

**Unveiling Shared Molecular Pathways Between COVID-19 and Multiple Sclerosis**

**Final Report**

|  |  |
| --- | --- |
| **Names** | **ID** |
| Ashraqat Abdelhamid | 221000836 |
| Farid Ghattas | 221000545 |
| Raghad Abdelkader | 221001278 |
| Malak Haitham Mohamed | 221001396 |

**Advanced programming/Data analysis**

**Supervised by:**

**Dr. Mohamed Elsayeh**

**Table of content**

*Abstract.............................................................................................................3*

*Introduction........................................................................................................4*

*Methodology ......................................................................................................7*

*Results.................................................................................................................8*

*Discussion...........................................................................................................18*

*Conclusion..........................................................................................................22*

*References...........................................................................................................23*

**Abstract**

COVID-19 and multiple sclerosis (MS) exhibit overlapping molecular mechanisms that may contribute to their disease progression. This study utilized gene expression datasets to identify differentially expressed genes (DEGs) and performed Gene Set Enrichment Analysis (GSEA) to uncover shared pathways. The analysis highlighted ferroptosis, starch and sucrose metabolism, cytokine-cytokine receptor interactions, and broader metabolic pathways as key overlaps. These pathways suggest common processes such as oxidative stress, immune dysregulation, and metabolic reprogramming, linking viral infections with neuroimmune conditions. These findings provide a molecular basis for understanding the shared biology between COVID-19 and MS, offering potential avenues for targeted therapeutic interventions.

**1.Introduction**

***1.1 COVID & Multiple sclerosis***

COVID-19 is a global pandemic with wide-ranging effects, from respiratory illness to immune and neurological complications. Key molecular mechanisms include oxidative stress, cytokine storms, and metabolic reprogramming, linking it to other diseases like autoimmune and neuroimmune disorders.

Multiple sclerosis (MS) is a chronic autoimmune disease that affects the central nervous system, leading to inflammation, demyelination, and neurodegeneration. It is characterized by symptoms such as fatigue, motor weakness, and cognitive impairment, driven by immune dysregulation and neuroinflammation.

***1.2 MS & COVID Relationship:***

The interrelation between **Multiple Sclerosis (MS)** and **COVID-19** is primarily rooted in the effects of COVID-19 on the immune system, which is already compromised in MS patients. The following points highlight the key interrelations based on recent research:

***1.2.1 Immune System Dysregulation***

MS is an autoimmune disease where the immune system mistakenly attacks the central nervous system, causing inflammation and damage to the myelin sheath. COVID-19 also triggers immune dysregulation, notably through cytokine storms, which may exacerbate immune system problems in MS patients. The inflammation caused by both diseases can worsen the symptoms of MS and potentially lead to relapses or flare-ups.

Both conditions involve autoimmune responses, where overactive immune responses can lead to increased neuronal damage.

***1.2.2 COVID-19 as a Trigger for MS***

Several studies suggest that COVID-19 infection may trigger new-onset MS or a relapse in individuals with pre-existing MS, particularly in those with severe COVID-19. The viral infection may act as a trigger for demyelination or worsen existing neurological damage in MS patients due to the immune system's heightened response to the infection.

***1.2.3 Increased Risk for MS Patients***

MS patients who are on immunomodulatory treatments (e.g., disease-modifying therapies) may be at increased risk for severe outcomes from COVID-19. These treatments can suppress the immune system, making it harder for the body to fight off infections like SARS-CoV-2. COVID-19 exacerbates the vulnerability of MS patients, especially those on immunosuppressive therapies, who may experience complications such as respiratory failure or higher rates of infection.

**1.2.4 Neurological Effects**

Both COVID-19 and MS are associated with neurological symptoms, with COVID-19 causing various complications such as cognitive dysfunction, loss of smell and taste, and headaches. In individuals with MS, COVID-19 may worsen cognitive decline and other neurological problems due to common mechanisms like neuroinflammation and oxidative stress. Inflammation in the brain caused by COVID-19 can resemble or even trigger symptoms like MS, complicating both diagnosis and treatment.

**1.2.5 Inflammatory Pathways**

Both diseases involve inflammatory pathways, such as cytokine release (e.g., IL-6, TNF-α). The inflammatory environment in MS is further aggravated by COVID-19, particularly in severe cases, leading to an increased risk of relapses or progression in MS patients. Additionally, both diseases involve oxidative stress and mitochondrial dysfunction, which play critical roles in neurodegeneration and disease progression.

**1.3 Treatment Challenges and long-term effects**

Multiple sclerosis (MS) treatments, particularly immunosuppressive drugs, present challenges for managing COVID-19, as some therapies may increase susceptibility to infections, making it harder for the body to respond to SARS-CoV-2. Conversely, MS patients not on immunosuppressants might experience milder forms of COVID-19 due to a less suppressed immune system. Additionally, growing evidence indicates that COVID-19 survivors, especially those with severe cases, may experience long-term neurological effects, such as fatigue, cognitive issues, and muscle weakness, which can overlap with MS symptoms. These neurological symptoms could mimic the relapsing-remitting nature of MS, complicating the understanding and management of long-COVID syndrome in MS patients.

**2.Methodology**

***2.1 Data Collection and Preprocessing***

We used Gene Expression Omnibus (GEO) to retrieve the datasets for SARS-COV-2 (COVID-19) and Multiple Sclerosis (MS). The accession numbers of the datasets were GSE172114 and GSE137143 respectively. Those datasets included RNA-seq data, and while multiple datasets were present regarding these two diseases, yet we specifically chose those due to the availability of RNA-seq counts and GEO2R option, as well as the large number of samples that enables us to derive accurate results. 60 samples were chosen from each dataset containing equal numbers of control and diseased samples. For both diseases, differential gene expression analysis (DEGs) was carried out after defining the groups in the sample using GEO2R software.

***2.2 Filtration of genes***

In the GEO2R step we obtained two lists of DEGs, one for each disease. However, to carry out the aim of this paper and integrate the genes, we used R which is a language and environment for statistical computing and graphics. The R script aimed to combine and filter the genes produced from GEO2R according to the p-value for significance, and the log2 fold change (log2FC) to differentiate between upregulated and downregulated gene expressions.

***2.3 ShinyGO analysis***

Moving forward, now that we obtained the csv files with the common genes, We used ShinyGO, a bioinformatics tool, to understand the functions of the shared genes. ShinyGO helped with enrichment analysis, building phylogenetic trees, studying protein-protein interactions and exploring gene functions. The phylogenetic analysis looked at evolutionary relationships, while the functional annotation identified the biological roles and processes these genes are involved in.

***2.4 Pathway Analysis***

For the pathway analysis we used 2 pathway databases: Reactome database and KEGG database. Starting with the reactome database, the genes’ symbols were copied to g profiler and the genes marked in the reactome segment were identified and searched by ID in the reactome database. Moreover, while doing the shinyGO analysis, there was a section named “KEGG”, this section provided additional pathways for us to explore.

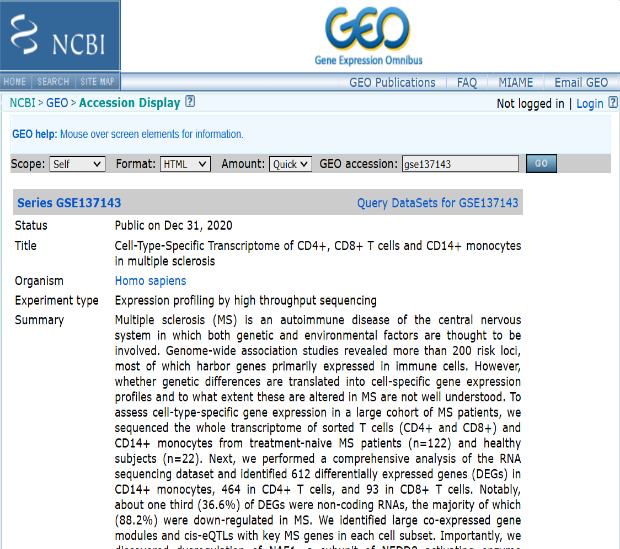
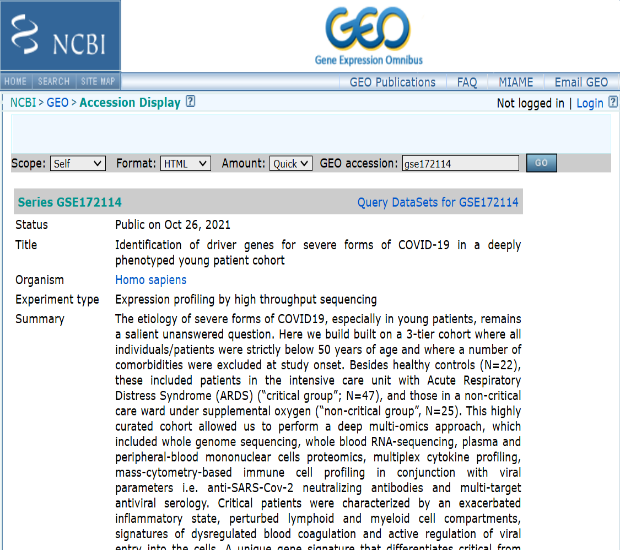
***2.5 Statistical Validation***

Finally, multiple plots were created using R, to validate the significance of our results, some of the plots targeted the p-value while others targeted the false discovery rate (FDR), to ensure the reliability and accuracy of the data and results.

**3. Results**

**3.1 Data Collection and Preprocessing**

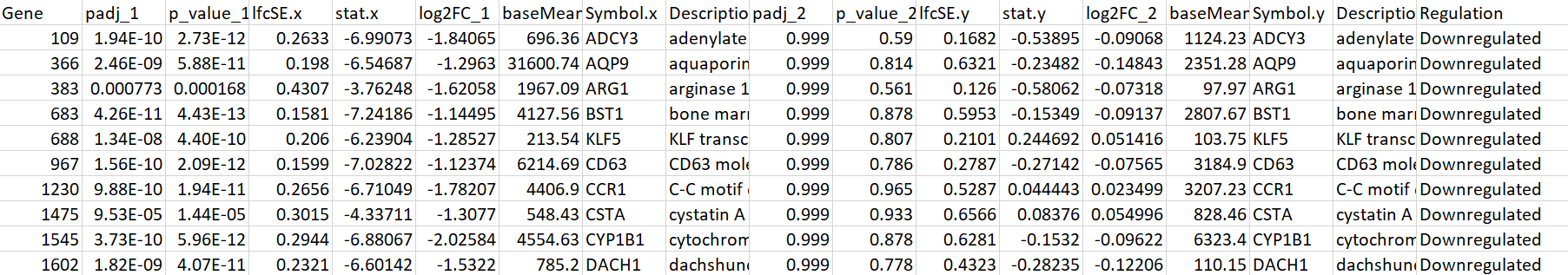
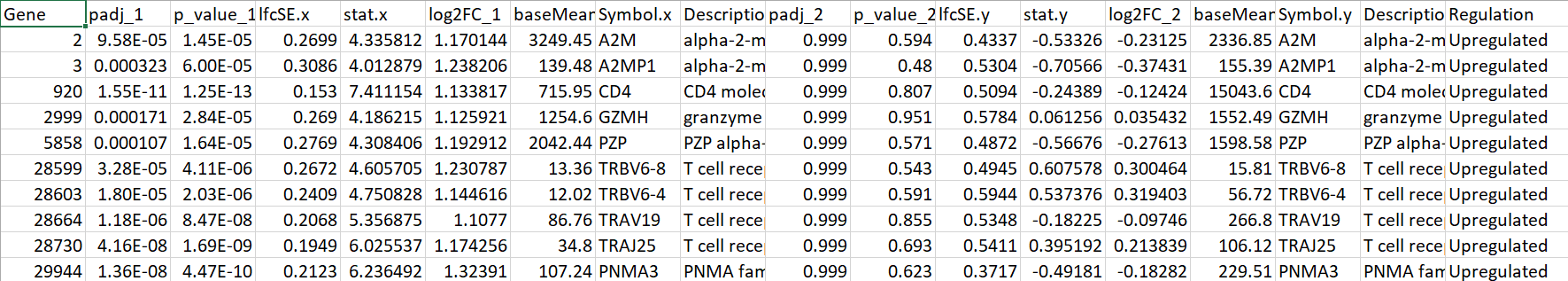
*GEO dataset*



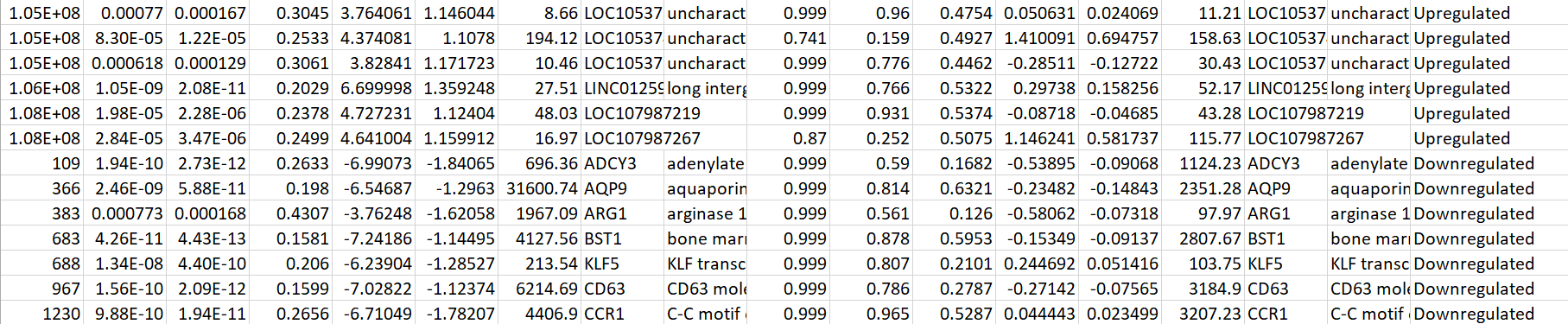
***Figure 1.*** *The data was collected from NCBI GEO to get the dataset of both diseases COVID and MS with their codes respectively gse172114 and gse137143.*

***3.2* Filtrationof genes**

*Upregulated and downregulated genes*

***Figure 2.*** *Upregulated genes may produce proteins involved in inflammation or immune responses while downregulated genes occur in situations where the body needs to reduce certain functions or pathways, like when trying to limit excessive inflammation.*

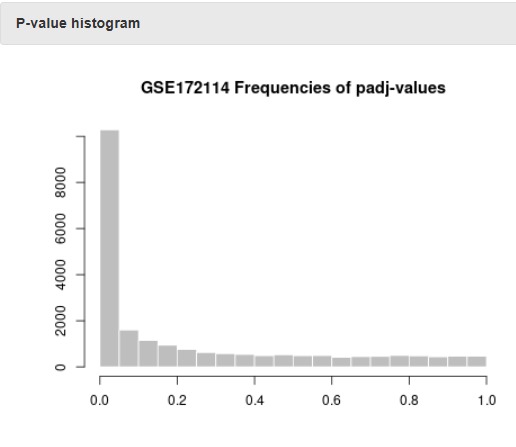
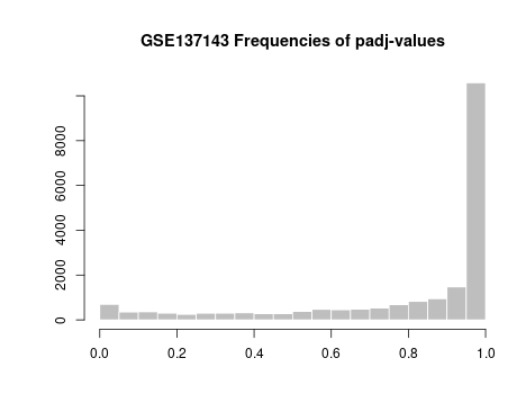
*Filtered genes*

***Figure 3.*** *The genes were filtered using R script and it arranged the statistical analysis in an excel sheet that express statistical significance. that helps to understand the mechanisms of diseases like COVID-19 and MS*.

## **3.3 Insights Plots for COVID-19 and Multiple Sclerosis Disease**

*Histogram comparison of p-value distributions between two diseases*

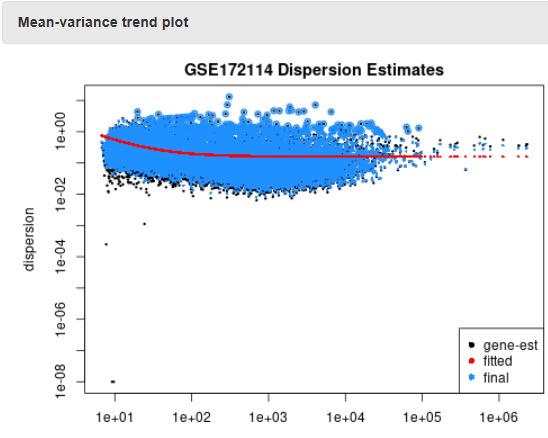
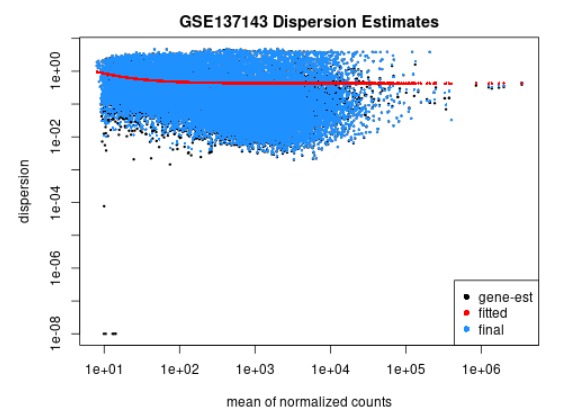
The histogram plot compares the distribution of adjusted p-values between the two diseases, highlighting differences in the number of significantly differentially expressed genes.

***Figure 4.*** *These histogram figures show the range of p value in x axis. In GSE172114 the high bar near 0 indicates a very low p values that suggest they are significantly differentially expressed. There are a considerable proportion of low p-values that are often expected in datasets with many differentially expressed genes. In GSE137143 the high bar shown in the figure have a high p value that is approximately 1.0*

*Comparison of mean-variance trends between two diseases*

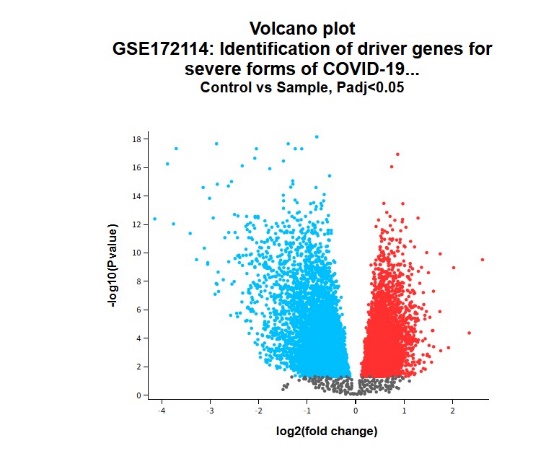
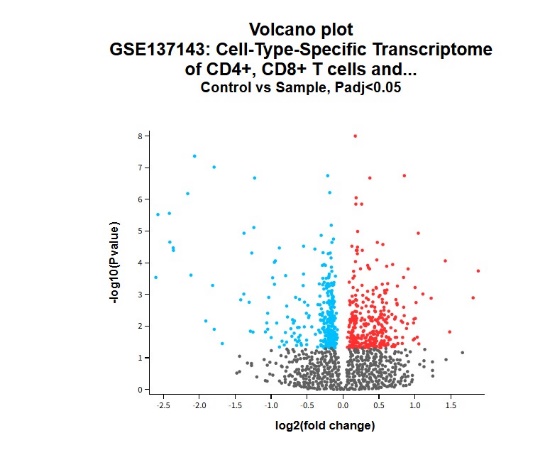
The mean-variance trend plot compares the dispersion estimates as a function of mean expression levels for the two diseases, highlighting differences in gene expression variability.

***Figure 5.*** *The mean variance plots show the dispersion level as a function of the mean expression values for each gene. The red line represents the final dispersion, while the blue dots show fitted values, and the black dots represent the gene-wise dispersion estimates. In both diseases it is shown that the dispersion estimation model fits the data well, with a typical decreasing trend of dispersion as mean expression increases.*

**Comparative analysis of two volcano plots**

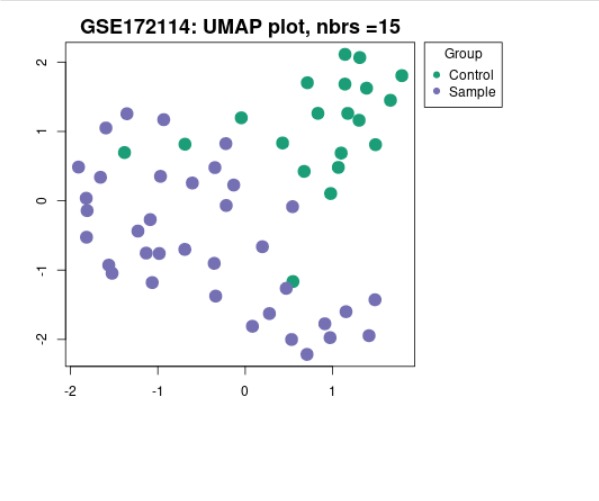
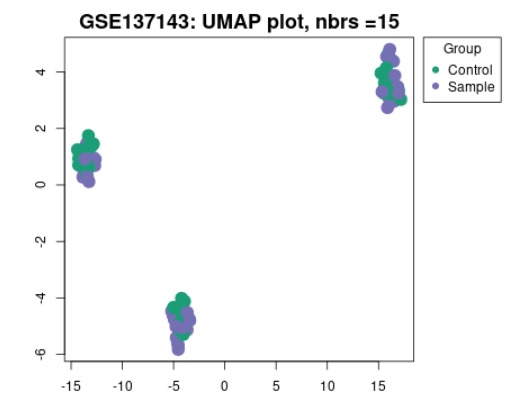
This comparison of two volcano plots provides a side-by-side visualization of the two diseases of gene expressions, highlighting the contrast in statistical significance and changes across two distinct datasets.

***Figure 6.*** *The x axis shows positive for up regulation, negative for down regulation, while the y axis shows the statistical significance. Each point represents a gene with its corresponding fold change and p-value. The red color represents the up regulated gene, blue represents the down regulated genes, and the grey color represents the non-significant genes, based on fold change and p-value. In GSE172114 it shows red and blue color with little amount of grey color that represent many genes showing significant changes. While GSE137143 shows fewer red and b*l*ue points, meaning fewer genes are showing substantial up regulation or down regulation.*

*UMAP visualization of gene expression across two diseases*

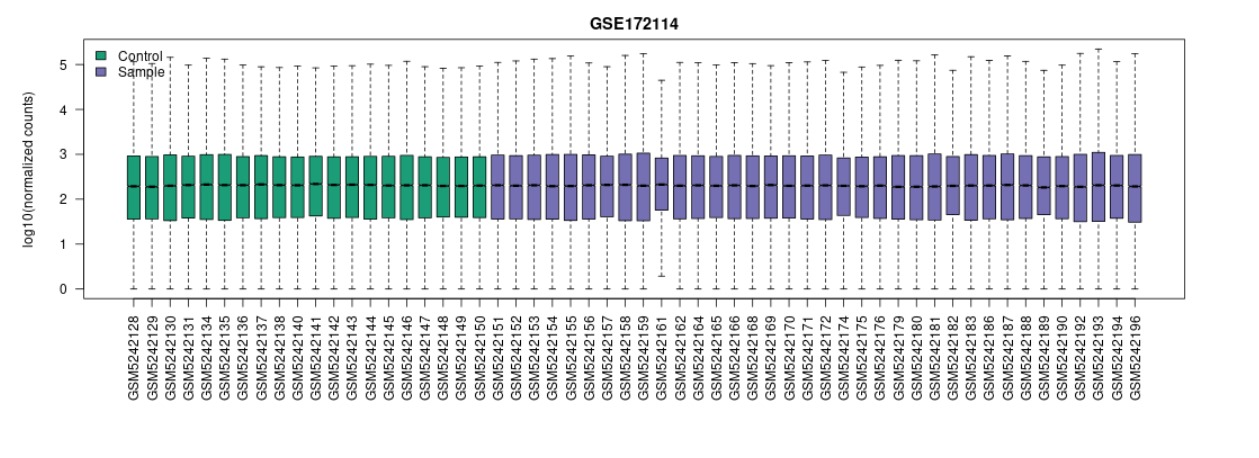
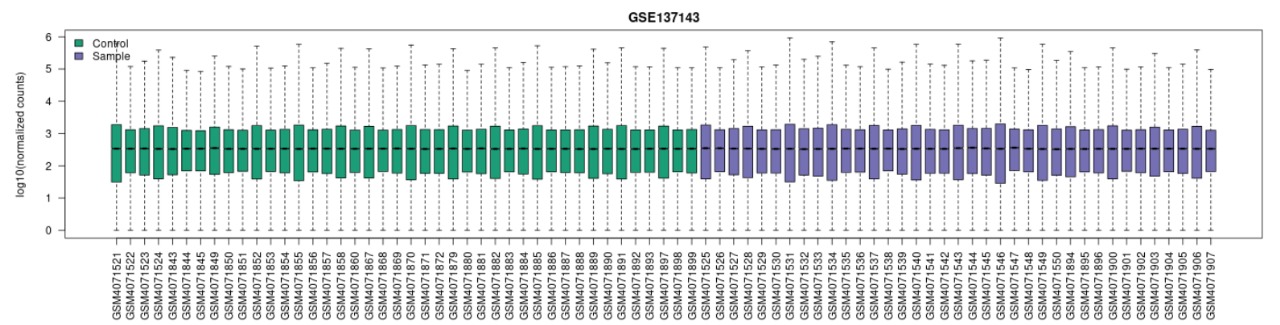
UMAP is a plot used to reduce complex data into a simpler format, making it easier to visualize patterns and groupings in the data of the two diseases (Covid, MS).

***Figure 7.*** *There are two groups of genes the green color represents as control, and the purple represents as sample which appear intermixed, suggesting some overlap between the groups in GSE172114. In GSE137143, the points are more tightly clustered into distinct groups, indicating less variability within the sample and it shows more defined cluster.*

**Analysis of boxplots**

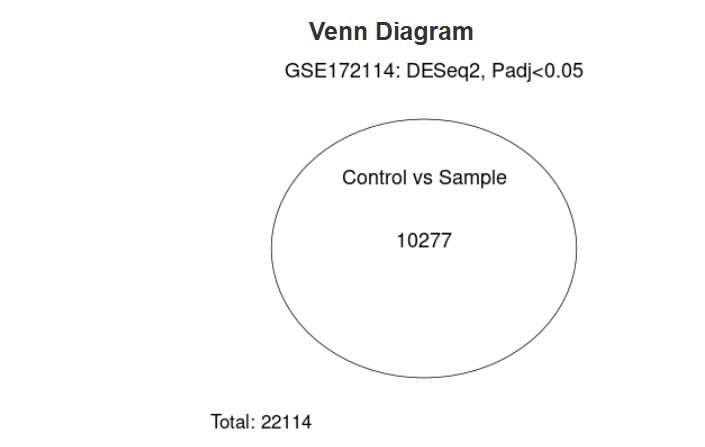
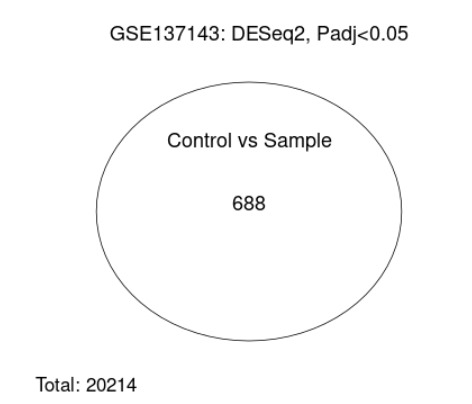
The boxplots illustrate the distribution and variability of "Control" and "Sample" groups in two different datasets, highlighting key differences in their medians and overlap.

***Figure 8.*** *The boxplot shows two groups: Control (green) and Sample (purple). Each dataset offers valuable insights into the central tendency, variability, and potential outliers within the "Control" and "Sample" groups, helping to understand their relationships and differences more clearly.*

**Venn diagram**

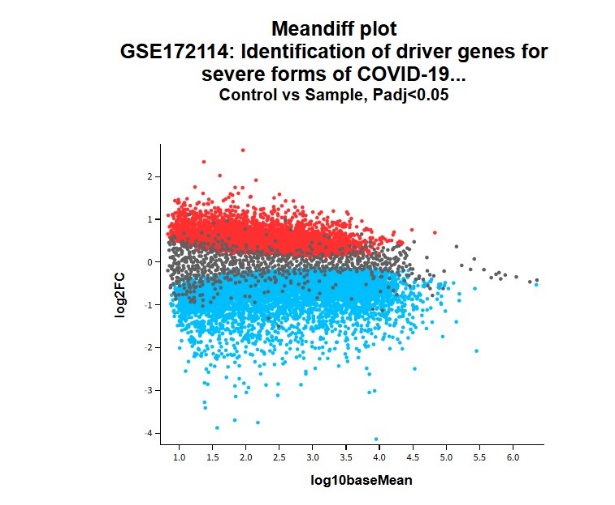
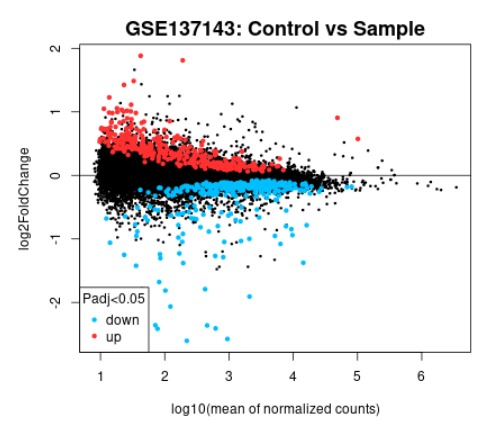
The comparison between GSE172114 and GSE137143 reveals a substantial difference in the number of significant genes.

***Figure 9.*** *In GSE172114, the total genes are 22114 while the significant genes are 10277 that show a higher level of significant genes. And in GSE137143 the total genes are 20214 while the significant are 688.*

*Meandiff plot*

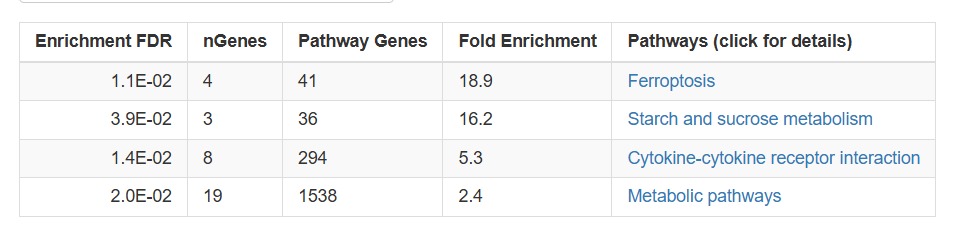
These plots identify genes that are significantly up regulated or down regulated in an experimental condition compared to a control.

***Figure 10.*** *The plots visualize the relationship between log2 fold change (log2FC) and the mean expression level for genes under the comparison of Control vs. Sample with a p value < 0.05 threshold*.

**3.4 ShinyGO analysis**

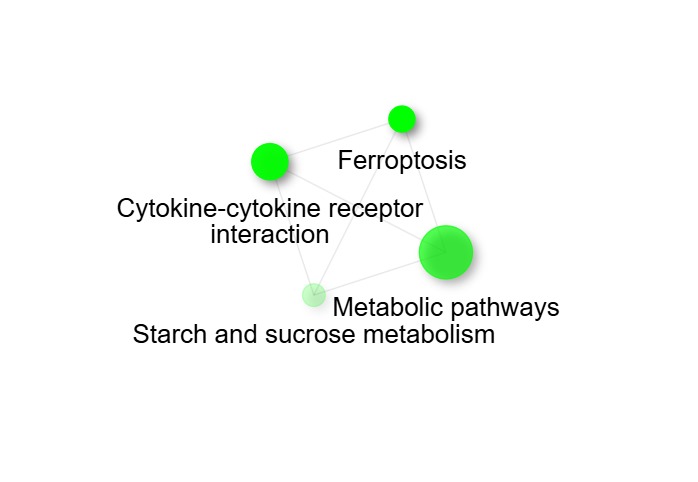
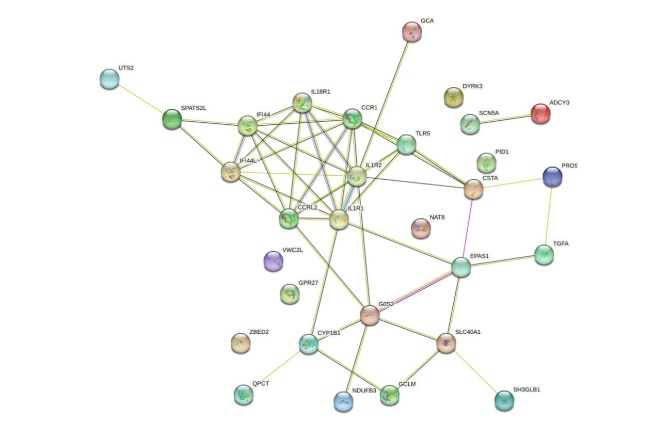
The table represents the result analysis of enrichment FDR of the pathways while they are Ferroptosis, Starch and sucrose metabolism, Cytokine-cytokine receptor interaction, and Metabolic pathways.



***Figure 11.*** *The ferroptosis shows a remarkably high fold enrichment, indicating a significant association with the genes analyzed. The low FDR (1.1E-02) suggests that this result is statistically significant. Starch and sucrose metabolism represents strong fold enrichment and a low FDR (3.9E-03), and cytokine-cytokine receptor interaction has a lower fold enrichment compared to the first two. The last pathway, which is metabolic pathway, has the largest number of genes involved, but the fold enrichment is lower*.

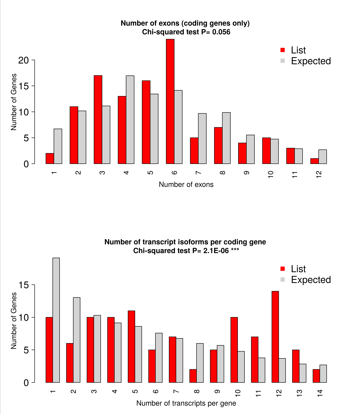
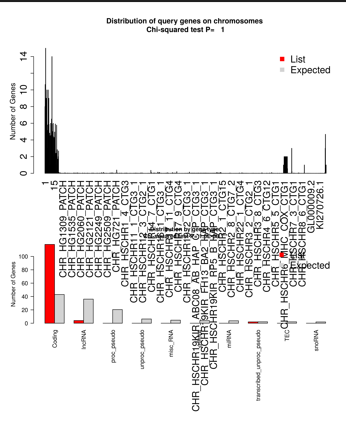
*Relation between pathways*

The following figures represent biological pathways and gene networks. The first image focuses on broader pathways, while the second delves into specific gene interactions.



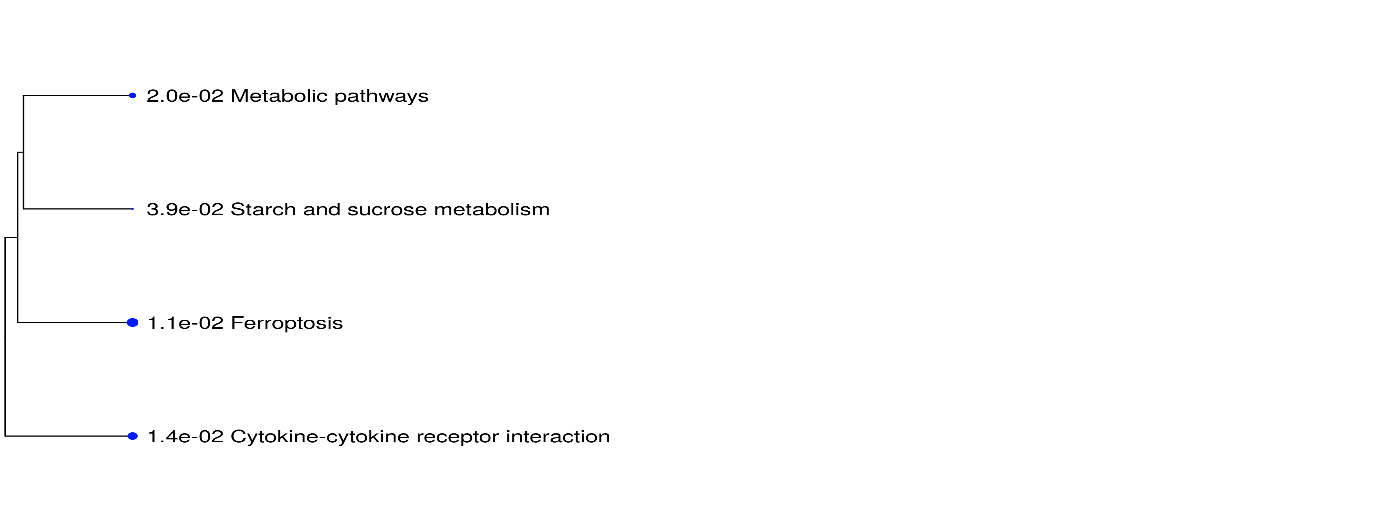
***Figure 12.*** *The networks show interactions with each other according to biological pathways. The proteins are represented as nodes while the interactions are represented as edges and the colors represent the functions of each proteins.*

*Characterizations of query genes*



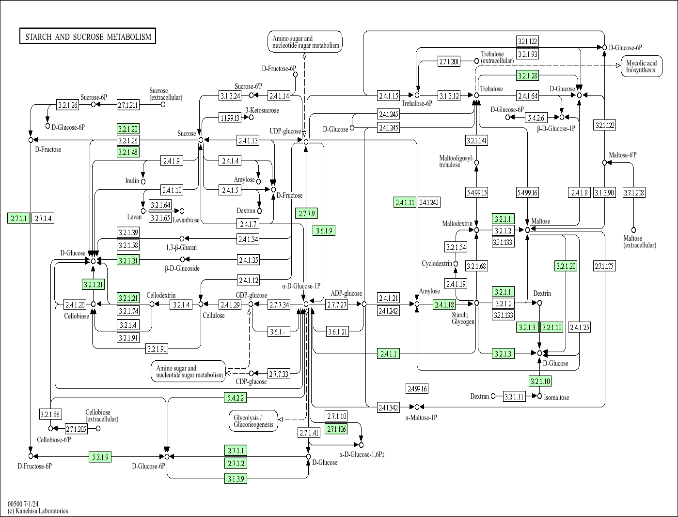
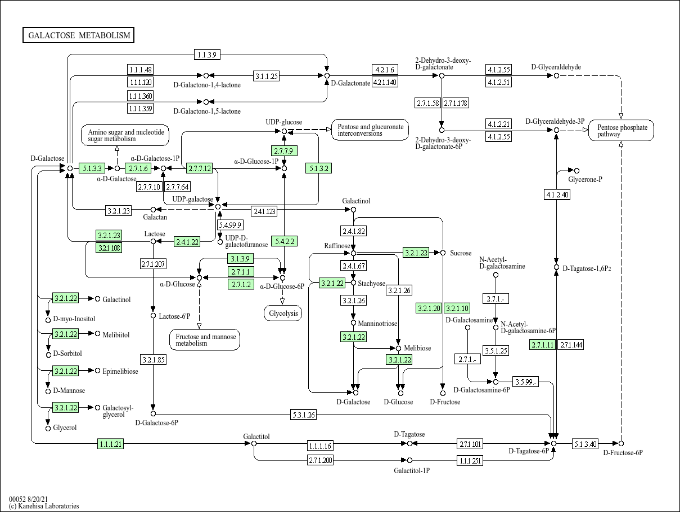
***Figure 13.*** *These graphs show the distribution of query genes, their exons, and transcript isoforms. In bargraph the red bar represents the actual number of genes across different chromosomes and the grey bars represent the expected numbers. The chi square test determines if the distributions are significantly different from the expected values.*

*Tree plot*

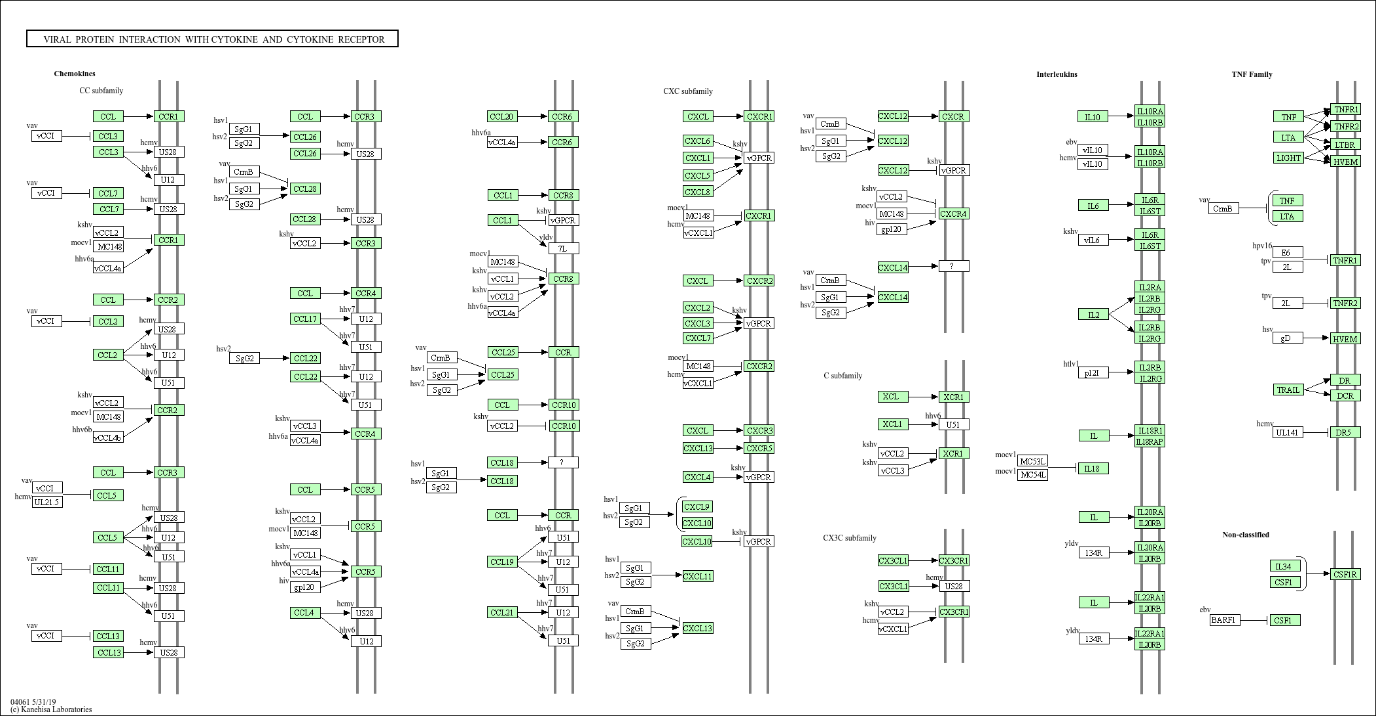
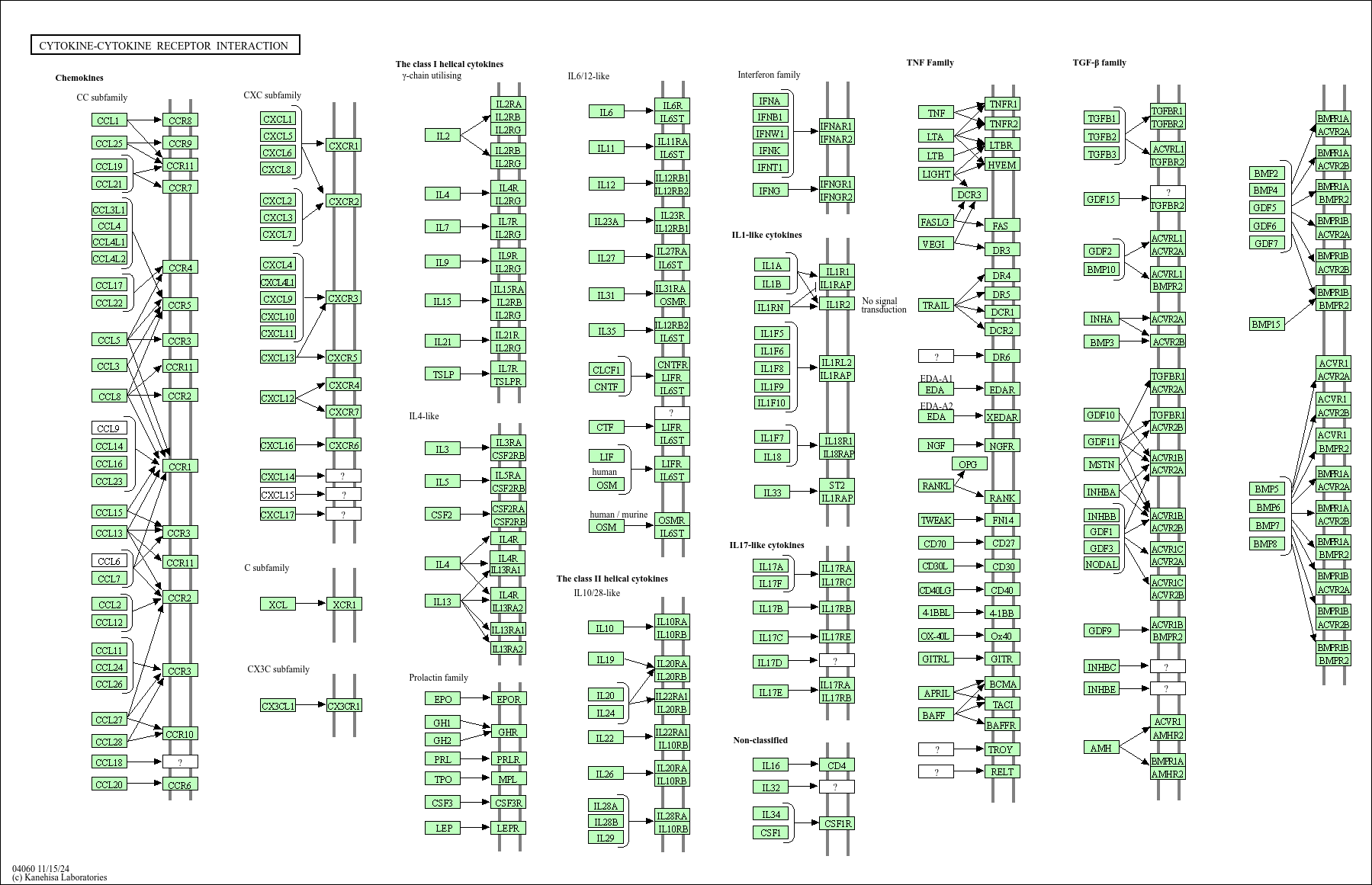
***Figure 14.*** *The tree plot shows the relation between pathways, the nodes represent the biological pathways (metabolic pathways, starch and sucrose metabolism, ferroptosis, and cytokine-cytokine receptor interaction). The number next to each pathway represents the significance value (p-value).*

***3.5 Kegg pathway analysis***

*Galactose, starch, and sucrose metabolism (hsa00052, hsa00500)*

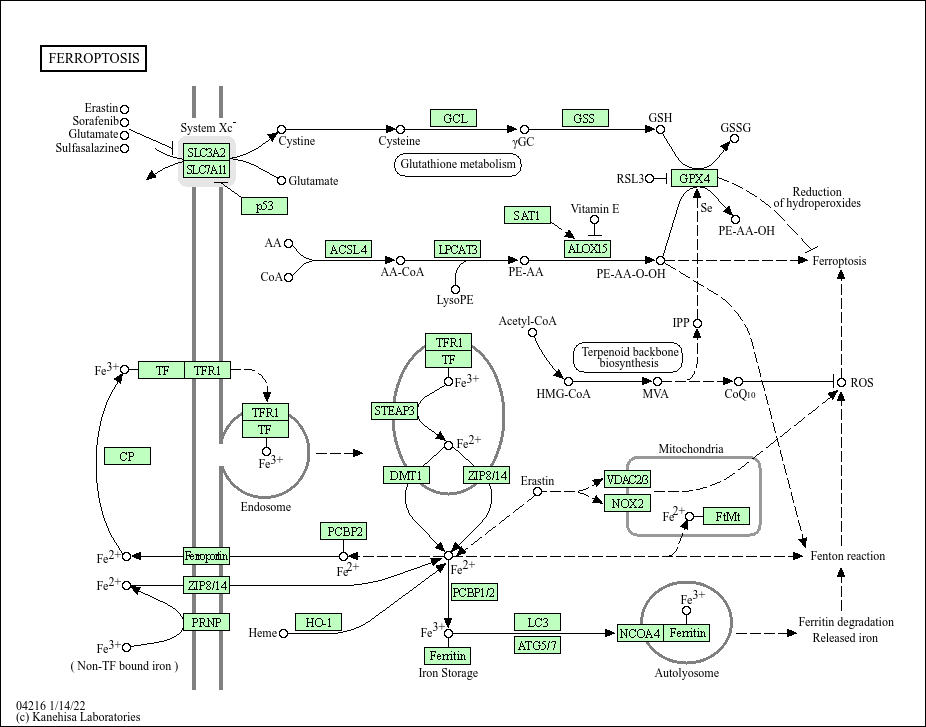


***Figure 15.*** *In COVID the viral infection alters carbohydrate metabolism and affects energy production, according to (hsa00500) the increased energy impact glycogen storage. and in MS the metabolic dysregulation may exacerbate neurodegeneration. The two pathways show impact in immune and nervous tissues that dysregulate sugar metabolism and dysfunction in energy homeostasis.*

*Cytokines interactions (hsa04060, hsa04061)*

***Figure 16.*** *In covid the pathways drive cytokine interaction and interfere cytokine signaling to invade immune responses. In MS, the pathways showed pro-inflammatory cytokines drive autoimmune neuroinflammation, so it disrupts the cytokine pathways. It leads to chronic inflammation and disruption in both hsa04060 and hsa04061.*

*Ferroptosis (hsa04216)*



***Figure 17.*** *In covid the oxidative stress triggers cell death in infected tissues. And in MS the oxidative stress contributes to neuronal death. So, there is a strong relation between both disease since they concentrate in oxidative stress that will lead to cell death.*

**4.0 Discussion**

This study reveals important molecular similarities between COVID-19 and Multiple Sclerosis (MS), showing how both diseases share pathways that may contribute to their progression. Our analysis highlighted specific processes like ferroptosis, altered sugar metabolism, and immune dysregulation through cytokine interactions as key links between the two.

***4.1 Shared Molecular Pathways***

Ferroptosis, a process associated with oxidative stress, was prominent in both diseases. In COVID-19, it leads to tissue damage, while in MS, it contributes to neuronal death and demyelination. This points to antioxidant therapies as a potential treatment avenue for both conditions.

Cytokine interactions also play a significant role in both diseases. COVID-19’s cytokine storms intensify inflammation, while in MS, pro-inflammatory cytokines fuel autoimmune damage. Drugs targeting cytokine pathways could help manage inflammation in both contexts, especially for MS patients with COVID-19.

Metabolic disruptions, such as changes in starch and sucrose metabolism, further connect the two diseases. COVID-19 alters energy production, while MS experiences metabolic dysregulation linked to neurodegeneration. These findings suggest that targeting metabolic pathways could help address symptoms in both diseases.

***4.2 Gene Expression Patterns***

Gene expression analysis showed more significant changes in COVID-19 compared to MS, reflecting the systemic nature of viral infections. Despite this, overlapping genes and similar patterns in expression suggest shared molecular mechanisms that affect immune and neurological functions in both diseases.

***4.3 Insights from Plots***

The data visualizations provide critical insights into the gene expression profiles and variability in COVID-19 and MS:

**4.3.1 Histogram and Volcano Plots**  
The histograms and volcano plots highlight distinct patterns of gene expression. COVID-19 (GSE172114) demonstrated a significant number of differentially expressed genes with low p-values, reflecting its systemic and acute impact on cellular pathways. MS (GSE137143), however, exhibited fewer significant changes, consistent with its chronic and localized nature. The volcano plots further illustrate these differences, with COVID-19 showing a larger number of genes undergoing significant upregulation or downregulation compared to MS.

**4.3.2 Mean-Variance and UMAP Analyses**  
The mean-variance plots revealed higher dispersion levels in COVID-19, which aligns with its variable clinical manifestations and immune responses. In contrast, MS demonstrated a more consistent and constrained variability. UMAP visualizations echoed this finding, showing intermixed control and sample clusters for COVID-19, suggesting broader variability, while MS exhibited tighter and more distinct clustering.

**4.3.3 Boxplots and Venn Diagram**  
Boxplots displayed differences in central tendency and variability between control and sample groups, aiding in identifying significant expression patterns. The Venn diagram further emphasized the disparity in the number of significant genes, with COVID-19 showing a higher number of impactful genes, reflecting its widespread biological effects.

**4.3.4 Meandiff Plot**  
These plots highlighted genes with substantial upregulation or downregulation, offering insight into the differential molecular responses to COVID-19 and MS. The more extensive changes observed in COVID-19 suggest its broader impact on cellular processes.

***4.4 ShinyGO Results***

The ShinyGO analysis elucidated critical molecular pathways and protein interactions shared between COVID-19 and MS:

**4.4.1 Pathway Enrichment**  
Four primary pathways—ferroptosis, starch and sucrose metabolism, cytokine-cytokine receptor interaction, and general metabolic pathways—were identified. Ferroptosis exhibited strong fold enrichment, reflecting its critical role in oxidative stress-mediated cell death in both diseases. Similarly, dysregulated carbohydrate metabolism impacts energy homeostasis in COVID-19 and MS, contributing to immune and neurological dysfunction.

**4.4.2 Cytokine-Cytokine Receptor Interaction**  
This pathway underscores shared immune dysregulation mechanisms. COVID-19 is characterized by cytokine storms, while MS involves chronic neuroinflammation mediated by cytokine signaling. The overlapping roles of cytokines highlight a shared disruption in immune system balance.

**4.4.3 Protein-Protein Interaction Networks and Hub Proteins**  
The protein-protein interaction (PPI) network analysis identified key hub proteins, including CCR1 and IL1R2, which showed multiple connections. These hub proteins are critical in coordinating shared molecular mechanisms between the two diseases. CCR1 and IL1R2 may regulate immune responses and inflammatory cascades, representing promising targets for therapeutic intervention. Understanding their roles and interactions provides valuable insights into the molecular crosstalk between COVID-19 and MS, potentially paving the way for strategies targeting these pivotal proteins.

**4.4.4 Characterization and Tree Plot Analysis**  
The distribution of query proteins across pathways and their evolutionary significance was explored through the tree plots. The interconnections among pathways, such as ferroptosis, cytokine-cytokine receptor interactions, and metabolic processes, further illustrate the biological interdependence of these mechanisms in disease progression.

***4.5 Treatment Implications***

Therapeutic opportunities arise from these findings. Ferroptosis inhibitors could reduce oxidative damage, while cytokine inhibitors, such as IL-6 blockers, may mitigate inflammation. Restoring energy metabolism might also provide benefits by addressing metabolic imbalances common to both conditions.

***4.6 Limitations and Future Work***

This study relied on publicly available datasets, which may not fully capture the complexity of these diseases. Validation in larger cohorts and experimental studies is needed. Additionally, exploring proteomic and metabolomic data alongside transcriptomics could deepen our understanding of how these diseases interact.

**5.0 Conclusion**

The present research represents significant molecular links between covid 19 and multiple sclerosis (MS) and showed new pathways and gene expressions levels. In the analysis of gene expression and pathway enrichment data, we identified processes such as ferroptosis, cytokine-cytokine receptor interactions, and endocrine metabolic pathways which were in both diseases. Exposure to viral infections such as COVID-19 appears to make autoimmune and neuroimmune disorders like MS. They share mechanisms that include oxidative stress, immune dysregulation, and neuroinflammation. This information improves our understanding of these diseases and provides designing intervention strategies towards the common pathways. However, additional studies are required to test these hypotheses and more accurately assess the apparent effects of these common mechanisms on the COVID-19 and MS diseases, which may allow for better treatment of MS patients infected with coronavirus.

**References:**

Hollist, M., Hollist, A., Au, K., Betts, C., Kirmani, M., Kirmani, M., Armour, B., Udeh, M. C., & Kirmani, B. F. (2024). Multiple Sclerosis and COVID-19: An Overview on Risk, Severity, and Association With Disease Modifying Therapies. *Neuroscience insights*, *19*, 26331055241265668. <https://doi.org/10.1177/26331055241265668>

Prosperini, L., Arrambide, G., Celius, E. G., Goletti, D., Killestein, J., Kos, D., Lavorgna, L., Louapre, C., Sormani, M. P., Stastna, D., Ziemssen, T., & Di Filippo, M. (2024). COVID-19 and multiple sclerosis: challenges and lessons for patient care. *The Lancet Regional Health - Europe*, *44*, 100979. https://doi.org/10.1016/j.lanepe.2024.100979