). Each cluster was then assigned with the most significant TF only if its Benjamini-Hochberg adjusted p-value < 0.05.

2.3. Results

2.3.1 Transcriptomes of single yeast cells are sufficiently sequenced

Using the scRNA-seq protocol, we sequenced the full-length transcriptomes of 117 yeast cells from four treatments in two sequencing batches: amino acids starvation (AAS) (n=20 in batch 1), isotonic (n=19 in batch 1; n=19 in batch 2), glucose starvation (GS) (n = 47 in batch 2) and hypotonic conditions (n=12 in batch 1). As summarized in Supplementary Table 2-1, an average of 9.5 million reads were generated in each cell, and an average of 36% and 11% of them were uniquely mapped to the genome and ERCC spike-in RNA sequences, respectively. Moreover, we sequenced a total of 59 RNA-seq libraries prepared using varying amounts (5pg, 10pg, 20pg, 100pg, 1,000pg and 10,000pg) of bulk RNA from yeast cells under AAS using the same protocol. As summarized in Supplementary Table 2-2, an average of 6.7 million reads were generated in bulk samples, and an average of 34% and 24% of them were uniquely mapped the genome and ERCC spike-in RNA sequences, respectively.

To see whether the sequencing depth was sufficient to cover all the captured mRNAs in a cell, we randomly sampled different numbers of mapped reads from the scRNA-seq libraries and computed the total number of genes to which reads were mapped. As shown in Figure 2-1A~E, the number of genes detected approached saturation when around 1 million reads were sampled for each cell, suggesting that for most of our scRNA-seq libraries, the sequencing depth should be more than sufficient to detect most transcribed

mRNAs. However, there is considerable variation in the number of genes detected between cells under the same treatment (Figure 2-1A~E). Comparing the two different batches of isotonic treatment (Figure 2-1A, E) reveals that batch effect contributes significantly to the difference in the number of genes detected, a well-known phenomenon for scRNA-seq libraries (Brennecke et al., 2013; Buttner, Miao, Wolf, Teichmann, & Theis, 2019; Haghverdi, Lun, Morgan, & Marioni, 2018; C. Jia et al., 2017). On the other hand, the more input bulk mRNA amount, the more genes detected in the bulk mRNA samples (Figure 2-1F). Therefore, the varying numbers of genes detected in single cells are most likely caused by varying amounts of mRNA that each cell expressed or released during lysis, which is characteristic of scRNA-seq libraries (G. Chen, Ning, & Shi, 2019), thus partially explaining the variation in detection rates of genes. Bootstrapping single cells by computationally pooling raw reads from a set of single cells under the same treatment shows, except for GS, the detection rate already reaches saturation when reads from as few as five cells were combined (Figure 2-1G), indicating that aggregating over the single cells can provide a comparable sample to that of bulk analysis.

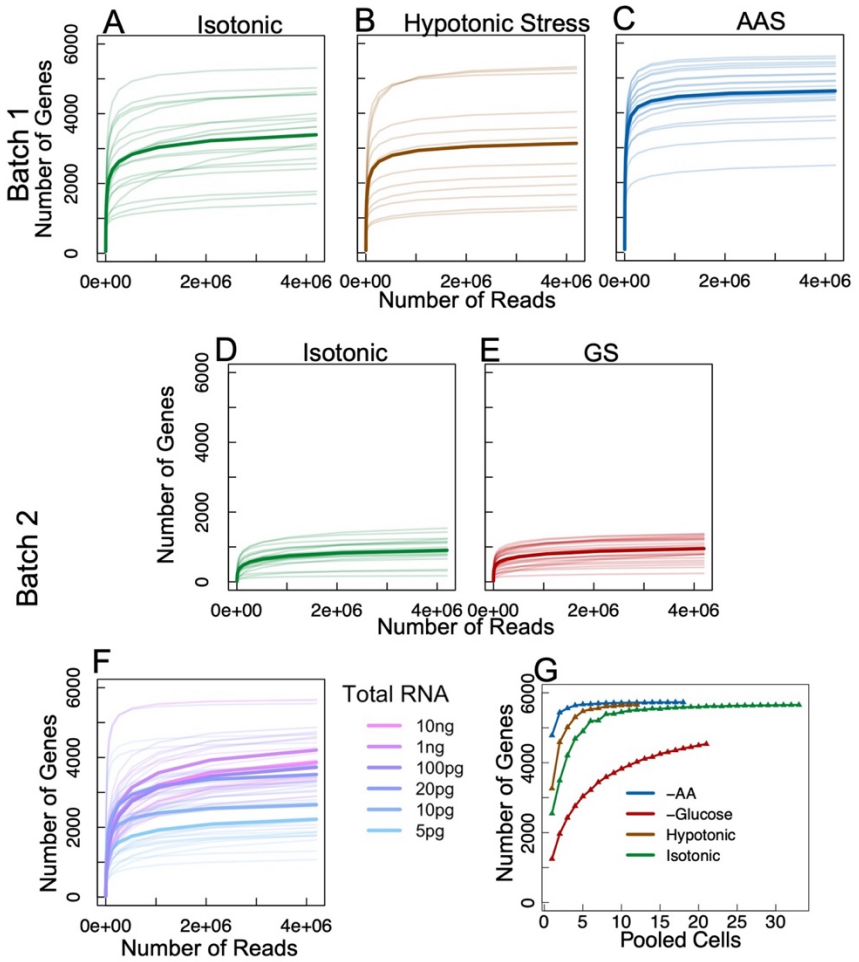


Figure 2-1 Saturation of detected genes in the libraries. A. Number of genes detected in each cell under isotonic condition from batch 1 as a function of number of mapped reads randomly sampled from the cell. **B.** Number of genes detected in each cell under hypotonic stress samples from batch 1 as a function of number of mapped reads randomly sampled from the cell. **C.** Number of genes detected in each cell under AAS from batch 1 as a function of number of mapped reads randomly sampled from the cell. **D.** Number of genes detected in each cell under GS from batch 2 as a function of number of mapped reads randomly sampled from the cell. **E.** Number of genes detected in each cell under isotonic condition from batch 2 as a function of number of mapped reads randomly sampled from the cell. **F.** Number of genes detected in each bulk sample as a function of number of mapped reads randomly sampled from the sample. The bulk libraries were prepared using different amount (5pg, 10pg, 20pg, 1,00pg, 1000pg and 10,000pg) of input mRNA extracted from a population of cells under AAS. **G.** Number of genes detected as a function of number of single cells randomly selected from those under the same conditions. Bold colored lines represent average detections.

An evaluation of the possible bias of read coverage along a gene body for all libraries show that the read coverage declines toward the 5'-end for all libraries (Figure 2-2). This result is consistent with the earlier findings that single-cell RNA-Seq data tend to be biased toward the 3' end, because of the oligo (DT) primers used in the first-strand cDNA synthesis(Tang, Barbacioru, Bao, et al., 2010; Tang, Barbacioru, Nordman, et al., 2010). We also noted that there is pronounced difference in biases between the two batches of isotonic samples (Figure 2-2A).

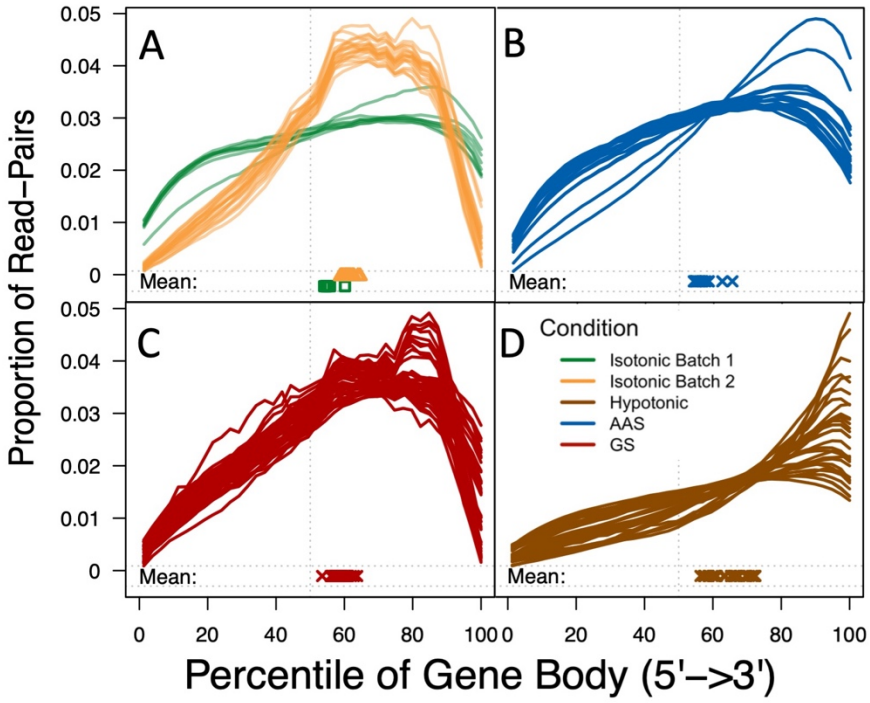


Figure 2-2 Reads coverage along the 5'-end to the 3'-end of the coding regions of genes. For each library, the averaged relative coverage is shown at each relative position along the length of coding regions of genes from the 5'-end to the 3'-end. **A**. Isotonic condition, batch 2 samples are colored orange here to distinguish from batch 1 samples (colored green). **B.** AAS. **C.** GS. **D.** Hypotonic stress.

2.3.2 Library quality control reduces technical artifacts

Since scRNA-Seq results are sensitive to multiple factors during library preparation, we evaluated each library for its quality by several assessment criteria: complexity, evenness of coverage, continuity of coverage, sensitivity, and bio-reads ratio (see Methods). The complexity of the libraries ranges from 0 to 0.80 with an average of 0.30 (Figure 2-3A). There is significant difference between the complexity of the two batches, with samples of batch 2 showing lower complexity, indicating the presence of a strong bias in fragment amplification and insufficient sampling of mRNA molecules in the individual cells (Figure 2-3A). This result is consistent with the considerably lower gene detection rate for batch 2 isotonic cells than for isotonic cells in batch 1 (Figure 2-1A, E). Most libraries have less than one gap on average, though a few exhibits high number of gaps. In general, these measurements of library quality from different aspects are highly correlated (Figure 2-3A). We filtered out low quality cells using a procedure as detailed in the Methods section. As shown in Figure 3B, 84 of the 117 libraries passed the filters set by the five metrics. Clearly, after filtering, the distributions of the metrics are more uniform, most notably in the case of the GS treatment. Thus, our subsequent analyses were based on these 84 cells. Furthermore, we included ERCC spike-in in most of our samples as a quality control for assessing the accuracy of our gene expression quantification. The assessment of quantification quality was carried out using all bulk/single cells samples with spike-in rate of > 2%. It can be seen in

Figure **2-4**A that the expression levels (TPM) of spike-ins added in single-cell samples correlate well with the concentrations of the spike-ins (PPC = 0.842), indicating that our

expression quantifications are reliable. Furthermore, we show that ERCC spike-in in bulk RNA-seq samples all show similar levels of correlation with known concentrations despite the different concentrations of spike-in added (

Figure **2-4**B-G). However, it is worth noting that as the concentration of added spike-ins increase, the quantified expressions show lower variance between samples as expected. We also applied the detection limit metric described in (Svensson et al., 2017) to our spike-in data, using samples with at least 2% spike-in rate and having at least 8 different spike-in with non-zero expressions for this calculation. For these samples, the detection limits found were similar to the levels described in (Svensson et al., 2017) for the SmartSeq2 protocol (

Figure **2-4**H), indicating our procedures are reliable in terms of the ability to detect RNA molecules. All quality metrics for biological reads and spike-ins as well as additional quality metrics such as Mitochondrial gene % and Ribosomal RNA % are also included in Supplementary Table 2-1, Supplementary Table 2-2. Pearson correlations were also calculated between our samples post filtering and those of previous studies described in the Methods section (

Figure **2-4**I) (Gasch et al., 2017; Nadal-Ribelles et al., 2019; J. Wang et al., 2022). Although these studies performed single cell RNA-seq on a slightly different cell line (BY4741) and used different protocols in their sample preparation, our batch 1 isotonic cells were able to achieve a mean Pearson correlation of 0.59 (max: 0.91, min:0.26) across all cells under a similar unstressed condition, indicating that the transcriptomic profiles of

our batch 1 isotonic yeast cells are similar to unstressed yeast cells from these previous studies. Moreover, the study of (J. Wang et al., 2022) utilized a slightly modified Smartseq-2 protocol, and thus our batch 1 isotonic cells correlate very well with their 2-hour cells (Pearson correlation max: 0.91, min:0.45, mean: 0.70). Differences in the degree of correlation is likely due to differences in sample preparation protocol, yeast strains used (S288C vs BY4741), and slight variations in the condition of the cells. On the other hand, our AAS and hypotonic cells had much lower correlations. Isotonic samples from Batch 2 correlated poorly to all the studies due to strong batch effects and technical artifacts in these samples, (Pearson correlation max: 0.45, min:0.04, mean: 0.23). It is worth noting that despite the poor correlation of Isotonic Batch 2 samples with unstressed cells from previous datasets, the degree of correlation was still consistently higher than the correlation between GS samples (also from Batch 2, Pearson correlation max: 0.33, min:0.006, mean: 0.12) and the unstressed cells from previous datasets (

Figure **2-4**I), again indicating that poor correlation likely originates from batch effects and that inherent biological differences caused by stress are still preserved in the transcriptomic profile within the same batch. Due to the presence of these strong batch effects, the technical differences between the two batches are too significant for aggregated analysis, thus subsequent analyses were performed in a batch-specific manner: isotonic vs AAS and isotonic vs hypotonic comparisons were conducted with cells from batch 1, while isotonic vs GS comparison was performed with cells from batch 2.

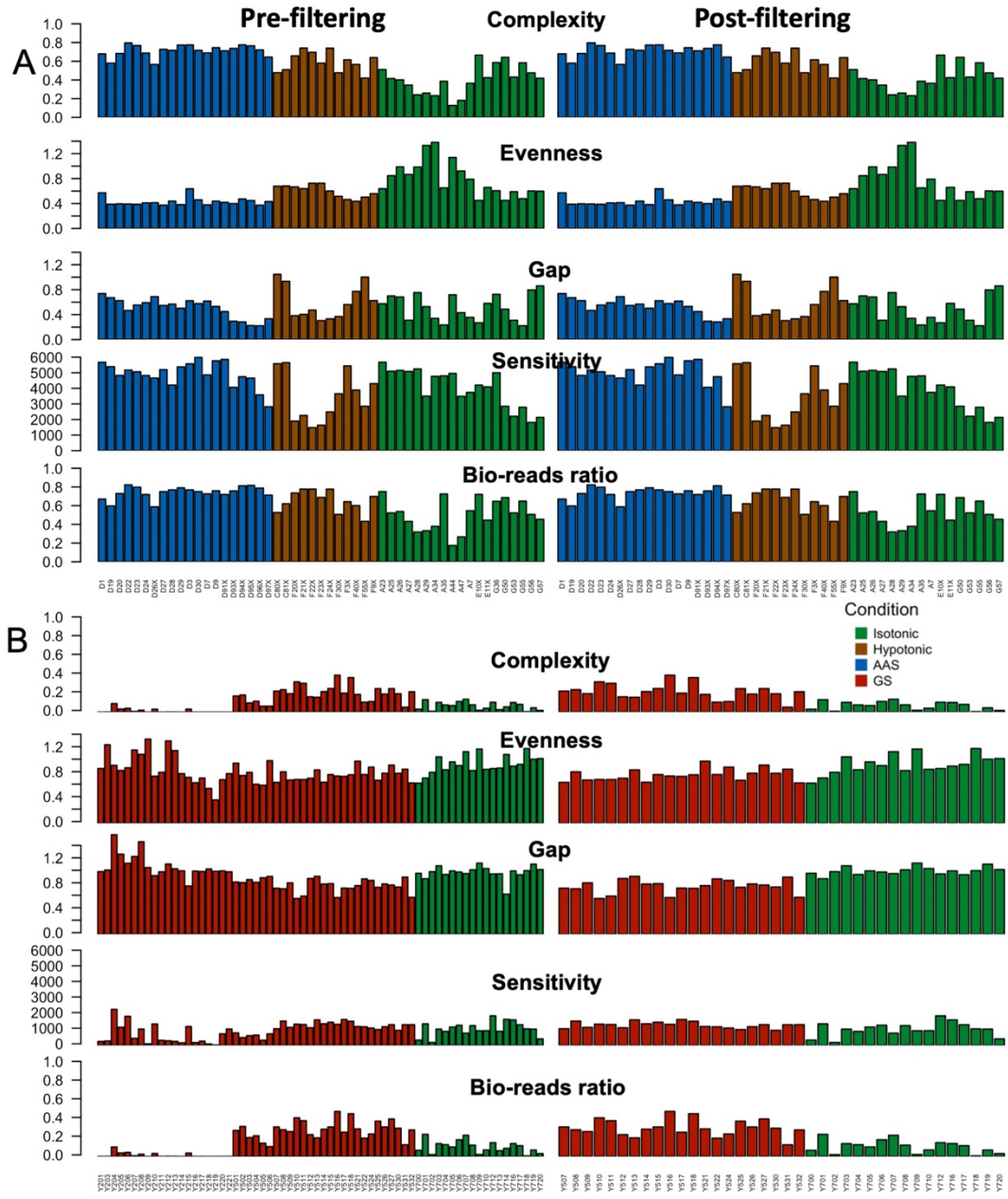


Figure 2-3 Quality control of scRNA-seq libraries using complexity, evenness, gap, sensitivity, and bio-reads ratio. In both panel **A** and **B**, cells pre-filtering are shown on the left panel, and the cells post-filtering are shown on the right panel. **A.** Batch 1 samples (51 samples pre-filter / 46 samples post-filter). **B.** Batch 2 samples (66 samples pre-filter / 38 samples post-filter).

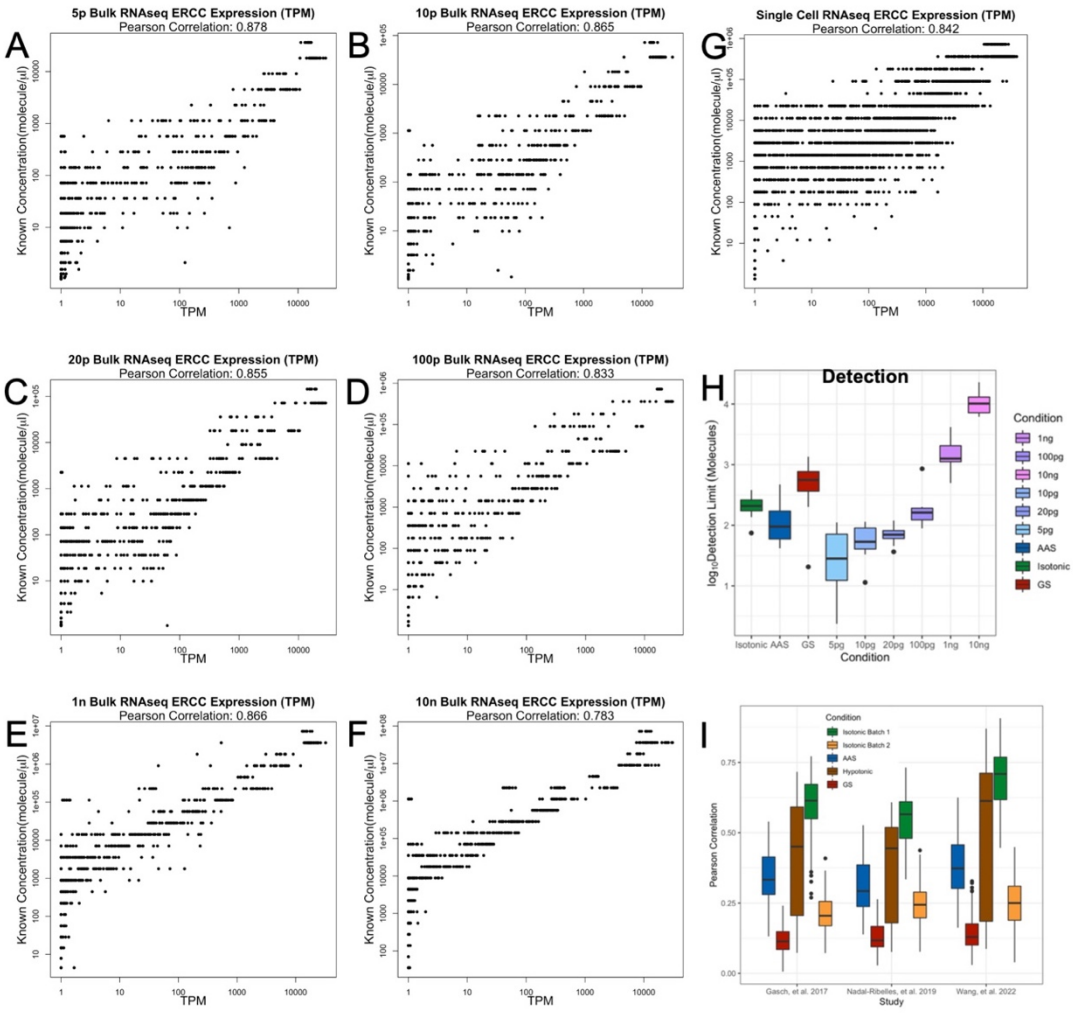


Figure 2-4 ERCC Spike-in Expression (TPM) vs Known Concentration in samples with at spike-in rate > 2%. A-F: 5pg, 10pg, 20pg, 100pg, 1ng, 10ng bulk-RNAseq sample ERCC expression vs known concentration(molecule/µl). G: single cell RNAseq ERCC expression vs known concentration (molecule/µl). H: Detection limit metric described in (Svensson et al., 2017). I: Pearson correlation with previous datasets(Gasch et al., 2017; Nadal-Ribelles et al., 2019; J. Wang et al., 2022) colored by condition and batch, black bars in box plot represent mean Pearson correlation.

2.3.3 AAS and GS treatments induce distinct transcriptomes while hypotonic stress does not

To characterize the transcriptome features of the cells under different treatments, we compared the transcription levels of the 5,419 genes (Supplementary Table 2-3) that

were expressed in at least 20% of the cells using principal component analysis (PCA) (Figure 2-5A-C), uniform manifold approximation and projection (UMAP) (Figures 4D~4F) and hierarchical clustering (Figure 2-5G~I). Clearly, cells under either AAS (Figure 2-5A, D, G) or GS (Figure 2-5B, E, H) treatments display distinct transcriptomes from cells under the isotonic condition, suggesting that both AAS and GS triggered relevant gene regulatory pathways in the cells. However, cells under hypotonic stress did not separate well from isotonic cells in PCA (Figure 2-5C), UMAP (Figure 2-5F) and Hierarchical Clustering (Figure 2-5I), suggesting that the hypotonic stress did not significantly alter gene transcription in the cells.

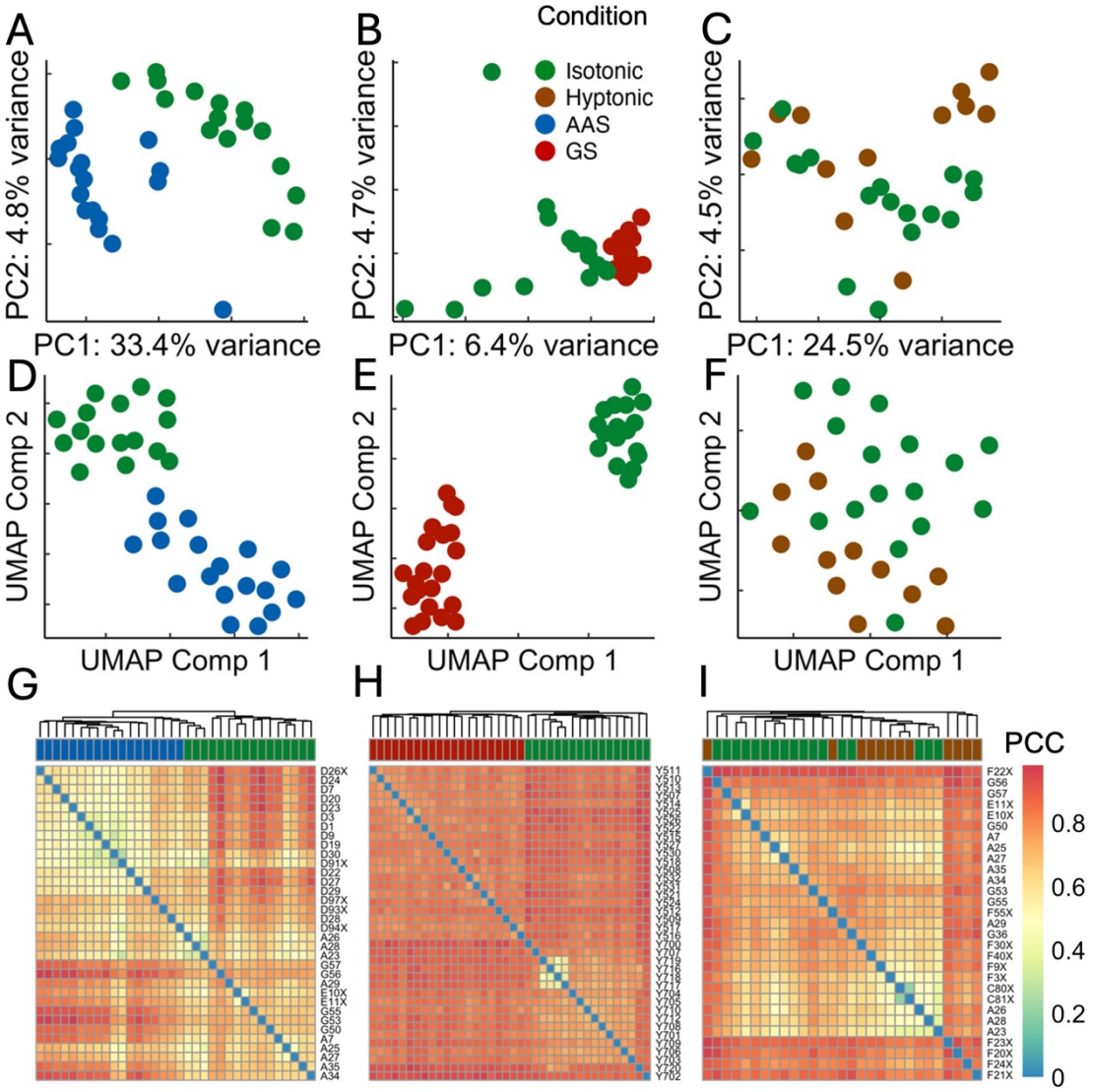


Figure 2-5 Comparison of transcriptomes of the cells under different stress conditions with those of cells under the isotonic condition. A, B, C. Visualization of the cells under AAS (A), GS (B) and hypotonic stresses (C) in comparison with cells under the isotonic condition using the first and second PCA components of their transcriptomes. D, E, F. Visualization of the cells under AAS (D), GS (E) and hypotonic (F) stresses in comparison with cells under the isotonic condition using the first and second UMAP components of their transcriptomes. G, H, I. Heatmaps of hierarchical clustering of the cells based on the Pearson correlation coefficients (PCC) of their transcriptomes using complete linkage. Color code shown in (B) applies to all the figures.

We next identified differentially expressed genes (DEGs) between cells under AAS, GS and hypotonic stresses to those under isotonic treatment using MAST (Finak et al., 2015) (FDR<5%, absolute (Log_2(fold-change (FC))> 0.5). We identify 409 genes that were significantly differentially expressed under AAS relative to the isotonic condition (Supplementary Table 2-4), of which 137 were down regulated while 272 were up regulated (Figure 2-6A, C). Under GS, 80 genes were significantly differentially expressed relative to the isotonic condition (Supplementary Table 2-5), of which, 27 and 53 were significantly upregulated and down regulated, respectively (Figure 2-6B, D). Contrastingly, no genes were found to be significantly differentially expressed in hypotonic condition relative to the isotonic condition. This is consistent with the results from the above PCA (Figure 2-5C), UMAP (Figure 2-5F) and clustering (Figure 2-5I) analyses, where cells under isotonic and hypertonic conditions are indistinguishable. In our experiments, cells under all treatments (AS, GS, hypotonic and isotonic) were exposed to 1 M sorbitol solution for about one hour after being harvested from the log-phase growth in YPD in a procedure to remove the cell wall. Subsequently, to induce hypotonic shock, isotonic cells were exposed to a sorbitol lacking solution for at least half hour before being collected, at which point the cells had adapted to the hypotonic environment according to the earlier study (Gasch et al., 2000).

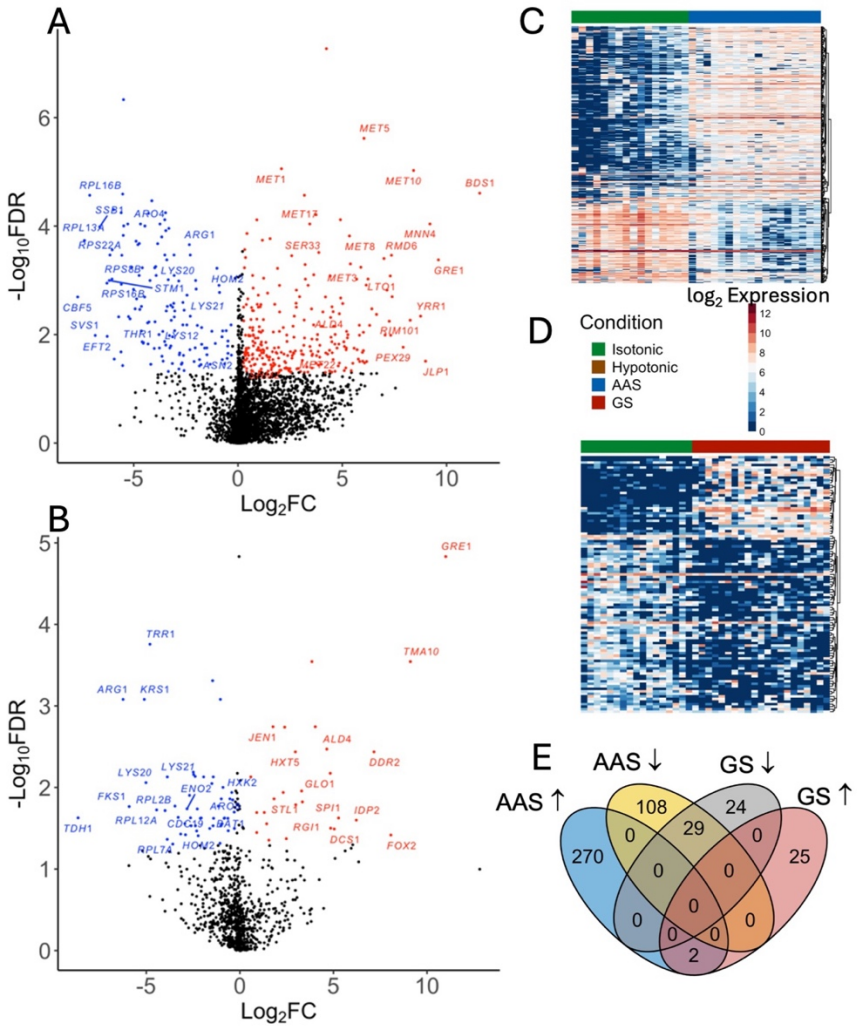


Figure 2-6 AAS and GS induce distinct gene expression patterns. A. Log_2FC vs $-Log_{10}$ (adjusted p-values) volcano plots of DEGs under AAS. B. Log_2FC vs $-Log_{10}$ (adjusted p-values) volcano plots of DEGs under GS. C. Heatmap of the expression levels of DEGs under AAS. D. Heatmap of the expression levels of DEGs under GS. Bars on the top of heatmaps indicate the treatments of cells according to the colors shown in the legend. The intensity of the heatmaps is colored according to $Log_2(TPM+1)$ values. E. Venn diagram of AAS upregulated (↑), AAS downregulated (↓), GS upregulated (↑) and GS downregulated (↓) genes.

The 272 upregulated genes under AAS are enriched for Sulfur metabolism and Sulfur amino-acid pathways, such as MET*1/3/5/8/10/17/22* as well as for the general

response *GRE1* (Figure 2-6A). The 137 down-regulated genes are enriched for ribosomal and translation functions, 77 of which encode components of a ribosomal subunit. Moreover, we found that genes involved in lysine (*LYS12/21/20*), threonine (*THR1, HOM2*), arginine (*ARG1)*, asparagine (*ASN2*) and aromatic amino acids (*ARO4*) biosynthesis are significantly downregulated (Figure 2-6A). These results indicate that there may be a preferential synthesis of Sulfur amino acids when the cell was limited for all amino acids. A more recent study has found that methionine functions as an anabolic signal to induce global gene expression changes when in excess and that the presence of methionine can trigger the synthesis of other amino acids under AAS to sustain anabolism (Walvekar, Srinivasan, Gupta, & Laxman, 2018). Furthermore, genes involved in the biosynthesis of glycine and serine (*SHM2*, *SER33)* were also found to be significantly upregulated. In this regard, glycine and serine have been known to be crucial one-carbon donor molecules in yeast and contribute to the synthesis of methionine (Piper, Hong, Ball, & Dawes, 2000).

Of the 53 downregulated genes under GS stress, 29 were also downregulated under AAS stress (Figure 2-6E). Specifically, 16 of the 53 downregulated genes encoded ribosomal subunits (Figure 2-6B). There is also a down-regulation of several amino acid biosynthesis genes such as *LYS20/21, ARO8, BAT1, HOM2* and *ARG1* (Figure 2-6B), possibly by Snf1p mediated repression of *GCN4* translation (Shirra et al., 2008). The other down-regulated genes are mainly involved in carbohydrate metabolism. For example, hexokinase isoenzyme 2 (Hxk2p) encoding gene *HXK2* showed decreased expression (Figure 2-6B). Previous studies have shown that under GS, activation of Snf1p prevents

the interaction between Mig1p and Hxk2p, leading to de-repression of carbon-source responsive element (CSRE) containing genes involved in the use of alternate carbon sources(Conrad et al., 2014). The reduction in *HXK2* transcription might be a possible alternate mechanism to derepress CSRE containing genes. Our results also show decreased expression of *ENO*2 and *CDC*19 (Figure 2-6B). Eno2p and Cdc19p catalyze the last two steps in glycolysis, respectively, where 2-phosphoglycerate is converted by Eno2p to phosphoenolpyruvate, which is then converted to pyruvate by Cdc19p (Stryer, 1988). It has been found that under GS, the activity of Cdc19p is quickly inactivated(Y. F. Xu et al., 2012). Furthermore, mRNA levels of *CDC*19 and *ENO*2 are also reduced through phosphorylated eIF4G regulation of the degradation (Y. Chang & Huh, 2018).

The 27 upregulated genes under GS are glucose-repressed or involved in the use of alternate carbon resources (Figure 2-6B). For instance, the expression of *HXT5*, *STL1* and *JEN1* (Figure 2-6B) were significantly increased (Supplementary Table 2-5), whereas *HXT5* encodes a hexose transporter that has affinity for glucose only under GS (Buziol et al., 2002); *Stl1,* a high affinity glycerol importer (Ferreira et al., 2005) and *JEN1,* a high affinity symporter for alternate carbon sources such as lactose, pyruvate, and acetate (Casal, Paiva, Andrade, Gancedo, & Leao, 1999; Chambers, Issaka, & Palecek, 2004). Moreover, the expression of *FOX2* (Figure 2-6B), which encodes a multifunctional enzyme in the peroxisomal fatty acid beta-oxidation pathway was significantly upregulated (Figure 2-6B). It has been shown that *FOX2* also is activated in autophagy (Hiltunen et al., 1992). These results are consistent with the earlier notion that under GS when the extra-cellular environment provides no other carbon resource, in order to survive the cells might need to

recycle cytoplasmic components (bulk autophagy) and utilize an alternative source of energy (most likely metabolism of lipids) from within the cell (Seo et al., 2017).

Interestingly, we only found two genes *ALD*4 and *GRE1* to be upregulated under both GS and AAS stresses (Figure 2-6A, B, E). Previous studies have shown that the expression of *GRE1* is controlled through the HOG pathway and its transcripts accumulate under osmotic, ionic, heavy metal, heat shock and oxidative stress (Garay-Arroyo & Covarrubias, 1999; J. Wu et al., 2004). Our finding that *GRE1* was induced significantly under both AAS and GS conditions suggests that it may play a role in nutrient-limiting stress responses as well. *ALD4* encodes mitochondrial acetaldehyde dehydrogenase, which is necessary for the growth of yeast on ethanol (Aranda & del Olmo Ml, 2003). Although it is known that *ALD*4 is glucose repressed, our finding that it was also significantly upregulated under AAS stress suggests that *ALD4* might play a role in general stress response.

The DEGs are also significantly enriched for relevant Wiki-pathways, Gene Ontology (GO) biological processes and KEGG pathways. Specifically, down-regulated genes under both AAS and GS stresses, are enriched for translation-related Wiki-pathways (Figure 2-7A, B), GO terms (Figure 2-7E, F) and KEGG pathways (Figure 2-7C, D). Up-regulated genes under AAS stress were enriched significantly for Sulfur metabolism (Figure 2-7C) and alpha-amino acid biosynthesis (Figure 2-7E). Both up- and down-regulated genes under GS were enriched for Wiki- (Figure 2-7B) and KEGG (Figure 2-7D) pathways and GO terms biological processes (Figure 2-7F) related to carbon metabolism, pyruvate metabolism, gluconeogenesis, and tricarboxylic acids cycle. Notably, under GS,

DEGs were enriched for GO terms related to biosynthesis of lysine, arginine, cysteine, and methionine (Figure 2-7F), and most of these genes showed decreased expression, indicating that glucose limitation shut down the synthesis of these amino acids.

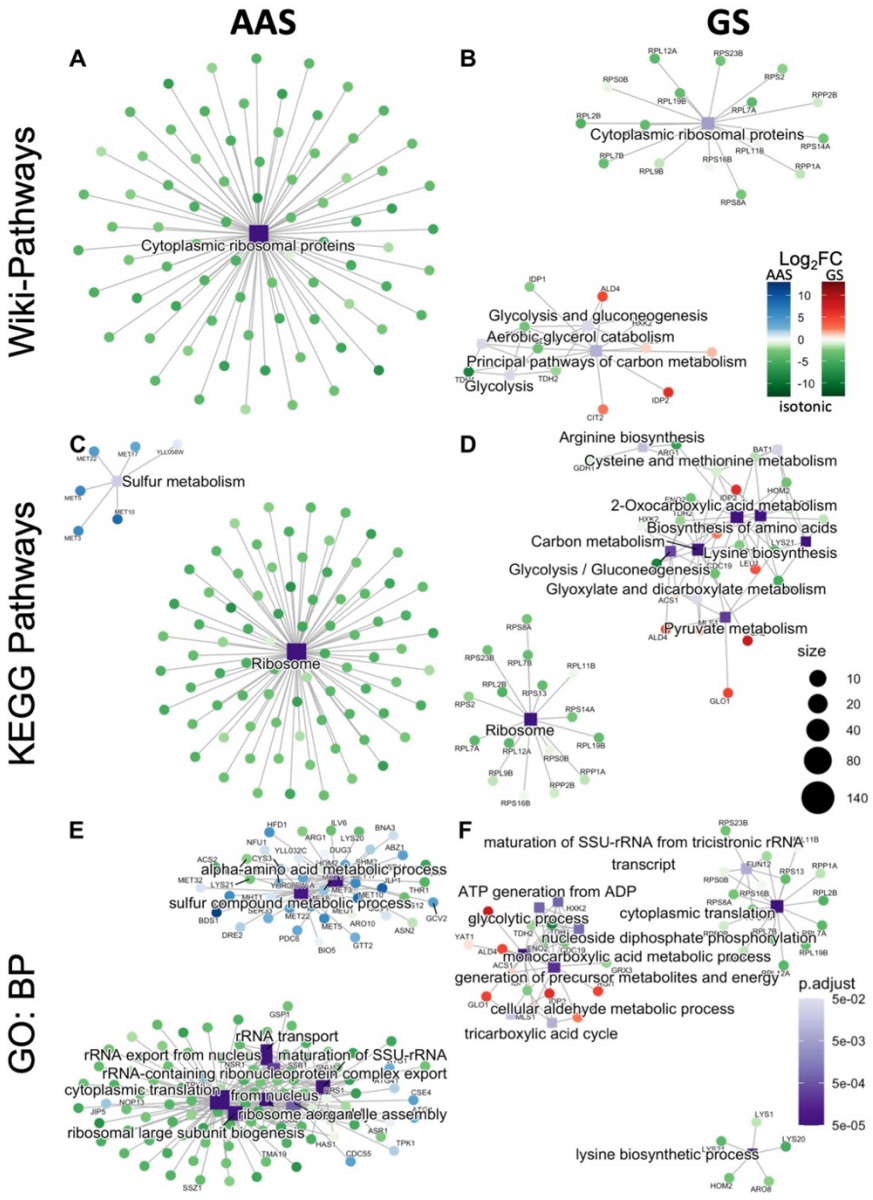


Figure 2-7 Enrichment of up- and down-regulated genes for known functional modules under AAS and GS. A, B. Wiki-Pathways enrichment under AAS (A) and GS (D). C, D. KEGG pathways enrichment under AAS (C) and GS (D). E, F. GO term biological process (BP) enrichment under AAS (E) and GS (F). Each square represents an

enriched functional module. Color of a circle reflects the adjusted p-value while size of the square reflects the number of genes in the enriched functional module. Each circle represents a gene that is connected by an edge to its belonging functional module. Genes are colored by their Log_2FC under the AAS or GS relative to the isotonic condition. False discovery rate (FDR<5%) was controlled using the Benjamini-Hochberg procedure.

2.3.4 Motifs found in the upstream regions of the DEGs reveal possible gene regulatory networks

To reveal possible gene regulatory networks responsible for the distinct expression patterns of gene induced under AAS and GS stresses, we identified possible TF binding sites (TFBSs) in the upstream regions of the DEGs under AAS and GS. We found six and eight motifs in the upstream regions of downregulated and upregulated genes under AAS, respectively. TomTom (Gupta et al., 2007) matched these six and eight motifs to known motifs of 14 and 23 TFs, respectively, in the Yeastract database (Supplementary Table 2-10). Such multiple hits are understandable, since it is well-known that TFs of the same protein family recognize highly similar motifs (Gordan et al., 2011; Inukai, Kock, & Bulyk, 2017). These TFs include amino acid starvation response TFs Gcn4p and methionine biosynthesis regulators such as Met4p, Met31p and Met32p. The details of the identified motifs in the upstream regions of the DEGs under AAS are shown in Supplementary Table 2-6. Moreover, we identified 12 motifs for 23 TFs and 10 motifs for 27 TFs in the upstream regions of down- and upregulated genes under GS, respectively. These TFs are involved in alternative carbon source utilization, such as Pdr3p, Ert1p, Sip4p and Aaca1p. The details of the identified motifs in the upstream regions of the DEGs under AAS are shown in Supplementary Table 2-7.

Based on the identified TFBSs and putative cognate TFs, we constructed gene regulatory networks for the DEGs under AAS consisting of 594 putative regulatory

relationships between 26 TFs and 218 genes (Figure 2-8A, for the details see Supplementary Table 2-8) and for DEGs under GS consisting of 114 putative regulatory relationship between 31 TFs and 55 genes (Figure 2-8B, for the details see Supplementary Table 2-9). Most of these inferred regulatory relationships are supported by existing data, while others might be novel findings. More specifically, 266 and 52 of our predicted TF-gene relationships under AAS and GS, respectively, are documented in the Yeastract database, while 235 and 47 of our predicted TF-gene relationships under AAS and GS, respectively, might be novel regulatory relationships (Figure 2-8C). For examples, we confirm multiple ribosomal protein genes such as *RPL16B/13B/18B, EPS0B/1B/18A* etc. (for the complete list, see Supplementary Table 2-8) that are regulated by TF Rap1p under both AAS and GS(Figure 2-8A, B), which has been shown to control the expression of ribosomal genes (Morse, 2000). We identified Spt15p binding sites in the upstream regions of multiple upregulated genes such as *HAP1, CDC55, HFD1*, etc. and downregulated genes such as *LYS20, LYS21, ACS2* etc. (Supplementary Table 2-6) under AAS (Figure 2-8A), and Spt15p is known for binding the TATA box in the promoters of many of these genes (Cormack & Struhl, 1992). We found binding sites of TF Sfp1p in upstream regions of upregulated genes (*JLP1, SUL1,* etc.) under AAS (Figure 2-8A). Interestingly, two of the inferred target genes of Sfp1p, *SUL1* and *JLP1* (Figure 2-8A), both are involved in the uptake of Sulfur, have not been documented in previous research, thus, might be novel findings. Interestingly, while we inferred that Sfp1p mainly upregulated genes under AAS (Figure 2-8A), it mainly down-regulate genes under GS (Figure 2-8B). We show that *JEN*1 might be a target gene of Gat4p under GS, although it has been reported that *JEN1* was

indirectly regulated by Gat4p (Chua et al., 2006). We also found that *GRX3*, coding a glutathione-dependent oxidoreductase that protects the cell from oxidative damage (Pujol-Carrion, Belli, Herrero, Nogues, & de la Torre-Ruiz, 2006), and *ATG42* (*YBR139W*), coding a vacuolar serine-type carboxypeptidase involved in the proteolytic processing and final steps of autophagy in yeast(Parzych, Ariosa, Mari, & Klionsky, 2018), might be novel target genes of GAT4. While we were able to confirm that Pdr1p and Pdr3p were regulators of *FOX2* under GS (Figure 2-8B) (Salin et al., 2008), we also inferred that Ert1p might be a novel regulator of *FOX2*, sharing a similar binding site as Pdr3p (Supplementary Table 2-7).

Interestingly, some TFs are predicted to upregulate some genes and downregulate some other genes under the same condition. We thus examined whether the motifs found in upregulated genes and downregulated genes differed for such TFs. As shown in Supplementary Table 2-10, most such TFs bind similar binding sites of the same motif in the upstream regions of both upregulated genes and downregulated genes. The different regulatory effects might have resulted from different locations of the binding sites and/or with different cooperative TFs (Latchman, 2001; Scully et al., 2000). However, Gcn4p recognizes quite distinct two motifs in the upstream regions of upregulated genes (Figure 2-8) and downregulated genes (Figure 2-8E) under AAS, suggesting different regulatory machineries of Gcn4p for upregulated genes than for down-regulated genes involved in amino-acid synthesis, a phenomenon that has been previously reported (Badis et al., 2009; Scully et al., 2000).

Furthermore, we attempted to corroborate the putative regulatory relationships

through a correlation analysis between the TFs and predicted target DEGs. In brief, the mean correlation between a TF and its predicted targets is tested for significance through a permutation test, see Methods section. The results showed that the expression of some TFs did not significantly correlate with its targets, indicating post-transcriptional regulation mechanisms of these TFs under stress, while some exhibit significant correlation, indicating regulation of those TFs at a transcriptional level. Notably, Met31p/Met32p/Cbf1p are known to be regulators for over 4,000 genes in Yeastract, however a permutation test finds all the three TFs significantly correlated with predicted targets under AAS (Figure 2-9A). Correlation between Met4p and its targets was also found to be significantly correlated with its DEG targets. These results seem to indicate that the response to AAS induces a transcriptional change in TFs that govern methionine biosynthetic genes. Gcn4p on the other hand was not shown to be significantly correlated with its predicted target DEGs, this reaffirms earlier studies showing Gcn4p to be controlled in a translational manner (16). Interestingly, Gcn4p is shown to be significantly correlated with its target TFs under GS. GS is also known to induce Gcn4p activation via a translational mechanism through Gcn2p, however specifics of the process differ from that under AAS (Yang, Wek, & Wek, 2000). Our results indicate that Gcn4p may also be transcriptionally regulated in response to GS (Figure 2-9B).

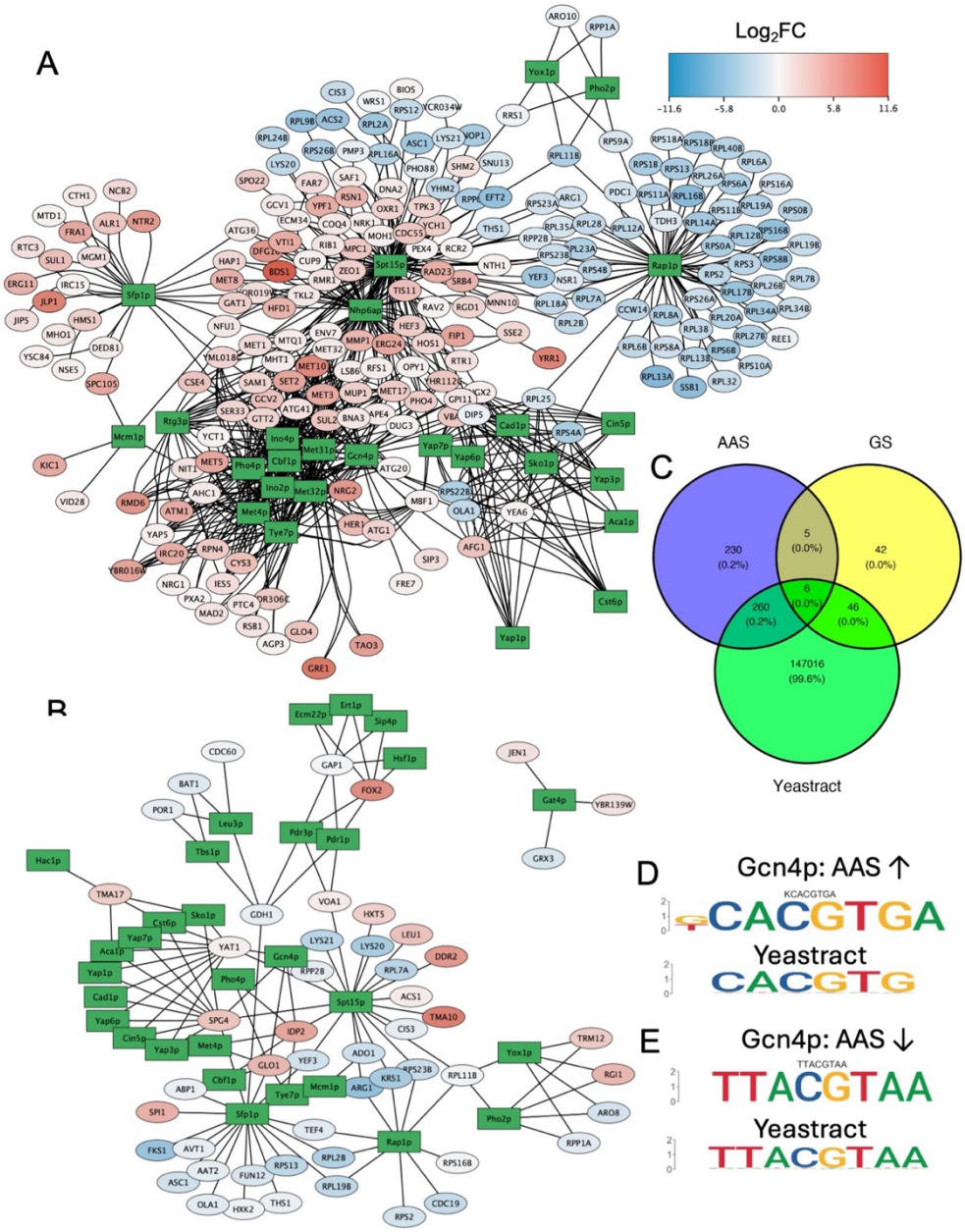


Figure 2-8 Inferred gene regulatory networks for stress-induced transcriptional changes. A. Putative gene regulatory networks for the DEGs under AAS. **B.** Putative gene regulatory networks for the DEGs under GS. TF genes are shown as green rectangles and non-TF genes as ovals. A TF and its inferred target gene are connected by an edge. Upregulated genes are colored red and downregulated genes blue. C. Venn diagram of predicted TF-gene relationships and known relationships on the Yeastract Database. D, E. Different motif usages of Gcn4p in AAS for upregulated genes (D) and for downregulated genes (E).

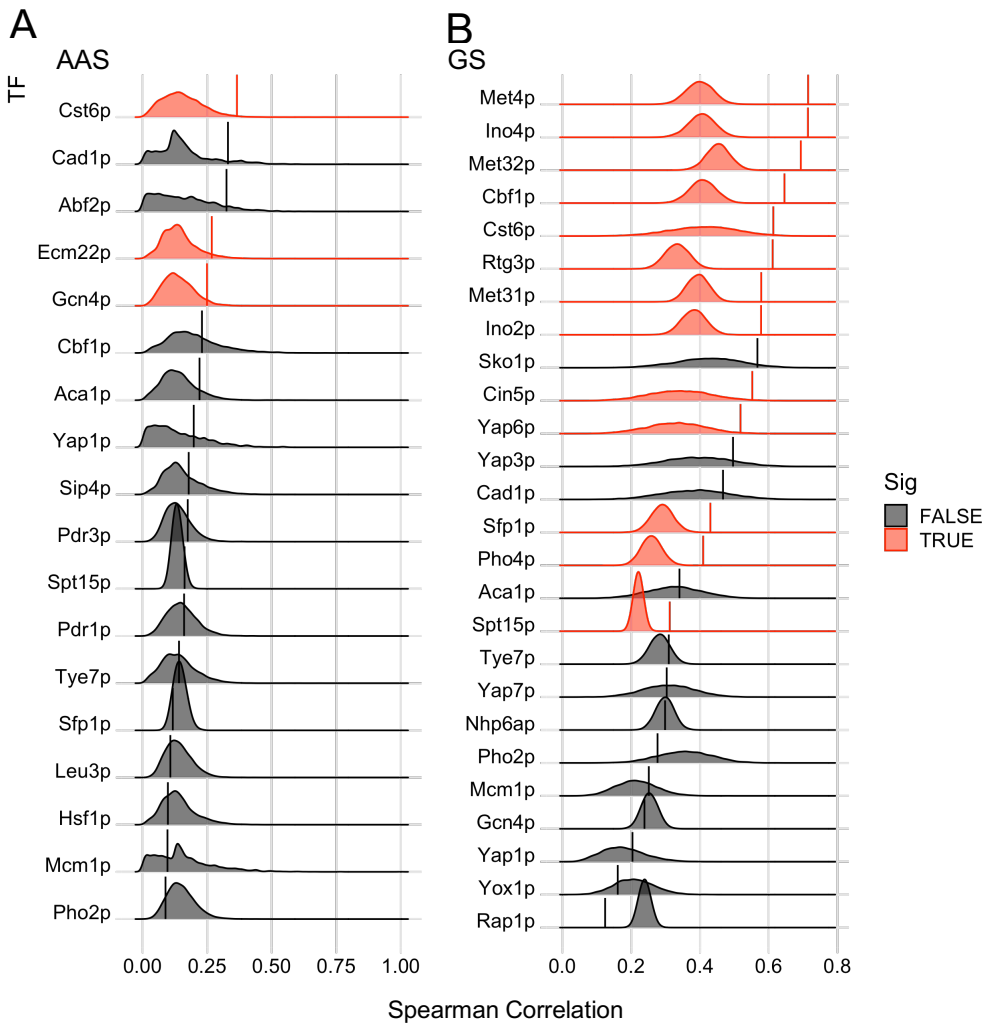


Figure 2-9 Spearman correlation significance test between predicted target DEGs and TFs. A. AAS TF mean spearman correlation with predicted target DEGs vs random gene sets of the same size. B. GS TF mean spearman correlation with predicted target DEGs vs random gene sets of the same size. Each ridge-plot shows the distribution of the mean spearman correlation between each TF and randomly sampled gene sets. The vertical bar is the mean spearman correlation between each TF and its predicted DEG targets. The distribution and vertical bars are colored red if the mean spearman correlation between the TF and predicted DEG targets is significantly (alpha = 0.05) greater than expected.

2.3.5 TFs contribute to cellular variability in a post-transcriptional manner

We also seek to elucidate whether the variability observed between single-cell transcriptomes of the same condition can be attributed to variation of transcriptional

regulatory machineries such as variation of TF expression levels. Our single cell study naturally provides such an advantage over the conventional population-based study, in which minute differences in transcriptional regulation are obfuscated by the pooling of thousands of cells. Using cells under the same conditions (AAS, Hypotonic and Isotonic Batch 1 for their better quality), we performed hierarchical clustering (see Methods.) of genes that are expressed in at least 10% of all cells. Though not all clusters showed significant enrichment for regulating TFs, many clusters were able to enrich for TFs, indicating genes associated with the same TFs exhibit similar patterns of variation under the same conditions, suggesting biological variability between cells under the same conditions is due to the transient differences in TF activation states, and such variabilities can possibly be utilized to discover regulatory relationships. (Figure A-1, Figure A-2, Figure A-3). Interestingly, the enriched regulating TFs are rarely present in their respective clusters, likely due to delayed effects of TF induced activity and target genes expression change, and post-transcriptional activation of TFs.

2.4. Discussion

Recently, scRNA-seq has become a powerful tool to address important biological problems including cell type identification, understanding mechanisms of gene transcriptional regulation and characterization of functionally related genes at the single cell level. Despite many applications of scRNA-seq to the most well-studied eukaryotic organism, scRNA-seq applied to the budding yeast are rare, likely due to the more technical demands associated with a tough cell wall and very low mRNA contents when preparing

scRNA-seq libraries for a yeast cell. In this study we overcame the technical obstacles by adapting an earlier full-length scRNA-seq method (Tang, Barbacioru, Bao, et al., 2010; Tang, Barbacioru, Nordman, et al., 2010), and profiled the transcriptomes of single yeast cells in isotonic, nutrients-rich medium and under three stress treatments hypotonic, AAS and GS. However, to remove the cell walls, we inevitably exert stresses to the cells by exposing them to water and softening medium to facilitate subsequent enzymatic cell wall digestion. To minimize the impacts on the cells, we equilibrated the cells in the isotonic, nutrients-rich spheroplast medium at 30°C for 60 min when digesting the cell with Zymolyase 100T, followed by 2X 5-min washes of the cells in the spheroplast medium. The transcriptional responses of the cells under subsequent stress treatments are like those seen in previous studies, indicating that our cell wall removal procedure had little effects on the cells.

Though scRNA-seq improves upon previous transcriptomics quantification methods, cell-to-cell variation may be confounded by factors such as cell size, state and technical factors that arise during sample preparation. The scRNA-seq libraries in this study exhibited varying degrees of technical variation and batch effects. Specifically, there exists significant technical differences and batch effects between our batch 2 and batch 1 samples, as shown in both QC metrics and correlation with previous studies. The lower quality of samples in batch 2 could be attributed to various technical factors. Therefore, a careful quality control is critical before any formal analysis to minimize the effects of technical variability. To this end, we filtered out low quality libraries using a stringent QC procedure based on five metrics, and only performed comparisons between transcriptome

collected from the same batch. We show that this QC procedure was able to not only enhance the reliability of our analysis but also yield more biologically meaningful results.

Another critical aspect of inference from scRNA-seq data is the depth of read coverage, which has an effect on the number of genes that can be detected as has been noted in cell population based libraries as well (Tarazona, Garcia-Alcalde, Dopazo, Ferrer, & Conesa, 2011). To evaluate the effect of read depth, we prepared a series of diluted bulk RNA libraries, starting from 5 pg which is near the lower bound of mRNA levels in small single cells. From the saturation analysis using the varying number of reads randomly sampled from the 5 pg bulk library (Figure 4-1), we deduced that a minimal sequencing depth of 4×10^6 reads is required to detect almost all genes that can be quantitatively characterized at this input quantity of mRNA. Cells under a certain treatment exhibited higher correlation of transcription levels, while different treatments led to distinct transcription of relevant genes in response to the treatments. This indicates that transcriptome analysis at the single-cell level is biologically meaningful regarding stress conditions.

Our observation that the isotonic and hypotonic cells cannot be clearly differentiated using their transcriptomes is consistent with a previous study using microarray to characterize gene expression on diverse environmental transitions, which found that when cells were transferred from standard isotonic to hypotonic solution, the change in the expression of the genes involved in environmental stress response is only transient (Gasch et al., 2000). Consequently, our DEG analysis was not able to find significantly differentially regulated genes, suggesting that gene expression levels were

indeed back to isotonic conditions.

As expected, individual cells under the same treatment show varying levels of variation due to intrinsic biological noise and unavoidable technical noise (Brennecke et al., 2013), though the latter has been largely reduced by our QC procedure. However, both PCA/UMAP and DEG analyses revealed larger variability among single cells under AAS vs isotonic and GS vs isotonic treatments. Under AAS stress, we observed significant down-regulation of translation related genes such as those encoding ribosomal proteins, consistent with the earlier findings (Gasch et al., 2000), where ribosomal genes were found to be repressed during general stress response. More genes were found to be upregulated, particularly, those involved in the methionine and cysteine biosynthetic pathways. Previous studies have shown that the AAS-induced response in yeast cell could be activated under the depletion of even a single type of amino acid(Hinnebusch, 1986, 1997). In this study, the yeast cells were transferred to a YPD medium without any type of amino acids. Our results show that when depleted of all amino-acids, there was a very strong transcriptional response of genes responsible for the synthesis of Sulfur-containing amino acids. In this regard, methionine, one of the two amino acids containing Sulfur, has been found to be an integral part of signaling pathways involved in the inhibition of autophagy, the regulation of tRNA thiolation that controls overall metabolic state, and cell proliferation (Sutter, Wu, Laxman, & Tu, 2013; Walvekar et al., 2018). Beyond its role as the initiation amino acid, methionine is also involved in increasing translation capacity through controlling upstream regulators in the TORC pathway (Sutter et al., 2013). Our finding that the DEGs under AAS stress are significantly enriched for pathways involved in Sulfur metabolism and

Sulfur-containing amino acid biosynthesis, provides a different perspective to the earlier results. It seems that under the limitation of all amino acids, apart from a drastic reduction in ribosomal and translation activity, the yeast cells preferentially synthesize methionine first as a mechanism to induce the synthesis of other amino acids and prepare for the restart of anabolism once conditions are viable again.

Under the GS stress, there is also a significant down-regulation of translation and ribosomal encoding genes, consistent with early findings (Causton et al., 2001). Reduced expression was also found in genes involved in the repression of other genes in the presence of glucose. Genes involved in biosynthesis of multiple amino acids were also found to be downregulated. The upregulated genes are mostly involved in the general stress response and processes related to the utilization of alternate carbon resources such as gluconeogenesis, hexose transporters and energy metabolism (Gasch et al., 2017; Nadal-Ribelles et al., 2019; Piper et al., 2000; Walvekar et al., 2018; J. Wang et al., 2022).

By identifying putative TFBSs in the upstream regions of the DEGs under AAS and GS, we were able to confirm known TF-target gene relationships as well as potential novel ones. Modes of regulation also seems to be different for shared TFs between different stresses, such as the case for Sfp1p, which is mainly associated with upregulated genes under AAS and downregulated genes under GS.

Enrichment of TFs in clusters of expressed genes was able to provide an explanation for the observed cellular variability between cells of the same condition. We showed that, with samples of decent quality, it is possible to identify gene modules regulated by the same TF using only intrinsic biological variability without the introduction

of perturbation. This is likely a result of stochastic variation in TF activation differences between cells of the same condition and the propagation of such variation or noise in gene regulatory networks (Pedraza & van Oudenaarden, 2005). This finding provides further credence(Pedraza & van Oudenaarden, 2005; Stewart-Ornstein, Weissman, & El-Samad, 2012), supporting the idea of using cellular variability/noise to study regulatory mechanisms.

CHAPTER 3 TRANSCRIPTOMIC ANALYSIS OF THE SPATIOTEMPORAL AXIS OF OOGENESIS AND FERTILIZATION IN *C. ELEGANS*

3.1 Introduction

With a transparent body of less than 1,000 somatic cells, a fully sequenced genome harboring 19,985 protein-coding genes (WS291 annotation), and about 14 hours of embryogenesis time and two weeks of life span, the *C. elegans* hermaphrodite provides an extraordinary model to understand various types of cell differentiation and organogenesis (Chu & Shakes, 2013; Consortium, 1998; Hillier et al., 2005; S. Kim, Spike, & Greenstein, 2013; Labouesse & Mango, 1999; Marcello, Singaravelu, & Singson, 2013; Robertson & Lin, 2013; Rose & Kemphues, 1998; Sulston & Horvitz, 1977; Sulston, Schierenberg, White, & Thomson, 1983; Wood & Edgar, 1994). Particularly, *C. elegans* gonad provides an excellent model to understand meiosis (Pazdernik & Schedl, 2013), gamete formation (Chu & Shakes, 2013; S. Kim et al., 2013) and fertilization (Marcello et al., 2013).

In the *C. elegans* hermaphrodite germline, oogenesis occurs independently in two sets of U-shaped gonads connected to a single shared uterus (Pazdernik & Schedl, 2013). Oocyte formation begins at the distal end of each gonad with mitotically proliferating germline stem cells near the single somatic distal tip cell (DTC). Proliferating germ cells away from the DTC begin to enter meiosis prophase I through a transition zone, after which germ cells move along the gonad while going through the pachytene, diplotene and diakinesis stages ending in the most proximal (-1) oocytes that awaits fertilization in the spermathecae for progression into metaphase I and the subsequent formation of the zygote. Apart from the proximal oocytes in diakinesis, most of the germline nuclei do not have fully enclosed membranes and form a syncytium, sharing a nucleus free cytoplasmic region

called the rachis, which facilitates the transport of RNAs and proteins to growing oocytes. Throughout this process, the germline also is enveloped by five pairs of gonadal sheath cells (Sh1-Sh5 from distal to proximal), each pair serving distinct functions through communication with the germline and promoting the oogenesis program (Pazdernik & Schedl, 2013).

However, for a long time this system is limited by its miniscule size, preventing a detailed dissection of the biochemistry in each part of the oocyte assembly line using techniques such as transcriptome profiling using microarray (Baugh, Hill, Slonim, Brown, & Hunter, 2003; Reinke, 2002; Walhout et al., 2002) or bulk-RNA sequencing (RNA-seq)(Gerstein et al., 2010; J. J. Li, Huang, Bickel, & Brenner, 2014; Spencer et al., 2011), and proteome profiling using mass spectrometry (Yuet et al., 2015), as all of these techniques require a descent quantity of RNA/protein from at least hundreds of thousand cells.

Recent studies have performed micro-dissections of the *C. elegans* gonad and profiled transcriptomes of the gonad segments using single-cell RNA-seq (scRNA-seq) techniques (Diag, Schilling, Klironomos, Ayoub, & Rajewsky, 2018). However, these analyses mainly focused on the post-transcriptional/translational regulation of germline transcripts via binding of 3'UTRs to RNA binding proteins and miRNAs. Although these studies provided expression estimates for genes from each segment as well, they did not focus on other aspects of transcriptomic changes between the segments that might also account for the progress of oogenesis. Consequently, we still lack a good understanding of

the machinery of the assembly line, such as key regulators and gene expression patterns along the temporal and spatial axis of the gonad.

To fill these gaps, we combined microdissection with scRNA-seq technique (Picelli, Bjorklund, et al., 2013; Ramskold et al., 2012; Tang, Barbacioru, Bao, et al., 2010; Tang, Barbacioru, Nordman, et al., 2010), and profiled the transcriptomes in the proliferative zone, pachytene zone, diplotene zone, early diakinesis zone (before -3 oocyte stage), later diakinesis zone (-3, -2, -1 oocytes), and the zygote. Our results revealed a highly dynamic picture of gene transcriptional regulation at each transitional time point throughout the oocyte assembly line. These results should provide a foundation to further understanding the molecular mechanisms of the oogenesis and fertilization processes.

3.2 Materials and Methods

3.2.1 Experimental Model

The AZ212 *C. elegans* strain was obtained from the *C. elegans* Genetics Center (University of Minnesota), and was maintained in E. coli OP50 lawn on an agar plate according to the standard protocol (Stiernagle, 2006).

3.2.2 Dissection of the gonad and harvest of samples

After a well-fed gravid hermaphrodite was immobilized in the egg salt solution (ESS) with 10% tetramisole (Sigma, St. Louis), a cut was made across the vulva using a 26G subcutaneous needle controlled by a micromanipulator (ROE-200, Sutter) under an inverted microscope (Olympus 1X71). This would release fertilized eggs and early-stage embryos from the uterus as well as sperms and at least portions of the two sides of the gonad. Each end of the gonad wrapped around by five pairs of sheath cells was completely

isolated by pushing its distal end as shown in Figure 3-1A. The distal proliferative zone (S1) was cut off around the transition zone (Figure 3-1A) and harvested in about 10 nl ESS by suction using a patch clamp pipette under controlled of another micromanipulator (ROE-200, Sutter). The pachytene zone (S2), the loop corresponding to the diplotene zone (S3) and diakinesis zone (S4) were sequentially cut off at the positions as shown in Figure 3-1A and similarly harvested. The -3 (F3), -2 (F2), and -1 (F1) oocytes were also isolated by cutting through their boundaries and similarly harvested (Figure 3-1A). The zygote (fertilized oocyte) also known as P0 was similarly harvested when the two pronuclei were fused at its center (Figure 3-1A). Unavoidably, sheath cells wrapped around the gonad segments and oocytes as well as released sperms could be harvested in the samples.

3.2.3 Preparation of RNA-seq libraries

We prepare a RNA-seq library for each harvested sample for Illumina platforms using a modified scRNA-seq method based on Tang et al as previously described (Su, Xu, Shea, Destephanis, & Su, 2023; Tang, Barbacioru, Bao, et al., 2010; Tang, Barbacioru, Nordman, et al., 2010) at the earlier stage of the project and using the Smart-seq2 protocol (Picelli, Bjorklund, et al., 2013) later on. The libraries were sequenced by 100 bp paired end reads on an Illumina HiSeq2000 or HiSeq2500 machine.

3.2.4 Transcriptome Mapping and Quantification

The *C. elegans* genome assembly (GCA_000002985.3) was obtained from NCBI Refseq, while the annotations were based on Wormbase version: WS291. Prior to mapping, raw reads were trimmed with Trim Galore(Krueger, 2015), with parameters (quality >= 10, length > 35bp). We quantified the expression levels of genes in two ways for different

subsequent analysis. For differential gene expression analysis, trimmed reads were mapped to the genome using HISAT2(D. Kim, Paggi, Park, Bennett, & Salzberg, 2019) with default settings, read counts were obtained by using HTSeq (S. Anders, Pyl, & Huber, 2015) with default settings based on the mapping results. The trimmed reads were also mapped to the genome using Salmon (Patro, Duggal, Love, Irizarry, & Kingsford, 2017) with default settings to obtain transcript per million (TPM) estimates for both genes and transcripts.

3.2.5 Quality Control

Sequenced libraries were then assessed for quality with custom scripts and quality metrics evaluated via the QoRTs package(Hartley & Mullikin, 2015). First, Salmon quantified TPM values for mitochondrial genes, spike-ins and sperm specific genes (See also, Figure B-1, Supplementary Table 3-1) were obtained. A sample was filtered out if it met any of the following criteria: i) over 5% reads (in terms of TPM) were from the mitochondrial genome; ii) over 5% reads (in terms of TPM) were from rRNA genes; iii) over 5% reads (in terms of TPM) were from sperm specific genes); iv) over 5% reads (in terms of TPM) were from intestine specific genes; v) HISAT unique reads mapping rate < 70%; vi) less than 50% of HISAT uniquely mapped reads were mapped to coding DNA sequences (detailed procedure and genes that were removed and used for filtering are shown in Figure B-1 and Supplementary Table 3-1). These criteria were set to remove samples that were of poor libraries quality or were heavily contaminated by sperm, intestinal tissue and/or exhibited reduced quality during sample collection. To further increase the robustness of subsequent analysis, samples were visualized using Uniform Manifold Approximation and Projection (UMAP), and those that largely deviated from

clustered groups of the same sample type were removed. In addition, we included the P0 (1-cell) samples of Tintori, et al (Tintori, Osborne Nishimura, Golden, Lieb, & Goldstein, 2016) in our analysis, and the samples were processed through the same pipeline as our own samples.

3.2.6 Comparison with Previous Datasets

Gene expression data from six previous studies were collected from the following sources and compared with our data. For all comparisons, we used filtered samples with all genes (genes were not filtered). Details of the datasets and comparisons are as follows:

1) Reinke et al. 2004 (Reinke et al., 2004) provided the first microarray-based list of oogenic genes. The list was retrieved from via their supplementary material.
2) Ortiz et al. 2014 (Ortiz et al., 2014) performed RNA-seq analysis on the gonad to distill a list of genes termed oogenic. These genes were acquired via their supplementary data, and genes marked oogenic were used for our subsequent comparisons.
3) Stoeckius et al. 2014 (Stoeckius et al., 2014) performed RNA-seq on proximal oocytes and 1 cell zygotes. Expression profiles were acquired via the instructions in their paper and genes with expression > 0.5 RPKM were deemed expressed.
4) West et al, 2018 (West et al., 2018) dissected the gonad into mitotic and meiotic sections, and oocytes. RNA-seq data of each sample was acquired via the supplementary material of the paper, and genes with a reads count > 0 were deemed expressed.
5) Tzur et al. 2018 (Tzur et al., 2018) utilized the Cel-seq protocol to sequence 10

segments of the *C. elegans* gonad, with 2 replicates per segment. Alignment of these 10 segments to our segments was based on diagrams presented in their study and rough estimates of where their dissection occurred. The exact alignments between their segments and ours are given in Table 3-1. Count matrices were acquired per the authors' instructions. Pearson correlation was performed with log transformed count values using all shared genes.

6) Diag et al, 2018 (Diag et al., 2018) performed cryo-dissection of the 3 posterior and 3 anterior gonads into 13-15 segments per gonad. This resulted in 85 slices sequenced via Cel-seq. Expression profiles for these samples were retrieved from GEO with accession number GSE115884. Samples with $< 10^4$ reads were discarded from correlation analysis with our samples. The authors (Diag et al., 2018) provided approximate slice label, slice size as well estimates size of each gonad region. Thus, we were able to derive a coarse conversion from their slices to our segments, as shown in Table 3-1. Pearson correlation was performed with log transformed count values using all shared genes.

3.2.7 Differential Gene Expression Analysis

We performed differential gene expression analysis between each two dissected neighboring stages along the developmental axis of the gonads as above-described and zygotes using Monocle2 (Qiu et al., 2017). Experimental batch and gene detection rate in each sample were included as covariates along with segment/cell-type to model normalized gene expression using the negbinomial.size model of Monocle2. Because Monocle2 does not produce Log2FoldChange values, we applied Bayesian shrinkage of gene model

coefficients using the apeglm (Zhu, Ibrahim, & Love, 2019) package to account for large fold-change values of genes with low expression and obtain shrunken Log_2FC values for each gene. A model of gene expression as a function of segment/cell-type was also fit to assess genes that were differentially expressed across all stages prior to fertilization (excluding P0). Genes with an FDR < 0.05 and a fold change increase/decrease of 1.5 were considered differentially expressed. ClusterProfiler(Yu et al., 2012) was used to perform Gene Set Enrichment Analysis (GSEA) with pre-ranked shrunken Log2FC values and gene sets from KEGG (Kanehisa et al., 2022), GO(Gene Ontology, 2021) Biological Pathways, Reactome (Milacic et al., 2024) and Wikipathways(Martens et al., 2021). Enrichment of each type of gene sets was performed separately, and the results were aggregated. Only gene sets containing more than 10 and less than 250 genes were considered, and those with an FDR < 0.05 were considered significantly enriched.

3.2.8 Clustering co-expressed genes

The union of DEGs identified in all pairwise comparisons were used for gene co-expression analysis. After the read count values of genes were variance stabilizing transformed using the vstExprs function of Monocle2 package, Pearson correlation coefficient between expression levels of the genes in the samples were calculated, and genes were hierarchically clustered using the ‘ward.D2’ method of the hclust function in R. Upon visual inspection of the resulting clustering heatmaps, the clusters were set at a hierarchical level. Each cluster was then subject to enrichment analysis for GO biological process (BP) terms using ClusterProfiler (Yu et al., 2012) to identify significantly enriched terms for the cluster. Gene expression as well as the respective clusters were visualized

with the ComplexHeatmap package(Gu, 2022), and the top three most significantly (fdr < 0.05 or p-value < 0.001) enriched GO terms were shown alongside the heatmap.

3.2.9 Differential Alternative Polyadenylation Analysis

3'UTR regions were extracted from the WS291 annotation via custom scripts to only include 3'UTR regions that did not overlap coding exons and other UTR regions. The samtools (H. Li et al., 2009) depth function was used to obtain pair-read aware coverage of the genome for each samples with HISAT2(D. Kim et al., 2019) aligned bam files. Coverage for each sample was normalized with DESeq2(Michael I Love et al., 2014) size factors before estimation of polyadenylation site and long/short 3'UTR coverage, and Percentage of Distal polyA site Usage Index (PDUI) was computed performed using DaPars2(Feng, Li, Wagner, & Li, 2018; L. Li et al., 2021). Modification to the DaPars2 program was made to begin polyadenylation site search starting from 25bp downstream of 3'UTR's 5' end. For each neighboring stages comparison, only 3'UTRs that belonged to a gene with a mean count > 10 across all compared samples and had PDUI values in at least 3 samples in both stages were tested for differential alternative polyadenylation. Fisher's exact test was performed with the average long/short 3'UTR coverage in compared stages, and the resulting p-values were corrected for false discovery rate (FDR) via the Benjamini Hochberg method. Genes that had FDR < 0.05 and |PDUI difference| > 0.05 were called for significantly differential alternative polyadenylation. ClusterProfiler(Yu et al., 2012) was used to perform GO BP (Ashburner et al., 2000) term enrichment analysis, and significant terms with FDR < 0.05 were called significantly enriched. Visualization was made with the trackViewer(Ou & Zhu, 2019) R package.

3.2.10 Differential Splicing Analysis

Differential splicing analysis was performed using rMATS (S. Shen et al., 2014) that calculated splicing Psi values and evaluated their statistical significance. rMATS classifies splicing events into five categories: alternative 3' splice site (A3SS), alternative 5' splice site (A5SS), intron retention (IR), mixed exon usage (MXE), skipped exon usage (SE). A splicing event with a Psi value change > 0.1 and an adjusted p-value < 0.05 was considered significant.

3.3 Results

3.3.1 Expression levels of detected genes correlate well with those from previous studies

We cut each isolated gonad into seven segments roughly corresponding to the stages of oocyte development (Figure 3-1A) (Materials and Methods), and the number of samples collected for each segment, oocyte and the zygote are shown in Figure 3-1B. To assess the quality of our RNA-seq libraries, we evaluated the similarity between the detected genes and their expression values and those from six previous studies (Diag et al., 2018; Ortiz et al., 2014; Reinke et al., 2004; Stoeckius et al., 2014; Tzur et al., 2018; West et al., 2018) (Materials and Methods). Four (Ortiz et al., 2014; Reinke et al., 2004; Stoeckius et al., 2014; West et al., 2018) of these studies largely quantified expression levels in entire gonads or large sections of the gonad, thus we aggregated gene expression in corresponding samples to allow reasonable comparisons. Our aggregated expression profiles recall over 90% of expressed genes in all the four datasets (Reinke et al., 99%;

Ortiz et al., 97%; West et al., 93% mitotic, 95% meiotic and 96% proximal oocyte; Stoeckius et al., 96% oocyte and 94% zygote) (Figure 3-1C).

Two of these studies (Diag et al., 2018; Tzur et al., 2018) dissected the *C. elegans* gonad into multiple segments and profiled the transcriptome of each segments using a variety of techniques including RNA-seq. As both studies cut the gonad in more segments than we did, we aggregated data from the segments of (Tzur et al., 2018) and (Diag et al., 2018) according to the alignments of the segments (Materials and Methods, Table 3-1), so that data from largely the same segments as ours were compared. Our detected genes in each segments and oocytes recall most of detected genes in the corresponding aggregated segments by (Tzur et al., 2018) and (Diag et al., 2018) (Figure 3-1C). Moreover, the expression levels of genes in our segments are largely correlated with those in the corresponding aggregated segments in the two prior studies (Figure 3-1D, E). These results suggest that we have largely correctly align the segments in both studies to ours. However, notably, our detected genes have higher recall rates for (Figure 3-1C) and higher correlation coefficients with (Figure 3-1D, E) those of (Tzur et al., 2018) than for and with those of (Diag et al., 2018). This might be due to the more similarity between our segments and those of (Tzur et al., 2018) than between our segments and those of (Diag et al., 2018). Taken together, these results suggest that our detected genes are largely consistent with those detected by previous studies.

Table 3-1 Correspondence between or gonad stages and that of (Diag et al., 2018) and (Tzur et al., 2018).

Our segments	**Tzur et al, 2018(Tzur et al., 2018)**	**Diag et al, 2018(Diag et al., 2018)**
S1	Segments 1,2	Segments 1,2
S2	Segments 3,4,5	Segments 4, 5, 6, 7
S3	Segments 6,7	Segment 8
S4	Segment 8	Segments 10,11
F3	Segment 9	Segment 12
F2	Segment 9	Segment 13
F1	Segment 10	Segments 14, 15
P0	NA	NA

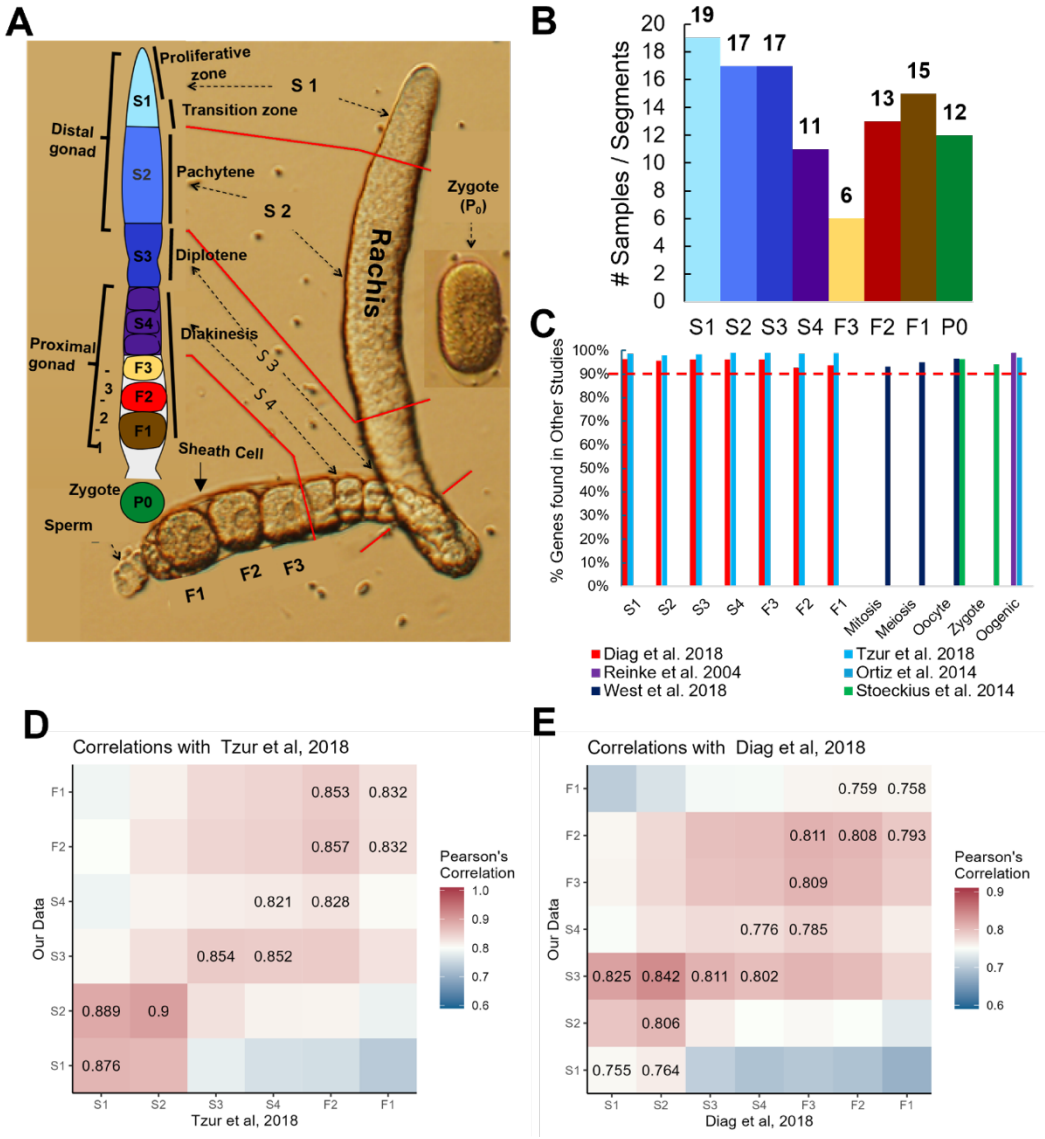


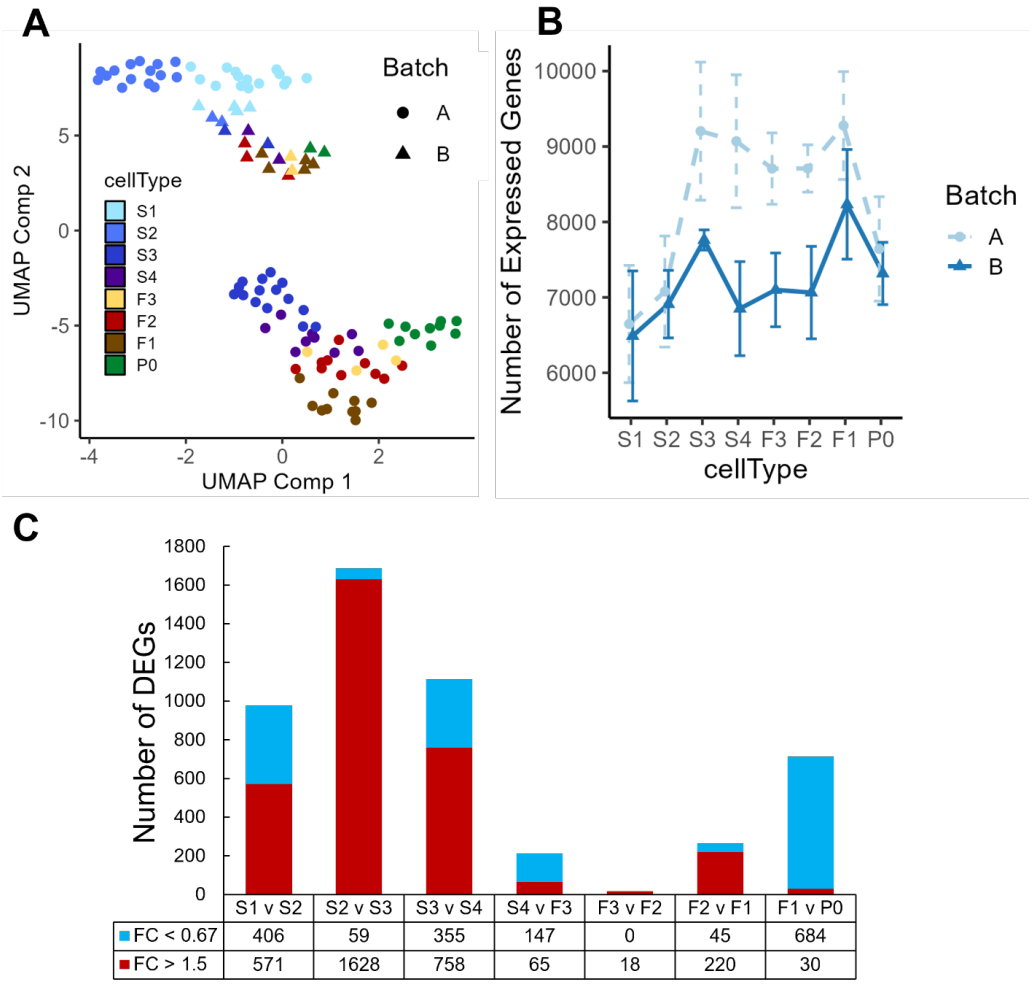
Figure 3-1 Comparison of our datasets with existing ones. A. A diagram of an isolated one side gonad together with a cartoon of one side gonad showing the dissection positions for the segments along the one side gonad. B. Number of samples from each stage of segments, oocytes and zygotes. C. Percentage of genes found expressed in each stage in previous studies that we found expressed in our study. D, E. Heatmap of Pearson correlation coefficient of our detected expressed genes in the segments with those of (Tzur et al., 2018) (D) and of (Diag et al., 2018) (E).

3.3.2 Differential gene expression occurs in early stages of oogenesis and mostly in proximal oocytes.

Dissection of the *C. elegans* hermaphrodite gonad is a delicate procedure that is prone to contamination from neighboring tissues due to the miniscule size of the gonad and proximity of neighboring cells such as sheath cells, intestine cell and released sperms. To mitigate the effects of such contaminations, we filtered sperm-, intestine- and stress-related genes as well as heavily contaminated samples (Materials and Methods, Figure B-1, Supplementary Table 3-1). Mover, the hermaphrodite gonad itself also contains somatic cells, most notably the five pairs of gonadal sheath cells that tightly enclose the germline. The sheath cells mainly function to provide germline maturation signals, move germ cells along the rachis and push proximal oocytes into the spermathecae. Due to the tight interactions between the sheath cells and the germline, complete removal of these cells was very difficult, especially for earlier stages (S1-S4). Thus, some differential gene expression results for these early stages are inevitably due to differences between gonadal sheath cells, albeit they seem to negligibly affect our results for these stages of comparison as described below.

We inspected the relationships among our samples via UMAP visualizations. As shown in (Figure 3-2A), the samples form into two distinct clusters, indicating strong batch effects in our datasets possibly due to the two different scRNA-seq library preparation protocols used at different stages of the project (Materials and Methods). Nonetheless, a trajectory from S1 samples to F1 and zygote samples is formed in both batches, which is in line with the developmental path of the germline. Thus, we account for batch effects in

subsequent analysis when possible. Inspection of the number of genes expressed in each segment/cell type shows a clear pattern, i.e., the number of expressed genes increased from S1 to S3, before dropping slightly in S4 and exhibiting only minor changes before another increase in the -1 oocyte (F1) and finally a large decrease in the fertilized oocyte (Figure 3-2B). Therefore, it appears that gene transcriptional regulation mostly occurs in early stages of oogenesis, particularly between the S2 (pachytene) and S3 (diplotene) transition and becomes progressively quieter as the oocyte goes through the S4, F3 and the F2 stages (Figure 3-2B). Gene transcriptions are reactivated in the F1 oocytes, probably preparing for fertilization (Figure 3-2B). To further reveal gene expression transitions alone the developmental axis of the gonad, we analyzed DEGs between each pair of neighboring stages with the earlier stage as the baseline reference in each comparison (Figure 3-2C). Transition from S2 to S3 invokes the largest number of up-regulated DEGs, and transition from F3 to F2 has the smallest number of DEGs, while fertilization triggers the largest number of downregulated DEGs in the zygotes (Figure 3-2C).



	S1 v S2	S2 v S3	S3 v S4	S4 v F3	F3 v F2	F2 v F1	F1 v P0
FC < 0.67	406	59	355	147	0	45	684
FC > 1.5	571	1628	758	65	18	220	30

Figure 3-2 UMAP display of samples and differential expression analysis of genes between neighboring stages. A. Both batches of samples are clustered according to their positions along the gonad developmental axis by UMAP based on their measured transcriptomes. B. Boxplot of numbers of genes detected in the samples in each developmental stage of the gonad and zygotes. C. Number of upregulated and downregulated genes detected between each pair of neighboring stages, see Supplementary Table 3-2 for details.

3.3.3 DEGs form distinct clusters that are significantly enriched for various functions related to oogenesis.

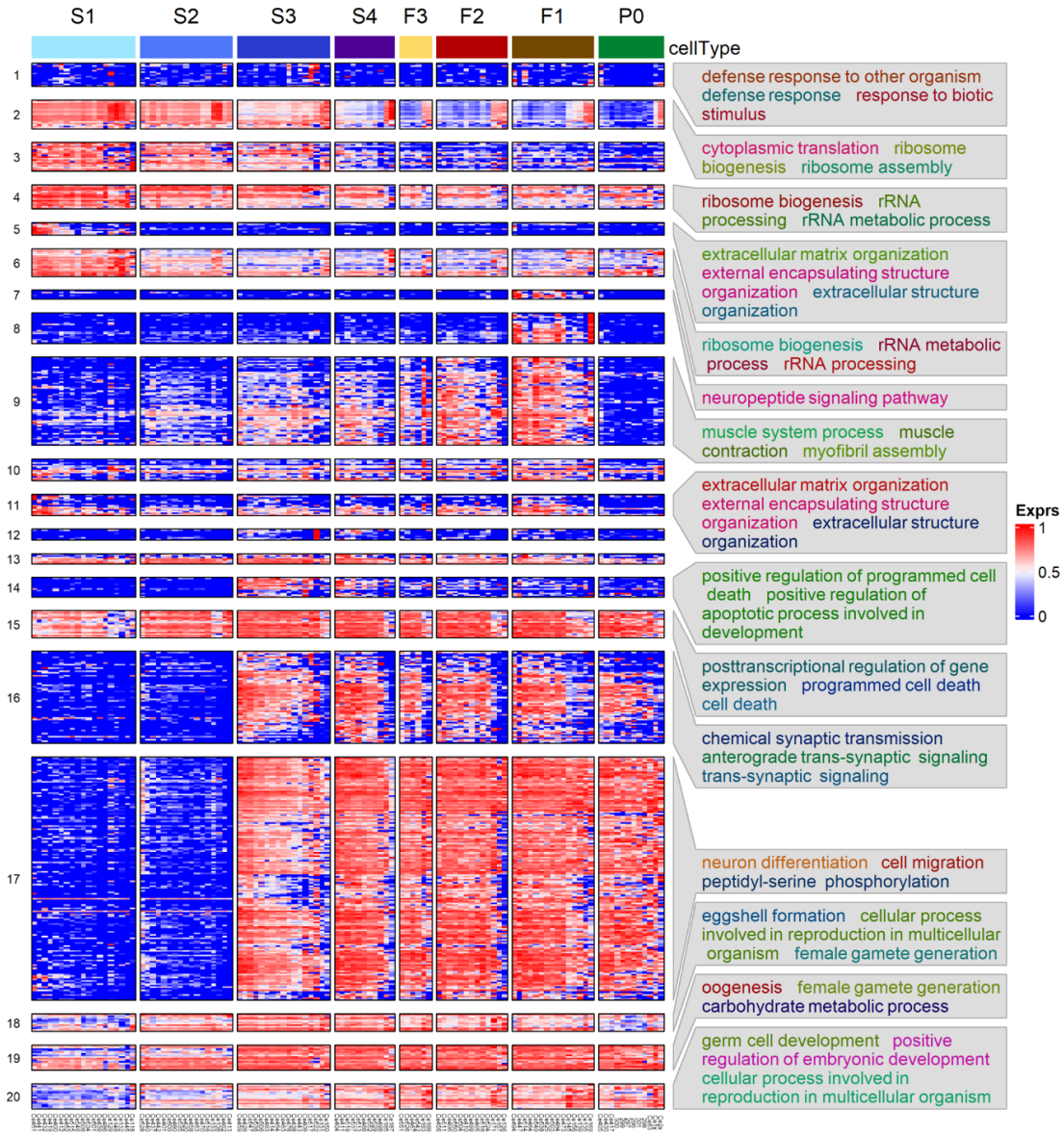
To reveal functional modules underlying the maturation process and fertilization of oocytes, we clustered the union of DEGs identified in all neighboring stages comparisons,

based on their expression levels in all analyzed samples. As shown in Figure 3-3, the DEGs form distinct clusters that are significantly enriched for various functional modules. For instance, clusters 2, 4 and 6 are significantly enriched for ribosomal and translation related processes. All these three clusters of genes exhibited a downregulating trend of expression, albeit with their largest decrease at different stages. Cluster 14 and 15 are enriched for genes involved in programmed cell death, with expression levels elevated in in the S3 stage corresponding to the diplotene loop. However, genes in cluster 14 were quickly downregulated after the S4 stage, while genes in cluster 15 retained similar transcription levels through the subsequent stages. Cluster 18 -20 are all enriched for processes related to oogenesis, e.g., eggshell formation and female gamete generation. Genes in these three clusters exhibited increasing trends of expression from S1 to -1 Proximal oocyte (F1), with the largest increases happening in the early stages (S1-S3). However, genes in cluster 18 experienced reduced expression after fertilization in the zygotes (P0 cells), while genes in cluster 19 and 20 remained similar expression levels. Furthermore, genes in cluster 18 are enriched for eggshell formation, suggesting that transcripts-related to eggshell formation begin degradation post-fertilization after their protein products are no longer needed. Most DEGs belonging to the larger clusters 16 and 17 exhibited similar increases in expression from S2 to S3 and maintained steady levels of expression throughout the later stages even post fertilization. These genes are involved in phosphorylation, synaptic transmission and signaling, positive regulation of transcription, neuronal differentiation, cell fate specification and cell migration. Gene involved in cell migration might be responsible for the mobility of oocytes along the rachis. Interestingly, cluster 17 is strongly enriched for

genes involved in neuronal development, suggesting common functional modules might be used in the differentiation processes of both neurons and oocytes. Cluster 9 is enriched for genes involved in muscle structures and myofibril assembly. As mentioned above, proximal gonadal sheath cells serve the role of pushing oocytes into the spermathecae and require many components like those of muscle cells. Thus, it is highly likely that genes of this cluster originate from contamination of proximal sheath cells wrapped around the proximal oocytes. It is also worth noting that gene expression pattern of cluster 9 differ from those of clusters 16 and 17 in that expression of genes in cluster 9 almost completely disappears in fertilized zygotes, likely due to the absence of sheath cells surrounding the isolated zygotes. Cluster 1 exhibits no obvious pattern of change in expression and the expression levels are generally low. These genes are enriched for defense response related processes and might be required at low levels along the gonad temporospatial axis. Both clusters 5 and 11 are enriched for extracellular matrix organization. It has been shown that many genes (*mig-6, mig-39, lag-2, let-2, epi-1*, etc.) in the two clusters (Supplementary Table 3-3) were preferentially expressed in the distal mitotic regions of the gonad and played roles in extracellular matrix organization and distal tip cell migration (Henderson, Gao, Lambie, & Kimble, 1994; C.-C. Huang et al., 2003; Kawano et al., 2009; Kikuchi et al., 2015). Consistently, expression levels of these genes were elevated in S1.

We also performed GSEA using shrunken Log2FC values of all genes evaluated between each pair of neighboring stages and the results are summarized in Supplementary Table 3-4. Although most of the GSEA results are in accordance with those observed in the gene clustering enrichments (Figure 3-3), surprisingly, GSEA finds upregulated genes

enriched for cell cycle activity, mitosis, transcription, mRNA splicing, mitochondrial translation, and ATP production in the F1 vs P0 comparison (Supplementary Table 3-4).



This suggests that transcriptional activation of cell division and energy production is present in the zygote.

Figure 3-3 Heatmap of hierarchical clustering of DEGs using their transcription levels across the seven stages of oogenesis and zygotes. Enriched GO biological processes in some clusters are shown. See Supplementary Table 3-3 for details.

3.3.4 Possible contaminations of sheath cells in proximal oocytes samples

As mentioned above, there are possible contaminations in the proximal oocytes (F1~F3) samples from surrounding somatic cells. This presents a challenge in deciphering whether expression changes originate from the germline or from the surrounding somatic tissues. As zygotes were often released in the medium once a cut was made across the vulva, and were always collected without obvious objects wrapped around, thus the zygote sample were unlikely contaminated by surrounding somatic cells. Therefore, we postulate that genes that are detected in proximal oocytes (F1~F3) samples but absent in zygote samples (such as those found in clusters 7-9 in Figure 3-3) are likely from contaminating tissues, and find many gene meet this criterion. Most notably, expression of *let-23* and *itr-1* were relatively stable between F2 and F1 prior to dropping significantly in the zygotes, while expression of lin-3 remained high and relatively unchanged between proximal oocytes and zygotes (Figure 3-4A). The contractile activity of sheath cells begins with major sperm protein signals to the proximal oocytes, which in turn produces and releases LIN-3 ligands that are received by the LET-23 receptor on proximal sheath cells (Miller et al., 2001). The LET-23 receptor then triggers signaling inside the sheath cells through PLC-3, which phosphorolyzes IP3 that binds to ITR-1 receptors on the ER, causing the release of calcium(Yin, Gower, Baylis, & Strange, 2004). In addition, sheath cell specific innexin channel encoding genes *inx*-8 and *inx*-9 (Starich, Hall, & Greenstein, 2014) maintained intermediate expression levels in S1~S4 stages, and were highly upregulated in proximal oocytes, but had negligible expression levels in the zygotes (Figure 3-4B). Furthermore, expression levels of sheath cell contractile activity related genes were also progressively

increased along the gonadal development axis, but almost vanished in zygote samples, such as genes *pat*-10, *mup*-2, *tni*-1 and *unc*-27 coding for the troponin complex (Obinata, Ono, & Ono, 2010; K. Ono & Ono, 2004)(Figure B-2A), and genes *unc*-54 and *myo*-3 coding for the myosin heavy chain (K. Ono & Ono, 2016; Shelton, Carter, Ellis, & Bowerman, 1999) (Figure B-2B).

Similar reasoning can be made with other genes that have previous evidence of somatic or germline origins. Searching the CenGEN database(Hammarlund, Hobert, Miller, & Sestan, 2018) revealed that genes *perm-2/4* encoding components of the eggshell(Gonzalez et al., 2018) had the highest expression levels in sheath cells. Expression of perm-2/4 is absent in P0 but high in F1 (Figure B-2C), while other known components of eggshell that are produced in the germline such as egg-1/2 do not exhibit such significant decreased expression in P0 (Kadandale et al., 2005). The expression of myosin light chains genes *mlc-1/2/3* involved muscle activity (Moerman, Fire, & Riddle, 1997; Rushforth, White, & Anderson, 1998) are all high in F1 but absent in P0, while the expression of the non-muscle myosin light chain gene *mlc-4* required for cytokinesis in zygotes, is present in P0 (Figure B-2D) (K. Ono & Ono, 2016; Shelton et al., 1999). Analysis of actin genes *act-1/2/3/4* (S. Ono, 2014; S. Ono & Pruyne, 2012) may even suggest that the expression of act-4 is not required in zygotes, as it is the only actin gene with negligible expression in P0 (Figure B-2E), an observation also supported by previous findings that *act-1/2/3* were expressed in both muscle and non-muscle cells, while *act-4* was expressed predominantly in body wall muscle(Stone & Shaw, 1993; Willis, Munro, Lyczak, & Bowerman, 2006).

3.3.5 Proximal oocyte expression profiles reveal interactions between the germline and the somatic gonad

We compared the distributions of Log2FC values for all the comparisons of the DEGs that exhibit significantly lower expression levels in P0 samples in the F1 vs P0 comparison. As shown in Figure 3-4C, a considerable number of the genes show a significant increase in expression between the F2 vs F1 comparison, as indicated by an additional small peak with higher Log_2FC values in the distribution compared to other comparisons. These genes might account for those in clusters 7 and 8 (Figure 3-3). This is interesting, as early studies indicate transcriptional inactivity or an overall presence of transcriptional silence as oocytes move to the proximal end (Starck, 1977; Walker, Boag, & Blackwell, 2007). Though this was the case for the F3 to F2 transition, however, clearly not for the F2 to F1 transition (Figure 3-2C). Two Uterine Lumen-Expressed (ule) genes *ule*-3 and *ule*-5 (Figure 3-5D) exhibited sudden increases in transcription from 10-fold to 100-fold between the F2 and F1 transition. This is different from expression patterns of genes of gonadal sheath origin that we described earlier, where the expression levels stay relatively stable in the proximal oocytes. It has been reported that *ule*-3/5 might play a role in driving the ageing of the reproductive system, though the origin of their expression is not clearly discernable (Zimmerman, Hinkson, Elias, & Kim, 2015). A more recent study utilizing fluorescent in situ hybridization (FISH) to track the origins of these transcripts suggests a mechanism by which the transcripts are produced in spermathecae and carried over into the proximal oocytes(Trimmer et al., 2023). Using the expression of *ule-3/5* as a reference, we discerned a set of 25 genes displaying the similar expression pattern (Table

3-2) by by the criteria: $Log_2FC < -7$ in the F1 vs P0 comparison; and $Log_2FC > 4$ in the F2 vs F1 comparison; and average Median Normalized Expression in F1 > 1000). From the NEXTDB database (Kohara, 2001), we were able to obtain in situ hybridization imaging for the products of 17 of these genes, of which 13 genes exhibited clear localization of corresponding transcripts in the spermathecae region (Figure 3-4E, Figure B-3). These results suggest possible interactions between transcriptionally silent oocytes and its somatic neighbors, where the transcriptional events happen in the surrounding spermathecae, and the proximal oocytes take up the transcripts and produce protein products.

Table 3-2 25 Putative Genes from Spermathecae
clec-222, D1054.10, ule-3, D1086.6, F53H4.2, F54F7.3, F55B11.2, F55B11.3, F57C2.4, K07A1.6, Y37D8A.19, Y57G11B.5, Y62H9A.3, Y62H9A.4, Y62H9A.5, ule-5, ZC373.2, E02H9.7, F17E9.4, ZK813.1, ZK813.3, D1086.11, H29C22.1, ZK813.7, F38A5.22

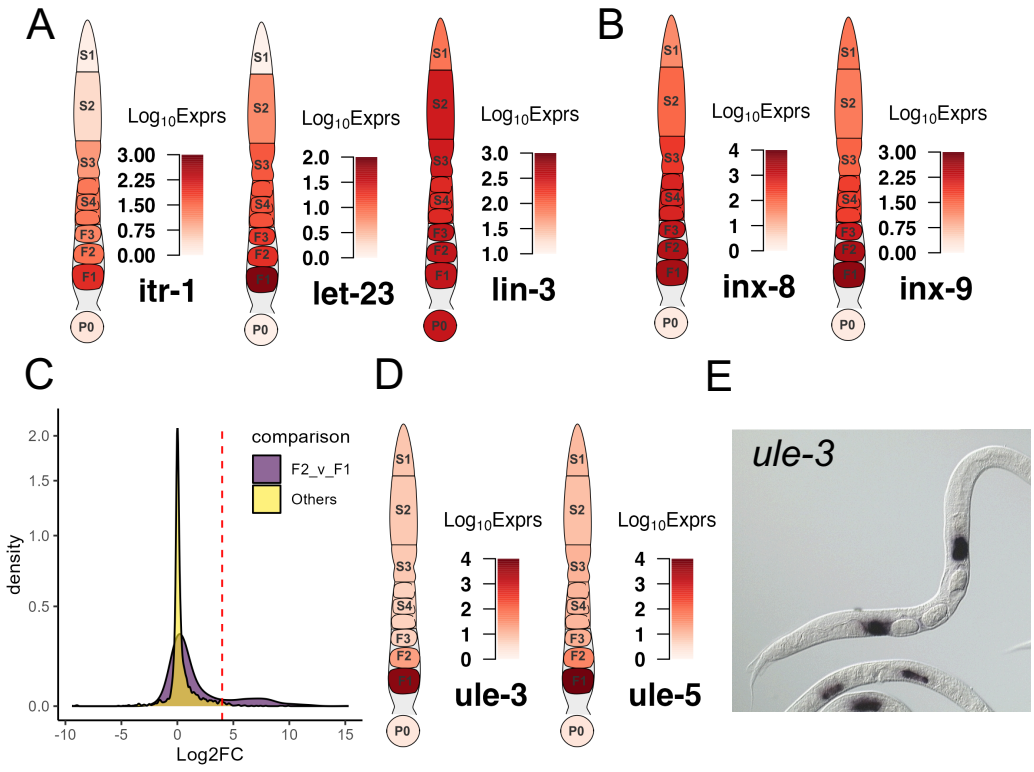


Figure 3-4 Examples of transcriptional dynamics of possible sheath cell genes along the gonad developmental axis. A. Genes coding for hemichannels (*inx*-8 and *inx*-9) of the somatic gonad. B. Genes coding for components signaling pathways between proximal sheath cells and oocytes. C. Distribution of Log_2FC values between neighboring stages of the DEGs that are significantly downregulated in the F1 vs P0 comparison. A small portion of these DEGs is significantly upregulated in the F2 vs F1 comparison as indicated by the right peak of the distribution compared to other comparisons, see Table 3-2 for details. D. Genes coding for ULE-3/5. E. NEXTDB in situ imaging of ule-3 expression in spermathecae. In each gonad diagram, the average expression levels of the genes in each segment or the zygote are shown.

3.3.6 DEGs mark transcriptional timing of the key events of oogenesis and fertilization

One of the early key events in the oogenesis process is the control of mitosis and meiosis. Thus, it is interesting to look into the transcription patterns of mitosis promoting

regulators FBF-1/2 (Crittenden et al., 2002) and meiosis promoting regulators GLD-1/2/3 (Eckmann, Crittenden, Suh, & Kimble, 2004). We found *fbf-1/2* showed high expression in the S1 stage, but lower expression in the S2 and S3 stages and beyond (Figure 3-5A), while *gld*-2/3 exhibited increased expression only between S1 and S2, and then along with *gld*-1 maintained high expression in the later stages (Figure 3-5A). In addition, we observed transcriptional regulation of key factors involved in the maintenance of germ cells in meiotic prophase I such as OMA-1/2 and LIN-41(Tsukamoto et al., 2017). Specifically, the expression of *oma*-1/2 and lin-41 gradually increased throughout the early stages (S1 and S2) of oogenesis followed by high elevations in the S3 stage, which were maintained even after fertilization, apart from *lin*-41, whose expression dropped after fertilization (Figure 3-5A). It has been suggested that LIN-41 could prolong prophase I and inhibit meiotic maturation after fertilization by a translational level regulatory mechanism(Spike et al., 2014; Tsukamoto et al., 2017), thus diminishment of the *lin-41* transcription in zygotes suggests that transcriptional degradation might also play a role in the exit of the oocyte from metaphase I upon fertilization.

We also found that many genes coding for eggshell components were upregulated in distal segments of the gonad, far before the complete formation of the eggshell that happened around the early-stage embryo (Stein, 2018). Genes coding for components of the vitelline layer (cbd-1)(Gonzalez et al., 2018), the chitin layer (chs-1, gna-2, egg-1/2/3) (Wendy L. Johnston & Dennis, 2012; Wendy L Johnston, Krizus, & Dennis, 2006; Kadandale et al., 2005; Maruyama et al., 2007; Y. Zhang, Foster, Nelson, Ma, & Carlow, 2005) and the proteoglycan layer (cpg-1/2) (Olson, Bishop, Yates, Oegema, & Esko, 2006)

all exhibit increased expression in early stages of the germline until after fertilization (Figure 3-5B). These results suggest that transcription of these eggshell genes occur mostly during the mitosis to meiosis transition and the pachytene, while translation and degradation of these transcripts might occur as a response to fertilization signaling.

Moving along the germline, another key event of oogenesis happens in the diplotene loop (S3) where germ cells undergo apoptosis. Here we find that genes regulating apoptosis form three distinct patterns of expression. The expression of genes encoding core apoptosis machinery such as apoptosis initiators CED-4/3(W. Huang et al., 2013) and apoptosis inhibitor CED-9 (Hengartner, Ellis, & Horvitz, 1992) were initially low in the S1 stage but elevated to steady states in the S2 (pachytene) stage (Figure 3-5C). The high expression levels of both *ced*-3 and *ced*-9 were largely maintained thereafter until after fertilization, while that of *ced*-4 was maintained thereafter until the F2 stage and then gradually decreased in F1 and zygotes (Figure 3-5C).

Expression of *ced*-8, which encodes a substrate of the CED-3 Caspase and is likely involved in regulating the timing of apoptosis (Y.-Z. Chen, Mapes, Lee, Robert Skeen-Gaar, & Xue, 2013), follows a different pattern, with significant upregulation in the S3 stage, and maintaining high expression until fertilization (Figure 3-5D). The sudden increase in *ced*-8 transcription in the S3 stage suggest that CED-8 might play an important role in initiating apoptosis in the germline. Other genes such as *skr*-7, *eor*-2 and *dre*-1 showed expression patterns like that of *ced*-8, with elevated expression starting from the pachytene (S3) loop onwards through fertilization (Figure 3-5D). SKR-7 has been implicated in inducing apoptosis(Gao et al., 2008), and DRE-1 has been found to interact

directly with CED-9 in regulating apoptosis (Chiorazzi et al., 2013). Early studies have found EOR-2, along with EOR-1 to induce apoptosis in neuronal cells (Hoeppner et al., 2004). However, we only observed upregulation of *eor-2* (Figure 3-5D) but not of *eor-1* in the germline, suggesting a different mechanism of EOR-2 induced apoptosis in the germline than in neuronal cells.

The third group of apoptosis related genes follow a different expression pattern that can be characterized by the expression profile of *egl-1*, which encodes a direct downstream target of CED-4 and inhibitor of CED-9, playing a critical role in DNA damage induced germline apoptosis (W. Huang et al., 2013). *Egl-1* exhibited a transient increase in transcription in the pachytene loop (S3) that did not go beyond the S4 stage (Figure 3-5E). Other apoptosis related genes such as *csp*-1 and *ces*-2 displayed expression patterns like that of *egl*-1 (Figure 3-5E). An earlier study has found that *csp-1* was expressed in late stage pachytene of the germline using FISH imaging (Denning, Hatch, & Horvitz, 2013). *Ces-2* has been implicated in the apoptosis of neuronal cells in *C. elegans*, though a previous study suggested that *ces*-2 was not essential for germline apoptosis (Metzstein, Hengartner, Tsung, Ellis, & Horvitz, 1996). However, the sudden upregulation of *ces-2* transcription in S3 strongly suggests a role of *ces-2* in apoptosis of the germline. Furthermore, since all 3 genes belong to cluster 14 (Figure 3-3), it is likely cluster 14 contains other genes that are related to apoptosis as well.

As shown in Figure 3-5F, genes encoding ribosome subunits and other translation-related proteins generally exhibited downtrends in transcription as oocytes matured and

prepared for fertilization, consistent with a previous observation (Diag et al., 2018). Interestingly, genes downregulated between S1 and S2 are mainly involved in ribosomal precursor production, such as *eif-6, rpoa-1/2, fib-1, nucl-1*, etc. (Supplementary Table 3-2) (Miluzio, Beugnet, Volta, & Biffo, 2009; D. Xu et al., 2023), while those downregulated between S4 and F3 mainly encode ribosomal protein subunits, such as *rla-0/1, rpl-1/2/3/4/5/7/9/10/13/14/15/16/17* and *rps-0/1/2/3/4/5/7/8/9/10/11/12/13/14/15* (Supplementary Table 3-2) (Nakao, 2004). This suggests that the preparation of ribosomal assembly machinery for oogenesis mainly occurs in the distal mitotic regions prior to entering pachytene (S2), but ribosomal proteins continue to be produced until diakinesis (S4). Moreover, the reduced levels of expression from the S4 stage and beyond indicating that all the transcription of translational machinery required for oocyte maturation are formed before the diakinesis stage.

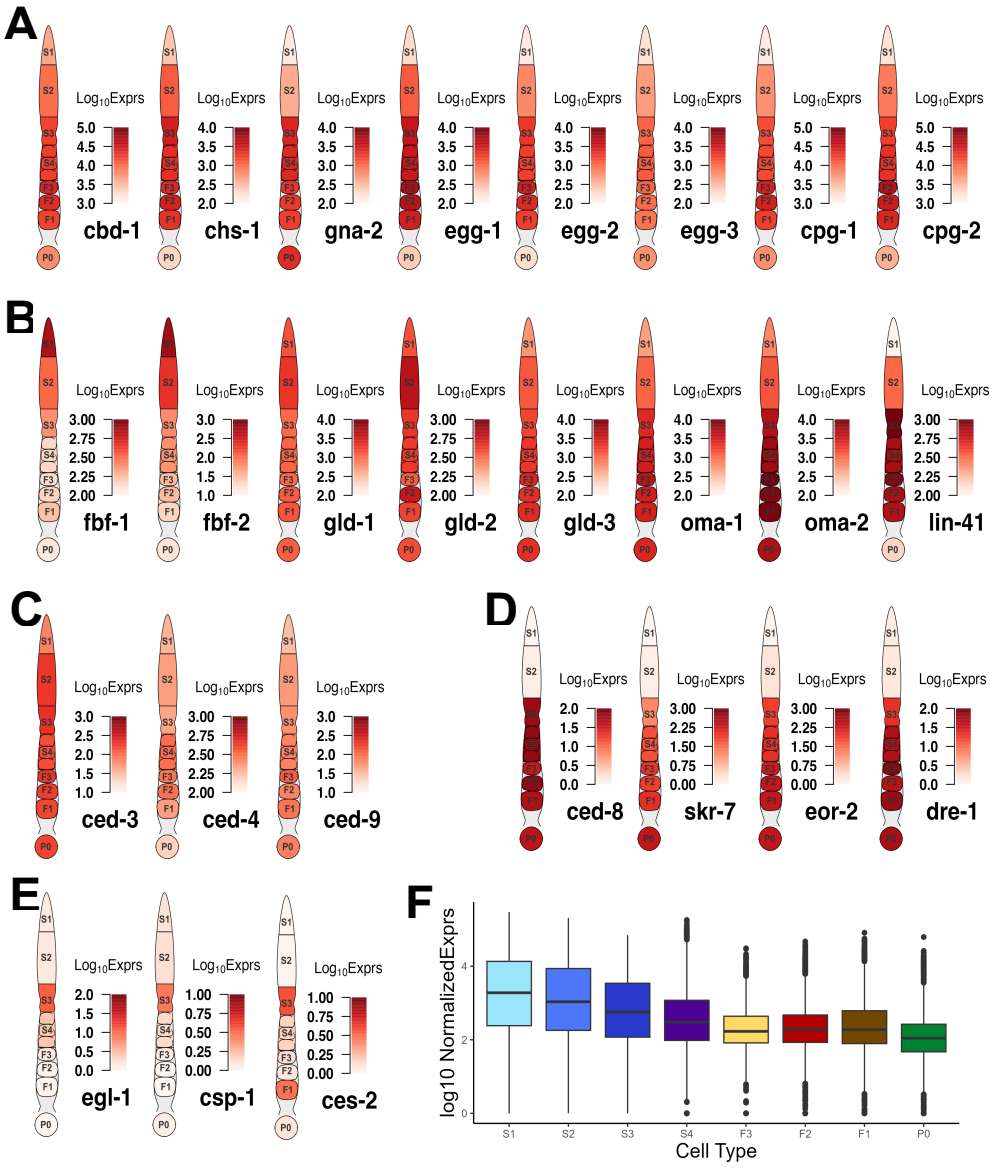


Figure 3-5 Examples of transcription of DEGs that are involved in key events of oogenesis and fertilization. A. Genes encoding different elements of the eggshell. B. Genes involved in mitosis- meiosis transition and meiotic maturation. C. Genes involved in apoptosis with expression throughout the gonad. D. Genes involved in apoptosis with elevated expression starting from the S3 stage. E. Genes involved in apoptosis showing transient expression in the S3 stage. F. Boxplot of transcription levels of genes coding for ribosomal subunit across each stage of gonad development and in zygotes.

3.3.7 Differential Alternative Polyadenylation activity resumes post-fertilization.

Though a great deal of literature has focused on regulation of translation through the 3'UTRs of transcripts by ribosomal binding proteins (RBPs), few have elucidated changes of the 3'UTRs themselves (Diag et al., 2018; M. Mangone et al., 2010; Merritt, Rasoloson, Ko, & Seydoux, 2008; Steber, Gallante, O'Brien, Chiu, & Mangone, 2019). Thus, we analyzed differential alternative polyadenylation (DAP) usage through the DaPars2 software (Feng et al., 2018; L. Li et al., 2021), which estimates changes in proportion of distal (lengthened 3'UTR) and proximal (shortened 3'UTR) polyadenylation sites used in two conditions. We found very few significant changes in distal versus proximal sites usage between neighboring stages, apart from the S4 vs F3, F3 vs F2 and F1 vs P0 comparisons (Figure 3-6A). GO term enrichment analysis found that only the F1 vs P0 comparison resulted in significant enrichment of genes with ADP for mitotic cell cycle related processes, with mostly shortened 3'UTRs (Figure 3-6B). For instance, we find that *cyb*-1/2.2 exhibit shortened 3'UTRs while *cdk*-1 exhibited lengthened 3'UTR (Figure 3-6C). CYB-1/2.2 along with CDK-1 regulate M phase entry of cell cycle in *C. elegans* (Rabilotta, Desrosiers, & Labbé, 2015). Though most DAP genes between F1 and P0 exhibit shortened 3'UTR (Figure 3-6B), it is not straightforward how usage of distal vs proximal sites functions in the regulation of the final expression of protein products. Furthermore, despite very few significant DAP genes in the early stages of oogenesis, we found par-5 gene to exhibit DAP in both the S1 vs S2 and F2 vs F1 comparisons (Figure 3-6C). In fact, 3'UTR length of the par-5 transcript gradually decreased until F3 stages

before increasing again (Figure 3-6C). It has been reported that PAR-5 regulates asymmetric cell division and alternative 3'UTR isoforms of par-5 confers different levels of PAR-5 protein abundance (Mikl & Cowan, 2014). Interestingly, most significant DAP genes between S4 and F3 exhibited an increase in 3'UTR length, while most DAP genes between F3 and F2 exhibit decreased in 3'UTR length (Figure 3-6A). However, it is unclear whether this is because F3 oocytes are fully cellularized and maintain a stable transcriptome or other factors.

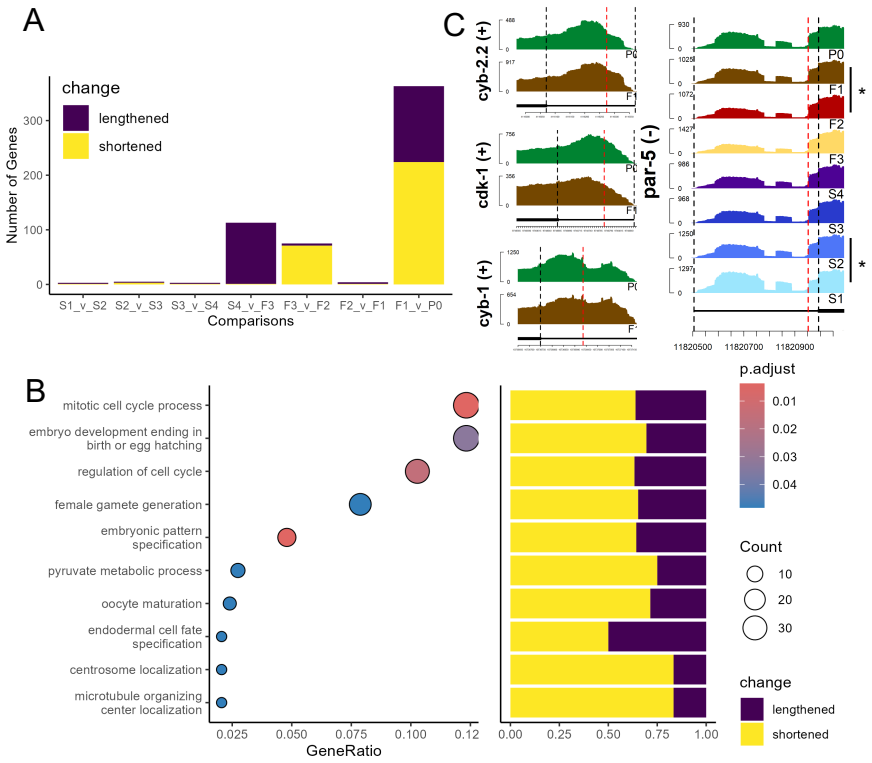


Figure 3-6 Differential alternative polyadenylation (DAP) analysis of genes between neighboring stages. A. Number of DAP genes between each neighboring stage, colors indicate lengthening (purple) or shortening (yellow) of 3'UTR lengths. B. GO term enrichment of significant DAP gene between F1 and P0 (left panel), and bar plot of percentage of significantly lengthened or shortened genes in each enriched gene set, GeneRatio is the proportion of differentially polyadenylated genes that belong to a known gene set. C. Coverage by RNA-seq reads of 3'UTRs of genes *cyb-1/2.2, cdk-1* and *par-5,* red lines mark the estimated proximal polyadenylation site.

3.3.8 Differential Splicing play roles in germline development

We further performed differential splicing analysis using rMATs (S. Shen et al., 2014) to look for differential transcription of alternatively spliced isoforms of genes between neighboring stages along the oocyte developmental axis. Since rMATs could not account for batch effects, we performed the analysis with samples from Batch A (Figure 3-2A). We identified varying numbers of genes exhibiting significant splicing signals defined by rMATs, i.e. alternative 3' splice site (A3SS), alternative 5' splice site (A5SS), intron retention (IR), mixed exon usage (MXE), skipped exon usage (SE), between neighboring stages. Most notably, the S1 vs S2 and the S4 vs F3 comparisons yielded the most differential splicing usage with 58 genes and 54 genes exhibiting differential splicing, respectively (Figure 3-7A). Genes with differential splicing usages between the S1 vs S2 comparison are enriched for GO terms related to mitosis (Figure 3-7B), which is expected, given the fact that S1 contains the TZ regions (Figure 3-1C). However, other neighboring stages comparisons yielded no significantly enriched GO Biological Process terms. A few interesting examples are detailed as follows.

Gene *inx*-14 was differentially spliced during the S1 to S2 transition (Figure 3-7C). Specifically, *inx*-14 was preferentially utilized for its longer 6th exon in the S2 stage compared to the S1 stage, resulting in increased proportions of its F07A5.1b isoform (Figure 3-7C). It has been documented that INX-14, along with INX-21/22 and INX-8/9 forms hemichannels that facilitate the communication between the somatic gonad and the germline and plays a role in meiosis to mitosis transitions by negatively regulating meiotic

maturation and promoting germline proliferation (Starich et al., 2014). UniProt designates the F07A5.1b isoform as the canonical isoform, differing from the alternative F07A5.1a isoform by 2 amino acids in the 406-407 positions. Our results present a possible mechanism by which INX-14 changes its association with either INX-21 or INX-22 (Starich et al., 2014).

Another notable event was a gradual increase in preference of *zen-4* skipping its 8th exon in the S1 to S3 transition (Figure 3-7D). Specifically, the most abundant *zen*-4 isoforms were M03D4.1a.1, M03D4.1c.1, M03D4.1d.1 and M03D4.1f.1 (Figure 3-7D). This is due to the lack of read coverage for the regions that are spanned by the other isoforms (Figure 3-7D). The exon skipping event is indicative of decreased preference for the M03D4.1d.1 and M03D4.1f.1 isoforms, which contain the skipped exon in the other isoforms (Figure 3-7D). ZEN-4 along with CYK-4 forms the centralspindlin complex, a conserved component of intercellular bridges that functions in the cellularization of cells during cytokinesis(K.-Y. Lee et al., 2018; White & Glotzer, 2012; K. Zhou, Rolls, & Hanna-Rose, 2013). Though a recent study suggested that ZEN-4 was not essential in the germline for the closure of the intercellular bridge(K.-Y. Lee et al., 2018), our results suggest that as the oocyte moves along the rachis into late pachytene stage, alternative isoforms of *zen*-4 may still play a role in the cellularization of maturing oocytes.

In addition, we found that *ife*-3 switched isoforms during the S1 to S2 transition (Figure 3-7E). The *ife*-3 gene encodes one of the *C. elegans* homologs for human translation initiation factors (eIFs) that plays critical role in regulating mRNA content along with microRNA and RNA binding proteins(Huggins et al., 2020). More specifically,

ife-3 functions as a repressor of *fem*-3 expression to promote production of oocytes in the germline(Huggins et al., 2020; Mangio, Votra, & Pruyne, 2015). Here, we showed a switch in *ife*-3 splicing preference for the B0348.6b and B0348.6c isoforms over the shorter B0348.6a isoform (Figure 3-7E). Along with a slight increase in *ife*-3 expression, these results hint at a possible mechanism of *ife*-3 regulation in the pachytene stage of oogenesis. Interestingly, *ife*-3 expression reduced significantly in proximal oocytes, where transcription became increasingly silent, thus obviating the need for mRNA regulation (Figure 3-7E).

Other genes worth pointing out include *tos-1* coding for a reporter of differential splicing (L. Ma, Tan, Teng, Hoersch, & Horvitz, 2011), and lev-11 coding for tropomyosin (Watabe, Ono, & Kuroyanagi, 2018). *Tos*-1 experiences loss of preference for the usage of its 3rd exon from F1 to zygote (Figure B-4A), which is further corroborated by the decreased coverage of its longer isoform in S4 (Figure B-4A). However, the differential splicing of *lev*-11 transcripts (Figure B-4B) might occur in sheath cells wrapped around the oocytes as we argued earlier. It has been shown that different isoforms of *lev*-11 exhibit different characteristics in terms of muscle assembly and function (Watabe et al., 2018). Since F2 oocytes are roughly covered by Sh4 and F1 oocytes by Sh5, it is likely that an alternative isoform switch of *lev*-11 contributes to the different functions of these two sheath cell types.

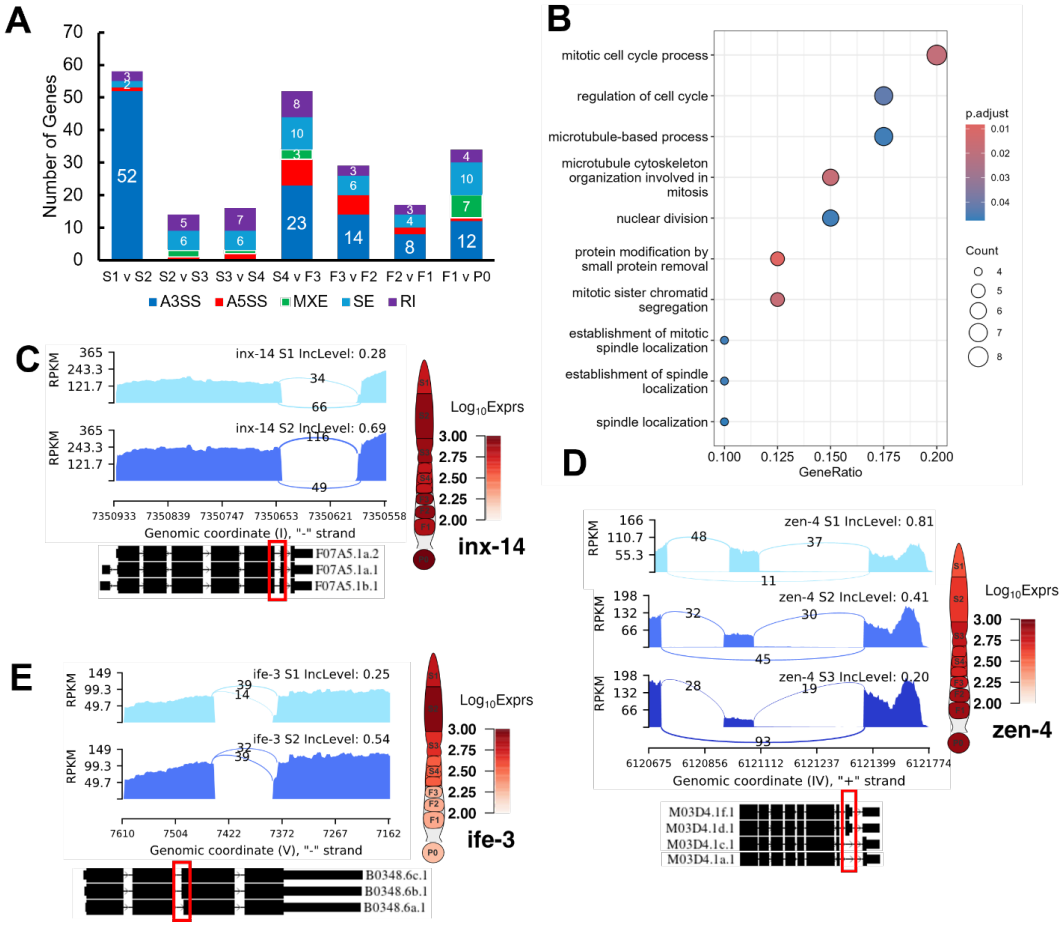


Figure 3-7 Examples of differential splicing usage of genes during germline development. A. Box plot of numbers of five splicing types (A3SS, A5SS, MXE, Se and RI, see main text for definitions) detected between each pair of neighboring stages of germline development. B. Enriched GO terms of genes with differential splicing events between the S1 and S2 stages, GeneRatio is the proportion of differentially spliced genes that belong to a known gene set. C. Differential splicing events of gene *inx*-14 between the S1 and S2 stages. D. Differential splicing events of gene *zen*-4 between the S1 and S2 as well as S2 and S3 stages. E. Differential splicing event of gene *ife*-3 between the S1 and S2 stages. Exact positions of splicing events are shown in the red box.

3.4 Discussion

With a spatial layout of cells that simultaneously mirrors the timeline of oogenesis, the *C. elegans* gonad can serve as a powerful model for uncovering mechanisms of oogenesis. However, the tiny size of the gonad also presents challenges for in-depth studies of the intricacies of this process. With the recent development of single cell methods, we utilize scRNA-seq techniques to decipher the transcriptomic landscape of different stages of oocyte formation as well as fertilization. Our transcriptomic dataset of the *C. elegans* gonad presents a good platform for research into the transcriptional landscape of oogenesis of animals. Our results not only are able to recall most of the oogenic genes designated by earlier research that utilized micro-arrays and bulk-RNAseq (Ortiz et al., 2014; Reinke et al., 2004), but also are highly correlated, through careful alignment of samples, with expression profiles of the different stages of the germline found by more recent studies that relied on single cell based techniques (Diag et al., 2018; Tzur et al., 2018).

Our expression profiles show a distinct pattern in the UMAP display, which is consistent with the developmental axis of the gonad, indicative of our successful capture of the transcriptomes underlying the oogenesis program. Though our dataset presents discernable batch effects, we either incorporated them into our analysis models or forfeited the smaller batch of samples when necessary. The number of biological repeats for each stage as well as the sequencing depth for each sample means the results are robust to the discarding of few samples.

Despite meticulous filtering of possible contamination of somatic tissues, the dissection of the tiny gonad presents a delicate problem, and it is difficult to fully avoid

contamination by surrounding tissues. In this study, we note that distal stages (S1-S3) inevitably contain transcripts originating from Sh1 and Sh2 sheath cells, due to the unenclosed and miniscule nature of the germline along the rachis. Thus, we focus on elucidating the transcriptomic changes of known germline associated genes to minimize false positive findings. Despite separating proximal stage oocytes as best as possible from shattered somatic tissue, e.g., sheath cells. complete removal of somatic components remained an elusive task. This prompted us to investigate patterns of expression that may arise from known somatic specific genes and take care in interpreting the results. We find that a great portion of genes that are drastically downregulated in zygotes relative to the F1 oocyte are of somatic origin, including many known markers of muscle cells and sheath cells. This allowed us to mark a large portion of genes as somatic in nature, especially in the proximal gonad samples. However, we argue that the expression profiles of these genes depleted in zygotes are not without merit. From their expression patterns throughout the gonad, these genes can be divided into roughly two groups. The first consisting of genes that have relatively stable transcription in the proximal oocytes before complete disappearance in zygotes, and a second group consisting of genes that are drastically upregulated in only the F1 oocyte. This second group of genes includes those whose transcripts have recently been found to be produced in the spermathecae but transported into F1 oocytes (Trimmer et al., 2023). We thus provide a list of genes that might undergo this process. Though the exact function and underlying mechanisms for this phenomenon remain to be elucidated, a few genes exhibiting this pattern have been shown to affect the ageing of *C. elegans* (Zimmerman et al., 2015).

We confirm previous findings (L.-W. Lee, Lee, Huang, & Lo, 2012) at the transcriptomic level that the growth of oocytes presents as a process in which ribosomal biogenesis and cellular activity gradually decreases. Moreover, we observed at the transcriptional level known dynamics of core regulators of the mitosis to meiosis switch and meiosis maturation. In addition, we find that many genes involved in the eggshell formation initiate transcription as early as in the S1 stage, and their transcripts are accumulated until post fertilization. The *C. elegans* germline also presents a remarkable model for studying germ cell apoptosis (Gumienny, Lambie, Hartwieg, Horvitz, & Hengartner, 1999). Our results not only capture distinct upregulation of apoptosis related genes in the pachytene loop (S2 stage), but also discover novel candidate genes for future studies of germ line apoptosis. Furthermore, our gene clustering and DEG results also reveal three distinct sets of apoptotic related genes, characterized by the expression pattern of ced-3/4/9 (Figure 3-5C), *ced*-8 (Figure 3-5D) and *egl*-1(Figure 3-5E), respectively. These different modes of transcription suggest that cross-talks occur between different genes at the transcriptional and post-transcriptional levels to induce apoptosis.

The previous report that RBP and the 3'UTRs are key players in a complex regulatory mechanism (Diag et al., 2018; Marco Mangone et al., 2010; Merritt et al., 2008) in the *C. elegans* germline prompted us to investigate whether significant changes in polyadenylation site usage occurred during oogenesis and fertilization. Though we were not able to find significantly enriched pathways regulated via changes in polyadenylation during oogenesis, we did find enrichment for cell cycle processes due to changes in polyadenylation site usage during fertilization. Our results suggest that polyadenylation

sites of transcripts remain relatively stable during oogenesis, and active regulation of alternative polyadenylation likely resumes in the zygote.

Finally, we find that alternative splicing events are present throughout the gonadal segments. We reveal significant changes in the usage of isoforms of hemi-channel gene *inx*-14. It is highly likely that products of different isoforms of *inx*-14 associate with germline hemichannels INX-21/22 or somatic hemichannels INX-8/9 to facilitate communication between the somatic gonad and germline. We also find differential splicing usage of genes in the germline. For instance, we observe differential splicing of *zen*-4 throughout the pachytene region. Though previous studies preclude the involvement of ZEN-4 in oocyte cellularization in the germline syncytium (K.-Y. Lee et al., 2018), ZEN-4 isoforms may still serve functions in oocyte growth in late pachytene. Future studies are needed to elucidate the roles of isoform usages in *C. elegans* oogenesis and the underlying mechanisms.

Taken together, our results paint a complex transcriptional landscape of the germline development, oogenesis and fertilization processes in *C. elegans* in finer detail than previous studies. Though contamination of somatic tissue presented challenges, we were able to discern putative somatic elements. We not only confirm previous findings, but also present many novel discoveries of transcriptional events along the temporospatial axis of the *C. elegans* germline and the zygote. Though much work remains to be done, particularly, with better dissection techniques to remove somatic contaminations, our study still presents a wealth of resources and gene candidates for future experimental investigation to reveal the underlying mechanisms of the oogenesis program.

CHAPTER 4 SINGLE-CELL RNA-SEQ ANALYSES OF OOCYTES AND ZYGOTES REVEAL EARLIER MATERNAL TO ZYGOTE TRANSITION IN MICE AND RATS

4.1. Introduction

Fusion of the oocyte and sperm forming the zygote marks the inception of all mammalian embryogenesis, during which the male and female genome come together to jumpstart the developmental process(McGrath & Solter, 1984). Despite providing half the DNA and being crucial for oocyte activation, the sperm brings few components in the zygote, and thus provides smaller sources in the early stages of embryogenesis than the oocyte. It is therefore mainly the contents of the maternal gamete that set up the suitable environment for successful polyspermy prevention and zygote genome activation (ZGA). During ZGA the control of cellular processes is gradually let go mainly from the products of the maternal genome and transferred to products of the zygote genome(Saunders et al., 2002; Sutovsky & Schatten, 2000). This process, also known as maternal to zygote transition (MZT) (Tadros & Lipshitz, 2009), has long been the subject of interest for researchers. Most foremost, the oocyte is a highly differentiated cell that quickly acquires pluripotency after fertilization. Though much research has been conducted in this area, it is still an elusive topic, and many intricacies remain to be uncovered.

An aspect that has been observed in MZT of all mammals has been the silence of the transcriptome in mature oocytes prior to fertilization up until ZGA after fertilization, which initiates the events of embryogenesis(M. T. Lee, Bonneau, & Giraldez, 2014). The length of this period of transcriptional silence varies in different species. This quiescent period that is conserved among species has not been fully understood, particularly, the mechanisms underlying ZGA may be different for each species(Schulz & Harrison, 2019).

Nonetheless, in all the mammalian species the machineries that carry on the progression of meiosis and remodel the genome depend solely on the maternal mRNAs accumulated during oocyte development(M. T. Lee et al., 2014; Schulz & Harrison, 2019). Thus prior to ZGA, there is also translation and degradation of many maternal mRNAs(J. Ma, Flemr, Strnad, Svoboda, & Schultz, 2013; Q. Q. Sha et al., 2020).

The mouse (*Mus musculus*) and rat (*Rattus norvegicus*) are two of the most used model mammals in biomedical research. Many aspects of the two organisms are very similar, such as genome size and number of protein-coding genes. It is estimated that at least 90% of genes in the mouse genome have at least a homolog in the rat genome, and vice versa (Srivastava et al., 2020). However, there remain considerable differences in the biology of the two organisms including reproduction(Mullins & Mullins, 2004). For instance, it is known that ZGA occurs in 2 waves in mice with a 'minor ZGA' in the late 1-cell stage and a 'major ZGA' in 2-cell stage. In contrast, it was reported that the onset of embryonic transcription activity in rats occur in late 2-cell stage(K. I. Abe et al., 2018; Zernicka-Goetz, 1994). However, even though major ZGA is known to occur at the 8-cell stage in human embryos, recent research has suggested that minor ZGAs occur in 1-cell human zygotes as well, though to a lesser degree than that of mice(Xue et al., 2013). Thus, it is possible that similar transcriptional events might also occur in 1-cell rat zygote.

Research into transcriptional events occurring during minor ZGA in mice has revealed insightful results. Early studies in mice using brUTP incorporation found that transcription occurred in the S stage of 1-cell zygote (Aoki, Worrad, & Schultz, 1997), and transcription of MuERVL (murine endogenous retrovirus-L) genes was verified via RT-

PCR studies (Kigami, Minami, Takayama, & Imai, 2003). Another study found that transcription was promiscuous across open chromatin regions of the mouse zygote genome, with many transcripts containing intronic and intergenic regions, indicating limited 3'end processing and splicing mechanisms involved, possibly to protect the zygote from precocious gene expression(K. I. Abe et al., 2018).

Although the understanding of transcriptional events in the zygote is still inconclusive in many species, post-transcriptional regulation of transcripts appears to be ubiquitous. A pivotal aspect of post-transcriptional modification is cleavage and polyadenylation of maternal mRNAs. The choice of polyadenylation cleavage sites in 3' untranslated regions (3' UTRs) is thought to influence various UTR binding components that in turn regulate the stability, translation, and degradation of transcripts. Post-transcriptional regulation via alternative poly-A cleavage sites as well as the length of poly-A tail plays vital roles in MZT in many organisms. For example, poly-A tail lengthening promotes global translation in *Drosophila* zygotes (Eichhorn et al., 2016). Removal of de-adenylation components in mouse oocytes leads to developmental arrest in pre-implantation embryos (Liu et al.; Pasternak, Pfender, Santhanam, & Schuh; Qian-Qian Sha et al.). A recent study has also found extensive remodeling of maternal mRNAs in human embryos via poly-A tail modifications such as changes in poly-A tails and 3'UTR lengths, which are essential for MZT in humans (Liu et al., 2023).

Furthermore, despite the plethora of studies into the embryonic development of mice, little is known about the maternal vs zygote transcriptomic landscape in rats, and how the shared and differing characteristics with mice contribute to early stages of their

respective pre-implantation embryos. The rise in application of single-cell transcriptomic studies has opened a front into studying the intricacies of developmental dynamics at the single-cell level. However, many such transcriptomic studies overlook alternative insights that can be made apart from gene expression. In this study, we present orthogonal forms of comparative analysis on the transcriptomic landscape of the early MZT by applying a full-length single-cell RNA sequencing (scRNA-seq) protocol (Picelli, Björklund Å, et al., 2013) to oocytes and zygotes in mice and rats. We identify RNA transcripts that are produced, modified, or degraded in the zygotes. We also reveal mechanisms that are unique and conserved in the two species.

4.2. Materials and Methods

4.2.1 Sample preparation

Harvest and preparation of oocytes and zygotes: Mouse and rat oocytes and zygotes were isolated at Horizon Discovery, St. Louis site, operated under approved animal protocols overseen by Horizon Discovery's Institutional Animal Care and Use Committee (IACUC). C57BL/6N mice and Sprague Dawley rats purchased from Taconic Farms were housed in standard cages and maintained on a 12-h light/dark cycle with *ad libitum* access to food and water. Three- to four-week-old female mice were injected with PMS (5 IU/mouse) 48 hr before hCG (5 IU/mouse) injection, followed by with or without mating to stud males. Fertilized eggs and oocytes were harvested 1 d later, respectively. Four- to five-week-old female rats were injected with 20 units of PMS, which was followed by an

injection of 50 units of hCG 48 h later, followed by with or without mating to stud males. Fertilized eggs and oocytes of both species were harvested 1 d later.

4.2.2 Construction of scRNA-seq libraries

Full-length double-strand cDNA for each oocyte or zygote was prepared using an sc-RNA-seq libraries SMART-Seq v4 Ultra Low Input RNA Kit (Clontech, Mountain View, CA) following the vendor's instruction, which was based on the Smartseq2 protocol (Picelli, Björklund Å, et al., 2013). cDNA-seq libraries were constructed using a Nextera® XT DNA Library Preparation Kit (Illumina, Sandiago, CA, Cat. Nos. FC-131-1024, FC-131-1096) following the vendor's instruction. The libraries were quantified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara) and sequenced on an Illumina 2500 machine with 150 bp paired end reads.

4.2.3 Transcriptome mapping and quantification

The mouse genome assembly (GCF_000001635.27) with annotation version GRCm39 and the rat genome assembly (GCF_015227675.2) with annotation version mRatBN7.2 were obtained from NCBI Genbank. Raw sequences were trimmed using Trim_galore (Krueger, 2015), with parameters length > 35 and q > 10, and subsequently mapped to the respective reference genome using HISAT2(D. Kim et al., 2019) with default settings. The resulting mappings were quantified using HTSeq (S. Anders et al., 2015) with default settings to obtain read counts. Trimmed reads were also mapped to the respective reference genome using Salmon(Patro et al., 2017) with default settings to obtain expression levels in per million (TPM) for both genes and transcripts. Quantification of read coverage of nascent transcripts was performed using kallisto(Bray, Pimentel,

Melsted, & Pachter, 2015)/bustools (Melsted, Ntranos, & Pachter, 2019) in the kb_python package (v0.28.0) (Melsted, Booeshaghi, et al., 2019; Sullivan et al., 2023). More specifically, trimmed reads were aligned to the reference genome with kb_python under the 'nac' workflow with default settings.

4.2.4 Quality control and exploratory analysis

Samples with more than 5% reads mapped to rRNA genes or 25% reads mapped to the mitochondrial genome were removed from further analysis. Furthermore, only genes whose biotypes belong to protein-coding, lncRNA and transcribed pseudogenes were kept for further analysis. Samples were visualized using Uniform Manifold Approximation and Projection (UMAP) based on log transformed DESeq(Michael I Love et al., 2014) median normalized counts, outliers were filtered out for further analysis .

4.2.5 Differentially expressed gene (DEG) analysis

Due to the inherent high variability intrinsic to single cell transcriptomic data, to ensure genes had sufficient coverage and replicates for DEG analysis**,** we only considered genes that were expressed in at least 6 samples. DEG analysis of both mouse and rat samples were carried out with count values from HISAT2(D. Kim et al., 2019)/(S. Anders et al., 2015). The raw counts were normalized to Counts per Million (CPM) and log transformed prior to analysis with MAST (Finak et al., 2015). In fitting the data to the MAST model, we enabled the **useContinuousBayes=TRUE** option to estimate $\log_2$(foldchange) values for genes that had no expression in oocytes or zygotes. For each species, genes with an FDR < 0.05 and a foldchange greater than 2 or less than 0.5 were considered differentially expressed.

4.2.6 Estimation of pre-mature mRNA transcripts

Estimation of nascent mRNA transcripts was carried out through kallisto/bustools (Melsted, Ntranos, et al., 2019; Sullivan et al., 2023) using the kb_python (version = 0.28) wrapper, with method set to '**nac**'. The method is based on the idea that the number of nascent mRNAs for each gene inside the cell reflects a snapshot in the cell's transcription dynamics, and thus by inferring nascent and mature mRNA levels for many cells, it is possible to delineate developmental trajectories by modeling the RNA production using differential equations. From henceforth, we shall use the terms un-spliced reads, nascent reads and intronic reads interchangeably to describe the reads that cover intronic portions of genes. A gene was considered to have pre-mature mRNA expression if it met the following criteria:

- mean read coverage of spliced transcripts across zygote OR oocyte samples > 5.
- spliced transcripts were expressed in 10% of zygote OR oocyte samples.
- mean read coverage of un-spliced transcripts across zygote OR oocyte samples > 0.
- un-spliced transcripts expressed in 10% of zygote OR oocyte samples.
- R^2 of linear model fit between spliced and un-spliced reads of all samples > 0.1.
- Kendall correlation between spliced and un-spliced reads of all samples > 0.1.
- Ratio of standard deviation of spliced reads over standard deviation of un-spliced reads was between 0.005 and 5.

This filtering criteria was set to ensure that estimation of nascent read coverage truly came from un-spliced intronic portions of transcripts. DEG analysis of un-spliced genes was performed with MAST, with un-spliced reads normalized by sample-wise CPM

normalization factors used in the DEG analysis. We define the nascent proportion of a gene as the ratio of the mean spliced read coverage over the mean un-spliced read coverage in a cell. To test differences in nascent proportions (DNP) of a gene between zygotes and oocytes, we applied a Fisher's exact test.

4.2.7 Differential transcript usage (DTU) analysis

DTU analyses were performed under the guidelines given by (Michael I. Love, Soneson, & Patro, 2018). Specifically, transcript abundances (TPM values) estimated by Salmon(Patro et al., 2017) were first filtered using Drimseq(Nowicka & Robinson, 2016) with default parameters. The filtered data was then analyzed using DEXseq(Simon Anders, Reyes, & Huber, 2012) with default settings. Instead of exons in the original applications (Simon Anders et al., 2012), each transcript of a gene was regarded as an exon in our case. The DEXseq results for each gene were aggregated and further analyzed using stageR(Van Den Berge, Soneson, Robinson, & Clement, 2017) to determine if it exhibited DTU while overall false discovery rate was controlled on the gene level.

4.2.8 Alternative polyadenylation analysis

HISAT2 aligned BAM files were first filtered using Samtools (H. Li et al., 2009) to only include concordantly and uniquely mapped reads. The resulting filtered bam files were then transformed into BEDGRAPH format using bedtools(Quinlan & Hall, 2010). Custom scripts were used to account for the pair-end and spliced natures of the read mappings. Primary lists of mouse and rat 3'UTRs were obtained by extracting 3'UTRs from the NCBI annotation files and only unique 3'UTRs were kept. A secondary list of 3'UTRs in each species was obtained using DaPars2 (Feng et al., 2018; W. Wang, Wei, &

Li, 2014) following the authors' instructions. The 3'UTRs in the secondary list may include coding sequences in the last exon. Thus, if the 3'UTRs in the primary and secondary lists of a species were from the same isoform, only the one in the primary list was kept. Though most UTRs were composed of a single continuous segment, there was still a significant portion of UTRs that contained introns. In these cases, we discarded all segments with length < 150 bp and treated the remaining ones as individual 3'UTRs. We identified genes with differential alternative polyadenylation (DAP) in each species using DaPars2 (Feng et al., 2018; W. Wang et al., 2014), which computed a Percentage of Distal poly-A site Usage Index (PDUI) by de novo estimating putative polyadenylation breakpoints via read coverage change on the lists of 3'UTRs. To ensure more accurate DAP analysis, we modified DaPars2 as follows:

1. Searching for breakpoints 25 bp instead of the default 150 bp downstream of the 3'UTR start point and only analyzed 3'UTRs > 150 bp.
2. Furthermore, instead of using read depth of each sample for normalization, we used DESeq2-estimated size factor of each sample's read counts as normalization factor.
3. Allowing DaPars2 to produce distal and proximal polyadenylation site coverages in addition to PDUI units for each UTR.

Moreover, the threshold for coverage was set to 0, thus, any read coverage was counted in the estimation of alternative polyadenylation. Prior to testing, we filtered out genes with less than 10 reads detected in any of the samples, as well as genes with no coverage in both distal and proximal 3'UTRs (PDUI == 'NA') in more than 5 samples. We

then performed Fisher's exact test to evaluate the differences in distal vs proximal polyadenylation preference between oocytes and zygotes using mean distal/proximal coverage in oocyte and zygote samples. Differences in mean PDUI between oocytes and zygotes were also calculated. To aggregate the results to gene level (many genes have multiple 3'UTRs), for each gene, we selected the UTR with the lowest p-value. If multiple different UTRs of a gene had a tied p-value, the one with the largest absolute difference in PDUI was chosen. The resulting p-values were then controlled for false discovery rate with the qvalue package in R. Finally, genes with UTRs that had an absolute mean expression difference > 0.2 and a false discovery rate < 0.05 were considered to have differential alternative polyadenylation. Gene-body coverage in each sample was normalized by the BEDGRAPH coverage as aforementioned, and the average was taken over the oocyte or zygote samples. The normalized BEDGRAPH coverage was processed and visualized using custom scripts and the trackViewer R package(Ou & Zhu, 2019).

4.2.9 Pathway and GO term enrichment analysis

The ClusterProfiler(Yu et al., 2012) R package was used to perform both gene set enrichment analysis and over-representation analysis with gene sets from KEGG (Kanehisa et al., 2022; Kanehisa & Goto, 2000), Gene Ontology Biological Processes (GO BP)(Gene Ontology, 2021), Reactome (Milacic et al., 2024) and Wikipathways(Martens et al., 2021).

4.2.10 Analysis of orthologs between mouse and rat

Orthologs between mice and rats were obtained from the MGI database(Blake et al., 2021). Specifically, rat and mouse genes belonging to the same homology class were considered as orthologs. Over-representation analysis was performed on orthologous genes

that had an absolute foldchange > 1.25 at an FDR < 0.05 in both species with all orthologous genes that were tested for differential expression in both species as the background set. The analysis was based on mouse gene annotations in the GO BP, KEGG, Reactome and Wikipathways databases. Visualization of gene networks was generated using clusterProfiler(Yu et al., 2012). All intersections of genes with DNP, DTU and DAP were performed with the respective results obtained in the prior analysis. All gene coverage plots were generated via the trackViewer(Ou & Zhu, 2019) package.

4.3 Results

4.3.1 Oocytes and zygotes of both mice and rats display distinct transcriptomic patterns

We sequenced the transcriptomes of a total of 17 mouse oocytes, 17 mouse zygotes, 15 rat oocytes, and 16 rat zygotes, with an average of 10,018,474 pairs of reads/cell (Supplementary Table 4-1). The vast majority (>90%) of the reads could be mapped to the respective genomes for most (53/81.52) samples (Supplementary Table 4-1). However, 6 mouse oocytes, 1 rat oocyte and 5 rat zygotes exhibited high levels of reads mapping to mitochondrial or ribosomal genes, we thus excluded them from further analyses (Supplementary Table 4-1) (Materials and Methods). This left us with 28 mouse samples (11 oocytes / 17 zygotes) and 25 rat samples (14 oocytes and 11 zygotes). Interestingly, the reads from mouse samples, particularly, from mouse oocytes, were strongly biased to the 3' ends of CDSs of genes, while the bias was not seen in both rat oocytes and zygotes (Figure C-1A). We detected transcripts for an average of 14,514 and 15,934 annotated genes in mouse oocytes and zygotes, respectively, and for an average of 12,826 and 13,535

annotated genes in rat oocytes and zygotes, respectively (Figure C-1B). Thus, there were more genes expressed in zygotes than in oocytes of both species. Furthermore, oocytes and zygotes of both species are clearly separated by estimated gene expression levels in UMAP displays (Figure 4-1A, 1B), indicating that oocytes and zygotes of both mice and rats contain distinct sets of gene transcripts.

4.3.2 ZGA in rats may begin earlier than previously believed

We first compared expression levels of genes in zygotes relative to those in oocytes of each species. We found 4,603 DEGs in mice, but only 842 DEGs in rat (Figure 4-1C). Of the mouse DEGs, a smaller portion (1,368 or 29.7%) were upregulated while the remaining larger portion (3,235 or 70.3%) were downregulated (Figure 4-1C). In contrast, of the rat DEGs, a larger portion (627 or 74.5%) were upregulated while the remaining smaller portion (215 or 25.5%) were downregulated (Figure 4-1C). Since the zygotes of both species were analyzed prior to the first cell division, the elevated transcriptional activities of the 1,368 upregulated DEGs in mice might reflect the previously reported initial transcriptional events in the mid-1cell stage zygote (K. I. Abe et al., 2018; Schultz, 1993) . Interestingly, even though at a smaller scale, the elevated transcriptional activities of the 627 upregulated DEGs in rats might unveil an early wave of transcriptional activation in rat zygotes, which was previously believed to not occur until later stage of embryogenesis (Zernicka-Goetz, 1994).

4.3.3 Distinct pathways are up- or down-regulated in both mouse and rat zygotes

Downregulated genes in mouse zygotes were significantly enriched for GO BP or pathways including cytoplasmic translation, rRNA processing, nonsense-mediated decay,

TCA cycle, energy production and chromatin modifying enzymes, etc. (Figure 4-1D, Figure 4-2A). For example, ribosomal protein subunits encoding genes such as *Rpl*3/9/12/18/19/26 (Nakao, 2004) were downregulated in mouse zygotes (Figure 4-1D), a result that is in excellent agreement with the earlier findings (Schultz, Letourneau, & Wassarman, 1979). Genes that function in the TCA cycle *such as Ndufa2/Nudfb7/Ndufa8/Ndufc1* (Vinothkumar, Zhu, & Hirst, 2014) and genes involved in energy production such as *Atpa1a/Atp5h* (Song et al., 2018) showed decreased transcript levels (Figure 4-1D). On the other hand, upregulated genes in mouse zygotes were significantly enriched for biological processes or pathways including mRNA transport, cell cycle, transcription, DNA replication, energy production, etc. (Figure 4-1D, Figure 4-2A). For instance, genes *Pdhb, Pdp1, Pdhx* and *Dlat* were upregulated, which encode components of the pyruvate dehydrogenase complex responsible for acetyl-CoA biosynthesis from pyruvate (Z. H. Zhou, Mccarthy, O'Connor, Reed, & Stoops, 2001)(Figure 4-1D). Interestingly, the role of the TCA cycle in ZGA was previously studied in early mouse embryos, and it was found that despite low metabolic requirements in early embryo, pyruvate was responsible for the nuclear localization of several TCA enzymes that might contribute to epigenetic regulations that were essential for major ZGA activation in the 2-cell stage(Nagaraj et al., 2017). RNA polymerase II subunit genes *Taf*6/9 (Cler, Papai, Schultz, & Davidson, 2009; Juven-Gershon, Hsu, Theisen, & Kadonaga, 2008), cell cycle genes *Cdc20, Mastl, Mad2l1*and *Ube2e1*(C. Guo et al., 2020; Rogers et al., 2018; Wheaton et al., 2017; Q. Wu et al., 2017), and general transcription factor II genes *Gtf2a2/2b*(Roeder, 1996) also exhibited upregulation in mouse zygotes (Figure

4-1D). Genes *Nup35/37/54* encoding nucleoporin (Nup), and gene *Nxt1* encoding the nuclear transport factor 2-like export factor (Xie & Ren, 2019) were upregulated, suggesting increased transport of mRNA and proteins between the nucleus and cytoplasm in mouse zygotes, in line with recent findings that knockdown of the nucleoporin gene *Nup37* led to reduced blastocyte formation rates (Q. Guo et al., 2022) and that massive remodeling of the nuclear envelope occurred during minor zygotic activation in bovine pre-implantation embryos (Popken et al., 2015). Moreover, we found genes encoding essential ZGA transcription factors *Obox1/2/5/7* were all upregulated in mouse zygotes (Figure 4-1D), suggesting that the elevation of transcription levels of these genes during minor ZGA may also be at least partially transcriptional and not solely of maternal origin as reported by a recent study (Ji et al., 2023).

Downregulated genes in rat zygotes were enriched for biological processes or pathways including non-sense mediated decay, RNA polymerase 1 promotor opening, structural constituent of chromatin, etc. (Figure 4-1E, Figure 4-2B). For instance, as in the case of mouse zygotes, genes encoding ribosomal proteins also were downregulated in rat zygotes (Figure 4-1E, Figure 4-2B). Moreover, Genes *NuRD* and *CHD* involved in chromatin remodeling were downregulated (Figure 4-1E). Upregulated genes in rat zygotes were enriched for four biological processes or pathways including carbon metabolism and the TCA cycle, etc. For instance, transcription of the following genes involved in the TCA cycle were all upregulated in rat zygotes: *Aco2* (aconitase 2) (Ciccarone, Vegliante, Di Leo, & Ciriolo, 2017), *Pdha1*(pyruvate dehydrogenase component) (Z. H. Zhou et al., 2001), *Idh1* (pyruvate dehydrogenase) (Ni et al., 2022), *Sdhc*/d (succinate dehydrogenase

subunits) (Rutter, Winge, & Schiffman, 2010), and *Dlst* (dihydrolipoamide S-succinyltransferase) (Berg, Tymoczko, & Stryer, 2002)(Figure 4-1E). This result contrasts with that seen in mouse zygotes where some TCA cycle genes were downregulated while some others were upregulated (Figure 4-1D, Figure 4-2A). Coupled with the aforementioned studies showing that several TCA cycle enzymes were transiently localized in the nucleus of mammalian embryos and essential for major ZGA (Kafkia et al., 2022; Nagaraj et al., 2017), the shared increased expression of the said enzymes in both species suggests that these enzymes may be regulated at the transcriptional level during minor ZGA. Taken together, these results once more suggest that ZGA in the rat might begin earlier than previously believed.

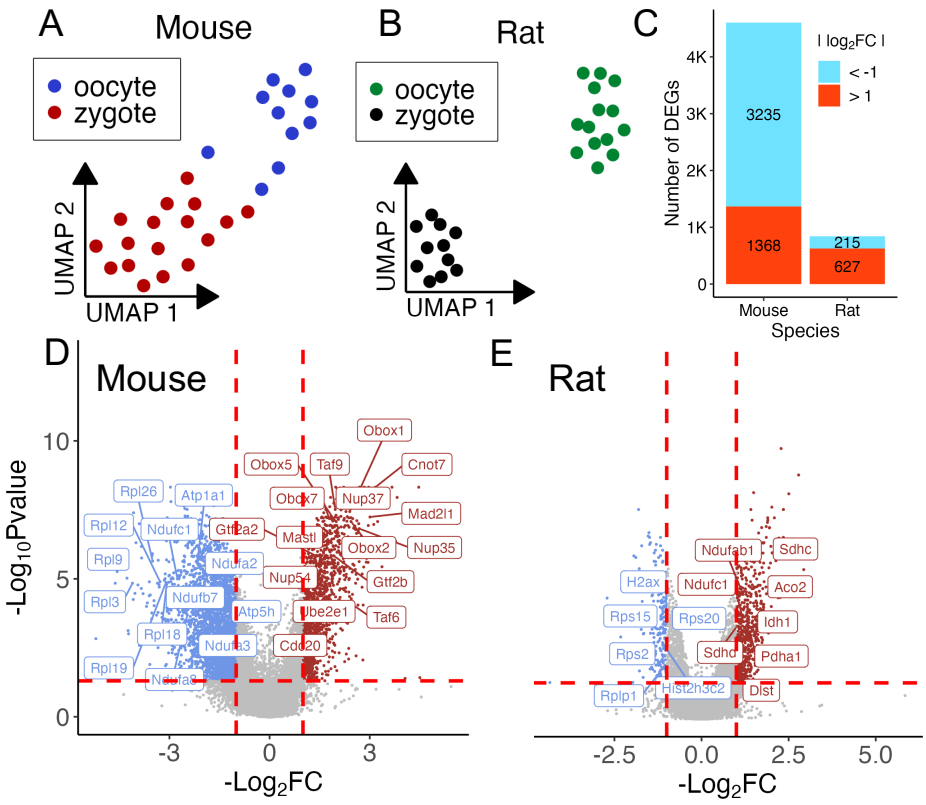


Figure 4-1 Oocytes and zygotes of both species contains distinct sets of transcripts. A, B UMAP visualization of single mouse (A) and rat (B) oocytes and zygotes based on expression levels of genes. C. Number of up- or down-regulated DEGs in mouse and rat zygotes relative to respective oocytes. D, E. Volcano plots of $-Log_{10}$Pvalues vs Log_2FC of mouse (D) and rat (E) genes. Genes marked red were significantly upregulated ($Log_2FC > 1$ and FDR < 0.05), while those marked blue were significantly downregulated ($Log_2FC < -1$ and FDR < 0.05) in zygotes relative to oocytes.

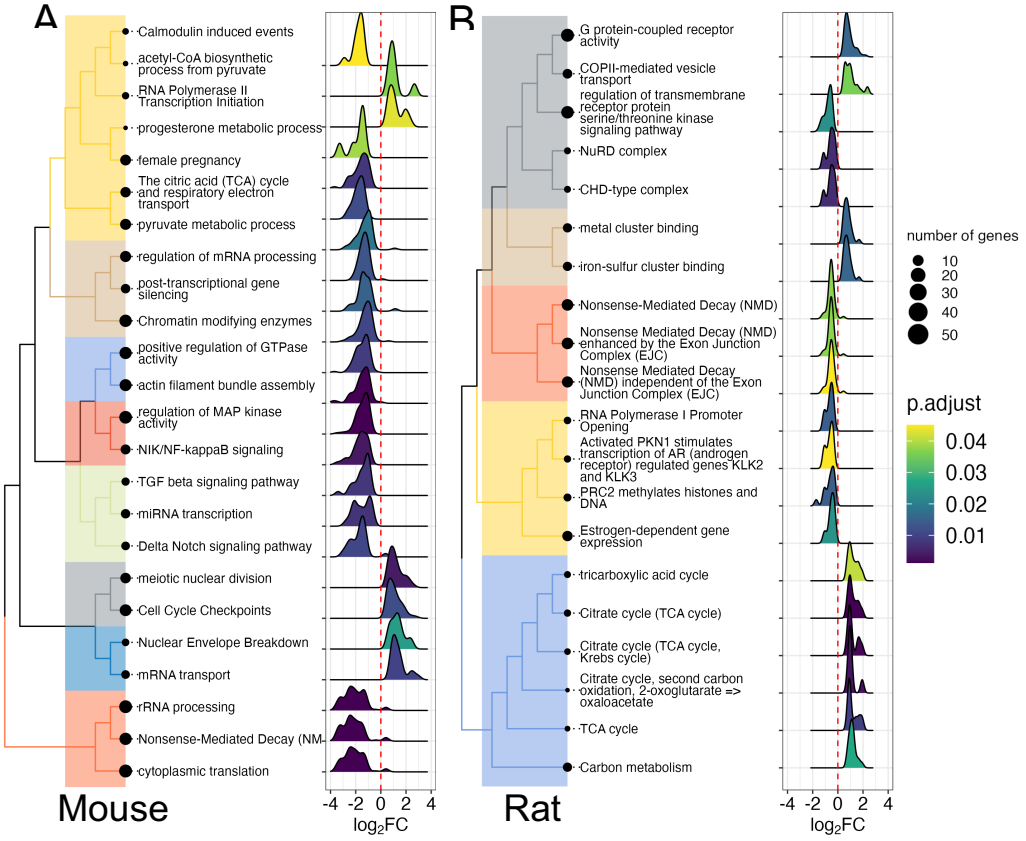


Figure 4-2 A, B. Enriched GO BP terms, Wikipathways, KEGG Pathways and Reactome at FDR < 0.05 in mouse (A) and rat (B). Significant pathways and gene sets are clustered into similar groups based on Jaccard's similarity for ease of viewing, density plots for each gene set/pathway show distribution of the log_2FC of the genes in the indicated GO BP or pathways.

4.3.4 pre-mature mRNA transcription is elevated in zygotes of both species

It has been reported that early transcriptional events in the zygote produced pre-mature (nascent) RNAs that originated from intronic as well as intergenic regions(Ken-Ichiro Abe et al., 2015). We thus estimated nascent (un-spliced) and mature (spliced) transcripts in the samples of both species (Materials and Methods). We detected a total of 26,514 and 21,009 genes with read coverage in the mouse and rat oocytes and zygotes, of which 3,639 (13.8%) and 2903 (13.7%) had nascent reads, respectively (Figure C-2A). In both species, zygotes contained higher proportions of nascent reads than oocytes (Figure 4-3B). We also quantified reads mapped to 1 kb binned intergenic regions as has been done previously(Ken-Ichiro Abe et al., 2015). Zygotes in both species had significantly larger numbers of bins with >1 read coverage than oocytes (Figure 4-3C). These results confirm the previous findings of increased expression of nascent transcripts in 1-cell mouse zygotes, while also suggesting that a similar phenomenon of transcription occurs in the 1-cell rat zygotes. We then performed a DGE analysis of genes with nascent transcripts in oocytes and zygotes (Materials and Methods). As shown in Figure 4-3C, Log_2FC values of nascent transcripts and Log_2FC values of spliced transcripts are strongly correlation, indicating that nascent transcripts mostly change in the same ways as their spliced counterparts.

To identify genes that exhibited significant changes in the proportion of nascent transcripts, we performed a DNP analysis between oocytes and zygotes in both species (Materials and Methods). We found that mouse had more genes with decreased nascent proportions than with increased nascent proportions (113 vs 86), while rat had more genes with increased nascent proportions than with decreased nascent proportions (50 vs 35)

(Figure 4-3D). Comparisons of DEGs and DNP genes in both species found only small overlaps between DEGs and DNPs (Figure 4-3E, 3F). However, in both species, the largest overlap was between upregulated DEGs and genes with decreased nascent proportions (Figure 4-3E, 3F). Although this was expected in rats due to a greater number of upregulated DEGs than downregulated DEGs (Figure 4-1C), it was surprising in mice, since mouse had many more down-regulated DEGs than upregulated DEGs (Figure 4-3E). This result suggests that some of upregulated DEGs in mice could be due in part to more mature transcripts in the zygote cytoplasm that could have been stored as unprocessed transcripts in the oocyte nucleus. Over-representation analysis of these DNP genes in both species yielded no significant pathways or GO BP terms when accounting for false discovery rate. However, we found changes in genes that played important roles in either oocytes or zygotes. For instance, the nascent proportion of the *Oog1/2* genes increased in mouse zygotes (Figure 4-3G), which is known to be required for oocyte development (Minami et al., 2003; Monti & Redi, 2009). The nascent proportion of the *Obox* genes increased in mouse zygotes (Figure 4-3G), consistent with our earlier observation that the *Obox* genes were upregulated in the zygotes as a part of the minor ZGA process, albeit in unprocessed transcript form. The *Nlrp4b/g* genes also exhibited decreased nascent proportions in mouse zygotes (Figure 4-3G), in agreement with the previous reports that the *Nlrp* transcripts were maternally derived and played essential roles in zygotic genome (B.-H. Chang et al., 2013; W. Li, Lin, Peng, & Zhang, 2015). Although DNP genes in rats were less evident, we found the *Zar1* gene to exhibit decreased nascent transcript

proportion (Figure 4-3G), consistent with the previous report that Zar1 played essential roles in zygotic genome activation by regulating maternal derived RNAs (X. Wu et al.).

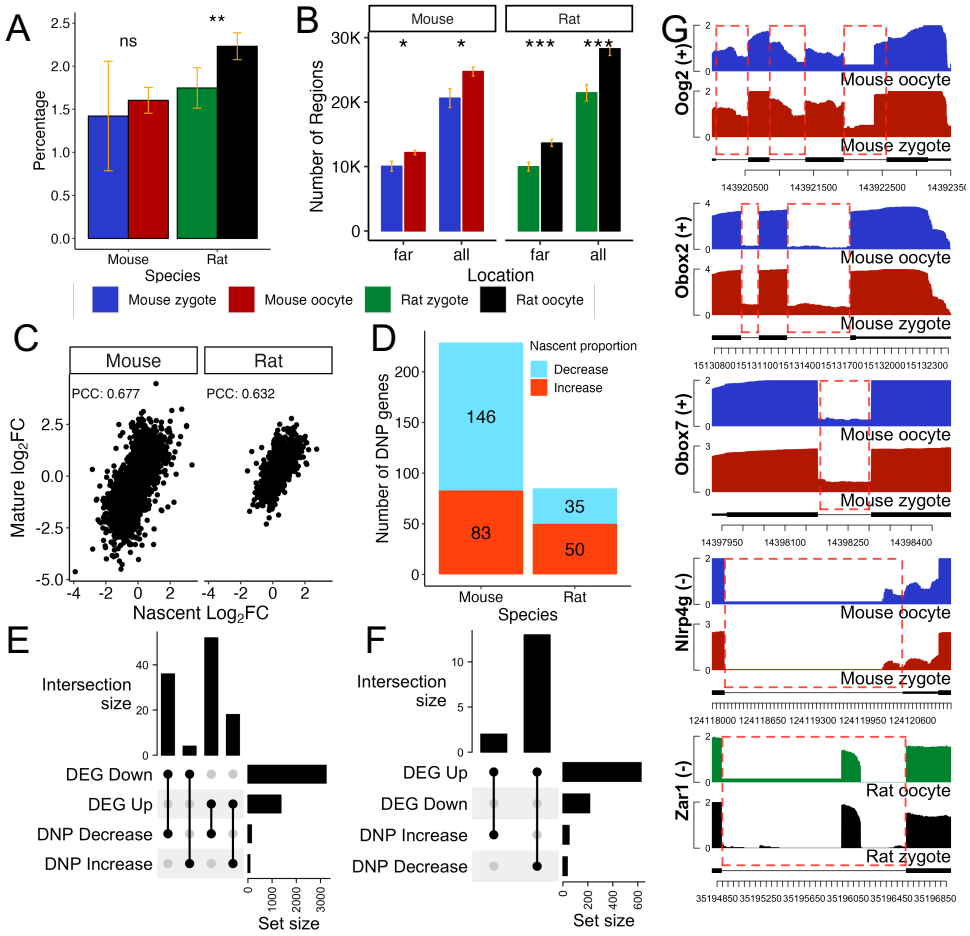


Figure 4-3 Expression of pre-mature mRNA transcripts in oocytes and zygotes of both species. A. Average percentage of total un-spliced reads out of all reads (spliced and un-spliced) in oocytes and zygotes of mice and rats. B. Number of all and far 1kb intergenic regions with reads in oocytes and zygotes of the two species. Far regions are >10kb downstream or upstream of any gene. C. Scatter plots of MAST estimated Log_2FC values of un-spliced reads and Log_2FC of spliced reads for all genes that have un-spliced reads in each species. Pearson's correlation coefficient (PPC) between the values is shown for each species. D. Number of genes with differential nascent proportion (DNP) for each species, which are up- or down-regulated in zygotes. **E, F.** Upset plots of increased/decreased DNP genes and upregulated/downregulated DEGs in mice (E) and rats (F). G. Average intronic coverage for *Oog2, Obox2, Obox7, Nlrp4g* and *Zar1* that were called DNP, in oocytes and zygotes, red boxes highlight regions with intronic coverage. * $p < 0.05$, ** $p<0.01$ and *** $p<0.001$, Student's t-test.

4.3.5 Differential splicing might play a role in MZT in both mice and rats

To uncover genes with splicing changes and isoform switches during MZT, we performed DTU analysis (Materials and Methods) on genes between oocytes and zygotes of both species. We found that 2,601 genes exhibited DTU between mouse oocytes and zygotes, while only 609 genes showed DTU between rat oocytes and zygotes (Figure 4-4A), once more suggests a much more dynamic transcriptional landscape in the mouse zygotes than in rat zygotes. In both species, differential transcript usage can be attributed to the differential expression of no more than two isoforms. (Figure C-2B). Only small portions of DEGs overlapped genes with DTUs and vice versa in both species (Figure 4-4B). Interestingly, the number (308) of DTU genes in mice that also are upregulated DEGs is close to the number (324) of DTU genes that also are downregulated DEGs, despite many more downregulated DEGs in this species (Figure 4-4B). This might suggest that some of the upregulated DEGs might be, at the very least, newly modified transcripts that preside in the cytoplasm of zygotes. DTU genes in mice significantly enriched for GO BP terms such as mitotic cell cycle and protein ubiquitination (Figure 4-4C), while DTU genes in rats did not significantly enrich for any GO BP terms (FDR < 0.05) (Figure 4-4D). Albeit top ranked GO BP terms by p-values (without multiple hypotheses test correction) for both mice and rats are related to cell cycle. DTU genes in both species also are involved in histone modification, chromatin organization, protein ubiquitination and modification (Figures 4C, 4D), suggesting a role of alternative splicing in ZGA associated chromatin remodeling. For instance, we found *Bap* and *Usp3* exhibiting DTU in mice (Figure 4-5A, B), and *Usp16* and *Rnf2* showing DTU in rats (Figure 4-5C, D). These results are consistent

with previous findings that *Usp3* and *Bap1* deubiquitinate H2AK119ub1 (Sharma et al., 2014; Thomas et al., 2023), *Usp16* is the dominant deubiquitinase of H2AK119ub1 in oocytes(Rong et al.), while Rnf2 monoubiquitinates H2A histone (H. Wang et al.). Chromatin undergoes drastic changes during MZT including redistribution of histone H2A among other epigenetic makers, and thus, the control of H2A is vital to ensure proper embryonic development (B. J. Wu et al.). Previous studies have shown that both accumulation and absence of H2A may result in embryonic arrest(Z. Chen, Djekidel, & Zhang, 2021; Rong et al., 2022). Furthermore, we found that some genes with DTU (such as *Usp*16) did not exhibit overall expression level changes in both mice and rats. Thus, the switches in isoform usage may originate from both newly transcribed and degraded transcripts and may play roles in MZT in both mice and rats.

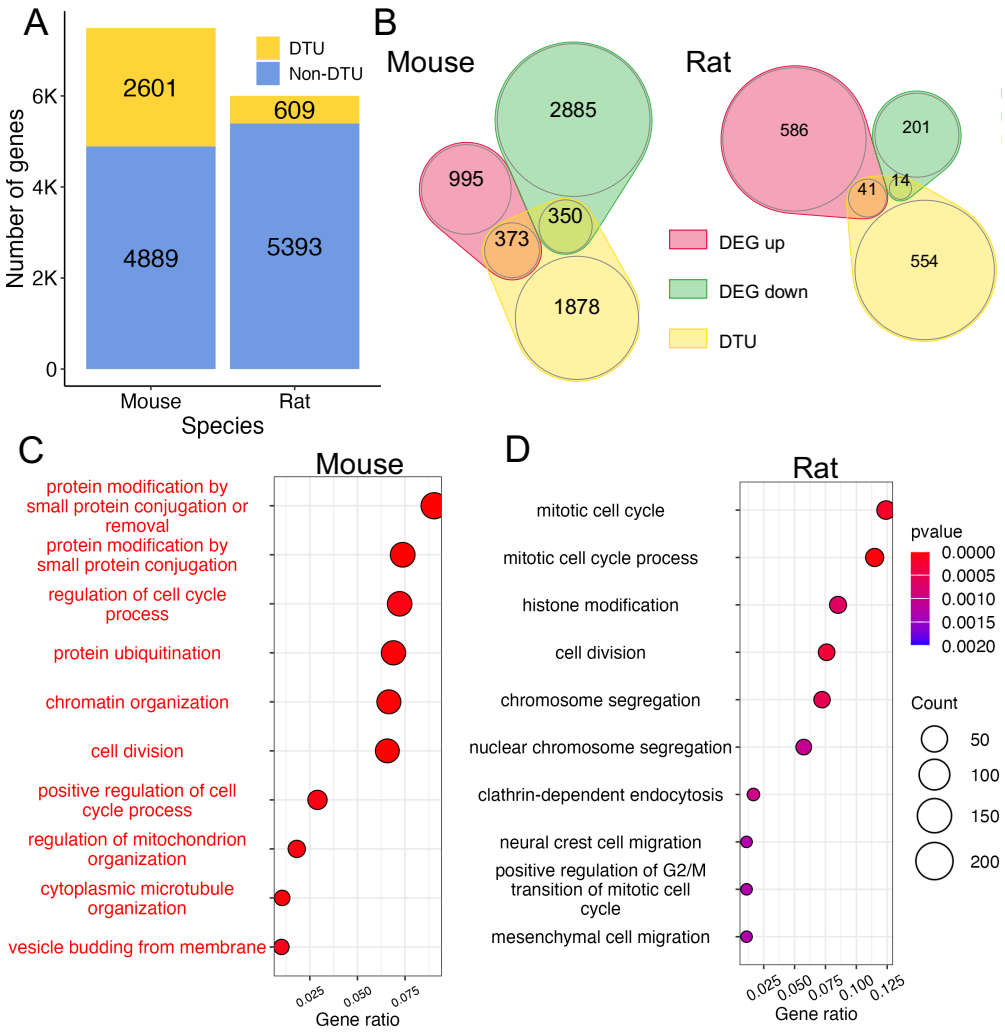


Figure 4-4 Differential splicing of transcripts between oocytes and zygotes of mice and rats. A. Number of transcribed genes with and without DTU (FDR < 0.05) in each species. C. Venn diagram of DTU genes and upregulated or downregulated DEGs in each species. D. GO BP term enrichment for mouse genes with DTU (top 10 terms are shown, names marked in red are significant (FDR < 0.05). E. Go BP term enrichment for rat genes with DTU (top 10 terms are shown). Gene ratio is the ratio of genes in a GO BP term that were called DTU over all genes in that term.

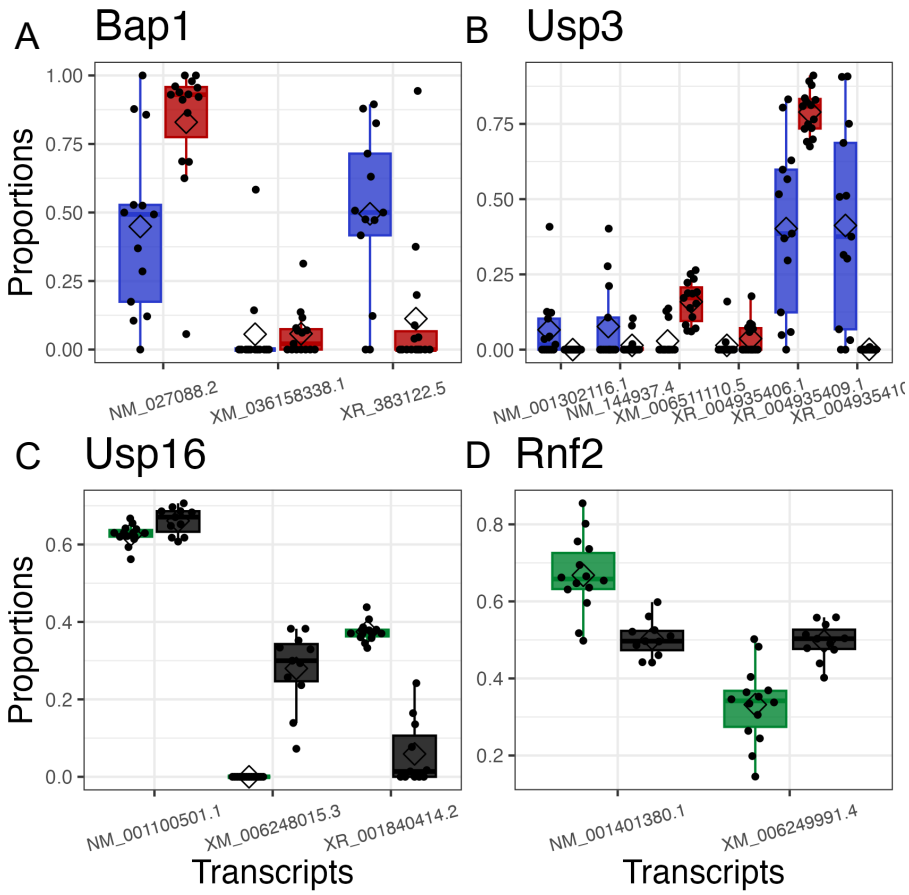


Figure 4-5 A-D Comparison of proportions of isoforms of *Bap1*(A)*, Usp3* (B), *Usp16* (C) and *Rnf2* (D) between oocytes and zygotes of mice and rats.

4.3.6 Genes transcripts in oocytes and zygotes undergo distinct 3'UTR polyadenylations in both mice and rats

Post-transcriptional modifications to mRNA molecules such as alternative use of polyadenylation sites in 3'UTRs, play a critical role in the regulation of mRNA stability and translation (Mayya & Duchaine, 2019; Mignone, Gissi, Liuni, & Pesole, 2002). Thus, we analyzed differences in 3'UTR usage of transcribed genes between oocytes and zygotes in both species by calculating a PDUI (Percentage of Distal poly-A site Usage Index) value (Feng et al., 2018; Xia et al., 2014) for each expressed gene (Materials and Methods).

Oocytes and zygotes in each species could be clearly differentiated based on their PDUI values in UMAP displays (Figure 4-6A, B), indicating that genes expressed in oocytes and zygotes underwent distinct 3'UTR polyadenylations. We then sought to uncover significant differences in 3'UTR usage between oocytes and zygotes (Materials and Methods). In mice, we found that 1020 genes exhibited significantly differential alternative polyadenylation (DAP) between oocytes and zygotes; of which, 785 (76.9%) showed increases in proximal adenylation site usage (preference for shorter 3'UTRs), while the remaining 186 (23.0%) showed preference for distal polyadenylation (preference for longer 3'UTRs) (Figure 4-6C). In rats, we found that 154 genes showed significant DAP between oocytes and zygotes; of which, 106 (55.2%) had increased preference for distal adenylation sites, while the remaining 86 (44.8%) showed increased preference for proximal sites (Figure 4-6C). Thus, the much greater proportion of genes with shortened 3'UTRs in mouse zygotes than in rat zygotes suggests that mouse zygotes undergo much more extensive de-polyadenylation modifications than rat zygotes.

We then investigated the relationships between DEGs and DAP genes. In mice, a considerable number of DAP genes were also DEGs. Specifically, of the mouse DAP genes with shortened 3'UTRs, 288 were upregulated DEGs, but only 35 were downregulated DEG, even though there were fewer upregulated DEGs than downregulated DEGs (Figure 4-1C). Of the mouse DAP genes with lengthened 3'UTRs, 14 were upregulated DEGs, and 37 was downregulated DEGs (Figure 4-6D). These results indicate that DAP genes in mice with shortened 3'UTRs tend to be upregulated DEGs (Figure 4-6D). In addition, we analyzed the Log2FC changes of all genes analyzed for DAP (Figure 4-6E) and found that

though genes with shortened 3'UTRs tended to be upregulated (skewed towards Log2FC > 0), there was no correlation between the magnitude of expression change and change of 3'UTR usage (PDUI difference). We also add that significant DAP genes with lengthened 3'UTRs showed no apparent association with upregulation/downregulation, and that downregulated genes also did not show apparent association with lengthened/shortened 3'UTRs. In rats, only a few DAP genes were also DEGs, with no association between the numbers of DEGs and DAPs (Figure C-3A), thus no confident conclusion could be drawn. DAP genes with shortened 3'UTRs in mice are enriched for mitotic cell cycle transition as well as meiotic division (Figure 4-6F). Although no GO terms are significantly enriched for the DAP genes after false discovery rate correction (FDR < 0.05), top ranked (by p-value) GO terms enriched by significant DAP genes in rats include protein ubiquitination, positive regulation of miRNA mediated gene silencing and positive regulation of G2/M transition of mitotic cell (Figure C-3B).

Many of DAP genes in mice or rats have been known to play various roles in the MZT. We describe two sets of such genes below. The first set are involved in deadenylation and readenylation of mRNAs. We found that *Btg4* and *Cnot7* had DAP with shortened 3'UTRs in mice (Figure 4-6G). It has been shown that CNOT7 along with CNOT6/6l/8 form the CCR4-NOT complex that controls deadenylation of mRNAs (Aslam, Mittal, Koch, Andrau, & Winkler, 2009; Horvat et al., 2018; Jun Ma, Fukuda, & Schultz, 2015; Qian-Qian Sha et al., 2018). It is recently reported that BTG4 functions as a mediator of deadenylation, playing a critical role in the production of substrates for maternal RNA re-adenylation during MZT in humans(Liu et al., 2016; Pasternak et al., 2016; Q. Q. Sha et

al., 2020). It is likely that these deadenylation related genes switch their transcription modes in zygotes to produce transcripts with shorter 3'UTRs via deadenylation followed by re-adenylation, thereby conferring higher translational efficiency of the genes.

The second set of genes are implicated in the regulation of mitotic progression and cell cycle. We found that *Nek7* had DAP with shortened 3'UTR in mouse zygotes (Figure 4-6G). An early study has shown that reduction in *Nek7* activity can cause cells to arrest in metaphase of mitosis (Salem et al., 2010). It is likely that *Nek7* serves a similar function in meiosis through an inhibitory mechanism due to a longer 3'UTR in MII oocytes. Interestingly, unlike the case in mice, several mitosis related DAP genes in rats exhibited lengthened 3'UTR usage. The *Mastl* gene, which encodes an essential regulator of cell cycle control (Rogers et al., 2018), exhibited increased usage of longer 3'UTR transcripts in rat zygotes (Figure 4-6H). Similarly, *Cdc25a* encoding a protein phosphatase that functions as a critical cell cycle regulator (T. Shen & Huang, 2012), exhibited increased usage of longer 3'UTRs (Figure 4-6H). These results suggest different regulatory mechanisms of mRNA stability and protein translation in rat and mouse zygotes during the first mitotic progression.

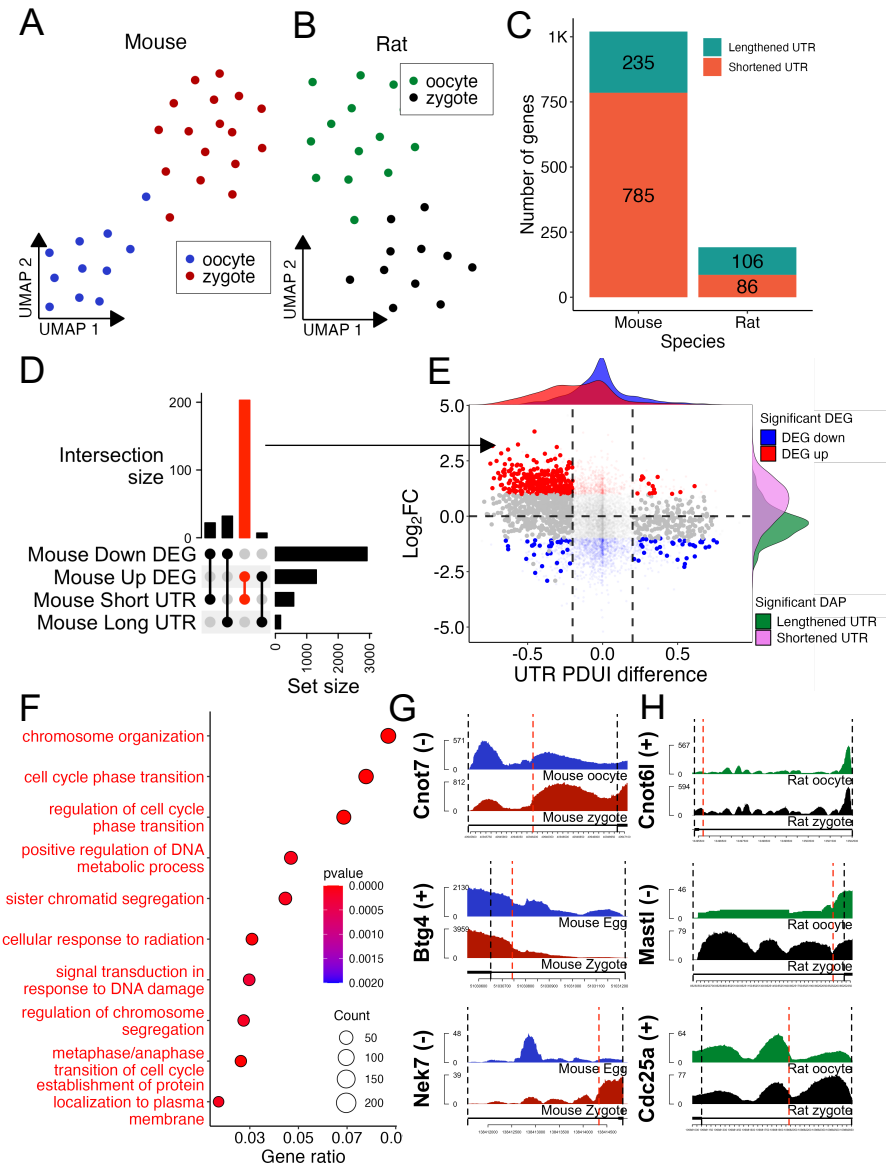


Figure 4-6 Differential alternative polyadenylation of transcripts between oocytes and zygotes of mice and rats. A, B UMAP display of mouse (A) and rat (B) oocytes and zygotes based on PDUI values of genes. C. Number of genes with lengthened and shortened UTRs in zygotes relative to in oocytes of mice and rats. D. Upset plot of DAP genes and DEGs in mice. E. Log_2FC expression vs PDUI difference for all genes that were analyzed for DAP; low transparency points are genes not significant for DAP, red/blue genes are significantly up/down regulated respectively; pink and green density plots show Log2FC distribution for significant DAP genes with lengthened and shortened UTRs; red and blue density plots show PDUI changes for significantly up/down regulated DEGs. F. Significantly enriched GO BP terms for DAP genes in mice (FDR < 0.05). G. Reads coverage on the 3'UTRs of mouse genes *Cnot7, Btg4* and *Nek7*. H. Read coverage on the 3'UTRs of rat genes *Cnot6l, Mastl* and *Cdc25a.* In both H) and I), red lines mark the predicted proximal polyadenylation sites.

4.3.7 Most of orthologous genes in mice and rats have similar while a small portion have opposite transcriptional patterns.

We sought to investigate to what extent DEG, DNP, DTU and DAP genes in mice and rats are conserved based on their orthologous relations. Surprisingly, even though the two species share many orthologous genes (21,824 genes), of which only about 4% (889) were found to be significantly differentially expressed (FDR < 0.05, $| \log_2 FC | > \log_2 1.25$) in zygotes relative to in oocytes in both species. However, most (627, or 70.5%) of the orthologous DEGs showed similar tendency of change (257 upregulated and 370 downregulated) (Figure 4-7A), while the rest small portion (262 or 29.5%) exhibited opposite changes, i.e., 74 upregulated DEGs in mice were downregulated DEGs in rats, and 188 downregulated DEGs in mice were upregulated DEGs in rats (Figure 4-7A). These results suggest that most of the orthologous genes that exhibit changes in both species behave similarly in terms of transcriptional change. These orthologous DEGs are enriched for multiple GO BPs that might be crucial to MZT in the two species (Figure 4-7B, C). Orthologous DEGs that were generally upregulated in both species were enriched for transport of transcripts and the cell cycle, while DEGs that were downregulated in both species were enriched for cytoplasmic translation and Nonsense-mediated decay. Orthologous DEGs that were upregulated in rats but downregulated in mice were enriched for the TCA cycle, RNA splicing and rRNA processing (Figure 4-7B, C).

There are only seven orthologous genes with DNPs in both mice and rats (Figure C-4A). Nonetheless, some may play essential roles in MZT. For instance, the *Oog1* gene

encoding oogenesin exhibited increased nascent proportions in both mouse and rat zygotes (Figure 4-7D). It has been shown that oogenesin plays critical roles in zygotic transcription (Minami et al., 2003). Moreover, *Nlrp4b* showed decreased nascent proportion in both mouse and rat zygotes (Figure 4-7D). Studies have shown that *Nlrp4a-g* may play important roles in early pre-implantation embryos, with high expression of NLRP4G and NLRP4E in both oocytes and zygotes (B.-H. Chang et al., 2013; W. Li et al., 2015).

There were 210 orthologous genes with DTU in both mice and rats (Figure C-4B), they were enriched for positive regulation of transcription elongation by RNA polymerase II (FDR < 0.05). Finally, a total of 53 orthologous genes had DAP in both species, with either lengthened or shortened 3'UTR in both species mostly aligned (Figure 4-7E). Close inspection finds many interesting orthologous DAP genes. For instance, *Eif4e* (Figure 4-7F) encoding a rate-limiting regulator of translation of (Duncan, Milburn, & Hershey, 1987; Hiremath, Webb, & Rhoads, 1985), exhibited preference for shortened 3'UTR in both rat and mouse zygotes (Figure 4-7F), implying higher translational efficiency of the transcripts to prepare the zygotes for higher levels of protein production in the later stages. It has also been shown that Eif4e interacts with CNOT7 via BTG4, thereby expediting maternal mRNA degradation during MZT (Liu et al., 2016; Pasternak et al., 2016). Moreover, both *Usp28* and *Bmi1* exhibited shortened 3'UTRs in both mouse and rat zygotes (Figure 4-7F). Though the products of both genes participate in the regulation of histone H2A, it has been shown that Usp28 is a deubiquitinase of H2A(F. Li et al., 2019), while Bmi1 forms the core protein of the Polycomb Repressive Complex 1 (PRC1), which mono-ubiquitinates H2A histones (Abdouh, Hanna, El Hajjar, Flamier, & Bernier, 2016). This result, along

with the DTU changes mentioned prior, suggests a role of post-transcriptional splicing and polyadenylation regulating chromatin states for ZGA in both species. In addition, both *Anapc1* and *Arid1a* exhibited lengthened 3'UTRs in both mouse and rat zygotes (Figure 4-7F). *Anapc1* encodes a subunit of the Anaphase Promoting Complex/Cyclosome (APC/C), a key ubiquitin-ligase in mitosis. It has recently been shown that delay in the APC/C activation extends mitosis in mouse embryos (Ajeawung et al., 2019). Thus, our results may provide a putative regulatory mechanism of *Anapc1* via 3'UTR extension, which may in turn limit translational efficiency and cause delays in subsequent mitosis (Ajduk, Strauss, Pines, & Zernicka-Goetz, 2017). *Arid1a* encodes a subunit of chromatin remodeling complexes, and its expression in mouse embryos has been found to accumulate during G0 phase and vanish by mitosis (Flores-Alcantar, Gonzalez-Sandoval, Escalante-Alcalde, & Lomelí, 2011). The lengthened 3'UTR of *Arid1a* transcripts in both mouse and rat zygotes may represent a snapshot in the first cell cycle of the embryo, implying a post-transcriptional regulating mechanism through extension of 3'UTR tails (Flores-Alcantar et al., 2011).

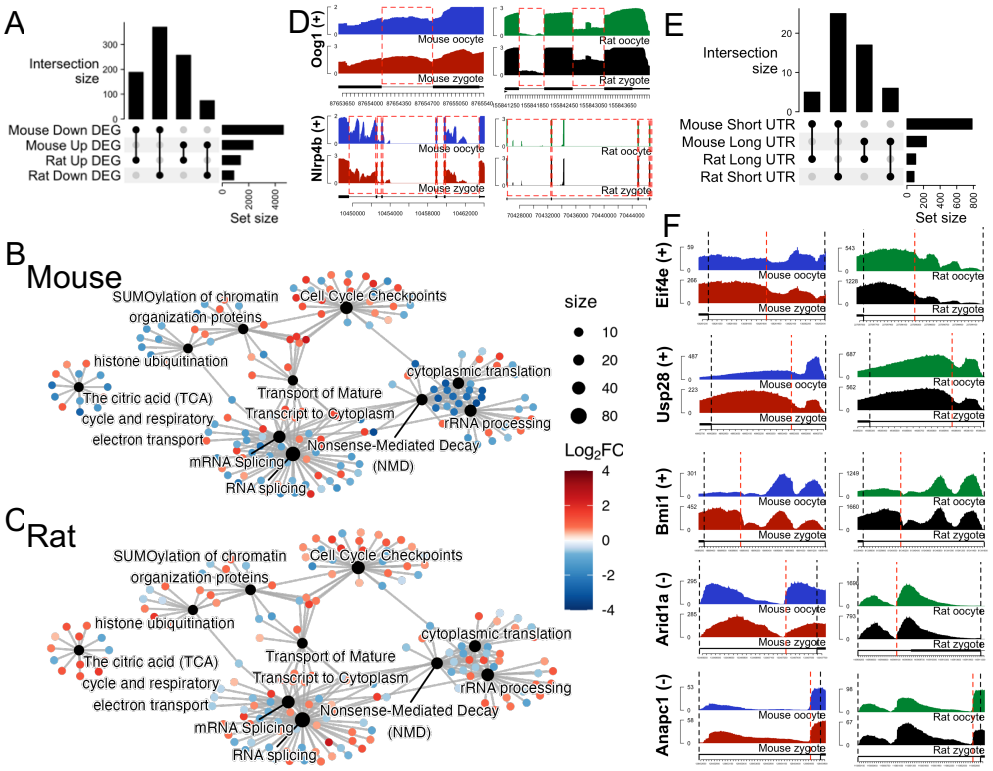


Figure 4-7 Orthologous genes in mice and rats may have similar or opposite transcriptional patterns. A. Upset plot of upregulated and downregulated orthologous DEGs in zygotes relative to oocytes of mice and rats. B, C. Enriched GO BP terms for orthologous DEGs in mice (B) and rats (C). D. Read coverages of orthologous DNP genes *Oog1* and *Nlrp4b* in oocytes and zygotes of mice and rats. E. Upset plot of orthologous DAP genes in mice and rats. F. 3'UTR expression of orthologous DAP genes *Eif4e, Usp28, Bmi1, Arid1a,* and *Anapc1* in mice and rats.

4.4 Discussion

Our analysis of DEGs between oocytes and zygotes of mice and rats revealed new different facets of the underlying workings of their early MZT. Most prominently, we found decreases in ribosomal protein and nonsense-mediated mRNA decay related transcripts in zygotes of both species, though to a lesser degree in rats. We also found significant increases in transcripts of genes related to the cell cycle, mRNA transport and transcription initiation in mouse zygotes. Although most transcripts in mouse zygotes undergo degradation upon fertilization, there are clearly newly produced transcripts. Moreover, the increased transcripts of nucleoporin genes in mouse zygotes suggest increased transport of transcripts into the cytoplasm from the nucleus, which could also explain observed upregulation of many genes even in the absence of true transcription. On the other hand, upregulated genes in rat zygotes are enriched for the TCA cycle, hinting at a requirement for more energy production after fertilization. Conversely, the TCA cycle genes involved in energy production were downregulated in mice, while the genes involved in acetyl-CoA synthesis from pyruvate were upregulated. These results could be due to differences in the timing of events, whereby the mouse zygote prepares its first mitosis earlier and in a more active manner, while the rat zygote has still yet begun degradation of maternal transcripts. However, our detection of numerous upregulated genes in zygotes relative to oocytes in rats, suggesting that MZT might occur in rat zygotes, though at a smaller scale than in mouse zygotes.

Our analysis on intronic transcripts of genes found overall increases in the proportion of intronic reads and number of genes with intronic transcripts in both species

after fertilization, indicating an increase of nascent transcription in zygotes. This was particularly true in mice, despite the majority of DEGs found to be downregulated. Previous studies have suggested that the increases in nascent transcription in the zygote is in part due to promiscuous transcription of genes and inefficient splicing during minor zygotic activation in mice(Ken-Ichiro Abe et al., 2015), and our results suggest that a similar process might occur in the rat zygotes as well. In addition, coupled with the increased expression activity related to RNA transport, particularly in mice, these changes in nascent and mature transcript proportions may also at least partially result from increased RNA transport from the nucleus to the cytosol.

The observed changes in isoform usage between oocytes and zygotes of both species indicate post-transcriptional regulation of genes in zygotes. Interestingly, cell cycle was a top enriched process for genes with differential transcript usage in both mice and rats. It is worth noting that several genes that regulate histone H2A show strong differential isoform usage and differential poly-adenylation in both mice and rats. These results advocate for similar epigenetic regulations of chromatins in both species by differential isoform usage. Though more details remain to be revealed, our results suggest similar post-transcriptional events in zygotes of mice and rats.

Our analysis of differential polyadenylation reveals that transcripts undergo distinct 3'UTR polyadenylations in oocytes and zygotes of both species. We would be remiss to not touch on the issue of differences in gene-body read distribution between oocytes and zygotes and the subsequent impact on alternative polyadenylation analysis. The stronger 3'-end bias of reads in mouse oocytes than in mouse zygotes is unlikely due to technical

factors but rather biological differences between mouse oocytes and zygotes, which is not the case in rat oocytes and rat zygotes. Such bias in mouse samples should be negligible on our DAP analysis, since similar 3'UTR preference of the same gene would result in similar bias observed in read coverage (exhibiting similar peaks in the 3'UTR region), while only differences in 3'UTR length preference (caused by differential poly-A cleavage sites) would result in peaks appearing in one group and not in the other. Another issue that has been brought up in previous studies regarding transcriptomic profiles of 1-cell stage zygotes obtained via poly-A enrichment techniques is inflation of expression values due to inherent biases towards longer Poly-A tails in poly-A capture RNA-seq methodologies (Viscardi & Arribere). Previous studies have also held conflicting views towards the length of poly-A tails in zygotic transcripts. For example, one study (K. Lee, Cho, Morey, & Cook-Andersen, 2024) reported extensive shortening of poly-A tails in the zygote, while another (Liu et al., 2021) suggested a shift towards longer poly-A tail lengths from MII oocytes to zygotes. Though we were not able to quantify for poly-A tail length in our study, we found an overwhelmingly larger overlap between shortened 3'UTRs and upregulated genes in mouse zygotes, compared to other overlaps between DAP and DEG genes. Yet, we note that there is no correlation between the changes in expression levels and the changes in 3'UTR preference measured by PDUI. Taken together, it is unclear whether the observed upregulation is indirectly caused by changes in poly-adenylation, and it appears that many other genes do not experience significant changes in polyadenylation but are upregulated regardless in mouse zygotes. The case in rats is even more prominent, with most upregulated genes experiencing no significant change in 3'UTR preference.

Although most orthologous genes in mice and rats had similar differential transcriptional patterns in term of DGEs, DNT, DTU and DAP during MZT, others showed different patterns. In summary, these orthogonal analyses on the transcriptomic changes during the oocyte to zygote transition in mice and rats strongly suggest minor zygotic activation is a vibrant process in both species, albeit a delayed and weaker process in rats. Much of the observed results also confirm significant post-transcriptional regulation, most notably in the form of alternative polyadenylation, in both species.

Future studies should involve a multi-omics approach for both species, with strong emphasis on post-transcriptional modifications such as poly-adenylation and epigenetic regulations of expression. Furthermore, the lack of transcript degradation in rat zygotes may imply that timing of rat zygotes and mouse zygotes differ in the progression of MZT. However, there was a much greater number of upregulated DEGs observed in the rat, perhaps owing to transient expressions of the maternal genome upon fertilization. Thus, a more comprehensive analysis could focus on finer-grained snapshots of the zygote in both mice and rats to capture transient expressions of genes, as well as timing of minor zygotic activation.

CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS

Gene expression regulation plays a crucial role in processes ranging from stress responses to development. In this dissertation, using full-length scRNA-seq methods, we investigated the dynamic nature of gene expression and RNA processing in four model organisms: *Saccharomyces cerevisiae* (yeast), *Caenorhabditis elegans* (*C. elegans*), *Mus musculus* (mouse) and *Rattus Norvegicus* (rat). Through comprehensive studies on the transcriptomic landscape of these organisms, we elucidate the intricate regulatory mechanisms that govern cellular responses to environmental cues and developmental signals. These studies emphasize the dynamic nature of gene expression regulation through transcriptional, post-transcriptional, and alternative polyadenylation (APA) mechanisms.

In chapter two, we demonstrate the possibility of sequencing single-cell full-length mRNA transcriptomes in small *S. cerevisiae* cells, generated a scRNA-seq data set in the organism under isotonic, hypotonic, AAS and GS conditions. Our analysis of these data provided insights into genes that are differentially expressed in *S. cerevisiae* under the stressors. Our results align closely with early studies into the stress response of *S. cerevisiae*, while also providing unique findings. We highlight the preferential increase in the expression of genes for the biosynthesis of methionine over other amino acids when all amino acids are absent. In addition, gene regulatory networks underlying the transcriptional responses are highlighted through the analysis of TFBSs for each of the DEGs. The DEGs and gene regulatory networks uncovered in this study provide a single cell based transcriptomic view for future studies of stress-induced responses in this model organism. Moreover, we show that biological variability between cells under the same conditions due

to stochastic events can be utilized to discover regulatory relationships.

In chapter three, we present a scRNA-seq dataset of the C. elegans oocyte oogenesis and fertilization via micro-dissection and full-length sequencing. Our analysis of these data show accordance with earlier studies while also revealing deeper insights into the complex transcriptional landscape of germline development, oogenesis, and fertilization in C. elegans. We take note of the intricacies of dissection and complexity of somatic contamination and take delicate steps into removing unwanted factors. Our results highlight distinct transcriptomic signals of oogenesis, such as different patterns of expression of apoptotic genes. In addition, we find significant changes to poly-adenylation sites involved in the transcripts encoding mitotic factors upon oocyte fertilization. Furthermore, we uncover significant changes in the usage of isoforms of hemichannels genes associated with germline and somatic communication.

In chapter four, we focus on the transcriptional dynamics in oocytes and zygotes of mice and rats. Via different aspects of the transcriptome, we find signals of transcription in the zygotes of rats, as well as provided a more detailed look into the MZT of both organisms. Upon further inspection of 3'UTR coverage of transcripts, we highlight a considerable overlap between genes showing shortened 3'UTRs and those that showed increased expression in mice, with a significant portion of these genes involved in mitosis related processes. Moreover, we perform comparative analysis of orthologous genes between the two species and find similar tendencies in differentially expressed genes, as well as some strikingly similar patterns of 3'UTR changes in genes involved in early stages of zygotic development.

Overall, whether it's stress response in yeast, germline development in C. elegans, or early embryogenesis in mice and rats, these processes are tightly regulated by complex transcriptional and post-transcriptional mechanisms. Single cell approaches and detailed analyses of various aspects of the transcriptome reveal how these regulatory layers contribute to cellular adaptation, differentiation, and development.

Future directions along this line of work can help further illuminate the finer details of these processes. Specifically, further refinement of single-cell RNA sequencing methods, especially for organisms with challenging cell structures like yeast, can provide even deeper insights into cellular heterogeneity and stress responses. Additionally, expanding studies of DAP/DEG/DTU and other regulatory mechanisms across different developmental stages will shed light on how these different aspects of the transcriptome contributes to critical windows of regulatory shifts, while performing these studies in more species will help elucidate evolutionary conservation and divergence in gene regulation and uncover fundamental principles applicable across taxa. Furthermore, application of multi-omics-based approaches these studies can provide a much more comprehensive understanding of the molecular underpinnings of development and stress responses. Undoubtedly, experimental validation of key regulatory genes and pathways identified in these studies will be crucial and applying traditional methods such as RT-PCR can help validate expression changes of candidate genes while techniques like CRISPR/Cas9-mediated gene editing and RNA interference can be employed to dissect the functional roles of candidate genes in vivo.

By pursuing these future directions, we can build on the foundational insights

provided by these studies, advancing our understanding of gene regulation in diverse biological contexts.

REFERENCES

Abdouh, M., Hanna, R., El Hajjar, J., Flamier, A., & Bernier, G. (2016). The Polycomb Repressive Complex 1 Protein BMI1 Is Required for Constitutive Heterochromatin Formation and Silencing in Mammalian Somatic Cells. *Journal of Biological Chemistry, 291*(1), 182-197. doi:10.1074/jbc.m115.662403

Abe, K. I., Funaya, S., Tsukioka, D., Kawamura, M., Suzuki, Y., Suzuki, M. G., . . . Aoki, F. (2018). Minor zygotic gene activation is essential for mouse preimplantation development. *Proc Natl Acad Sci U S A, 115*(29), E6780-E6788. doi:10.1073/pnas.1804309115

Abe, K. I., Yamamoto, R., Franke, V., Cao, M., Suzuki, Y., Suzuki, M. G., . . . Aoki, F. (2015). The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3′ processing. *The EMBO Journal, 34*(11), 1523-1537. doi:10.15252/embj.201490648

Ajduk, A., Strauss, B., Pines, J., & Zernicka-Goetz, M. (2017). Delayed APC/C activation extends the first mitosis of mouse embryos. *Scientific Reports, 7*(1), 9682. doi:10.1038/s41598-017-09526-1

Ajeawung, N. F., Nguyen, T. T. M., Lu, L., Kucharski, T. J., Rousseau, J., Molidperee, S., . . . Campeau, P. M. (2019). Mutations in ANAPC1, Encoding a Scaffold Subunit of the Anaphase-Promoting Complex, Cause Rothmund-Thomson Syndrome Type 1. *The American Journal of Human Genetics, 105*(3), 625-630. doi:10.1016/j.ajhg.2019.06.011

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics, 31*(2), 166-169. doi:10.1093/bioinformatics/btu638

Anders, S., Reyes, A., & Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Research, 22*(10), 2008-2017. doi:10.1101/gr.133744.111

Aoki, F., Worrad, D. M., & Schultz, R. M. (1997). Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev Biol, 181*(2), 296-307. doi:10.1006/dbio.1996.8466

Aranda, A., & del Olmo Ml, M. (2003). Response to acetaldehyde stress in the yeast Saccharomyces cerevisiae involves a strain-dependent regulation of several ALD genes and is mediated by the general stress response pathway. *Yeast, 20*(8), 747-759. doi:10.1002/yea.991

Arndt, K., & Fink, G. R. (1986). GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. *Proc Natl Acad Sci U S A, 83*(22), 8516-8520. doi:10.1073/pnas.83.22.8516

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet, 25*(1), 25-29. doi:10.1038/75556

Ashe, M. P., De Long, S. K., & Sachs, A. B. (2000). Glucose depletion rapidly inhibits translation initiation in yeast. *Mol Biol Cell, 11*(3), 833-848. doi:10.1091/mbc.11.3.833

Aslam, A., Mittal, S., Koch, F., Andrau, J.-C., & Winkler, G. S. (2009). The Ccr4–Not Deadenylase Subunits CNOT7 and CNOT8 Have Overlapping Roles and Modulate Cell Proliferation. *Molecular Biology of the Cell, 20*(17), 3840-3850. doi:10.1091/mbc.e09-02-0146

Bacher, R., Chu, L. F., Leng, N., Gasch, A. P., Thomson, J. A., Stewart, R. M., . . . Kendziorski, C. (2017). SCnorm: robust normalization of single-cell RNA-seq data. *Nat Methods, 14*(6), 584-586. doi:10.1038/nmeth.4263

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., . . . Bulyk, M. L. (2009). Diversity and complexity in DNA recognition by transcription factors. *Science, 324*(5935), 1720-1723. doi:10.1126/science.1162327

Bahn, Y. S., Xue, C., Idnurm, A., Rutherford, J. C., Heitman, J., & Cardenas, M. E. (2007). Sensing the environment: lessons from fungi. *Nat Rev Microbiol, 5*(1), 57-69. doi:10.1038/nrmicro1578

Batiza, A. F., Schulz, T., & Masson, P. H. (1996). Yeast respond to hypotonic shock with a calcium pulse. *J Biol Chem, 271*(38), 23357-23362. doi:10.1074/jbc.271.38.23357

Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L., & Hunter, C. P. (2003). Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. *Development, 130*(5), 889-900.

Berg, J. M., Tymoczko, J. L., & Stryer, L. (2002). *Biochemistry, Fifth Edition*: W.H. Freeman.

Blake, J. A., Baldarelli, R., Kadin, J. A., Richardson, J. E., Smith, L., Cynthia, Bult, C. J., . . . Zhu, Y. S. (2021). Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology. *Nucleic Acids Research, 49*(D1), D981-D987. doi:10.1093/nar/gkaa1083

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. S. (2015). Near-optimal RNA-Seq quantification. *ArXiv, abs/1505.02710*.

Brennecke, P., Anders, S., Kim, J. K., Kolodziejczyk, A. A., Zhang, X., Proserpio, V., . . . Heisler, M. G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods, 10*(11), 1093-1095. doi:10.1038/nmeth.2645

Brunquell, J., Morris, S., Lu, Y., Cheng, F., & Westerheide, S. D. (2016). The genome-wide role of HSF-1 in the regulation of gene expression in Caenorhabditis elegans. *BMC Genomics, 17*(1). doi:10.1186/s12864-016-2837-5

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., . . . Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature, 523*(7561), 486-490. doi:10.1038/nature14590

Buttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A., & Theis, F. J. (2019). A test metric for assessing single-cell RNA-seq batch correction. *Nat Methods, 16*(1), 43-49. doi:10.1038/s41592-018-0254-1

Buziol, S., Becker, J., Baumeister, A., Jung, S., Mauch, K., Reuss, M., & Boles, E. (2002). Determination of in vivo kinetics of the starvation-induced Hxt5 glucose

transporter of Saccharomyces cerevisiae. *FEMS Yeast Res, 2*(3), 283-291. doi:10.1016/S1567-1356(02)00113-7

Casal, M., Paiva, S., Andrade, R. P., Gancedo, C., & Leao, C. (1999). The lactate-proton symport of Saccharomyces cerevisiae is encoded by JEN1. *J Bacteriol, 181*(8), 2620-2623. doi:10.1128/JB.181.8.2620-2623.1999

Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., . . . Young, R. A. (2001). Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell, 12*(2), 323-337. doi:10.1091/mbc.12.2.323

Chambers, P., Issaka, A., & Palecek, S. P. (2004). Saccharomyces cerevisiae JEN1 promoter activity is inversely related to concentration of repressing sugar. *Appl Environ Microbiol, 70*(1), 8-17. doi:10.1128/AEM.70.1.8-17.2004

Chang, B.-H., Liu, X., Liu, J., Quan, F.-S., Guo, Z.-K., & Zhang, Y. (2013). Developmental expression and possible functional roles of mouse Nlrp4e in preimplantation embryos. *In Vitro Cellular & Developmental Biology - Animal, 49*(7), 548-553. doi:10.1007/s11626-013-9638-9

Chang, Y., & Huh, W. K. (2018). Ksp1-dependent phosphorylation of eIF4G modulates post-transcriptional regulation of specific mRNAs under glucose deprivation conditions. *Nucleic Acids Res, 46*(6), 3047-3060. doi:10.1093/nar/gky097

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software, 61*(6), 1 - 36. doi:10.18637/jss.v061.i06

Chen, G., Ning, B., & Shi, T. (2019). Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front Genet, 10*, 317. doi:10.3389/fgene.2019.00317

Chen, Y.-Z., Mapes, J., Lee, E.-S., Robert Skeen-Gaar, R., & Xue, D. (2013). Caspase-mediated activation of Caenorhabditis elegans CED-8 promotes apoptosis and phosphatidylserine externalization. *Nature Communications, 4*(1). doi:10.1038/ncomms3726

Chen, Z., Djekidel, M. N., & Zhang, Y. (2021). Distinct dynamics and functions of H2AK119ub1 and H3K27me3 in mouse preimplantation embryos. *Nature Genetics, 53*(4), 551-563. doi:10.1038/s41588-021-00821-2

Chiorazzi, M., Rui, L., Yang, Y., Ceribelli, M., Tishbi, N., Maurer, C. W., . . . Staudt, L. M. (2013). Related F-box proteins control cell death in Caenorhabditis elegans and human lymphoma. *Proceedings of the National Academy of Sciences, 110*(10), 3943-3948. doi:10.1073/pnas.1217271110

Chu, D. S., & Shakes, D. C. (2013). Spermatogenesis. *Adv Exp Med Biol, 757:171-203.*(doi), 10.1007/1978-1001-4614-4015-1004_1007.

Chua, G., Morris, Q. D., Sopko, R., Robinson, M. D., Ryan, O., Chan, E. T., . . . Hughes, T. R. (2006). Identifying transcription factor functions and targets by phenotypic activation. *Proc Natl Acad Sci U S A, 103*(32), 12045-12050. doi:10.1073/pnas.0605140103

Ciccarone, F., Vegliante, R., Di Leo, L., & Ciriolo, M. R. (2017). The TCA cycle as a bridge between oncometabolism and DNA transactions in cancer. *Seminars in Cancer Biology, 47*, 50-56. doi:10.1016/j.semcancer.2017.06.008

Cler, E., Papai, G., Schultz, P., & Davidson, I. (2009). Recent advances in understanding the structure and function of general transcription factor TFIID. *Cellular and Molecular Life Sciences, 66*(13), 2123-2134. doi:10.1007/s00018-009-0009-3

Conrad, M., Schothorst, J., Kankipati, H. N., Van Zeebroeck, G., Rubio-Texeira, M., & Thevelein, J. M. (2014). Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae. *FEMS Microbiol Rev, 38*(2), 254-299. doi:10.1111/1574-6976.12065

Consortium, C. e. S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science, 282*(5396), 2012-2018.

Cormack, B. P., & Struhl, K. (1992). The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell, 69*(4), 685-696. doi:10.1016/0092-8674(92)90232-2

Crittenden, S. L., Bernstein, D. S., Bachorik, J. L., Thompson, B. E., Gallegos, M., Petcherski, A. G., . . . Kimble, J. (2002). A conserved RNA-binding protein controls germline stem cells in Caenorhabditis elegans. *Nature, 417*(6889), 660-663. doi:10.1038/nature754

Davis, P., Zarowiecki, M., Arnaboldi, V., Becerra, A., Cain, S., Chan, J., . . . Sternberg, P. W. (2022). WormBase in 2022-data, processes, and tools for analyzing Caenorhabditis elegans. *Genetics, 220*(4). doi:10.1093/genetics/iyac003

Denning, D. P., Hatch, V., & Horvitz, H. R. (2013). Both the Caspase CSP-1 and a Caspase-Independent Pathway Promote Programmed Cell Death in Parallel to the Canonical Pathway for Apoptosis in Caenorhabditis elegans. *PLoS Genetics, 9*(3), e1003341. doi:10.1371/journal.pgen.1003341

Dever, T. E., Feng, L., Wek, R. C., Cigan, A. M., Donahue, T. F., & Hinnebusch, A. G. (1992). Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. *Cell, 68*(3), 585-596. doi:10.1016/0092-8674(92)90193-g

Diag, A., Schilling, M., Klironomos, F., Ayoub, S., & Rajewsky, N. (2018). Spatiotemporal m(i)RNA Architecture and 3' UTR Regulation in the C. elegans Germline. *Dev Cell, 47*(6), 785-800 e788. doi:10.1016/j.devcel.2018.10.005

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Duncan, R., Milburn, S. C., & Hershey, J. W. (1987). Regulated phosphorylation and low abundance of HeLa cell initiation factor eIF-4F suggest a role in translational control. Heat shock effects on eIF-4F. *J Biol Chem, 262*(1), 380-388.

Dunn, B., & Wobbe, C. R. (2001). Preparation of protein extracts from yeast. *Curr Protoc Mol Biol, Chapter 13*, Unit13 13. doi:10.1002/0471142727.mb1313s23

Eckmann, C. R., Crittenden, S. L., Suh, N., & Kimble, J. (2004). GLD-3 and control of the mitosis/meiosis decision in the germline of Caenorhabditis elegans. *Genetics, 168*(1), 147-160. doi:10.1534/genetics.104.029264

Eichhorn, S. W., Subtelny, A. O., Kronja, I., Kwasnieski, J. C., Orr-Weaver, T. L., & Bartel, D. P. (2016). mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. *Elife, 5*. doi:10.7554/eLife.16955

Feng, X., Li, L., Wagner, E. J., & Li, W. (2018). TC3A: The Cancer 3′ UTR Atlas. *Nucleic Acids Research, 46*(D1), D1027-D1030. doi:10.1093/nar/gkx892

Ferreira, C., van Voorst, F., Martins, A., Neves, L., Oliveira, R., Kielland-Brandt, M. C., . . . Brandt, A. (2005). A member of the sugar transporter family, Stl1p is the glycerol/H+ symporter in Saccharomyces cerevisiae. *Mol Biol Cell, 16*(4), 2068-2076. doi:10.1091/mbc.e04-10-0884

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., . . . Gottardo, R. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol, 16*, 278. doi:10.1186/s13059-015-0844-5

Flores-Alcantar, A., Gonzalez-Sandoval, A., Escalante-Alcalde, D., & Lomelí, H. (2011). Dynamics of expression of ARID1A and ARID1B subunits in mouse embryos and in cells during the cell cycle. *Cell and Tissue Research, 345*(1), 137-148. doi:10.1007/s00441-011-1182-x

Gao, M. X., Liao, E. H., Yu, B., Wang, Y., Zhen, M., & Derry, W. B. (2008). The SCFFSN-1 ubiquitin ligase controls germline apoptosis through CEP-1/p53 in C. elegans. *Cell Death & Differentiation, 15*(6), 1054-1062. doi:10.1038/cdd.2008.30

Garay-Arroyo, A., & Covarrubias, A. A. (1999). Three genes whose expression is induced by stress in Saccharomyces cerevisiae. *Yeast, 15*(10A), 879-892. doi:10.1002/(SICI)1097-0061(199907)15:10A<879::AID-YEA428>3.0.CO;2-Q

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., . . . Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell, 11*(12), 4241-4257. doi:10.1091/mbc.11.12.4241

Gasch, A. P., & Werner-Washburne, M. (2002). The genomics of yeast responses to environmental stress and starvation. *Funct Integr Genomics, 2*(4-5), 181-192. doi:10.1007/s10142-002-0058-2

Gasch, A. P., Yu, F. B., Hose, J., Escalante, L. E., Place, M., Bacher, R., . . . McClean, M. N. (2017). Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress. *PLoS Biol, 15*(12), e2004050. doi:10.1371/journal.pbio.2004050

Gene Ontology, C. (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res, 49*(D1), D325-D334. doi:10.1093/nar/gkaa1113

Genomics X. 1.3 million brain cells from E18 mice. (2017).

Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., . . . Waterston, R. H. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science, 330*(6012), 1775-1787. doi:science.1196914 [pii]

10.1126/science.1196914 [doi]

Girardot, F., Monnier, V., & Tricoire, H. (2004). Genome wide analysis of common and specific stress responses in adult drosophila melanogaster. *BMC Genomics, 5*, 74. doi:10.1186/1471-2164-5-74

Gonzalez, D. P., Lamb, H. V., Partida, D., Wilson, Z. T., Harrison, M. C., Prieto, J. A., . . . Olson, S. K. (2018). CBD-1 organizes two independent complexes required for eggshell vitelline layer formation and egg activation in C. elegans. *Dev Biol, 442*(2), 288-300. doi:10.1016/j.ydbio.2018.08.005

Gordan, R., Murphy, K. F., McCord, R. P., Zhu, C., Vedenko, A., & Bulyk, M. L. (2011). Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. *Genome Biol, 12*(12), R125. doi:gb-2011-12-12-r125 [pii]

10.1186/gb-2011-12-12-r125 [doi]

Gu, Z. (2022). Complex heatmap visualization. *iMeta, 1*(3). doi:10.1002/imt2.43

Gumienny, T. L., Lambie, E., Hartwieg, E., Horvitz, H. R., & Hengartner, M. O. (1999). Genetic control of programmed cell death in the *Caenorhabditis elegans* hermaphrodite germline. *Development, 126*(5), 1011-1022. doi:10.1242/dev.126.5.1011

Guo, C., Kong, F., Lv, Y., Gao, N., Xiu, X., & Sun, X. (2020). CDC20 inhibitor Apcin inhibits embryo implantation in vivo and in vitro. *Cell Biochemistry and Function, 38*(6), 810-816. doi:10.1002/cbf.3550

Guo, Q., Liu, Q., Wang, N., Wang, J., Sun, A., Qiao, J., & Yan, L. (2022). The function of Nucleoporin 37 on mouse oocyte maturation and preimplantation embryo development. *Journal of Assisted Reproduction and Genetics, 39*(1), 107-116. doi:10.1007/s10815-021-02330-x

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., & Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biol, 8*(2), R24. doi:10.1186/gb-2007-8-2-r24

Haghverdi, L., Lun, A. T. L., Morgan, M. D., & Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol, 36*(5), 421-427. doi:10.1038/nbt.4091

Hammarlund, M., Hobert, O., Miller, D. M., & Sestan, N. (2018). The CeNGEN Project: The Complete Gene Expression Map of an Entire Nervous System. *Neuron, 99*(3), 430-433. doi:10.1016/j.neuron.2018.07.042

Haque, A., Engel, J., Teichmann, S. A., & Lonnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med, 9*(1), 75. doi:10.1186/s13073-017-0467-4

Harris, T. W., Arnaboldi, V., Cain, S., Chan, J., Chen, W. J., Cho, J., . . . Sternberg, P. W. (2019). WormBase: a modern Model Organism Information Resource. *Nucleic Acids Research*. doi:10.1093/nar/gkz920

Hartley, S. W., & Mullikin, J. C. (2015). QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics, 16*, 224. doi:10.1186/s12859-015-0670-5

Hashimshony, T., Wagner, F., Sher, N., & Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep, 2*(3), 666-673. doi:10.1016/j.celrep.2012.08.003

Henderson, S. T., Gao, D., Lambie, E. J., & Kimble, J. (1994). *Lag-2* may encode a signaling ligand for the GLP-1 and LIN-12 receptors of *C. elegans*. *Development, 120*(10), 2913-2924. doi:10.1242/dev.120.10.2913

Hengartner, M. O., Ellis, R. E., & Horvitz, H. R. (1992). Caenorhabditis elegans gene ced-9 protects cells from programmed cell death. *Nature, 356*(6369), 494-499. doi:10.1038/356494a0

Hillier, L. W., Coulson, A., Murray, J. I., Bao, Z., Sulston, J. E., & Waterston, R. H. (2005). Genomics in C. elegans: so many genes, such a little worm. *Genome Res, 15*(12), 1651-1660.

Hiltunen, J. K., Wenzel, B., Beyer, A., Erdmann, R., Fossa, A., & Kunau, W. H. (1992). Peroxisomal multifunctional beta-oxidation protein of Saccharomyces cerevisiae. Molecular analysis of the fox2 gene and gene product. *J Biol Chem, 267*(10), 6646-6653.

Hinnebusch, A. G. (1986). The general control of amino acid biosynthetic genes in the yeast Saccharomyces cerevisiae. *CRC Crit Rev Biochem, 21*(3), 277-317. doi:10.3109/10409238609113614

Hinnebusch, A. G. (1997). Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome. *J Biol Chem, 272*(35), 21661-21664. doi:10.1074/jbc.272.35.21661

Hiremath, L. S., Webb, N. R., & Rhoads, R. E. (1985). Immunological detection of the messenger RNA cap-binding protein. *J Biol Chem, 260*(13), 7843-7849.

Hoeppner, D. J., Spector, M. S., Ratliff, T. M., Kinchen, J. M., Granat, S., Lin, S. C., . . . Hengartner, M. O. (2004). eor-1 and eor-2 are required for cell-specific apoptotic death in C. elegans. *Dev Biol, 274*(1), 125-138. doi:10.1016/j.ydbio.2004.06.022

Hohmann, S. (2009). Control of high osmolarity signalling in the yeast Saccharomyces cerevisiae. *FEBS Lett, 583*(24), 4025-4029. doi:10.1016/j.febslet.2009.10.069

Hohmann, S., & Mager, W. H. (2007). *Yeast stress responses* (Vol. 1): Springer Science & Business Media.

Horvat, F., Fulka, H., Jankele, R., Malik, R., Jun, M., Solcova, K., . . . Svoboda, P. (2018). Role of *Cnot6l* in maternal mRNA turnover. *Life Science Alliance, 1*(4), e201800084. doi:10.26508/lsa.201800084

Huang, C.-C., Hall, D. H., Hedgecock, E. M., Kao, G., Karantza, V., Vogel, B. E., . . . Wadsworth, W. G. (2003). Laminin α subunits and their role in *C. elegans* development. *Development, 130*(14), 3343-3358. doi:10.1242/dev.00481

Huang, W., Jiang, T., Choi, W., Qi, S., Pang, Y., Hu, Q., . . . Shi, Y. (2013). Mechanistic insights into CED-4-mediated activation of CED-3. *Genes & Development, 27*(18), 2039-2048. doi:10.1101/gad.224428.113

Huggins, H. P., Subash, J. S., Stoffel, H., Henderson, M. A., Hoffman, J. L., Buckner, D. S., . . . Keiper, B. D. (2020). Distinct roles of two eIF4E isoforms in the germline of *Caenorhabditis elegans*. *Journal of Cell Science, 133*(6), jcs237990. doi:10.1242/jcs.237990

Inukai, S., Kock, K. H., & Bulyk, M. L. (2017). Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev, 43*, 110-119. doi:10.1016/j.gde.2017.02.007

Ji, S., Chen, F., Stein, P., Wang, J., Zhou, Z., Wang, L., . . . Xie, W. (2023). OBOX regulates mouse zygotic genome activation and early development. *Nature, 620*(7976), 1047-1053. doi:10.1038/s41586-023-06428-3

Jia, C., Hu, Y., Kelly, D., Kim, J., Li, M., & Zhang, N. R. (2017). Accounting for technical noise in differential expression analysis of single-cell RNA sequencing data. *Nucleic Acids Res, 45*(19), 10978-10988. doi:10.1093/nar/gkx754

Jia, M. H., Larossa, R. A., Lee, J. M., Rafalski, A., Derose, E., Gonye, G., & Xue, Z. (2000). Global expression profiling of yeast treated with an inhibitor of amino acid biosynthesis, sulfometuron methyl. *Physiol Genomics, 3*(2), 83-92. doi:10.1152/physiolgenomics.2000.3.2.83

Johnston, W. L., & Dennis, J. W. (2012). The eggshell in the C. elegans oocyte-to-embryo transition. *genesis, 50*(4), 333-349. doi:10.1002/dvg.20823

Johnston, W. L., Krizus, A., & Dennis, J. W. (2006). The eggshell is required for meiotic fidelity, polar-body extrusion and polarization of the C. elegansembryo. *BMC Biology, 4*(1), 35. doi:10.1186/1741-7007-4-35

Jona, G., Choder, M., & Gileadi, O. (2000). Glucose starvation induces a drastic reduction in the rates of both transcription and degradation of mRNA in yeast. *Biochim Biophys Acta, 1491*(1-3), 37-48. doi:10.1016/s0167-4781(00)00016-6

Juven-Gershon, T., Hsu, J.-Y., Theisen, J. W., & Kadonaga, J. T. (2008). The RNA polymerase II core promoter — the gateway to transcription. *Current Opinion in Cell Biology, 20*(3), 253-259. doi:10.1016/j.ceb.2008.03.003

Kadandale, P., Stewart-Michaelis, A., Gordon, S., Rubin, J., Klancer, R., Schweinsberg, P., . . . Singson, A. (2005). The Egg Surface LDL Receptor Repeat-Containing Proteins EGG-1 and EGG-2 Are Required for Fertilization in Caenorhabditis elegans. *Current Biology, 15*(24), 2222-2229. doi:10.1016/j.cub.2005.10.043

Kafkia, E., Andres-Pons, A., Ganter, K., Seiler, M., Smith, T. S., Andrejeva, A., . . . Patil, K. R. (2022). Operation of a TCA cycle subnetwork in the mammalian nucleus. *Science Advances, 8*(35). doi:10.1126/sciadv.abq5206

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci, 28*(11), 1947-1951. doi:10.1002/pro.3715

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2022). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. doi:10.1093/nar/gkac963

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res, 28*(1), 27-30. doi:10.1093/nar/28.1.27

Kawano, T., Zheng, H., Merz, D. C., Kohara, Y., Tamai, K. K., Nishiwaki, K., & Culotti, J. G. (2009). *C. elegans mig-6* encodes papilin isoforms that affect distinct aspects of DTC migration, and interacts genetically with *mig-17* and collagen IV. *Development, 136*(9), 1433-1442. doi:10.1242/dev.028472

Kigami, D., Minami, N., Takayama, H., & Imai, H. (2003). MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos1. *Biology of Reproduction, 68*(2), 651-654. doi:10.1095/biolreprod.102.007906

Kikuchi, T., Shibata, Y., Kim, H. S., Kubota, Y., Yoshina, S., Mitani, S., & Nishiwaki, K. (2015). The BED finger domain protein MIG-39 halts migration of distal tip cells

in Caenorhabditis elegans. *Dev Biol, 397*(2), 151-161. doi:10.1016/j.ydbio.2014.10.008

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol, 37*(8), 907-915. doi:10.1038/s41587-019-0201-4

Kim, S., Spike, C., & Greenstein, D. (2013). Control of oocyte growth and meiotic maturation in Caenorhabditis elegans. *Adv Exp Med Biol, 757:277-320.*(doi), 10.1007/1978-1001-4614-4015-1004_1010.

Klein, C. J. L., Olsson, L., & Nielsen, J. (1998). Glucose control in Saccharomyces cerevisiae: the role of Mig1 in metabolic functions. *Microbiology (Reading), 144 (Pt 1)*, 13-24. doi:10.1099/00221287-144-1-13

Kohara, Y. (2001). [Systematic analysis of gene expression of the C. elegans genome]. *Tanpakushitsu Kakusan Koso, 46*(16 Suppl), 2425-2431.

Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The technology and biology of single-cell RNA sequencing. *Mol Cell, 58*(4), 610-620. doi:10.1016/j.molcel.2015.04.005

Krueger, F. (2015). Trim galore. *A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, 516*, 517.

Labouesse, M., & Mango, S. E. (1999). Patterning the C. elegans embryo: moving beyond the cell lineage. *Trends Genet, 15*(8), 307-313. doi:S0168-9525(99)01750-3 [pii]

Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., . . . Davis, R. W. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci U S A, 94*(24), 13057-13062. doi:10.1073/pnas.94.24.13057

Latchman, D. S. (2001). Transcription factors: bound to activate or repress. *Trends Biochem Sci, 26*(4), 211-213. doi:10.1016/s0968-0004(01)01812-6

Lee, K., Cho, K., Morey, R., & Cook-Andersen, H. (2024). An extended wave of global mRNA deadenylation sets up a switch in translation regulation across the mammalian oocyte-to-embryo transition. *Cell Reports, 43*(2), 113710. doi:10.1016/j.celrep.2024.113710

Lee, K.-Y., Green, R. A., Gutierrez, E., Gomez-Cavazos, J. S., Kolotuev, I., Wang, S., . . . Oegema, K. (2018). CYK-4 functions independently of its centralspindlin partner ZEN-4 to cellularize oocytes in germline syncytia. *Elife, 7*, e36919. doi:10.7554/eLife.36919

Lee, L.-W., Lee, C.-C., Huang, C.-R., & Lo, S. J. (2012). The Nucleolus of *Caenorhabditis elegans*. *Journal of Biomedicine and Biotechnology, 2012*, 1-11. doi:10.1155/2012/601274

Lee, M. T., Bonneau, A. R., & Giraldez, A. J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol, 30*, 581-613. doi:10.1146/annurev-cellbio-100913-013027

Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., . . . Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods, 7*(9), 709-715. doi:10.1038/nmeth.1491

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*, 323. doi:10.1186/1471-2105-12-323

Li, F., Han, H., Sun, Q., Liu, K., Lin, N., Xu, C., . . . Zhao, W. (2019). USP28 regulates deubiquitination of histone H2A and cell proliferation. *Exp Cell Res, 379*(1), 11-18. doi:10.1016/j.yexcr.2019.03.026

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, J. J., Huang, H., Bickel, P. J., & Brenner, S. E. (2014). Comparison of D. melanogaster and C. elegans developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Res., 24*(7), 1086-1101. doi: 1010.1101/gr.170100.170113.

Li, L., Huang, K.-L., Gao, Y., Cui, Y., Wang, G., Elrod, N. D., . . . Li, W. (2021). An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nature Genetics, 53*(7), 994-1005. doi:10.1038/s41588-021-00864-5

Li, S., Dean, S., Li, Z., Horecka, J., Deschenes, R. J., & Fassler, J. S. (2002). The eukaryotic two-component histidine kinase Sln1p regulates OCH1 via the transcription factor, Skn7p. *Mol Biol Cell, 13*(2), 412-424. doi:10.1091/mbc.01-09-0434

Li, W., Lin, X., Peng, H., & Zhang, W. (2015). Expression and localization of Nlrp4g in mouse preimplantation embryo. *Zygote, 23*(6), 846-851. doi:10.1017/S0967199414000525

Li, Y., Ni, P., Zhang, S., Li, G., & Su, Z. (2019). ProSampler: an ultrafast and accurate motif finder in large ChIP-seq datasets for combinatory motif discovery. *Bioinformatics, 35*(22), 4632-4639. doi:10.1093/bioinformatics/btz290

Lieberman, B., Kusi, M., Hung, C. N., Chou, C. W., He, N., Ho, Y. Y., . . . Chen, C. L. (2021). Toward uncharted territory of cellular heterogeneity: advances and applications of single-cell RNA-seq. *J Transl Genet Genom, 5*, 1-21. doi:10.20517/jtgg.2020.51

Liu, Y., Lu, X., Shi, J., Yu, X., Zhang, X., Zhu, K., . . . Li, L. (2016). BTG4 is a key regulator for maternal mRNA clearance during mouse early embryogenesis. *Journal of Molecular Cell Biology, 8*(4), 366-368. doi:10.1093/jmcb/mjw023

Liu, Y., Nie, H., Zhang, C., Hou, Z., Wang, J., & Lu, F. (2021). *Poly(A) tail length is a major regulator of maternal gene expression during the mammalian oocyte-to-embryo transition*. Cold Spring Harbor Laboratory.

Liu, Y., Zhao, H., Shao, F., Zhang, Y., Nie, H., Zhang, J., . . . Lu, F. (2023). Remodeling of maternal mRNA through poly(A) tail orchestrates human oocyte-to-embryo transition. *Nature Structural & Molecular Biology, 30*(2), 200-215. doi:10.1038/s41594-022-00908-2

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology, 15*(12). doi:10.1186/s13059-014-0550-8

Love, M. I., Soneson, C., & Patro, R. (2018). Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. *F1000Research, 7*, 952. doi:10.12688/f1000research.15398.3

Ma, J., Flemr, M., Strnad, H., Svoboda, P., & Schultz, R. M. (2013). Maternally recruited DCP1A and DCP2 contribute to messenger RNA degradation during oocyte maturation and genome activation in mouse. *Biol Reprod, 88*(1), 11. doi:10.1095/biolreprod.112.105312

Ma, J., Fukuda, Y., & Schultz, R. M. (2015). Mobilization of Dormant Cnot7 mRNA Promotes Deadenylation of Maternal Transcripts During Mouse Oocyte Maturation1. *Biology of Reproduction, 93*(2), 48-48. doi:10.1095/biolreprod.115.130344

Ma, L., Tan, Z., Teng, Y., Hoersch, S., & Horvitz, H. R. (2011). In vivo effects on intron retention and exon skipping by the U2AF large subunit and SF1/BBP in the nematode *Caenorhabditis elegans*. *RNA, 17*(12), 2201-2211. doi:10.1261/rna.027458.111

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., . . . McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell, 161*(5), 1202-1214. doi:10.1016/j.cell.2015.05.002

Mangio, R. S., Votra, S., & Pruyne, D. (2015). The canonical eIF4E isoform of C. elegans regulates growth, embryogenesis, and germline sex-determination. *Biol Open, 4*(7), 843-851. doi:10.1242/bio.011585

Mangone, M., Manoharan, A. P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S. D., . . . Kim, J. K. (2010). The landscape of C. elegans 3'UTRs. *Science, 329*(5990), 432-435. doi:10.1126/science.1191244

Mangone, M., Manoharan, A. P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S. D., . . . Kim, J. K. (2010). The Landscape of C. elegans 3′UTRs. *Science, 329*(5990), 432-435. doi:10.1126/science.1191244

Marcello, M. R., Singaravelu, G., & Singson, A. (2013). Fertilization. *Adv Exp Med Biol, 757:321-50.*(doi), 10.1007/1978-1001-4614-4015-1004_1011.

Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D. N., Hanspers, K., . . . Kutmon, M. (2021). WikiPathways: connecting communities. *Nucleic Acids Res, 49*(D1), D613-D621. doi:10.1093/nar/gkaa1024

Martinez-Pastor, M. T., Marchler, G., Schuller, C., Marchler-Bauer, A., Ruis, H., & Estruch, F. (1996). The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO J, 15*(9), 2227-2235.

Maruyama, R., Velarde, N. V., Klancer, R., Gordon, S., Kadandale, P., Parry, J. M., . . . Singson, A. (2007). EGG-3 Regulates Cell-Surface and Cortex Rearrangements during Egg Activation in Caenorhabditis elegans. *Current Biology, 17*(18), 1555-1560. doi:10.1016/j.cub.2007.08.011

Mayya, V. K., & Duchaine, T. F. (2019). Ciphers and Executioners: How 3'-Untranslated Regions Determine the Fate of Messenger RNAs. *Front Genet, 10*, 6. doi:10.3389/fgene.2019.00006

Mcghee, J. (2007). The C. elegans intestine. *WormBook*. doi:10.1895/wormbook.1.133.1

McGrath, J., & Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell, 37*(1), 179-183. doi:10.1016/0092-8674(84)90313-1

Melsted, P., Booeshaghi, A. S., Gao, F., Beltrame, E., Lu, L., Hjorleifsson, K. E., . . . Pachter, L. (2019). *Modular and efficient pre-processing of single-cell RNA-seq*. Cold Spring Harbor Laboratory.

Melsted, P., Ntranos, V., & Pachter, L. (2019). The barcode, UMI, set format and BUStools. *Bioinformatics, 35*(21), 4472-4473. doi:10.1093/bioinformatics/btz279

Merritt, C., Rasoloson, D., Ko, D., & Seydoux, G. (2008). 3' UTRs are the primary regulators of gene expression in the C. elegans germline. *Curr Biol, 18*(19), 1476-1482. doi:10.1016/j.cub.2008.08.013

Metzstein, M. M., Hengartner, M. O., Tsung, N., Ellis, R. E., & Horvitz, H. R. (1996). Transcriptional regulator of programmed cell death encoded by Caenorhabditis elegans gene ces-2. *Nature, 382*(6591), 545-547. doi:10.1038/382545a0

Mignone, F., Gissi, C., Liuni, S., & Pesole, G. (2002). *Genome Biology, 3*(3), reviews0004.0001. doi:10.1186/gb-2002-3-3-reviews0004

Mikl, M., & Cowan, R., Carrie. (2014). Alternative 3′ UTR Selection Controls PAR-5 Homeostasis and Cell Polarity in C. elegans Embryos. *Cell Reports, 8*(5), 1380-1390. doi:10.1016/j.celrep.2014.08.004

Milacic, M., Beavers, D., Conley, P., Gong, C., Gillespie, M., Griss, J., . . . D'Eustachio, P. (2024). The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Research, 52*(D1), D672-D678. doi:10.1093/nar/gkad1025

Miller, M. A., Nguyen, V. Q., Lee, M. H., Kosinski, M., Schedl, T., Caprioli, R. M., & Greenstein, D. (2001). A sperm cytoskeletal protein that signals oocyte meiotic maturation and ovulation. *Science, 291*(5511), 2144-2147. doi:10.1126/science.1057586

Miluzio, A., Beugnet, A., Volta, V., & Biffo, S. (2009). Eukaryotic initiation factor 6 mediates a continuum between 60S ribosome biogenesis and translation. *EMBO reports, 10*(5), 459-465. doi:10.1038/embor.2009.70

Minami, N., Aizawa, A., Ihara, R., Miyamoto, M., Ohashi, A., & Imai, H. (2003). Oogenesin is a novel mouse protein expressed in oocytes and early cleavage-stage embryos. *Biol Reprod, 69*(5), 1736-1742. doi:10.1095/biolreprod.103.018051

Moerman, D., Fire, A., & Riddle, D. I. (1997). C. elegans II. In: Cold Spring Harbor Laboratory Press, New York.

Monteiro, P. T., Oliveira, J., Pais, P., Antunes, M., Palma, M., Cavalheiro, M., . . . Teixeira, M. C. (2020). YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. *Nucleic Acids Res, 48*(D1), D642-D649. doi:10.1093/nar/gkz859

Monti, M., & Redi, C. (2009). Oogenesis specific genes (Nobox, Oct4, Bmp15, Gdf9, Oogenesin1 and Oogenesin2) are differentially expressed during natural and gonadotropin-induced mouse follicular development. *Mol Reprod Dev, 76*(10), 994-1003. doi:10.1002/mrd.21059

Morano, K. A., Grant, C. M., & Moye-Rowley, W. S. (2012). The response to heat shock and oxidative stress in Saccharomyces cerevisiae. *Genetics, 190*(4), 1157-1195. doi:10.1534/genetics.111.128033

Morse, R. H. (2000). RAP, RAP, open up! New wrinkles for RAP1 in yeast. *Trends Genet, 16*(2), 51-53. doi:10.1016/s0168-9525(99)01936-8

Mullins, L. J., & Mullins, J. J. (2004). Insights from the rat genome sequence. *Genome Biol, 5*(5), 221. doi:10.1186/gb-2004-5-5-221

Nadal-Ribelles, M., Islam, S., Wei, W., Latorre, P., Nguyen, M., de Nadal, E., . . . Steinmetz, L. M. (2019). Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast populations. *Nat Microbiol, 4*(4), 683-692. doi:10.1038/s41564-018-0346-9

Nagaraj, R., Sharpley, M. S., Chi, F., Braas, D., Zhou, Y., Kim, R., . . . Banerjee, U. (2017). Nuclear Localization of Mitochondrial TCA Cycle Enzymes as a Critical Step in Mammalian Zygotic Genome Activation. *Cell, 168*(1-2), 210-223.e211. doi:10.1016/j.cell.2016.12.026

Nakao, A. (2004). RPG: the Ribosomal Protein Gene database. *Nucleic Acids Research, 32*(90001), 168D-170. doi:10.1093/nar/gkh004

Natarajan, K., Meyer, M. R., Jackson, B. M., Slade, D., Roberts, C., Hinnebusch, A. G., & Marton, M. J. (2001). Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. *Mol Cell Biol, 21*(13), 4347-4368. doi:10.1128/MCB.21.13.4347-4368.2001

Ni, Y., Shen, P., Wang, X., Liu, H., Luo, H., & Han, X. (2022). The roles of IDH1 in tumor metabolism and immunity. *Future Oncology, 18*(35), 3941-3953. doi:10.2217/fon-2022-0583

Nowicka, M., & Robinson, M. D. (2016). DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. *F1000Research, 5*, 1356. doi:10.12688/f1000research.8900.2

Obinata, T., Ono, K., & Ono, S. (2010). Troponin I controls ovulatory contraction of non-striated actomyosin networks in the *C. elegans* somatic gonad. *Journal of Cell Science, 123*(9), 1557-1566. doi:10.1242/jcs.065060

Olson, S. K., Bishop, J. R., Yates, J. R., Oegema, K., & Esko, J. D. (2006). Identification of novel chondroitin proteoglycans in Caenorhabditis elegans: embryonic cell division depends on CPG-1 and CPG-2. *J Cell Biol, 173*(6), 985-994. doi:10.1083/jcb.200603003

Ono, K., & Ono, S. (2004). Tropomyosin and Troponin Are Required for Ovarian Contraction in the *Caenorhabditis elegans* Reproductive System. *Molecular Biology of the Cell, 15*(6), 2782-2793. doi:10.1091/mbc.e04-03-0179

Ono, K., & Ono, S. (2016). Two distinct myosin II populations coordinate ovulatory contraction of the myoepithelial sheath in the *Caenorhabditis elegans* somatic gonad. *Molecular Biology of the Cell, 27*(7), 1131-1142. doi:10.1091/mbc.e15-09-0648

Ono, S. (2014). Regulation of Structure and Function of Sarcomeric Actin Filaments in Striated Muscle of the Nematode Caenorhabditis elegans. *The Anatomical Record, 297*(9), 1548-1559. doi:10.1002/ar.22965

Ono, S., & Pruyne, D. (2012). Biochemical and cell biological analysis of actin in the nematode Caenorhabditis elegans. *Methods, 56*(1), 11-17. doi:10.1016/j.ymeth.2011.09.008

Ortiz, M. A., Noble, D., Sorokin, E. P., & Kimble, J. (2014). A new dataset of spermatogenic vs. oogenic transcriptomes in the nematode Caenorhabditis elegans. *G3 (Bethesda), 4*(9), 1765-1772. doi:10.1534/g3.114.012351

Ou, J., & Zhu, L. J. (2019). trackViewer: a Bioconductor package for interactive and integrative visualization of multi-omics data. *Nat Methods, 16*(6), 453-454. doi:10.1038/s41592-019-0430-y

Parzych, K. R., Ariosa, A., Mari, M., & Klionsky, D. J. (2018). A newly characterized vacuolar serine carboxypeptidase, Atg42/Ybr139w, is required for normal vacuole function and the terminal steps of autophagy in the yeast Saccharomyces cerevisiae. *Mol Biol Cell, 29*(9), 1089-1099. doi:10.1091/mbc.E17-08-0516

Pasternak, M., Pfender, S., Santhanam, B., & Schuh, M. (2016). The BTG4 and CAF1 complex prevents the spontaneous activation of eggs by deadenylating maternal mRNAs. *Open Biology, 6*(9), 160184. doi:10.1098/rsob.160184

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods, 14*(4), 417-419. doi:10.1038/nmeth.4197

Pazdernik, N., & Schedl, T. (2013). Introduction to germ cell development in Caenorhabditis elegans. *Adv Exp Med Biol, 757:1-16.*(doi), 10.1007/1978-1001-4614-4015-1004_1001.

Pedraza, J. M., & van Oudenaarden, A. (2005). Noise propagation in gene networks. *Science, 307*(5717), 1965-1969. doi:10.1126/science.1109090

Picelli, S., Björklund Å, K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods, 10*(11), 1096-1098. doi:10.1038/nmeth.2639

Picelli, S., Bjorklund, A. K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods, 10*(11), 1096-1098. doi:10.1038/nmeth.2639

Picelli, S., Faridani, O. R., Bjorklund, A. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc, 9*(1), 171-181. doi:10.1038/nprot.2014.006

Piper, M. D., Hong, S. P., Ball, G. E., & Dawes, I. W. (2000). Regulation of the balance of one-carbon metabolism in Saccharomyces cerevisiae. *J Biol Chem, 275*(40), 30987-30995. doi:10.1074/jbc.M004248200

Popken, J., Graf, A., Krebs, S., Blum, H., Schmid, V. J., Strauss, A., . . . Cremer, T. (2015). Remodeling of the Nuclear Envelope and Lamina during Bovine Preimplantation Development and Its Functional Implications. *PLoS One, 10*(5), e0124619. doi:10.1371/journal.pone.0124619

Price, C. W., Fawcett, P., Ceremonie, H., Su, N., Murphy, C. K., & Youngman, P. (2001). Genome-wide analysis of the general stress response in Bacillus subtilis. *Mol Microbiol, 41*(4), 757-774. doi:10.1046/j.1365-2958.2001.02534.x

Pujol-Carrion, N., Belli, G., Herrero, E., Nogues, A., & de la Torre-Ruiz, M. A. (2006). Glutaredoxins Grx3 and Grx4 regulate nuclear localisation of Aft1 and the oxidative stress response in Saccharomyces cerevisiae. *J Cell Sci, 119*(Pt 21), 4554-4564. doi:10.1242/jcs.03229

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., & Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell developmental trajectories. *bioRxiv*.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842. doi:10.1093/bioinformatics/btq033

Rabilotta, A., Desrosiers, M., & Labbé, J.-C. (2015). CDK-1 and Two B-Type Cyclins Promote PAR-6 Stabilization during Polarization of the Early C. elegans Embryo. *PLoS One, 10*(2), e0117656. doi:10.1371/journal.pone.0117656

Ramskold, D., Luo, S., Wang, Y. C., Li, R., Deng, Q., Faridani, O. R., . . . Sandberg, R. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol, 30*(8), 777-782. doi:10.1038/nbt.2282

Reinke, V. (2002). Functional exploration of the C. elegans genome using DNA microarrays. *Nat Genet, 32 Suppl*, 541-546.

Reinke, V., Gil, I. S., Ward, S., & Kazmer, K. (2004). Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans. *Development, 131*(2), 311-323. doi:10.1242/dev.00914

Robertson, S., & Lin, R. (2013). The oocyte-to-embryo transition. *Adv Exp Med Biol, 757:351-72.*(doi), 10.1007/1978-1001-4614-4015-1004_1012.

Rodriguez, M., Snoek, L. B., De Bono, M., & Kammenga, J. E. (2013). Worms under stress: C. elegans stress response and its relevance to complex human disease and aging. *Trends Genet, 29*(6), 367-374. doi:10.1016/j.tig.2013.01.010

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci, 21*(9), 327-335.

Rogers, S., Mccloy, R. A., Parker, B. L., Gallego-Ortega, D., Law, A. M. K., Chin, V. T., . . . Burgess, A. (2018). MASTL overexpression promotes chromosome instability and metastasis in breast cancer. *Oncogene, 37*(33), 4518-4533. doi:10.1038/s41388-018-0295-z

Rolland, F., Winderickx, J., & Thevelein, J. M. (2002). Glucose-sensing and -signalling mechanisms in yeast. *FEMS Yeast Res, 2*(2), 183-201. doi:10.1111/j.1567-1364.2002.tb00084.x

Rong, Y., Zhu, Y.-Z., Yu, J.-L., Wu, Y.-W., Ji, S.-Y., Zhou, Y., . . . Sha, Q.-Q. (2022). USP16-mediated histone H2A lysine-119 deubiquitination during oocyte maturation is a prerequisite for zygotic genome activation. *Nucleic Acids Research, 50*(10), 5599-5616. doi:10.1093/nar/gkac468

Rose, L. S., & Kemphues, K. J. (1998). Early patterning of the C. elegans embryo. *Annu Rev Genet, 32*, 521-545. doi:10.1146/annurev.genet.32.1.521 [doi]

Rushforth, A. M., White, C. C., & Anderson, P. (1998). Functions of the Caenorhabditis elegans Regulatory Myosin Light Chain Genes mlc-1 and mlc-2. *Genetics, 150*(3), 1067-1077. doi:10.1093/genetics/150.3.1067

Rutter, J., Winge, D. R., & Schiffman, J. D. (2010). Succinate dehydrogenase – Assembly, regulation and role in human disease. *Mitochondrion, 10*(4), 393-401. doi:10.1016/j.mito.2010.03.001

Salem, H., Rachmin, I., Yissachar, N., Cohen, S., Amiel, A., Haffner, R., . . . Motro, B. (2010). Nek7 kinase targeting leads to early mortality, cytokinesis disturbance and polyploidy. *Oncogene, 29*(28), 4046-4057. doi:10.1038/onc.2010.162

Salin, H., Fardeau, V., Piccini, E., Lelandais, G., Tanty, V., Lemoine, S., . . . Devaux, F. (2008). Structure and properties of transcriptional networks driving selenite stress response in yeasts. *BMC Genomics, 9*, 333. doi:10.1186/1471-2164-9-333

Saunders, C. M., Larman, M. G., Parrington, J., Cox, L. J., Royse, J., Blayney, L. M., . . . Lai, F. A. (2002). PLC zeta: a sperm-specific trigger of Ca(2+) oscillations in eggs and embryo development. *Development, 129*(15), 3533-3544. Retrieved from

Schmitt, A. P., & McEntee, K. (1996). Msn2p, a zinc finger DNA-binding protein, is the transcriptional activator of the multistress response in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A, 93*(12), 5777-5782. doi:10.1073/pnas.93.12.5777

Schultz, R. M. (1993). Regulation of zygotic gene activation in the mouse. *Bioessays, 15*(8), 531-538. doi:10.1002/bies.950150806

Schultz, R. M., Letourneau, G. E., & Wassarman, P. M. (1979). Program of early development in the mammal: changes in the patterns and absolute rates of tubulin and total protein synthesis during oocyte growth in the mouse. *Dev Biol, 73*(1), 120-133. doi:10.1016/0012-1606(79)90142-8

Schulz, K. N., & Harrison, M. M. (2019). Mechanisms regulating zygotic genome activation. *Nature Reviews Genetics, 20*(4), 221-234. doi:10.1038/s41576-018-0087-x

Scully, K. M., Jacobson, E. M., Jepsen, K., Lunyak, V., Viadiu, H., Carrière, C., . . . Rosenfeld, M. G. (2000). Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science, 290*(5494), 1127-1131. doi:10.1126/science.290.5494.1127

Seo, A. Y., Lau, P.-W., Feliciano, D., Sengupta, P., Le Gros, M. A., Cinquin, B., . . . Lippincott-Schwartz, J. (2017). AMPK and vacuole-associated Atg14p orchestrate μ-lipophagy for energy production and long-term survival under glucose starvation. *Elife, 6*, e21690.

Sha, Q. Q., Yu, J. L., Guo, J. X., Dai, X. X., Jiang, J. C., Zhang, Y. L., . . . Fan, H. Y. (2018). CNOT6L couples the selective degradation of maternal transcripts to meiotic cell cycle progression in mouse oocyte. *The EMBO Journal, 37*(24), e99333.

Sha, Q. Q., Zhu, Y. Z., Li, S., Jiang, Y., Chen, L., Sun, X. H., . . . Fan, H. Y. (2020). Characterization of zygotic genome activation-dependent maternal mRNA clearance in mouse. *Nucleic Acids Res, 48*(2), 879-894. doi:10.1093/nar/gkz1111

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., . . . Regev, A. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature, 498*(7453), 236-240. doi:10.1038/nature12172

Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., . . . Regev, A. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature, 510*(7505), 363-369. doi:10.1038/nature13437

Sharma, N., Zhu, Q., Wani, G., He, J., Wang, Q.-E., & Wani, A. A. (2014). USP3 counteracts RNF168 via deubiquitinating H2A and γH2AX at lysine 13 and 15. *Cell Cycle, 13*(1), 106-114. doi:10.4161/cc.26814

Shelton, C. A., Carter, J. C., Ellis, G. C., & Bowerman, B. (1999). The Nonmuscle Myosin Regulatory Light Chain Gene *mlc-4* Is Required for Cytokinesis, Anterior-Posterior Polarity, and Body Morphology during *Caenorhabditis elegans* Embryogenesis. *The Journal of Cell Biology, 146*(2), 439-451. doi:10.1083/jcb.146.2.439

Shen, S., Park, J. W., Lu, Z.-X., Lin, L., Henry, M. D., Wu, Y. N., . . . Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences, 111*(51), E5593-E5601. doi:10.1073/pnas.1419161111

Shen, T., & Huang, S. (2012). The Role of Cdc25A in the Regulation of Cell Proliferation and Apoptosis. *Anti-Cancer Agents in Medicinal Chemistry, 12*(6), 631-639. doi:10.2174/187152012800617678

Shirra, M. K., McCartney, R. R., Zhang, C., Shokat, K. M., Schmidt, M. C., & Arndt, K. M. (2008). A chemical genomics study identifies Snf1 as a repressor of GCN4 translation. *J Biol Chem, 283*(51), 35889-35898. doi:10.1074/jbc.M805325200

Smits, G. J., Kapteyn, J. C., van den Ende, H., & Klis, F. M. (1999). Cell wall dynamics in yeast. *Curr Opin Microbiol, 2*(4), 348-352. doi:10.1016/s1369-5274(99)80061-7

Song, K.-H., Kim, J.-H., Lee, Y.-H., Bae, H. C., Lee, H.-J., Woo, S. R., . . . Kim, T. W. (2018). Mitochondrial reprogramming via ATP5H loss promotes multimodal cancer therapy resistance. *Journal of Clinical Investigation, 128*(9), 4098-4114. doi:10.1172/jci96804

Spencer, W. C., Zeller, G., Watson, J. D., Henz, S. R., Watkins, K. L., McWhirter, R. D., . . . Miller, D. M., 3rd. (2011). A spatial and temporal map of C. elegans gene expression. *Genome Res, 21*(2), 325-341. doi:gr.114595.110 [pii]
10.1101/gr.114595.110 [doi]

Spike, C. A., Coetzee, D., Eichten, C., Wang, X., Hansen, D., & Greenstein, D. (2014). The TRIM-NHL Protein LIN-41 and the OMA RNA-Binding Proteins Antagonistically Control the Prophase-to-Metaphase Transition and Growth of *Caenorhabditis elegans* Oocytes. *Genetics, 198*(4), 1535-1558. doi:10.1534/genetics.114.168831

Srivastava, A., Malik, L., Sarkar, H., Zakeri, M., Almodaresi, F., Soneson, C., . . . Patro, R. (2020). Alignment and mapping methodology influence transcript abundance estimation. *Genome Biol, 21*(1), 239. doi:10.1186/s13059-020-02151-8

Starck, J. (1977). Radioautographic study of RNA synthesis in Caenorhabditis elegans (Bergerac variety) oogenesis. *Biologie Cellulaire, 30*(2), 181-182.

Starich, T. A., Hall, D. H., & Greenstein, D. (2014). Two classes of gap junction channels mediate soma-germline interactions essential for germline proliferation and

gametogenesis in Caenorhabditis elegans. *Genetics, 198*(3), 1127-1153. doi:10.1534/genetics.114.168815

Steber, H. S., Gallante, C., O'Brien, S., Chiu, P. L., & Mangone, M. (2019). The C. elegans 3' UTRome v2 resource for studying mRNA cleavage and polyadenylation, 3'-UTR biology, and miRNA targeting. *Genome Res, 29*(12), 2104-2116. doi:10.1101/gr.254839.119

Stein, K. K. (2018). The C. elegans eggshell. *WormBook*, 1-36. doi:10.1895/wormbook.1.179.1

Stewart-Ornstein, J., Weissman, J. S., & El-Samad, H. (2012). Cellular noise regulons underlie fluctuations in Saccharomyces cerevisiae. *Mol Cell, 45*(4), 483-493. doi:10.1016/j.molcel.2011.11.035

Stiernagle, T. (2006). Maintenance of C. elegans. *WormBook*, 1-11. doi:10.1895/wormbook.1.101.1 [doi]

Stoeckius, M., Grun, D., Kirchner, M., Ayoub, S., Torti, F., Piano, F., . . . Rajewsky, N. (2014). Global characterization of the oocyte-to-embryo transition in Caenorhabditis elegans uncovers a novel mRNA clearance mechanism. *EMBO J, 33*(16), 1751-1766. doi:10.15252/embj.201488769

Stone, S., & Shaw, J. E. (1993). A Caenorhabditis elegans act-4::lacZ fusion: use as a transformation marker and analysis of tissue-specific expression. *Gene, 131*(2), 167-173. doi:10.1016/0378-1119(93)90290-j

Stryer, L. (1988). *Biochemistry, Third Edition*.

Su, Y., Xu, C., Shea, J., Destephanis, D., & Su, Z. (2023). Transcriptomic changes in single yeast cells under various stress conditions. *BMC Genomics, 24*(1). doi:10.1186/s12864-023-09184-w

Sullivan, D. K., Min, K. H. J., Hjörleifsson, K. E., Luebbert, L., Holley, G., Moses, L., . . . Pachter, L. (2023). *kallisto, bustools, and kb-python for quantifying bulk, single-cell, and single-nucleus RNA-seq*. Cold Spring Harbor Laboratory.

Sulston, J. E., & Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev Biol, 56*(1), 110-156. doi:0012-1606(77)90158-0 [pii]

Sulston, J. E., Schierenberg, E., White, J. G., & Thomson, J. N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol, 100*(1), 64-119. doi:0012-1606(83)90201-4 [pii]

Sutovsky, P., & Schatten, G. (2000). Paternal contributions to the mammalian zygote: fertilization after sperm-egg fusion. *Int Rev Cytol, 195*, 1-65. doi:10.1016/s0074-7696(08)62703-5

Sutter, B. M., Wu, X., Laxman, S., & Tu, B. P. (2013). Methionine inhibits autophagy and promotes growth by inducing the SAM-responsive methylation of PP2A. *Cell, 154*(2), 403-415. doi:10.1016/j.cell.2013.06.041

Svensson, V., Natarajan, K. N., Ly, L. H., Miragaia, R. J., Labalette, C., Macaulay, I. C., . . . Teichmann, S. A. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat Methods, 14*(4), 381-387. doi:10.1038/nmeth.4220

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development, 136*(18), 3033-3042. doi:10.1242/dev.033183

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., . . . Surani, M. A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell, 6*(5), 468-478. doi:10.1016/j.stem.2010.03.015

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V. I., . . . Surani, M. A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc, 5*(3), 516-535. doi:10.1038/nprot.2009.236

Tao, W., Deschenes, R. J., & Fassler, J. S. (1999). Intracellular glycerol levels modulate the activity of Sln1p, a Saccharomyces cerevisiae two-component regulator. *J Biol Chem, 274*(1), 360-367. doi:10.1074/jbc.274.1.360

Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A., & Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. *Genome Res, 21*(12), 2213-2223. doi:10.1101/gr.124321.111

Thomas, J. F., Valencia-Sánchez, M. I., Tamburri, S., Gloor, S. L., Rustichelli, S., Godínez-López, V., . . . Armache, K.-J. (2023). Structural basis of histone H2A lysine 119 deubiquitination by Polycomb repressive deubiquitinase BAP1/ASXL1. *Science Advances, 9*(32). doi:10.1126/sciadv.adg9832

Tintori, S. C., Osborne Nishimura, E., Golden, P., Lieb, J. D., & Goldstein, B. (2016). A Transcriptional Lineage of the Early C. elegans Embryo. *Dev Cell, 38*(4), 430-444. doi:10.1016/j.devcel.2016.07.025

Trimmer, K. A., Zhao, P., Seemann, J., Chen, S.-Y., Mondal, S., Ben-Yakar, A., & Arur, S. (2023). Spatial single-cell sequencing of meiosis I arrested oocytes indicates acquisition of maternal transcripts from the soma. *Cell Reports, 42*(6), 112544. doi:10.1016/j.celrep.2023.112544

Tsukamoto, T., Gearhart, M. D., Spike, C. A., Huelgas-Morales, G., Mews, M., Boag, P. R., . . . Greenstein, D. (2017). LIN-41 and OMA Ribonucleoprotein Complexes Mediate a Translational Repression-to-Activation Switch Controlling Oocyte Meiotic Maturation and the Oocyte-to-Embryo Transition in *Caenorhabditis elegans. Genetics, 206*(4), 2007-2039. doi:10.1534/genetics.117.203174

Tzur, Y. B., Winter, E., Gao, J., Hashimshony, T., Yanai, I., & Colaiacovo, M. P. (2018). Spatiotemporal Gene Expression Analysis of the Caenorhabditis elegans Germline Uncovers a Syncytial Expression Switch. *Genetics, 210*(2), 587-605. doi:10.1534/genetics.118.301315

Van Den Berge, K., Soneson, C., Robinson, M. D., & Clement, L. (2017). stageR: a general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biology, 18*(1). doi:10.1186/s13059-017-1277-0

Vinothkumar, K. R., Zhu, J., & Hirst, J. (2014). Architecture of mammalian respiratory complex I. *Nature, 515*(7525), 80-84. doi:10.1038/nature13686

Viscardi, M. J., & Arribere, J. A. (2022). Poly(a) selection introduces bias and undue noise in direct RNA-sequencing. *BMC Genomics, 23*(1). doi:10.1186/s12864-022-08762-8

Walhout, A. J., Reboul, J., Shtanko, O., Bertin, N., Vaglio, P., Ge, H., . . . Vidal, M. (2002). Integrating interactome, phenome, and transcriptome mapping data for the C. elegans germline. *Curr Biol, 12*(22), 1952-1958.

Walker, A. K., Boag, P. R., & Blackwell, T. K. (2007). Transcription reactivation steps stimulated by oocyte maturation in C. elegans. *Developmental Biology, 304*(1), 382-393. doi:10.1016/j.ydbio.2006.12.039

Walvekar, A. S., Srinivasan, R., Gupta, R., & Laxman, S. (2018). Methionine coordinates a hierarchically organized anabolic program enabling proliferation. *Mol Biol Cell, 29*(26), 3183-3200. doi:10.1091/mbc.E18-08-0515

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature, 431*(7010), 873-878. doi:10.1038/nature02985

Wang, J., Sang, Y., Jin, S., Wang, X., Azad, G. K., McCormick, M. A., . . . Huang, Y. (2022). Single-cell RNA-seq reveals early heterogeneity during aging in yeast. *Aging Cell, 21*(11), e13712. doi:10.1111/acel.13712

Wang, W., Wei, Z., & Li, H. (2014). A change-point model for identifying 3′UTR switching by next-generation RNA sequencing. *Bioinformatics, 30*(15), 2162-2170. doi:10.1093/bioinformatics/btu189

Wang, X., He, Y., Zhang, Q., Ren, X., & Zhang, Z. (2021). Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. *Genomics, Proteomics & Bioinformatics, 19*(2), 253-266.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet, 10*(1), 57-63. doi:10.1038/nrg2484

Watabe, E., Ono, S., & Kuroyanagi, H. (2018). Alternative splicing of the *Caenorhabditis elegans lev-11* tropomyosin gene is regulated in a tissue-specific manner. *Cytoskeleton, 75*(10), 427-436. doi:10.1002/cm.21489

West, S. M., Mecenas, D., Gutwein, M., Aristizabal-Corrales, D., Piano, F., & Gunsalus, K. C. (2018). Developmental dynamics of gene expression and alternative polyadenylation in the Caenorhabditis elegans germline. *Genome Biol, 19*(1), 8. doi:10.1186/s13059-017-1369-x

Wheaton, K., Sarkari, F., Stanly Johns, B., Davarinejad, H., Egorova, O., Kaustov, L., . . . Sheng, Y. (2017). UbE2E1/UBCH6 Is a Critical in Vivo E2 for the PRC1-catalyzed Ubiquitination of H2A at Lys-119. *Journal of Biological Chemistry, 292*(7), 2893-2902. doi:10.1074/jbc.m116.749564

White, E. A., & Glotzer, M. (2012). Centralspindlin: At the heart of cytokinesis. *Cytoskeleton, 69*(11), 882-892. doi:10.1002/cm.21065

Willis, J. H., Munro, E., Lyczak, R., & Bowerman, B. (2006). Conditional Dominant Mutations in the *Caenorhabditis elegans* gene *act-2* Identify Cytoplasmic and Muscle Roles for a Redundant Actin Isoform. *Molecular Biology of the Cell, 17*(3), 1051-1064. doi:10.1091/mbc.e05-09-0886

Wood, W. B., & Edgar, L. G. (1994). Patterning in the C. elegans embryo. *Trends Genet, 10*(2), 49-54.

Wu, B. J., Dong, F. L., Ma, X. S., Wang, X. G., Lin, F., & Liu, H. L. (2014). Localization and expression of histone H2A variants during mouse oogenesis and

preimplantation embryo development. *Genetics and Molecular Research, 13*(3), 5929-5939. doi:10.4238/2014.august.7.8

Wu, J., Zhang, N., Hayes, A., Panoutsopoulou, K., & Oliver, S. G. (2004). Global analysis of nutrient control of gene expression in Saccharomyces cerevisiae during growth and starvation. *Proc Natl Acad Sci U S A, 101*(9), 3148-3153. doi:10.1073/pnas.0308321100

Wu, Q., Li, Z., Huang, Y., Qian, D., Chen, M., Xiao, W., & Wang, B. (2017). Oxidative Stress Delays Prometaphase/Metaphase of the First Cleavage in Mouse Zygotes via the MAD2L1-Mediated Spindle Assembly Checkpoint. *Oxidative Medicine and Cellular Longevity, 2017*, 1-9. doi:10.1155/2017/2103190

Wu, X., Viveiros, M. M., Eppig, J. J., Bai, Y., Fitzpatrick, S. L., & Matzuk, M. M. (2003). Zygote arrest 1 (Zar1) is a novel maternal-effect gene critical for the oocyte-to-embryo transition. *Nature Genetics, 33*(2), 187-191. doi:10.1038/ng1079

Xia, Z., Donehower, L. A., Cooper, T. A., Neilson, J. R., Wheeler, D. A., Wagner, E. J., & Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nature Communications, 5*(1), 5274. doi:10.1038/ncomms6274

Xie, Y., & Ren, Y. (2019). Mechanisms of nuclear mRNA export: A structural perspective. *Traffic, 20*(11), 829-840. doi:10.1111/tra.12691

Xu, D., Chen, X., Kuang, Y., Hong, M., Xu, T., Wang, K., . . . Guang, S. (2023). rRNA intermediates coordinate the formation of nucleolar vacuoles in C. elegans. *Cell Reports, 42*(8), 112915. doi:10.1016/j.celrep.2023.112915

Xu, Y. F., Zhao, X., Glass, D. S., Absalan, F., Perlman, D. H., Broach, J. R., & Rabinowitz, J. D. (2012). Regulation of yeast pyruvate kinase by ultrasensitive allostery independent of phosphorylation. *Mol Cell, 48*(1), 52-62. doi:10.1016/j.molcel.2012.07.013

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C. Y., Feng, Y., . . . Fan, G. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature, 500*(7464), 593-597. doi:10.1038/nature12364

Yang, R., Wek, S. A., & Wek, R. C. (2000). Glucose limitation induces GCN4 translation by activation of Gcn2 protein kinase. *Mol Cell Biol, 20*(8), 2706-2717. doi:10.1128/MCB.20.8.2706-2717.2000

Yin, X., Gower, N. J. D., Baylis, H. A., & Strange, K. (2004). Inositol 1,4,5-Trisphosphate Signaling Regulates Rhythmic Contractile Activity of Myoepithelial Sheath Cells in *Caenorhabditis elegans*. *Molecular Biology of the Cell, 15*(8), 3938-3949. doi:10.1091/mbc.e04-03-0198

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS, 16*(5), 284-287. doi:10.1089/omi.2011.0118

Yuet, K. P., Doma, M. K., Ngo, J. T., Sweredoski, M. J., Graham, R. L., Moradian, A., . . . Tirrell, D. A. (2015). Cell-specific proteomic analysis in Caenorhabditis elegans. *Proc Natl Acad Sci U S A., 112*(9), 2705-2710. doi: 2710.1073/pnas.1421567112. Epub 1421562015 Feb 1421567117.

Zaman, S., Lippman, S. I., Zhao, X., & Broach, J. R. (2008). How Saccharomyces responds to nutrients. *Annu Rev Genet, 42*, 27-81. doi:10.1146/annurev.genet.41.110306.130206

Zernicka-Goetz, M. (1994). Activation of embryonic genes during preimplantation rat development. *Mol Reprod Dev, 38*(1), 30-35. doi:10.1002/mrd.1080380106

Zhang, W., Yu, Y., Hertwig, F., Thierry-Mieg, J., Zhang, W., Thierry-Mieg, D., . . . Fischer, M. (2015). Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol, 16*, 133. doi:10.1186/s13059-015-0694-1

Zhang, Y., Foster, J. M., Nelson, L. S., Ma, D., & Carlow, C. K. (2005). The chitin synthase genes chs-1 and chs-2 are essential for C. elegans development and responsible for chitin deposition in the eggshell and pharynx, respectively. *Dev Biol, 285*(2), 330-339. doi:10.1016/j.ydbio.2005.06.037

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One, 9*(1), e78644. doi:10.1371/journal.pone.0078644

Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., . . . Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat Commun, 8*, 14049. doi:10.1038/ncomms14049

Zhou, K., Rolls, M. M., & Hanna-Rose, W. (2013). A postmitotic function and distinct localization mechanism for centralspindlin at a stable intercellular bridge. *Developmental Biology, 376*(1), 13-22.

Zhou, Z. H., Mccarthy, D. B., O'Connor, C. M., Reed, L. J., & Stoops, J. K. (2001). The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. *Proceedings of the National Academy of Sciences, 98*(26), 14802-14807. doi:10.1073/pnas.011597698

Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics, 35*(12), 2084-2092. doi:10.1093/bioinformatics/bty895

Zimmerman, S. M., Hinkson, I. V., Elias, J. E., & Kim, S. K. (2015). Reproductive Aging Drives Protein Accumulation in the Uterus and Limits Lifespan in C. elegans. *PLoS Genet, 11*(12), e1005725. doi:10.1371/journal.pgen.1005725

APPENDIX A: SUPPLEMENTAL FIGURES FOR CHAPTER 2

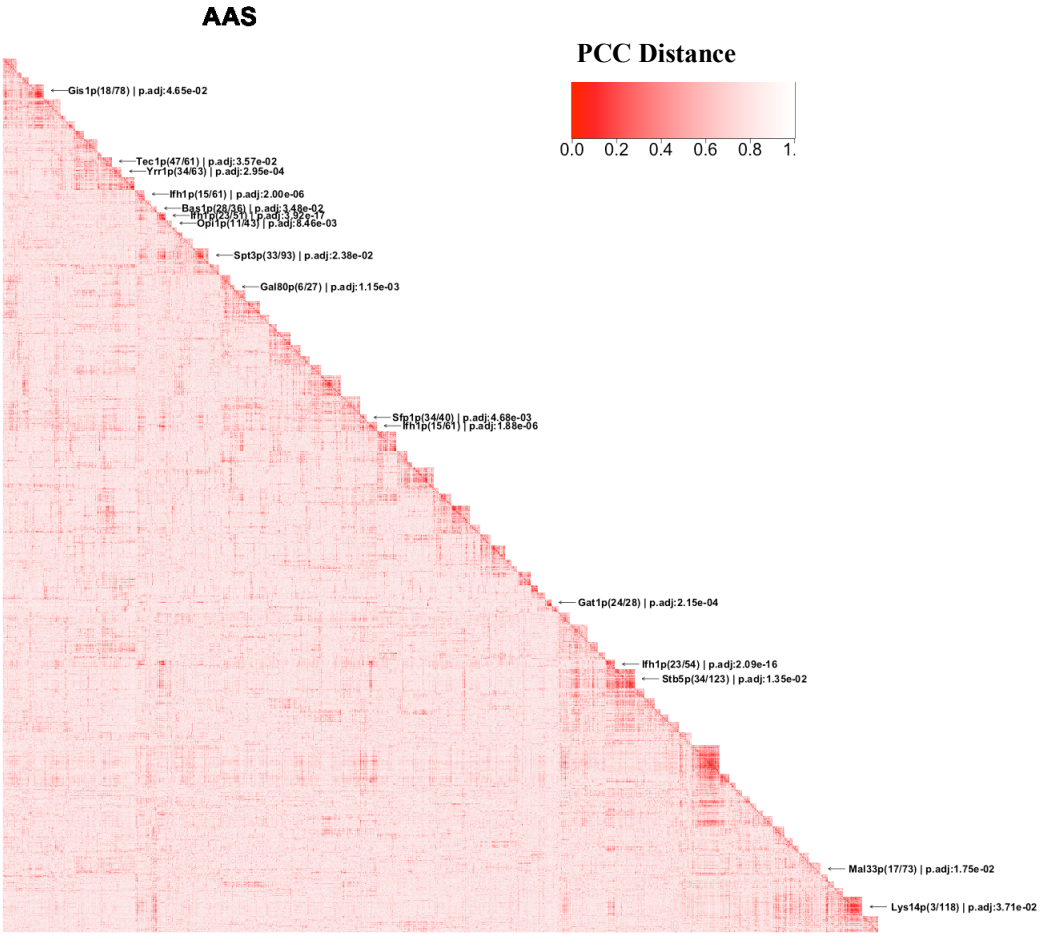


Figure A-1 Hierarchical clustering of genes under AAS condition and enrichment of transcription factors for each cluster. Name of the most enriched TF is shown next to cluster if its p.adjust < 0.05. Next to the TF name, the following is shown: (#Genes targets of TF found in cluster)/(#Genes in Yeastract Database found in cluster) | BH adjusted p-value.

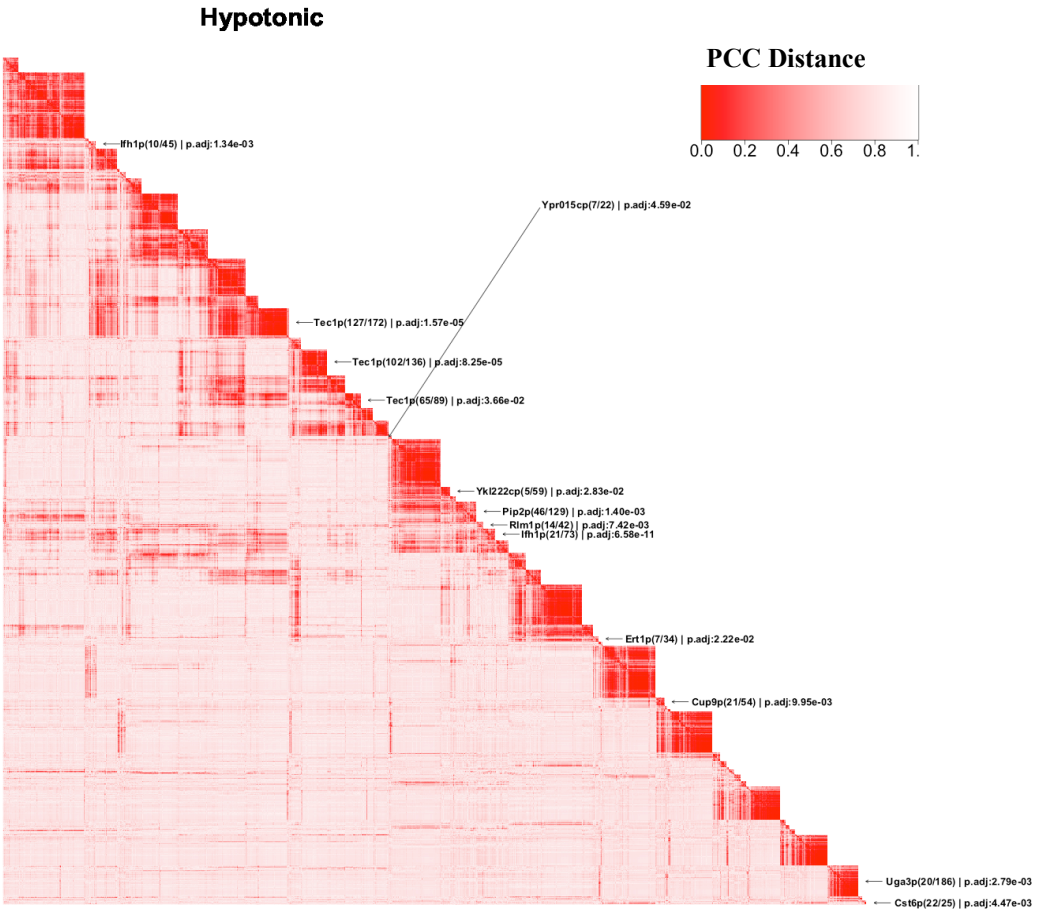


Figure A-2 Hierarchical clustering of genes under hypotonic condition and enrichment of transcription factors for each cluster. Name of the most enriched TF is shown next to cluster if its p.adjust < 0.05. Next to the TF name, the following is shown: (#Genes targets of TF found in cluster)/(#Genes in Yeastract Database found in cluster) | BH adjusted p-value.

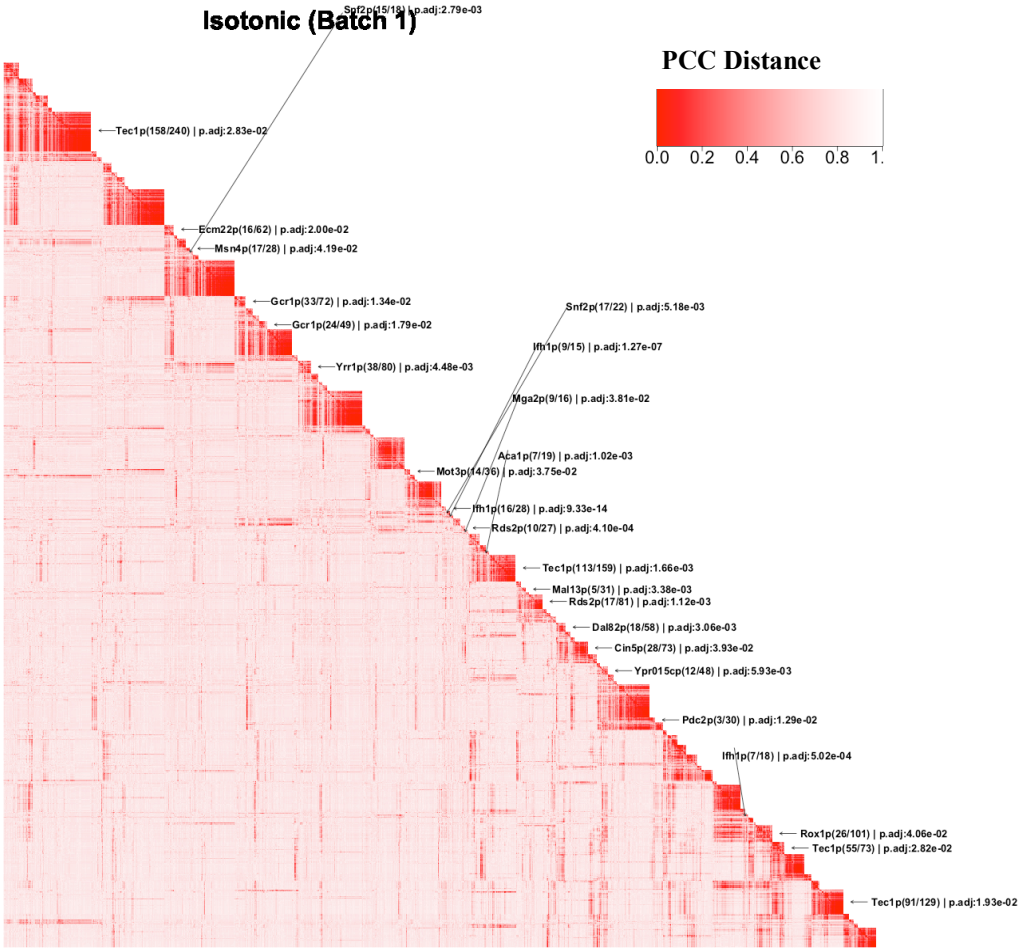


Figure A-3 Hierarchical clustering of genes under isotonic (Batch 1) condition and enrichment of transcription factors for each cluster. Name of the most enriched TF is shown next to cluster if its p.adjust < 0.05. Next to the TF name, the following is shown: (#Genes targets of TF found in cluster)/(#Genes in Yeastract Database found in cluster) | BH adjusted p-value.

APPENDIX B: SUPPLEMENTAL FIGURES FOR CHAPTER 3

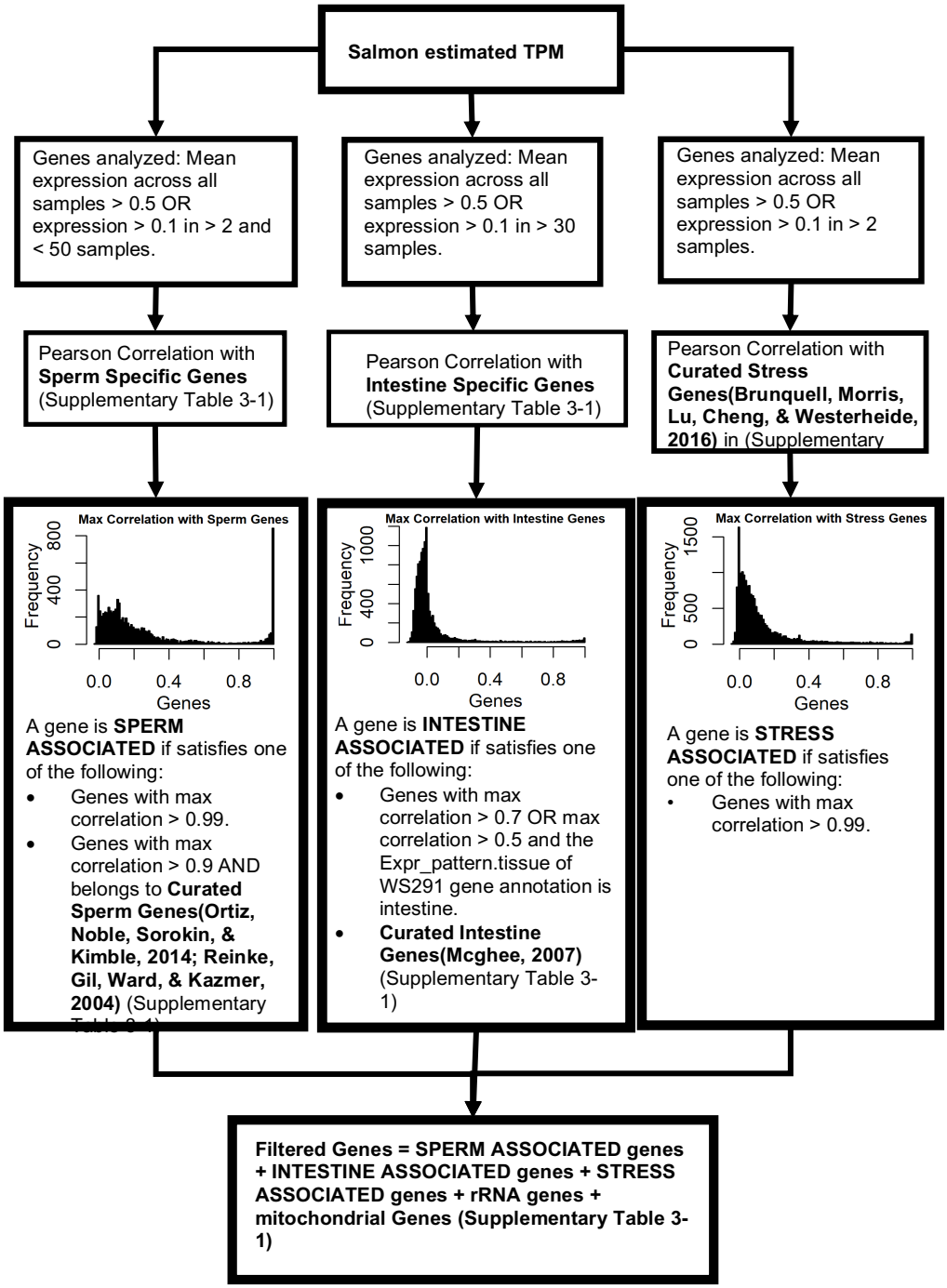


Figure B-1 Gene filtering pipeline for mitochondrial, rRNA, intestine, sperm and stress associated genes, related to Methods and Materials. WS291 gene annotations are obtained via the Wormbase SimpleMine tool(Davis et al., 2022; Harris et al., 2019).

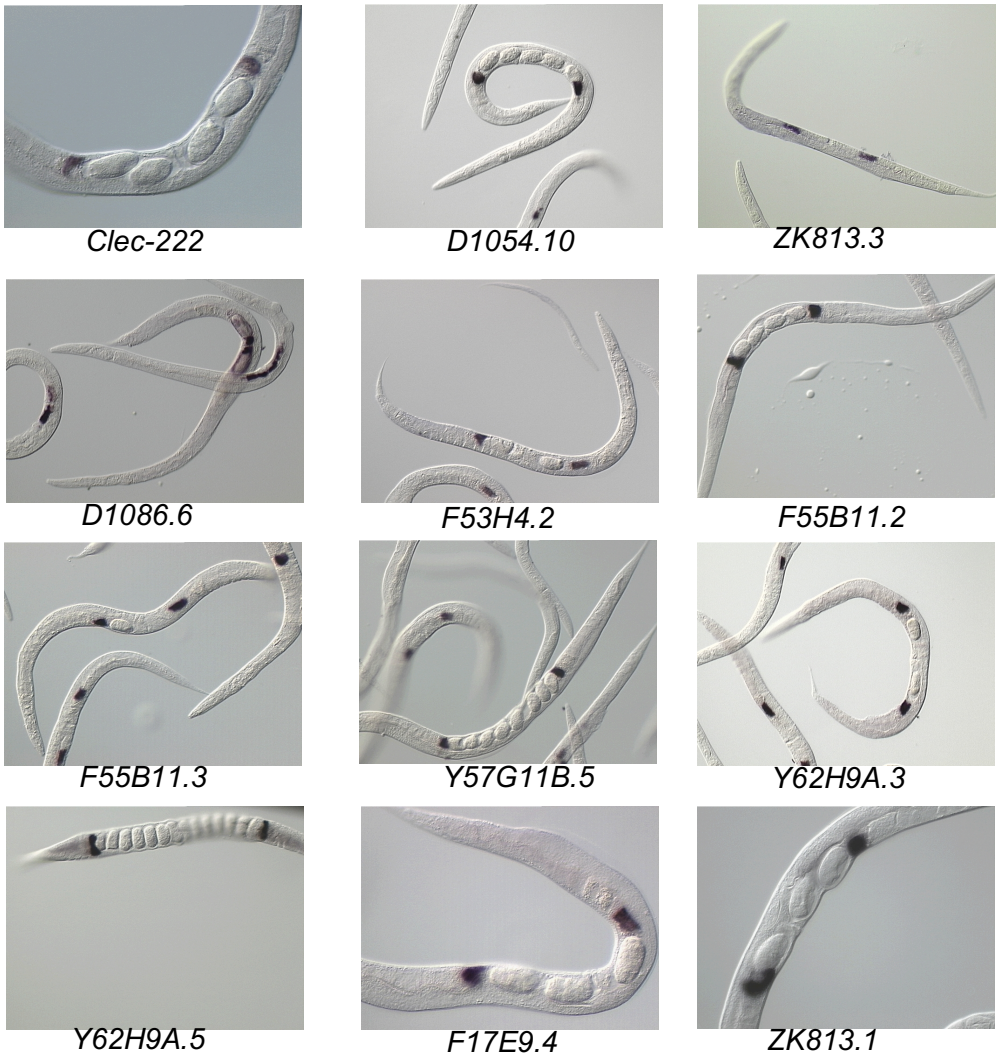


Figure B-2 A. genes encoding for components of the troponin complex. B. Genes coding for myosin heavy chain proteins (unc-54 and myo-3). C. Genes coding eggshell components perm-2/4. D. Genes coding for four myosin light chains encoding genes. E. Genes coding for four actin encoding genes.

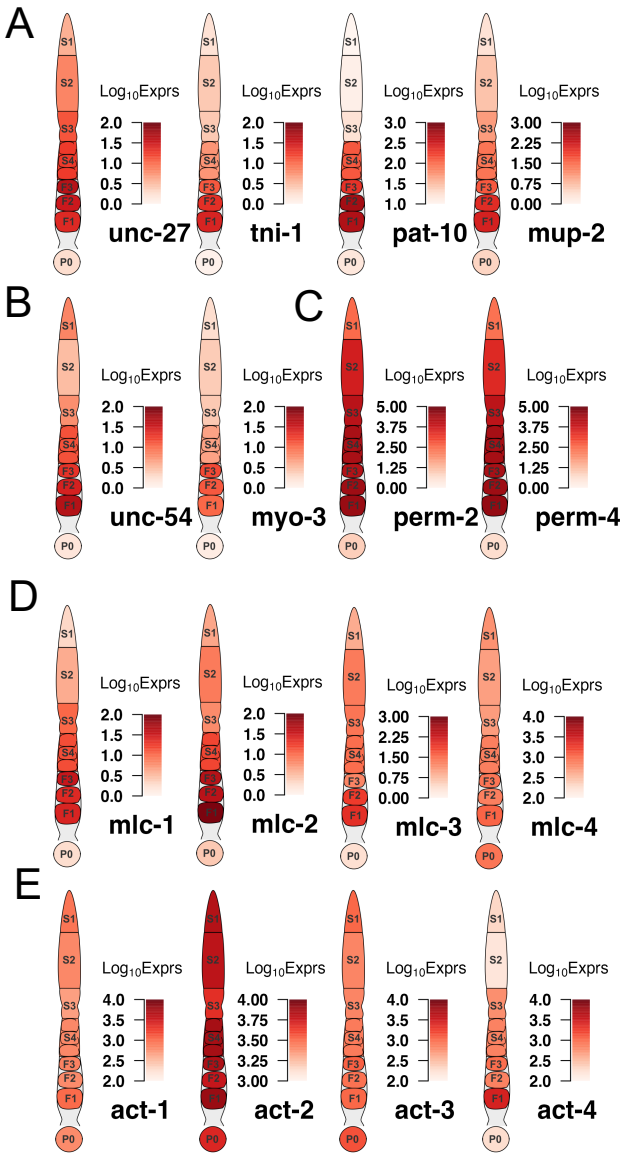


Figure B-3 NEXTDB(Kohara, 2001) in situ imaging of 13 of the 25 putative genes that originate from the spermathecae.

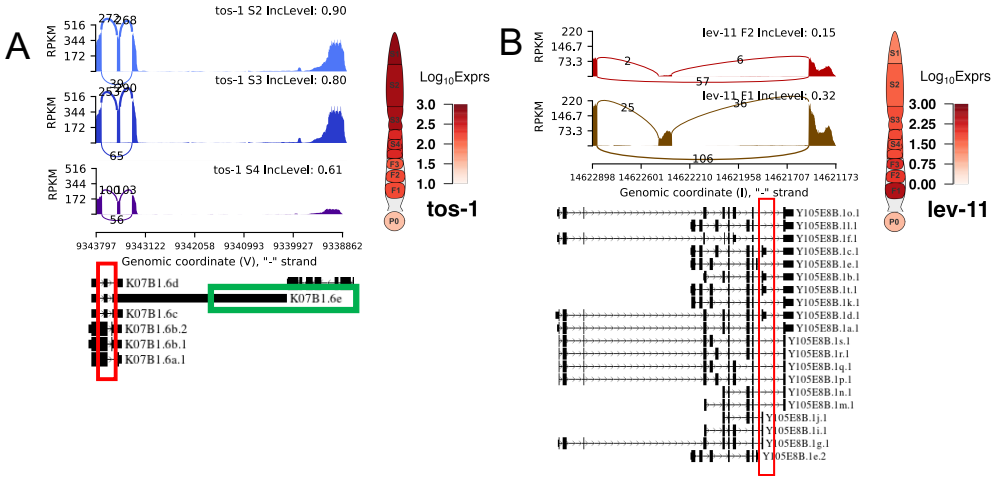


Figure B-4 A. Differential splicing events of tos-1 between S2 and S3 as well as S3 and S4 stages. B. Differential splicing event of lev-11 between F2 and F1 oocytes. Exact positions of splicing events are shown in the red box, long isoform of tos-1 is shown in green box.

APPENDIX C: SUPPLEMENTAL FIGURES FOR CHAPTER 4

Figure C-1 A. Gene body coverage plots of mouse zygotes and oocytes B. Average number of genes expressed in oocytes and zygotes of each species.

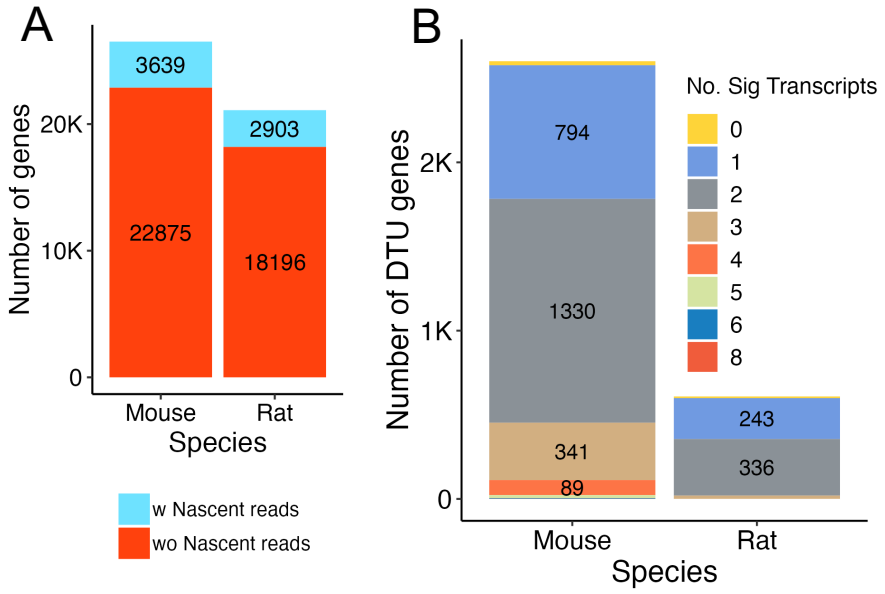


Figure C-2 A. Number of genes with and without un-spliced read coverage in oocytes and zygotes of mice and rats. B. Number of differentially expressed transcripts associated with DTU genes in each species.

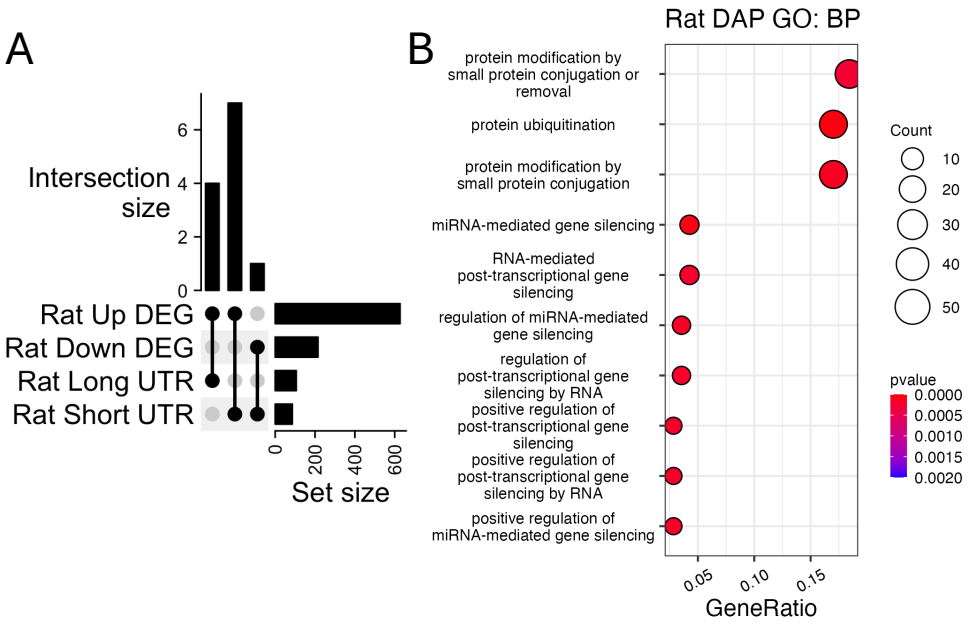


Figure C-3A. Upset plot of rat DEGs and DAP genes. B. GO BP enrichment of rat DAP genes.

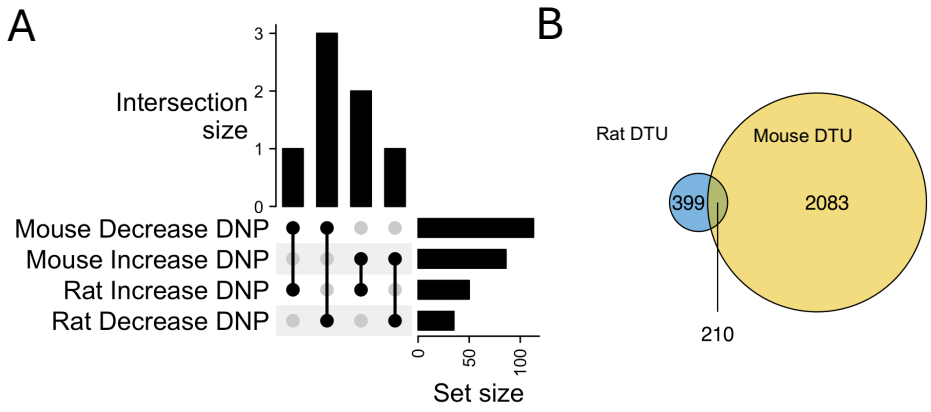


Figure C-4 A. Upset plot of orthologous genes with DNP in mice and rats. B. Ven diagram of orthologous DTU genes in mice and rats.

APPENDIX D: LINK TO SUPPLEMENTARY MATERIALS

Supplementary Tables and Materials can be found at: https://github.com/bio-info-guy/Dissertation/