Short Paper

Elliot Todd Martin*,[a], Noor Kotb[a,b], Gabriele Fuchs[a], Prashanth Rangan*,[a]

[a] *Department of Biological Sciences/RNA Institute, University at Albany SUNY, Albany, NY 12202, USA.*
[b] *Department of Biomedical Sciences, School of Public Health, University at Albany SUNY, Rensselaer, NY, 12144, USA.*

Abstract

Proper stem cell differentiation requires both transcriptional and post-transcriptional changes. Previous work at characterizing these changes have primarily focused either on cell culture systems, which lack biological context. In-vivo work on stem cell systems has biological context, but these models are less tractable to high throughput methods especially for post-transcriptional changes. Here, we have compiled and developed tooling for mRNA level and post-transcriptional data that represent several stages of differentiation of the female *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. We confirmed the expression of genes that have been described previously and elucidate...

Introduction:

The female *Drosophila* germline provides a powerful system to study stem cell differentiation in an in-vivo setting. *Drosophila* are one of the most genetically tractable model organisms available and hundreds of thousands of mutant, RNAi, overexpression, and reporter lines are commercially available. Additionally, 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. 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 promises to bridge this gap however, it has several limitations which have not been overcome as of the writing of this manuscript. First, single cell-seq is limited to mRNA level data, it cannot yet be applied to techniques such as polysome or ribosome sequencing. Second, it has a higher limit of detection than bulk mRNA-seq, preventing the detection of rare, but potentially important mRNAs. Here, we use *Drosophila* genetics in order to

*Corresponding Author
**Equal contribution
Email addresses: etmartin@albany.edu (Elliot Todd Martin), prangan@albany.edu (Prashanth Rangan)

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 Ovary-App.

Results:

Ovary-App.

We present the tool Ovary-App 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. Ovary-App consists of three modules, ovary-map, ovary-heatmap, and ovary-violin. Each module of Ovary-App 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 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. Researchers can use these datasets to enhance their hypothesis generation or to confirm expression patterns they have observed from other methods.

mRNA Expression Validation.

To determine if the bulk mRNA-seq data we have collected 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 FACS step so that only are 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. 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 observe recapitulate we used in-situ probes targeting RpS19b. 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 n).

Translation Efficiency Validation.

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 Bru1 because it has been previously observed to dramatically increase in expression during the cyst stages of GSC differentiation at the protein level. Our Ovary-App data suggests that Bru1 mRNA levels are relatively consistent during early oogenesis, both from bulk mRNA-seq (Fig n) and SC-seq (Fig n), until the egg chamber stages where Bru1 levels dramatically increase. However, our polysome-seq data is consistent with the observation that Bru1 expression increases during the cyst stages, at the level of translation. This led us to predict that Bru1 mRNA levels would be consistent GSC differentiation, but the protein expression would increase during the cyst stages, implying a change in the translation status of Bru1 mRNA. 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.

Bibliography styles

There are various bibliography styles available. You can select the style of your choice in the preamble of this document. These styles are Elsevier styles based on standard styles like Harvard and Vancouver. Please use BibTeX to generate your bibliography and include DOIs whenever available.

Here are two sample references: (Dirac, 1953; Feynman and Vernon Jr.; 1963).

References

Dirac, P.A.M., 1953. The lorentz transformation and absolute time. Physica 19, 888–896. doi:10.1016/S0031-8914(53)80099-6

Feynman, R.P., Vernon Jr., F.L., 1963. The theory of a general quantum system interacting with a linear dissipative system. Annals of Physics 24, 118–173. doi:10.1016/0003-4916(63)90068-X