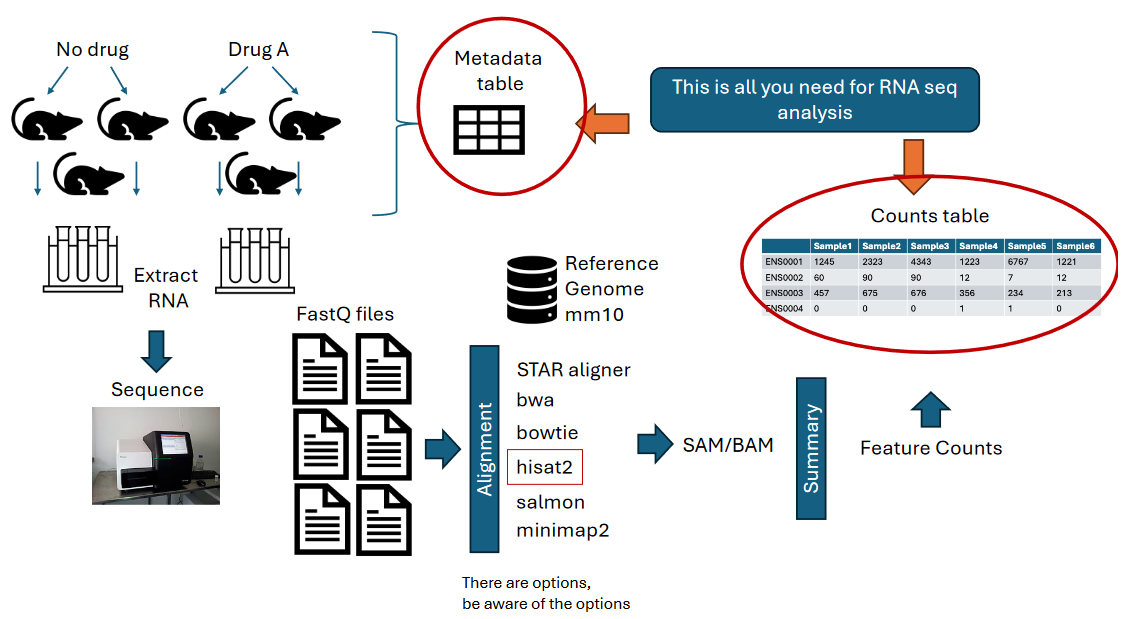
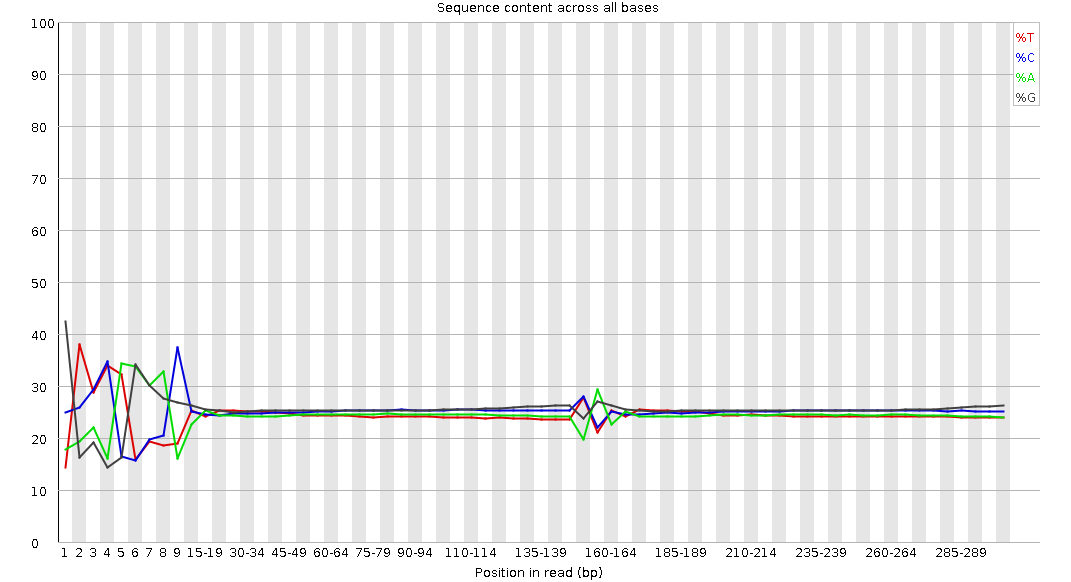
Joseph Uchytil

Bioinformatics in R

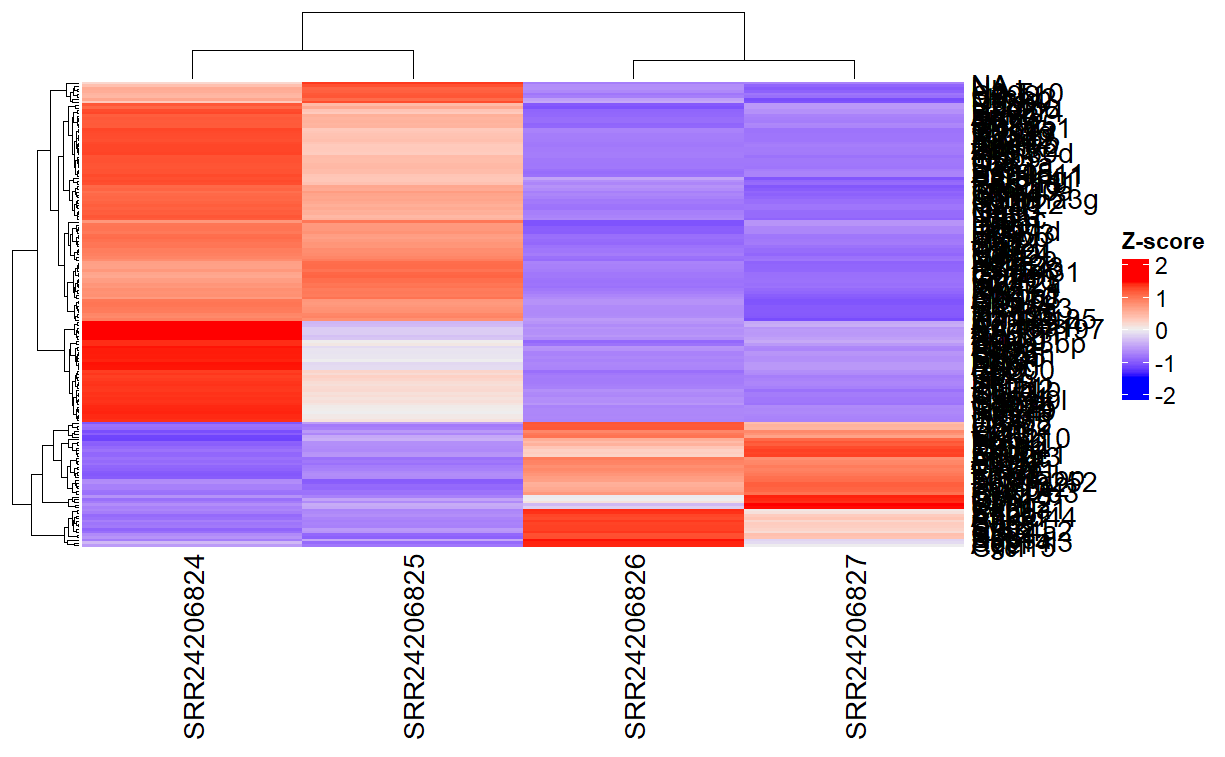
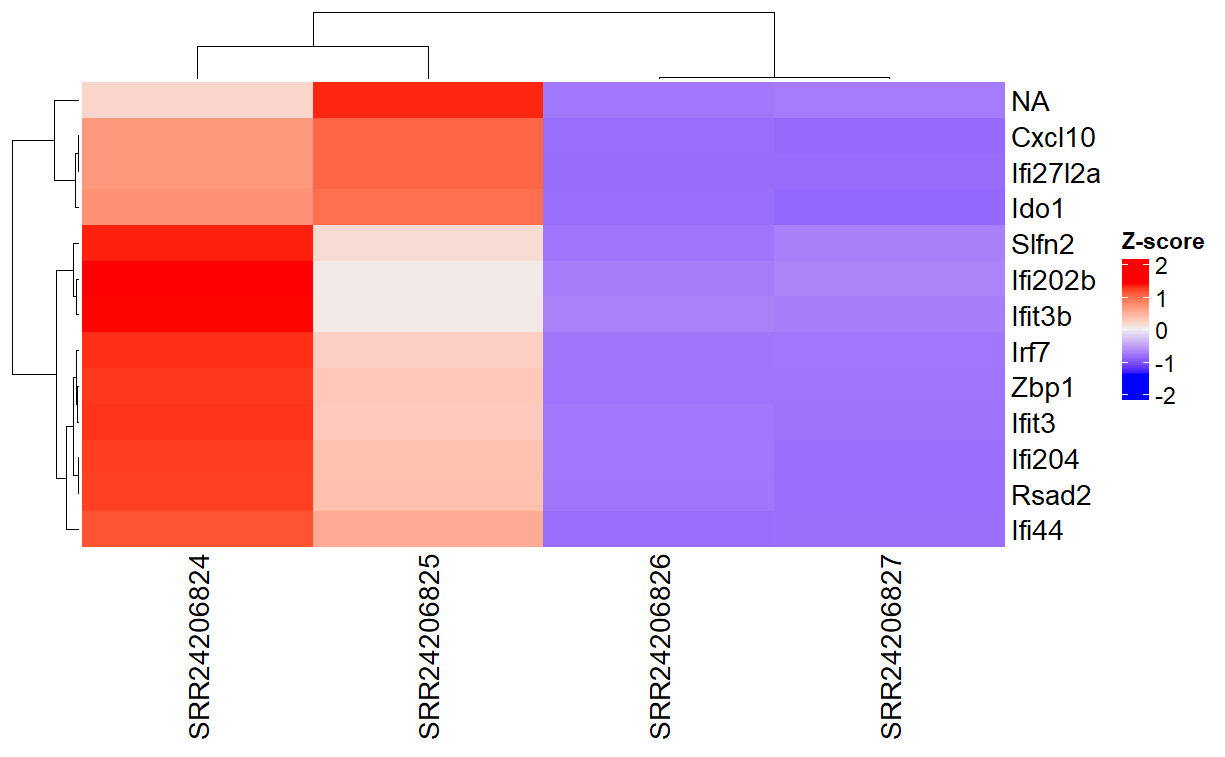
Project Draft

1. Understanding of the data
   1. I chose project 2, for this experiment they isolated and purified the pulmonary endothelial cells of mice. They split up the cells into four categories, young mice without the treatment, middle-aged mice without the treatment, young mice with the treatment, and middle-aged mice with the treatment. In this case the treatment was a mouse adapted SARS-CoV-2 strain. They extracted lung tissue from all the mice, isolated and amplified the RNA using qRT-PCR. Lastly, they used an Illumina prep kit and Illumina NovaSeq 6000 to run RNA-Seq on the RNA data.
   2. With this data, questions we would want to address are how SARS-CoV-2 affects the gene expression profile of the pulmonary endothelial cells. Important gene expression levels to look at would be the immune response genes and chronic lung disease genes.
   3. Potential challenges for this project would be that we only have the RNA seq data of one mouse for each different age and treatment so gene expression differences could be mouse specific rather than treatment specific and we wouldn’t be able to know without more data.
   4. Potential outcomes and future directions would be looking at genomic differences like methylation patterns.



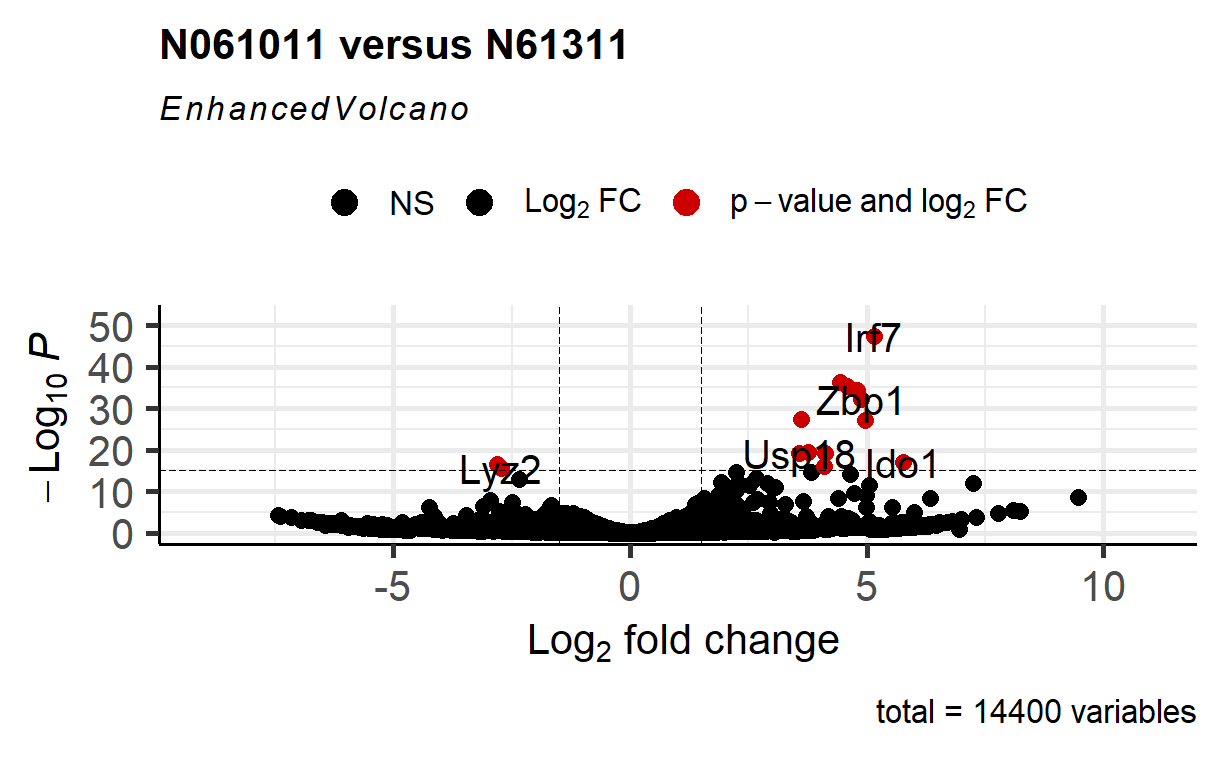
1. Data Analysis
   1. We did perform FastQC, and after running it we can see that the genetic content of the first 15 bases was different from the norm so we decided to trim that section off.
   2. 
   3. Alignment code at <https://colab.research.google.com/drive/14Xu45Oo6c27iR4vLNmINDnM6EEjmxuvP?usp=sharing>

* 1. DEseq2 Pipeline code can be seen in attached R file
  2. GO ontology enrichment can be seen in attached R file

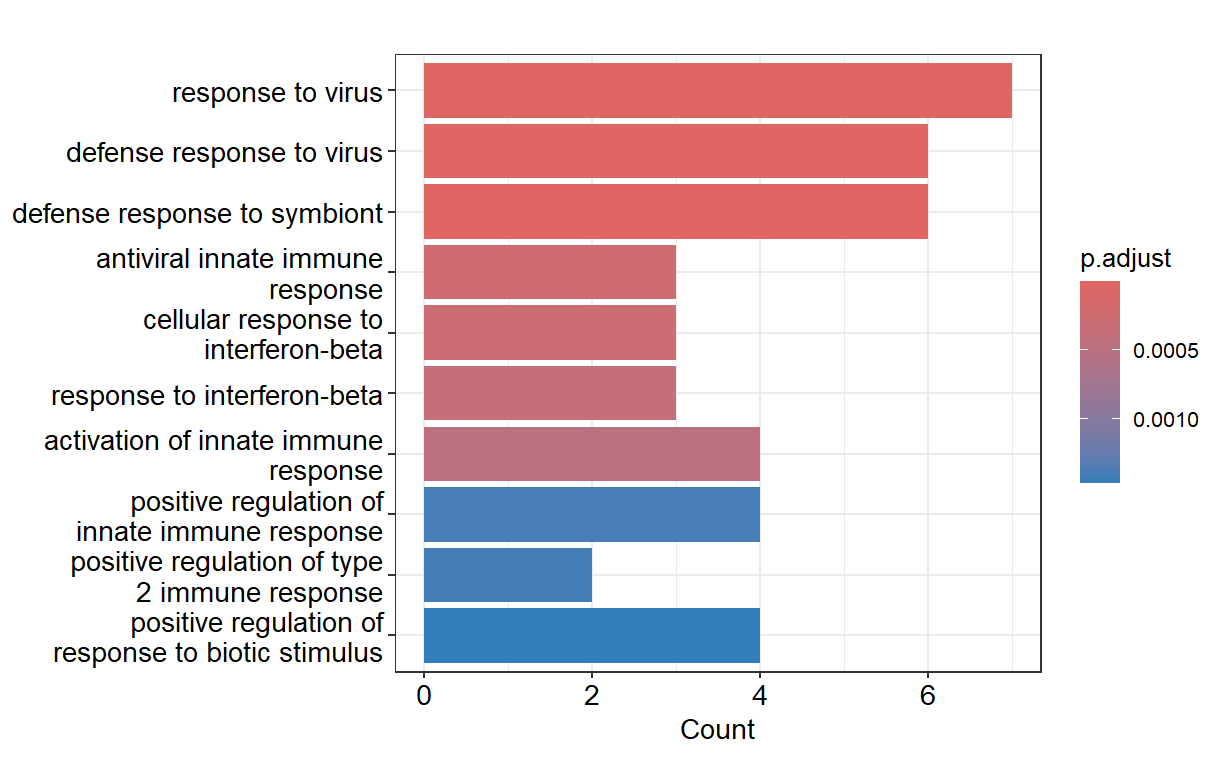
1. Interpretation and Conclusion
2. Presentation and Clarity
   1. Heatmaps

This heat map shows the 13 genes that had the most significant difference in genetic expression depending on the treatment that the sample had. The 2 samples on the left were treated with SARS-CoV-2 while the 2 on the right were the controls. These genes would probably be the main ones to investigate further.

This heat map shows every gene from the RNA seq data and compared the expression levels of those genes based on the treatment of our 4 samples. The 2 samples on the left were treated with SARS-CoV-2 while the 2 on the right were the controls. Red means overexpression and blue means under expressed in relation to eachother.

* 1. Volcano plots

This volcano plot shows the Log2 fold change of every gene in the counts table plotted against the -Log10P. The genes plotted as black dots are not significant while the genes plotted as red dots are significant. If plotted to the right it means it's overexpressed in our treatment, if plotted to the left it’s under expressed in out treatment.

* 1. GO Analysis figure

This GO Analysis takes the counts of all our genes, looks to see if they’re over or under expressed in our treatment, then groups them together based on what biological process the gene is involved in. After they have been grouped together, we can see what biological processes are over or under expressed as a whole based on our treatment. Our figure is showing that response to virus, defense response to virus, and defense response to symbiont are overexpressed in out treatment mice. Response to biotic stimulus, positive regulation of type 2 immune response, and positive regulation of innate immune response are under expressed.