Translation control tunes *Drosophila* oogenesis

Elliot T. Martin

May 2022

All dynamic biological processes require control over transcription, translation, or post-translation products. Stem cells in particular require dynamic control of gene expression. My work has focused on characterizing this control, primarily at the translation level, to better understand how stem cell differentiation occurs. Stem cells are cells with the unique ability to develop into more specialized cell types in a process called differentiation (Morrison & Spradling, 2008; Allan Spradling, Drummond-Barbosa, & Kai, 2001). Some stem cell, including those focused on in my work, also have the ability to “self-renew,” a process that allows one stem cell to copy itself giving rise to two stem cells (Cinalli, Rangan, & Lehmann, 2008; Allan Spradling, Drummond-Barbosa, & Kai, 2001). These processes must be carefully balanced as excess self-renewal will result in cells that do not give rise to differentiated cells necessary for further development or biological function (Cinalli, Rangan, & Lehmann, 2008). However, excess differentiation will result in the lack of an available pool of stem cells, preventing future differentiation and development (Cinalli, Rangan, & Lehmann, 2008).

The decision of a stem cell to either self-renew or differentiate is controlled by specific cellular pathways that can act at the level of transcription, translation, or post-translation (Blatt et al., 2020; Blatt, Martin, Breznak, & Rangan, 2020; Flora, Wong-Deyrup, et al., 2018b; Flora, McCarthy, Upadhyay, & Rangan, 2017; McCarthy et al., 2021; Sarkar et al., 2021; Seydoux & Braun, 2006). To study the regulation of these pathways in-vivo I have used the female *Drosophila* germline as a model system. The female *Drosophila* germline is contained within two pairs of ovaries. Ovaries consist of two main types of tissue, soma and germline (Roth, 2001; Schüpbach, 1987; Xie & Spradling, 2000). Each ovary is made up of strands called ovarioles. Ovarioles represent an assembly line of successive development. At the anterior tip of each ovariole a structure called a germarium is present. At the anterior of the germarium two to three stem cells are housed in a somatic niche (Eliazer & Buszczak, 2011; Roth, 2001; Schupbach & Wieschaus, 1989; Xie & Spradling, 2000). These germline stem cells (GSCs) can self-renew, or differentiate giving rise to a daughter cell called a cystoblast (CB) (D. Chen & McKearin, 2003b; D. McKearin & Ohlstein, 1995; Xie & Spradling, 1998). The CB turns on a differentiation factor called bag of marbles (*bam*) (D. Chen & McKearin, 2003a; D. McKearin & Ohlstein, 1995). This CB then undergoes four incomplete cellular divisions, resulting in interconnected cysts consisting of two, four, eight, and finally sixteen cells (D. M. McKearin & Spradling, 1990; D. McKearin & Ohlstein, 1995). One of these cells is designated as the oocyte while the rest of the cells will become nurse cells (Bastock & St Johnston, 2008; J.-R. Huynh & St Johnston, 2004; J. Huynh & St Johnston, 2000; Navarro, Lehmann, & Morris, 2001; Navarro, Puthalakath, Adams, Strasser, & Lehmann, 2004; A. C. Spradling et al., 1997; Theurkauf, Alberts, Jan, & Jongens, 1993). The sixteen cell cyst is then encapsulated by somatic cells, forming egg chambers (Bastock & St Johnston, 2008; A. J. Forbes, Lin, Ingham, & Spradling, 1996; Xie & Spradling, 2000). Egg chambers successively grow in size in fourteen stages (A. Spradling, 1993). During this time the nurse cells produce mRNAs and proteins that are transported to the oocyte (Lilly & Duronio, 2005; Royzman & Orr-Weaver, 1998). The oocyte continues to grow, while the nurse cells eventually die, dumping their contents into the oocyte (Guild, Connelly, Shaw, & Tilney, 1997; Jacob & Sirlin, 1959). Once the oocyte reaches the final, 14th stage it is known as an egg (A. Spradling, 1993).

In concert with GSC differentiation, the differentiating progeny of GSCs also transition from a mitotic cell cycle to a meiotic cell cycle, in order to eventually undergo reductional cell division to form an egg (Kimble, 2011; McKim, Jang, & Manheim, 2002). Several of the currently know factors that control this transition have been characterized as RNA binding proteins that likely facilitate the mitotic to meiotic transition by changing the translation landscape of the differentiating cysts (Blatt, Martin, Breznak, & Rangan, 2020; Carreira-Rosario et al., 2016; Flora, Wong-Deyrup, et al., 2018b; Kim-Ha, Kerr, & Macdonald, 1995; Y. Li, Minor, Park, McKearin, & Maines, 2009; Slaidina & Lehmann, 2014; Tastan, Maines, Li, Mckearin, & Buszczak, 2010). However, although some of the mRNAs these RNA binding proteins regulate have been identified, it is likely that many of the mRNA targets remain unidentified.

Each of the steps from GSC to egg require changes in cellular pathways. These changes can occur at the level of transcription, post-transciption, translation, or post-translation (Blatt, Martin, Breznak, & Rangan, 2020; Flora, McCarthy, Upadhyay, & Rangan, 2017). Decades of research has elucidated many of the changes to gene that occur during oogenesis, however, many players in this process still remain mysterious. My work has helped to identify and characterize novel developmental mechanisms that are required for the successive developmental transitions that take place during oogenesis. I have leveraged RNAseq and polysome-seq to probe the global transcription and translation landscape over development and used the power of *Drosophila* genetics in concert with these sequencing techniques to identify and characterize misregulated pathways.

A crucial participant in translation control is the ribosome, which is the molecular machine that carries out translation (Brombin, Joly, & Jamen, 2015; Gabut, Bourdelais, & Durand, 2020; Genuth & Barna, 2018). Stem cells generally have high levels of ribosomes and ribosome biogenesis components, but relatively low levels of global translation (Gabut, Bourdelais, & Durand, 2020; Sanchez et al., 2016; Woolnough, Atwood, Liu, Zhao, & Giles, 2016; Zahradkal, Larson, & Sells, 1991; Q. Zhang, Shalaby, & Buszczak, 2014). When ribosome biogenesis is perturbed stem cells can differentiate inappropriately, at least sometimes in part because specific mRNAs become misregulated. This can result in tissue specific diseases called ribosomopathies. The tissue specific nature of these diseases has long been a question of study, but recently several examples have uncovered that in general these diseases arise from misregulation of stem cell differentiation (Corsini et al., 2018; Fortier, MacRae, Bilodeau, Sargeant, & Sauvageau, 2015; Khajuria et al., 2018; Q. Zhang, Shalaby, & Buszczak, 2014).

We have discovered a link between the efficient biogenesis of the translation apparatus, the ribosome, and the translation of the proteins constituent proteins of the ribosome. We found that three RNA helicases, Aramis, Athos, and Porthos, which were previously uncharacterized in *Drosophila* are all required for pre-rRNA processing and successful ribosome biogenesis. We found that proper ribosome biogenesis ensures that ribosomal proteins are translated at normal levels by preventing a translation inhibitor called La-related protein (Larp) from binding its targets, which primarily consist of ribosomal proteins. We found that one of the mRNAs repressed by Larp is Novel nucleolar protein 1 (*Non1*), which prevents cell cycle arrest in a p53 dependent manner. Therefore we discovered a novel connection between ribosome biogenesis and cell cycle. This resolves a longstanding question of why most genes involved in ribosome biogenesis all share the same phenotype when knocked down in *Drosophila* ovaries. Our work demonstrates that this likely occurs because when when aspect of ribosome biogenesis is perturbed, translation of core ribosomal proteins are reduced to compensate for this loss in an attempt to balance ribosome biogenesis. This mechanism also results in a cell cycle arrest giving rise to a characteristic stem-like cyst where the GSC fails to divide from its progeny. More broadly, this connection has important implications in how stem cells regulate ribosome production which is known to play a crucial role in stem cell differentiation.

Additionally, I have developed a tool called Oo-cyte to allow researchers to investigate changes in gene expression at the mRNA level and post-transcriptionally over the course of GSC differentiation. Oo-site can aid the research and hypothesis generation of other researchers in the fieldby democratizing access to our labs stage specific mRNAseq and polysome-seq data, as well as integrates publicly available single-cell seq data. This tool allows non-bioinformaticians to quickly and easily view expression data across *Drosophila* GSC differentiation and development. This work has revealed that *Ord*, a key meiotic gene is controlled post-transcriptionally, at the level of translation and suggests that other key genes involved in the transition of a GSC from a mitotic to a meiotic fate may be controlled though modulating their translation.

Finally, I attempted to characterize the regulatory role the ribosome plays by performing a screen of post-translational modifying enzymes with the hypothesis that some of these enzymes might act on ribosomeal proteins. This screen made use of a dual-luciferase reporter to attempt to monitor changes in translation status. Some limitations of this work have thus far prevented any firm conclusions, however, future work in this area could help in understanding what role the ribosome might play in directly regulating translation, which is an emerging area of interest.

Overall, my work has emphasized the role the ribosome plays regulating stem cell differentiation. This regulation occurs both directly and indirectly. The ribosome regulation stem cell differentation directly in that sufficient ribosome levels are requried in order to overcome cell cycle blocks that ensure differentiation occors properly. Indirectly, the ribosome carries out translation, which my work has demonstrated is a key point of regulation during stem cell differentiation. Moving forward, discovering the factors that enact translation regulation during differentiation is of critical importance to fully understanding stem cell differentation and therefore differentation related disease states. Future work should focus on understanding what role ribosomes play in guiding translation through alternative ribosmal protein usages, post-transcriptional modification of rRNA, and post-translational modification of ribosomal proteins.

The work herein was only able to be completed thanks to the contribution of others. Foremost, my wife Allison Martin, without whom I would have given up countless times along the way to my PhD. She has been a sounding board, a life-coach, and my best friend for the years this work has taken.

Secondly, my family including, Levi, who from childhood supported my curiosity and enabled me to pursue my interests and passions. Knowing that I have always had them to fall back on provided a cushion that has helped me from struggling in undergrad to the completion of my PhD.

For direction, motivation, and guidance, I thank my mentors Dr. Prash Rangan and Dr. Gaby Fuchs. They agreed to mentor a disorganized student with less than stellar academics. Since that point they have helped me not only in developing a successful project, but also in maturing as an academic, a bench scientist, and generally, into adulthood.

A thank you to my labmates who were always there to talk me through a failed experiment or get excited about an interesting result.

To my collaborators, Elaine Nguyen, Roni Lahr, Dr. Andrea Berman, Dr. Shamsi Emtenani, and Dr. Daria Siekhaus, that contributed to this work I thank you for your expertise and beautiful results.

Finally, to my committee members, Dr. Thomas Begley, Dr. Paolo Forni, and Dr. Joesph Wade for their guidance and advice throughout my graduate studies.

1. **Chapter 1: “Post-transcriptional gene regulation mediates critical cell fate transitions during Drosophila oogenesis”** was published as Blatt P, Martin ET, Breznak SM, Rangan P. 2020. Post-transcriptional gene regulation regulates germline stem cell to oocyte transition during Drosophila oogenesis. Current Topics in Cell Biology 140: 3–34.
2. Except for minimal re-organization of the figures, the entire **Chapter 2: “A translation control module coordinates germline stem cell differentiation with ribosome biogenesis during *Drosophila* oogenesis”** was published as Martin, E.T., Blatt, P., Nguyen, E., Lahr, R., Selvam, S., Yoon, H.A.M., Pocchiari, T., Emtenani, S., Siekhaus, D., Berman, A.J., Fuchs, G., and Rangan, P. 2021. A translation control module coordinates germline stem cell differentiation with ribosome biogenesis during Drosophila oogenesis. bioRxiv.
3. **Chapter 3: “Oo-site: A dashboard to visualize gene expression during *Drosophila* oogenesis reveals meiotic entry is regulated post-transcriptionally”** was published as as Martin, Elliot Todd, Kahini Sarkar, Alicia McCarthy, and Prashanth Rangan. 2022. Oo-Site: A Dashboard to Visualize Gene Expression during Drosophila Oogenesis Reveals Meiotic Entry Is Regulated Post-Transcriptionally. bioRxiv.

These studies or articles are being included because they were part of the programmatic line of research that comprised the dissertation and that including them provides a coherent and appropriately sequenced investigation.

# 3 Oo-site: A dashboard to visualize gene expression during *Drosophila* oogenesis reveals meiotic entry is regulated post-transcriptionally

Elliot T. Martin1\*, Kahini Sarkar1, 4, Alicia McCarthy1, and Prashanth Rangan1,4\*

1Department of Biological Sciences/RNA Institute, University at Albany SUNY, Albany, NY 12202

4Black Family Stem Cell Institute, Department of Cell, Developmental, and Regenerative Biology, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, NY 10029, USA

\*Co-corresponding authors

Email: [etmartin@albany.edu](mailto:etmartin@albany.edu) and [prashanth.rangan@mssm.edu](mailto:prashanth.rangan@mssm.edu)

## 3.1 Summary

Determining how stem cell differentiation is controlled has important implications for understanding the etiology of degenerative disease and designing regenerative therapies. *In vivo* analyses of stem cell model systems have revealed regulatory paradigms for stem cell self-renewal and differentiation. The germarium of the female *Drosophila* gonad, which houses both germline and somatic stem cells, is one such model system. Bulk mRNA sequencing (RNA-seq), single-cell (sc) RNA-seq, and bulk translation efficiency of mRNAs are available for stem cells and their differentiating progeny within the *Drosophila* germarium. However, visualizing those data is hampered by the lack of a tool to spatially map gene expression and translational data in the germarium. Here, we have developed Oo-site (<https://www.ranganlab.com/Oo-site>), a tool for visualizing bulk RNA-seq, scRNA-seq, and translational efficiency data during different stages of germline differentiation, that makes these data accessible to non-bioinformaticians. Using this tool, we recapitulated previously reported expression patterns of developmentally regulated genes and discovered that meiotic genes, such as those that regulate the synaptonemal complex, are regulated at the level of translation.

## 3.2 Introduction

The *Drosophila* ovary provides a powerful system to study stem cell differentiation *in vivo* (Bastock & St Johnston, 2008; Eliazer & Buszczak, 2011; Lehmann, 2012; Allan Spradling, Fuller, Braun, & Yoshida, 2011). The *Drosophila* ovary consists of two main cell lineages, the germline, which ultimately gives rise to eggs, and the soma, which surrounds the germline and plays a supportive role in egg development (Eliazer & Buszczak, 2011; Roth, 2001; Schüpbach, 1987; Xie & Spradling, 2000). Each stage of *Drosophila* female germline stem cell (GSC) differentiation is observable and identifiable, allowing GSC development to be easily studied (Bastock & St Johnston, 2008; Lehmann, 2012; Xie & Spradling, 1998). Specifically, female *Drosophila* GSCs undergo an asymmetric division, giving rise to another GSC and a cystoblast (CB) (**Figure 3.1A**) (D. Chen & McKearin, 2003b; D. McKearin & Ohlstein, 1995; Xie & Spradling, 1998). The GSC and CB are marked by a round structure called the spectrosome (**Figure 3.1A**) (De Cuevas & Spradling, 1998; Zaccai & Lipshitz, 1996). The CB then undergoes four incomplete divisions resulting in 2-, 4-, 8-, and finally 16-cell cysts (CC), which are marked by an extended structure called the fusome (**Figure 3.1A**) (D. Chen & McKearin, 2003a; D. Chen & McKearin, 2003b; De Cuevas & Spradling, 1998). In the 16-CC, one of the cyst cells is specified as the oocyte, while the other 15 cells become nurse cells, which provide proteins and mRNAs to support the development of the oocyte (**Figure 3.1A**) (Bastock & St Johnston, 2008; Carpenter, 1975; J.-R. Huynh & St Johnston, 2004; J. Huynh & St Johnston, 2000; Navarro, Lehmann, & Morris, 2001; Theurkauf, Alberts, Jan, & Jongens, 1993). The 16-CC is encapsulated by somatic cells and buds off from the germarium, forming an egg chamber (**Figure 3.1A**) (Bastock & St Johnston, 2008; A. J. Forbes, Lin, Ingham, & Spradling, 1996; Xie & Spradling, 2000). In each chamber, the oocyte grows as the nurse cells synthesize and then deposit mRNAs and proteins into the oocyte, which eventually gives rise to a mature egg (Bastock & St Johnston, 2008; J. Huynh & St Johnston, 2000).

Expression of differentiation factors, including those that regulate translation, results in progressive differentiation of GSCs to an oocyte (Blatt, Martin, Breznak, & Rangan, 2020; Slaidina & Lehmann, 2014). In the CB, Bag-of-marbles (Bam) expression promotes differentiation and the transition from CB to 8-CC stage (D. Chen & McKearin, 2003a; D. McKearin & Ohlstein, 1995; Ohlstein & McKearin, 1997). In the 8-CC, RNA-binding Fox protein 1 (Rbfox1) promotes exit from the mitotic cell cycle into meiosis (Carreira-Rosario et al., 2016). Both the differentiation factors Bam and Rbfox1 affect the translation of mRNAs to promote differentiation (Carreira-Rosario et al., 2016; Y. Li, Minor, Park, McKearin, & Maines, 2009; Tastan, Maines, Li, Mckearin, & Buszczak, 2010). In addition, in 8-CCs, recombination is initiated across many cyst cells and then eventually is restricted to the specified oocyte (Hinnant, Merkle, & Ables, 2020; J. Huynh & St Johnston, 2000). Neither the mRNAs that are translationally regulated during this progressive differentiation nor how recombination is temporally regulated is fully understood (Cahoon & Hawley, 2016; Carreira-Rosario et al., 2016; Flora, Wong-Deyrup, et al., 2018a; Rubin, Macaisne, & Huynh, 2020; Slaidina & Lehmann, 2014; Tanneti, Landy, Joyce, & McKim, 2011; Youheng Wei et al., 2014).

Within the germarium, the germline is surrounded by and relies on distinct populations of somatic cells for signaling, structure, and organization (Roth, 2001; Schüpbach, 1987; Xie & Spradling, 1998, 2000). For example, the terminal filament, cap, and anterior-escort cells act as a somatic niche for the GSCs (Decotto & Spradling, 2005; Haifan Lin & Spradling, 1993; X. Wang & Page-McCaw, 2018; Xie & Spradling, 2000). Once GSCs divide to give rise to CBs, posterior escort cells guide CB differentiation by encapsulating the CB and the early-cyst stages (Kirilly, Wang, & Xie, 2011; Shi et al., 2021; Upadhyay et al., 2016). Follicle stem cells (FSCs), which are present towards the posterior of the germarium, divide and differentiate to give rise to follicle cells, (FCs) which surround the late-stage cysts that give rise to egg chambers (Margolis & Spradling, 1995; Nystul & Spradling, 2010; Rust et al., 2020). FSCs also give rise to stalk cells and polar cells which connect the individual egg chambers that comprise the ovariole (Margolis & Spradling, 1995; Nystul & Spradling, 2010; Rust et al., 2020; Sahai-Hernandez, Castanieto, & Nystul, 2012).

While there is a wealth of bulk RNA-seq, single-cell mRNA-seq (scRNA-seq), and translational efficiency data from polysome-seq experiments for the cells in the germarium, there are several hurdles for easy utilization of this data:

1. scRNA-seq has exquisite temporal resolution but it can miss some lowly expressed transcripts which are better captured by bulk RNA-seq (Lähnemann et al., 2020). However, there is no easy way to compare these two data sets.
2. While scRNA-seq provides mRNA levels, it does not indicate if these mRNAs are translated, especially in the germline where translation control plays an important role (Blatt, Martin, Breznak, & Rangan, 2020; Slaidina & Lehmann, 2014).
3. Lastly, there is a barrier to the visualization of the data for those who are not experienced in bioinformatics.

Here, we have developed a tool that we call Oo-site which integrates bulk RNA-seq, scRNA-seq, and polysome-seq data to spatially visualize gene expression and translational efficiency in the germarium.

## 3.3 Results

To make bulk RNA-, scRNA-, and polysome-, seq data accessible to the community, we have collated and reprocessed previously published sequencing datasets of ovaries enriched for GSCs, CBs, cysts, and egg chambers (**Figure 3.1B**). Notably, each genetically enriched sample had matched bulk RNA-seq and polysome-seq libraries prepared, allowing for simultaneous read-out of mRNA level and translation status (**Figure 3.2A**). One limitation is that the enriched cyst stages do not resolve each distinct stage of cyst development, instead, these samples represent a mixture of cyst stages. Therefore to supplement the enrichment data, we have integrated scRNA-seq data from Slaidina *et al*. which provides a more discrete temporal resolution of the cyst stages (Slaidina, Gupta, Banisch, & Lehmann, 2021). We present these data as a tool called Oo-site (<https://www.ranganlab.com/Oo-site>), a collection of interactive visualizations that allows researchers to easily input a gene or collection of genes of interest to determine their expression pattern(s).

Figure 3.1: **Oo-site integrates and provides an interface for interacting with multi-omic data covering major stages of *Drosophila* GSC differentiation.**

(**A**) Schematic illustrating developmental stages of germline development. (**B**) Summary of the samples used for bulk RNA-seq and polysome-seq and the cell types these samples are enriched for. (**C**) Screenshot of Oo-site dashboard, indicating: (1) “Take a Tour!” function, which guides the user through the functionality and operation of Oo-site. (2) The available seq datasets which the user can view, including RNA-seq of ovaries genetically enriched for developmental stages (bulk RNA-seq), polysome-seq of ovaries genetically enriched for developmental stages (Polysome-seq), single-cell seq of germline stages (Single-Cell seq: Germline), and single-cell seq of somatic stages in the germarium (Single-Cell seq: Soma). (3) the available visualizations which the user can use, including viewing the expression of genes over development at the level of a single gene (Developmental Progression), viewing all significantly changing genes as heatmaps (Heatmap), and viewing groups of genes either derived from GO-term categories or supplied by the user (Gene Groups). (4) The control panel, which the user can use to control the current visualization, and (5) the Generate Report Function, which can be used to download a PDF report of either the current visualization or all active visualizations.

Oo-site consists of three modules: ovary-map, ovary-heatmap, and ovary-violin (**Figure 3.1C**). Each module allows users to visualize expression from matched mRNA-seq and polysome-seq data of genetically enriched stages of early GSC differentiation as well as previously published scRNA-seq data (Slaidina, Gupta, Banisch, & Lehmann, 2021). Additionally, we have integrated scRNA-seq expression data for genes that cluster in somatic cell populations that reside in the germarium (Slaidina, Gupta, Banisch, & Lehmann, 2021), however, here we focus on the germline (Slaidina, Gupta, Banisch, & Lehmann, 2021). Ovary-map allows users to visualize the expression of a single gene over the course of differentiation in the framework of a germarium schematic, which contextualizes staging for those less familiar with *Drosophila* oogenesis. Ovary-heatmap consists of a clustered, interactive heatmap of genes determined to be differentially expressed that allows users to explore expression trends over-development (**Figure 3.1B, Figure 3.2B-C'**). Finally, ovary-violin allows users to visualize the expression of multiple genes over the course of differentiation (**Figure 3.1C**). These groups of genes can be selected either by a GO-term of interest or a custom list of genes supplied by the user. The user can download a spreadsheet of gene expressions corresponding to the subset of selected or input genes. Finally, Oo-site incorporates a reporting tool that generates a downloadable report of the visualization(s) in a standardized format to facilitate their use for publication (**Figure 3.1C**). Researchers can use these datasets to enhance hypothesis generation or to confirm expression patterns observed from other methods.

Figure 3.2: **Sequencing strategy and clustered heatmaps of differential expression, related to Figure 3.1.**

(**A**) Schematic of strategy used to obtain input mRNA samples and matched polysome-seq libraries of ovaries genetically enriched for developmental milestones. (**B-B'**) Clustered heatmaps of (B) bulk RNA-seq and (B') log2(TE) from bulk polysome-seq of the developmental milestones indicated on the X-axis. Each row in the heatmap indicates a gene that is differentially expressed in at least one of the milestones compared to all others in a pairwise fashion. Color scale denotes average relative expression. (**C**) scRNA-seq of early germline cells and (**C'**) scRNA-seq of somatic cells in the germarium. X-axis denotes cell-type and each row in the heatmap indicates a gene that is differentially expressed in at least one of the cell-types compared to all others in a pairwise fashion.

Using Oo-site, we first determined if the bulk RNA-seq data that was acquired by enriching for specific stages of germline development is representative of the gene expression patterns from purified cell types. We compared bulk RNA-seq data obtained by enriching for GSC and CB cell types without purification from somatic cells (**Figure 3.1C**) to the GSC and CB data from Wilcockson *et al.* where they included a fluorescent-assisted cell sorting (FACS) step to eliminate somatic cells so that a pure population of these germline cells was sequenced (Wilcockson & Ashe, 2019). We analyzed the expression of genes that Wilcockson *et al.* identified as 2-fold or more down- or upregulated with a p-value < 0.01. We found that in the enriched bulk RNA-seq data these genes followed similar trends as identified by Wilcockson *et al.*, indicating that despite the lack of FACS purification, enrichment of cell types reproduces meaningful mRNA expression changes over these stages (**Figure 3.4A-A'**).

Figure 3.3: **Oo-site allows for visualization of dynamically regulated genes.**

(**A-B**) Visualization of expression of *RpS19b* over germline development from (A) developmentally enriched stages and (B) single-cell seq data indicate that the mRNA level of *RpS19b* decreases starting in the cysts and is dramatically decreased in early egg chambers. Color indicates relative expression and values indicate the (A) mean TPM±standard error or (B) the normalized expression of *RpS19b* in each given stage. (**C-C''**) Confocal images of ovaries with *in situ* hybridization of *RpS19b* (green, middle greyscale) and stained for DAPI (blue, right greyscale) demonstrate that the mRNA level of *RpS19b* decreases starting in the cyst stages and are dramatically lower in early egg chambers consistent with the seq data. (**D-D''**) Confocal images of ovaries expressing RpS19b::GFP, visualizing (D') GFP (green, middle greyscale), (D'') Vasa staining (blue, right greyscale), and 1B1 (red) demonstrate that the protein expression of RpS19b::GFP is consistent with its mRNA levels. (**E-E'**) Quantifications of normalized mean intensity of staining, X-axis represents the distance in microns from the niche, Y-axis represents mean intensity normalized to the maximum mean intensity per germarium of (E) *RpS19b* mRNA or (E') RpS19b::GFP. The line represents fit using a loess regression, shaded area represents the standard error of the fit. (n=5 germaria).

To determine if the bulk RNA-seq data recapitulates genuine changes in gene expression, we compared the expression of *ribosomal small subunit protein 19b* (*RpS19b*) in bulk RNA-seq to scRNA-seq data. Our bulk RNA-seq data, as well as the available scRNA-seq data indicated that *RpS19b* was highly expressed in GSCs, decreased during differentiation in the cyst stages and was greatly decreased in expression in early egg chambers, consistent with previous reports (**Figure 3.3A-B**) (McCarthy et al., 2021; Sarkar et al., 2021). To further validate this expression pattern, we probed the expression of *RpS19b* *in vivo* using *in situ* hybridization as well as an RpS19b::GFP line that is under endogenous control elements (McCarthy et al., 2021). We found that *RpS19b* was present in the GSCs and diminishes in the cyst stages both at the mRNA and protein level (**Figure 3.3C-E'**). Additionally, RpS19b::GFP expression resembled its mRNA expression indicating that its dynamic expression is achieved primarily through modulating the mRNA level of *RpS19b*, consistent with its moderate to high translational efficiency in early stages (**Figure 3.3C-D, Figure 3.4B**). Thus, enriching for specific germline stages captures changes to gene expression in the germline. However, we note that care should be taken in interpreting bulk RNA-seq results as the data may be influenced by the somatic cells present in the samples. However, simultaneous comparison with scRNA-seq can alleviate this problem.

Figure 3.4: **Bulk RNA-seq recapitulates previously observed expression patterns of gene expression, related to Figure 3.3.**

(**A-A'**) Violin plots of expression from bulk RNA-seq of genes 2-fold or more (A) down or (A') upregulated in bam RNAi germline cells compared to UAS-TKV overexpressing germline cell with a p-value < 0.01 over germline development from Wilcockson *et al.* demonstrate that bulk RNA-seq identifies similar trends in gene expression compared to the FACS based method employed by Wilcockson *et al*. Values above plots represent Holm-Bonnferroni adjusted p-values from a Welch's t-test between the indicated genotypes. (**B**) Visualization of expression of *RpS19b* over germline development from polysome-seq data. Color indicates TE and values indicate the log2 mean TE±standard error *RpS19b* TE is relatively consistent during early oogenesis and decreases in the egg chambers.

To determine the groups of genes that change as the GSCs differentiate into an egg, we used gene ontology (GO)-term analysis to probe for pathways that change at the level of RNA using bulk RNA-seq data. We did not identify any significant GO-terms in genes that are differentially expressed between GSCs and CBs. We found that genes with lower expression in GSCs compared to differentiating cysts are enriched in the GO-term polytene chromosome puffing which is consistent with GO-terms identified in Wilcockson *et al*. for genes that are expressed at lower levels in GSCs than in differentiating cysts than GSCs (**Figure 3.5A**). We also identified the polytene chromosome puffing GO-term in genes downregulated in CBs compared to cysts. Additionally, we observed that several GO-terms involving peptidase activity were enriched in genes upregulated in GSCs and CBs compared to cysts (**Figure 3.5B**). This is consistent with findings suggesting that peptidases can be actively regulated during differentiation and can influence stem cell fate (Han, Wang, Bachovchin, Zukowska, & Osborn, 2015; Perišić Nanut, Pečar Fonović, Jakoš, & Kos, 2021; Tiaden et al., 2012). We found that two GO-terms related to glutathione transferase activity were enriched in genes downregulated in GSCs and CBs compared to ovaries from young-wildtype (young-WT) flies and in CBs compared to differentiating cysts, suggesting that metabolic processes may be altered during GSC differentiation. Additionally, comparison of CBs and differentiating cysts to young-WT, which contain early egg chambers, indicated that downregulated genes were enriched in GO-terms involving vitelline and eggshell coat proteins (**Figure 3.5A**).

Figure 3.5: **GO-terms enriched from differentially expressed genes between genetically enriched developmental milestones.**

(**A-B**) Heatmaps of top five significant GO-terms by fold enrichment resulting from each pairwise comparison of significantly (A) downregulated or (B) upregulated genes in the first genotype listed relative to the second genotype listed in the x-axis from bulk RNA-seq of each developmentally enriched stage. Comparisons that did not generate any significant GO-terms are omitted.

Next, to determine if our data could resolve large-scale expression changes that occur during oogenesis we examined the expression of genes in the GO-term meiotic cell cycle. Meiosis is initiated during the cyst stages of differentiation and therefore we would expect genes in the category, in general, to increase in expression in the >*bam* RNAi; hs-*bam* samples (Carpenter, 1979; Tanneti, Landy, Joyce, & McKim, 2011). We were surprised to find no significant change in the mean mRNA expression of genes in this GO-term in any of our enriched stages compared to enriched GSCs, though this does not preclude gene expression changes for individual genes (**Figure 3.7A**). However, this is consistent with the observation that several factors that promote meiosis I are transcribed in the GSCs and the cells that follow (McCarthy et al., 2021). This suggests that, in general, a transition from a mitotic state to a meiotic state is not driven by large changes in mRNA levels of meiotic genes.

As we did not see overall changes to mRNA levels of genes in the GO-term meiotic cell cycle, we next examined the polysome-seq data of those genes to determine if changes in expression might occur at the level of translation. Polysome-seq uses polysome profiling to separate mRNAs that are associated with polysomes which form by mRNAs engagement with multiple ribosomes. To quantify the degree to which an mRNA is associated with polysome fractions, we compared the relative abundance of mRNAs from the polysome fractions to their relative expression using corresponding input lysates to calculate a metric referred to as translational efficiency (TE). Indeed, genes in the meiotic cell cycle GO-term had a significant increase in translation efficiency in CBs and a more dramatic increase in cysts despite no significant changes to the overall mRNA level of these genes (**Figure 3.7A-B**). Based on scRNA-seq data, the expression of meiotic cell cycle genes increased slightly but significantly in the 4-CC cluster with a median increase in expression of 1.25 fold (**Figure 3.7C**). This suggests that some genes in the meiotic cell cycle GO-term may be regulated at the mRNA level, but as a group this regulation is modest. This is likely because genes in this GO-term are robustly expressed even in GSCs as the median mRNA level of meiotic cell cycle genes in enriched GSCs is 36.1 TPM, which exceeds the 70th expression percentile among all genes in enriched GSCs.

Figure 3.6: **Ord expression is controlled post-transcriptionally.**

(**A-B**) Visualization of expression of *ord* over germline development from (A) bulk RNA-seq of developmentally enriched stages and (B) polysome-seq of developmentally enriched stages indicates that the mRNA level of *ord* is consistent from GSCs to cysts, until decreasing in early egg chambers, but the translation efficiency of *ord* increases during the cyst stages compared to other stages. Color indicates (A) relative expression or (B) TE and values indicate the (A) mean TPM±standard error or (B) the log2 mean TE±standard error (**C-C''**) Confocal images of ovaries expressing Ord::GFP with *in situ* hybridization of *gfp* mRNA (red, middle greyscale) and stained for GFP protein (green, right greyscale) and DAPI (blue) demonstrate that the mRNA level of Ord::GFP is consistent throughout the germarium. (**D-D'**) Quantification of normalized mean intensity of stainings (C-C''). X-axis represents the distance in microns from the niche, Y-axis represents mean intensity normalized to the maximum mean intensity per germarium of *ord* mRNA (D) or Ord protein (D). The line represents fit using a loess regression, shaded area represents the standard error of the fit. (n=8 germaria).

To validate this finding, we examined *orientation disrupter* (*ord)* because it is a well-characterized gene, is required for sister chromatid cohesion, and has previously been reported to peak in expression as meiosis begins in *Drosophila* (S. E. Bickel, Wyman, Miyazaki, Moore, & Orr-Weaver, 1996; Sharon E. Bickel, Wyman, & Orr-Weaver, 1997; Khetani & Bickel, 2007). Our Oo-site results suggested that *ord* mRNA was expressed before meiosis, both from bulk RNA-seq (**Figure 3.6A**) and scRNA-seq (**Figure 3.7D**) consistent with reports that chromosome pairing initiates before meiotic entry (Christophorou, Rubin, & Huynh, 2013; Joyce, Apostolopoulos, Beliveau, & Wu, 2013). However, polysome-seq data were consistent with the observation that Ord protein expression increases during the cyst stages due to translation (**Figure 3.6B**). This led us to predict that *ord* mRNA would be expressed before meiosis, and that Ord protein expression would increase during the cyst stages as previously observed, implying a change in the translation status of *ord* mRNA. To test this, we performed fluorescent *in situ* hybridization against GFP in a fly expressing Ord-GFP under the control of the *ord* promoter and 5'UTR. We visualized both the GFP protein and the mRNA and observed increased expression of Ord::GFP protein but consistent *ord::GFP* mRNA expression, indicating that Ord is controlled post-transcriptionally, likely at the level of translation based on our polysome-seq data (**Figure 3.6C-D'**). This finding also underscores the utility of Oo-site in exploring post-transcriptional gene expression changes.

Figure 3.7: **Genes involved in meiotic cell cycle, including *ord*, may be controlled post-transcriptionally, related to Figure 3.6.**

(**A**) Violin plots of gene expression from RNA-seq of genes in the GO-term category meiotic cell cycle. No significant overall change occurs to expression of these genes at any of the developmental milestones compared to GSCs. Values above plots represent Holm-Bonnferroni adjusted p-values from a Welch's t-test between the indicated genotypes. (**B**) Violin plots of TE from polysome-seq of genes in the GO-term category meiotic cell cycle. Overall TE increases in CBs and cysts significantly compared to GSCs indicating that meiotic entry may be partially controlled post-transcriptionally. Values above plots represent Holm-Bonnferroni adjusted p-values from a Welch's t-test between the indicated genotypes. (**C**) Violin plot of expression of genes in the GO category “meiotic cell cycle” from scRNA-seq. Overall expression of these genes increases in CBs, cysts, and young-WT ovaries compared to the GSC/CB/2CC cluster. Values above plots represent Holm-Bonnferroni adjusted p-values from a Welch's t-test between the indicated genotypes. (**D**) scRNA-seq data indicate that the mRNA level of *ord* is highest in the GSC/CB/2CC cluster, but remains relatively consistent in its expression starting in the 4-CC through 16-CC 2ab clusters and is dramatically decreased in early egg chambers. Color and values indicate the normalized expression of *ord* in each given stage.

To further determine if meiosis is regulated post-transcriptionally, we examined the expression of genes in the GO-term “Double-strand break repair,” which is known to occur during meiosis 1 (Hughes, Miller, Miller, & Hawley, 2018; Page & Hawley, 2003). Double-stranded breaks are resolved before egg chamber formation (Hughes, Miller, Miller, & Hawley, 2018; Mehrotra & McKim, 2006; Page & Hawley, 2003). At the level of input mRNA, we found no significant changes in the expression of genes in this category compared to enriched GSCs (**Figure 3.8A**). From scRNA-seq data, the median expression of double-strand break repair genes significantly increases, but the median increase was only 1.05 fold in 4-CCs and 1.06 in 8-CCs compared to the GSC/CB/2CC group (**Figure 3.8B**). This suggests that double-strand break repair gene transcription begins in GSC stages and increases modestly during the cyst stages.

In contrast, we found a significant increase in the median translational efficiency of double-strand break repair genes, with a 1.20 fold increase in the median translational efficiency in enriched CBs and a 1.56 fold increase in enriched cysts compared to enriched GSCs (**Figure 5C**). In young-WT the median fold change in translational efficiency decreased slightly but significantly compared to enriched GSCs at 0.95 fold. This is consistent with the observed progression of double-stranded break repair that occurs *in vivo*. This demonstrates that Oo-site can be used to derive insights into biological processes that may be changing during early oogenesis (Mehrotra & McKim, 2006; Page & Hawley, 2003). That key processes related to meiosis and differentiation are controlled post-transcriptionally is consistent with the importance of proteins that regulate translation such as Bam and Rbfox1 in differentiation and meiotic commitment during *Drosophila* oogenesis (Blatt, Martin, Breznak, & Rangan, 2020; Carreira-Rosario et al., 2016; Flora, Wong-Deyrup, et al., 2018a; Kim-Ha, Kerr, & Macdonald, 1995; Y. Li, Minor, Park, McKearin, & Maines, 2009; Slaidina & Lehmann, 2014; Tastan, Maines, Li, Mckearin, & Buszczak, 2010).

Figure 3.8: **Genes involved in double-strand break repair may be controlled post-transcriptionally.**

(**A**) Violin plot of expression of genes in the GO category “Double-strand break repair” from bulk RNA-seq. No significant overall change in expression of these genes occurs comparing each genetically enriched developmental stage to GSCs. (**B**) Violin plot of expression of genes in the GO category “Double-strand break repair” from scRNA-seq. Overall expression of these genes increases in CBs, cysts, and young-WT ovaries compared to the GSC/CB/2CC cluster. Values above plots represent Holm-Bonnferroni adjusted p-values from a Welch's t-test between the indicated genotypes (**C**) Violin plot of expression of genes in the GO category “Double-strand break repair” from polysome-seq. Overall expression of these genes increases in CBs, cysts, and young-WT ovaries compared to GSCs. Values above plots represent Holm-Bonnferroni adjusted p-values from a Welch's t-test between the indicated genotypes.

## 3.4 Discussion

We have developed an application that facilitates analysis of bulk RNA-seq, sc RNA-seq, and polysome-seq data of early *Drosophila* oogenesis that is accessible to non-bioinformaticians. We have demonstrated its utility in representing expression at the mRNA and translation level. Additionally, we have demonstrated that it can be used to visualize the expression of groups of genes over development to facilitate hypothesis development. As with all sequencing data, care should be taken to validate findings from Oo-site as sequencing can be influenced by a myriad of factors.

We have used Oo-site to discover that key meiosis regulators such as proteins of the synaptonemal complex and proteins of the double-strand break machinery are regulated at the level of translation. This adds to our understanding of the mechanisms regulating the mitotic to meiotic transition. In future work, identifying the factors mediating the widespread post-transcriptional regulation of crucial meiotic genes and mechanistically how it drives the mitotic to meiotic transition is of high importance.

High-throughput sequencing has enabled researchers to generate more data than ever before However, the development of analysis tools that are usable without bioinformatics training that enable users to make sense of these data to generate hypotheses and novel discoveries has lagged (Shachak, Shuval, & Fine, 2007). Oo-site allows for hypothesis generation and discovery using the powerful model system of *Drosophila* oogenesis. We believe Oo-site might also have utility as a teaching and demonstration tool to introduce students to the power of genomics in developmental biology. The open-source nature of this software facilitates future tool development, which will be crucial as more researchers delve into more data-intensive scRNA-seq, where visualization tools are limited and produce plots that may be difficult to interpret for those not versed in bioinformatics. Oo-site can be supplemented in the future to include additional data such as Cut and Run for various chromatin marks, nascent mRNA transcription using transient transcriptome sequencing or similar techniques, or protein levels from mass-spectroscopy to further extend its utility in hypothesis development.

## 3.5 Acknowledgements

We thank the Drs. Ruth Lehmann and Maija Sladina for sharing scRNA-seq data with us before publication of the manuscript. We are grateful to all members of the Rangan laboratory for discussion and comments on the manuscript. We thank Noor Kotb for naming the dashboard Oo-site. We also thank Dr. Florence L. Marlow for critically reading and editing the manuscript. P.R. is funded by the National Institutes of Health NIGMS (RO1GM11177 and RO1GM135628).

## 3.6 Materials and Methods

The following RNAi stocks were used in this study; *ord-GFP* (Bickel Lab), *Rps19b::GFP* (McCarthy et al., 2021), *UAS-Dcr2;nosGAL4* (Bloomington stock #25751), *bam* RNAi (Bloomington #58178), *hs-bam*/TM3 (Bloomington #24637),

### 3.6.1 Sequencing data

Polysome-seq data were obtained from previous studies conducted by the Rangan lab. Data are available via the following GEO accession numbers:

>UAS-*tkv* GSE171349

>*bam* RNAi GSE171349, GSE166275

>*bam* RNAi; hs-*bam* GSE143728, GSE195893

Young-WT GSE119458

Single-cell sequencing data were obtained from Slaidina *et al*., GEO accession: GSE162192

### 3.6.2 Code Availability

All code used in the preparation of this manuscript is available on GitHub at <https://github.com/elliotmartin92/Developmental-Landscape/tree/master/Paper>

The codebase underlying Oo-site is available on GitHub at <https://github.com/elliotmartin92/Developmental-Landscape/tree/master/ShinyExpresionMap>

### 3.6.3 Antibodies

Mouse anti-1B1 1:20 (DSHB 1B1), rabbit anti-GFP 1:2000 (abcam, ab6556), rabbit anti-Vasa 1:4000 (Upadhyay et al., 2016), chicken anti-Vasa 1:4000 (Upadhyay et al., 2016)

### 3.6.4 Polysome-seq

Flies ready for heat shock were placed at 37°C for 2 hours, moved to room temperature for 4 hours, and placed back into 37°C for 2 additional hours. Flies were then left overnight at room temperature and the same heat shocking procedure was repeated for a total of 2 days. Flies were then dissected in 1x PBS. Polysome-seq was performed as previously described (McCarthy et al., 2021).

### 3.6.5 Polysome-seq data processing

Reads were mapped to the *Drosophila* genome (dm6.01) using STAR version 2.6.1c. Mapped reads were assigned to features also using STAR. Translation efficiency was calculated as in (Flora et al., 2018) using an R script which is available in the Oo-site Github repo. Briefly, TPMs (transcripts per million) values were calculated The log2 ratio of TPMs between the polysome fraction and total mRNA was calculated as such to prevent zero counts from overly influencing the data and to prevent divide by zero errors: . This ratio represents TE, TE of each replicate was averaged and standard error about the calculated average for each gene was calculated.

### 3.6.6 Differential Expression

Differential expression analysis between all bulk RNA-seq samples in a pairwise manner was performed using DEseq2 (Love, Huber, & Anders, 2014). Differential expression was considered as Foldchange > |4| fold, FDR < 0.05.

Differential expression analysis between all polysome-seq samples in a pairwise manner was performed using DEseq2 (Love, Huber, & Anders, 2014) using the model ~ type + genotype + genotype:type with LRT (reduced = ~ type + genotype) to test for changes in polysome counts controlling for input counts. Differential expression was considered as (Foldchange > |2| fold, pvalue < 0.05)

Differentially expressed genes between all germline clusters from scRNA-seq was determined using the FindAllMarkers function from Seurat (Hao et al., 2021). Cutoff was logfc.threshold = 0.75.

Differentially expressed genes between all germarium soma clusters from scRNA-seq was determined using the FindAllMarkers function from Seurat (Hao et al., 2021). Cutoff was logfc.threshold = 0.75.

### 3.6.7 GO term heatmaps

GO-term enrichment analysis was performed using Panther (release 20210224) using the default settings for an Overrepresentation Test of genes differentially expressed between Input samples. Top 5 GO-terms based on fold enrichment of each category were plotted using ggplot2 (Wickham, 2016).

### 3.6.8 Fluorescent *in situ* hybridization

A modified *in situ* hybridization procedure for Drosophila ovaries was followed from Sarkar *et al*. (Sarkar et al., 2021). Probes were designed and generated by LGC Biosearch Technologies using Stellaris® RNA FISH Probe Designer, with specificity to target base pairs of target mRNAs. Ovaries (3 pairs per sample) were dissected in RNase free 1X PBS and fixed in 1 mL of 5% formaldehyde for 10 minutes. The samples were then permeabilized in 1mL of Permeabilization Solution (PBST+1% Triton X-100) rotating in RT for 1 hour. Samples were then washed in wash buffer for 5 minutes (10% deionized formamide and 10% 20x SSC in RNase-free water). Ovaries were covered and incubated overnight with 1ul of the probe in hybridization solution (10% dextran sulfate, 1 mg/ml yeast tRNA, 2 mM RNaseOUT, 0.02 mg/ml BSA, 5x SSC, 10% deionized formamide, and RNase-free water) and primary antibody at 30°C. Samples were then washed 2 times in 1 mL wash buffer with 1ul of corresponding secondary antibody for 30 minutes each and mounted in Vectashield (VectaLabs).

### 3.6.9 Quantification of Stainings

Stainings were quantified using the Fiji Measure tool. Images were aligned and cropped to place the stem cell niche at x=0. Individual cells were outlined within the germarium and Measure was used to calculate the Mean intensity of staining within the cell as well as the X coordinate of the centroid of the cell. Values were normalized to 1 by dividing Mean Intensity values by the maximum of the Mean Intensity per germarium. Data were plotted using ggplot2 and a fit line was added using ggplot2 geom\_smooth with a “loess” function with default settings. The shaded area around the line represents standard error.

# 4 Screen of post-translational modifying enzymes for their effects on translation

Elliot Martin, Siu Wah Wong-Deyrup, Gaby Fuchs, and Prashanth Rangan

## 4.1 Abstract

The ribosome has been demonstrated to play an active role in regulating translation. We hypothesized that post-translational modifications on ribosomal proteins might modulate the activity of the ribosome. Therefore, we attempted to screen for enzymes known to deposit post-translational modifications with the goal of identifying genes that play a role in regulating translation using a dual-luciferase assay which monitors the relative amount of Internal Ribosome Entry Site (IRES) translation compared to cap-based translation. We identified several genes that had a significant effect on the translation of the dual luciferase reporter, however, these affects were modest. We also found that ovaries enriched for stem cells and stem cell daughters had significantly higher relative IRES translation, however, we did not observe the expected result when we knocked down RpS25, which is known to be required for efficient IRES translation. Due to these limitations, we did not pursue these findings further.

## 4.2 Introduction

In the past, ribosomes have been considered as passive players in the process of translation. However, evidence has begun to accumulate that ribosomes can play a regulatory role in translation (Gościńska & Topf, 2020). There are several mechanisms through which ribosomes have been found to regulate translation. One example of this is Rack1 which is a core ribosomal protein (Gerbasi, Weaver, Hill, Friedman, & Link, 2004). Despite the role of Rack1 as a core ribosomal protein, it is dispensable for cell viability, but required for efficient Internal Ribosome Entry Site (IRES) translation (Coyle, Gilbert, & Doudna, 2009; LaFontaine, Miller, Permaul, Martin, & Fuchs, 2020; Majzoub et al., 2014). Furthermore, Rack1 activity can be regulated through post-translational phosphorylation which can be exploited during viral infection to allow for translation initiation on polyA-leader sequences (Jha et al., 2017).

Other mechanisms of translation control involving the ribosome have not been conclusively demonstrated but show promise. For instance, under stress conditions, ribosomal proteins have been shown to carry different post-translational modifications than under basal conditions (Jha et al., 2017; Mukhopadhyay et al., 2008; Simsek & Barna, 2017). The modifications in some cases can alter the propensity of the ribosome to translate certain mRNAs, which could allow the cell to better adapt to varying conditions (Jha et al., 2017).

There are two main modes of translation initiation, canonical cap-based translation and cap-independent translation. Cap-based translation is a complex many-stepped process in which the 43S ribosome pre-initiation complex (43S PIC) is recruited to an mRNA through interacting with cap-binding protein (eIF4E) as well as other components (Jackson, Hellen, & Pestova, 2010). This initiates scanning where the 43S PIC moves along the mRNA until a suitable start codon is identified and translation begins once a 60S ribosomal subunit is recruited (Jackson, Hellen, & Pestova, 2010). Translation can also be initiated cap-independently (Jackson, Hellen, & Pestova, 2010). Several mechanisms can allow translation to occur without the presence of eIF4E, but of interest to this work is IRES-based initiation (Shatsky, Terenin, Smirnova, & Andreev, 2018). IRESs are secondary structures of RNA of which several classes have been described which are classified based on the initiation factors they require as well as their secondary structure (Mailliot & Martin, 2018). IRESs bypass at least the use of eIF4E and can bypass the use of any initiation factors at all, requiring only the 40S and 60S ribosomal subunits to perform initiation (Mailliot & Martin, 2018; Shatsky, Terenin, Smirnova, & Andreev, 2018). In general, IRESs are used at a higher rate under stress conditions such as those initiated by viral infection, but also under other stress conditions (Mailliot & Martin, 2018; Shatsky, Terenin, Smirnova, & Andreev, 2018). However, how cellular stress changes the usage of IRES translation vs cap-based translation is not fully understood.

We performed a screen to attempt to identify enzymes that post-translationally modify proteins and have the potential to alter the translation landscape. To perform this screen, we used the *Drosophila* ovary as it has several advantages. First, *Drosophila* have thousands of available RNAi lines that can be used to knock down genes of interest (del Valle Rodríguez, Didiano, & Desplan, 2012). Second, many of the genes we chose to screen are histone modifying enzymes that may be required for cell survival, by using the ovary as our tissue of interest, we were able to deplete those genes in the germline without affecting the viability of the flies. The last major advantage to working in the *Drosophila* ovary in this case is that endogenous control elements have been identified allowing for expression of a reporter specifically in the germline (Rørth, 1998; Serano, Cheung, Frank, & Cohen, 1994). This in combination with germline specific RNAi allowed us to assay for changes in translation only in the germline tissue.

## 4.3 Results and Conclusions

To assess the status of translation we adapted a dual luciferase assay commonly used to monitor translation status. This reporter contains two ORFs encoded in a single transcript which allowed us to assay for changes in translation while eliminating confounding differences in transcription. The reporter transcript was under the control of a Nos promoter, which is only expressed in the germline. The reporter had a Nos 5'UTR and a K10 3'UTR which have both been demonstrated to have a neutral effect on translation (Gavis & Lehmann, 1992, 1992; Serano, Cheung, Frank, & Cohen, 1994). The first ORF encodes *Renilla* luciferse under the control of the Nos 5'UTR and its production monitors for cap-based translation, this ORF is followed by hairpins to prevent ribosomal read-through. The second ORF encodes a firefly luciferase under the control of the Cricket Paralysis Virus intergenic (CrPV) IRES . These luciferases can be monitored sequentially as they require different conditions in order to produce luminescence and in this way can be used to monitor changes in the relative amount of IRES to cap-based translation.

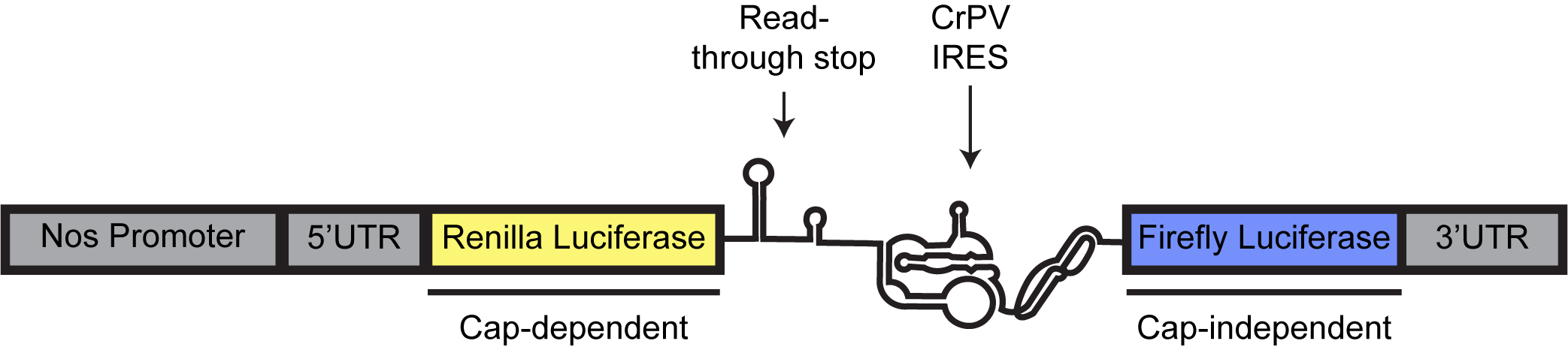


Figure 4.1: **Schematic of dual luciferase reporter used for screen.**

The dual luciferase reporter consists of a Nos promoter and 5’UTR controlling translation initaiation of *Renilla* luciferase followed by a read-though stop consisting of hairpins and the CrPV-IGR IRES which controls initiation of Firefly luciferase.

We performed a screen of 35 genes known to perform post-translational modification to determine if any of them played a role in regulating IRES vs cap-based translation. We also screened several control genes to ensure our reporter system was working as intended. We knocked down RpS25 which is known to be required for IRES, but not cap-based translation, however, we did not find a statistically significant difference between the knockdown and control ovaries (Hertz, Landry, Willis, Luo, & Thompson, 2013; Landry, Hertz, & Thompson, 2009). There are a several explanations of this finding. First, knockdown of RpS25 had a severe phenotype, which was unexpected as previous work indicated that RpS25 is required for IRES, but not cap-based translation (Landry, Hertz, & Thompson, 2009). Therefore, it could be that developmental differences between the control and RpS25 RNAi masked any change in relative IRES translation. Second, it is possible RpS25 is not efficiently depleted or that the RNAi line used has off-target effects. Because of this lack of a result from RpS25 RNAi, our positive control, the remainder of our findings must be taken in context. Several genes were found to have a significant affect on the amount of relative IRES translation, including Not, HDAC3, Lsd1, nsl-1, and Set8.

We additionally included two developmental controls to determine if a phenotype which enriched for stem cells (UAS-*tkv*) or stem cell daughters (*bam* RNAi) would alter the relative IRES translation . Indeed, we found that the relative IRES translation was slight, but significantly elevated. This could suggest that IRES translation occurs more in the stem cells and stem cell daughters of the ovary.

Lastly, we identified five genes that when knocked down resulted in significantly lower relative IRES translation, including *Set8*, *nsl-1*, *Lsd1*, *HDAC3*, and *Not*. However, the affect size observed when these genes were knocked down was modest, with Set8 having the most dramatic affect with a 23.6% reduction in the relative amount of IRES translation. However, it should also be noted that as is widely accepted, no multiple test-correction was performed as for screens multiple-test correction leads to a high degree of type I error and screens typically are followed by confirmatory studies (Gaus, 2015). In this case the results have not been followed-up for the aforementioned limitations regarding RpS25 as well as how the subtle changes in relative IRES translation were.

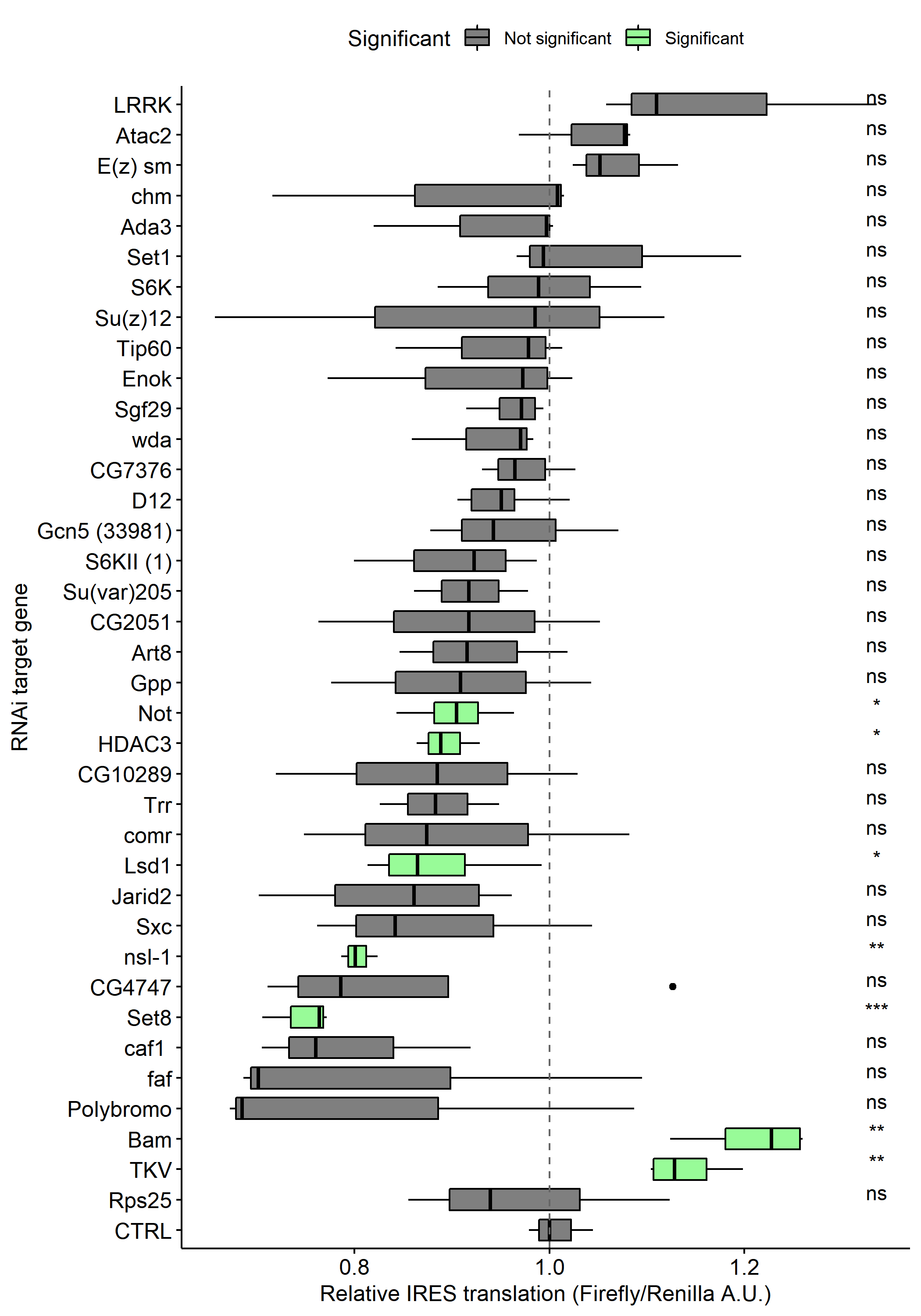


Figure 4.2: **Screen of post-translational modification enzymes.**

Boxplot of relative IRES translation normalized to control ovaries resulting from knockdown of indicated target proteins. Boxes in green represent significant changes in relative IRES-mediated translation compared to control.

As one of the largest changes to relative IRES translation was in *bam* RNAi, which increased relative IRES translation by 23%, it could indeed be that IRES translation is regulated over the course of stem cell differentiation. Future work should be aimed at this aspect of these findings. A spatial version of this reporter could be made using fluorescent reporters to assay the relative IRES translation over stem cell differentiation in the ovary.

## 4.4 Methods

### 4.4.1 Flylines

*nosGAL4* (Bloomington stock 7303)

Generated by this work: *nosGAL4*; *dLuciferase*

Protein targets screened and corresponding RNAi line number used:

|  |  |
| --- | --- |
| Protein Target | Line Number |
| Gpp | v110264 |
| D12 | v29954 |
| Art8 | v100228 |
| Su(z)12 | 33402 |
| Sgf29 | 36637 |
| Atac2 | 32890 |
| CG2051 | 34730 |
| chm | 32484 |
| Tip60 | 28563 |
| Set1 | 33704 |
| chm | 32484 |
| Atac2 | 53918 |
| Lsd1 | V16235 |
| Ada3 (dik) | 28905 |
| Caf1 | 31714 |
| CG10289 | 35597 |
| Spt7 (comr) | 42552 |
| E(z) | 36068 |
| Enok | 29518 |
| faf | 35728 |
| Gcn5 | 33981 |
| Hdac3 | 31633 |
| Jarid2 | 32891 |
| Ubp8 (not) | 28725 |
| nsl1 (wah) | 32561 |
| Polybromo | 32840 |
| Set8 | 35322 |
| Su(var)205 | 33400 |
| Sxc | 50909 |
| Trr | 29563 |
| Wda | 31125 |
| Rps25 | V101342 |
| LRRK | 32457 |
| S6K | 41895 |
| S6KII | 41895 |
| TKV | 36537 |

### 4.4.2 Cloning

A gBlock corresponding to a codon optimized version of *Renilla* and firefly luciferase as well as the readthrough block and CrPV IGR IRES were purchased from IDT. This gBlock was cloned using Gibson cloning according to manufacturer's instructions (#E2611S) into pCasper2 containing a Nos promoter and K10 3'UTR. Colonies were picked, cultured, and plasmids purified and sequenced by Eton Bioscience Inc. to conform the correct sequence had been cloned. A midi-prep scale of the plasmid was prepared using standard methods and plasmids were sent to BestGene Inc. for microinjection.

### 4.4.3 Dual luciferase assay

Ovaries were dissected on 1X PBS in microcentrifuge tubes. Excess PBS was aspirated and ovaries were stored at -20°C. Ovaries were lysed in 25 µl of passive lysis buffer (Promega) using a plastic pestle. 20 µl of lysate was added to a white 96 well plate and luminance was measured using a Glomax 20/20 luminometer (Promega) using the dual luciferase assay reagent (Promega #E1910). At least three biological replicates for Firefly/*Renilla* luminescence values of each RNAi line were generated.

# 5 Conclusion

My work has shined a spotlight on the role of post-transcriptional control in GSC differentiation. Specifically, the ribosome is a nexus for gene expression regulation in three ways. The ribosome is the subject of post-transcriptional gene regulation in order to balance the production of ribosomal proteins with the production of ribosomal RNA, ribosomal proteins are regulated post-transcriptionally. This role oddly ouroboric because the mechanism behind this regulation works by inhibiting translation initiation by the ribosome onto ribosomal mRNAs. Second, the ribosome plays its “traditional” role as a factory to translate genetic instructions into protein products. Although the role of the ribosome in this aspect is “passive” mRNAs must make their way from the nucleus to the ribosome and are subject to degradation, sequestration, and interaction with hundreds or thousands of potential regulators on their journey. Finally, less explored by my work, is the role of post-translational modifications on ribosomal proteins and mRNAs and the role that these modifications play in gene regulation. My work has specifically has highlighted the importance of the first two facets in controlling GSC differentiation and future work should seek to understand the role of post-translational modifications is GSC differentiation and stem cell differentiation more broadly.

This work has developed tools to study and described the crucial role of post-transcriptional gene regulation in GSC differentiation and entry into meiosis in *Drosophila*. Work in other systems has underscored the importance of translation control in stem cell differentiation in general (Gabut, Bourdelais, & Durand, 2020; Sanchez et al., 2016; Woolnough, Atwood, Liu, Zhao, & Giles, 2016; Zahradkal, Larson, & Sells, 1991; Q. Zhang, Shalaby, & Buszczak, 2014). Historically, study of post-transcriptional control has lagged behind that of control at the level of transcription. This is in part due to the lack of equity in tools and techniques between the two areas of study. One salient example is the lag in single-cell sequencing of mRNA which was first published in 2009 compared to the first published example of single-cell Ribo-seq, the first example of which was published in 2021, a 12 year lag (F. Tang et al., 2009; VanInsberghe, van den Berg, Andersson-Rolf, Clevers, & van Oudenaarden, 2021). This exemplifies that those interested in understanding translation control must continue to develop tools and use those tools to better understand developmental systems, regeneration, and disease states. Without equity in understanding between the domains of transcription control and translation control, gaps in our knowledge will prevent our understanding of fundamental biological questions. However, as our toolkits grow we must also remember that every year we generate more data than the previous, but only a subset of the scientific community has the skills necessary to process that data. Therefore, the we must attempt to democratize access to the high-throughput data we generate to empower the research of others.

Future work should aim to understand the pathways that underlie stem cell differentiation and meiotic entry, in particular the interplay and feedback between the Torc1 pathway and the synthesis of ribosomes. Our work has demonstrated that a feedback loop exists between Larp activity and ribosome biogenesis, but future work in other systems should determine if this loop is conserved and if Torc1 acts upstream of Larp as has been shown in other systems. More importantly, how the balance of the Larp-ribosome axis informs differentiation and what levers upstream of Larp might play a role in the initiation of differentiation are of great interest to better understand differentiation, regeneration, and developmental diseases. Emerging techniques such as single-cell Ribo-seq will no doubt allow for these questions to be studied in complex tissues including developmental systems and disease states.

Additionally, future work should examine whether the Larp-ribosome axis is tissue specific. So far we have demonstrated its importance in GSCs, but it remains an open question whether this mechanism acts in other stem cell populations or perhaps in unipotent cells. With our collaborators, we have found certain translational changes related to mitochondrial function that result when perturbing ribosome biogenesis in S2 cells. As Larp as been previously implicated in playing a role in mitochondrial translation in spermatogenesis, this may speak to a tissue specific mode of regulation, however, additional work is required to solidify the linkage between ribosome biogenesis, Larp, and Larp’s targets outside of GSCs. These questions are of great interest as understanding whether Larp may have tissue specific targeting or activity could help explain the tissue specific nature of ribosomeopathies, which has been intensely studied for decades.