Differentiating diseased and healthy human gene expression for MERS and COVID-19 using published data from 2 experiments.

Lucas Kopecky Bobadilla

Introduction

A patient's fight against a Coronavirus is largely dependent on the tenacity of the immune system hardcoded in their genome. A major challenge is determining how the virus will interact with the human immune system and influence gene expression. Identification of changes in gene expression during infection may focus research efforts toward looking for gene expression patterns in patients suffering from the virus.

MERS and SARS are both large positive-sense RNA viruses, part of the beta coronavirus group and fall into a highly pathogenic group that settles in the lower part of the respiratory tract. Rate of person to person transmissibility is a major difference between MERS and SARS, with SARS-CoV-2 (COVID from here on) being more infectious. Rabaan et al. (2020) compared the disease characteristics of MERS and SARS groups. COVID incubation period of 7-14 days is the longest, followed by SARS-CoV (2-7 days) and MERS (5-6 days). MERS includes symptoms of fever, nonproductive cough, sore throat and diarrhea has a high mortality rate of 34.4%. SARS has a lower fatality rate and more symptoms like fever, fatigue, nonproductive cough, myalgia, dyspnea, expectoration, sore throat, diarrhea, dizziness, headache, nausea or vomiting.

Coronaviruses are large RNA viruses that have their positive strand RNA contained within a nucleocapsid and outer envelope covered in glycoprotein spikes. Upon entry to the host cell, the host's RNA polymerase can synthesize the virus' mRNA that codes for spike (S), membrane (M), envelope (E), nucleocapsid (N), and possibly hemagglutinin-esterase (HE) proteins. MERS and SARS enter the cell by different mechanisms. COVID enters a cell when the S-protein of the virus binds to the metalloprotease angiotensin-converting enzyme 2 (ACE2) receptor on the host endothelial cells as a result of the action of protease (suggested to be TMPRSS2) that cleaves the S-protein of SARS and allow the virus to enter the cell (Rabi et al. 2020; Bassendine et al. 2020). ACE2 receptors are also on the surface of cells throughout the body allowing the virus to invade many organ systems. MERS binds to dipeptidyl-peptidase 4 (DPP4) to enter human cells, where DPP4 is a glycoprotein expressed on a variety of cells, implicated in disease states of inflammation and diabetes and involved in glucose homeostasis, signaling and immune cell activation (Röhrborn, Wronkowitz, and Eckel 2015). Bassendine et al. (2020) notes that modeling studies of COVID suggest that it may bind to both ACE2 and DPP4.

Endothelial cells line the surface of blood vessels and lymphatic vessel and manage blood or lymph movement into the tissue. Phenotypic change may increase production of by producing signaling molecules like cytokines to communicate with other cells or by expressing adhesion molecules to allow attachment of pathogen fighting leukocytes (Pober and Sessa 2007). Channappanavar and Perlman (2017) noted MERS and SARs present high serum levels of pro-inflammatory cytokines and chemokines yet lack anti-inflammatory cytokines.

Cell culture experiments help define the time course of gene expression early in the stage of infection. The Calu-3 cell is a lung epithelial cell used for viral respiratory illness studies. It is an epithelial cell derived from a metastatic site lung adenocarcinoma of a 25-year-old male who had received prior therapy with cytoxan, bleomycin and adriamycin. The Calu-3 cells express angiotensin-converting enzyme 2 (ACE-2), the functional receptor of SARS-CoV, on their apical surface and form monolayers in culture (Tseng et al. 2005).

Diagnostic tools are available to diagnose SARS and MERS infections but have limitations. PCR tests may not be able to detect the viral RNA if the viral load is low or if sample handling errors occur (Zainol Rashid et al. 2020). Rashid et al evaluated 9 rapid detection tests (RDTs) which all used detection of IgG and IgM antibodies (Zainol Rashid et al. 2020) and noted a predictable pattern with IgM appearing before IgG, an antibody that signals the activation of the acquired immunity system. However, serology tests are not ideal test systems as cross-reactivity with other coronaviruses can occur. Large scale testing platforms using CRISPR-based nucleic acid detection in microarray technologies like CARMEN hold potential for low-cost testing for presence of pathogens from one sample (Ackerman et al. 2020).

The challenge to understand COVID-19 and develop treatment rapidly have spurred data sharing. The types of COVID-19 related data made open include scholarly articles, de-identified patient records, digital pathology image and genomics data. For scholarly articles, many scientific publishers have made articles on COVID-19 free of charge. The data analytics platform Kaggle initiated a COVID-19 Open research dataset challenge (CORD-19) (https://kaggle.com/allen-institute-for-ai/CORD-19-research-challenge). It posted a dataset with over 59,000 scholarly articles and 10 tasks to guide the participants' effort. The COVID-19 research database project opened a collection of de-identified medical and pharmacy claims, electronic health record, and mortality data(https://covid19researchdatabase.org/). However, users must register with institutional email and go through an approval process. In terms of genomics data, NCBI's virus repository has accumulated more than 2000 SRAS-Cov-2 sequences from all over the world. In comparison, the number of RNA-seq and expression counts datasets is relevantly small. Only 14 datasets are available as of May 5th. However, with intensive research activities around the world, we expect to see more RNA-seq data deposited in the future.

Having open data encourages data reuse. NextStrain's analyses of the combined 2000+ virus sequence data has helped users to track the spread around the world(https://nextstrain.org/). The website also actively disputes misinformation with narratives in multiple languages. On the other hand, analysis of combined COVID-19 gene expression data has not been seen, due to the scarcity of the RNA-seq data deposited so far. Potential benefits from such analysis include detecting signals that are too weak to be detected in one experiment(Shah, Balakrishnan, and Wainwright 2016), discovering most consistently differentially expressed genes(Rung and Brazma 2013), and establishing unified theory across disease types(Baschal et al. 2020). When combining RNA-seq datasets for analysis, one key step is to remove the batch effect from different studies, and the `ComBat` function in `SVA` package has been the most commonly used tool for such purpose(Wang et al. 2018; Danielsson et al. 2015).

In this study we took samples from two GEO datasets (GSE147507 and GSE122876). Both datasets were single read RNA-seq files from Calu-3 cells infected with the virus after 24 hours, one infected with MERS and the other infected with COVID. We plan to investigate the difference between their gene expression profile and build a machine learning model to automatically classify disease types by gene expression profile.

Materials and Methods

Data source and pre-processing

We extracted data for this comparison from two GEO datasets (GSE147507 and GSE122876). Both datasets were single read RNA-seq files from Calu-3 cells infected with the virus after 24h. We conducted both experiments using triplicates for each condition. We downloaded raw fastq files for both datasets from NCBI using the SRA Toolkit and assess we used fastQC reports to data quality. All raw fastq files were trimmed using `Trimmomatic` (Bolger, Lohse, and Usadel 2014) set for LEADING:28, TRAILING:28, SLIDINGWINDOW:4:30 and MINLEN:50. After trimming low quality reads, libraries were pseudo-aligned to the homo sapiens genome GRCh38 using `Kallisto` without bootstrapping to get the transcript abundance(Bray et al. 2016).

Differential expression analysis

We conducted a differential expression analysis at the gene level to compare the differences and similarities about gene expression between the two types of Coronavirus. Analysis was possible using the package `EdgeR` (Robinson, McCarthy, and Smyth 2010). We selected the function `ComBat_seq` from package `sva` to

adjust for batch effects using an empirical Bayes framework (Leek et al. 2019). We compared the control treatments and common differentially expressed genes between disease. We imported abundance data into `EdgeR` using the package `tximport` (Love, Soneson, and Robinson 2017) and summarized results at the gene level using ensembl ID queried from `biomaRt` (Smedley et al. 2009). A TMM method normalized the data for library sizes. Common, Trended and Tagwise Negative Binomial dispersions estimated dispersions by weighted likelihood empirical Bayes and robustified data against outliers. We created a model matrix based on the disease condition to compare to the mock treatment. Gene count was fitted into Negative Binomial Generalized Linear Models.
We adjusted the p-value using the false discovery rate method (FDR) and differential expression significance was set to FDR < 0.001.

PCA

We used principal component analysis to reduce the dimensionality of the data so we could identify clusters in the data. We batch-corrected the abundance files using `ComBat` (Leek et al. 2019) and then ran a principle components analysis to reduce the dimensionality of the data to create a better visualization of the groups.

Enrichment analysis

We conducted an enrichment analysis using `GO-seq` package to check the biological process involved in the differentially expressed genes that were in common between COVID-19 and MERS.

Volcano plots

Volcano plots are scatter plots used to represent p-values against the differences in expression levels between the two samples (termed log2-fold changes) compared. Points of interest were determined by looking for points that display a significant p-value and a large fold change. For this analysis two volcano plots were made: one for comparing COVID to the control untreated cells and one comparing MERS to the control untreated cells.

Classification: Logistic regression and Support Vector Machines

We tried 2 main classification methods on the gene abundance data frame created by `biomaRt` and `tximport`: regularized logistic regression and support vector machines (linear and non-linear). For classification we combined the COVID ($n = 3$) and MERS ($n = 3$) observations into one condition: diseased ($n = 6$). Each diseased observation had an associated healthy observation ($n = 6$).
To control for batch effects we included a blocking factor to group each diseased observation with the corresponding healthy observation.

Logistic regression

We conducted a logistic regression using binomial errors in `glmnet` (Friedman, Hastie, and Tibshirani 2010). We forced the blocking factor to have an unpenalized coefficient by using the `penalty.factor` argument within `glmnet` and setting the penalty for the blocking effect coefficient to be zero. We used `cv.glmnet` to select for an optimal value of λ (the regularization parameter). Due to the small dataset we chose to use the leave one out cross validation method by setting the `nfolds` argument to 12 (the number of observations in our data). We also used a `for` loop to select an optimal value for α. The tuning parameter α determines whether we use ridge regression, LASSO, or Elastic Net (a mixture of ridge and LASSO). We tried a list of 11 α values from 0 to 1 by 0.1.

To assess the prediction accuracy of our model we constructed a `for` loop to separate the data into testing and training sets. This was essentially leave-one-out cross validation. We created our model based on the training set. Each iteration of the loop assessed whether prediction worked with the testing set correctly. Due to the low sample size the testing set only ever had 1 observation to predict. Finally, we created a confusion matrix based from the observed vs. predicted data using a probability cutoff of 0.5.

Support Vector Machines (Linear and Kernel-based (non-linear))

The Support Vector Machine (SVM) is a supervised learning method of classification. We can think of a situation like this: In a two-dimensional coordinate axis, we draw every observation on it as a single dot. SVM is a method that allows us to find a line which can separate these dots into two groups with the lowest misclassification rate. The linear SVM gets its name because a straight-line separates data. Non-linear SVM creates a non-linear separation in the points and is often implemented using kernel densities. If we throw every gene into this model we could get a 100% accuracy, so instead we tested the accuracy of SVM when only a handful of random genes were selected. This is the same idea used in random forest to select random predictors.

Results

Differential expression analysis and PCA

For all differentially expressed genes (DEG) for both diseases SARS and MERS infections both induced expression of 3505 genes when compared to the control (Figure 1). A total of 2175 genes were unique to COVID and 4225 were unique to MERS. We generated 2 volcano plots to visualize the different expression pattern between the diseases (Figures 2 and 3). Results show that MERS has a larger distribution between down and up regulated genes while for COVID we can see a vast majority of differentially expressed genes up regulated.

By looking at the top 400 genes from both diseases, 80 genes are shared between the diseases with similar patterns. From the 400 most differentially expressed genes for each disease, COVID presented 337 genes up-regulated and 63 down-regulated while MERS presented 223 genes up-regulated and 177 genes down-regulated. The PCA results also suggest a high amount of correlation between the genes (Figure 4).

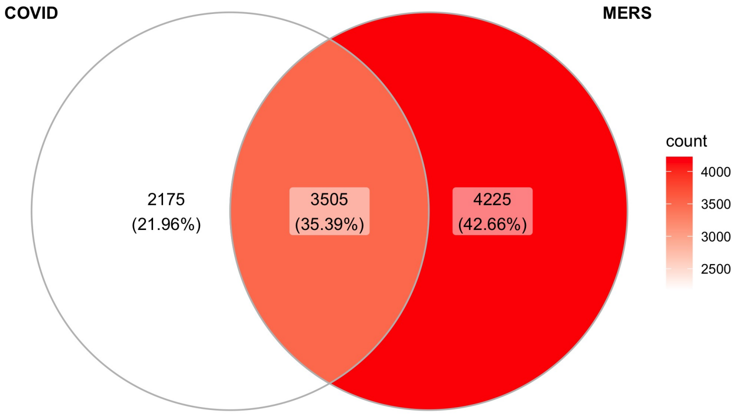

Figure 1: Venn diagram illustrates the gene expression shared by MERS and COVID

The gene expression of GBP 1, 2 and 3 were decreased in COVID samples compared to the control, while the expression of GBP 1,2,3 were increased for the MERS infected cells (Figure 5). Gene from the tumor necrosis family and interleukin family also generally showed higher rates of expression in both COVID and MERS when compared to the control (Figures 6 and 7).

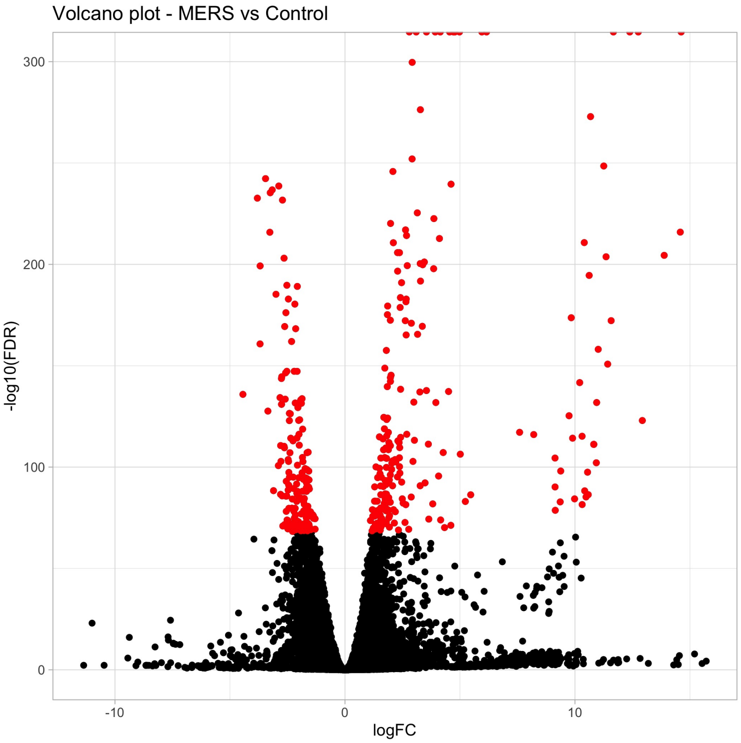

Figure 2: A volcano plot of MERS versus control

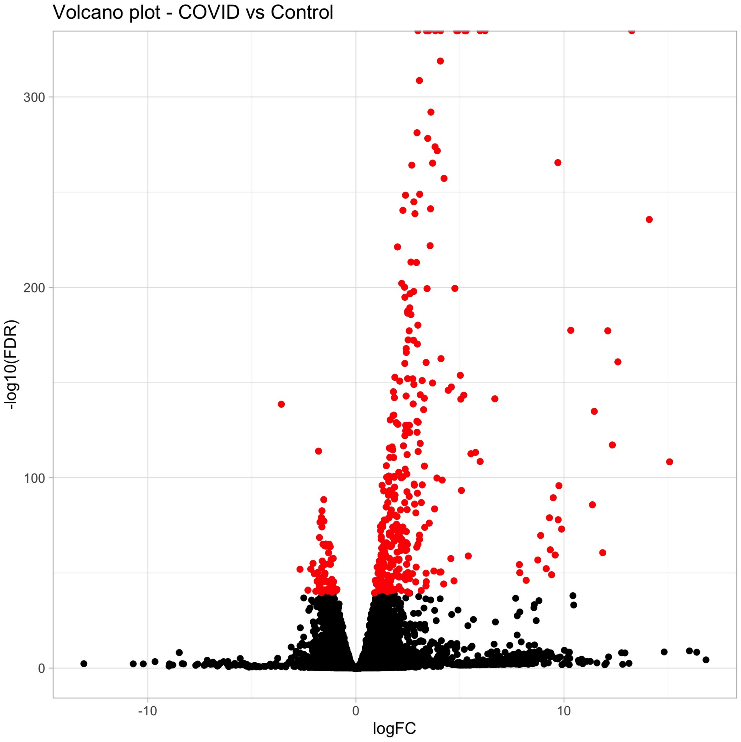

Figure 3: A volcano plot of COVID versus control

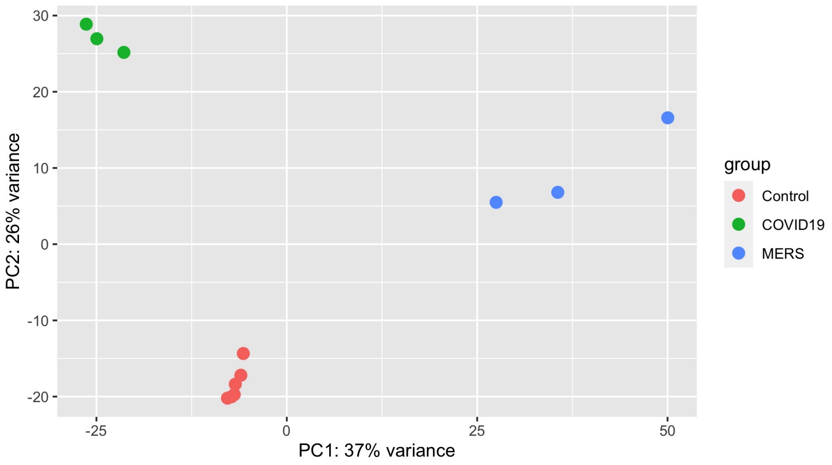

Figure 4: PCA results using the first two principal components. Strong evidence that we can seperate MERS, COVID, and healthy tissue based on human gene expression.

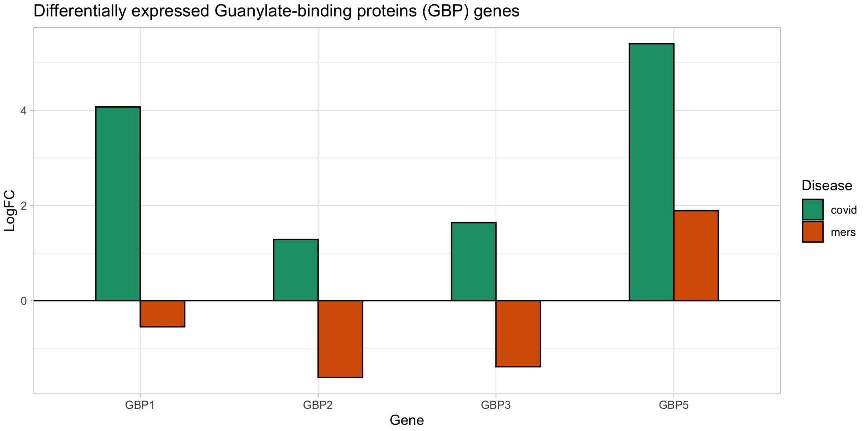

Figure 5: Comparison of differentially expressed Guanylate-binding protein genes

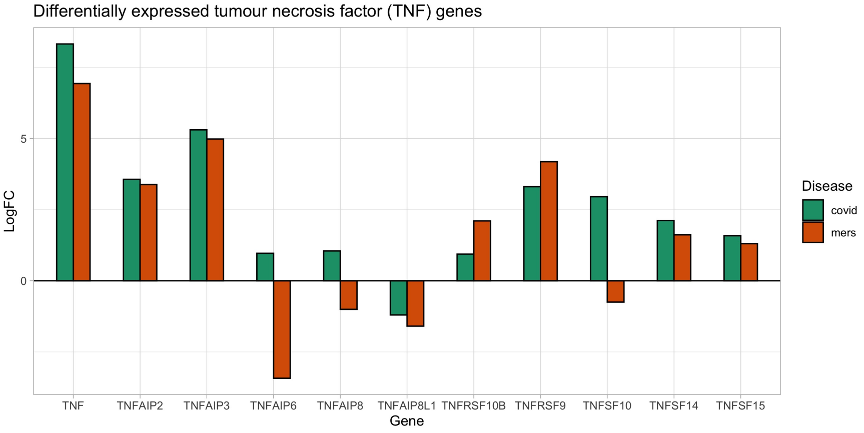

Figure 6: Differentially expressed genes in Tumor Necrosis Family

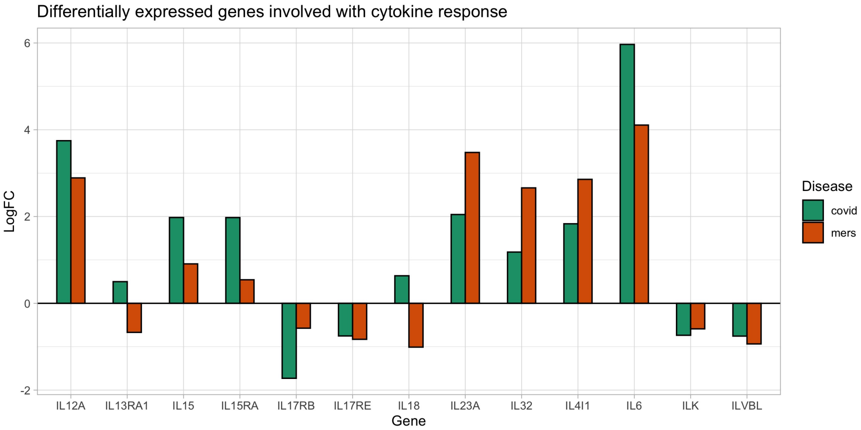

Figure 7: Differentially expressed genes in the interleukin family of genes

Enrichment analysis

The results of our enrichment analysis indicate that GO terms involved with metabolic processes are the most significant and that terms invloved with apoptosis and cytokinin regulation processes were also highly significant (Figure 8).

Logistic regression

Based on the cross validation results we chose $\alpha = 1$ (LASSO) and $\lambda = 0.04$. 6 of 6 infected cases were predicted correctly while 5 of 6 healthy cases were predicted correctly. LASSO reduced all predictor coefficients to zero except for gene “ENSG00000255967” (not including the intercept and blocking factor).

SVM

Classification accuracy was very high using both linear (Figure 9) and non-linear (not shown) SVM when 20 or more genes are selected randomly. Including around 10 random genes resulted in average classification accuracies of 95%.

Discussion:

These results strongly support the idea that RNA seq data of human level expression can be used to discriminate between MERS, COVID, and uninfected tissues.

35% of genes differentially expressed were in common between both diseases. Although they are both from the coronavirus family, MERS and COVID use different receptors to enter the cell, and thus enact different cell responses. Such difference may explain why the two viruses share only 35% of differentially expressed genes. Cells have an innate immunity response following infection and the 35% could represent within this common group of genes.

Among those genes, the two most significant up-regulated genes that was common to both diseases was the gene GBP5 (guanylate binding protein 5) that belongs to the TRAFAC class of dynamin-like GTPase superfamily and the gene IFNB1 (interferon beta 1) which encodes a cytokine that belongs to the interferon family of signaling proteins. GBP5 encodes a protein that activates NLRP3 inflammasome assembly and has a role in innate immunity and inflammation. The protein encoded by IFNB1 belongs to the type I class of interferons, which are important for defense against viral infections pathogenesis(Schoeman and Fielding

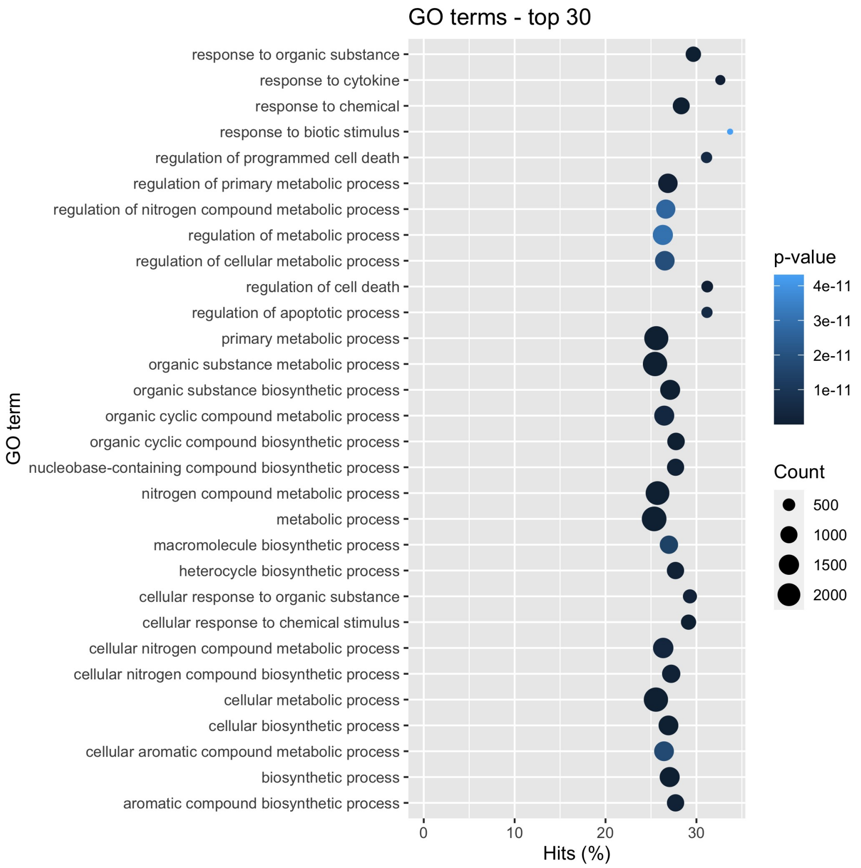

Figure 8: Enrichment analysis results. These are the metabolic processes of genes commonly expressed between COVID and MERS.

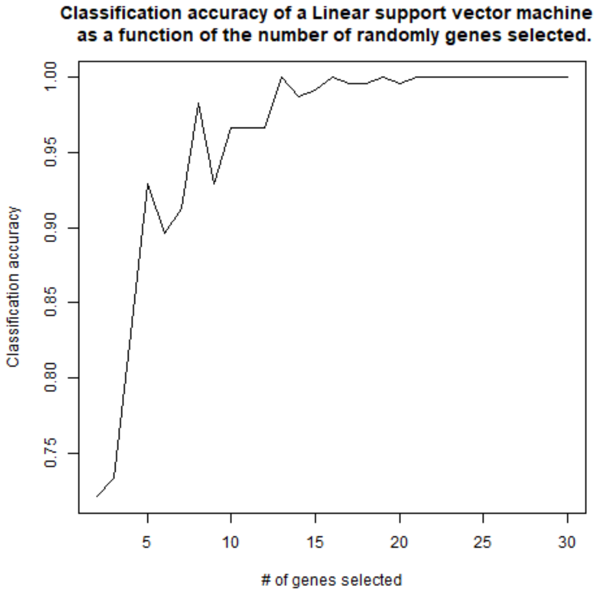

Figure 9: Classification accuracy of a linear SVM when a random subset of genes is used for prediction. The high amount of accuracy at low gene numbers indicates a high correlation between genes which is consistent with our PCA results.

2019). In the TNF subset of genes, TNFAIP6 and 8, TNFSF10 expression differed from controls depending on the infection. MERS showed a decrease while COVID showed an increase. TNFSF10 is highly expressed in COVID may be related to apoptosis in the cell.

GBP5 is part of a group of GBPs and was noted to play a role in reducing the infectious rate of the dsRNA retrovirus HIV-1 (Krapp et al. 2016). GBP5 is termed as a restriction factor and the expression levels of GBP5 may determine the production of infectious HIV-1 in macrophages, by interfering with Env processing. In our comparison (see Figure 5), the gene expression of GBP 1, 2 and 3 were decreased in COVID samples compared to the control, while the expression of GBP 1,2,3 were increased for the MERS infected cells. GBP5 was also found to be associated with expressions in CD14+ monocytes, a type of immune cell [@ Johri2020_04_26_20081182].

IFNB1 belongs to the family of interferon signaling proteins as part of the innate immune response and to type 1 which defend the host against viral infections. Both GBP5 and IFNB1 have antiviral properties. In the GO analysis many of the top classes of differentially expressed genes that had the most hits pertained to programmed cell death, apoptosis, and response to cytokines. The cytokine storm syndrome influences the severity of COVID illnesses. The body's dysregulation of the cytokine pathways typically includes a decrease in the antiviral cytokines (Sarzi-Puttini et al. 2020).

The severe symptoms seen in a many subgroup of COVID patients may be suffering from cytokine storm syndrome, where the bodies production of cytokines may lead to ARDS (Mehta et al. (2020)). The storm refers to an excessive level of proinflammatory cytokines produced by the host with the intention of destroying the virus, but damages the host tissues, and is believed to cause ARDS (Acute Respiratory Distress Syndrome) (Yang et al. 2020). Yang et al. (2020) evaluated 53 plasma samples from COVID cases for the presence of 48 cytokines and noted increased concentrations of P-10, MCP-3 and IL-1ra were associated with negative health outcomes. Some treatments for COVID have used immunosuppression to tame the hyperinflammation that is damaging the body leading to ARDs.

The success of the principal components analysis to separate these groups suggests a high degree of correlation between the genes which is supported by the randomized-genes support vector machine results. Multivariate methods like PCA generally work best with more data, but these results strongly suggest that these methods would work well for discriminating between MERS-infected, COVID-infected, and healthy tissues.

The logistic regression was also able to identify all infected tissues while only misidentifying one uninfected individual as infected, which is an appropriate type of error when the goal is to protect public health. Even though LASSO was the best regularization method with these specific data it may not be the best with a larger sample size. Ridge regression and elastic net tend to be better with highly correlated predictors. LASSO may have performed so well here because dimension reduction was so important based on the sample size. Elastic Net may turn out to be a better choice with more data.Based on the PCA and the classification results, RNA seq using human cells could be an important method of discriminating between healthy and sick individuals moving forward.

The gene selected by LASSO was ENSG00000255967 which is a component of the HADHA pseudogene 2 (HADHAP2): a mitochondrial protein ("HADHA Gene - Genetics Home Reference - Nih," n.d.). HADHA associates with the dicer, a dsRNA endoribonuclease, involved in RNA silencing and Kakumani suggests an auxiliary role for HADHA in miRNA biogenesis (Kakumani et al. 2015) and may play roles in RNA silencing and post-transcriptional regulation of gene expression. It's certainly possible that there's a relation here to fever, however we would need more data to know for sure.

Limitations

This study reflects gene expression from the Calu-3 cells sourced from a lung adenocarcinoma which may not be representative for other types of cells in the human body. Although Calu-3 is a functional model to study lung epithelial cells, care must be taken when extrapolating the results to healthy epithelial cells. Martens noted differences in the immune response of Calu-3 cells compared to primary nasal epithelial cells (Martens, Hellings, and Steelant 2018) with respect to cell permeability and cytokine production (IL8). The small sample size is another problem, which may make our results less generalizable. Finally, using existing studies rather than conducting new experiments is slightly more like an observational study than a true experiment. Our results indicate possibilities but they have to be investigated further.

Conclusion

We have successfully used two sets of publicly available RNA-seq data, one from COVID infected Calu-3 cell and one from MERS infected Calu-3 cell, to carry out our own analysis. We found that 35% of genes differentially expressed were in common between both diseases. The most DE genes were GBP5 and IFNB1, both of which have antiviral properties. Classification results suggest that we can differentiate between untreated, and COVID or MERS infected tissues based on human gene expression.

References

Ackerman, Cheri M, Cameron Myhrvold, Sri Gowtham Thakku, Catherine A Freije, Hayden C Metsky, David K Yang, Simon H Ye, et al. 2020. "Massively Multiplexed Nucleic Acid Detection Using Cas13." Springer Science; Business Media LLC.

Baschal, Erin E, Eric D Larson, Tori C Bootpetch Roberts, Shivani Pathak, Gretchen Frank, Elyse Handley, Jordyn Dinwiddie, et al. 2020. "Identification of Novel Genes and Biological Pathways That Overlap in Infectious and Nonallergic Diseases of the Upper and Lower Airways Using Network Analyses." *Frontiers in Genetics* 10. Frontiers: 1352.

Bassendine, Margaret F, Simon H Bridge, Geoffrey W McCaughan, and Mark D Gorrell. 2020. "Covid-19 and Co-Morbidities: A Role for Dipeptidyl Peptidase 4 (Dpp4) in Disease Severity?"

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20.

Bray, Nicolas L., Harold Pimentel, Páll Melsted, and Lior Pachter. 2016. "Near-Optimal Probabilistic RNA-Seq Quantification." *Nature Biotechnology* 34 (5): 525–27.

Channappanavar, Rudragouda, and Stanley Perlman. 2017. "Pathogenic Human Coronavirus Infections: Causes and Consequences of Cytokine Storm and Immunopathology." *Seminars in Immunopathology* 39 (5): 529–39. https://doi.org/10.1007/s00281-017-0629-x.

Danielsson, Frida, Tojo James, David Gomez-Cabrero, and Mikael Huss. 2015. "Assessing the Consistency of Public Human Tissue RNA-Seq Data Sets." *Briefings in Bioinformatics* 16 (6): 941–49. https://doi.org/10.1093/bib/bbv017.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1): 1–22. http://www.jstatsoft.org/v33/i01/.

"HADHA Gene - Genetics Home Reference - Nih." n.d. *U.S. National Library of Medicine.* National Institutes of Health. https://ghr.nlm.nih.gov/gene/HADHA.

Kakumani, Pavan Kumar, Rajgokul K Shanmugam, Inderjeet Kaur, Pawan Malhotra, Sunil K Mukherjee, and Raj K Bhatnagar. 2015. "Association of Hadha with Human Rna Silencing Machinery." *Biochemical and Biophysical Research Communications* 466 (3). Elsevier: 481–85.

Krapp, Christian, Dominik Hotter, Ali Gawanbacht, Paul J McLaren, Silvia F Kluge, Christina M Stürzel, Katharina Mack, et al. 2016. "Guanylate Binding Protein (Gbp) 5 Is an Interferon-Inducible Inhibitor of Hiv-1 Infectivity." *Cell Host & Microbe* 19 (4). Elsevier: 504–14.

Leek, Jeffrey T., W. Evan Johnson, Hilary S. Parker, Elana J. Fertig, Andrew E. Jaffe, John D. Storey, Yuqing Zhang, and Leonardo Collado Torres. 2019. *Sva: Surrogate Variable Analysis.*

Love, Michael I., Charlotte Soneson, and Mark D. Robinson. 2017. "Importing Transcript Abundance Datasets with Tximport." *Dim (Txi. Inf. Rep $ infReps $ Sample1)* 1 (178136): 5.

Martens, Katleen, Peter W Hellings, and Brecht Steelant. 2018. "Calu-3 Epithelial Cells Exhibit Different Immune and Epithelial Barrier Responses from Freshly Isolated Primary Nasal Epithelial Cells in Vitro." *Clinical and Translational Allergy* 8 (1). Springer: 40.

Mehta, Puja, Daniel F McAuley, Michael Brown, Emilie Sanchez, Rachel S Tattersall, and Jessica J Manson. 2020. "COVID-19: Consider Cytokine Storm Syndromes and Immunosuppression." *The Lancet* 395 (10229). Elsevier: 1033–4.

Pober, Jordan S, and William C Sessa. 2007. "Evolving Functions of Endothelial Cells in Inflammation." *Nature Reviews Immunology* 7 (10). Nature Publishing Group: 803–15.

Rabaan, Ali A, Shamsah H Al-Ahmed, Shafiul Haque, Ranjit Sah, Ruchi Tiwari, Yashpal Singh Malik, Kuldeep Dhama, M Iqbal Yatoo, D Katterine Bonilla-Aldana, and Alfonso J Rodriguez-Morales. 2020. "SARS-Cov-2, Sars-Cov, and Mers-Cov: A Comparative Overview." *Le Infezioni in Medicina* 2: 174–84.

Rabi, Firas A, Mazhar S Al Zoubi, Ghena A Kasasbeh, Dunia M Salameh, and Amjad D Al-Nasser. 2020. "SARS-Cov-2 and Coronavirus Disease 2019: What We Know so Far." *Pathogens* 9 (3). Multidisciplinary Digital Publishing Institute: 231.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. "edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26 (1): 139–40.

Röhrborn, Diana, Nina Wronkowitz, and Juergen Eckel. 2015. "DPP4 in Diabetes." *Frontiers in Immunology* 6. Frontiers: 386.

Rung, Johan, and Alvis Brazma. 2013. "Reuse of Public Genome-Wide Gene Expression Data." *Nature Reviews Genetics* 14 (2). Nature Publishing Group: 89–99.

Sarzi-Puttini, Piercarlo, Valeria Giorgi, Silvia Sirotti, Daniela Marotto, Sandro Ardizzone, Giuliano Rizzardini, Spinello Antinori, and Massimo Galli. 2020. "COVID-19, Cytokines and Immunosuppression: What Can We

Learn from Severe Acute Respiratory Syndrome?” *Clinical and Experimental Rheumatology* 38 (2): 337–42.

Schoeman, Dewald, and Burtram C Fielding. 2019. “Coronavirus Envelope Protein: Current Knowledge.” *Virology Journal* 16 (1). BioMed Central: 69.

Shah, Nihar B, Sivaraman Balakrishnan, and Martin J Wainwright. 2016. “A Permutation-Based Model for Crowd Labeling: Optimal Estimation and Robustness.” *arXiv Preprint arXiv:1606.09632*.

Smedley, Damian, Syed Haider, Benoit Ballester, Richard Holland, Darin London, Gudmundur Thorisson, and Arek Kasprzyk. 2009. “BioMart–Biological Queries Made Easy.” *BMC Genomics* 10 (1): 22.

Tseng, Chien-Te K, Jennifer Tseng, Lucy Perrone, Melissa Worthy, Vsevolod Popov, and Clarence J Peters. 2005. “Apical Entry and Release of Severe Acute Respiratory Syndrome-Associated Coronavirus in Polarized Calu-3 Lung Epithelial Cells.” *Journal of Virology* 79 (15). Am Soc Microbiol: 9470–9.

Wang, Qingguo, Joshua Armenia, Chao Zhang, Alexander V Penson, Ed Reznik, Liguo Zhang, Thais Minet, et al. 2018. “Unifying Cancer and Normal Rna Sequencing Data from Different Sources.” *Scientific Data* 5. Nature Publishing Group: 180061.

Yang, Yang, Chenguang Shen, Jinxiu Li, Jing Yuan, Minghui Yang, Fuxiang Wang, Guobao Li, et al. 2020. “Exuberant Elevation of Ip-10, Mcp-3 and Il-1ra During Sars-Cov-2 Infection Is Associated with Disease Severity and Fatal Outcome.” *MedRxiv*. Cold Spring Harbor Laboratory Press.

Zainol Rashid, Z., S. N. Othman, M. N. Abdul Samat, U. K. Ali, and K. K. Wong. 2020. “Diagnostic Performance of COVID-19 Serology Assays.” *The Malaysian Journal of Pathology* 42 (1): 13–21.