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

Elliot Martin 1†, Kahini Sarkar, Noor Kotb, Prashanth Rangan1†

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

†Co-corresponding authors

Email: [etmartin@albany.edu](mailto:etmartin@albany.edu), [prangan@albany.edu](mailto:prangan@albany.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)-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 sc-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 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. 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. 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. Specifically, the female *Drosophila* GSC undergoes an asymmetric diving giving rise to another GSC and a cystoblast (CB) (Figure 1A). The GSC and the CB are marked by a round structure called the spectrosome (Figure XX). 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. 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. The 16 cell cysts is encapsulated by somatic cells and buds off from the germarium, forming an egg chamber. 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.

Expression of differentiation factors that regulate translation results in progressive differentiation of GSCs to an oocyte. The CBs express differentiation factor Bam which promotes transition from a CB to an 8 cell cyst stage. In the 8-cell cyst, expression of Rbfox1 promotes exit from the mitotic cell cycles into meiosis. Both the differentiation factors Bam and Rbfox affect translation of mRNAs to promote differentiation. 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. What are all the mRNAs are translationally regulated during this progressive differentiation is not 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. In germarium, terminal filament, cap and anterior-escort cells act as somatic niche for the GSCs. Once GSCs divide to give rise to CBs the posterior escort cells guide CB differentiation by encapsulating it and the early-cysts stages. 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. FSCs also give rise to stalk cells and polar cells which connect the egg chambers to form the ovariole.

While wealth of bulk-, sc- and polysome- seq data are available for the cells in the germarium there are several hurdles for easy utilization of the data:

1. Sc-seq has exquisite temporal resolution but it can miss lowly expressed transcripts which can be captured by bulk-seq. There is no easy way to compare these two data sets.

2. While sc-seq gives us 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 sc-seq, bulk-seq and polysome-seq data to visualize expression of a gene of interest in the germarium and translational efficiency in the germline.

**Results and discussion**

To make data bulk-, sc-, 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**). Each genetically enriched sample had matched mRNAseq and polysome-seq libraries prepared, allowing for simultaneous read-out of mRNA level and translation status. Additionally, we have integrated sc-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. 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 XXX). 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 sc-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 XXX). Finally, ovary-violin allows users to visualize the expression of multiple genes over the course of differentiation (Figure XXX). 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 XXX). Researchers can use these datasets to enhance their hypothesis generation or to confirm expression patterns they have observed from other methods.

Using this tool, 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 specific purified cell types. We compared our bulk-seq mRNA-seq data to previously reported mRNA-seq data generated from similar genetic enrichment strategies that included a fluorescent-assisted cell sorting (FACS) step so that only a pure population of germline cells were sequenced. Comparing GSC and CB enriched data set to GSC and CB purified data set, we find that genes 2-fold or more down or upregulated with a p-value < 0.01, from the previously published data follow similar trends in our data, indicating that despite the lack of FACS our data reproduces meaningful expression changes at the mRNA level for these stages (Supplemental Figure 2A-A').

To validate our bulk- seq data for the genetically enriched stages for which no previous purified mRNA seq libraries have been published, we compared expression of select genes such as RpS19b in sc-seq data. Our mRNA-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 (Fig 3A-B). 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. We found that RpS19b is present in the GSCs and diminishes in the cyst stages via both enriched bulk-seq and sc-seq (Figure 3C-D). Thus, enriching for specific germline stages captures changes to gene expression by purifying specific cell types during oogenesis. (caveat here)

To determine what groups of genes are changing as the GSCs differentiate into an egg, we used GO-term analysis. We found that genes expressed more highly in young WT, which contains egg chambers, than in CBs and differentiating cysts are enriched in GO-terms involving vitelline and eggshell coat proteins. Additionally, 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.(!!!) what does this means?!! Fast DNA divisions? 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. 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. This suggests……though in single cell seq data they see this? we need to say as you have for DSBs

Need an intro of what thispoly-seq data is….

However, we did see that genes in this GO-term had a significant increase in translation efficiency in CBs and a more dramatic increase in cysts. To validate this finding we chose to examine ord and c(3)G because they are well characterized genes, required for sister chromatid cohesion, and have previously been reported to be expressed as meiosis begins in *Drosophila*. Our Oo-site data suggests that Ord and c(3)G mRNA is expressed prior to meiosis, both from bulk mRNA-seq (Fig 4A) and SC-seq (Fig n). However, our 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 a dramatic difference in the expression of *ord::GFP* mRNA as compared to the expression of Ord::GFP protein, indicating that Ord is controlled post-transcriptionally, likely at the level of translation based on our polysome-seq data.

To determine if changes 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.

**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. Can other data acquired by added to this dashboard?etc

**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, ?

bam RNAi; HS-bam GSE143728, ?

Young wild-type GSE119458

Single cell sequencing data was obtained from Slaidina et al., GEO accession: GSE162192

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.