Project Report - Trained Resistance to Infections

Group: Justin Hui, Arham Khan, Vishvak Seenichamy, Sophie Lesher

Github: https://github.com/Vishvak365/Bioinformatics-Trained-Resistance-to-Infections

# **I. Abstract**

Intergenerational inheritance of phenotypic characteristics has been demonstrated repeatedly in plants and invertebrates, including resistance to infections. The study of the intergenerational inheritance of infection resistance is crucial in our understanding of how immunity propagates in populations of vertebrates - including humans. Here we make use of Mus Musculus gene expression data collected from parents and offspring to observe inheritance of resistance to fungal infection. Because of the great diversity and complexity of the biological and genetic makeup of different species and infections, it is often difficult to make a determination as to whether immunity to infections can be passed down through generations. In this study that we chose to examine, the inheritance of resistance to a sublethal infection of Candida albicans from parent to offspring in Mus Musculus. The results of the study and our analyses show that inheritance does in fact pass down resistance to specific fungal infections such as Candida albicans. This finding may further advance the discussion for understanding basic infectious disease behavior, creating more efficient epigenetic immune responses, comparing cell and gene functions among different species, and predicting the spread of novel infections.

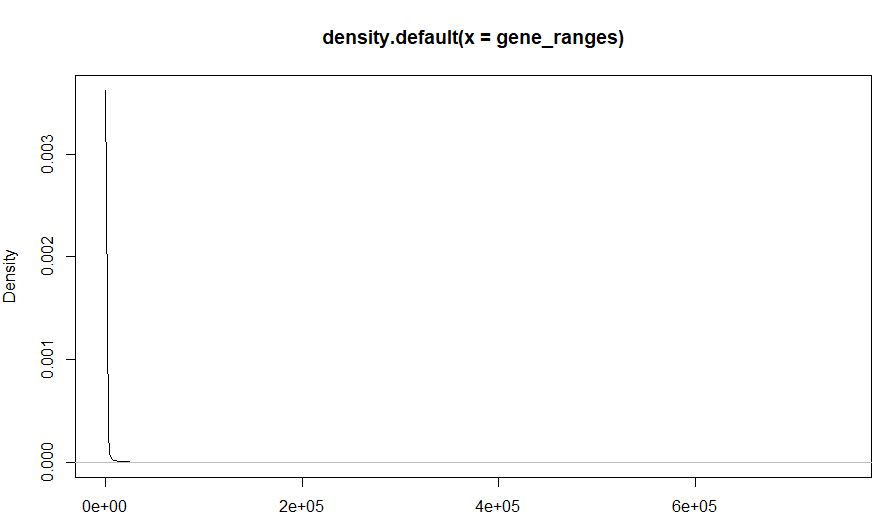
# **II. Introduction**

From the NCBI online database, we used the publicly available data gathered from the Myeloid Cell Biology lab in Germany to perform our analyses. Here, mice were infected with a sublethal systemic Candida albicans infection and immune responsiveness was measured. The question we were trying to answer from this data was whether or not immunity to a fungal infection in male mice gets passed down to their offspring. In a broader sense, we wanted to observe whether infection immunity could be genetically inherited in the context of work such as [1-4] where other intergenerationally transmitted characteristics are studied. Previous work has touched upon the effects of genetic inheritance upon physical characteristics such as olfactory calibration and physiological resistance to the environment, but we focus mainly on the intergenerational dynamics of infection [5-7]. It has been shown that genetic factors can lead to immune priming in various species, including by analyzing specific patterns in paternal and maternal lineage [8-11]. Further work has focused on the study of internal and environmental factors in infection resistance, here we focus purely on genetic factors [12,13]. We observed the infected and control groups alongside their offspring to measure the degree to which infection resistance propagates to the mice’s progeny. We utilized RNA-seq gene expression data from twenty-six samples to perform enrichment analysis and measure differential gene expression between groups. We utilized principal component analysis and t-SNE projection in order to perform data analysis and we enriched our data using various gene and disease ontology databases. We used unsupervised techniques to produce our final clustering for analysis, blending insights from KMeans, Hierarchical Clustering, Consensus Clustering, and Gaussian Mixture models. Ultimately we conclude that infection resistance to Candida Albicans was transmitted intergenerationally. This has implications in broader areas of research concerning pandemic control, inoculation propagation, and in treating genetic autoimmune disorders.

# **III. Methods**

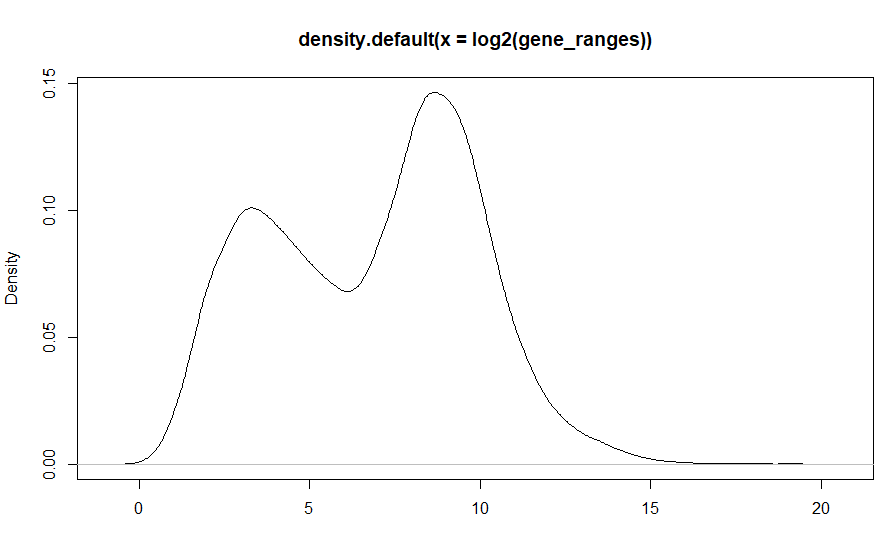
# Per Gene Expression Ranges:

Density plot



*FIGURE 1*

Density Plot log2 Scaling

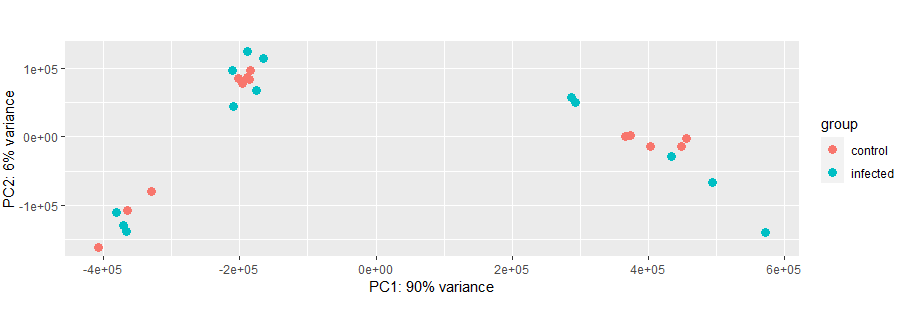


*FIGURE 2*

The graph shows that most of the expression ranges per gene fall between 2^0 and 2^15 with peaks at 2^4 and 2^9.

* Similarity Between 2 Groups:

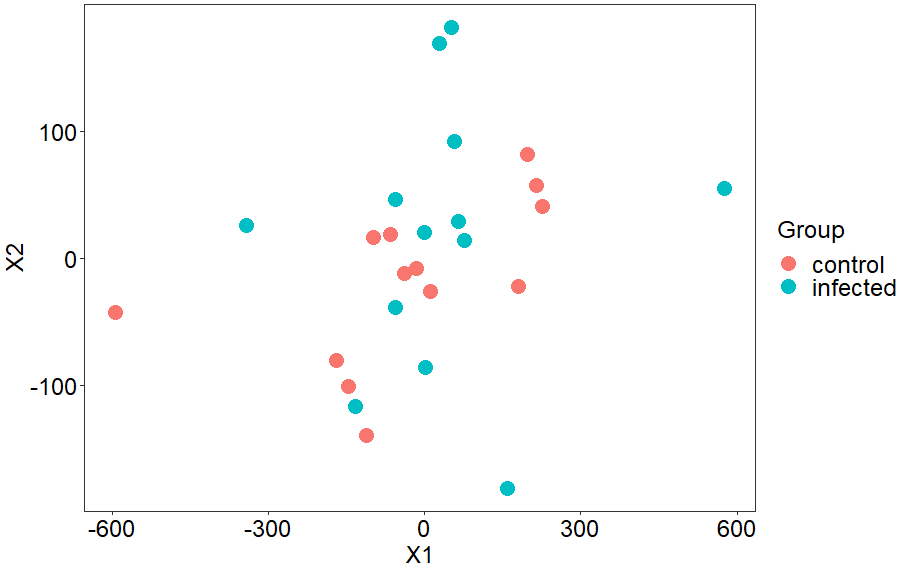
PCA Plot



*FIGURE 3 - PCA plot with two components*

PCA generates two feature dimensions where there is a somewhat-strong clustering of infected and control samples. Interestingly, there are visually distinct clusters of both control and infected samples, though within those clusters, items of the same class tend to be plotted closely together.

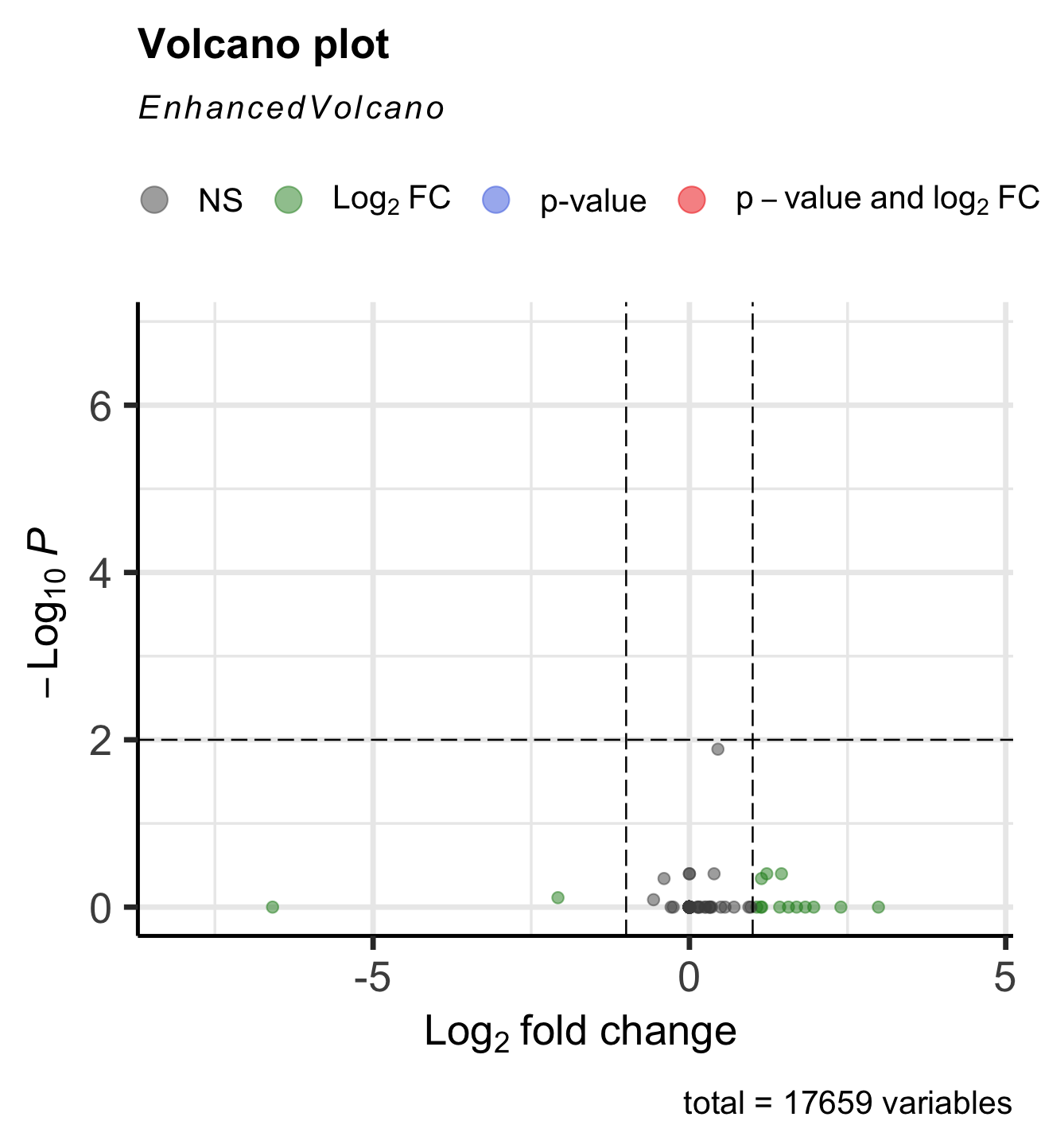
Tsne Plot



*FIGURE 4*

The t-SNE plot also reduced the data to two feature dimensions. The plot is less organized than that of PCA, with fewer distinct clusters. Even though infected samples still do tend to cluster close to one another, the separation is not as strong as shown in PCA.

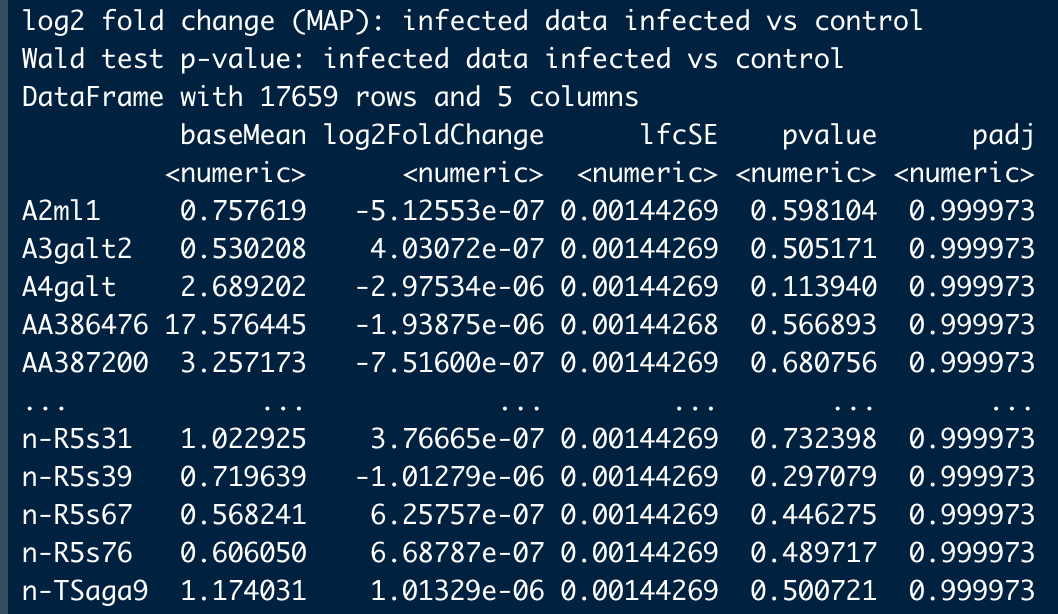
* Differential Analysis:



*FIGURE 5*

Based on the dataset that we chose and the results of the volcano plot, we had no data point that was significantly significant. However, based on a discussion with the Professor, we have concluded that this is acceptable in our case.

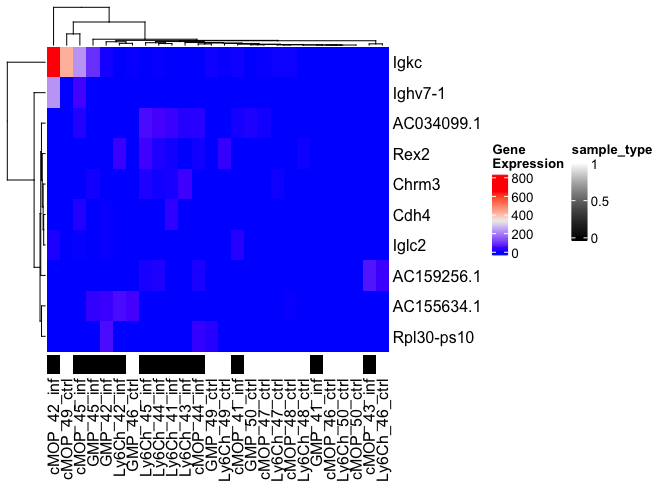
Differential Expression Analysis:

****

*FIGURE 6*

* Significantly Expressed genes

Heatmap

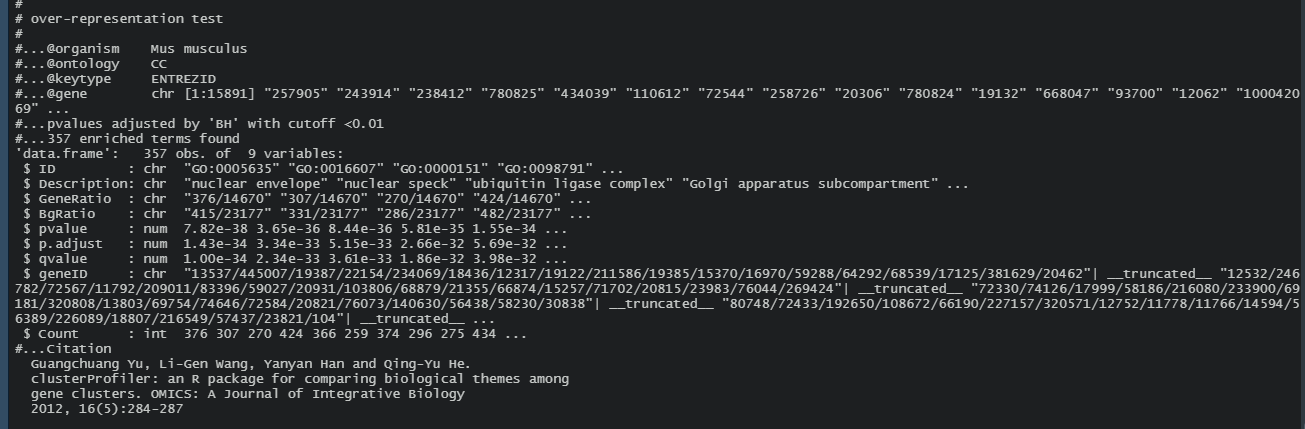


*FIGURE 7*

Since this volcano plot and differentially expressed gene matrix only had 1 value out of 17000+ that was above the threshold, the heatmap would have just been one solid color and not very useful. To display a more meaningful graph, we provided a table with the dataset containing the single significant gene and the next few top significant genes and their values. The heatmap shown above summarizes the top 10 significantly differentially expressed genes from the samples. The black-colored samples represent the infected group and white-colored for the control group.

* Enrichment analysis

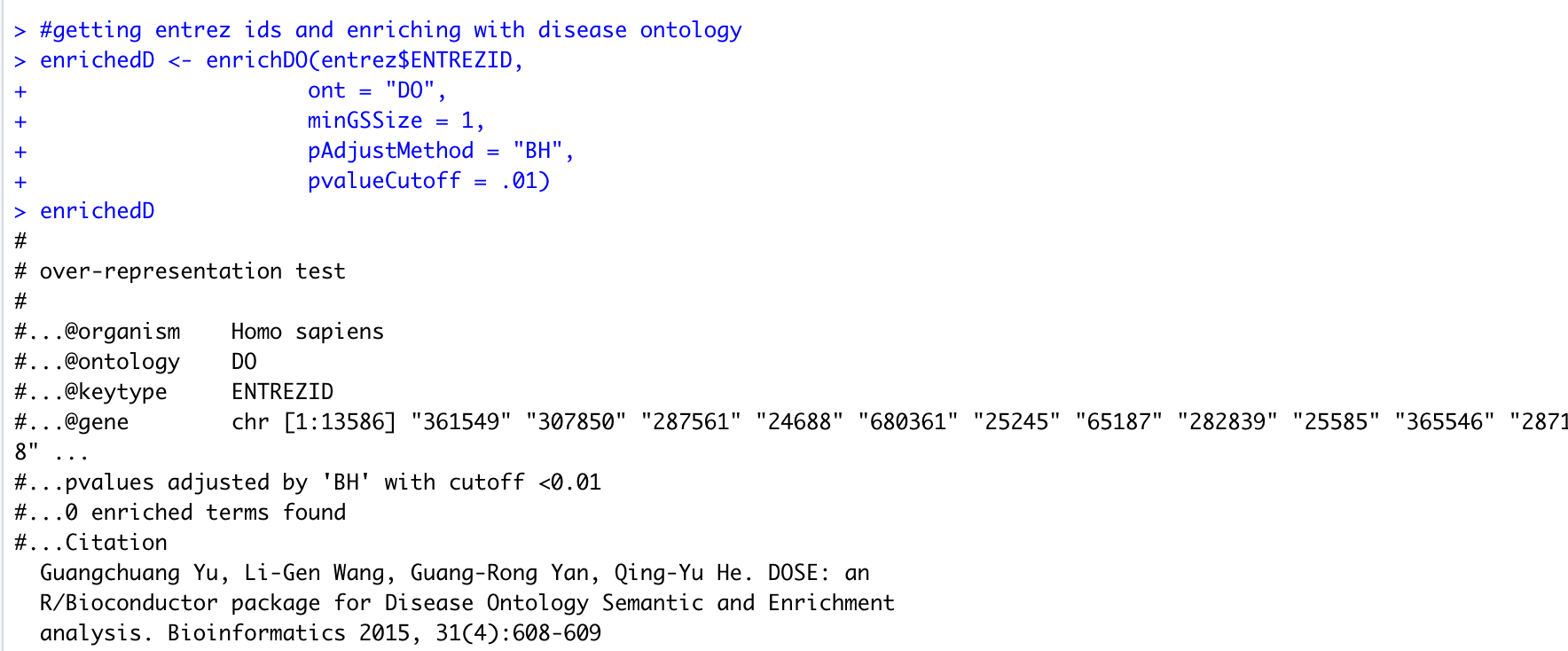
clusterProfiler (Gene Ontology)



*FIGURE 8*

Through using cluster profiler and gene ontology, we found 357 enriched observations.

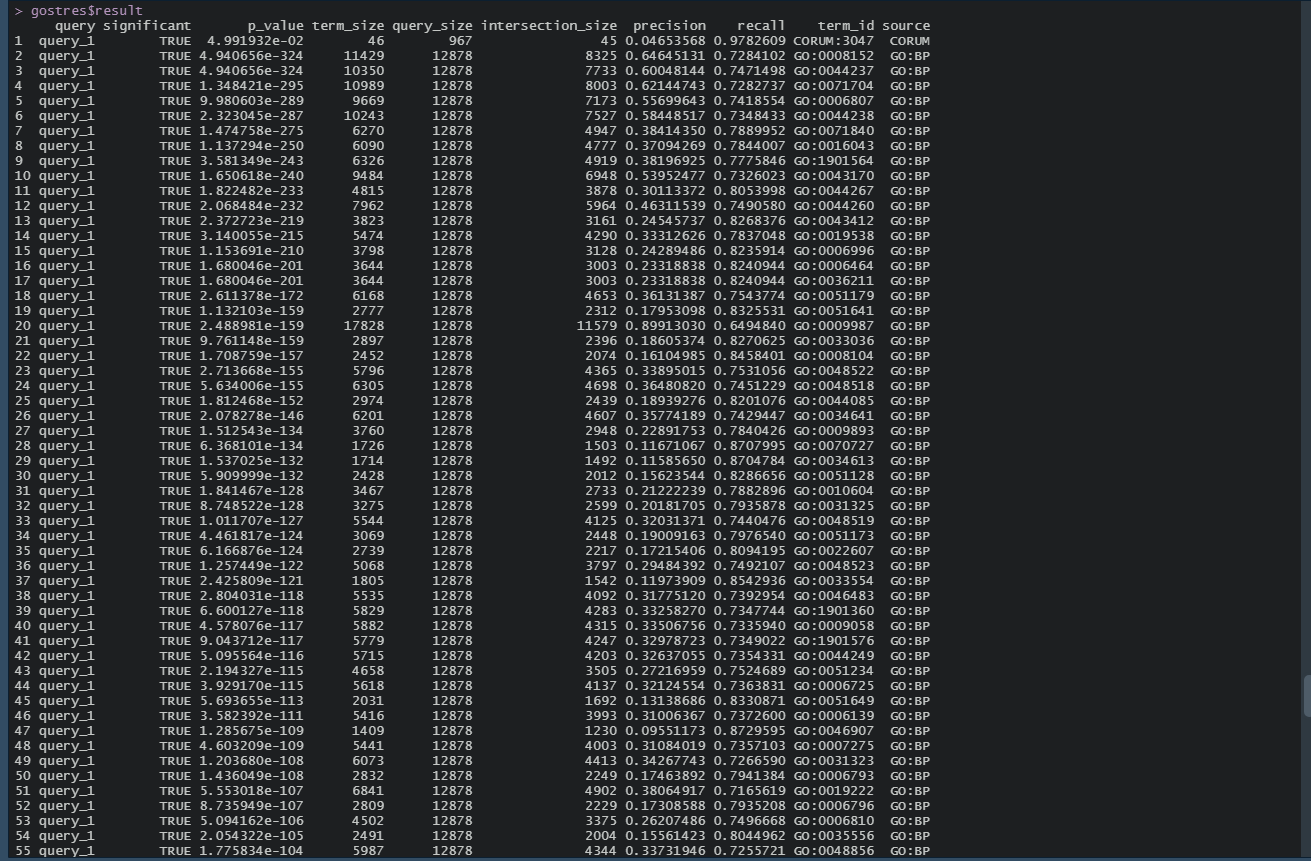
clusterProfiler (Disease Ontology)



*FIGURE 9*

According to this data, there are no found diseases found from our samples in the clusterProfiler DOSE library. Even when the p-value is increased from .01 to .1 and our minimum value is set to 1, there are still 0 enriched terms found. This may be because our dataset is somewhat unusual and/or underpowered. Therefore, there is no table of values that can be generated for the enriched processes found for disease ontology based on p-value.

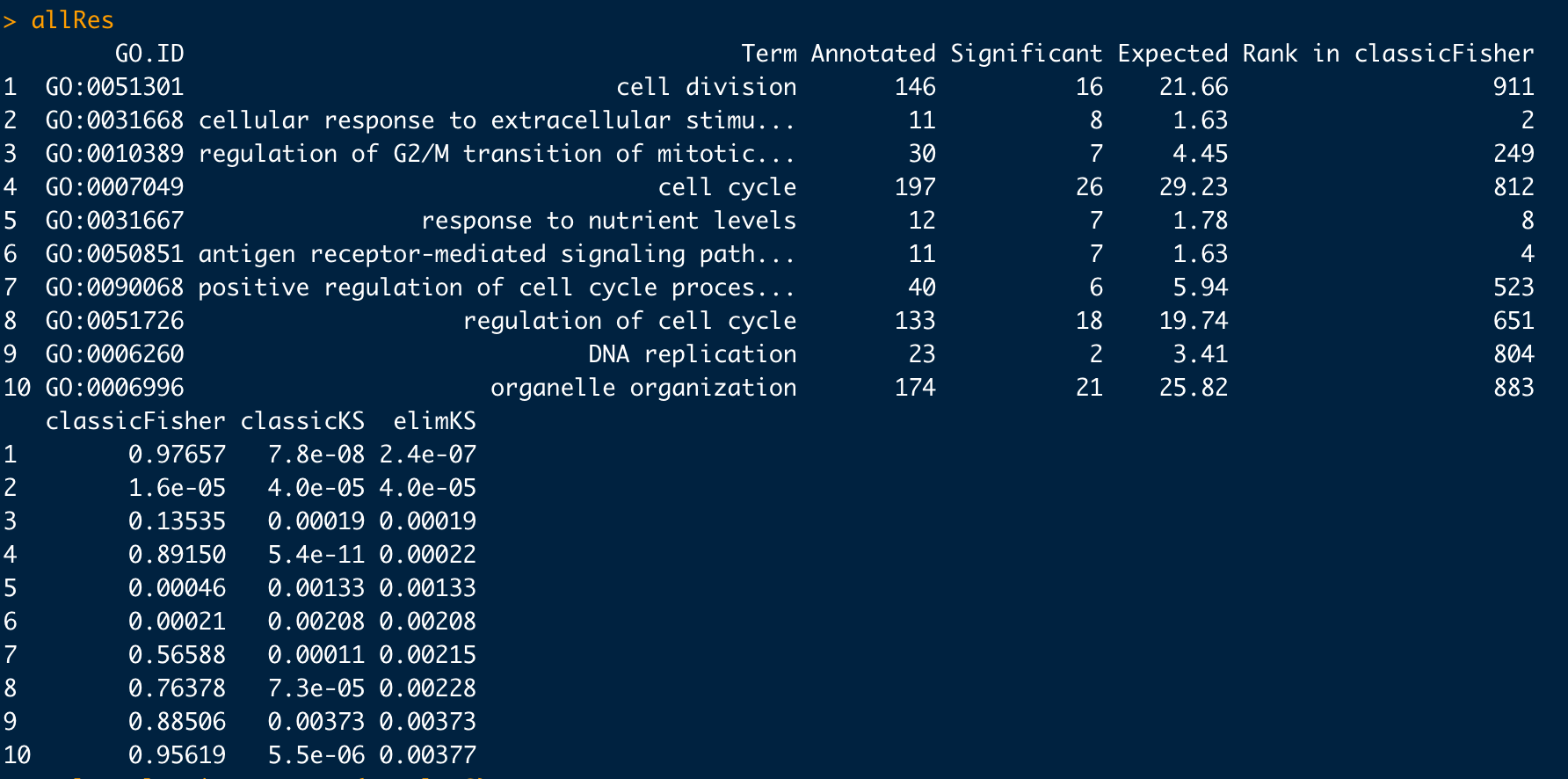
gProfiler2 (gene ontology)



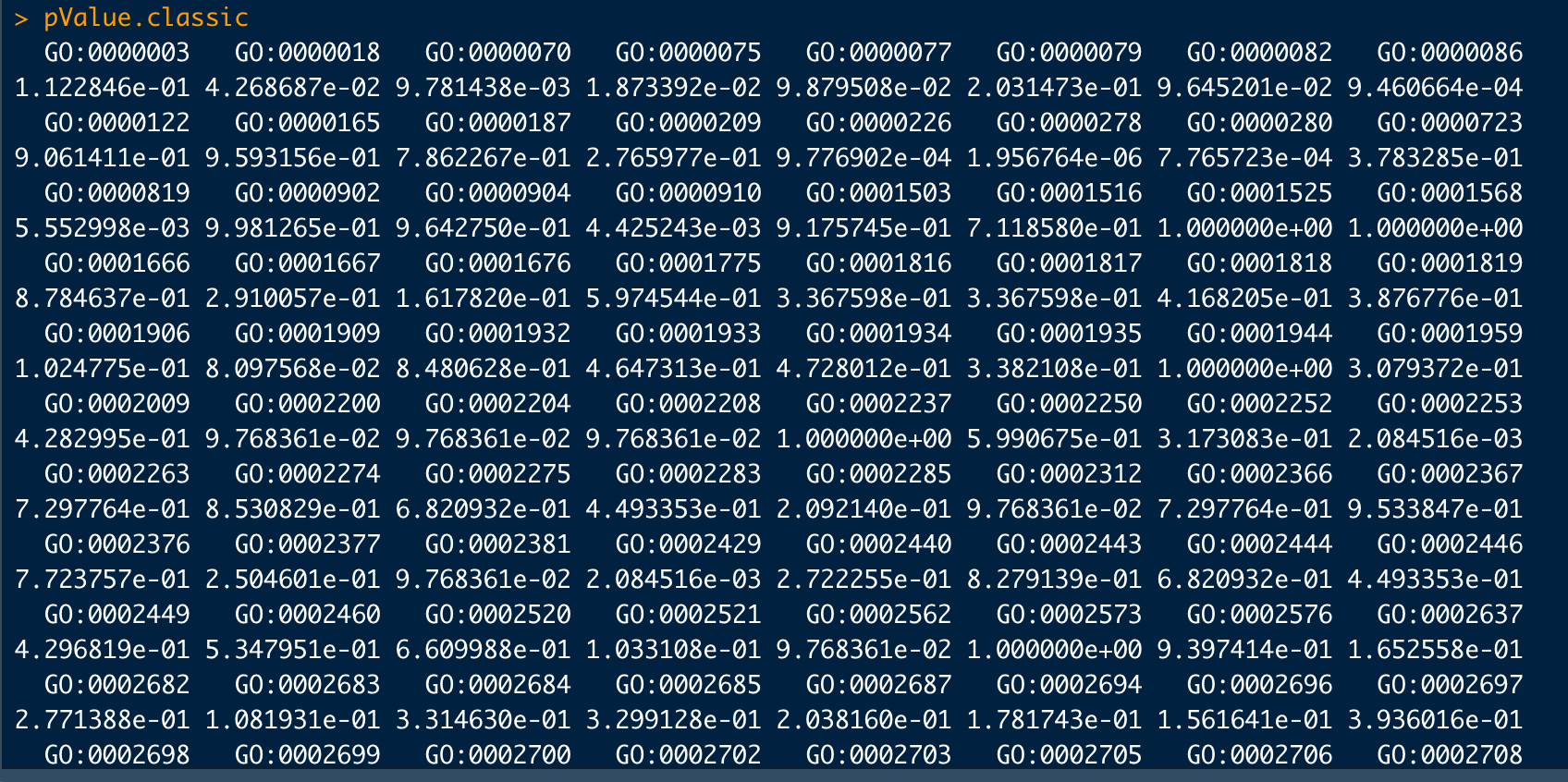
*FIGURE 10*

We utilized gProfiler2 to perform gene ontology analysis. We found many samples that presented as significant when using a p-value of 0.05 - the default in gProfiler2 - using a p-value of 0.01 would have returned similar results as per the table.

topGO (Gene Ontology)



*FIGURE 11*

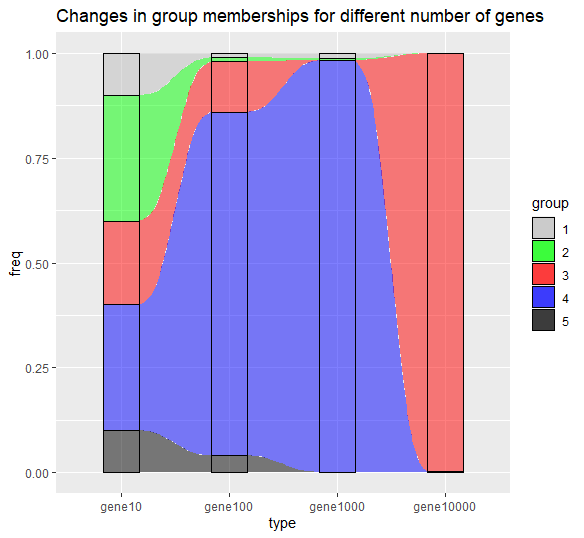


*FIGURE 12*

Instead of using GenomicSuperSequencer, as the documentation was lacking in a few areas, we decided to use the topGO method for Gene ontology for our additional method/ontology. The results appear to match the enrichment analysis of the other gene ontology methods.

* Clustering algorithms

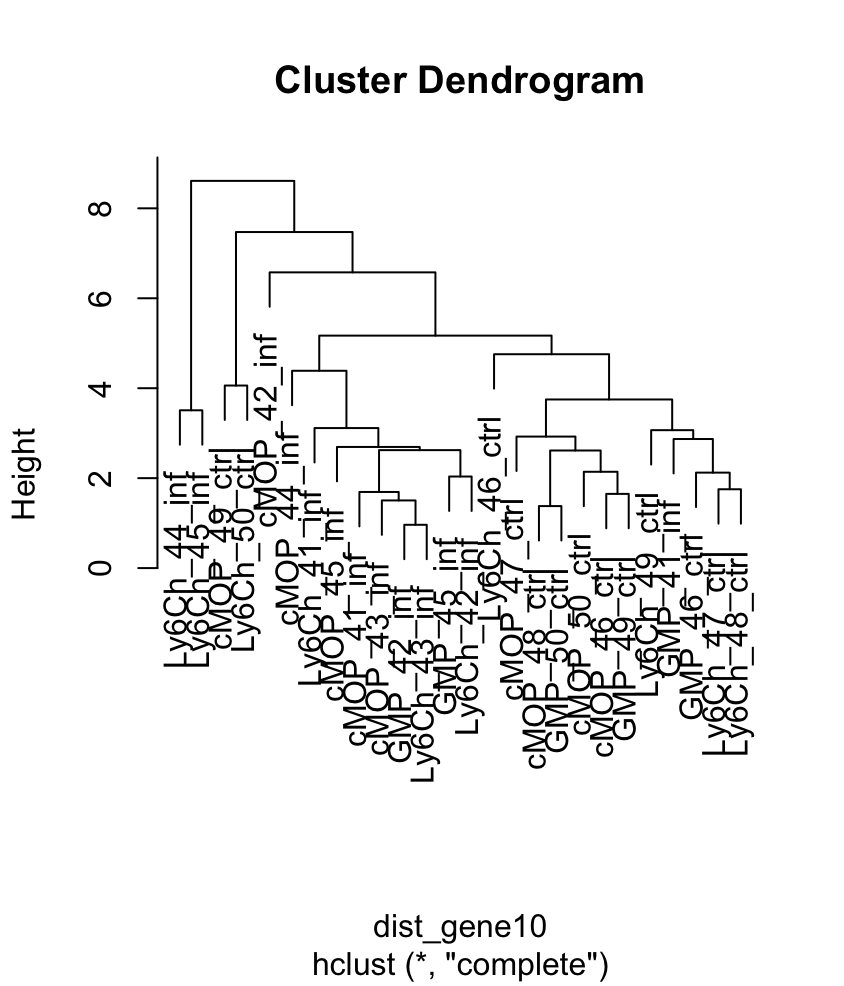
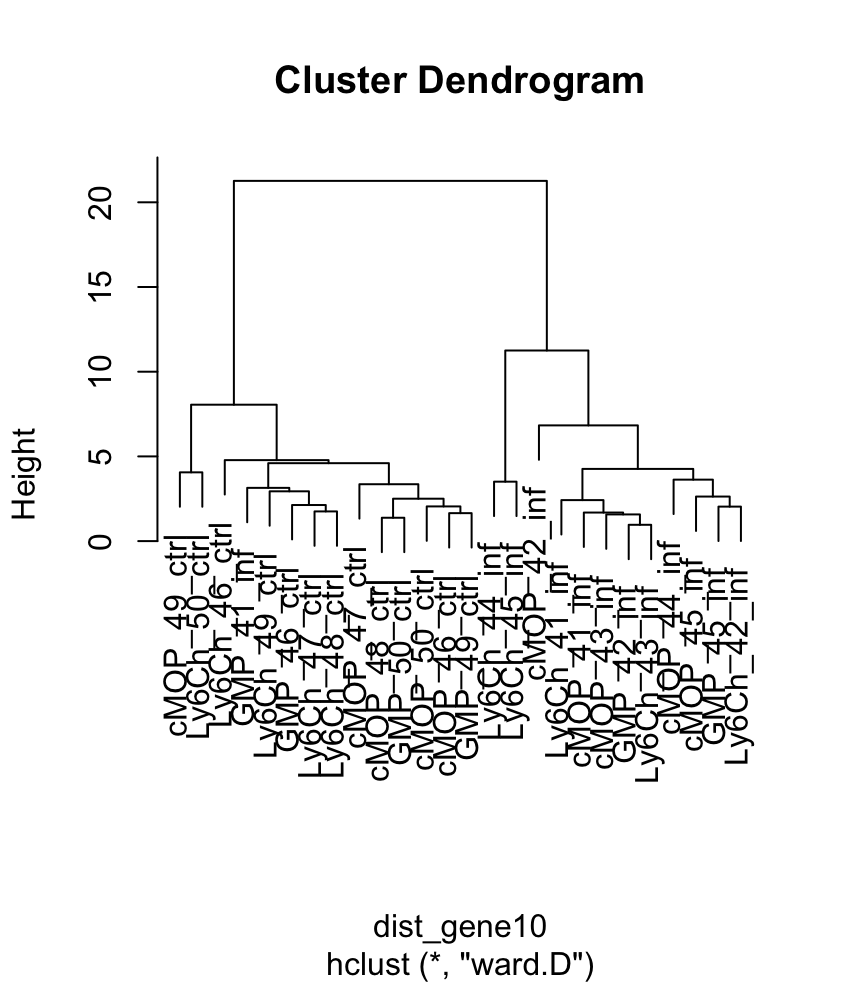
K Means

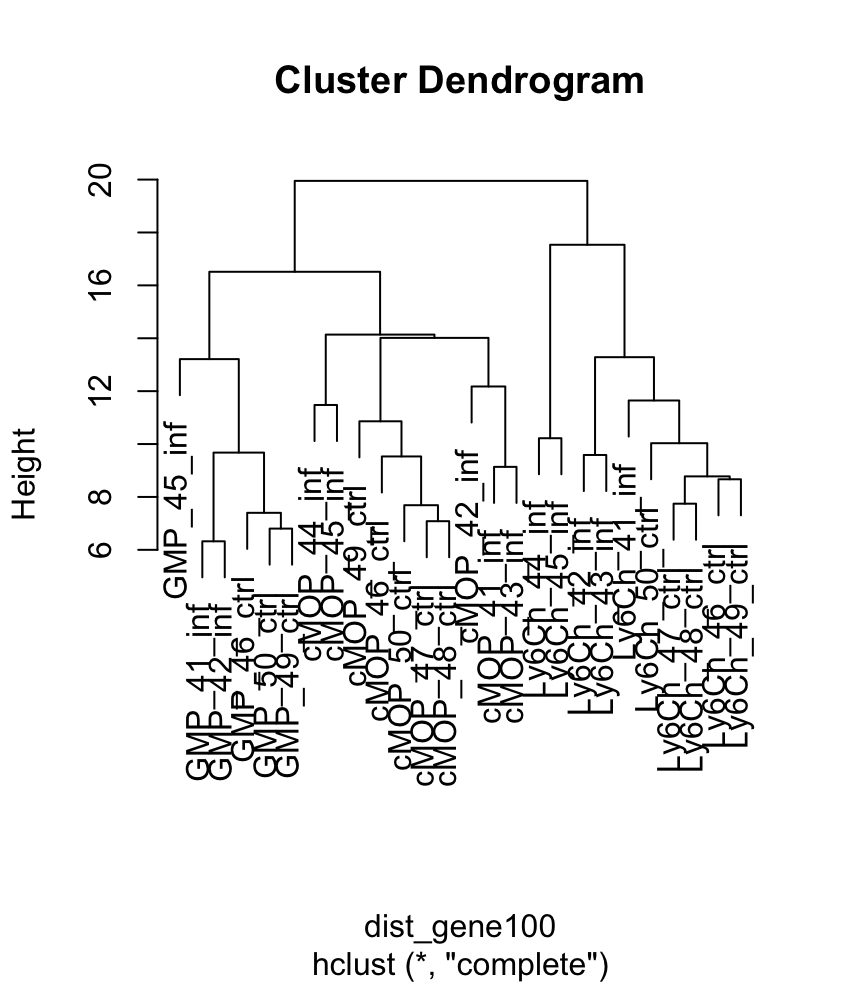
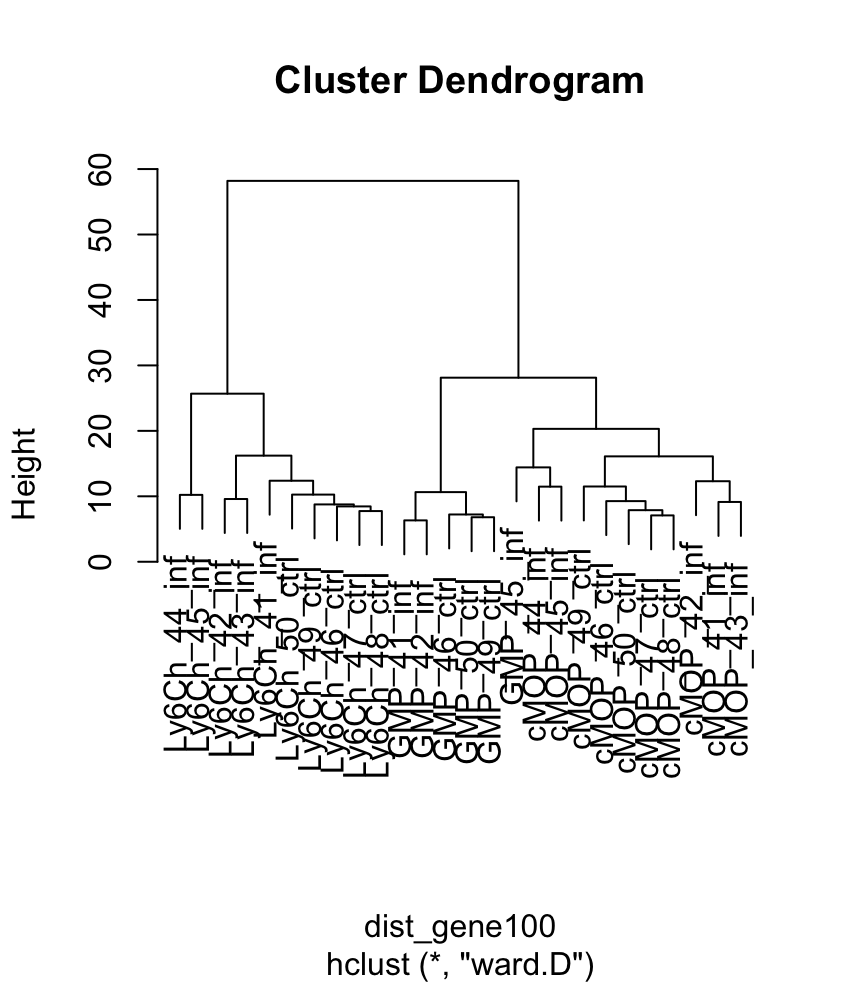


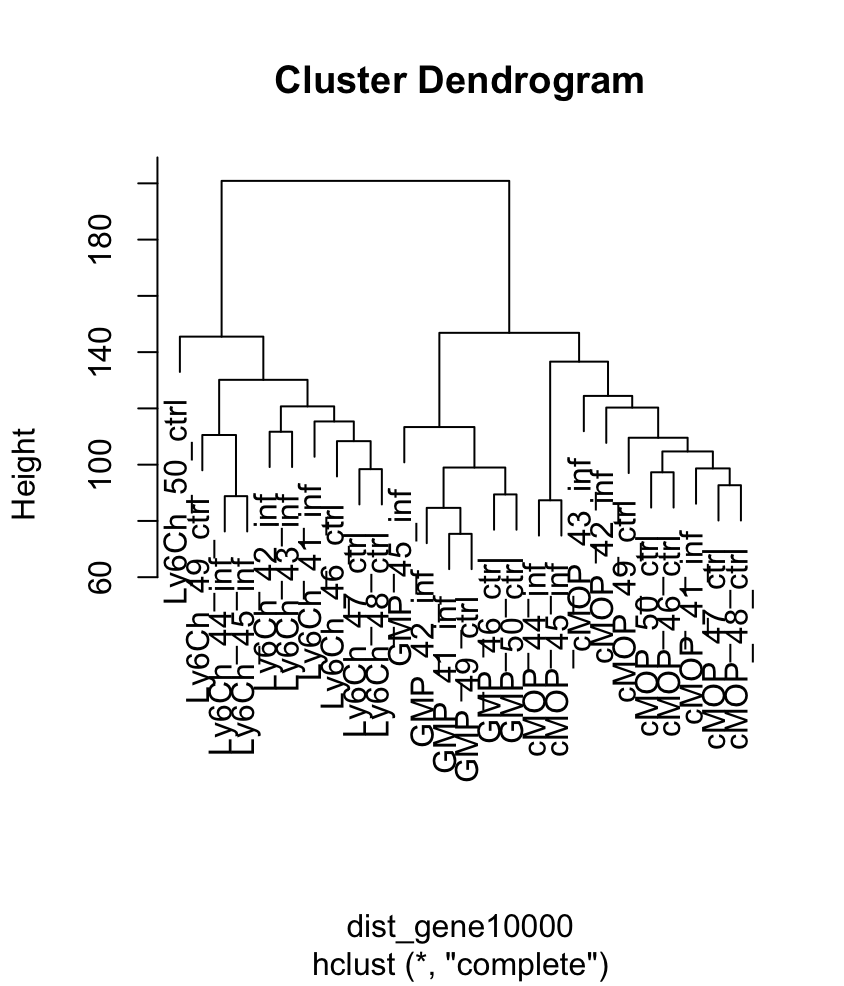
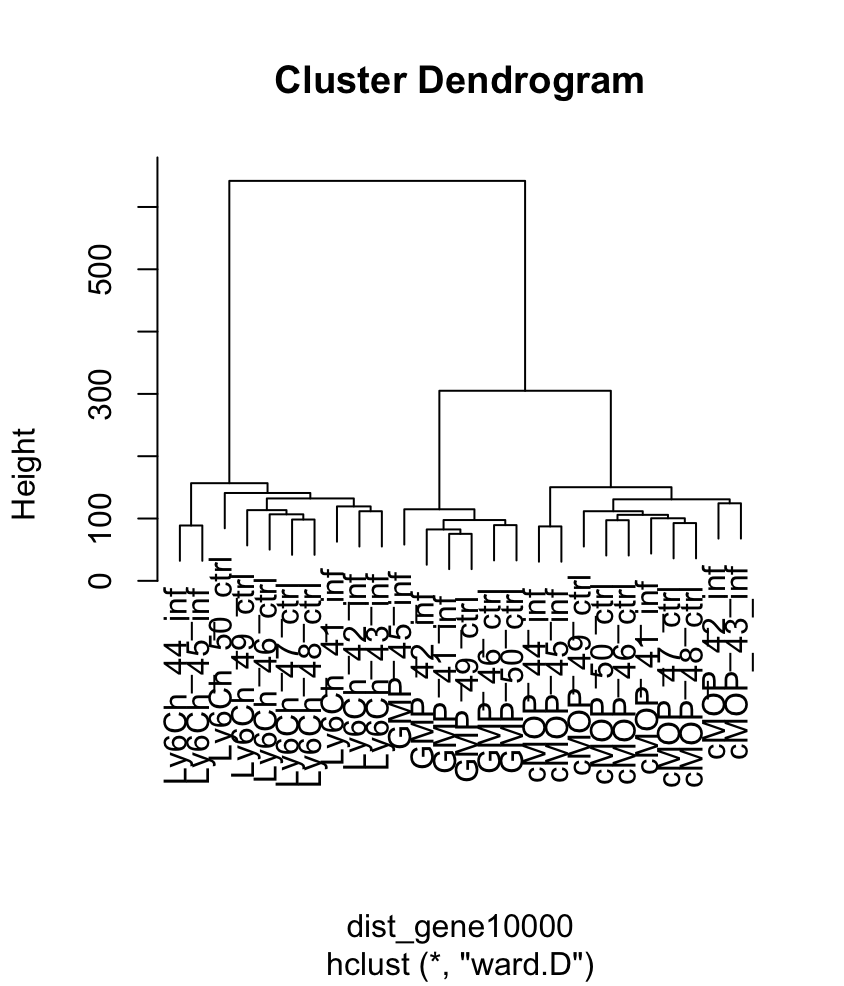
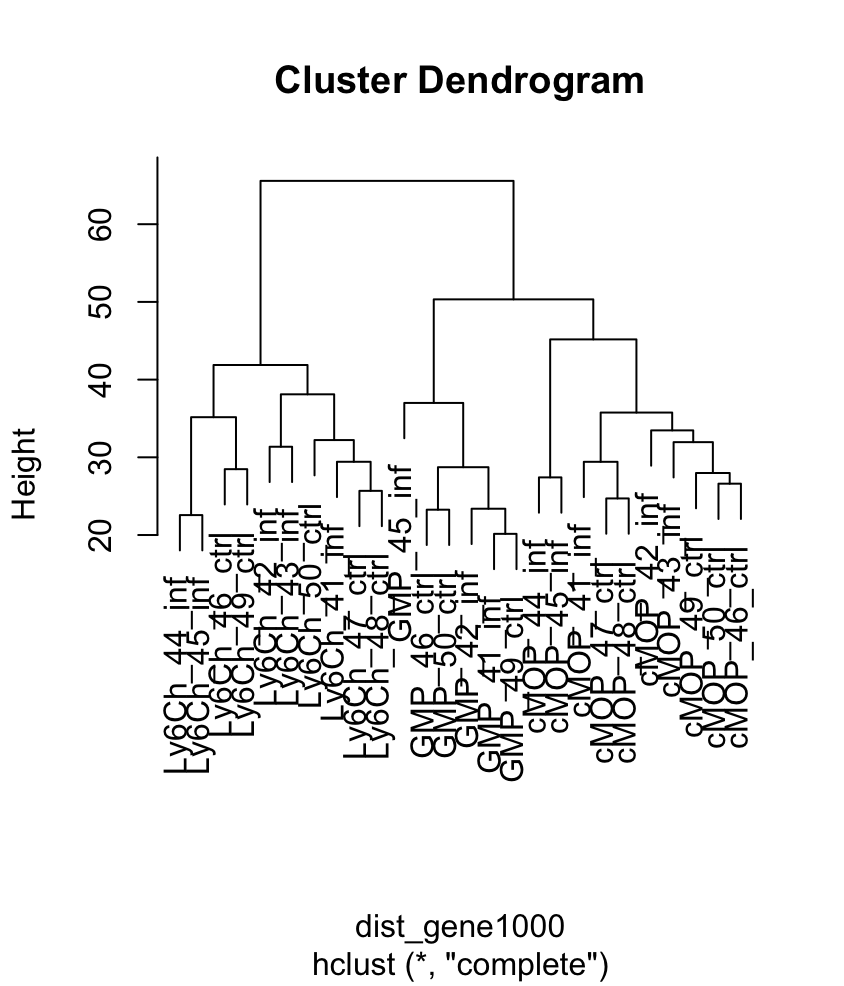
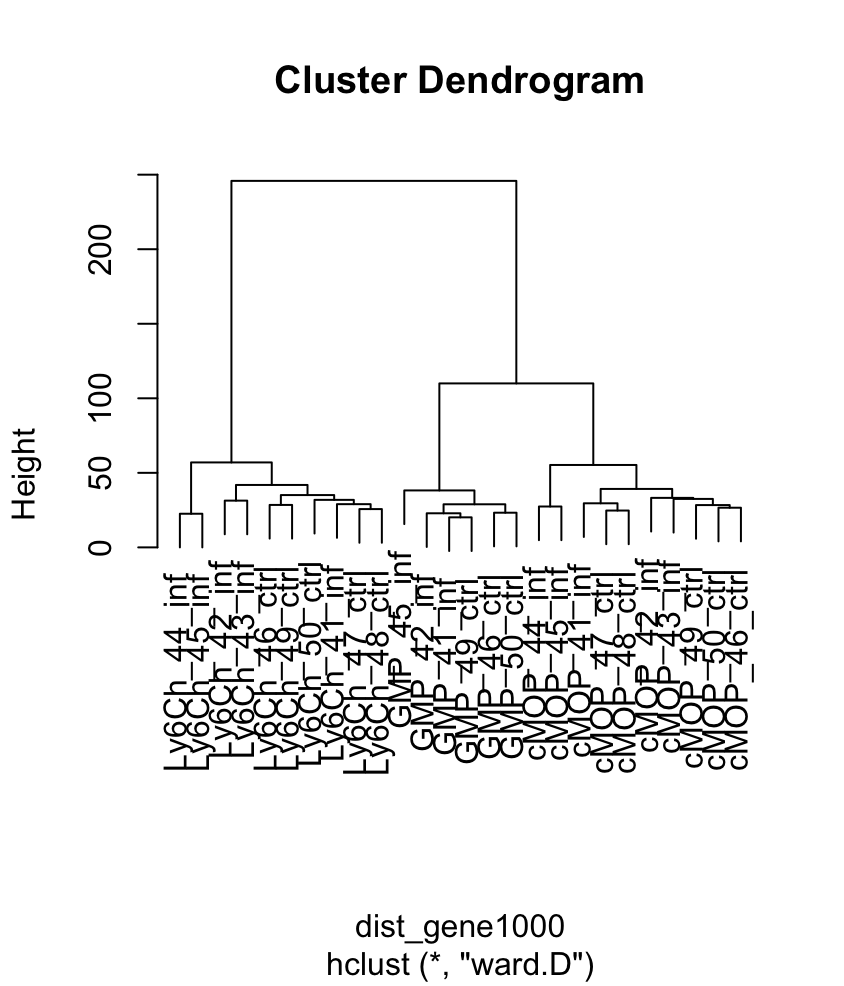
*FIGURE 13*

This alluvial graph shows how group membership and the centers of the groups can change when using more genes. Using a lower number of genes shows that the distribution of the genes among the groups are around even but using more genes seems to increase the membership of a single group.

Hierarchical Clustering

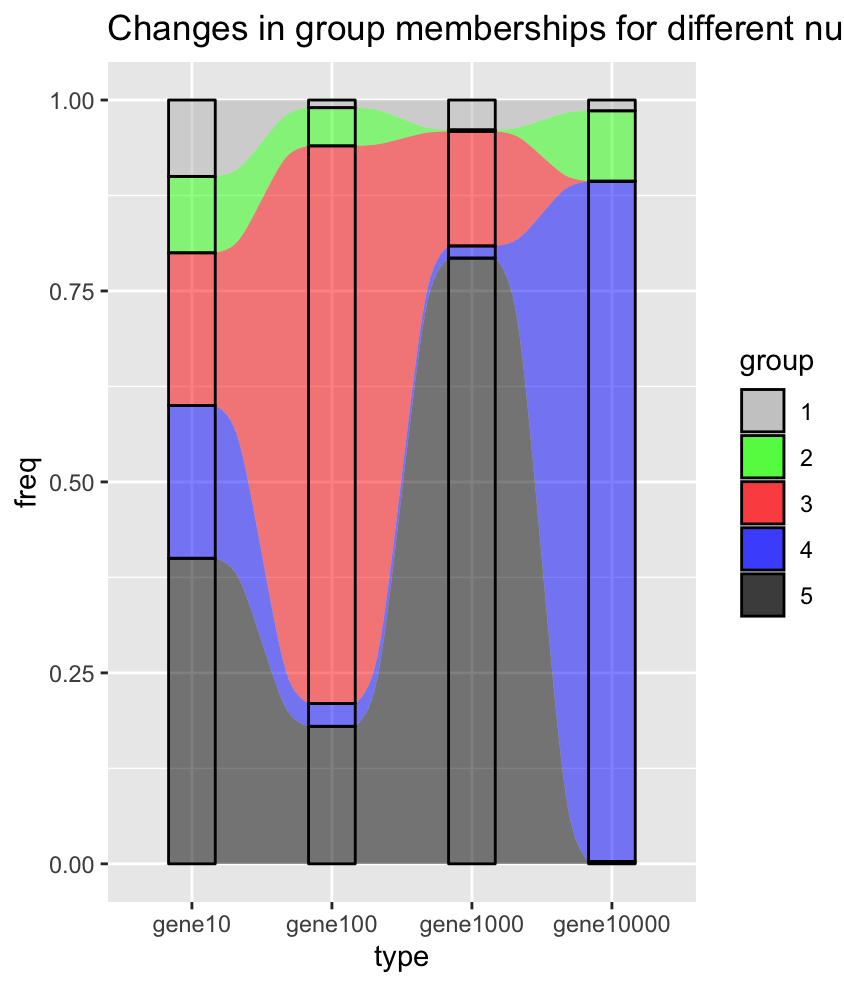






*FIGURE 14*

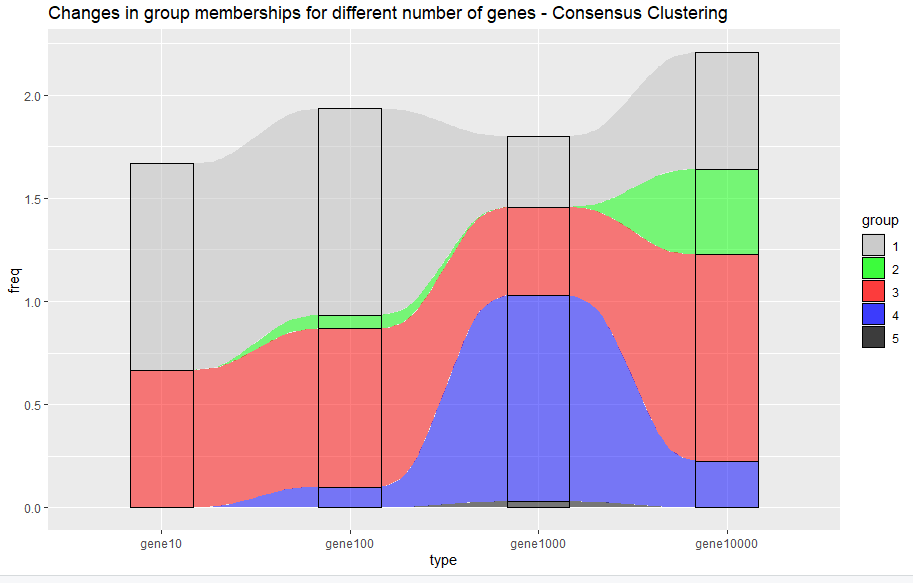
Hierarchical clustering performs an analysis on a set of dissimilarities using a distance function, using the function “dist”. Here, the different hclust methods “ward.D” and “complete”. We can observe that the two hclust methods used vary slightly as the algorithm calculated the clusters differently. Testing different numbers of genes, we can see that increasing the value from 10 to 100, 1000 and 10000 also increases the amount and precision of clusters.



*FIGURE 15*

The sankey plot above compares the groups at different amounts of genes for the “ward.D” method. The variation between the different amounts of genes follows the trend showing that cluster size grows as the amount of genes increases. This makes sense because replication and more data will yield more precise results.

Consensus clustering

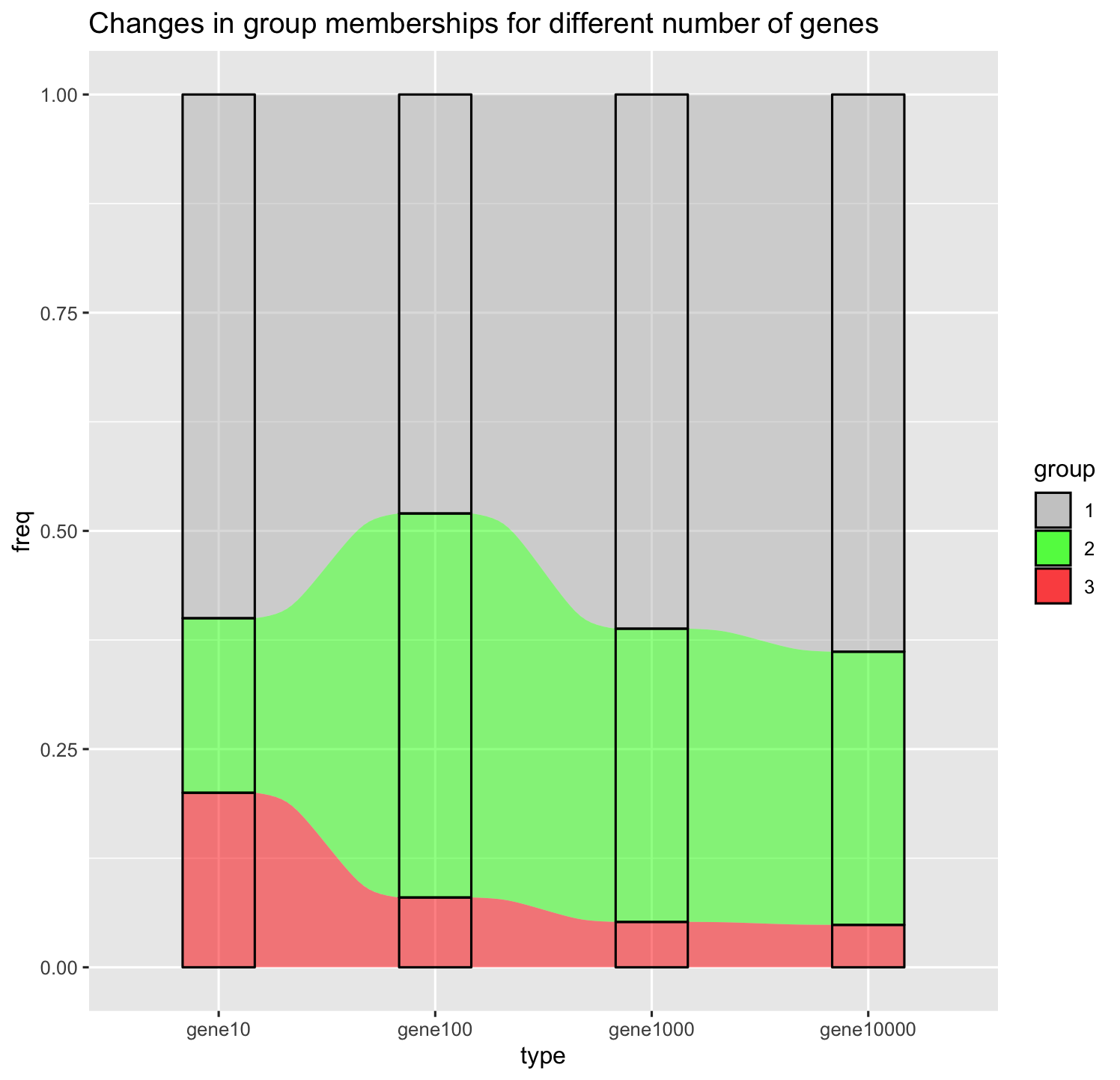


*FIGURE 16*

Cluster membership shifts slightly as we increase the number of genes. Notably, previously underrepresented clusters gain more representation as we increase the number of genes that we consider. This is expected as we are increasing the number of samples we consider, thereby increasing the probability that less frequent expression features will be observed.

Gaussian mixture

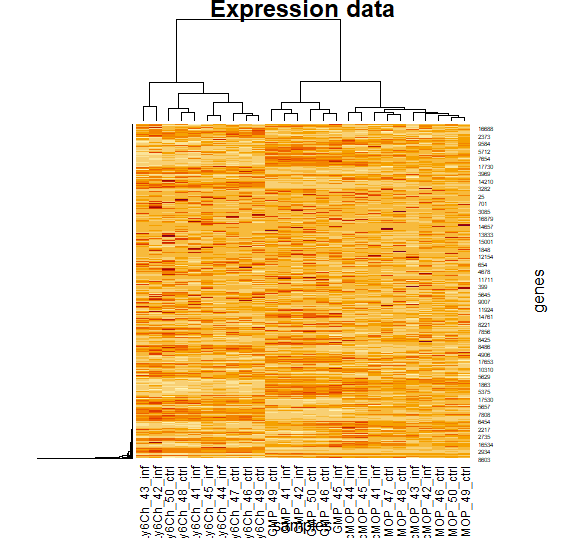
(cont. on next page)



*FIGURE 17*

This is the Sankey/Alluvial Plot for the varying number of genes for GMM and as mentioned above, lies within the expectations of the model. We are not going to see massive shifts in cluster memberships as we increase sample size.

* Heatmap of Genes Used in Clustering



*FIGURE 18*

The heatmap shows the relationship between our different groups and their gene expression. The darker shading indicates a higher expression value.

# **IV. Results**

Our original question was determining whether immunity to a fungal infection in male mice gets passed down to their offspring. We were able to answer this question with the following supporting arguments following the data we observed.

* Based on our Principal Component Analysis (PCA) visualization, we identified regions where cluster membership skewed towards the infected class. Similarly, our t-SNE plot displays regions where infected samples cluster together.
* Our Gaussian Mixture Model computed consistent clusterings, placing samples into three distinct groups. The clustering membership stayed similar when adding more genes.
* Additionally, hierarchical clustering generated cluster dendrograms (Figure 14), which place infected parents and children close together. Similarly, control parents and children are clustered closely.

Unexpected Findings

* Some of our clustering methods reflected odd results in the various group memberships. For example, consensus clustering had several groups in which the membership was 0% or near 0%, which could mean one of several things, but all of them did not aid in drawing a conclusion on the data.
* Another strange observation was that in our Volcano plot, none of the data points displayed differential expression that was statistically significant.

Future work would address this question with a larger number of generations to observe how resistance propagates. Additionally, this work could be extended to study viral and bacterial infections. There are subtle bioethical concerns we have in infecting lab mice, however, we attempted to minimize these concerns by utilizing only sublethal doses of our infection.

# **V. Conclusion**

Originally, based on the data that we chose, we asked the following question: does immunity to a fungal infection in male mice get passed down to their offspring? Our team, building off of research, hypothesized that immunity to fungal infection gets passed down to offspring. The information we presented in our statistical analyses supports this claim. Our various cluster and gene ontology methods were able to distinguish two separate groups, control and infected, at a significant threshold. Our predictive methodology found similarities between generations from both sample groups, suggesting a correlation between infection rate and genetics. Furthermore, the distinct clusters formed from the Principal Component Analysis, Gaussian Mixture Model, and hierarchical clustering generated cluster dendrograms indicate that infection rates were more easily predictable among the different generations of mice and that immunity is recognizably hereditary. Based on the various Bioinformatics methods, Machine Learning Methods, and Statistical Methods, our team is able to conclude that immunity does, in fact, get passed down to offsprings. We arrived at this conclusion through the results presented above as well as other analytical methods, which on their own neither prove nor disprove our hypothesis, but altogether combined give us confidence in supporting our conclusion.

**References**

1. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* 401, 157–161 (1999).
2. Luna, E., Bruce, T. J. A., Roberts, M. R., Flors, V. & Ton, J. Next-generation systemic acquired resistance. *Plant Physiol.* 158, 844–853 (2012).
3. Belicard, T., Jareosettasin, P. & Sarkies, P. The piRNA pathway responds to environmental signals to establish intergenerational adaptation to stress. *BMC Biol.* 16, 103 (2018).
4. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23, 314–318 (1999).
5. Dias, B. G. & Ressler, K. J. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat. Neurosci.* 17, 89–96 (2014).
6. Rowe, A. H. & Rowe, M. P. Physiological resistance of grasshopper mice (*Onychomys* spp.) to Arizona bark scorpion (*Centruroides exilicauda*) venom. *Toxicon* 52, 597–605 (2008).
7. Cramer, S. D., Ferree, P. M., Lin, K., Milliner, D. S. & Holmes, R. P. The gene encoding hydroxypyruvate reductase (GRHPR) is mutated in patients with primary hyperoxaluria type II. *Hum. Mol. Genet.* 8, 2063–2069 (1999).
8. Eggert, H., Kurtz, J. & Diddens-de Buhr, M. F. Different effects of paternal trans-generational immune priming on survival and immunity in step and genetic offspring. *Proc. Biol. Sci.* 281, 20142089 (2014).
9. Hernández López, J., Schuehly, W., Crailsheim, K. & Riessberger-Gallé, U. Trans-generational immune priming in honeybees. *Proc. Biol. Sci.* 281, 20140454 (2014).
10. Tidbury, H. J., Pedersen, A. B. & Boots, M. Within and transgenerational immune priming in an insect to a DNA virus. *Proc. Biol. Sci.* 278, 871–876 (2011).
11. Yue, F. et al. Maternal transfer of immunity in scallop *Chlamys farreri* and its trans-generational immune protection to offspring against bacterial challenge. *Dev. Comp. Immunol.* 41, 569–577 (2013).
12. Nankabirwa, V. et al. Child survival and BCG vaccination: a community based prospective cohort study in Uganda. *BMC Public Health* 15, 175 (2015).
13. Kovacs, E. J. et al. Aging and innate immunity in the mouse: impact of intrinsic and extrinsic factors. *Trends Immunol.* 30, 319–324 (2009).