Using bioinformatics algorithms to detect differentially expressed genes in *Schistosoma mansoni* miracidia

Brandy Corwin and Nicolas J. Wheeler

Department of Biology
University of Wisconsin – Eau Claire

Introduction

- *Schistosoma mansoni* are a parasitic worm that causes chronic schistosomiasis.
- *S. mansoni* trigger an immune response that leads to large tissue granulomas, causing fibrosis, organomegaly, and tissue damage.
- *S. mansoni* have a complex life cycle that is shown in figure 1.

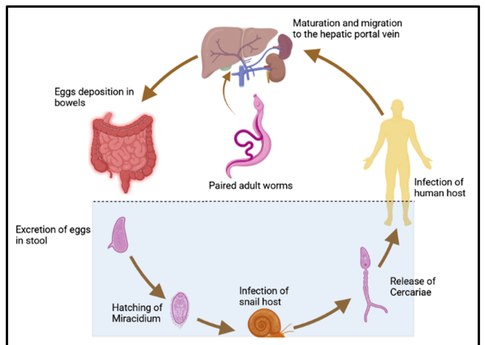

Figure 1. Life cycle of *Schistosoma mansoni*.

- Some portion of the eggs don't go through the intestine, but instead can end up in other areas, particularly the liver. The eggs that make it to the intestines continue their life cycle, while the others are trapped and eventually die.
- Since most of our previous knowledge is based on studies looking at eggs trapped in the liver instead of the eggs that move through the intestines, we want to understand the differences, if there are any, between miracidia from the liver versus the intestines.
- A past study was done that determined that was a difference in gene expression between *S. mansoni* eggs isolated from the liver versus the intestines.
- Our goal is to see if the miracidia also exhibit a difference in gene expression between those stuck in the liver and those in the intestines.

Methods

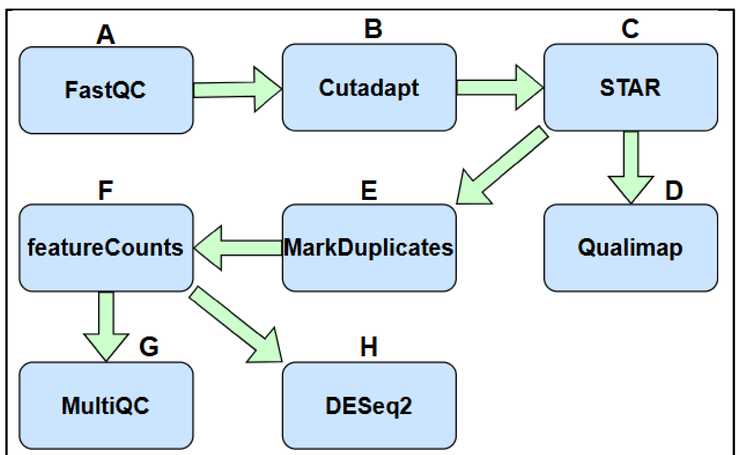

Figure 2. Flowchart of methods used in detecting differential gene expression in miracidia.

A. FastQC. Checked the quality of raw sequence data. [4]

B. Cutadapt. Forward and reverse adapters were trimmed as well as the first 10 bp cue to a barcode. [5]

C. Star. Aligned the sequence reads to the reference genome. [6]

D. Qualimap. Checked the quality control of alignment sequencing data. [7]

E. MarkDuplicates. Tagged duplicate reads based on sequence similarity and quality scores. [8]

F. FeatureCounts. Counted mapped reads for each type of genomic feature such as genes, exons, etc. [9]

G. MultiQC. Created an HTML report of all samples. [10]

H. DESeq2. Used to test for differential gene expression of read counts. [11]

Results

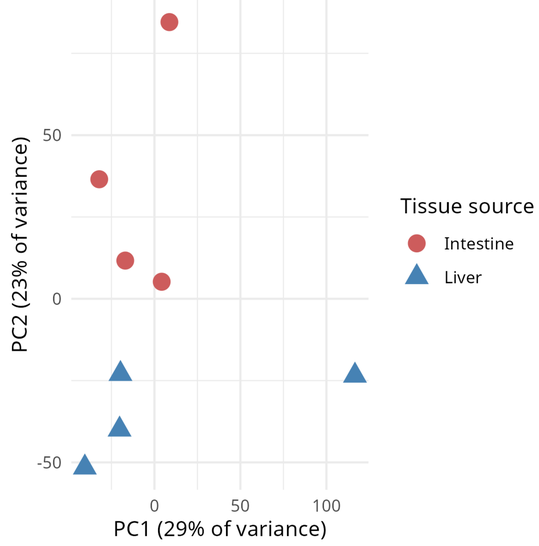

Figure 3. PCA plot of miracidia samples from the liver and the intestines. There is clear clustering of the samples by tissue source, indicating that there is a difference in gene expression.

Key Findings

- There is a noticeable difference in gene expression between the miracidia from different tissue sources. (Fig. 3)
- There are 14 genes that are significantly different in terms of gene expression. (Fig. 4)
- One statistically significant function found was cell adhesion. (Fig. 6)

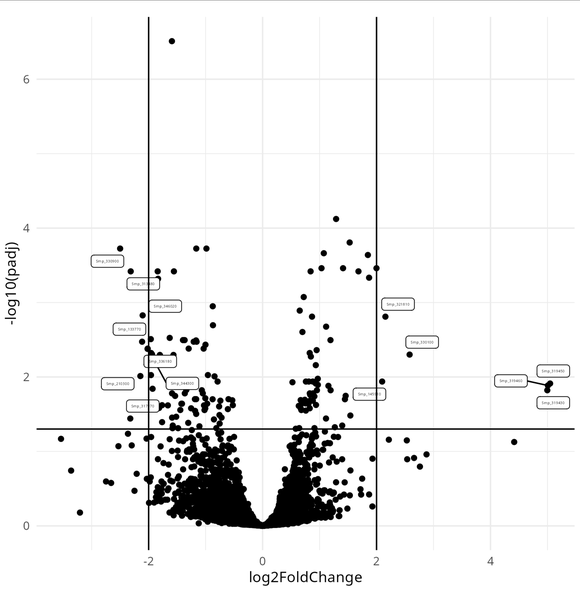

Figure 4. Volcano plot showing genes that were statistically significantly differentially expressed. We determined genes to be statistically significant when padj <= $-\log_{10}(0.05)$ and the $\log_2$ fold change > 2 or < -2.

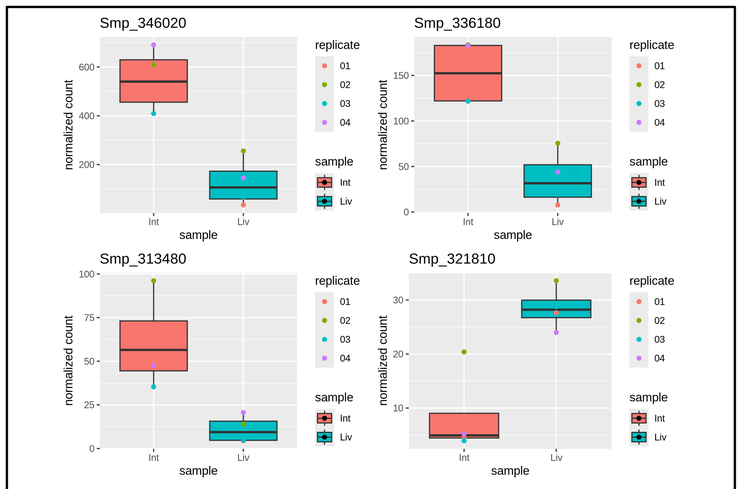

Figure 5. Boxplots of four selected differentially expressed genes. These four genes were selected from a total of 14 genes that all had a padj <= 0.05 and a $\log_2$ fold change > 2 or < -2. Each replicate point represents one miracidia sample.

Future Directions

- We have determined that certain genes are differentially expressed in miracidia in liver versus the intestines, so future work could look at how those differences affect the miracidia.
- More future work could be done to repeat previously done studies that used samples from the liver, but this time use samples from the intestines and see if the results change.

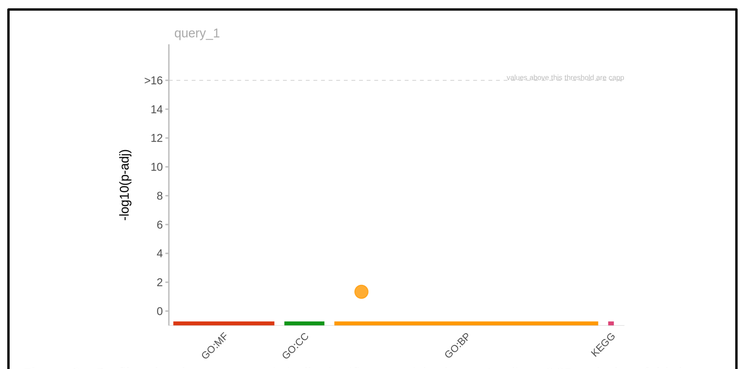

Figure 6. gProfiler plot shows one statistically significant enriched term (padj <= 0.05 and a $\log_2$ fold change > 2 or < -2). The function of this one point is cell adhesion with a p-value of 0.046.

References

1. Peterková K, Konečný L, Macháček T et al. Winners vs. losers: Schistosoma mansoni intestinal and liver eggs exhibit striking differences in gene expression and immunogenicity. 2024.
2. Hudak et al. Open OnDemand: A web-based client portal for HPC centers. Journal of Open Source Software. 3(25), 622. 2018.
3. Saber S, Alomar S, Yahya G. Blocking prostanoid receptors switches on‧multiple immune responses and‧cascades of‧inflammatory signaling against‧larval stages in‧snail fever. Environmental Science and Pollution Research. 2022.
4. Andrew S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics. 2010.
5. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Stichting.
6. Dobin A, Davis C, Schlesinger F et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013.
7. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data." Bioinformatics. 2015.
8. Picard toolkit. Broad Institute. 2019.
9. Liao Y, Smyth G, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014.
10. Ewels P, Magnusson M, Lundin S et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 2016.
11. Love MI, Love MI, Huber W et al. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. BioMed Central.

Acknowledgements

The computational resources of the study were provided by the Blugold Center for High-Performance Computing under NSF grant CNS-1920220. Infected tissues was provided by the Schistosomiasis Resource Center of the Biomedical Research Institute (Rockville, MD) through NIH-NIAID Contract HHSN272201700014I.