Analysis of RNA-Sequencing Data Methods to Understand Quorum Sensing in *Candida albicans*

Edward Sukarto, Deepika Gunasekaran, Melanie Ikeh, PhD, Cen Jiang, and Clarissa J. Nobile, PhD
School of Natural Sciences, University of California, Merced

SUMMER UNDERGRADUATE RESEARCH FELLOWSHIP

Abstract

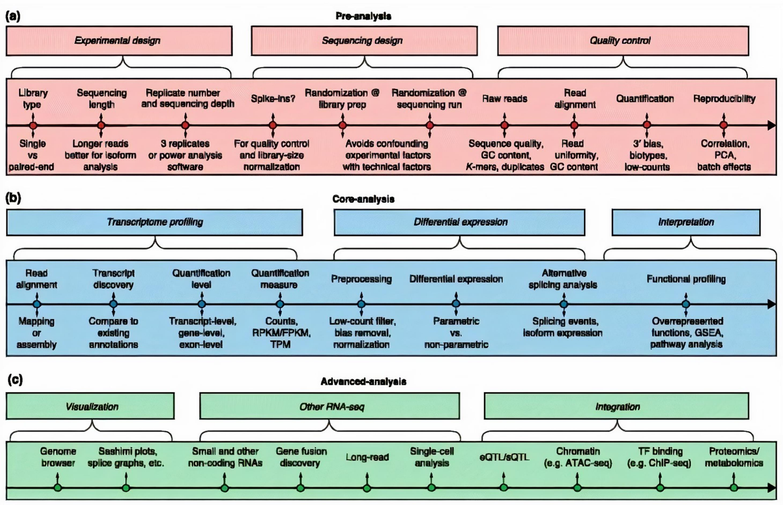
RNA-Sequencing (RNA-Seq) is a technique used to determine the transcriptional profile of cells under specific conditions. Different methods exist to analyze RNA-Seq data, especially in terms of determining differentially expressed genes in the conditions of interest. We have found that RNA-Seq data using different RNA-Seq analysis methods has generated very different results in terms of the differentially expressed genes identified. The goal of my project is to understand the limitations of these different methods in identifying differentially expressed genes. We are comparing two widely used methods, DESeq2 and Limma, to determine the extent of the differences observed between the differentially expressed genes identified using the two tools. We are looking at differences in Log2 fold change values, adjusted P values, and CPM values, to understand the sensitivity and precision of the different tools. Our results are revealing the utility of these tools and identifying the best-practices for analyzing RNA-Seq data in the field. These findings are being applied to existing RNA-Seq data for *Candida albicans* in the presence of the quorum sensing compound, farnesol. Analysis of this RNA-seq data will lead to the identification of the transcriptional profile of *C. albicans* during exposure to farnesol. This work will ultimately shed new light on how cells of this important human fungal pathogen communicate with one another.

Objectives

- Compare counts per million (CPM) values, log2FoldChanges, and adjusted p-values between differentially expressed genes identified by the 2 most widely used tools, Limma and DESeq2.
- Employ a functional enrichment analysis on RNA-Seq data to determine the processes affected by the presence of Farnesol.

Background

1.



RNA-Seq is a technique that measures gene expression profile under different conditions. RNA-Seq computational analyses are outlined above and it should be noted that there are no standardized methods that can be used for a comprehensive functional understanding of the results. The over-arching goal of this project is to put together the best practices and understand the limitations of RNA-Seq analysis.

Results

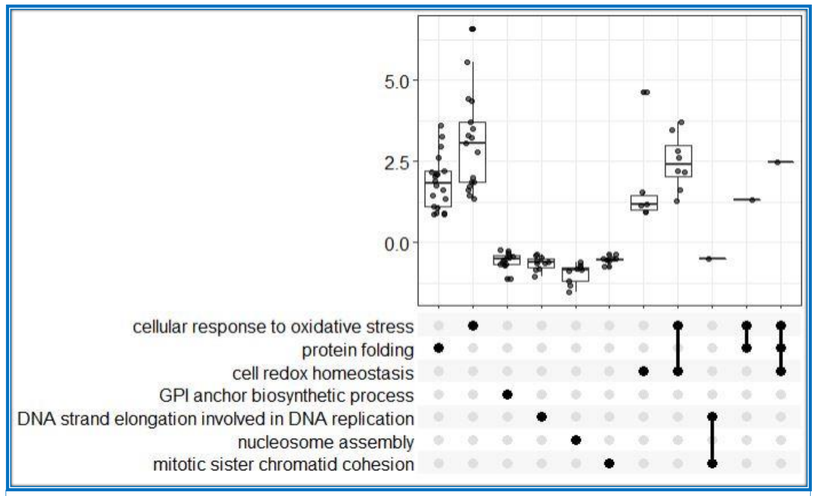


Figure 1. UpSet plot of enriched GO Terms comparing log2 fold changes of genes and the biological processes in the presence of Farnesol.
Each point represents the log2 fold change of each gene in a gene set. The greatest valued points are the most upregulated genes that are enriched the given biological process(es).

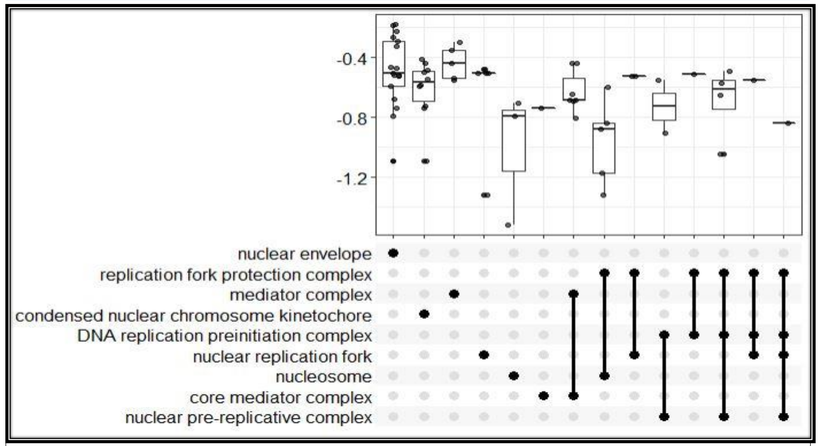


Figure 2. UpSet plot enriched GO Terms comparing log2 fold changes of genes and the cellular components in the presence of Farnesol.
Each point represents the log2 fold change of each gene in a gene set. The lowest valued points are the most downregulated gene that are enriched in the given cellular component(s).

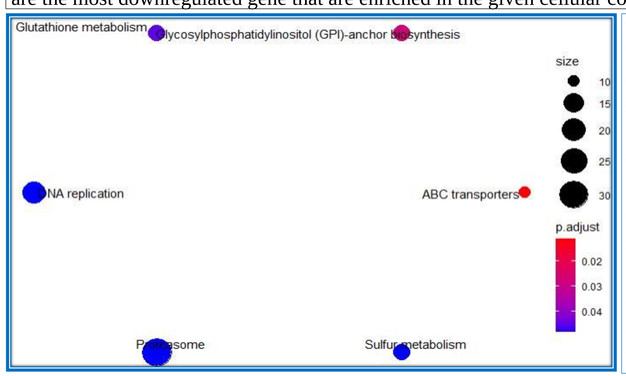


Figure 3. Enriched KEGG pathways of enzymatic network interactions.
Size represents amount of enriched genes in the given pathway in the presence of Farnesol. P-values represented by color hues indicate the significance of enrichment of the pathway.

Methods

- CPM differences between genes identified only by DESeq2 and Limma were compared using distribution plots and boxplots.

DESeq2

- Tool that identifies differentially expressed genes with the use of count data from RNA-seq. Uses shrinkage estimators for extreme values and fold changes to control dispersion of data. Identifies genes with high sensitivity and precision.

Functional Enrichment

- The two functional enrichment analyses that were conducted are gene set enrichment analysis (GSEA) and KEGG pathway enrichment.
- GSEA ranks log2 fold changes of genes of interests and creates an UpSet plot that shows the relationship between the log2 fold changes of genes and which biological processes and cellular components are enriched.
- KEGG pathway enrichment shows which metabolic pathways are being enriched and the number of genes involved in the presence of Farnesol.

Conclusion

- DESeq2 identifies a wider range of CPM differences between conditions compared to Limma, therefore it also identifies instances where the differences are low between the Farnesol treated and untreated conditions.
- DESeq2 identified 354 differentially expressed upregulated genes and 39 downregulated genes. Limma identified 48 differentially expressed upregulated genes and 6 downregulated genes.
- KEGG pathway enrichment suggests that six metabolic pathways are differentially regulated in the presence of Farnesol in *C.albicans.*
- GO Term and KEGG pathway enrichment concordantly indicate that proteasome complex and GPI-anchor biosynthesis are differentially regulated in the presence of Farnesol.

Future Work

- Perform a weighted gene correlation network analysis (WGCNA) on the remaining RNA-Seq data which will create a network by grouping similar gene expression profiles between all given samples. Then build modules where similarly expressed genes are correlated to phenotypes of interest to understand the network of genes roles in the specific phenotype.
- Utilize CRISPR/CAS9 to knockout genes of interest identified by DESeq2 as differentially expressed genes in the presence of Farnesol to further understand Quorum Sensing in *C.albicans.*

References

1. Conesa, A., Madrigal, P., Tarazona, S. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol* **17,** 13 (2016). https://doi.org/10.1186/s13059-016-0881-8

2. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Acknowledgements

I would like to acknowledge Gabriel Viramontes for insightful help on random questions I had and my childhood dog, Clovis. Thank you all!