**Oo-site: Dashboard to visualize gene expression in the *Drosophila* germarium**

Elliot Martin 1†, Kahini Sarkar, Prashanth Rangan1†,#

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

#Current Address: 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), [prashanth.rangan@mssm.edu](mailto:prashanth.rangan@mssm.edu)

**Summary:** Understanding how stem cells control their differentiation is important for understanding etiology of degenerative disease and as well in designing regenerative therapies. *In vivo* analysisof several stem cell model systems have revealed several paradigms for regulation of stem cell self-renewal and differentiation. The germarium of the female *Drosophila* gonad, which houses both germline and somatic stem cells and their niches, is one such model system. Bulk RNA sequencing (seq), single cell (sc) RNA-seq, and bulk translation efficiency for mRNAs is available for stem cell and their differentiating progeny in the *Drosophila* germarium. However, an easy visualization tool to determine the gene expression and translational efficiency in the germarium is lacking. Here, we have compiled and developed a tool for visualizing bulk- and scRNA-seq and translational efficiency data that represent several stages of germline differentiation to make the data accessible to non-bioinformaticians through a browser that we call Oo-site. Using this tool, we were able to confirm previously reported expression patterns of developmentally regulated genes. Excitingly, we also discovered that meiotic genes such as those that regulate the synaptonemal complex are regulated at the level of translation.

**Introduction:**

The female *Drosophila* ovary provides a powerful system to study stem cell differentiation in an *in vivo* setting (Lehmann 2012; Spradling et al. 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 (Schüpbach 1987; Roth 2001; Xie and Spradling 2000). Each of stage of differentiation of *Drosophila* female germline stem cell (GSCs) are observable and identifiable allowing for temporal changes over GSC development to be easily studied (Xie and Spradling 1998; Dansereau and Lasko 2008; Lehmann 2012). Specifically, the female *Drosophila* GSC undergoes an asymmetric diving giving rise to another GSC and a cystoblast (CB) (Figure 1A) (McKearin and Ohlstein 1995; Chen and McKearin 2003b; Xie and Spradling 1998). The GSC and the CB are marked by a round structure called the spectrosome (Figure 1A) (De Cuevas and Spradling 1998). The CB undergoes four incomplete divisions resulting in 2-, 4-, 8-, and finally 16-cell cysts, which are marked by an extended structure called the fusome (De Cuevas and Spradling 1998; Chen and McKearin 2003a; 2003b). In the 16 cell cysts, one of the cells in the cyst is specified as the oocyte, while the other 15 cells remain as nurse cells, which will provide proteins and mRNAs to support the development of the oocyte (Carpenter 1975; J. Huynh and St Johnston 2000; J.-R. Huynh and St Johnston 2004; Theurkauf et al. 1993; Navarro, Lehmann, and Morris 2001). The 16 cell cysts is encapsulated by somatic cells and buds off from the germarium, forming an egg chamber (Narbonne-Reveau et al. 2006; Xie and Spradling 2000; Forbes et al. 1996). In each chamber, the oocyte grows as the nurse cells synthesize mRNAs and proteins and deposits these into the oocyte which eventually gives rise to a mature egg (J. Huynh and St Johnston 2000; Narbonne-Reveau et al. 2006).

Expression of differentiation factors that regulate translation results in progressive differentiation of GSCs to an oocyte (Blatt et al. 2020; Slaidina and Lehmann 2014). The CBs express differentiation factor Bam which promotes transition from a CB to an 8-cell cyst stage (Chen and McKearin 2003a; McKearin and Ohlstein 1995; Ohlstein and McKearin 1997). In the 8-cell cyst, expression of Rbfox1 promotes exit from the mitotic cell cycles into meiosis (Carreira-Rosario et al. 2016). Both the differentiation factors Bam and Rbfox affect translation of mRNAs to promote differentiation (Li et al. 2009; Carreira-Rosario et al. 2016). In addition, in the 8-cell cyst stage, recombination is initiated in many cells of the cysts and then eventually restricted to the specified oocyte (Hinnant, Merkle, and Ables 2020; J. Huynh and St Johnston 2000). The mRNAs that are translationally regulated during this progressive differentiation are not all known nor is how recombination is temporally regulated.

Somatic cells surround the developing germline, and the germline relies on these cells for signaling, structure, and organization (Schüpbach 1987; Roth 2001; Xie and Spradling 1998; 2000). In germarium, terminal filament, cap and anterior-escort cells act as somatic niche for the GSCs (Decotto and Spradling 2005; Xie and Spradling 2000). Once GSCs divide to give rise to CBs the posterior escort cells guide CB differentiation by encapsulating it and the early-cysts stages (Upadhyay et al. 2016; Shi et al. 2021; Kirilly, Wang, and Xie 2011). Follicle stem cells (FSCs) which are present towards posterior of the germarium divide and differentiate to give rise to follicle cells, (FCs) which surround late-stage cysts that give rise to egg chambers (Nystul and Spradling 2010; Margolis and Spradling 1995). FSCs also give rise to stalk cells and polar cells which connect the egg chambers to form the ovariole (Margolis and Spradling 1995).

While wealth of bulk-, sc- and translational efficiency that was obtained from polysome- seq experiments are available for the cells in the germarium, there are several hurdles for easy utilization of the data:

1. scRNA-seq has exquisite temporal resolution but it can miss some lowly expressed transcripts which can be captured by bulk-seq (Lähnemann et al. 2020). In addition, there is no easy way to compare these two data sets.

2. While scRNA-seq provides mRNA levels, this is does not mean that these mRNAs are translated especially in the germline where translation control plays an important role.

3. Lastly, there is a barrier to visualization of the data for people who are not experienced in bioinformatics.

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

**Results and discussion**

To make data bulk-, 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 1B**). Notably, Each genetically enriched sample had matched mRNAseq and polysome-seq libraries prepared, allowing for simultaneous read-out of mRNA level and translation status (**Supplemental Figure 1A**). Additionally, we have integrated scRNA-seq data from Slaidina et al. which provides finer temporal resolution in the cyst stages but was unable to cluster GSCs from CBs or 2-cell cysts (Slaidina et al. 2021). We present this data as a tool called Oo-site which consists of a collection of user-interactable visualizations allowing researchers to easily determine the expression pattern of a gene of interest or the expression pattern of a collection of genes provided by the user. Oo-site consists of three modules, ovary-map, ovary-heatmap, and ovary-violin (**Figure 1C**). Each module of Oo-site 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. Ovary-map allows users to visualize the expression of a single gene over the course of differentiation in the form of a germarium schematic, which eases understanding of 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 1B, Supplemental Figure 1B-C’**). Finally, ovary-violin allows users to visualize the expression of multiple genes over the course of differentiation (**Figure 1C**). These groups of genes can be selected either by a GO-term of interest or a custom list of genes can be supplied by the user. The user can download a spreadsheet of gene expression corresponding to the subset of genes that they have provided or defined by the GO-term the user has selected. Finally, Oo-site incorporates a report tool which generates a downloadable report of the visualization(s) in a standardized format to facilitate their use for publication (**Figure 1C**). Researchers can use these datasets to enhance their hypothesis generation or to confirm expression patterns they have observed from other methods.

Using Oo-site, we first asked if the bulk mRNA-seq data that was acquired by enriching for specific stages of germline development is representative for gene expression patterns from purified cell types. We compared bulk-seq data that we obtained by enriching GSC and CB cell types but were not purified away from somatic cells (**Figure 1C**) to the GSC and CB data from Wilcockson et al where they included a fluorescent-assisted cell sorting (FACS) step such that only a pure population of these cells were sequenced (Wilcockson and 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 bulk-seq data these genes follow similar trends as identified by Wilcockson *et al*, indicating that despite the lack of FACS purification, our data reproduces meaningful expression changes at the mRNA level for these stages (**Supplemental Figure 2A-A'**).

To validate that our bulk- seq data recapitulates genuine changes in gene expression, we compared expression of select genes such as RpS19b in scRNA-seq data. Our bulk-seq data as well as the available SC-seq data indicates that RpS19b is highly expressed in GSCs that decreases over differentiation in the cyst stages with greatly decreased expression in early egg chambers consistent with previous reports (**Fig 2A-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* is present in the GSCs and diminishes in the cyst stages in both enriched bulk-seq and scRNA-seq (**Figure 2C-D**). Additionally, RpS19b::GFP expression resembles it’s mRNA expression indicating that its dynamic expression is achieved primarily through modulating the mRNA level of RpS19b, consistent with its moderate to high TE in early stages (**Supplemental Figure 2B**). Thus, enriching for specific germline stages captures changes to gene expression in the germline. However, care must still be taken in interpreting bulk-seq results from the data may be influenced by the somatic cells present in the samples.

To determine what groups of genes are changing as the GSCs differentiate into an egg, we used GO-term analysis to probe for pathways that change at the level of RNA using bulk-seq data. We did not identify any significant GO-terms in 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 more highly in differentiating cysts than GSCs (Figure 3A). We also identified the polytene chromosome puffing GO term in genes downregulated in CBs compared to cysts. 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-WT flies and in CBs compared to differentiating cysts, suggesting that metabolic processes may be altered during GSC differentiation. Additionally, we found that downregulated genes in CBs and differentiating cysts compared to young-WT, which contain egg chambers, are enriched in GO-terms involving vitelline and eggshell coat proteins (Figure 3A).

Next, to examine if our data can resolve large scale expression changes that occur during oogenesis we examined the expression of genes in the GO-term Meiosis I. Meiosis I is initiated at 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; *bam*-HS samples (ref). 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. This is consistent with the observation that several factors that promote meiosis 1 are transcribed in the GSCs and the stages therein (McCarthy et al. 2021). This suggests that, in general, transition from a mitotic state to a meiotic state is not driven by changes to mRNA level of genes associated with meiosis I.

As we did not see overall changes to mRNA levels of genes in the GO-term meiosis I, we next examined the polysome-seq data of genes in the meiosis I GO-term to determine if changes in expression of these genes might occur at the level of translation. Polysome-seq uses polysome profiling to separate mRNAs that are associated to polysomes which are formed by mRNAs engaged with multiple ribosomes. To quantify the degree to which an mRNA is associated with the polysome fractions, we sequenced mRNAs from the polysome fractions and compared their relative expression to their relative expression from corresponding input lysate yielding a metric referred to as translational efficiency. Indeed, genes in the meiosis I GO-term had a significant increase in translation efficiency in CBs and a more dramatic increase in cysts, but did not observe significant changes to the overall mRNA level of these genes (**Figure S4A-B**). From single cell seq data, the median expression of genes in this category increases slightly but significantly in clusters 4CC, 8CC, 16CC 2a I, 16CC 2a 2, and 16CC 2ab, but the median increase in expression is at most 0.124 fold in cluster 16CC 2a I. This suggests that some genes in the meiosis I GO-term may be regulated at the mRNA level, but as a group this regulation is modest. (are they made in GSCs? If so you should say that)

To validate this finding, we chose to examine *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*. Our Oo-site data suggests that Ord mRNA is expressed prior to meiosis, both from bulk mRNA-seq (**Fig 4A**) and SC-seq (**Fig S4C**) consistent with reports that chromosome pairing initiates prior to meiotic entry (**ref**). However, polysome-seq data is consistent with the observation that Ord protein expression increases during the cyst stages, at the level of translation. This led us to predict that *ord* mRNA would be expressed prior to meiosis, but the 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 expression as well as the mRNA expression and found an increase in the expression of Ord::GFP proteinas compared to the expression of *ord::GFP* mRNA, indicating that Ord is controlled post-transcriptionally, likely at the level of translation based on our polysome-seq data. This finding also underscores the utility of Oo-site in exploring post-transcriptional gene expression changes.

To determine if changes to biological processes could be detected using our tool, we examined the expression of genes in the GO-term “double-strand break repair” as this process is known to occur as a product of meiotic recombination. Specifically, during meiotic stages, double stranded breaks occur and are repaired. These double stranded breaks are resolved prior to egg chamber formation. At the level of input mRNA, we found no significant changes in the expression of genes in this category compared to enriched GSCs. From single cell seq data, the median expression of genes in this category significantly increases, but the median increase is only 1.05 fold in 4-CCs and 1.06 in 8-CCs compared to the GSC/CB/2-CC group.

However, at the level of translation, we found a significant increase in the median TE of genes in the category “double-strand break repair” with a 1.20 fold increase in the median TE of genes in enriched CBs and a 1.56 fold increase in enriched cysts compared to enriched GSCs. In young-WT the median fold change in TE decreases 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*, demonstrating our data can be used to derive insights into biological processes that may be changing during early oogenesis. 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, Rbfox1, and Bruno1 in differentiation and meiotic commitment during *Drosophila* oogenesis. However, the targets of these regulators remain an active area of study and we speculate that the post-transcriptional regulation that we observe may involve these key regulators.

**Conclusions and future directions:**

We have developed an application that facilitates analysis of mRNA-seq, polysome-seq, and single cell-seq data of early *Drosophila* oogenesis that is geared towards non-bioinformaticians. We have demonstrated its utility in recapitulating the 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.

We have discovered that a key meiosis regulator and a member of the synaptonemal complex is also regulated post-transcriptionally, at the level of translation. This adds to our understanding regarding how the mitotic to meiotic transition in regulated. We leave to future work to determine how widespread post-transcriptional regulation of crucial meiotic gene drives the mitotic to meiotic transition and mechanistically what factors underly this regulation.

High-throughput sequencing has enabled researched to generate more data than ever before. However, the development of analysis tools that are usable without bioinformatics training to make sense of this data and allow for hypothesis generation and novel discoveries has lagged behind. Oo-site allows for hypothesis generation and discovery using the powerful model system of *Drosophila* oogenesis. Additionally, the open-source ­­nature of this software allows for future tool development which will be crucial as more researchers delve into single-cell sequencing, where even more data is generated than traditional RNA-seq, but visualization tools are limited and produce plots that may be difficult to interpret to those not versed in bioinformatics and in particular single-cell sequencing. Oo-site can be supplemented by future work to include additional data such as Cut and Run for various chromatin marks, nascent mRNA transcription using TT-seq or similar techniques, or protein levels from mass-spectroscopy to further enhance its utility in hypothesis development.

**Figure Legends:**

**Figure 1: titles .**

(A) Developmental stages of germline development. (B) Summary of the samples used for input mRNAseq and polysome-seq and the cell types these samples are enriched for. C

**Figure 3**

(A-B) Visualization of expression of RpS19b over germline development from (A) developmentally enriched stages and (B) single-cell seq data indicates that the mRNA level of RpS19b decreases starting in the cysts and is dramatically decreased in early egg chambers. Color indicates relative expression and displayed values indicate the mean TPM of RpS19b in each given stage. (C-C’’) Confocal images of ovaries with in-situ hybridization of RpS19b (green, middle greyscale) XXX 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, stained for (D’) GFP (green, middle greyscale), (D’') Vasa (blue, right greyscale), and 1B1 (red) demonstrate that the protein expression of RpS19b::GFP is consistent with its mRNA levels.

**Figure 2**

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

**Figure 4**

(A-B) Visualization of expression of Ord over germline development from (A) RNAseq 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. (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.

**Figure 5**

(A) Violin plot of expression of genes in the GO category Double-strand break repair from input mRNAseq. 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 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 resulting from a one-sided t-test with mu=0. (C) 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 resulting from a one-sided t-test with mu=0.

**Figure S1**

(A) Schema 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 input mRNAseq 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.

**Figure S2**

(A-A’) Violin plots of expression from bulk mRNAseq 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 mRNAseq 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 resulting from a one-sided t-test with mu=0. (C) Visualization of expression of RpS19b over germline development from (C) polysome-seq data. Color indicates relative expression and values indicate the mean TE of RpS19b in each given stage. RpS19b TE is relatively consistent during early oogenesis and decreases in the egg chambers.

**Figure S4**

(A) Violin plots of gene expression from mRNAseq of genes in the GO-term category Meiosis I. 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 resulting from a one-sided t-test with mu=0. (B) Violin plots of TE from polysome-seq of genes in the GO-term category Meiosis I. 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 resulting from a one-sided t-test with mu=0. (C) Single-cell seq data indicates that the mRNA level of *ord* increases starting in the cysts and is dramatically decreased in early egg chambers. Color indicates relative expression and values indicate the normalized expression of *ord* in each given stage.

**Materials and Methods:**

Fly lines:

ord-GFP (Bickel Lab)

*nosGAL4*

smFISH:

Sequencing data:

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

UAS-tkv GSE171349

bam RNAi GSE171349, New

bam RNAi; HS-bam GSE143728, New

Young wild-type GSE119458

Code Availibility:

All code used in preparation of this manuscript as well as the code that powers OO-site are available on GitHub at <https://github.com/elliotmartin92/Developmental-Landscape>

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 a custom R script. Briefly, CPMs (counts per million) values were calculated. Any gene having zero reads in any library was discarded from further analysis. The log2 ratio of CPMs between the polysome fraction and total mRNA was calculated and averaged between replicates. This ratio represents TE, TE of each replicate was averaged. Targets were defined as transcripts falling greater or less than one standard deviation from the median TE in *aramis* RNAi, but not in either of the two developmental controls (NosGAL4>UAS-Tkv or NosGAL4>bam RNAi). Only targets meeting a conservative expression cutoff of log2(TPM) expression greater than five were considered to exclude more lowly expressed genes as they are highly influenced by noise in polysome-seq.

GO term heatmaps:

GO-terms were obtained from Panther release 20210224 using the default settings for an Overrepresentation Test of genes differentially expressed between Input samples compared in a pairwise manner (Foldchange > |2| fold, FDR < 0.1). Top 5 GO-terms based on fold enrichment of all categories were plotted.