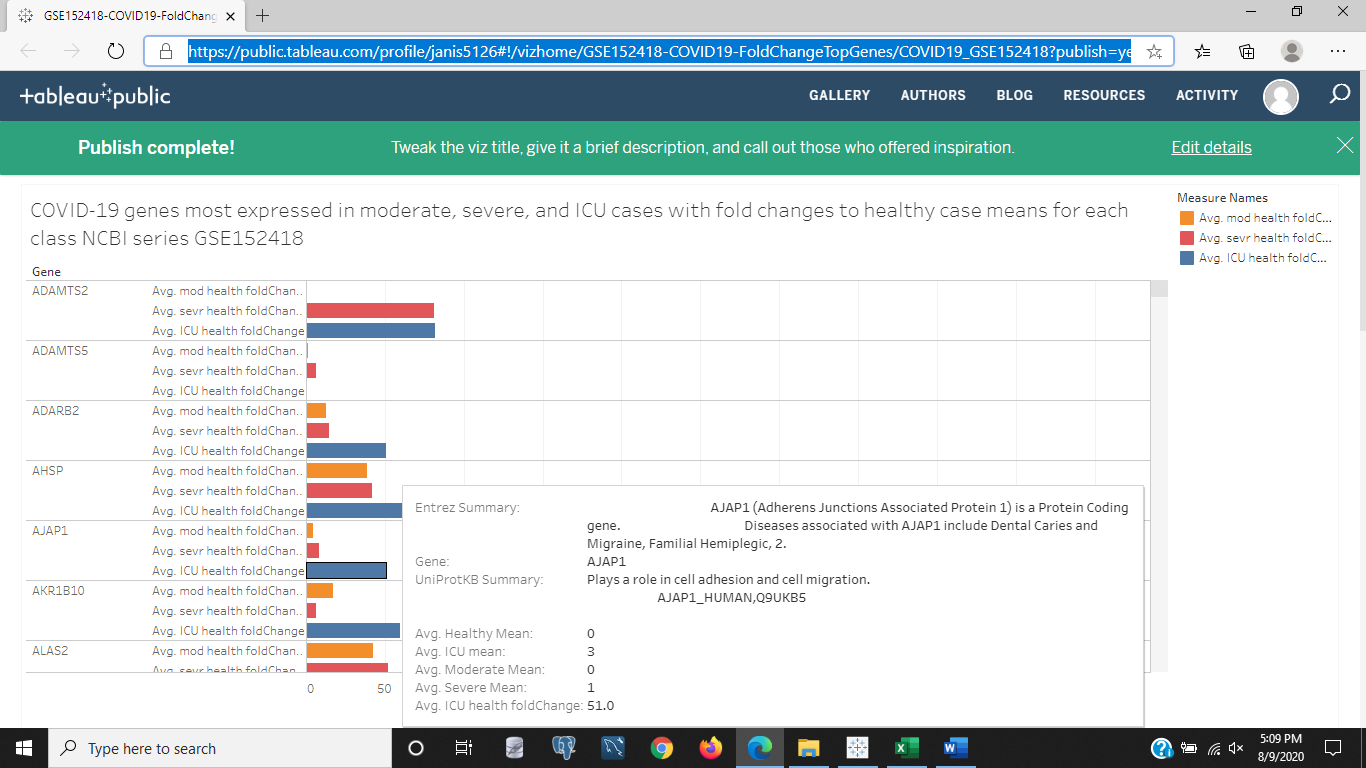
In this blog, I was able to find some COVID-19 gene expression data taken from peripheral blood mononuceated cells (PBMC) from 76 patients who ranged from 23-91 years of age and mixed genders with 18 females and 16 males in 34 total samples. The data was collected from patients who had three different cases of COVID-19 as 'moderate','severe',or 'ICU', and some healthy non-COVID-19 infected blood samples. The fold change of the sample means in each group of healthy sample means, moderate sample means, severe sample means, and ICU sample means. The diseased state over the healthy state ratio was the fold change measure. Then the patients were divided by age within each group keeping the gender mixed if it was because the moderate COVID-19 cases were all females. And also the groups were divided by gender of male or female while keeping mixed ages. The median age was 56, so the group ages were divided by those 56 and younger to those older than 56.

Some interesting results. The genes numbered in the 56 thousand plus range, but only the Ensembl ID of the gene was given, not any allele copy number variants in the sequence motifs to perform any other analysis for those genes having mutations that are more likely to show COVID-19 severe or ICU cases. When looking at the images below, they all link to the interactive dashboards, and only the top 100 genes were taken from the fold change values of only the fold change in the larger set of ICU/healthy means that were the 50 highest increase in expression or the 50 lowest decrease in expression. Some genes had zero means and when avoiding the 'not a number' (NaN) when dividing by zero, since there was no change those genes were forced to have fold change values of 1 to not show change and those genes with fold change values where the disease mean was a value greater than zero but the healthy mean was zero, were forced to keep the disease mean to show the change from healthy to disease state in the gene expression without having an 'Inf' when dividing by a very small number like 0.00000000001 which if you know your math increases indefinitely without bounds to infinity. Rstudio was used and the genecards.org definitions pulled from the web and combined to the Ensemble gene ID in favor of their gene symbol using the genecards2.R script as a source file make in previous files at my Github repository: <https://github.com/JanJanJan2018/RA-simulating-COVID-study-analysis> as well as other files that are under 25 mb in size. The study used for this analysis is GSE152418 from NCBI's grant funded in full or partially by the NIH in research for data commons and research for the public. That link is: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152418>.

Note if you are new to analysis, look for changes or outliers in the comparisons, or those genes on the outskirts of the common pattern. They usually hold clues to genes leading to pathogenesis of a disease or treatment effect. They could also show up more or less defined in gender groups or elderly versus younger groups. The hovering details will give the Entrez, genecards, or uniprotkb gene summary to show what studies have discovered about those genes so far. Every image links to the Tableau Public repository where they were uploaded. The title is descriptive of what is being shown. In the interactive charts you can play around with the values to magnify by excluding certain values, or download Tableau and all the data by right clicking and selecting the option to download the data, open up your own Tableau Public Server free edition and make sure the data types are how you want them to be. The visualizations created with Tableau are self explanatory if you watch an hour or two of the free tutorials Tableau offers.

Have fun exploring these genes that could hold clues to gender and age differences in case severity of COVID-19 immunity and pathogenesis.

<https://public.tableau.com/profile/janis5126#!/vizhome/GSE152418-COVID19-FoldChangeTopGenes/COVID19_GSE152418?publish=yes>

Figure 1a: Bar chart of the fold change values of each case of COVID-19 as moderate, severe, or ICU compared to healthy samples.

<https://public.tableau.com/profile/janis5126#!/vizhome/GSE152418-COVID19-FoldChangeTopGenes/COVID19_GSE152418?publish=yes>

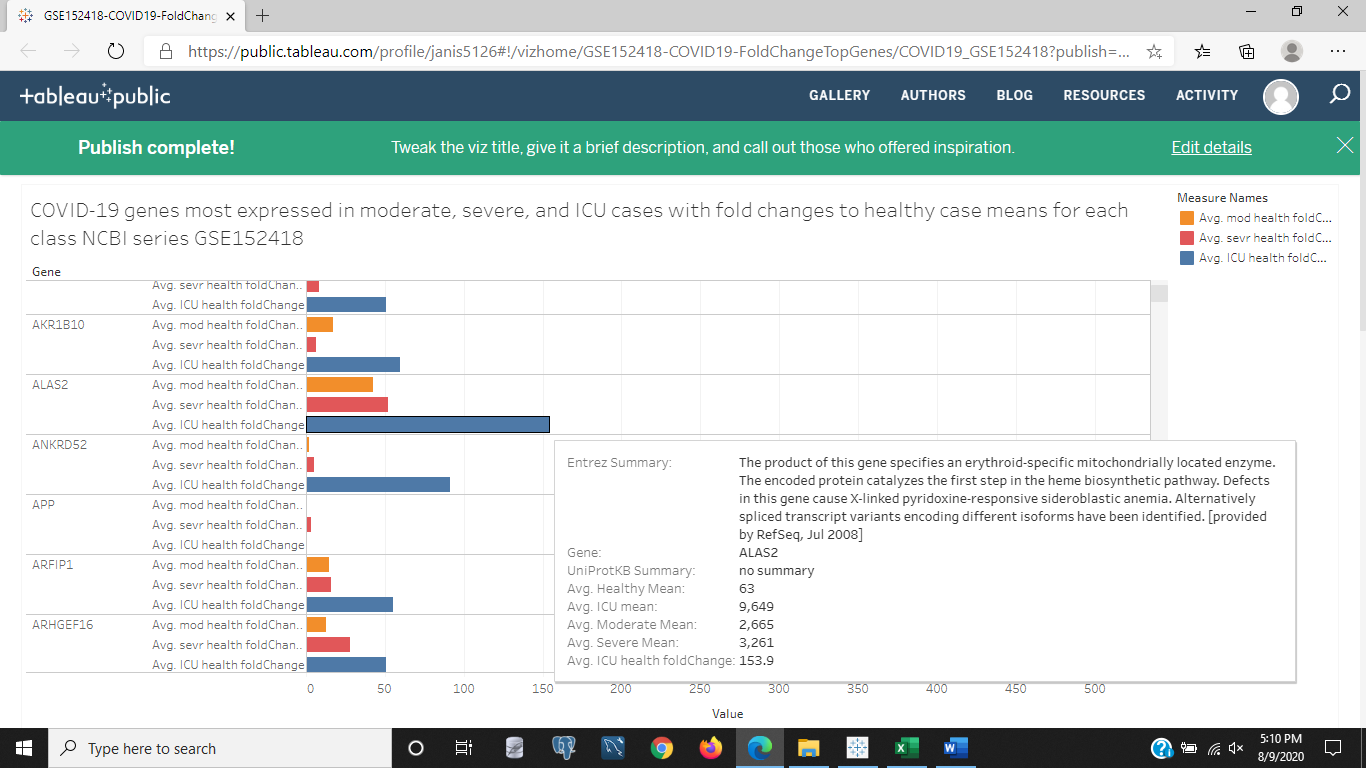


Figure 1b: Same bar chart as in Figure 1a except a different gene is being hovered on for details. The ICU/healthy fold change of ALAS2 shows a dramatic change in the ICU case versus the moderate and severe cases. This is a blood gene for red blood cells that occurs in the cell for blood biosynthesis.

<https://public.tableau.com/profile/janis5126#!/vizhome/GSE152418-COVID19-FoldChangeTopGenes/COVID19_GSE152418?publish=yes>

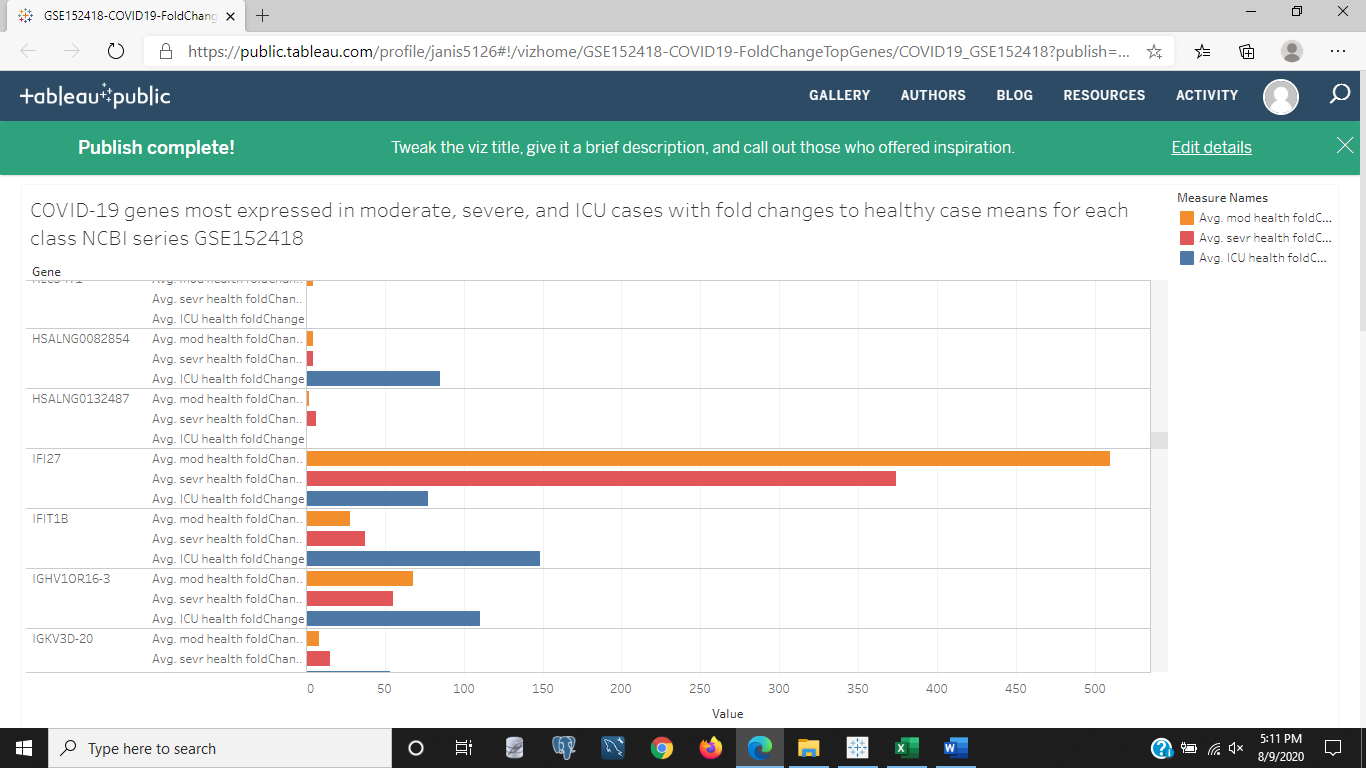


Figure 1c: This bar chart is also the same bar chart in Figures 1a and 1b, but scrolled down to show a gene in the moderate/healthy fold change that is really high and so is the severe/healthy value compared to the much lower ICU/healthy fold change value. This gene is IFI27.

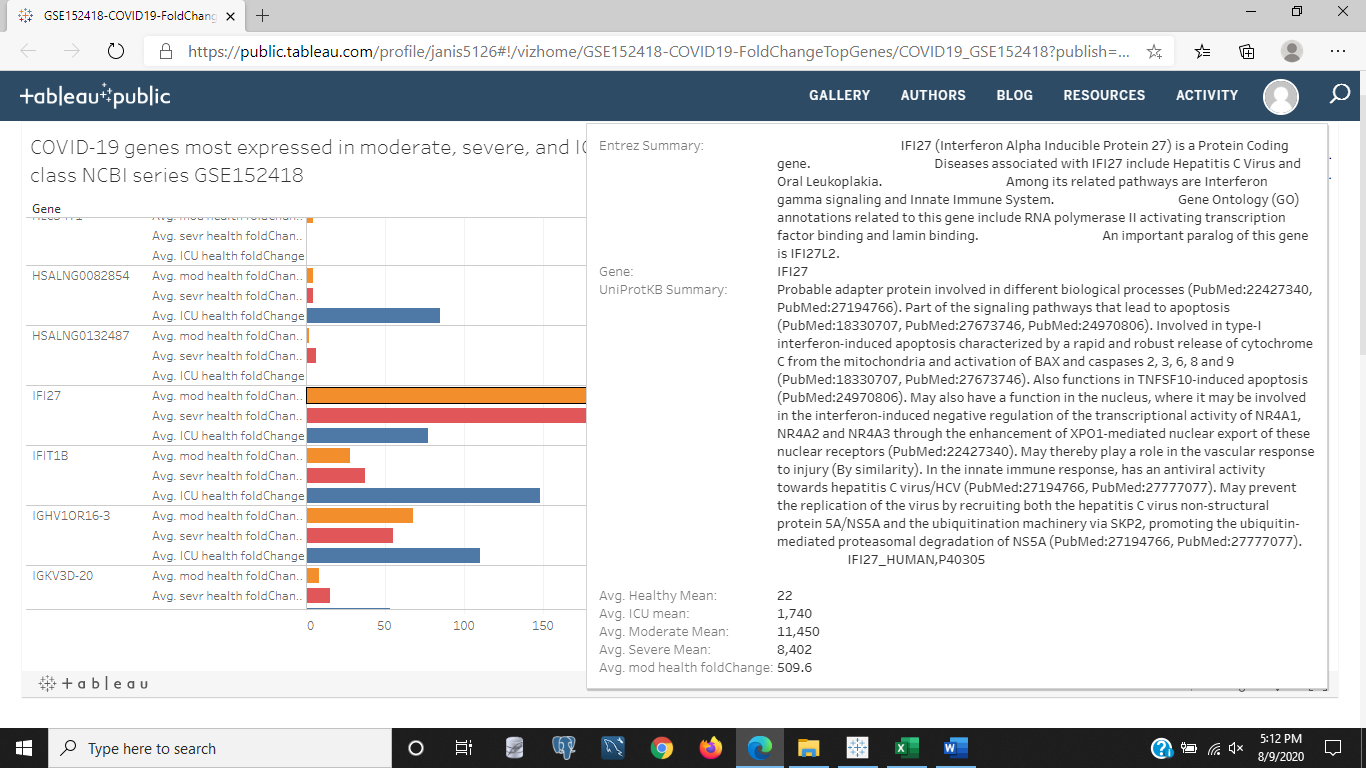


Figure 1d: In this image, the definition for IFI27 is given when hovering over the image in Figure 1c. This gene is said to be part of the signaling pathway that leads to apoptosis or cell death. This means the body in these early stages of COVID-19 know they have been breached by an antigen and need other cells to destroy themselves and any antigens they might have in them by likely sending in clast cells and other inflammatory interleukins to clean up the antigen after it is destroyed and the healthy cells that were ordered to destroy themselves to protect the surrounding tissue and health. Blood is made in the bone marrow.

<https://public.tableau.com/profile/janis5126#!/vizhome/GSE152418-COVID19-ICU-and-healthymeansscatter/ICU_healthy_scatterLOG2?publish=yes>



Figure 2a: This image is of a scatter plot of the log2 scaled mean values of the healthy mean gene expression values to the ICU grade COVID-19 mean values. Not many cells deviate from the center line diagonally crossing the image between scatter points of genes, but there are some genes in the lower left that stay the same in healthy cases, but increase more in the ICU grade COVID-19 as they form a horizontal line along the x-axis representing the ICU means in this image the diseased state.

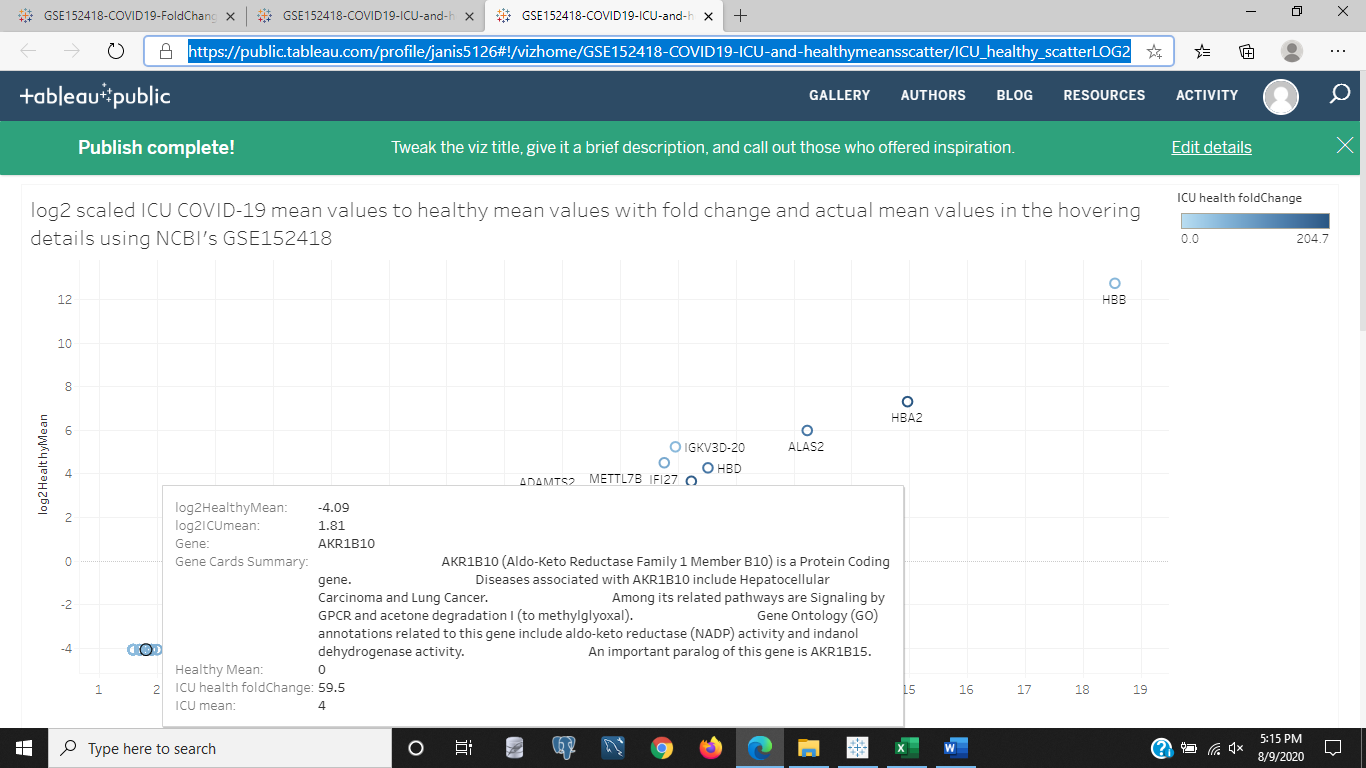


Figure 2b: We can see one of those genes in Figure 2a in the above image as a gene that increased in ICU compared to healthy conditions. This gene AKR1B10 has disease associations with liver and lung cancers in its likely mutated states or in high expressions. That latter information wasn't made clear in the gene summary from Gene Cards when hovering over this scatter point in the lower left of the chart.

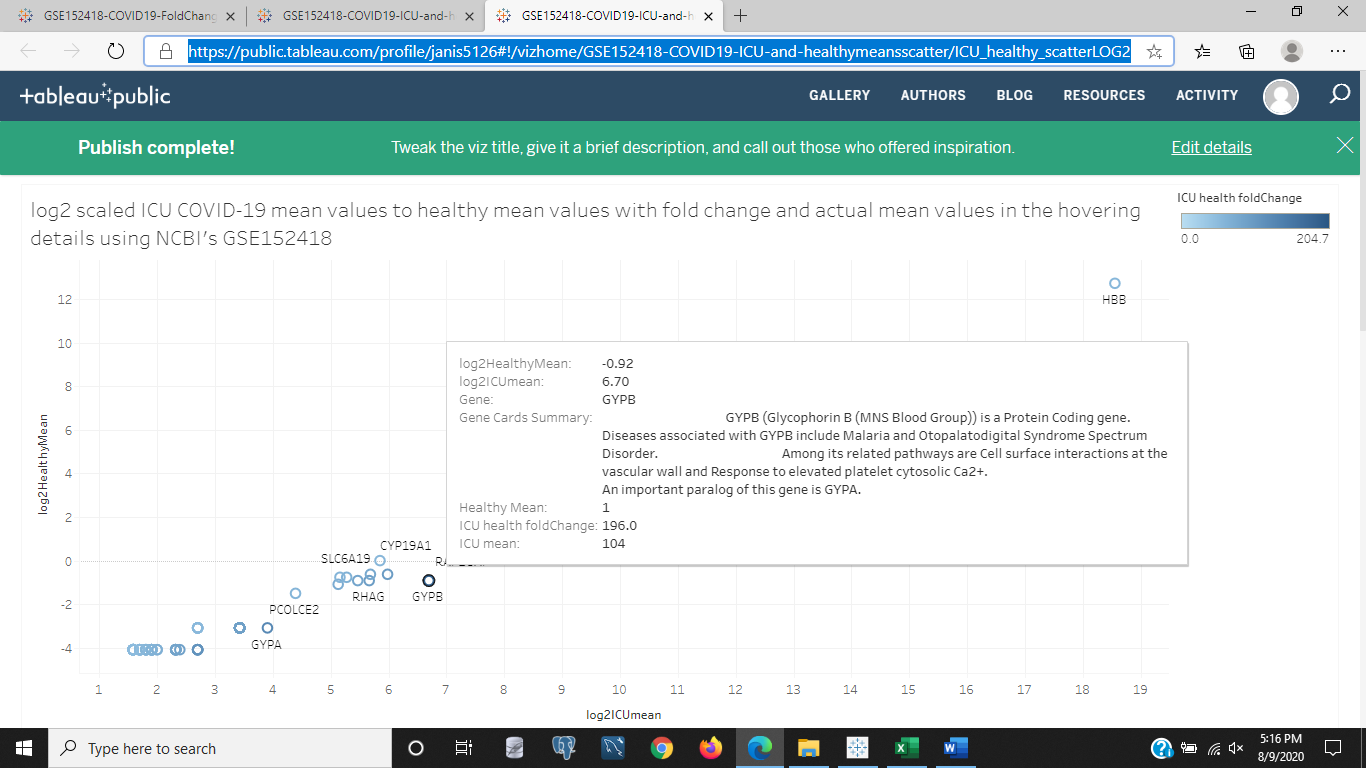


Figure 2c: In the above image, it is the same chart as Figure 2a and 2b but because the fold change value is indicated on this particular scatter point as high for the ICU/healthy fold change it was selected. This gene is GYPB and it is associated with vascular cell wall interactions, elevated blood platelets, and Malaria. Since it is elevated in the diseased state compared to the healthy state, this could make it important for fighting COVID-19 or identifying a patient as possibly having COVID-19.

<https://public.tableau.com/profile/janis5126#!/vizhome/GSE152418-COVID19-log2meansmoderatetohealthy/moderate_healthy_scatterLOG2?publish=yes>

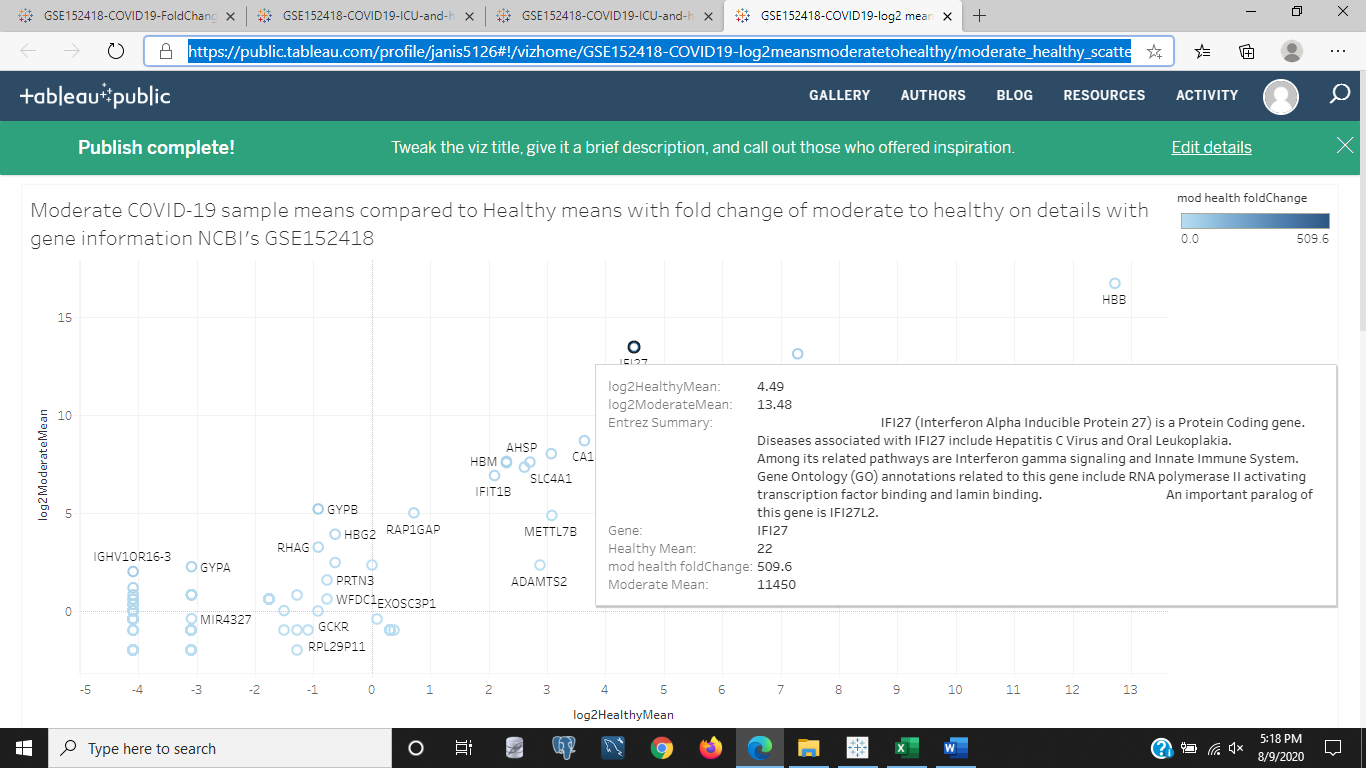


Figure 3a:This is a scatter of the Moderate cases of COVID-19 to the healthy cases by mean ratio comparisons per gene. All of the moderate cases were females and so couldn't be compared in upcoming images by gender to male gene expressions. The gene being highlighted and hovered on is IFI27 because it has a slightly high fold change indicated by the scatter point colors being darker than their neighbors. So the moderate COVID-19 cases had more of this gene being expressed than the healthy cases over all. This gene is part of the naturally born immune system functions and diseases associated with this gene's mutations are involving mouth sores due to white (leuk) plaque and liver disease called Hepatitis C viral infections. The over expression of this gene doesn't mean the patient has those diseases, but it is elevated in COVID-19 infection and it is also an interfeuron or immune cell for combating infections in the body.

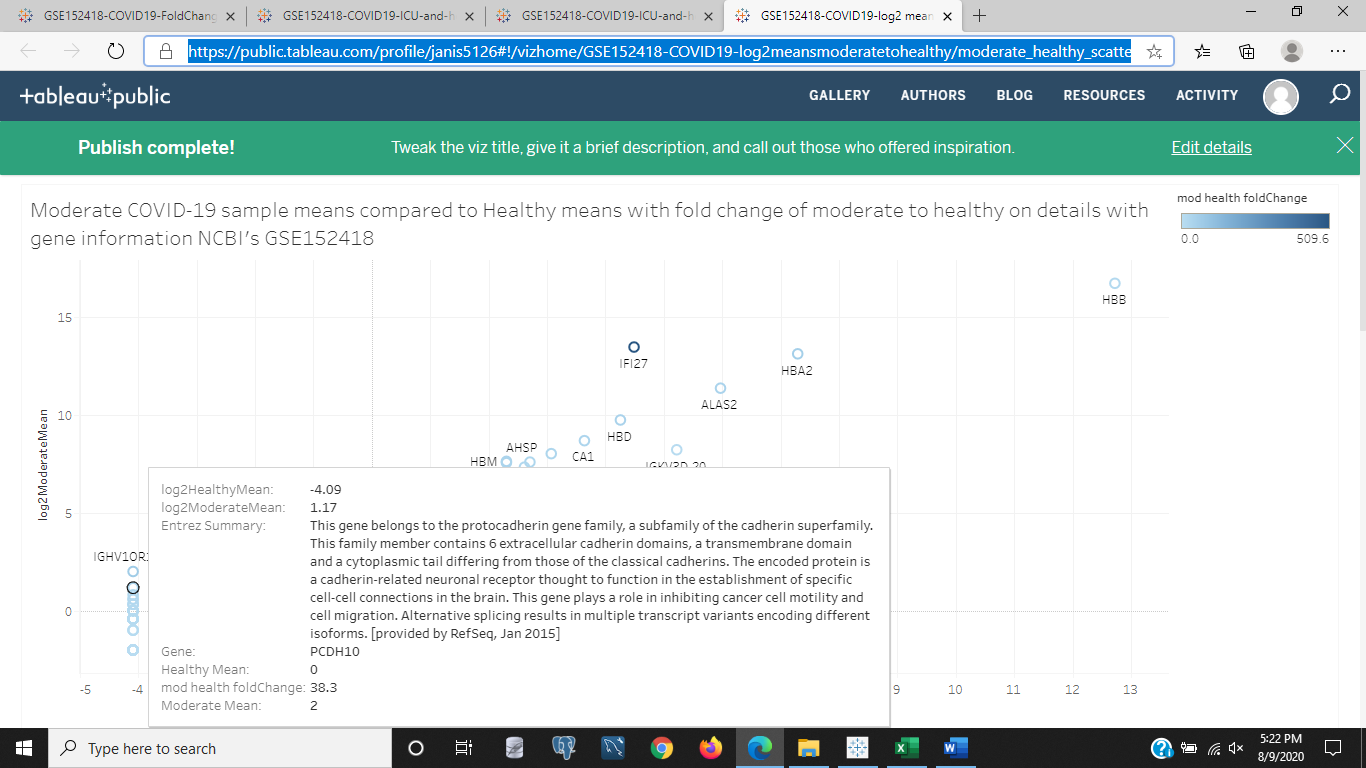


Figure 3b:This image is the same chart as Figure 3a, the moderate/healthy means with darker scatter points having higher fold change values or higher moderate mean gene expression values across all moderate samples compared to the mean gene expression values across all healthy samples. The gene being looked at is one of the far left genes that stay the same in healthy samples but increases in the moderate samples. All values are log2 scaled to normalize the skewed outliers as are the images in Figure 3a. This gene, PCDH10, is said to be a cadherin gene that connects certain brain neuron activity that specifically have to do with inhibiting cancer cell motility and cell migration. implying it needs to be present to help prevent malignant cancers from spreading to neighboring cells, tissues, and organs.

<https://public.tableau.com/profile/janis5126#!/vizhome/GSE152418-COVID19-log2meansseveretohealthyscatter/severe_healthy_scatterLog2?publish=yes>



Figure 4a:This image is of the chart comparing the severe COVID-19 cases to the healthy mean cases, but the axis is flipped so that the healthy means of the genes are on the y-axis instead of the x-axis as was shown in Figure 3a and 3b. We see that most of the genes are showing more gene expression in the lower left of the image and that the darker genes seem to only include IF27 for having higher severe/healthy fold change ratios. We saw in Figure 3a and 3b that IF27 has to do with preventing cancer cells from spreading. It is elevated in the moderate and severe cases of COVID-19. All values are the log 2 scaled mean values of the genes in the severe and healthy cases.

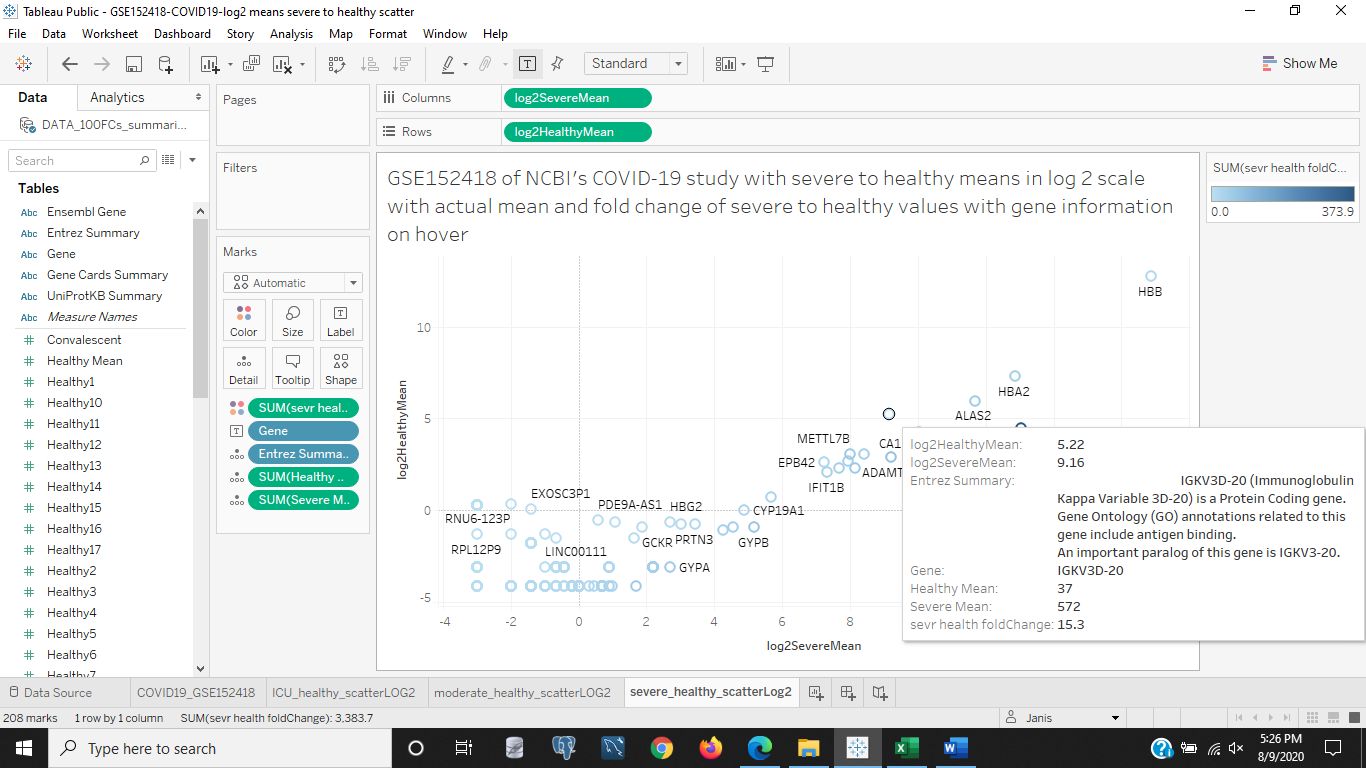


Figure 4b: The image above is the same chart as the image in Figure 4a, but with the IGKV3D-20 gene hovered over. This gene is an antigen binding gene which means it looks for foreign invaders or infectants in the body that could be causing the patient to be sick and binds to it perhaps to destroy it, copy it, send the information to other immune network cells in the body, etc. Since it is a globulin gene it deals with protein, like hema (blood) globulin (protein), that are related to the immune system.

<https://public.tableau.com/profile/janis5126#!/vizhome/age_FCs_severeCOVID19/age_FCs_severeCOVID19?publish=yes>

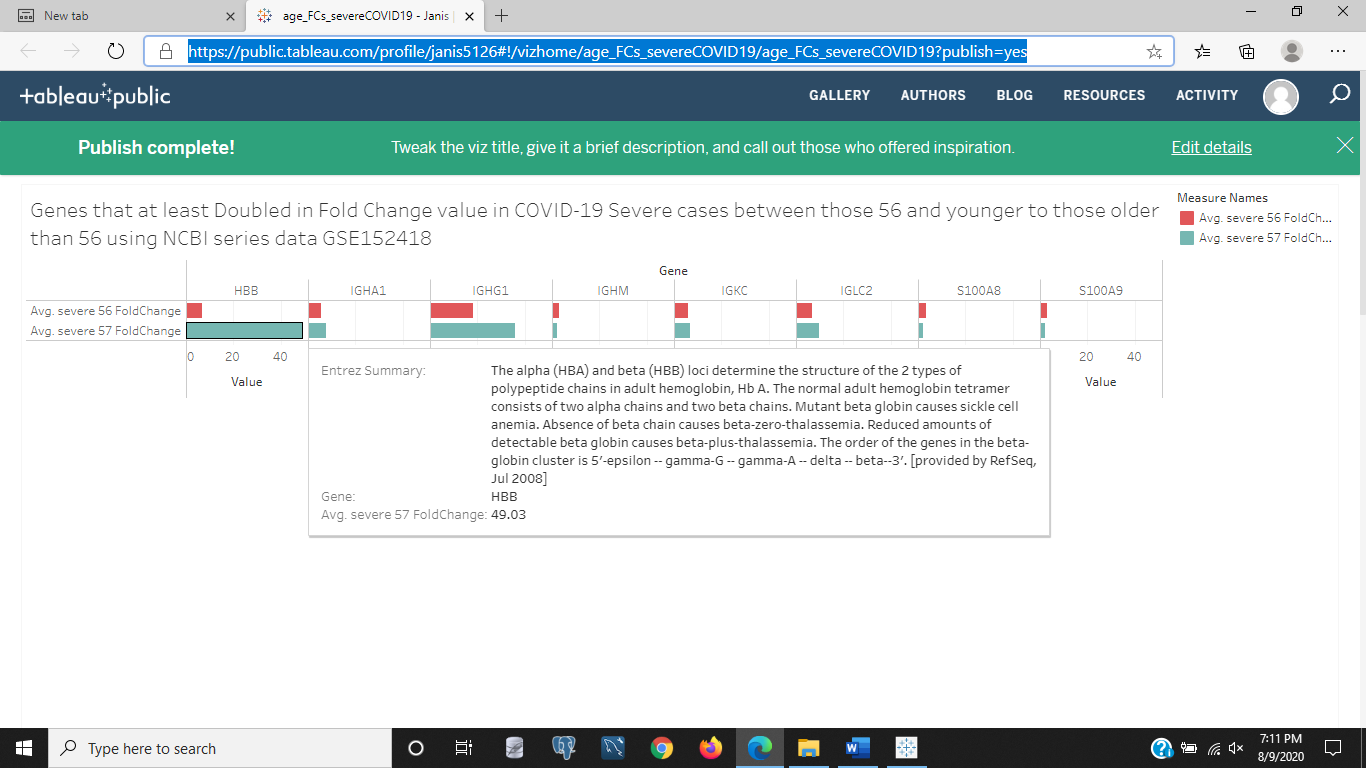


Figure 5a: The above image is a bar chart of the two age groups of those 56 and younger and those older than 56 in our samples of healthy patients and those with Severe COVID-19 cases. These values are fold change values, and not log2 scaled gene expression means as the previous images. Only those genes having a fold change of higher than 2 in value of the severe/healthy mean values are shown. We see that some genes have a much higher expression in the 57 and older group of patients, such as HBB, a gene associated with hemoglobin using the Entrez gene summary. Another gene is IGHG1 that hasn't been highlighted and IGKC. Only two genes are expressed more in those 56 and younger than those older than 56, S100AB and S100A9.

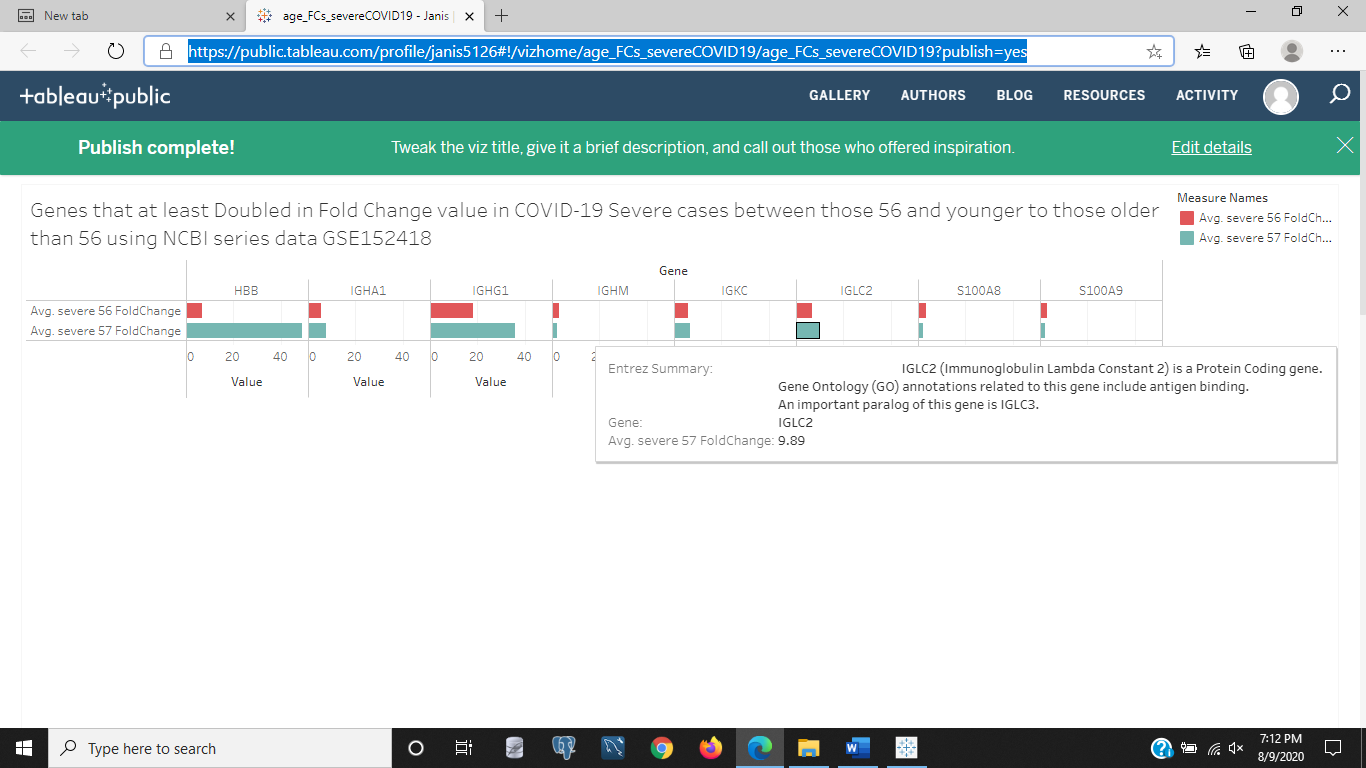


Figure 5b: The image above is for the same bar chart in Figure 5a of the severe/healthy fold change genes that at least doubled in COVID-19 severe cases compared to healthy cases. The gene highlighted is IGLC2. The Entrez gene summary says this gene is an immune protein that binds to foreign invaders or antigens in the body.

<https://public.tableau.com/profile/janis5126#!/vizhome/age_log2_severeCOVID19/age_log2_severeCovid19?publish=yes>

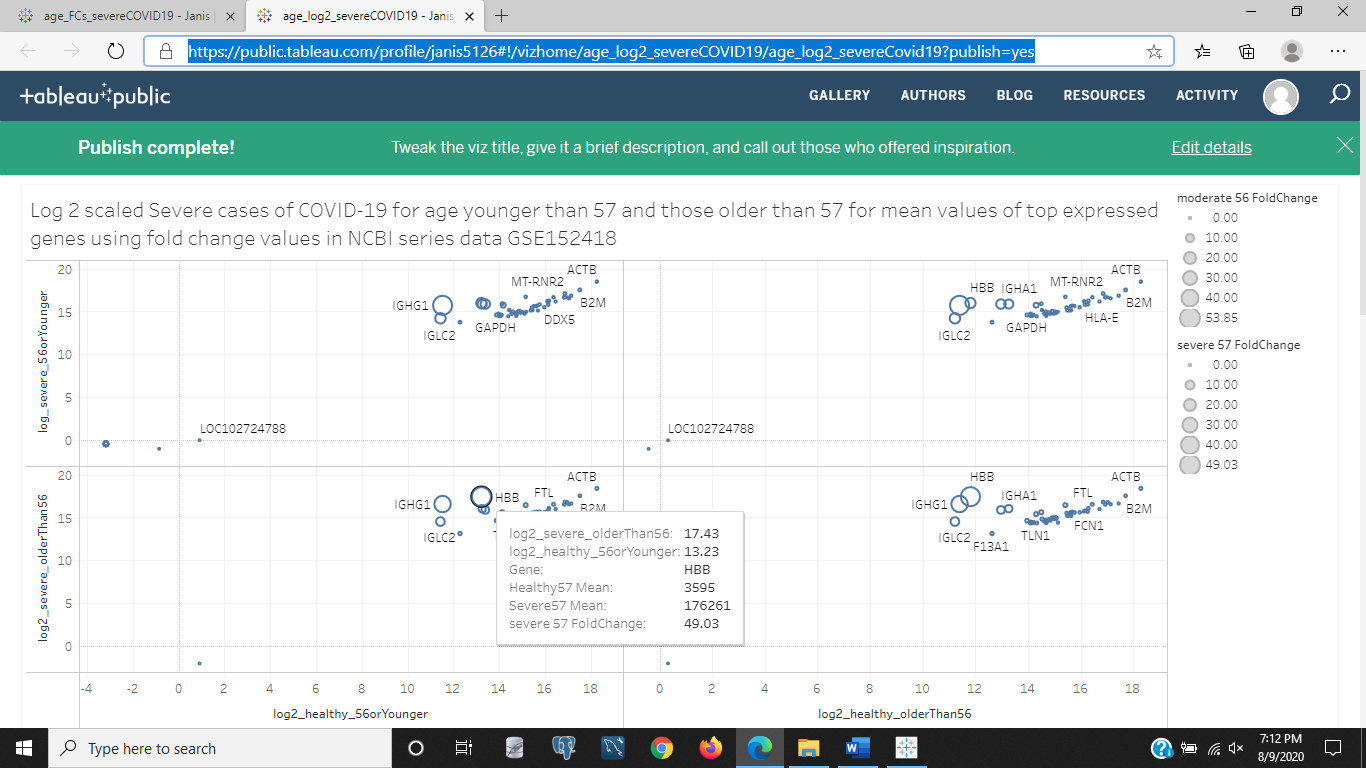


Figure 6a: The above image is a log2 scaled scatter plot of the gene expression means in two age groups on the healthy (x-axis) of those 56 and younger and those older than 56 to the severe COVID-19 log2 scaled gene expression means in the same two age groups. In each of the four segments of the chart, the lower segment in the left shows the gene expression means scaled for the younger group to the older group with HBB highlighted for hovering details. This gene is a larger scatter indicating those older than 56 show more of this gene expressed than those in the healthy group younger than 56 or 56. We saw that this gene HBB in Figures 5a is a hemoglobin gene using the Entrez gene summary. Why is there more of it expresssed in the elderly? A blood carrying protein. Maybe because they are taking more iron rich vitamins than younger patients and have more of this protein in their system. It might not be related to COVID-19, but more blood or red cells carry more oxygen and more processes in the body require oxygen to bind to proteins and do whatever work they do like break down cellular waste (osteoclasts) or build and repair damaged tissue (osteoblasts) as examples.

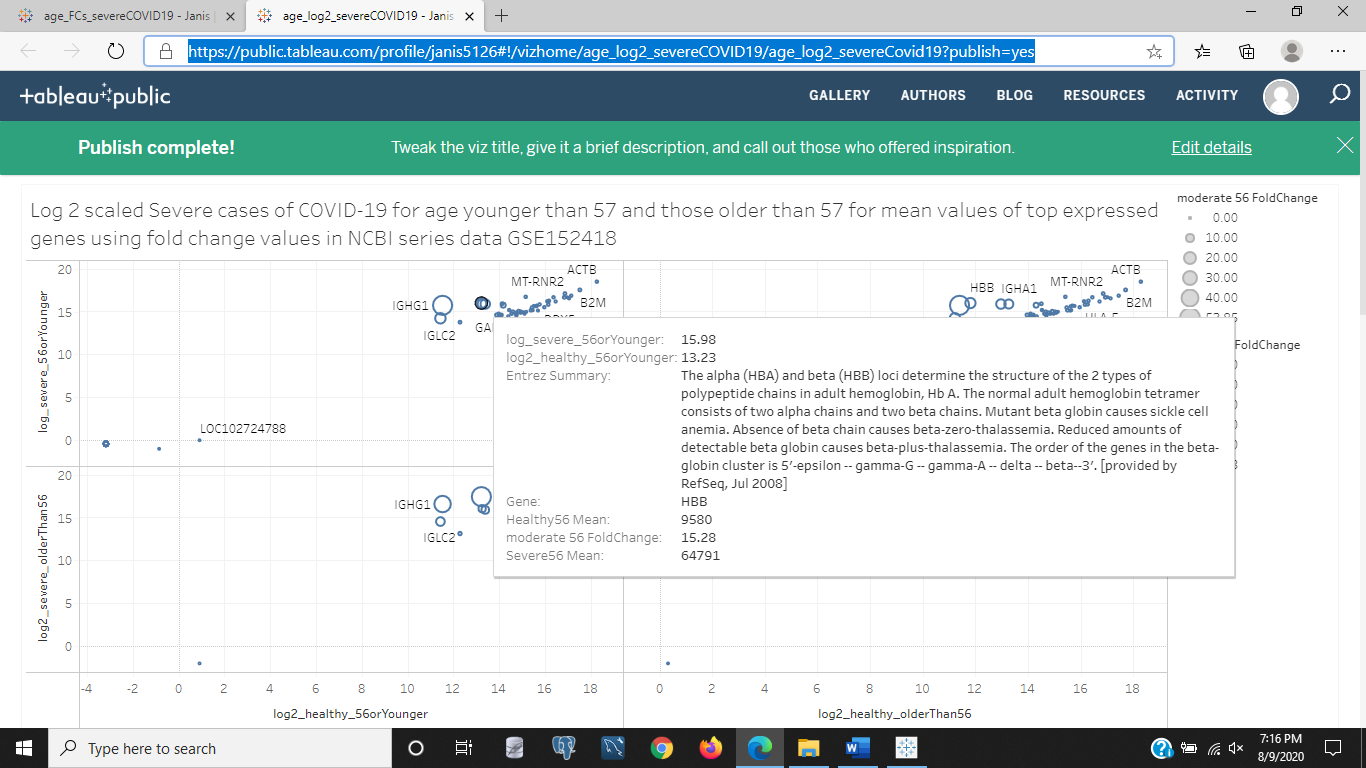


Figure 6b: In the above image, it is the same scatter chart of age group comparisons for the Severe COVID-19 cases to the healhty cases in those younger than 57 and those 57 and older. This time we see the upper left segment of the chart hovered over HBB in the younger age group compared to the healthy age group. It is much smaller in fold change than the older severe COVID-19 cases shown in figure 6a.

<https://public.tableau.com/profile/janis5126#!/vizhome/age_FCs_ICU_Covid19/age_FCs_ICU_Covid19?publish=yes>

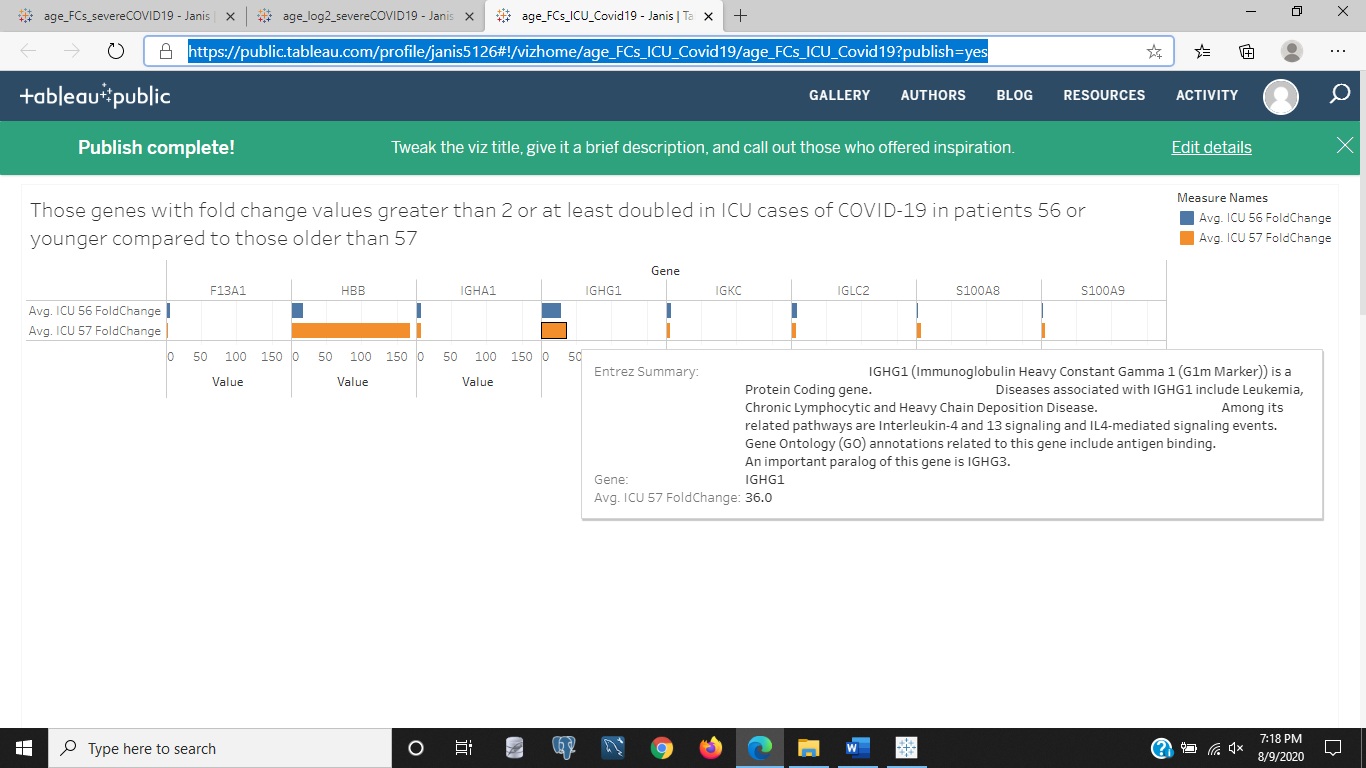


Figure 7: The above image is a bar chart of ICU COVID-19 cases compared to healthy cases in the two age groups in Figure 6a and 6b of those 56 and younger to those older than 56. We see the age groups color coded for fold change values, and only those genes having a fold change of 2 or greater to show the genes that doubled in the disease to healthy state are shown. We see that in the older group of ICU cases the HBB and IGHG1 genes are significantly higher than in the younger group. The other genes don't show much change, but S100A8 and S100A9 do show a slightly increased gene expression in the older group as well. The gene highlighted is IGHG1 that has an Entrez gene summary that says this gene is an immunity gene that involves interleukin signaling and is also related to the blood cancer Leukemia. This gene is an antigen binding gene like IGI27. So far, most genes having higher than normal gene expression out of 56,000+ genes are immunity genes.

<https://public.tableau.com/profile/janis5126#!/vizhome/age_log2_ICU_Covid19/age_log2_ICU_Covid19?publish=yes>

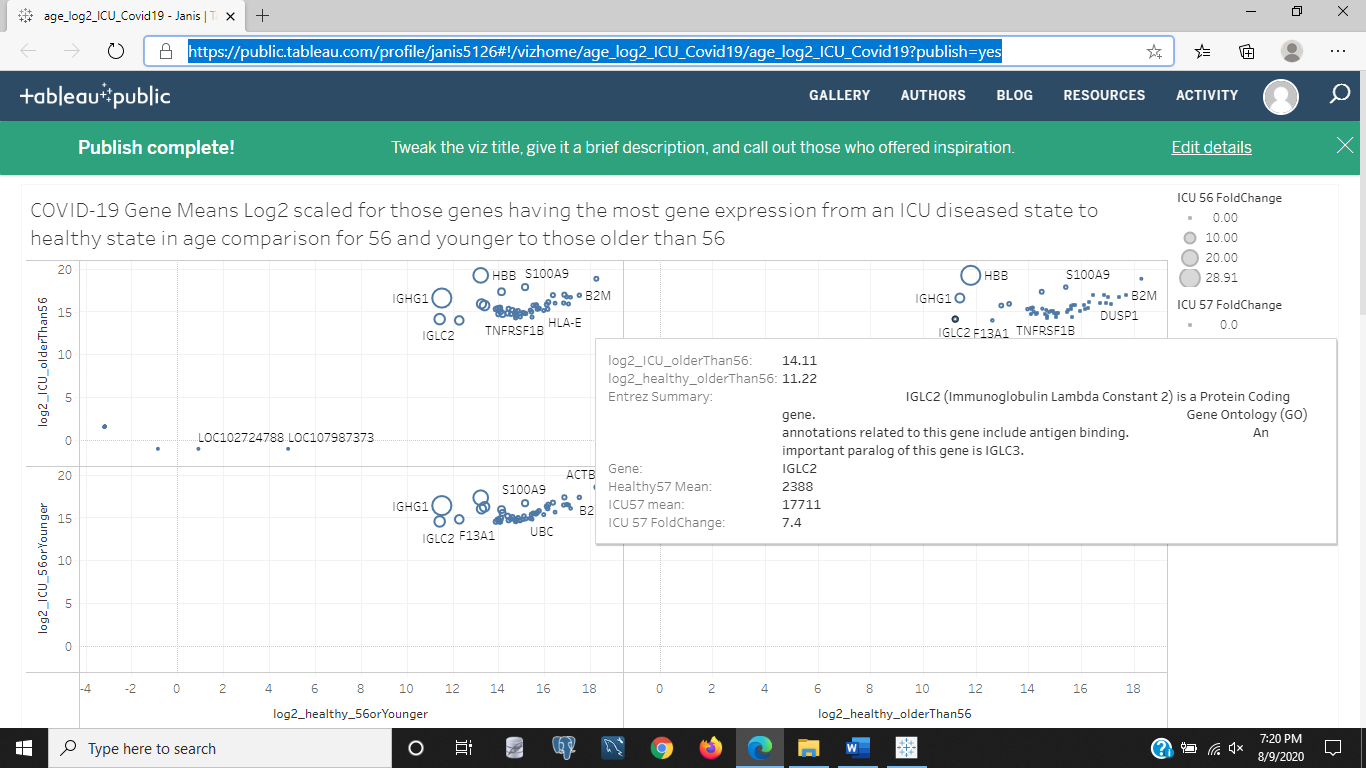


Figure 8a: The image above is similar to the image in Figure 7, because it is a scatter plot by age group but it is comparing the ICU COVID-19 cases to healthy cases as a ratio of log2 scaled gene expression means with the healthy groups along the x-axis and the diseased ICU age groups on the y-axis. We see that IGLC2 is hovered on in the upper right corner. The younger diseased state is on the bottom in this image, and the older group of ICU diseased state is on the top. This segment of the chart compared the ICU/healthy gene expression values of those patients older than 56. IGLC2 is an immune protein involved in antigen binding to help remove contaminents and foreign invaders in the body.

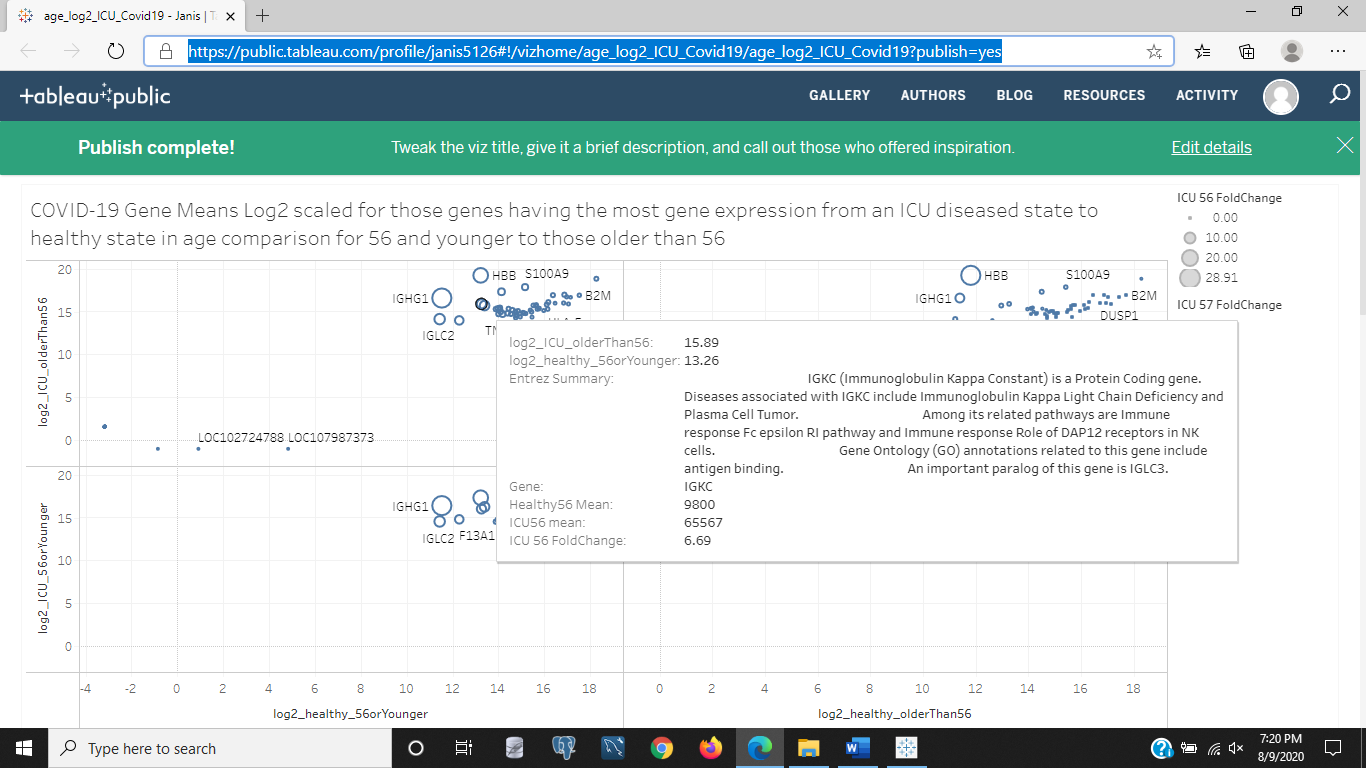


Figure 8b: The image above is the same ICU/healthy age group gene expression log2 scaled values with size indicated fold change values as an added dimension. The highlighted gene is in the segment of the chart that compares the ICU older group to the healthy younger group in the upper right of the image. IGKC is the gene highlighted, and its Entrez summary says it is also an immunity protein involved in natural killer (NK) immune cell signaling and antigen binding to detect and remove foreign invaders and contaminants in the body.

<https://public.tableau.com/profile/janis5126#!/vizhome/age_FCs_moderate_Covid19/age_FCs_moderate_Covid19?publish=yes>

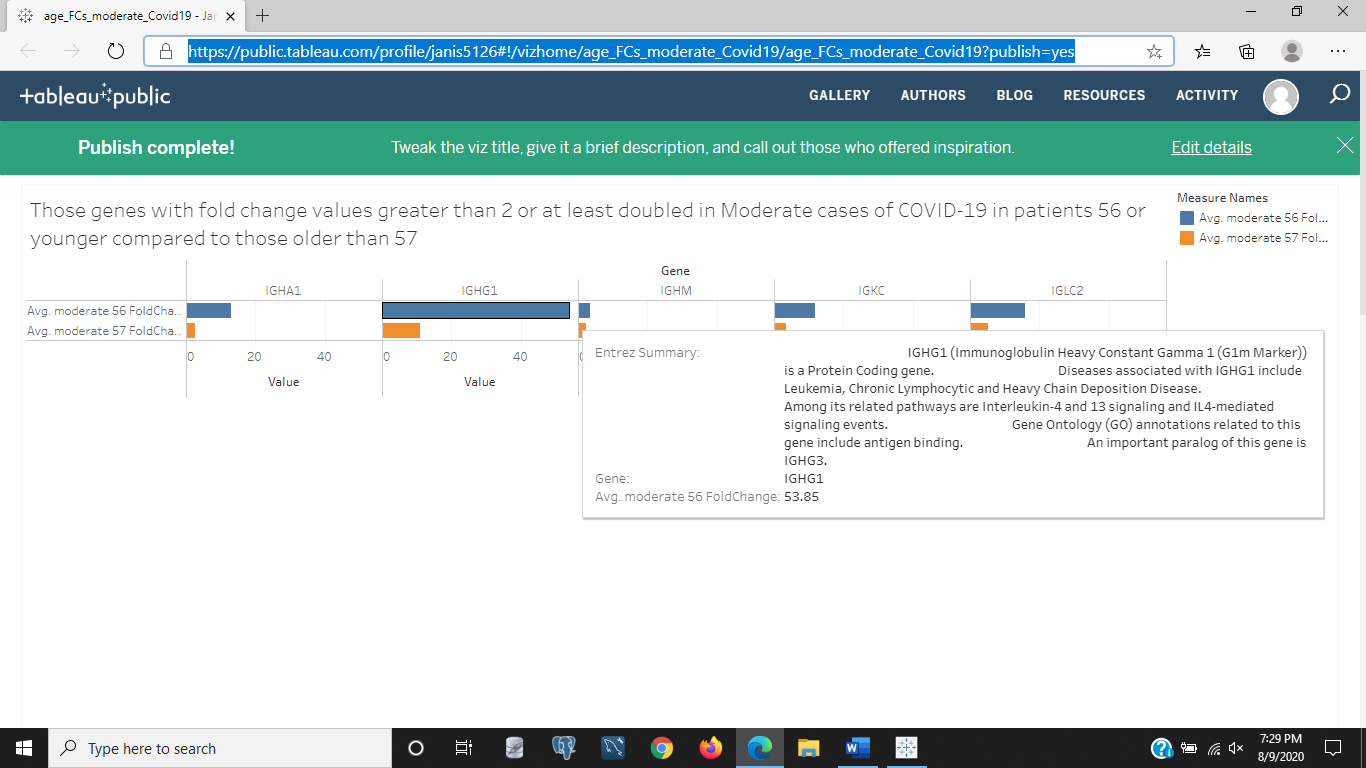


Figure9a: The above image is of the chart comparing those genes having fold change of moderate COVID-19/healthy gene expression means in the two age groups of 56 and younger to those older than 56. In the above image we see that in the moderate cases, the top genes having more change in the moderate COVID-19 cases are different than those shown in the ICU and severe COVID-19 cases and with the younger group having much more gene expression in the diseased to healthy state comparisons. One gene is hovered on for details, IGHG1, an immune protein that is related to Leukemia in its variant mutations and is also an antigen binding gene that locates foreign invaders and contaminants in the body to be eliminated. Another gene related to leukemia, we saw in the earlier images, .was this same gene but in the older age group of ICU COVID-19 to healthy fold change values in Figure 7. We see in moderate cases the younger group has more of this gene being expressed than the older group and the reverse for the ICU COVID-19 cases.

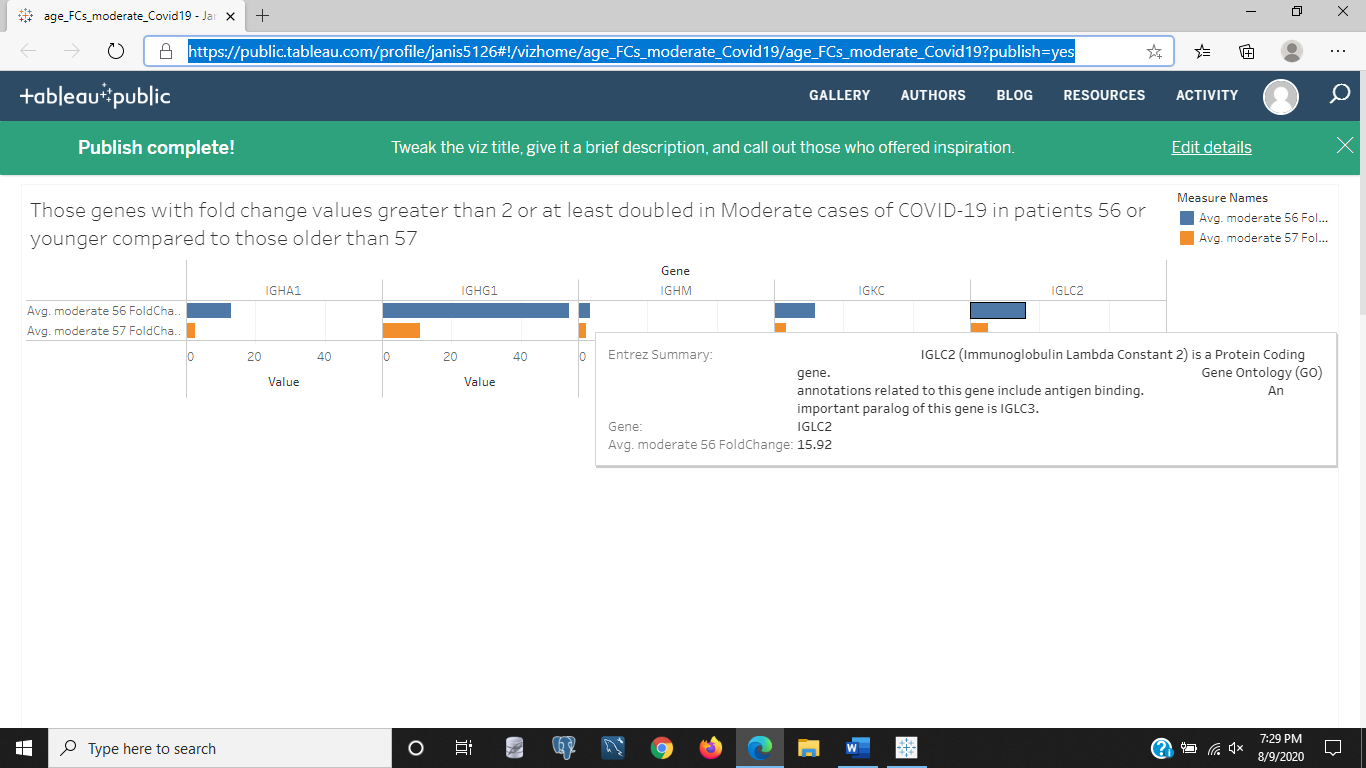


Figure 9b: The image above is the same bar chart comparing moderate COVID-19 to healthy fold change values in Figure 9a. But the gene being hovered on for details is IGLC2, which is another immune protein having a fold change value that at least doubled in the moderate to healthy cases. All displayed genes in the above image are showing much more gene expression increases in the younger group than the older group. Moderate is much less problematic as far as symptoms go for COVID-19 severity, and that all these genes are immune genes, and they are elevated in the younger group compared to the older group shows that the younger group is recognizing and fighting the virus early and that the older group might not be recognizing the virus to fight it early. It would be on interest to know if the timeline of these older patients with moderate COVID-19 symptoms end of developing more severe symptoms. The study didn't show a timeline for GSE152418, and only took measures of the perirpheral blood for each case of COVID-19, so we don't have that information to make an assumption on older patients' immune response not working as effectively or working well and not causing deadly inflammation by over-responding. There are a few differnt scenarios that these results imply that we don't have the data on determining.

<https://public.tableau.com/profile/janis5126#!/vizhome/age_log2_moderate_Covid19/age_log2_moderate_Covid19?publish=yes>



Figure 10a: The above image is the moderate COVID-19 to healthy comparison of log2 scaled gene expression mean values per gene in one of two age groups of either 56 and younger or older than 56. The healthy group in on the x-axis, and the moderate cases are on the y-axis. The older group on top and younger on the bottom of the y-axis while the younger group of healthy patients are in the lower left and older healthy patients to the lower right. The highlighted gene is IGHA1, an immune protein that detects antigens or foreign contaminants to eliminate from the body. There is more gene expression of this gene in the younger group of moderate COVID-19 cases than those healthy older patients. The comparison is in the lower right of the chart comparing healthy older patients to younger moderate COVID-19 patients' blood samples. This gene has an added characteristic to the other immune protein antigen binding genes that says it responds to elevated platelets and at the vascular wall has cell surface interactions. There are different layers of an artery that begin with 'tunica' and medial, intermedial, and outer, but also thinned artery walls can cause artery bulges called aneurysms that risk bursting and causing blood leakage and consequently a lack of oxygen from the transporting red blood cells or platelets that carry oxygen to the cells. But that could have nothing to do with why this gene that associates to platelets and removing foreign contaminants is over expressed in younger moderate COVID-19 cases compared to healthy older cases. Cell to cell surface interaction in vascular systems could mean the capillary supplying and removing waste from organs and muscles.



Figure 10b: The above image is the same chart as in Figure 10a, except that the IGHA1 gene is highlighted in the upper right segment for the moderate COVID-19 cases in the older group compared to teh healthy older group. We see this gene is also elevated in older moderate COVID-19 cases as it was in the younger moderate COVID-19 cases when compared to the healthy older group, but by less increase than the younger group. This gene is an immune gene that removes foreign invaders or detects them by binding to antigens, and has the added quality that it is involved in cell to cell interactions in the vascular wall and responds to elevated platelets or red blood cells that carry oxygen.

<https://public.tableau.com/profile/janis5126#!/vizhome/fem_mod_COVID19/fem_mod_COVID19?publish=yes>

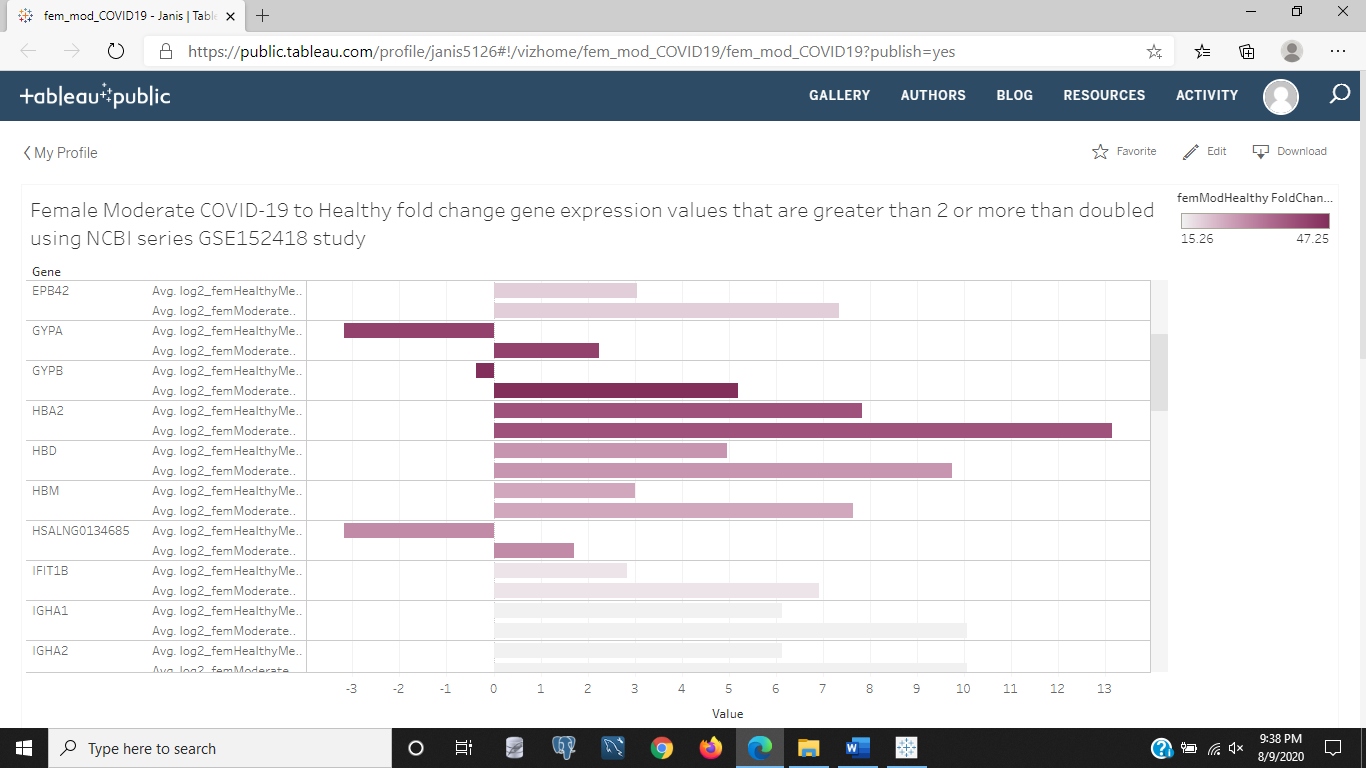


Figure 11a: The above image is the moderate COVID-19 fold change values as a color bar for the genes with higher fold change values of moderate/healthy being darker on the gender differences of only those females who had moderate COVID-19 symptoms compared to only those females in the healthy group by log 2 scaling of those separate gene expression means. The negative values, for those not accustomed to the log scale just means the e^x exponent of x is a negative value, but the gene expression value is not actually negative. This is a scaling device or method to normalize the data, when hovering on the genes you will see the actual gene expression values for each group of females of either moderate COVID-19 or healthy females. I.E. (e^(-1)) is 1/e or 1/2.72 approximately, hence a value greater than zero, but the exponent, x, is a negative value. This is a blog so I throw that informal education in to this explanation for why or how gene expression values are negative, because it is the log2 scaled value, so that outliers that are in 10s of thousands of measures higher than the average range don't skew the visualized data and make all the values a big blob with just that one outlier or even a few outliers compared to hundreds of other values in a much smaller range. Looking at the data, we see that HBA2 in the moderate female cases is much higher than the healthy cases, and taht the other gene HSALNG0134685 is also higher than the healthy cases. We don't see what those genes are, but know from having read the previous image summaries on HBB that HBA is also an immune protein, and if we click the image to visit the chart, we can hover over that gene and discover its known annotations or gene descriptions.

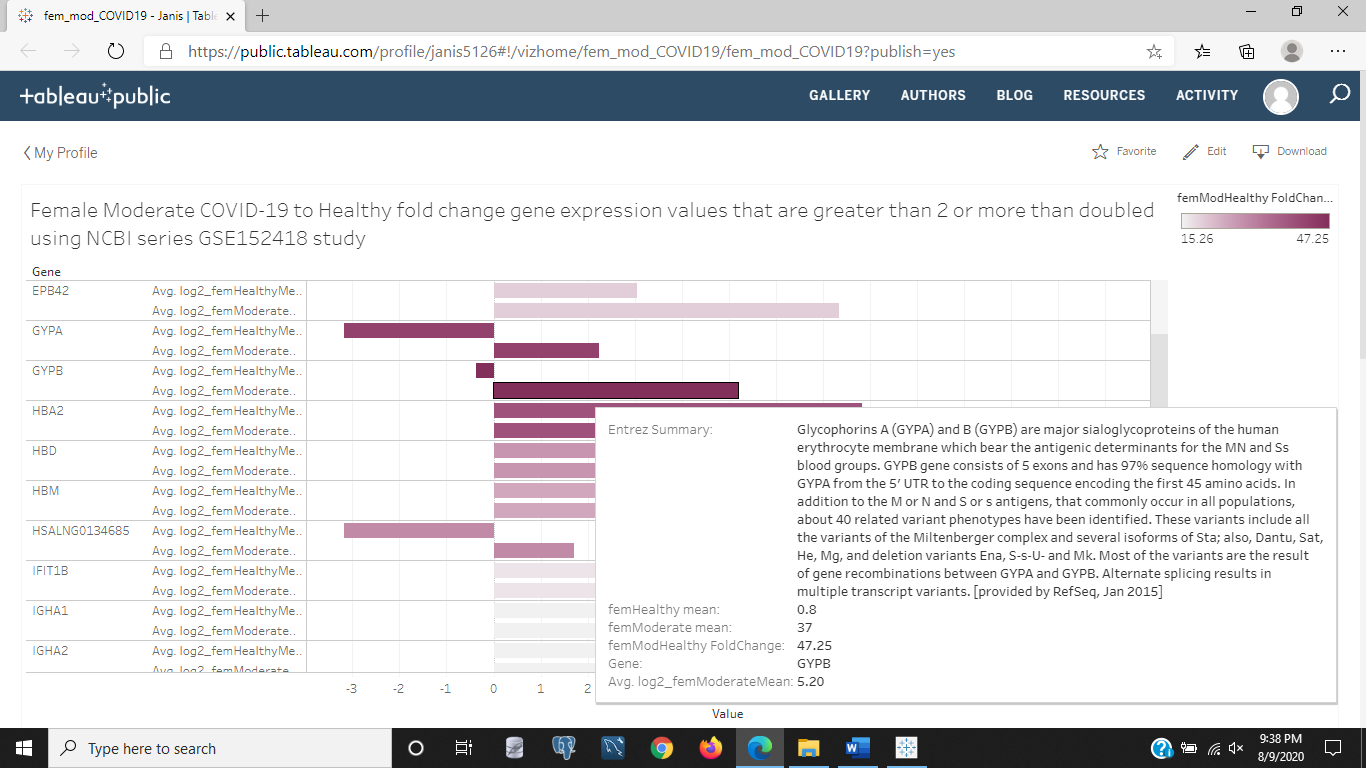


Figure 11b: The above image is the same chart as in Figure 11a, of the female moderate to healthy cases on a log2 scale. The hovered gene for detailed information shown in the image is GYPB, involved in the erythrocyte membrane or miniscule red blood cell membrane as well as other genetic identifiers for how it builds protein from its selection of amino acids.

<https://public.tableau.com/profile/janis5126#!/vizhome/fem_mod_COVID19_scatter/fem_mod_COVID19_scatter?publish=yes>



Figure 12: In the above image the moderate COVID-19 female cases to the healthy cases for gene expression means on a log2 scale of each gene's mean value across the samples is shown. The gene selected is EPB42, a red blood cell gene that is likely related to the red blood cell shape and mechanical functions. Red blood cells transport oxygen to cells and are called erythrocytes. There is almost 20x more gene expression in moderate COVID-19 cases compared to healthy cases.

<https://public.tableau.com/profile/janis5126#!/vizhome/fem_mod_COVID19_dashboard/fem_mod_COVID_dashboard?publish=yes>



Figure 13a: In the above image, the dashboard showing female moderate COVID-19 log2 scaled gene expression means with a color bar for fold change of moderate/healthy cases being darker in color is shown. The highlighted gene is EPB42, with the gene summary muted or not added. This is a dashboard and the bottom bar chart for this same data of genes will display the gene summary in Figure 13b a different gene will show its details when hovered over. We do see in the above image that it is expressed nearly 20x that of healthy cases in females. We saw this gene summary in Figure 12 of the chart of this data instead of the dashboard. It is an erythrocyte gene having to deal with the shape and mechanics of red blood cells.



Figure 13b: This is the same dashboard as in Figure 13 a, except it shows the bar chart portion with the log2 scaled gene expression means and color bar fold change for darker bars having higher fold change values in moderate/healthy gene expression means in females. The gene highlighted here is HSALNG0134685. There isn't much of a summary for this gene but it has a nearly 30% increase in gene expression by fold change value in moderate cases of COVID-19 compared to healthy female cases.

<https://public.tableau.com/profile/janis5126#!/vizhome/fem_severe_ICU_covid19_scatter/fem_severe_ICU_covid19_scatter?publish=yes>

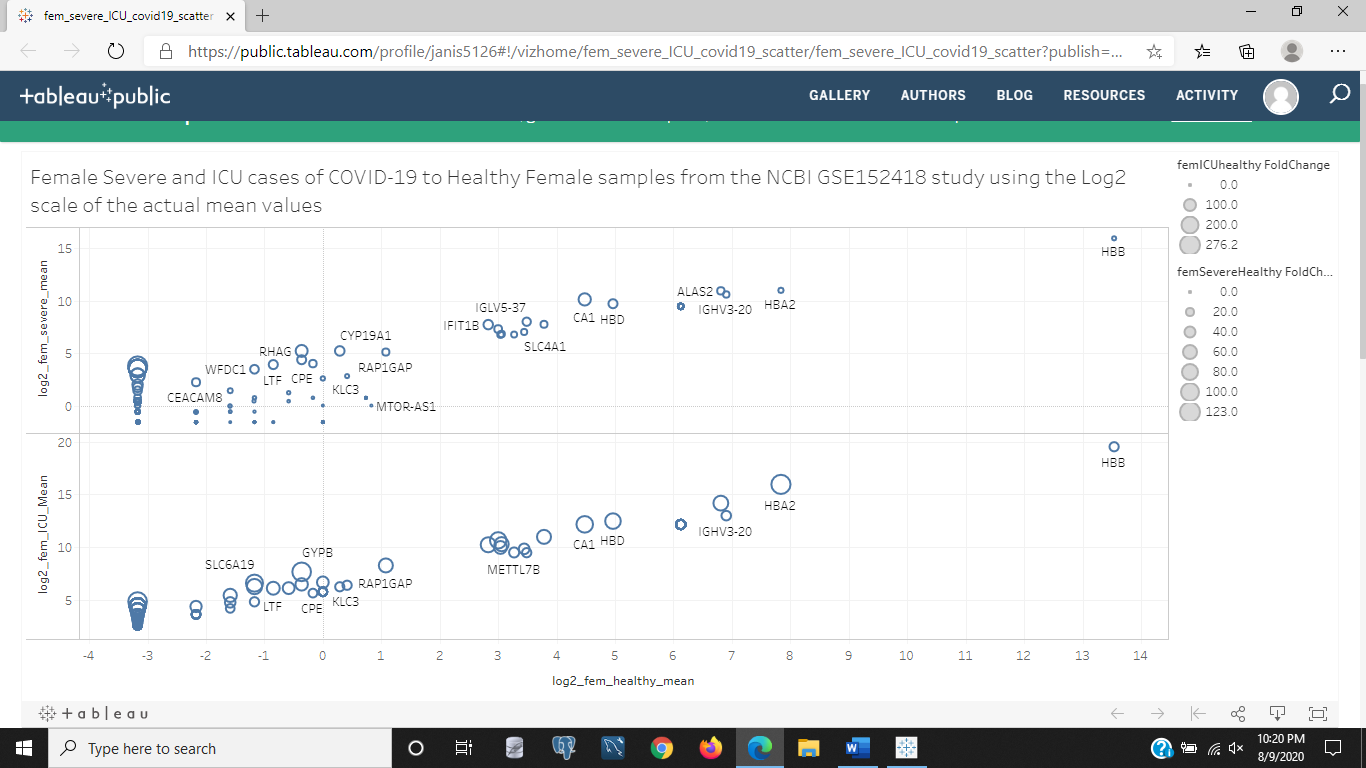


Figure 14a: In the above figure we see the ICU and severe COVID-19 log2 scaled gene expression means as scatters compared to the healthy gene expression means as scatters for females in this GSE152418 study on NCBI. Visually, we can inspect the differences and see that the lower segment for ICU cases has a much higher HBA2 fold change and larger fold change values in the center genes like METTL7B, and that there is a higher increase in the far left genes that stay the same value in healthy cases but increase in COVID-19 cases much more in the severe than ICU grades of COVID-19.

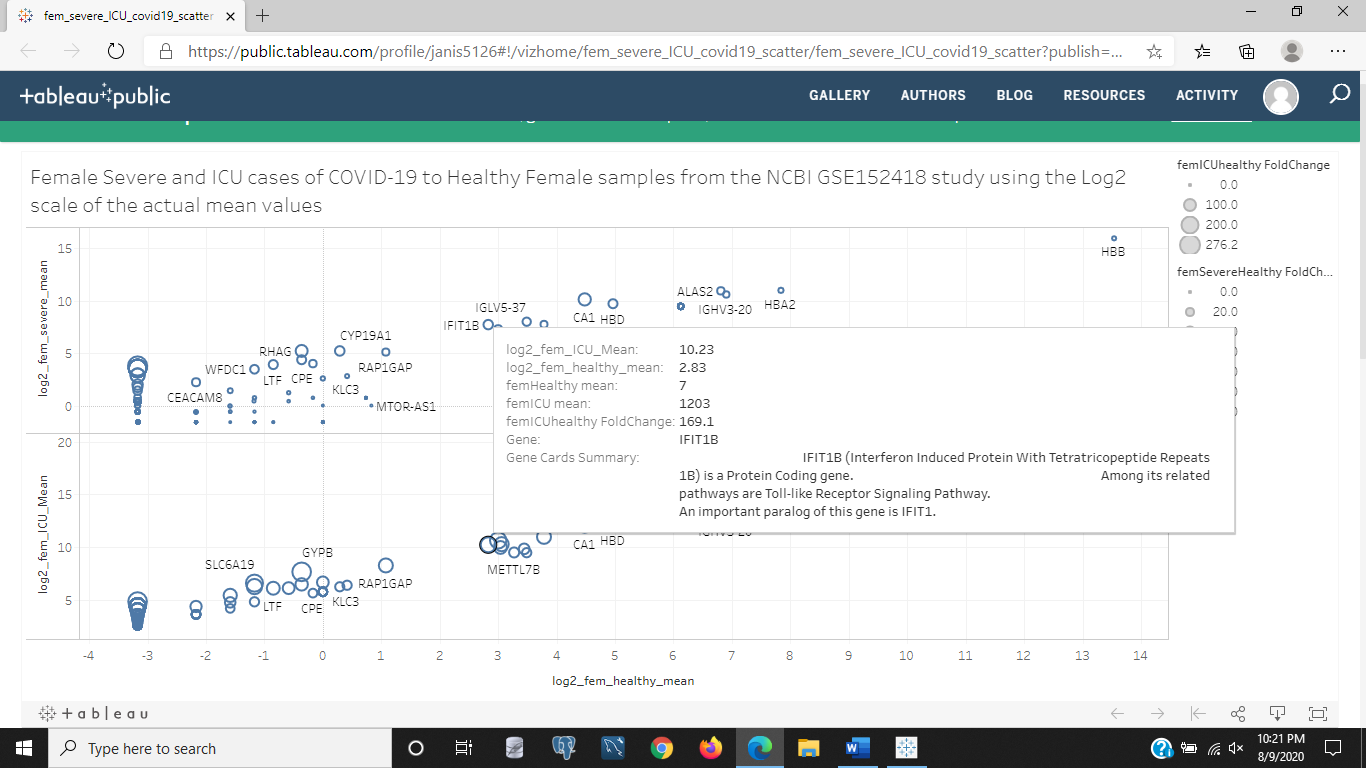


Figure 14b: The above image is the same scatter chart comparing female log2 scaled gene expression means of ICU and severe to healthy cases with size as an added dimension for actual mean value fold changes across samples. The highlighted gene is IFIT1B in the lower segment representing the ICU/healthy female gene expression values. This gene is one of the larger sized shapes so it has more expression in the disease state than the healthy state. This gene is an interferon protein encoding gene involved in the immune system. It is also involved in immune response signaling. This gene is more expressed in ICU COVID-19 cases possibly because it needs to signal to other genes to fight the antigens of COVID-19 and alert the host to danger and possibly incite symptoms that derail the host and keep them from doing any activities. Just an assumption, but interferons are usually the cells of the immune system like cytokines that are painful or uncomfortable to alert the host to be cautious, not do much, or keep them from doing much by making the host feel a lot of uncomfortable and possibly painful symptoms. They are analagous to you realizing you don't have your cell phone then being overtaken by anxiety and other emotions that make it an absolute priority to find it because of the valuable information on it that wasn't downloaded, backed up, or saved, like the photos, logged in accounts and social media hosts, phone numbers of clients and freinds you don't call frequently enough to remember, etc. They immobilize you like your foot getting hit by a hammer or stubbed in the door to protect you and keep you from doing any more damage.

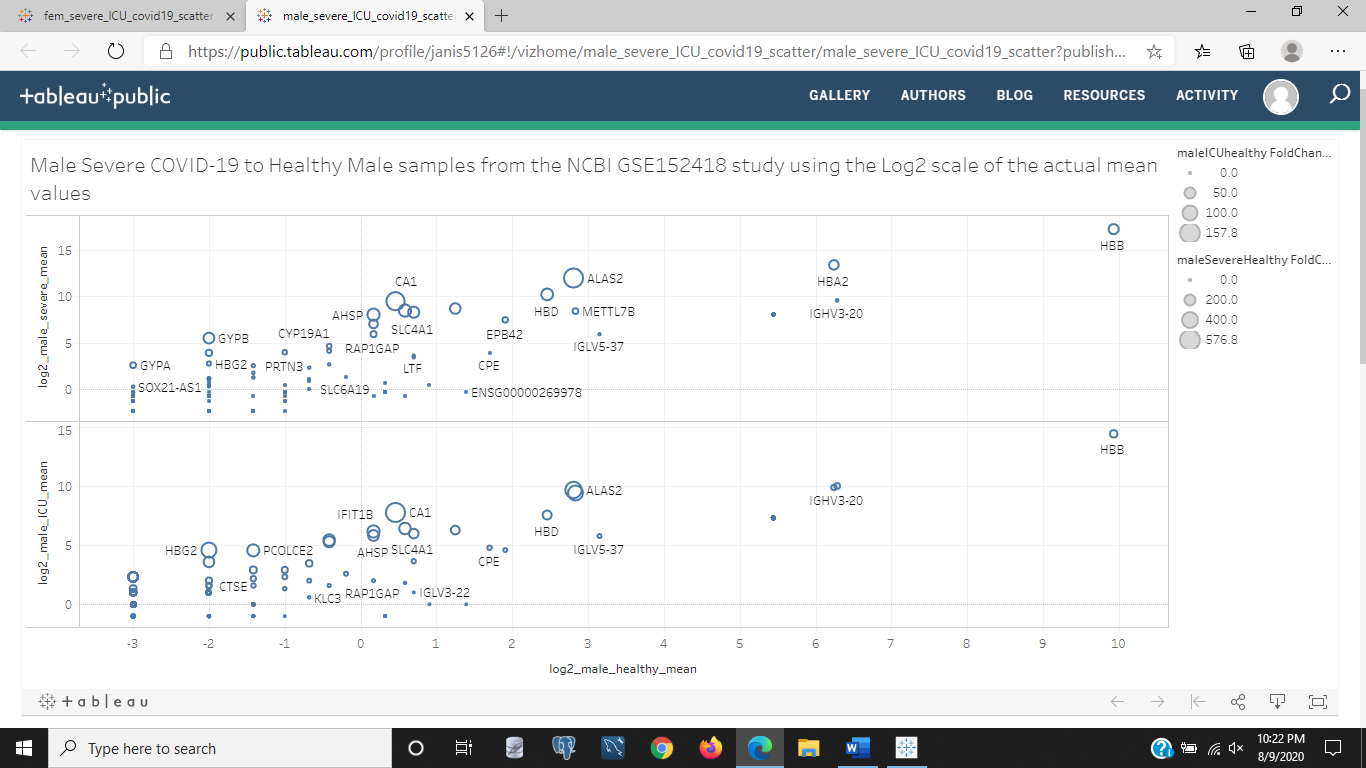
<https://public.tableau.com/profile/janis5126#!/vizhome/male_severe_ICU_covid19_scatter/male_severe_ICU_covid19_scatter?publish=yes>

Figure 15a: The above image is of the males in the study that compares the log2 scaled gene expression means of the ICU and severe COVID-19 to healthy male gene expression means with a size dimension added to indicate those genes having a higher fold change in the disease to healthy ratio. None of the scatter points are hovered on for details. But looking at the chart above, what differences are noticeable to you? We see the far left genes are much bigger in fold change as they are the same spot in healthy males along the x-axis, but increase vertically on the y-axis for the two disease states of ICU or severe COVID-19 cases. Many of the genes have increased as they are closer together in the disease states. ICU COVID-19 seems to be more severe of a class than the severe COVID-19 class. There aren't a whole bunch of changes though, and they seem to be very close in gene expression between the two cases that are both more severe than the moderate case. As a note, this data didn't have any males in the moderate group, they were all females. This was mentioned earlier, but added as a reminder as this is the 15th chart displayed and analyzed in this blog.

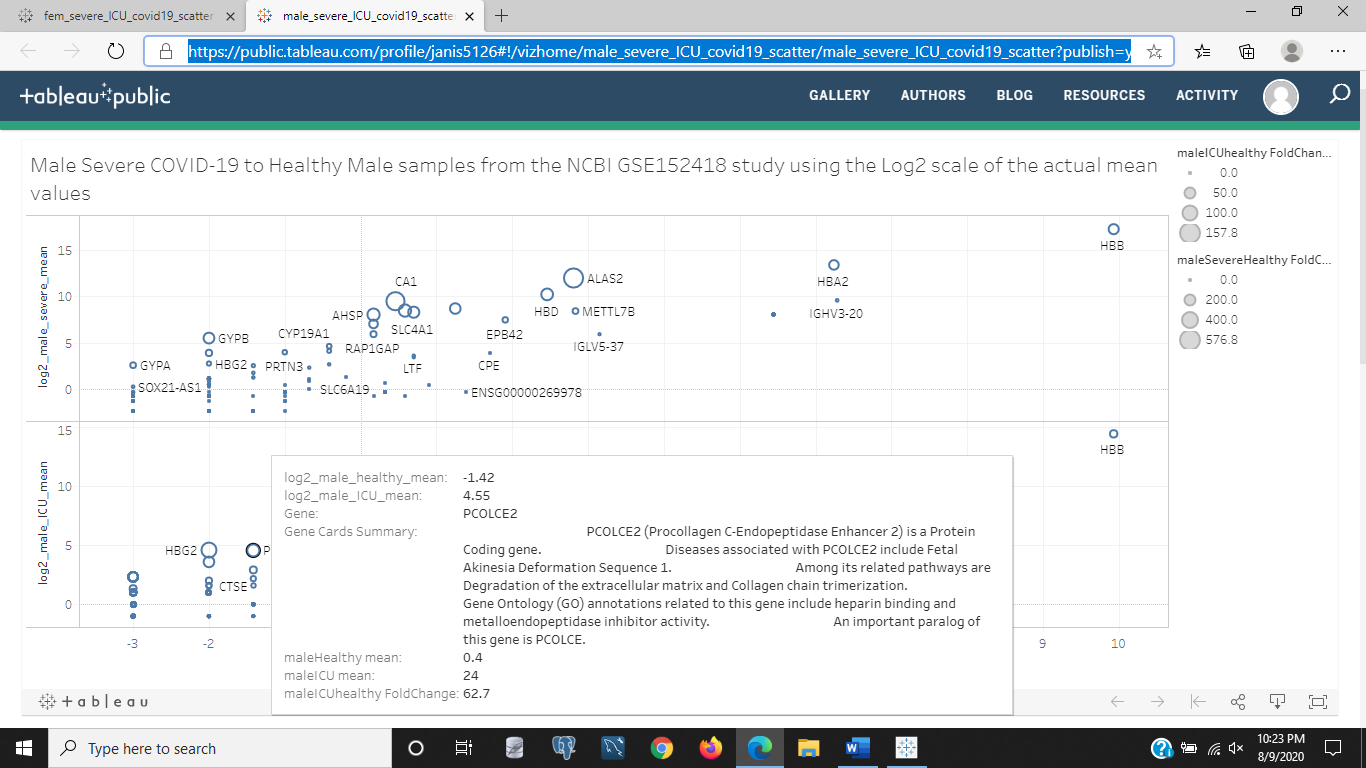


Figure 15b: The above image is the same chart as was in Figure 15a, of the male group comparison of log2 scaled gene expression means of the ICU and severe COVID-19 cases compared to the healthy cases in our study. The highlighted and hovered over scatter point is the gene, PCOLCE2, a gene associated with cellular matrix degeneration, collagen , heparin binding (hepatic is for liver), and metalloendopeptidase inhibitor activity (meta-endo-pep; beyond-inside-digest) possibly slowing down or stopping internal digestion. This gene affects the cellular matrix, and what ever its goal or intentions are in breaking down the cellular matrix could be because it received a signal that an invader has infiltrated the network and needs to be destroyed within the matrix or slow down the chemical breakdown of some protein. It could also mean that the virus has invaded the host cells and is now destroying the cell and making room to take over and replicate.

<https://public.tableau.com/profile/janis5126#!/vizhome/male_severe_ICU_covid19_barchart/gender_COVID19_barchart_FCs?publish=yes>

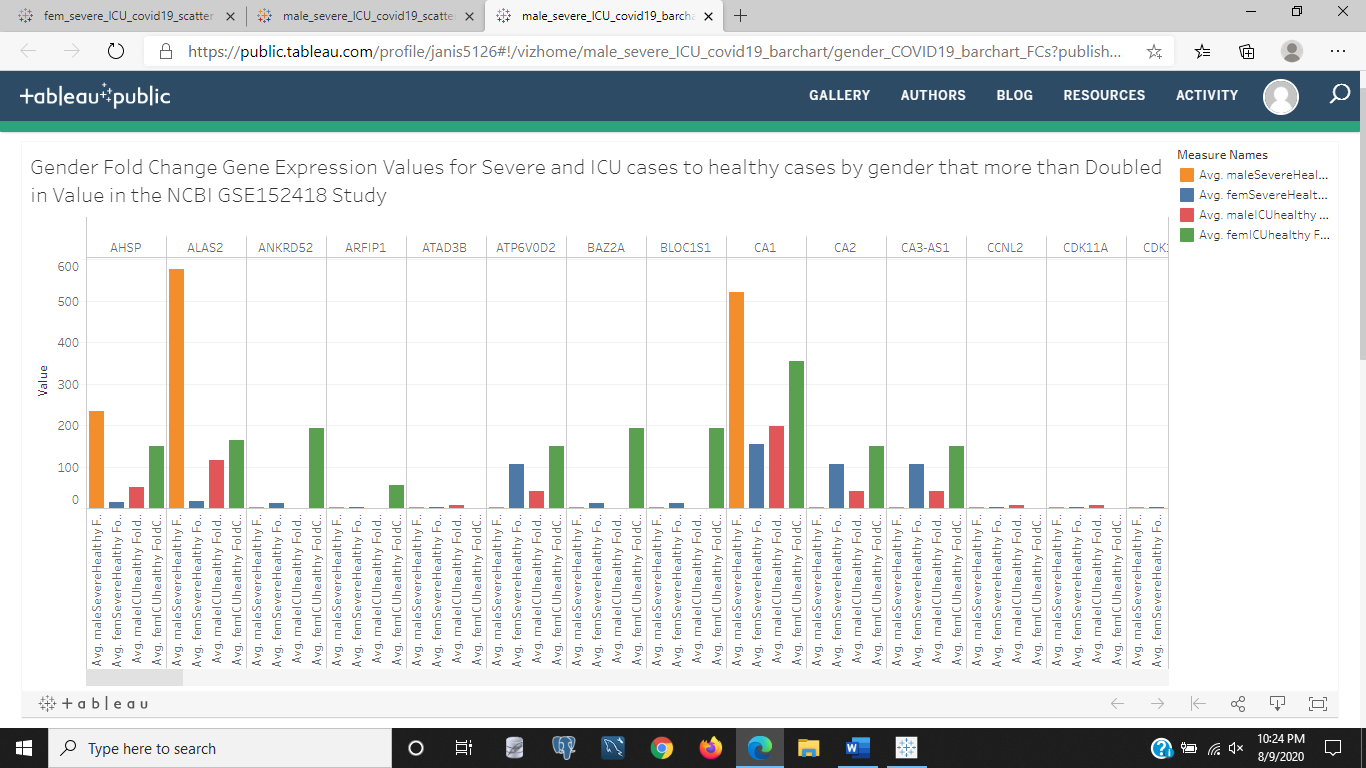


Figure 16a: In the above image we see a bar chart of the gene expression fold changes as an average, but the average of one value is that one value. In Tableau to set up this visualization the class of the variable has to be a measure and not a dimension by Tableau standards, and so the average instead of the sum, count, etc. was taken. This way we see a side by side comparison within each gene instead of a side by side comparison of each variable for the gene. Each of the categories are of the genders of male or female fold changes of severe or ICU COVID-19 gene expression means within each gender group compared to the healthy gender means of each group. We can visually see that Male Severe cases dominate in AHSP, ALAS2, and CA1 and that the female ICU cases dominate in most all the displayed genes as the runner up to male severe cases of AHSP, ALAS2, and CA1, or as the sole dominating group for genes ANKRD52, ARFIP1, ATP6V0D2, BAZ2A, BLOC151, CA2, and CA3-AS1.

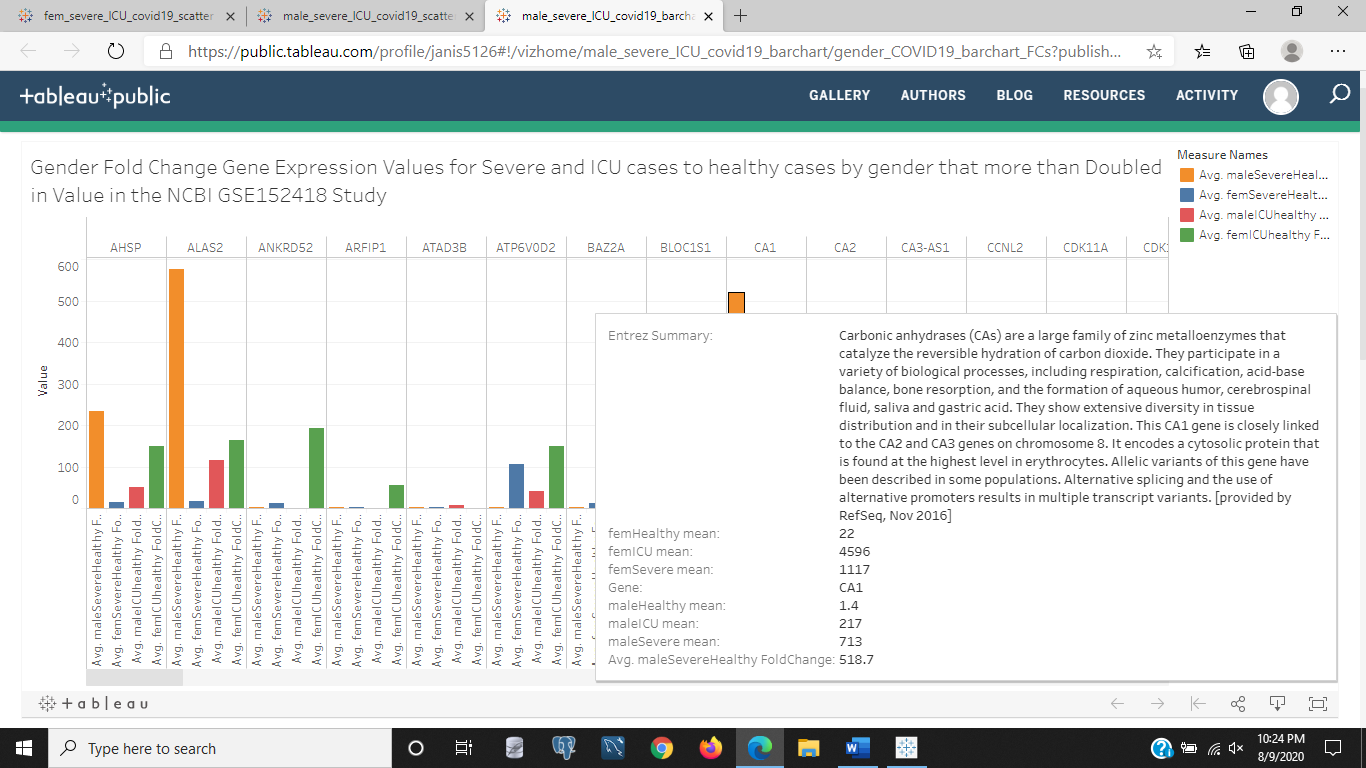


Figure 16b: The above image is the same chart as in Figure 16a, except that one gene was hovered on for details. The gene CA1 that we saw in Figure 16a had a high fold change for male severe and female ICU COVID-19 cases compared to healthy cases. This is a carbonic anhydrase that catalyze the reversible hydration of carbon dioxide, which to me and others not familiar with this terminology gives the idea that a biproduct of oxygen synthesis in the body to carbon dioxide adds a carbon atom, but to catalyze like starting or igniting a chain of events, the process of hydration of CO2 is reversible . Honestly, is heavily verbose. Fortunately, the other information in the gene summary given by the Entrez gene summary is more informative. It is involved in respiration or breathing, calcification or building bone, acid base balance or balancing the pH in our bodies, and balancing bone resorption or I guess recycling calcium for preventing osteoporosis, and balances cerebrospinal fluid, saliva, and gastric acid used in peristalsis or digestion in the parasympathetic nervous system. Other noticeable information about this gene given by the details and not having to translate the gene summary say that this gene, CA1, is related to the other CA2 and CA3 genes, and is heavily up regulated in females and males for the ICU and severe cases of COVID-19. But more up regulated in female ICU than female severe cases and also more up regulated or expressed in male severe than ICU cases. Somewhat of a flip in increased gene expression. Assuming the ICU COVID-19 cases are more severe than the severe COVID-19 cases. This wasn't a detail in the NCBI GSE152418 detail.

<https://public.tableau.com/profile/janis5126#!/vizhome/male_severe_ICU_covid19_barchart_side/gender_COVID19_barchart_FCs?publish=yes>

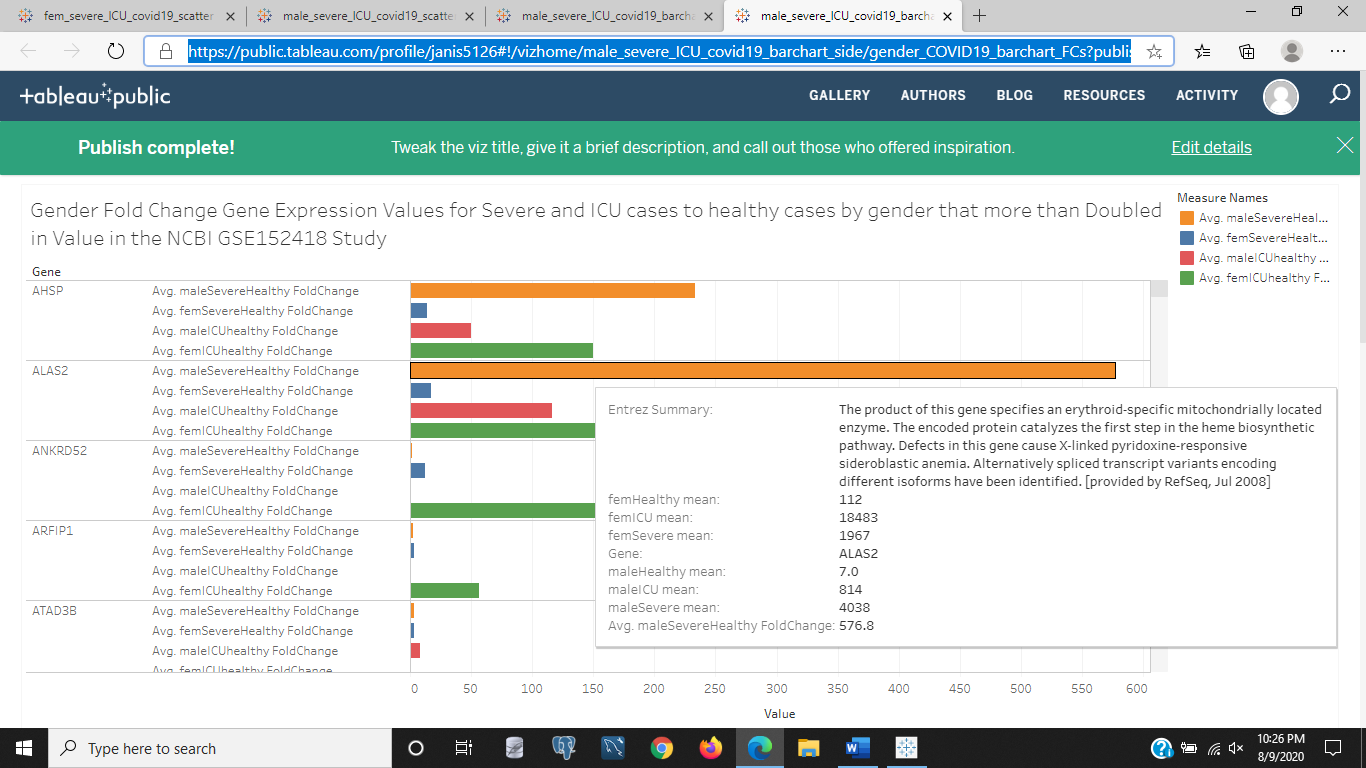


Figure 17: The above image is a bar chart of the fold change of gene expression means across samples by gender and severity of COVID-19 for ICU or severe cases having more than doubled in fold change of disease to healthy state. The ICU is assumed to be more severe than the 'severe' COVID-19 cases. These genes shown are those selected to have at least a fold change of 2 or having doubled in the disease to healthy ratio of fold change across samples by gender and severity within each gender. The details displayed are on the hovered bar representing the ALAS2 gene for the male severe to healthy fold change values. This gene is extremely over expressed compared to the other classes of COVID-19 and gender group. This gene deals with a mitochondrial located red blood cell enzyme. The mitochondria is where proteins and lipids are processed in the cell. Defects in this gene are related to a type of fever induced anemia (pyridoxine-responsive, pyr-'heat or fire or fever') and 'X-linked' has to deal with the X-chromosome or the female chromosome. One defect of the details on this gene for the sideroblastic (blast-'build and repair' like osteoblasts) could mean it responds to anemia and builds and repairs defective red blood cells. The male severe cases of COVID-19 are much higher than the more severe ICU COVID-19 cases and could mean more of this gene is needed to help fight the COVID-19 virus in males and prevent it from destroying any red blood cells The gene isn't known to be defective in these samples as motif and allele or copy number variants weren't available to differentiate the number of variations of this gene, but it is highly expressed in the earlier stage of COVID-19 than the later 'ICU' stage of COVID-19. I just want to point out, since this is a blog, or informal discourse on a subject, I have a minor in my Masters of Science in data science and bioinformatics, but I want to make it very clear that I was not given details on these meanings, my instructor always recited the principles than had us do research and 'synthesize' the information in our words. So, it wasn't a door unlocked to unmassed secret information, we had to use our own deductive reasoning and knowledge of what we found to explain and interpret and discover meaning from our research. You will see this same method is the reason higher ups in education and business and even parenting always fight, because there is no one way to do something and like any person will replicate identically the same exact way of discovering or searching for reasoning or logic behind something. So, feel free to disagree with me on interpretation or my own deductive reasoning and explain your points in the comments if you want to add something or point out a different interpretation based on other evidence or facts from other sources. Its a free country, just avoid harassment, because that is a punishable civil crime against another human being.

<https://public.tableau.com/profile/janis5126#!/vizhome/male_severe_ICU_covid19_barchart_side10X/gender_COVID19_barchart_FCs?publish=yes>

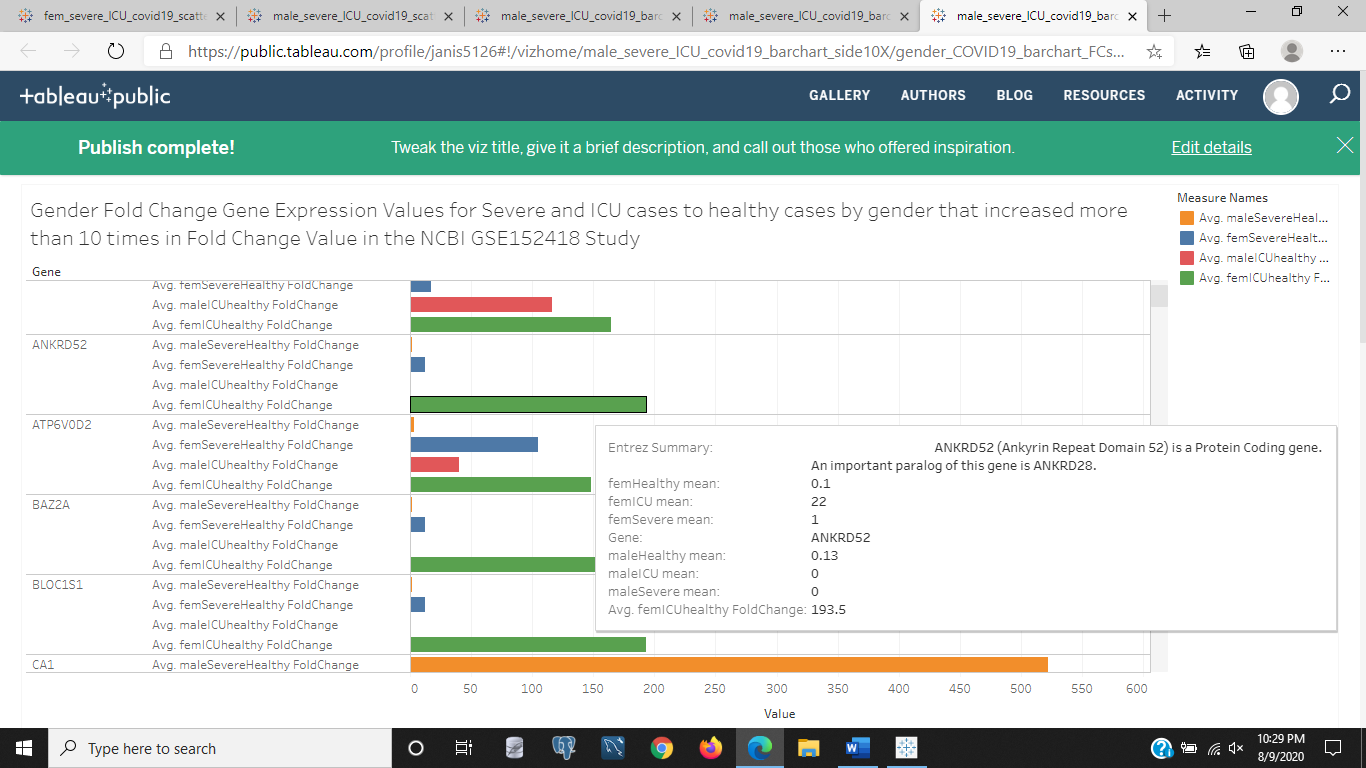


Figure 18: The above image is the bar chart of the gender comparisons of severe and ICU COVID-19 cases having fold change values greater than 10 for disease to healthy case ratios. We can see in the image above that There are quite a few genes that more than increased 10 times that of the healthy mean values in each separated group by gender. The highlighted gene displayed above is the ANKRD52 gene that is a protein coding gene according to its Entrez gene summary, so much more isn't known about this gene. However, it is expressed a great deal more in female ICU COVID-19 cases than female severe COVID-19 cases and also the male cases of COVID-19 don't seem to have much of an increase in this gene other than the severe male case of COVID-19 by a very miniscule amount visually. The healthy female mean was 0.1 and the ICU female mean was 22. A great increase in the female groups of the most severe case of COVID-19. This gene could be responsible for the deadly symptoms or aiding the virus as a protein coding gene to build itself and compete with other cells in the hosts body. Perhaps, as a suggestion, this gene could be knocked down or removed and see if it has an effect on the females with ICU grade COVID-19 to see if their symptoms and healty improve.

<https://public.tableau.com/profile/janis5126#!/vizhome/genderChartCovid19/genderChartCovid19?publish=yes>

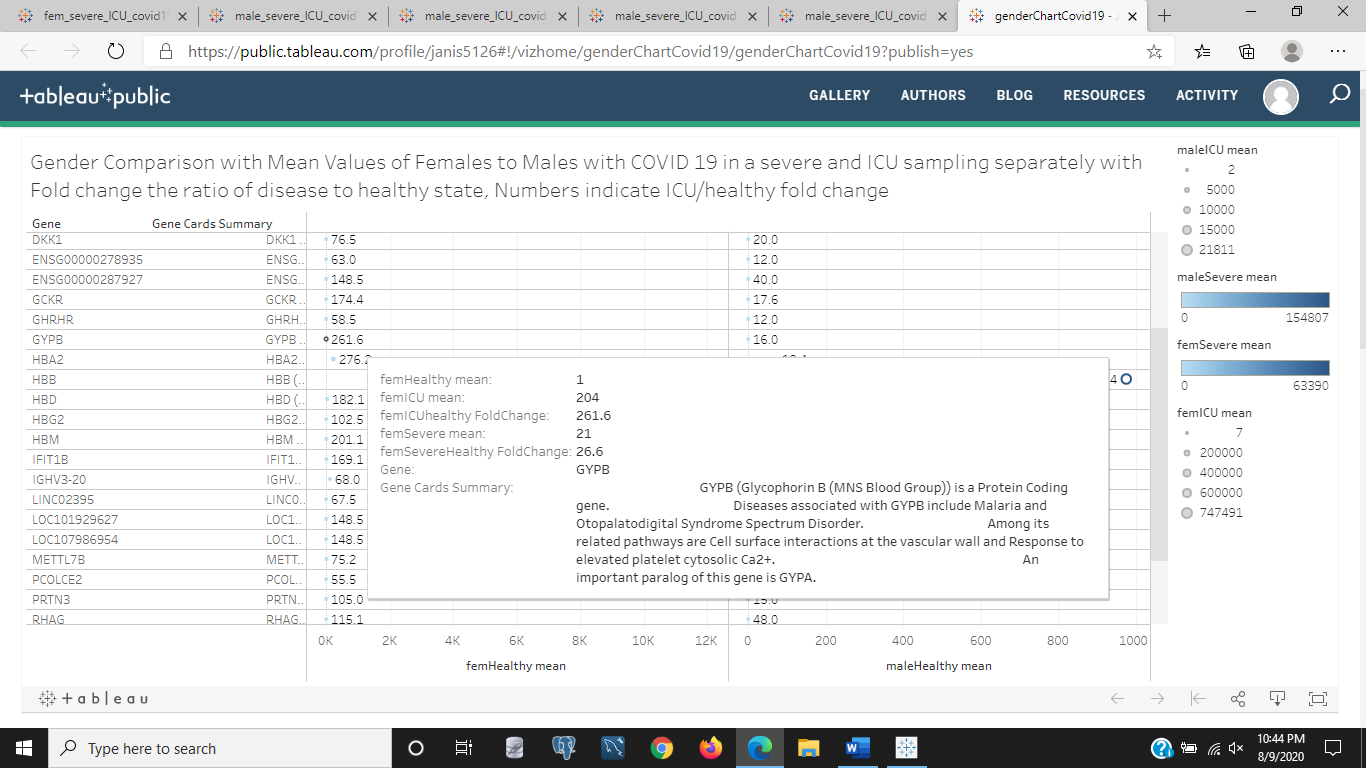


Figure 19: The above image is a table chart of the gene expression mean values of females to males with COVID-19 in the severe and ICU cases of COVID-19. The chart displays the numeric mean of those genes for the healthy females and males as columns and genes as rows. But the color bar indicates which of those genes have a higher severe COVID-19 gene expression value if it is darker blue than the others, and size of the scatter point indicates the ICU gene expression mean with larger points indicating higher ICU gene expression means. The hovered gene details is for GYPB, we saw earlier, a gene associated with Malaria and an Otopalatodigital Syndrome Spectrum Disorder. It is involved in cell surface interactions at the vascular wall and responds to elevated platelet (thrombocytes or red blood cell clotting components to heal wounds) cytosolic (inside the cytoplasm of a cell) calcium ion levels.

<https://public.tableau.com/profile/janis5126#!/vizhome/genderChartCovid19_100XfemICU_FCs/genderChartCovid19?publish=yes>

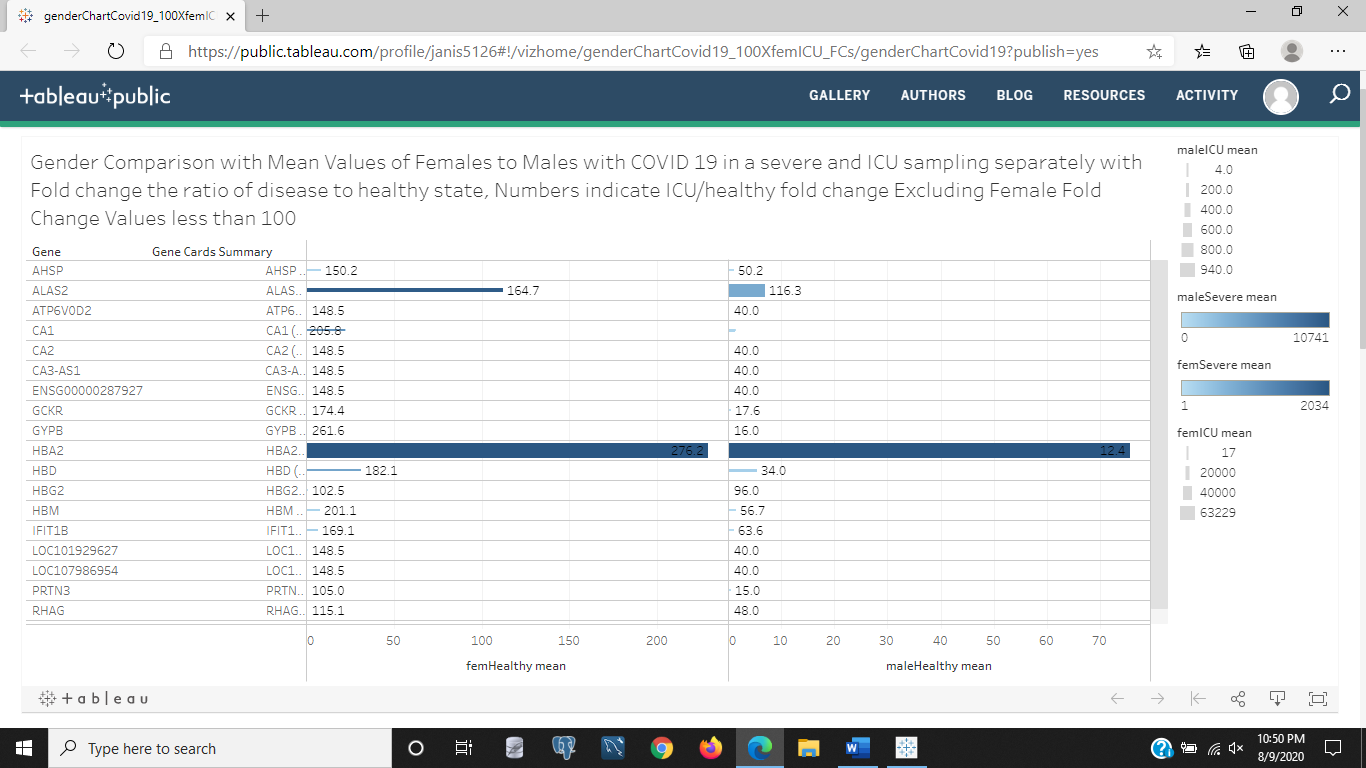


Figure 20: The above image is a variation of the chart in Figure 19, except instead of scatter points it uses bars to represent the size of the ICU gene expression mean values in each gender and the same color bar for severe COVID-19 cases with darker blue representing higher gene expression in the severe gene expression means for that gene. This chart excludes those genes that females had ICU/healthy fold change values less than 100. We can see the gender to gender differences in gene expression in the above chart. Look at GCKR, and see that the healthy female gene expression is 174.4, but the male healthy gene expression levels are 17.6, almost 10 times lower than the female levels as a comparison of the healthy groups of females to males. But also, due to our added dimensions for size and color indicating ICU and sever cases respectively, that the darker bar is HBA2, an immune gene and ALAS2. Also, look at CA1, in Healthy females it is 205 and negligible or zero for males, but the size indicates for ICU samples in females the change was significant, while not so much with the males. Click the image to explore other genes and the hovering details in the Tableau chart this image is a screen shot of.

Thank you for reading through this analysis of genes affecting COVID-19 patients and exploring the genes that could be possible gene targets of COVID-19 pathogenesis. Feel free to add in your own interpretations in the comments or explore the data behind the interactive notebooks by clicking each image that links to its relative chart.

Most genes displayed were immune genes or protein coding genes, we did see some differences by age and also by gender in certain stages of COVID-19, assuming these three classes of COVID-19 are stages and not different strains of COVID-19 as 'moderate','severe', and 'ICU'. Russia has recently released its own vaccine, but it has yet to be determined if it works. I discovered this while on my way back from my MBLEx exam at around 9:40 am this morning while listening to 103.5 radio and hearing the dj Ellen say that Putin already gave the vaccine to his daughters. Hopefully, it works.