Investigation on difference in level of expression of cellular receptor for SARS-CoV-2, Angiotensin-converting Enzyme 2(ACE2), regarding age, gender and organ

Abstract:

Angiotensin-Converting Enzyme 2(ACE2) has been known as the host cell receptor for SARS-CoV-2, the virus that is causing the recent global pandemic. Some studies have suggested that the gene expression level of ACE2 might vary by age, gender and the specific tissue type the cell belongs to. This project analyzed RNA-sequencing data of gene coding for ACE2 across samples from different age group, gender and tissue types. Analysis is done using GEO2R platform and DESeq2 package to obtain results of differential gene expression analysis. Overall, this project validated the previous findings that ACE2 expression varies across different age groups, with the expression being higher in older individuals, and this project noticed that there is no significant difference between ACE2 expression across gender based on the datasets used.

Introduction:

The recent COVID-19 pandemic has caused devastating effects to the economies, societies, and people's lives around the world. The number of cases around the globe has ramped up to more than three million and it has cost more than 200 thousand lives. [1].

In order to work towards finding treatments to COVID-19 and potential vaccines to SARS-CoV-2, it is important to understand the virology of SARS-CoV-2 and the disease mechanism for COVID-19 pneumonia. To date, a number of scientists and research labs have reported that the human cellular receptor, angiotensin-converting enzyme 2(ACE2) is required for SARS-CoV-2 to enter cells [2, 3, 4]. ACE2 is a cellular receptor responsible for cells' communication with its extracellular environment, and under normal circumstances, it plays a very important role in the regulation of blood pressure. [5] In the context of COVID-19 infection, the viral protein of SARS-CoV-2 is able to bind to ACE2 on host cells' plasma membrane and gain entrance to host cells. In a way, it can be said that the presence of ACE2 on cell surface membrane facilitates the entrance of SARS-CoV-2 into host cells, thus aiding the viral infection of SARS-CoV-2 to human body.

Several researchers hypothesized that there could be a positive correlation between the level of expression of the ACE2 gene and susceptibility to SARS-CoV-2 and the disease COVID-19. [6, 7] Leng et. al. even suggested that a stem cell transfer involving ACE2 gene modification improves the outcome of COVID-19 pneumonia patients.

There has been many attempts trying to draw a correlation between expression level of ACE2 and potential susceptibility to COVID-19. Corley and Ndhlovu analyzed ACE2 expression level using DNA methylation data [8]. According to their study, ACE2 gene loci are hypomethylated in lung epithelial tissue cells compared to cells in other organs or tissues. They also reported lower level of DNA methylation of ACE2 gene loci in female as compared to male, signaling a higher gene expression level of ACE2 in females.

However, their study only considered DNA methylation data, which is a form of gene expression regulation that takes place in the chromosome level. DNA methylation blocks the binding site of some of the transcription factors, hence reducing the level of transcription of genes. It is assumed that the higher level of DNA methylation, the lower the level of expression of that particular gene. Notably, after DNA methylation, there are still gene expression regulation at transcription level, post-transcription level, translation level, and post translation level, hence DNA methylation data might not be able to represent the full picture of gene expression of individual genes and it is less conclusive in suggesting higher or lower level of expression for ACE2 gene across different cells or individuals.

In addition, their study only analyzed ACE2 expression with regard to the age factor on a small scale of 4 individuals, which is not sufficient to suggest any convincing correlations between age and expression of ACE2.

As part of the attempt to better understand the expression level of ACE2 in human, my project aims to build on and improve from the above-mentioned project by Corley and Ndhlovu. This project aims to investigate the differences in level of expression of ACE2 gene across age, gender and organs, using public datasets of RNA-sequencing data that are available online. RNA-sequencing refers to sequencing of the transcriptome of cells, mostly consisting of mRNA, which is what will actually be translated into proteins by the cell. Thus, RNA-sequencing analysis might be able to provide a better picture of the actual gene expression level, as compared to using DNA methylation analysis.

Specifically, this project is looking at the p-value, which is the probability of the differences in level of ACE2 gene expression being due to chance, across pairwise comparisons between different groups of samples within the same dataset. Those p-values are then tabulated and discussed.

Methodology:

RNA sequencing analysis is a commonly used tool for studying gene expression. The analysis is performed on data of sequenced RNA transcripts from cells, usually through the platform Illumina HiSeq or several other platforms. The raw RNA sequences then undergo pre-processing until a count matrix is produced. Further downstream analysis is then done afterwards.

Data acquisition

All three datasets analyzed in this project are from Genome Expression Omnibus (GEO). They are chosen primarily because they come with the desired labels or annotations for samples. For comparison of ACE2 expression across different age groups, GEO accession GSE19804 is used. It provides RNA sequencing data of lung tissue for individuals with age ranging from mid-30s to 80s. For comparison between genders, GSE66499 is used as it contains RNA-seq data for a few hundred individuals across both genders. GEO accession GSE120795 contains RNA-sequencing data with labels of the tissues from which these samples are obtained and is thus used for ACE2 analysis across cells from different organs.

Analysis using GEO2R

In this project, GEO2R, a web-based platform from GEO, is used to analyze datasets that do not come with a raw count matrix. GEO2R directly accesses and compares processed data of submitted to GEO and obtains the p-value for pairwise comparison of genes. Data from GSE19804 and GSE66499 are analyzed using GEO2R and the p-values across each independent variable are compared.

Analysis with DESeq2

In addition, raw count matrix is obtained from the supplementary files of GEO accession GSE120795, which includes the gene expression data of cells from different organs in human.

The count matrix comes in the raw form and each entry comes with the sample number and the counts for the respective genes. Manipulation is done on the count matrix using Python to map each sample to the respective label of the organ that it is taken from to generate the header file required by DESeq2. The DESeq2 package is loaded into RStudio, and the count matrix itself, and the header with organ labels, are then fed into DESeq2 for analysis of differential gene expression. In particular, the expression data of ACE2, among all genes, are extracted and studied.

Results and Discussion:

RNA-sequencing analysis result across age and gender

It has been suspected that age could play a role in affecting susceptibility to COVID-19. To investigate whether age has a huge impact on gene expression level of ACE2, the RNA-sequencing data from different individuals across different age groups from <40 years old to >70 years old are analyzed using GEO2R, an in-built analysis platform to Genome Expression Omnibus (GEO). The samples are divided into five different age groups, namely <40, 40 to 50, 50 to 60, 60 to 70, and older than 70. The p-values of pair-wise comparison between each two age groups specific to ACE2 (ENSG00000130234) are listed in the table below.

age/ p-values	**<40**	**40-50**	**50-60**	**60-70**	**70+**
<40					
40-50	0.517				
50-60	0.569	0.550			
60-70	0.024	0.057	0.516		
70+	0.021	0.052	0.611	0.565	

Table 1. P-value of pairwise comparison of level of expression of ACE2 across different age groups

As can be seen in the table, the p-value is not significant (> 0.05) for age groups that are close to each other (ie. only 10 years apart). However, significant p-values (< 0.05) are observed in comparison between age group < 40 years old and the age group 60-70 y.o. and 70+ y.o., suggesting a significant difference in the level of ACE2 expression across these pairs.

The result above suggests that the ACE2 expression level is higher in individuals with age above 60 years old as compared to individuals who are much younger, such as those who are younger than 40 y.o. or in the 40 to 50 years old range. This validates Corley and Ndhlovu's result regarding differences in ACE2 expression across different age.

In addition, to investigate the influence of gender on the expression of ACE2, the RNA-sequencing data of lung tissue of 110 males and 75 females are analyzed using GEO2R, and the gene expression of ACE2 gene (ID 8171449) is studied. The p-value is found out to be 0.374, which means that there is a 37.4% probability that the differences are merely due to chance. Thus, the data is not sufficient to suggest any correlation between gender and expression level of ACE2. Thus, further studies should be done to investigate the significance of gender in causing discrepancies in ACE2 expression among individuals.

DESeq2 results across different organs

To have a better idea of how the level of ACE2 expression varies across different organs, the normalized count of RNA-transcript for ACE2 for different cell types is first obtained using DESeq2 and plotted in figure 1 below. Normalized count refers to the proportion of mRNA each particular gene takes up in the transcriptome [10]. In this case, it is the proportion of mRNA that the ACE2 transcript takes up in cells from different organs.

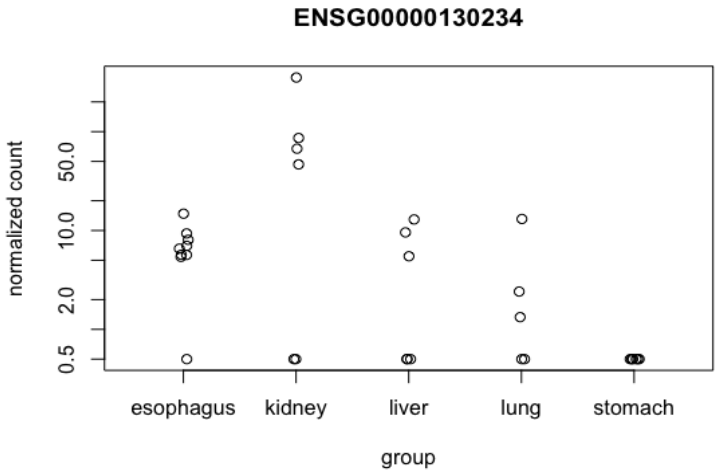

Figure 1. normalized count of ACE2 transcript across cells from different organs

As seen in figure 1, kidney cells have the highest expression of ACE2 in general according to this dataset. Esophagus cells have the second highest count on average, followed by lung and liver. Stomach cells have the lowest level of expression of ACE2 across the organs analyzed here. There is some variance in normalized count within liver and lung, potentially due to the differences in specific tissues those samples are taken from (which we do not have more information about).

Those findings are mostly consistent with the literature and clinical observations. [11] According to clinicians, lung is the primary attack target of SARS-CoV-2, and some severe patients subsequently develop kidney failures and/or gastrointestinal disorders. [12]

In addition to the normalized count plot, to study the significance of differences of ACE2 expression across these different organs, pair-wise comparisons are performed using DESeq2 to obtain the p-values, which can be found in table 2 below.

p-value	**lung**	**liver**	**esophagus**	**kidney**	**stomach**
lung					
liver	7.86E-01				
esophagus	2.76E-01	5.84E-01			
kidney	4.50E-05	3.30E-04	8.08E-10		
stomach	3.31E-01	2.54E-01	3.84E-04	2.57E-07	

Table 2. P-value of pairwise comparison of expression of ACE2 across cells from different organs

As can be seen in the table, the p-values for comparison between kidney and all other cell types are significant, indicating that kidney cells have a significantly higher level of ACE2 expression compared to other cell types studied here. This can be explained by the fact that ACE2 is responsible for regulating blood pressure in normal cells, thus they are present in bulk in kidney where blood pressure plays an important part in the normal functionality. [13]

Besides RNA-sequencing data of ACE2, the RNA-seq data of all other genes are also studied in pairs across the 5 different organs, as can be seen in figure 2. the volcano plot and figure 3. the PCA plot across cells from different organs.

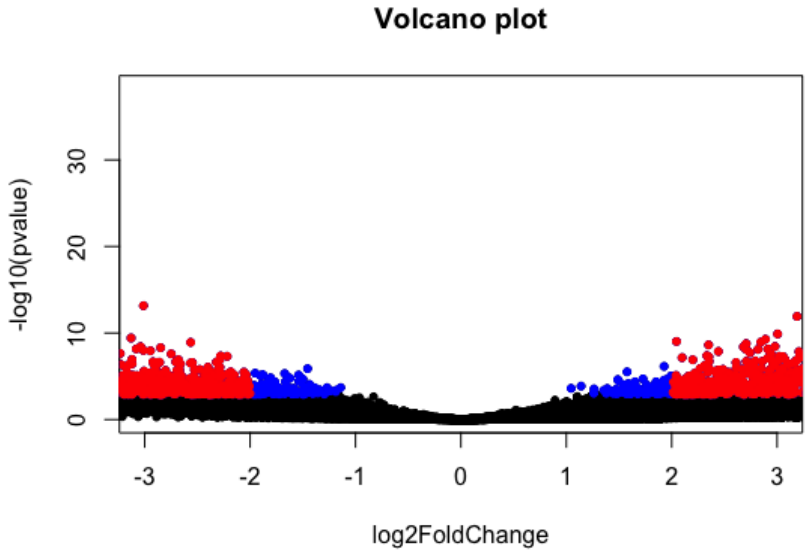

Figure 2. volcano plot of gene expression across lung, liver, esophagus, kidney and stomach

According to the volcano plot, most genes have a -log10fold of < 10. This suggests that the differential expression of ACE2 between kidney and esophagus (p-value = 8.08E-10) and between kidney and stomach (p-value =2.57E-07) are very significant, and the differences between ACE2 expression of those cell types (kidney and esophagus, kidney and stomach) are worth noting.

The overall gene expression data across the 5 different organs are visualized in a PCA plot as shown below. Based on the PCA, the general gene expression level of lung, stomach and kidney displays more similarity compared to the general gene expression level of liver and esophagus. The discrepancy between the gene expression level of ACE2 gene and overall gene expression across organs is not surprising and can be explained by the differences in functionalities of cells across organs.

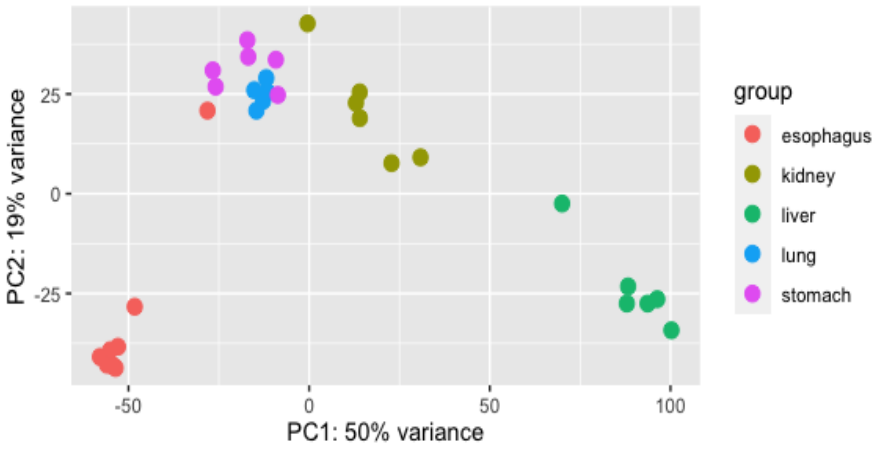

Figure 3. PCA plot of gene expression across lung, liver, esophagus, kidney and stomach

Limitation and conclusion

In summary, the results of this analysis show that ACE2 gene expression indeed differs by age. Specifically, older individuals displayed higher level of expression of ACE2 gene. This project also found that there is no significant difference in ACE2 expression level across gender. Based on the result of this analysis, ACE2 expression differs across organs, with kidney cells having the highest level of expression and stomach the lowest among the five organs being studied.

This project does have a few limitations.

First of all, this project only analyzed one dataset for each variable being investigated. More datasets could be analyzed to obtain a more reliable conclusion. However, due to limitations on the amount of public RNA-sequencing datasets available, this project only analyzed the three datasets mentioned above.

In addition, due to the difficulties in obtaining RNA-sequencing data of COVID-19 patients, this project only has access to RNA-sequencing data of the general population. Thus, only observations about ACE2 expression across different individuals can be made, yet no conclusion about how those factors might correlate to susceptibility to SARS-CoV-2 or COVID-19.

Bibliography

1. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) (2020). Retrieved from: https://coronavirus.jhu.edu/map.html
2. Li, X., Geng, M., Peng, Y., Meng, L., & Lu, S. (2020). Molecular immune pathogenesis and diagnosis of COVID-19. *Journal of Pharmaceutical Analysis*.
3. Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., ... & Choe, H. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, *426*(6965), 450-454.
4. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... & Chen, H. D. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270-273.
5. Oudit, G. Y., Crackower, M. A., Backx, P. H., & Penninger, J. M. (2003). The role of ACE2 in cardiovascular physiology. Trends in cardiovascular medicine, 13(3), 93-101.
6. Leng, Z., Zhu, R., Hou, W., Feng, Y., Yang, Y., Han, Q., ... & Fan, J. (2020). Transplantation of ACE2-mesenchymal stem cells improves the outcome of patients with COVID-19 pneumonia. Aging and disease, 11(2), 216-228.
7. South, A. M., Diz, D., & Chappell, M. C. (2020). COVID-19, ACE2 and the Cardiovascular Consequences.
8. Corley, M. J., & Ndhlovu, L. C. (2020). DNA methylation analysis of the COVID-19 host cell receptor, angiotensin I converting enzyme 2 gene (ACE2) in the respiratory system reveal age and gender differences.
9. Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biology, 15, 550. doi: 10.1186/s13059-014-0550-8.
10. Evans, C., Hardin, J., & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Briefings in bioinformatics, 19(5), 776–792. https://doi.org/10.1093/bib/bbx008
11. Burrell, L. M., Johnston, C. I., Tikellis, C., & Cooper, M. E. (2004). ACE2, a new regulator of the renin–angiotensin system. Trends in Endocrinology & Metabolism, 15(4), 166-169.
12. Cheng, Y., Luo, R., Wang, K., Zhang, M., Wang, Z., Dong, L., ... & Xu, G. (2020). Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney international.
13. Jafar, T. H., Stark, P. C., Schmid, C. H., Landa, M., Maschio, G., De Jong, P. E., ... & Levey, A. S. (2003). Progression of chronic kidney disease: the role of blood pressure control, proteinuria, and angiotensin-converting enzyme inhibition: a patient-level meta-analysis. Annals of internal medicine, 139(4), 244-252.