**Oosite: Dashboard to visualize gene expression in the *Drosophila* germarium**

Elliot Martin 1† and 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 has important implications in disease states and in regenerative therapy. *In vivo* analysisof several model systems using high throughput methods to parse gene expression have revealed several paradigms for regulating stem cell state. One model system that has given us tremendous insight for understanding stem cell biology is the germarium of the female *Drosophila* gonad, which houses several kinds of stem cells and their niches. Bulk and single cell sequencing data as well bulk translation efficiency for mRNAs are available for the *Drosophila* germarium. However, there is a lack of an easy visualization tool to determine the expression of a gene of interest in the germarium. Here, we have compiled and developed a tool for visualizing mRNA levels and post-transcriptional data that represent several stages of differentiation in the germarium. *Drosophila* germline stem cell (GSC) differentiation program. We have developed visualization tools to make this data accessible to non-bioinformaticians that are accessible through a browser. Using this tool we were able to confirm previously reported expression patterns of developmentally regulated genes. Excitingly we used our tool to search for meiotic genes that may be regulated at the translation level and found two synaptonemal complex genes are indeed regulated post-transcriptionally, un.

**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 from a single ovary 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 CB undergoes four incomplete divisions resulting in 2, 4, 8, and finally 16 cell cysts (CCs). During these divisions, the cysts transition from a mitotic cell cycle to a meiotic cell cycle and undergo recombination. Additionally, in the 16-CC 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-CC is encapsulated by somatic cells and buds off from the germarium, forming an egg chamber and subsequently the oocyte develops into a mature egg. Somatic cells surround the developing germline and the germline relies on these cells for signaling, structure, and organization. These cells are derived from follicle stem cells (FSCs) which are present towards posterior of the germarium. FSCs divide and differentiate to give rise to several cell types including cap cells (CCs), which provide a niche to the GSCs, escort cells (ECs) which provide signaling to the early germline, follicle cells, (FCs) which surround the oocyte and nurse cells and encapsulate the egg chambers, as well as stalk cells and polar cells which connect the egg chambers to form the ovariole.

However, one weakness to studying stem cell differentiation within a tissue has been the inaccessibility to high throughput methods that typically require a relatively homogenous collection of cells. Single cell seq has bridged this gap however, it has several limitations which have not been overcome as of the writing of this manuscript. First, single cell-seq has historically been limited to mRNA level data, and has only recently been applied to ribosome profiling. Second, it has a higher limit of detection than bulk mRNA-seq, preventing the detection of rare, but potentially important mRNAs. Finally, single cell-seq provides far more data than bulk sequencing methods, increasing its inaccessibility for non-bioinformaticians to visualize spatial expression of genes.

Here, we use *Drosophila* genetics in order to circumvent the limitations of single cell seq and apply bulk sequencing techniques to tissue enriched for several stages of GSC differentiation. We present this data, alongside previously published single cell-seq data from Drosophila ovaries in a tool called Oosite.

**Results and discussion**

Paragraph here about what is available :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. Single cell seq? What is not known?

In order to make this data as accessible to the community as possible, we present this data as a tool called Oosite 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. Oosite consists of three modules, ovary-map, ovary-heatmap, and ovary-violin. Each module of Oosite 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 single-seq data. Ovary-map allows users to visualize the expression of a single gene over the course of differentiation in the form of a cartoon germarium, 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. Finally, ovary-violin allows users to visualize the expression of multiple genes over the course of differentiation. 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 is able to 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, Oosite incorporates a report tool which generates a downloadable report of the visualization(s) in a standardized format to facilitate their use for publication. Researchers can use these datasets to enhance their hypothesis generation or to confirm expression patterns they have observed from other methods.

To determine if the bulk mRNA-seq data is representative of previously observed expression patterns, we first compared our mRNA-seq data to previously reported mRNA-seq data generated from similar genetic enrichment strategies to enrich for GSCs and GSC-daughter cells, however, the previously published data included a fluorescent-assisted cell sorting (FACS) step so that only a pure population of germline cells were sequenced. Indeed, we find that genes identified as being differentially expressed 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 (Supplemental Figure 2A-A'). To validate our mRNA seq data for the genetically enriched stages for which no previous mRNA seq libraries have been published and to ensure that the mRNA seq results we examined RpS19b which is known to be dynamically expressed during oogenesis. Our mRNA-seq data as well as the available SC-seq data indicates that RpS19b is highly expressed in GSCs that decreases over differentiation with greatly decreased expression in egg chambers (Fig 3A-B). We validated the expression of RpS19b in-vivo using in-situ hybridization as well as an RpS19b::GFP line under endogenous control elements and found that RpS19b seems to be regulated primarily at the mRNA level as we would expect from our sequencing data (Figure 3C-D).

To determine what groups of genes are changing during these developmental transitions, we used GO-term analysis .

Next, to determine if the polysome-seq data is representative of biologically meaningful changes in translation status, we (either do staining/in situ or refer) examined the expression pattern of a gene that is known to have dynamic expression during GSC differentiation. We chose to examine orientation disrupter (ord) because it is a well characterized gene, required for sister chromatid cohesion and has previously been reported to be expressed as meiosis begins in *Drosophila*. Our Oosite data suggests that Ord mRNA is expressed prior to meiosis, both from bulk mRNA-seq (Fig n) 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.

Another mRNA whose translation is well characterized and tightly regulated during GSC differentiation is *pgc. pgc* is a maternally deposited mRNA that encodes a protein that silences RNA Pol II via phosphorylation of the Ser2 residue of Pol II. During GSC differentiation, *pgc* mRNA is expressed, but only translated in a short burst in stem cell daughters, prior to Bam expression. Interestingly, from our mRNA-seq data, we observe a dramatic dip in the mRNA level of *pgc* in stem cell daughters and a simultaneous increase in the translation efficiency. Although the reason for the lowered mRNA expression in stem cell daughter cells is unknown, the increase in *pgc* translation is consistent with its reported expression. Perhaps the RNA is translated and destroyed? If we say this – we kind of say pooja’s paper is sort of wrong…….

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 in order to visualize the expression of groups of genes over development to facilitate hypothesis development.

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. Oosite 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.