FACULTY OF HEALTH & LIFE SCIENCES

BACHELOR'S IN BIOINFORMATICS (HONS)

SYSTEM ANALYSIS & DESIGN METHOD LCS20401P

FINAL PROJECT:

SYSTEM DESIGN ANALYSIS OF DRUG AGAINST COVID-19

NAME: DHANYASHRI A/P GURUPARAN ID:

012022020020

LECTURER: DR SURESH KUMAR

SUBMISSION DATE: 12/8/2024

1.0 INTRODUCTION

COVID-19 is, undoubtedly, an infectious disease caused by a new coronavirus that shook the world since it first emerged in late 2019. COVID-19 is caused by the SARS-CoV-2 virus, a member of the coronavirus family. According to Ciotti et al. (2020), the coronavirus represents a vast family of viruses that include both common cold and serious pathologies, such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) (Ciotti et al., 2020). It is supposed that SARS-CoV-2 has initially originated from bats, being subsequently transmitted to humans by an intermediate host, probably including wild animals sold at the seafood market in Wuhan, China.

The virus was first discovered in December 2019 after several cases of pneumonia of unknown cause had been reported in Wuhan. By January 2020, the World Health Organization had declared it a Public Health Emergency of International Concern and escalated it to a pandemic by March 2020. Up to early April 2020, about 1,197,405 confirmed cases of COVID-19 had been reported globally. The majority of the infected showed mild symptoms, while some of them had severe complications, and others were critical. As of late August 2023, the number of reported cases has risen above 770 million, while the number of deaths is above 6.9 million globally (Weekly Epidemiological Update on COVID-19 - 10 August 2023, n.d.).This rapid spread of COVID-19 has been contributed to by human-to-human transmission via droplets but majorly by contact with contaminated surfaces. The virus has a high transmission rate, contributing to its global spread and the significant public health challenges it has posed.

As the pandemic began to unfold, it had a huge impact on communities and healthcare systems across the world, causing extensive social distancing measures to be enacted, leading to widespread closures of schools and businesses, potentially exposing and increasing Access gaps and resource disparities. Admissions due to COVID-19 healthcare facilities were thus put under severe strain, including postponed medical care for non-COVID conditions and remarkable burnout among healthcare workers (Bloom et al., 2023). Even so, the crisis has also created a catalyst for telemedicine innovations, increased public health awareness, and given impetus to community resilience, underlining at once the challenge of preparedness and adaptability in global health management.

On this very note, the development of COVID-19 therapeutic drugs presents a frontline global health strategy, which comes with critical treatment options reducing the severity and the length of illness. These therapies not only reduce hospitalization and mortality rates, in particular, for high-risk groups, but also alleviate some stress from health systems all over the world. Successfully realized treatments not only back up the existing vaccines through

treating active cases but also stimulate economic recovery due to the early return of people to normal life. Furthermore, the continuous development of antiviral drugs is an essential element of the pharmaceutical response to COVID-19, mainly small molecule compounds.

Thus, this system design approach in developing these antiviral drugs is based on fast identification, synthesis, and scaling up of certain effective treatments. It pictures high-throughput screening using existing manufacturing processes with compounds that act on key viral proteins, acting against any type of viral replication. Small molecules will be quite effective due to their oral bioavailability and tissue penetration capability. Inexpensive, and intersubjectively amenable to accelerated regulatory approval—in particular, for repurposing drugs—small molecule-based therapies would be easily distributable globally. They could also be adapted for combination therapies, which would greatly increase their reach against evolving viral threats and potential future pandemics.

2.0 OBJECTIVES

The primary objective of this study is to design and synthesis small molecule compounds that manage to successfully target key viral proteins of the SARS-CoV-2 virus. An important component of this objective is ensuring enhanced oral bioavailability and tissue penetration for these antiviral drugs; these are critical factors in their effectiveness for treating COVID-19. Another major objective is accelerating the approval life cycle for these drugs by relevant regulatory authorities. It will, therefore, prioritize repurposing of existing compounds that can be very fast-tracked to licensure and accessibility in global markets. Ensuring also that these treatments are affordable and accessible worldwide is very key, especially in low-resource settings where the impact of the pandemic has been bad enough. To this end, the last objective of this project is the exploration of combination therapies. On this respect, the approach tends to acquire more effectiveness against the various strains of the virus by using some strategic mix of new drugs that are developed and minimize the possibility of drug resistance. This is in line with what is expected: dynamic and adaptive treatment strategies respond dynamically to an evolving pandemic like COVID-19.

3.0 METHODOLOGY

Data Sourcing

In the initial phase of this study, the SMACC database (https://smacc.mml.unc.edu/) was resorted to as a key resource for sourcing the overlap phytochemical dataset. The SMACC database is known to have one of the largest repertoires of diverse phytochemical datasets, very useful and often solicited to jump-start studies cutting across multiple disciplines. More specifically, in this case, it has successfully gathered, harmonized, and collated chemogenomic data from ChEMBL on 13 emerging viruses recognized as serious threats to human health. This integration took place with many challenges in data annotation accuracy to ensure the creation of a highly curated, well-annotated database. This is a database of compounds tested either in phenotypic or target-based assays. In this study, after downloading the dataset in an Excel file, the focus falls on the SMILES notation data by selecting 35 relevant entries involved in both the active phenotypic assays and target-based assays for detailed subsequent analysis (Martin et al., 2023).

Gene Expression Analysis

An analysis of gene expression was done to identify genes that were either up- or down-regulated in correlation with the canonical SMILES, obtained from the previous method. Each canonical SMILES entry was put through an in silico DIGEP-Pred platform, (https://www.way2drug.com/GE/), which predicts compound effects on gene expression based on structural formulas. DIGEP-Pred utilizes PASS, a tool predicting the biological activity of compounds by the Bayesian approach with a large training set of biologically active compounds. Filtering in results of protein-based prediction, with a probability > 0.5 only, Pa > 0.5, ensures reliability. The software ranked these compounds as likely to be either up- or downregulated at the gene level. This gave insight into the molecular mechanisms of actions and potential therapeutic targets of these compounds. Such output was organized into a table, with genes up- and down-regulated listed accordingly for further analysis (Lagunin et al., 2013).

Correlation Analysis of Gene Regulation with COVID-19 Genes

To determine the correlation between the results from the upregulated genes compared to downregulated genes with COVID-19-related genes, the upregulated genes compared to downregulated genes results obtained in the previous section were gathered together in a dataset known as Set A. On the other hand, COVID-19-related genes were retrieved from the Comparative Toxicogenomics Database (CTD), an integrated resource that provides information about the relationships between diseases and genes and gene product interactions with chemicals (https://ctdbase.org/). A sum of 16,857 COVID-19-related genes was downloaded in Excel format to produce Set B. Both gene sets were formatted to be comparable in a Venn diagram tool that elaborated on the relationship between the gene sets (Davis et al., 2023). The Venn diagram visually compared the commonalities and differences, displaying the number of shared genes, set A versus set B, indicating those up- and down-regulated genes identified from the SMILES analysis which were also found for COVID-19.

Druglikeness Analysis

Druglikeness was evaluated for the 35 pertinent entries using SMILES with the help of the related web tool SwissADME (http://www.swissadme.ch/). All 35 entries have been treated individually by input into the site SwissADME for evaluation of key druglikeness parameters such as Lipinski's Rule of Five, Ghose filter, Veber filter, Egan filter, and Muegge filter. These filters function to predict whether a compound would be potentially good for drug development (Daina et al., 2017). At the end, the results returned detailed reports for each of the entries on whether it satisfied all the druglikeness criteria. Only those SMILES entries that received a "Yes" under all the criteria were chosen appropriate for further study, and those with any violations were not considered. The SMILES entries, which scored all the checks for druglikeness, were combined into a list, hence making it a point that only counterparts with favorable pharmacokinetic properties and drug developmental potential were taken under consideration in the final research.

Enrichment Analysis

The sets of upregulated and downregulated genes were further taken for gene enrichment analysis using the Enrichr web tool. Enrichment analysis is a computational method to gain knowledge about input gene sets by comparing them with annotated gene sets that represent prior biological knowledge. This was done to verify if the input set of genes does significantly overlap with another annotated gene set. The answer to whether such genes

overlap with known human pathways was provided by using the KEGG database of human pathways. Enrichr provided enrichment results, including various pathways, like their associated genes and p-values for statistical relevance in overlap, as depicted. The same gene set was carried forward for gene ontology category analysis with GO Biological Processes, GO Cellular Components, and GO Molecular Functions to identify the functions of genes within cellular contexts and biological functions. Top 10 results from each pathway and ontology category were tabulated, including genes along with their p-values.

Network Analysis

An interaction network between proteins of the upregulated and downregulated genes was further visualized and analyzed by putting the identified genes into Cytoscape, a software platform for visualizing complex networks and integrating attribute data (Shannon et al., 2003). Construction of the PPI network was done on the base of STRING interaction data, depicting molecular interactions inside the cells. The most influential nodes of the PPI network were identified using the CytoHubba plugin and ranked top 10 in terms of centrality and influence within the network by the Maximum Clique Centrality algorithm. The key proteins, major regulators, and top-ranked nodes—the ones that have places in this top ten—are noted down to further understand the biological significance and their potential for roles in these genes' regulatory mechanisms.

Validation of Top Genes Through Enrichr Analysis for COVID-19 Association and Virus-Host Interactions

The top 10 genes identified in Cytohubba were used as an input to further validate, for extra enrichment analysis. Precisely, it was to assess whether certain key genes that played a vital role in COVID-19 expression were overlapping with disease-related gene sets, especially COVID-19 Related Gene Sets and Virus-Host PPI P-HIPSTer 2020 database. The COVID-19 Related Gene Sets in Enrichr were genes related to COVID-19, according to experimental and computational studies that give insights into the ways by which the identified genes might be involved in the pathology of COVID-19. Overlap with such gene sets has been checked for the identification of potential associations between top-ranked genes and COVID-19, hence their involvement in the context of this disease. Added to this, the database Virus-Host PPI P-HIPSTer 2020 is a store of protein–protein interactions that have already taken place between viral proteins and host proteins, including key interactions necessary for viral infection and replication. Using this comparison with the database Virus-Host PPI P-HIPSTer

2020 provided an opportunity to determine whether any such genes interact with viral proteins, exploring a set of mechanisms by which their genes exert their effect on viral processes.

The outcomes from these analyses provided valuable insights into the significance or relevance of the genes identified concerning the infection context and their possible roles during disease mechanisms, hence validating them across broader biological and pathological landscapes.

4.0 RESULTS

Gene Expression Analysis

The gene expression analysis identified a total of 55 genes that were either up-or down-regulated in response to the tested compounds. Of these genes, 40 were up-regulated while 17 others were down-regulated by gene expression analysis. Yet, taking into account the duplicates found in both lists of up-regulated and down-regulated genes leads to a sum of 53 distinct genes. This distribution thus suggests that the compounds might be more biased toward suppression than induction. There were also several genes for which no significant changes in expression were observed, either due to the specificity of the compounds or due to threshold setting for considering changes significant. These results can provide very valuable insight into possible molecular mechanisms and therapeutic targets of the compounds (Table 1).

Correlation Analysis of Gene Regulation with COVID-19 Genes

The Venn diagram comparison established that 53 out of 55 genes involved with Set A were associated with genes in Set B relevant to COVID-19. This ensures a huge overlap and suggests that indeed it is very likely these are genes for COVID-19. Only the 2 genes FDNC4 and SMN2 did not show a similar concordance between the two sets. This point only highlights the major association between the genes we identified and COVID-19, supporting their potential importance in understanding the disease mechanisms (Figure 1).

Druglikness Analysis

The results indicated that only 3 of the 35 SMILES entries agreed with all of the druglikeness criteria and had all parameters at "Yes" (CHEMBL1324, CHEMBL1458891, and CHEMBL3818159). The rest of the entries had violations in at least one of the parameters; hence, they were not fit for further analysis. This outcome demonstrates very few compounds with favorable pharmacokinetic properties and potential for drug development out of the screened compounds, which is oriented to focus on these four promising lead compounds for the final analysis.

Enrichment Analysis

These findings indicated that not all of the 55 upregulated and downregulated genes were annotated for the three analyses: KEGG pathways, GO Biological Processes, GO Cellular Components, and GO Molecular Functions. Only a subset had important annotations in such categories, while others repeated the same gene in the results. While many genes in this set are very well annotated to match the known pathways and functions, others either lack such annotations or were redundantly represented in this analysis (Table 2 – Table 5).

Network Analysis

The constructed protein-protein interaction (PPI) network denoted complicated molecular interactions within the cells (Figure 2). The top 10 of the PPI network were ranked concerning its centrality and influence according to the Maximum Clique Centrality algorithm (Figure 3). The major regulators among these top-ranked nodes were key proteins; their biological significance and probable roles in regulatory mechanisms regarding the genes under study were noted for further understanding. This network formation and top 10 node identification gave insight into the critical components and interactions in this studied gene set.

Validation of Top Genes Through Enrichr Analysis for COVID-19 Association and Virus-Host Interactions

These results showed that from among these highest-ranked genes, PLAU, CCL2, HMOX1, TIMP1, CTNNB1, and PLAT were significantly overlapping with disease-related gene sets and interactions crucial for viral infection and replication (Table 6 & Table 7). This overlap indicated likely involvement in COVID-19 pathology and interaction with viral proteins; such genes are likely to be of importance for the understanding of mechanisms of disease and their regulatory roles.

5.0 DISCUSSION

5.1 Overview of the Phytochemical Likely Genes and Their Categories

The results for the "Phytochemical-Likely" genes, 40 of which were upregulated by specific phytochemicals and 17 downregulated. To further understand these genes, it has been classified according to their biological functions, pathways, or diseases with which they are associated into broad categories such as immune response, metabolism, signal transduction, cell cycle and DNA repair, oxidative stress response, apoptosis and cell death, and others.

Genes that induce the activation of immune response and that regulate the body's defense in those upregulated includes genes such as CD86 involved in the activation of T-cells (Lim et al., 2012), while CCL2 recruits other immune cells to areas of inflammation (*Pozzi & Satchi-Fainaro, 2024*). Other genes are CD83 and CD14. In this case, the latter is specific to the immune response of dendritic cells, and the former acts as a co-receptor to detect bacterial lipopolysaccharides. Furthermore, TNFRSF1A is a receptor that has a role in apoptosis induction and the host immune response (Preedy et al., 2024). The downregulated genes such as NOS2 and FLT1 are found to be involved in nitric oxide production and modulate host immune responses (Bogdan et al., 2000).

Many genes from the subset of metabolism genes relate to basic cellular activities concerned with energy and nutrient management. Upregulated genes include ADIPOQ, which is involved in glucose homeostasis and fatty acid breakdown (Wang et al., 2020), and ABCA1, essential for cholesterol efflux. GH1 was also highly represented and has important functions in growth and metabolism (Vijayakumar et al., 2011). Other upregulated genes have been reported to link to metabolic control and reproductive growth possibly: FNDC4 and INHBE. On the other hand, CAT, which is responsible for fighting oxidative stress (Nandi et al., 2019), and SREBF2, involved in lipid metabolism, are downregulated genes, underscoring their diminished roles in metabolism in some conditions. DHFR, necessary in DNA synthesis and repair through its role in folate metabolism, is also downregulated (*Askari & Krajinovic, 2010*).

Additionally, signal transduction pathways act through genes that regulate cellular responses to exogenous signals. Upregulated genes, including RARA and VDR, are involved in the regulation of transcription and calcium/phosphate levels, respectively (Tan et al., 2021). REN participates in blood pressure regulation via the renin-angiotensin system (Kanugula et al., 2023), while PLAT takes part in the breakdown of blood clots(The Basics of Blood Clots:

What You Need to Know | NIH MedlinePlus Magazine, n.d.). Participating in signal transduction and cell skeleton organization is RAC1. As for KRT18, it belongs to those necessary factors for both cellular structure and integrity. ID1 participates in the regulation of cell differentiation and proliferation (Zhao et al., 2020). Some of the downregulated genes included NR3C1, a gene involved in regulating metabolism and immune responses, and MMP7, which participates in degradation of the extracellular matrix component (Quatrini & Ugolini, 2020).

Meanwhile, genes involved in cell-cycle regulation and DNA repair assure proper cell division and maintenance of genomic stability. Among these upregulated genes are CDK4 and TP73, cell-cycle regulators linked to apoptosis (Baker et al., 2022) , while SMN2 functions in RNA processing. On the other hand, there are CHEK1, TOP2A, and CTNNB1 genes whose expressions are down-regulated related to cell cycle arrest, DNA replication, transcription, and cell proliferation that argue, in this way, for reduced activity under certain conditions (Zhou et al., 2020).

The genes protect the cell against oxidative damage and may need to be upregulated in the stressed tissue such as NFE2L2, while PRDX4 and HMOX1 are genes protecting against oxidative stress (Chiang et al., 2021). On the other hand, KEAP1 were downregulated, suggesting reduced gene expression of the oxidative stress response. Apoptosis is programmed cell death influenced by genes, such as SERPINA3 and CLU, both upregulated with roles in inflammation and apoptosis. CLU has also been implicated in neurodegenerative diseases. AR has tissue-specific impacts on apoptosis. SIVA1 was downregulated, showing it had a decreased role in regulating apoptosis (Lee et al., 2023).

However, some of the genes, do not fit neatly into the above categories but also in important ways play their roles in various biological processes. The upregulated genes include the following: WIPI1, involved in autophagy; TIMP1, mechanistically affecting the behavior of cells and matrix remodeling; and PLAU, involved in proteolysis, cell migration, and tissue remodeling. On the other hand, CBR1 participates in drug metabolism and detoxification, while EPAS1 responds to oxygen levels and TNNI3 in heart muscle contraction. The downregulated genes include IVL, associated with the formation of the cell envelope of epidermal cells, and SELL, associated with lymphocyte adhesion and trafficking.

5.2 Gene Association with Covid-19

In the analysis, all the genes identified as upregulated or downregulated are related to the gene set that characterizes COVID-19, which could implicate them in a host response against the virus. These genes could be relevant to the complex physiological changes during a COVID infection in their activity modulating immune response, metabolism, signal transduction, regulation of the cell cycle, response to oxidative stress, and apoptosis. These genes are hence very significantly correlated with the COVID-19 gene set, underlining their relevance to understanding the molecular mechanisms of this disease.

But two specific genes, FDNC4 and SMN2, did not correlate with the COVID-19 gene set. Indeed, FDNC4, playing a metabolic control and reproductive growth role, and SMN2, being linked with RNA processing, are outstanding since they do not implicate the same way as other genes in response to COVID-19. Indeed, the absence of said relationships can suggest that either the pathways or biological processes governed by FDNC4 and SMN2 do not in some way interrelate or make a considerable contribution to the pathogenesis of the course of the disease caused by the SARS-CoV-2 virus.

The absence of FDNC4 and SMN2 in the COVID-19 gene set could either indicate that these genes are not responsive to the virus or be engaged in pathways irrelevant to the life cycle of the virus and to the immediate host response to the viral invasion. Part of the explanation could be that FDNC4 is peripheral to acute metabolic demands or stress responses induced by COVID-19. Similarly, while RNA processing is an important cellular activity, the contribution of SMN2 specifically may not overlap with any of the RNA-related activities co-opted or modified by the SARS-CoV-2 virus upon infection.

It makes a difference because this narrowing down of focus differentiates those genes in the context of COVID-19 that are most relevant and possibly cues future investigations toward those pathways and processes more directly implicated in it. On the other hand, understanding how FDNC4 and SMN2 do not appear to have relevance to the set of COVID-19 genes could provide myriad insights into the specific biological mechanisms of their action and probably explain why they might stay outside the response network being induced by the virus.

5.3 Druglikeness Analysis

Druglikeness was utilized to define whether a compound may become a good drug by structures and chemical properties. The three compounds that passed the parameters of druglikeness are CHEMBL1324, CHEMBL1458891, and CHEMBL3818159 where its molecular weight was satisfactory and lipophilicity balanced, with optimal hydrogen bond donors and acceptors. Taken as a whole, these properties tend to predict good oral bioavailability and effective absorption, with the potential for interaction with biological targets in these compounds, which will be of interest for further drug development.

On the other hand, concerning gene associations, CHEMBL1324 is associated with 14 different genes, reflecting its wide range of potential biological activities. These genes include proteins involved in the regulation of the extracellular matrix, such as PLAT, a participant in blood clotting, and CAT, a cell protector from oxidative damage. In addition, CHEMBL1324 affects ADIPOQ, a gene related to the regulation of glucose levels and fatty acid breakdown, pointing at its possible potential for metabolic pathways. That relation of CHEMBL1324 to the very wide gene association range could suggest that it would have the potential to affect several biological processes and, hence, needs to be considered a prospective drug candidate for the treatment of some rather complex conditions, such as diverse mechanisms. This bodes too high an activity potential that implies a greater potential for side effects through its interaction with such a large number of targets.

CHEMBL1458891 has much more a concentrated gene association, majorly to ADIPOQ, suggesting an influence over metabolic processes. This could then imply that CHEMBL1458891 might be beneficial in treating metabolic disorders, for example, obesity or type 2 diabetes. This kind of focused action on ADIPOQ would suggest this compound would have less off-target activity compared with compounds having broader activity profiles and therefore a potentially better safety profile.

Meanwhile, CHEMBL3818159 shows a dose range of genes involved in the DNA damage response, including CHEK1, a central cell cycle regulator and key factor during DNA repair; genes involved in immune function; and genes involved in metabolic processes, including ADIPOQ. This suggests that this compound may have therapeutic benefits for conditions in which there is DNA damage, immune dysregulation, or metabolic disturbance. Its balanced interaction with multiple biological pathways could make this compound versatile

for the treatment of many different diseases.

One very prominent commonality in all these three compounds is their co-expression with ADIPOQ, indicating that they might have some potential shared influence on metabolic pathways. Thus, these compounds might exert beneficial effects on metabolism, and this could be important in metabolic disorders. CHEMBL1324 and CHEMBL3818159 have quite divergent gene associations, which implicates pleiotropic effects—a single compound influencing multiple pathways. This may be good when fighting complex diseases, but it also increases the risk of side effects.

Although the results of computational predictions for druglikeness of such compounds are quite promising, their full pharmacokinetic, efficacy, and safety profiles will need to be comprehensively established in experimental studies as potential drug candidates.

5.4 Enrichment Analysis

Several of the pathways identified participate in important biological processes and diseases at KEGG. For example, upregulated pathways, such as "Lipid and Atherosclerosis" and "Complement and Coagulation Cascades", reveal high activities of lipid metabolism and blood clotting mechanisms, respectively, in which ABCA1, PLAT, and CCL2 were key genes in their respective areas of influence. ABCA1 and PLAT upregulation would indicate enhanced lipid transport and fibrinolysis, while CCL2 would indicate heightened inflammatory responses. On the other hand, downregulated pathways that include "Fluid Shear Stress and Atherosclerosis" and "Negative Regulation of Blood Coagulation" may indicate disturbances in lipid metabolism, regulation of cardiovascular function, and immune response. The genes whose downregulation was detected are ABCA1, NPPB, ADIPOQ, and CD86. The expression products of these genes are responsible for lipid transport and blood pressure regulation, and also for immune function. This differential regulation underlines the resilience of the complex interplay between cardiovascular risk, inflammation, and metabolic dysfunction—not to mention potential targets for therapeutic intervention. These are complex interaction pathways as potential therapeutic targets for cardiovascular health, cancer, modulation of immune responses during ageing processes, and oxidative stress management.

In the "GO Biological Process", it is seen that changeable processes are significant with blood coagulation and inflammatory responses. These up-regulated processes, like "Negative Regulation of Blood Coagulation," include genes that can have functions attributed to PLAT, CCL2, and MMP7 in module-enhanced clot breakdown and inflammation.

Downregulated processes, such as "Collagen-Containing Extracellular Matrix" and "Platelet Alpha Granule Lumen," likely point to decreased remodelling of the extracellular matrix and impaired blood clotting functions. These genes include ABCA1, whose expression is reduced, along with that of NPPB and ADIPOQ. This differential regulation suggests shifts in vascular health, tissue repair, immune reaction, and interrelation among inflammation, coagulation, and cardiovascular function.

Whereas in the "GO Cellular Component" category, upregulated components such as the "Collagen-Containing Extracellular Matrix" and "Platelet Alpha Granule Lumen" include genes like MMP7 and PLAT, thus suggesting increased extracellular matrix remodeling and enhanced blood clotting components. On the other hand, the downregulated components such as "Fibrinolysis" and "Platelet Alpha Granule Lumen" depict reduced activity in these very areas with genes like ABCA1, NPPB, and ADIPOQ showing less activity. These changes likely show up at the cellular structural level, pertinent to tissue repair and blood clotting, and could be pointing toward some potential vascular or extracellular matrix dynamic disruptions.

Finally, in GO Molecular Function," upregulated processes such as "Serine-Type Endopeptidase Activity and Receptor Ligand Activity" point out to higher enzymatic activities and that of receptors, with genes such as PROC and CD86 at their center. Such functional enrichments are associated with either immune activity or enzyme activity. The downregulated functions include "Serine-Type Peptidase Activity" and "Receptor Binding," indicating decreased activity in these areas as seen for genes CAT, NOS2, and RARA. This means that vital cellular survival and immune response signaling functions are turned down, hinting at possible flaws in the regulation of oxidative stress and immune functions.

5.5 Analysis of Top Regulatory Proteins in PPI Network

The top 10 nodes of the PPI network correspond to the most central and, hence, most influential proteins, without which no regulation mechanism could be understood. By MCC algorithm, these proteins each play important roles in certain biological processes.

For instance, CTNNB1 participates in the Wnt signaling pathway, which controls the growth and differentiation of cells, often related to the occurrence of cancer. REN is critical for blood pressure but could also have effects on cell growth pathways. LEP is involved in energy balance and inflammation. CCL2 recruits' monocytes to sites of inflammation, and TIMP1 inhibits breakdown of the extra-cellular matrix, influencing tissue remodeling and metastasis.

ADIPOQ, the anti-inflammatory adipokine, is critical for insulin sensitivity and lipid metabolism; this project passed the drug-likeness test. HMOX1 degrades the pro-oxidant hem during stress responses, while NFE2L2 (NRF2) regulates antioxidant defences. PLAU and PLAT participate in fibrin clot breakdown.

Among the ten nodes identified in the PPI network, all are of significance by their centrality and influence for a network; however, they need to be considered in regard to potential use as therapeutic targets. One of the very key properties which ensures the drugability of a molecule is drug likeness. Out of the top 10 identified nodes in this study, above, it is shown that only PLAT and ADIPOQ passed the test for drug likeness. These two proteins hence fall under as highly potent candidates targeted for drugs, with therapeutic potential to be acted upon. The development of drugs that target other proteins, despite their central roles, may be difficult due to issues of drug-likeness, such as bioavailability, toxicity, or problems in designing small molecules active against them.

5.6 Association of Key Genes in Covid-19 Gene set and Viral Host Interactions

The "Top10 vs Covid-19 Geneset" contains genes up-regulated upon SARS-CoV-2 infection under various experimental conditions, and that of "Virus Host Interaction" with specific interactions of host genes against viral proteins. Their overlap may indicate that some genes contribute to essential functions in the host response to viruses beyond SARS-CoV-2. These genes may thus play critical roles in viral pathogenesis and could be required for the entry, replication, or transmission of the causative virus while evading host immunity.

Among these genes, PLAU (plasminogen activator, urokinase) can be found a number of times in the Covid-19 Geneset, which shows that it is constantly upregulated because of SARS-CoV-2 infection. It means that PLAU takes part in the host response, more precisely in the processes of inflammation or tissue remodeling, both of which are key during viral infection and associated lung damage. Besides, there is prominent appearance of PLAU in the table "Virus Host Interaction", where it interacts with a variety of viral proteins from different viruses, for example, Human cosavirus and Vaccinia virus. That PLAU recurs within both datasets suggests it may represent a common target for viral manipulation across different families, implicating this gene in pathways that viruses commonly exploit to promote infection or to evade immune responses (Fang et al., 2021).

Similarly, the protein PLAT (Plasminogen Activator, Tissue) is also shared by both gene sets, but much less represented than its closest homolog PLAU. Its presence could suggest an involvement in responses to SARS-CoV-2, possibly pertaining to fibrinolysis, the decomposition of blood clots—an important process in severe viral infections, leading sometimes to coagulopathies. PLAT was found to bind specifically to some of the viral proteins, including a Vaccinia virus protein. This is intriguing, as this binding might similarly indicate that PLAT has functions of interest to viruses, being modulated towards a favourable coagulation system for effective viruses spreading/survival, just like PLAU (Bydoun et al., 2018). Thus, as PLAT is bound by some viral proteins, such as the Vaccinia virus, it has achieved limited status as a therapeutic target due to these druglikeness concerns.

The gene ADIPOQ, also known as adiponectin, is a significant factor that differs from the previous information. It contributes to glucose level regulation and fatty acid breakdown while being linked to inflammatory responses and immune system modulation. When examining viral infections such as SARS-CoV-2, this gene can potentially impact disease progression and severity by influencing metabolic rates in conjunction with every other aspect of an individual's immunity response. By involving itself in infection-fighting mechanisms within humans undergoing inflammation or attacks on their immune systems through viruses like COVID19; it could help researchers identify additional therapies beyond traditional treatment methods for patients struggling during these times due not only because there are limited resources but finding what genetics sets certain people into extreme symptoms would be beneficial overall too all medical fields when dealing with similar issues outside infected cases specifically associated with Covid-19.

Therefore, PLAU stands out as a noteworthy gene concerning SARS-CoV-2 infection and viral manipulation tactics; however, due to drug-like properties obstacles, the potential of PLAT is restricted. Conversely, ADIPOQ presents an alternative angle on host response that could unveil other therapeutic pursuits or knowledge into viral pathogenesis and related pathological circumstances.

6.0 CONCLUSION

In summary, PLAU and ADIPOQ are very strong candidates for drug development against COVID-19 when considering overall gene expression, druglikeness, and pathway enrichment. While their expression patterns are up-regulated by infection with SARS-CoV-2, they are also critical for key functions related to immune response, inflammation, and tissue remodeling, processes more relevant to the life-threatening complications of COVID-19, including coagulopathies and excess inflammation.

Such a strategic system design approach can effectively target PLAU and PLAT. Small-molecule inhibitor/modulator design allows one to interact with either of the protein products. As a gene for tissue remodeling and fibrinolysis, intervention in PLAU can be adequately done at an early point before excessive tissue injury from inflammatory responses. Modulation of ADIPOQ, a protein involved in glucose homeostasis, fatty acid catabolism, and modulation of inflammatory processes, may be a potential target for intervention. This would mean that, by controlling inflammation and metabolic disturbance, therapeutic targeting of ADIPQ could be facilitated in severe COVID-19.

Looking ahead, the identification of PLAU and ADIPOQ as key genes for the treatment of COVID-19 opens up new opportunities in drug design. Further efforts in the preclinical and clinical evaluations of PLAU and ADIPOQ inhibitors or modulators, combination therapies, and personalized approaches in which the choice to use treatment is based on genetic backgrounds, should be pursued. To this end, the long-term effects of such gene modulation will need to be understood, particularly with respect to post-COVID-19 conditions for safety and efficacy in therapeutic strategies.

REFERENCES

Askari, B. S., & Krajinovic, M. (2010). Dihydrofolate Reductase Gene Variations in Susceptibility to Disease and Treatment Outcomes. *Current Genomics*, *11*(8), 578. https://doi.org/10.2174/138920210793360925

Baker, S. J., Poulikakos, P. I., Irie, H. Y., Parekh, S., & Reddy, E. P. (2022). CDK4: a master regulator of the cell cycle and its role in cancer. *Genes & Cancer*, *13*, 21. https://doi.org/10.18632/GENESANDCANCER.221

Bloom, N., Bunn, P., Mizen, P., Smietanka, P., & Thwaites, G. (2023). The Impact of COVID-19 on Productivity. *The Review of Economics and Statistics*, 1–45. https://doi.org/10.1162/REST_A_01298

Bogdan, C., Röllinghoff, M., & Diefenbach, A. (2000). The role of nitric oxide in innate immunity. *Immunological Reviews*, *173*, 17–26. https://doi.org/10.1034/J.1600-065X.2000.917307.X

Bydoun, M., Sterea, A., Weaver, I. C. G., Bharadwaj, A. D., & Waisman, D. M. (2018). A novel mechanism of plasminogen activation in epithelial and mesenchymal cells. *Scientific Reports 2018 8:1*, *8*(1), 1–17. https://doi.org/10.1038/s41598-018-32433-y

Chiang, S. K., Chen, S. E., & Chang, L. C. (2021). The Role of HO-1 and Its Crosstalk with Oxidative Stress in Cancer Cell Survival. *Cells*, *10*(9). https://doi.org/10.3390/CELLS10092401

Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W. C., Wang, C. Bin, & Bernardini, S. (2020). The COVID-19 pandemic. *Critical Reviews in Clinical Laboratory Sciences*, 365–388. https://doi.org/10.1080/10408363.2020.1783198

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports 2017 7:1*, *7*(1), 1–13. https://doi.org/10.1038/srep42717

Davis, A. P., Wiegers, T. C., Johnson, R. J., Sciaky, D., Wiegers, J., & Mattingly, C. J. (2023). Comparative Toxicogenomics Database (CTD): update 2023. *Nucleic Acids Research*, *51*(D1), D1257–D1262. https://doi.org/10.1093/NAR/GKAC833

Fang, L., Che, Y., Zhang, C., Huang, J., Lei, Y., Lu, Z., Sun, N., & He, J. (2021). PLAU directs conversion of fibroblasts to inflammatory cancer-associated fibroblasts, promoting esophageal squamous cell carcinoma progression via uPAR/Akt/NF-κB/IL8 pathway. *Cell Death Discovery 2021 7:1*, *7*(1), 1–14. https://doi.org/10.1038/s41420-021-00410-6

Kanugula, A. K., Kaur, J., Batra, J., Ankireddypalli, A. R., & Velagapudi, R. (2023). Renin-Angiotensin System: Updated Understanding and Role in Physiological and Pathophysiological States. *Cureus*, *15*(6). https://doi.org/10.7759/CUREUS.40725

Lagunin, A., Ivanov, S., Rudik, A., Filimonov, D., & Poroikov, V. (2013). DIGEP-Pred: web service for in silico prediction of drug-induced gene expression profiles based on structural formula. *Bioinformatics (Oxford, England)*, *29*(16), 2062–2063. https://doi.org/10.1093/BIOINFORMATICS/BTT322

Lee, H., Aylward, A. J., Pearse, R. V., Lish, A. M., Hsieh, Y. C., Augur, Z. M., Benoit, C. R., Chou, V., Knupp, A., Pan, C., Goberdhan, S., Duong, D. M., Seyfried, N. T., Bennett, D. A., Taga, M. F., Huynh, K., Arnold, M., Meikle, P. J., De Jager, P. L., … Young-Pearse, T. L. (2023). Cell-type-specific regulation of APOE and CLU levels in human neurons by the Alzheimer's disease risk gene SORL1. *Cell Reports*, *42*(8), 112994. https://doi.org/10.1016/j.celrep.2023.112994

Lim, T. S., Goh, J. K. H., Mortellaro, A., Lim, C. T., Hämmerling, G. J., & Ricciardi-Castagnoli, P. (2012). CD80 and CD86 differentially regulate mechanical interactions of T-cells with antigen-presenting dendritic cells and B-cells. *PloS One*, *7*(9). https://doi.org/10.1371/JOURNAL.PONE.0045185

Martin, H. J., Melo-Filho, C. C., Korn, D., Eastman, R. T., Rai, G., Simeonov, A., Zakharov, A. V., Muratov, E., & Tropsha, A. (2023). Small molecule antiviral compound collection (SMACC): A comprehensive, highly curated database to support the discovery of broad-spectrum antiviral drug molecules. *Antiviral Research*, *217*, 105620. https://doi.org/10.1016/J.ANTIVIRAL.2023.105620

Nandi, A., Yan, L. J., Jana, C. K., & Das, N. (2019). Role of Catalase in Oxidative Stress- and Age-Associated Degenerative Diseases. *Oxidative Medicine and Cellular Longevity*, *2019*(1), 9613090. https://doi.org/10.1155/2019/9613090

Pozzi, S., & Satchi-Fainaro, R. (2024). The role of CCL2/CCR2 axis in cancer and inflammation: The next frontier in nanomedicine. *Advanced Drug Delivery Reviews*, *209*, 115318. https://doi.org/10.1016/J.ADDR.2024.115318

Preedy, M. K., White, M. R. H., & Tergaonkar, V. (2024). Cellular heterogeneity in TNF/TNFR1 signalling: live cell imaging of cell fate decisions in single cells. *Cell Death & Disease 2024 15:3*, *15*(3), 1–12. https://doi.org/10.1038/s41419-024-06559-z

Quatrini, L., & Ugolini, S. (2020). New insights into the cell- and tissue-specificity of glucocorticoid actions. *Cellular & Molecular Immunology 2020 18:2*, *18*(2), 269–278. https://doi.org/10.1038/s41423-020-00526-2

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/GR.1239303

Tan, Y., Wang, X., Song, H., Zhang, Y., Zhang, R., Li, S., Jin, W., Chen, S., Fang, H., Chen, Z., & Wang, K. (2021). A PML/RARα direct target atlas redefines transcriptional deregulation in acute promyelocytic leukemia. *Blood*, *137*(11), 1503–1516. https://doi.org/10.1182/BLOOD.2020005698

The basics of blood clots: What you need to know | NIH MedlinePlus Magazine. (n.d.). Retrieved August 12, 2024, from https://magazine.medlineplus.gov/article/the-basics-of-blood-clots-what-you-need-to-know/

Vijayakumar, A., Yakar, S., & LeRoith, D. (2011). The Intricate Role of Growth Hormone in Metabolism. *Frontiers in Endocrinology*, *2*(SEP). https://doi.org/10.3389/FENDO.2011.00032

Wang, Q., Ren, D., Bi, Y., Yuan, R., Li, D., Wang, J., Wang, R., Zhang, L., He, G., & Liu, B. (2020). Association and functional study between ADIPOQ and metabolic syndrome in

elderly Chinese Han population. *Aging (Albany NY)*, *12*(24), 25819. https://doi.org/10.18632/AGING.104203

Weekly epidemiological update on COVID-19 - 10 August 2023. (n.d.). Retrieved July 31, 2024, from https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19 10-august-2023

Zhao, Z., Bo, Z., Gong, W., & Guo, Y. (2020). Inhibitor of Differentiation 1 (Id1) in Cancer and Cancer Therapy. *International Journal of Medical Sciences*, *17*(8), 995–1005. https://doi.org/10.7150/IJMS.42805

Zhou, T., Luo, P., Wang, L., Yang, S., Qin, S., Wei, Z., & Liu, J. (2020). CTNNB1 Knockdown Inhibits Cell Proliferation and Aldosterone Secretion Through Inhibiting Wnt/β-Catenin Signaling in H295R Cells. *Technology in Cancer Research & Treatment*, *19*. https://doi.org/10.1177/15330338209796

TABLES

Table 1: The targets of each phytoconstituent identified using DIGEP-Pred for proteins at the pharmacological activity (Pa) > 0.5.

SMILES	Molecule ChEMBL ID	Downregulated Genes	Upregulated Genes
Oc1ccc(cc1C(=O)Nc1cccc(Cl)c1)N=Nc1ccc(cc1Cl)N(=O)=O	CHEMBL4437334		RARA
CCCCCOCc1c(Nc2ccc(cc2)S(N)(=O)=O)nc(nc1C(=O)NCCOc1ccccc1)N1CCOCC1	CHEMBL4454990		REN
			ATG5
Cc1ccc(cc1)C(=O)c1cc(O)c(O)c(c1)N(=O)=O	CHEMBL1324	MMP7	PLAT
		CAT	ADIPOQ
		SELL	ABCA1
		FLT1	CD86
		NOS2	CCL2
			CYP1A1
			RARA
			NFE2L2
			CD83
CC(=O)NC1C(NC(N)=N)C=C(OC1C(O)C(O)CO)C(=O)=O	CHEMBL222813		GH1
			NPPB
			CD83
CC(=C)C1CCC2(CCC3(C)C(CCC4C5(C)CCC(O)C(C)(C)C5CCC34C)C12)C(O)=O	CHEMBL269277	IVL	CD83
			VDR
COc1cccc(COc2ccc(cc2)C(NC(=O)C(CCCCN)NC(=O)C(CCCNC(N)=N)NC(=O)c2ccc(s2)-c2cccs2)C(N)=O)c1	CHEMBL3741713		
Oc1ccc2cc(Oc3ccccc3C(=O)Nc3nc4cc(O)c(O)cc4s3)ccc2c1	CHEMBL4440832		CYP1A2
			PLAT
			RARA
Oc1cc2nc(NC(=O)C3CCCN3S(=O)(=O)c3ccc(cc3)N(=O)=O)sc2cc1O	CHEMBL4462325		
Oc1cc2nc(NC(=O)C3CCCN3S(=O)(=O)c3ccc(cc3)N(=O)=O)sc2cc1O	CHEMBL4583315		
CN1CCN(CC1)c1ccc(NC(=O)c2c3CN(Cc4ccccc4)CCc3nc3ccccc23)c(C)c1	CHEMBL1370977		PROC
Cc1ccc(cc1)C1=NC(=O)C(S1)=Cc1cc(O)c(O)c(Br)c1	CHEMBL1458891		ADIPOQ
CC(=O)OC1CC(C)(C)C(=C=CC(C)=CC=CC=CC=C(C)C=C2OC(=O)C(C=CC34OC3(C)CC(O)CC4(C)C)=C2)C(C)(O)C1	CHEMBL1980535		RARA
			KRT18
			RAC1
			ID1
			NPPB
			VDR
			CD14
Nc1ncnc2n(ccc12)C1OC(CO)C(O)C1O	CHEMBL267099	CHEK1	NPPB
		NR3C1	CLU
			AR
NCCCCC(NC(=O)C(Cc1ccc(cc1)C(N)=N)NC(=O)CCc1ccccc1)C(=O)NC(C(N)=O)c1ccccc1	CHEMBL3628278		
NCCCCC(NC(=O)C(CCCNC(N)=N)NC(=O)c1ccccc1)C(=O)NC(C(N)=O)c1ccc(OCc2ccccc2)cc1	CHEMBL3741422		
CCC(CC)OCC1CC(=CC(N)C1NC(C)=O)C(O)=O	CHEMBL3818159	CHEK1	CD83
		TOP2A	CD86
			SMN2

			RAC1
			ADIPOQ
			NPPB
CC(=O)NC1C(NC(N)=N)C=C(OC1C(O)C(O)CNC(=O)C1CC1)C(O)=O	CHEMBL3818654		NPPB
CCCCCCCCCCCCCCCC(=O)NCCCCC(NC(=O)CCNC(=O)C(CCCCN)NC(=O)C(CCCCN)NC(=O)C(CCC(N)=O)NC(=O)C(NC(=O)C(NC(=O)C(CCC(O)=O)NC(=O)C(CCC(O)=O)NC(=O)C(NC(=O)C(CCCCN)NC(=O)C(CCCCN)NC(=O)C(CCC(N)=O)NC(=O)C(NC(=O)C(CCC(O)=O)NC(=O)C(CCC(O)=O)NC(=O)C(NC(=O)C(CCCCN)NC(=O)C(CCCCN)NC(=O)C(CCC(N)=O)NC(=O)C(NC(=O)C(CCC(O)=O)NC(=O)C(CCC(O)=O)NC(=O)C(NC(=O)C(CCCCN)NC(=O)C(CCCCN)NC(=O)C(CCC(N)=O)NC(=O)C(NC(=O)C(CCC(O)=O)NC(=O)C(CCC(O)=O)NC(=O)C(NC(C)=O)C(C)CC)C(C)CC)C(C)CC)C(C)CC)C(C)CC)C(C)CC)C(C)CC)C(C)CC)C(C)CC)C(N)=O	CHEMBL4206744		
OCc1c(C[N-][N+]#N)cnc(C(=O)NO)c1O	CHEMBL4439416	CHEK1	CLU
		LEP	CDK4
		ADIPOQ	PRDX6
		NR3C1	
COc1cccc(Cn2c3ccc(cc3c3cc(ccc23)-c2cccc(c2)C(N)=N)-c2cccc(c2)C(N)=N)c1	CHEMBL4446364		CBR1
COc1cccc(Cn2c3ccc(cc3c3cc(ccc23)-c2cccc(c2)C(N)=N)-c2cccc(c2)C#N)c1	CHEMBL4447165		CBR1
Cc1ccc2nc(sc2c1S(O)(=O)=O)-c1ccc2nc(sc2c1)-c1ccc(NC(=O)c2ccc3nc(sc3c2)-c2ccc(N)cc2)cc1	CHEMBL4447800		RARA
Cn1c2ccc(cc2c2cc(ccc12)-c1cccc(c1)C(N)=N)-c1cccc(c1)C(N)=N	CHEMBL4448497	TOP2A	CBR1
			PROC
			GH1
CCOC(=O)c1cnn(c1N(S(=O)(=O)c1ccc(cc1)N(=O)=O)S(=O)(=O)c1ccc(cc1)N(=O)=O)-c1ccccc1	CHEMBL4449109	CAT	ABCA1
NC(=N)c1cccc(c1)-c1ccc2n(c3ccc(Br)cc3c2c1)S(=O)(=O)c1cccc(c1)C(N)=N	CHEMBL4457232	KEAP1	FNDC4
		DHFR	WIPI1
		SIVA1	SERPINA3
		SREBF2	HPN
			INHBE
Oc1cc2nc(NC(=O)c3ccccc3Sc3ccc(cc3)C(F)(F)F)sc2cc1O	CHEMBL4522006		
OC(=O)C(Cc1ccc(O)cc1)NC(=O)C1=C2Sc3ccc(OC4CCCCC4)cc3N2C(=O)C(=C1)c1ccccc1	CHEMBL4526128		TNNI3
COc1cccc(Cn2c3ccc(Br)cc3c3cc(ccc23)-c2cccc(c2)C(N)=N)c1	CHEMBL4532866		CBR1
NCCCCC(NC(=O)C(CCCNC(N)=N)NC(=O)c1ccccc1)C(=O)NC(C(=O)NC1CN(C1=O)c1ccc(F)cc1)c1ccc(OCc2ccccc2)cc1	CHEMBL4576745		
OCc1c(CS(=O)(=O)c2ccccc2)cnc(C(=O)NO)c1O	CHEMBL4587069	NR3C1	CLU
			AR
NCCCCC(NC(=O)Cc1ccccc1)C(=O)NC(CCCCN)C(=O)NC(CCCNC(N)=N)C=O	CHEMBL522355		GH1
OC1C(COP(O)(=O)OP(O)(=O)OP(O)(O)=O)OC(C1O)N1C=NC2NC(=O)NNC(=O)C12	CHEMBL610669		VDR
NC1=NC(=N)NC(=N)C2N=CN(C3OC(COP(O)(=O)OP(O)(=O)OP(O)(O)=O)C(O)C3O)C12	CHEMBL611003	NR3C1	NPPB
			VDR
CC1OC(OC2=C(Oc3cc(O)cc(O)c3C2=O)c2ccc(O)c(O)c2)C(O)C(O)C1O	CHEMBL82242	MMP7	TIMP1
		CHEK1	PLAU
		NOS2	PLAT
		NFE2L2	TNFRSF1A
		CTNNB1	PRDX4

			HMOX1
			SMN2
			TP73
			KRT1
			NPPB
			CBR1
			EPAS1
			AR
			NFE2L2

Table 2: KEGG Pathway Identification for Differential Gene Expression with Enrichr

Term	P-value	Adjusted P-value	Genes
Upregulated			
Lipid and atherosclerosis	2.08E-07	2.51E-05	ABCA1
			CYP1A1
			CCL2
			RAC1
			CD14
			TNFRSF1A
			NFE2L2
Fluid shear stress and atherosclerosis	3.20E-07	2.51E-05	CCL2
			HMOX1
			PLAT
			RAC1
			TNFRSF1A
			NFE2L2
Complement and coagulation cascades	2.47E-05	0.001293	PLAU
			PROC
			PLAT
			CLU
Transcriptional misregulation in cancer	3.88E-05	0.001521	CD86
			PLAU
			RARA
			PLAT
			CD14
Pathways in cancer	7.79E-05	0.002446	AR
			EPAS1
			CDK4
			RARA
			HMOX1
			RAC1
			NFE2L2
Chemical carcinogenesis	1.09E-04	0.002807	CBR1
			AR
			VDR
			CYP1A2
			CYP1A1
Shigellosis	1.25E-04	0.002807	WIPI1
			RAC1
			CD14
			ATG5
			TNFRSF1A
Metabolism of xenobiotics by cytochrome P450	4.71E-04	0.00924	CBR1
			CYP1A2
			CYP1A1
Kaposi sarcoma-associated herpesvirus infection	5.85E-04	0.010212	CD86
			CDK4
			RAC1
			TNFRSF1A

Prostate cancer	9.59E-04	0.013159	AR
			PLAU
			PLAT
Downregulated			
Fluid shear stress and atherosclerosis	2.08E-04	0.013613987	KEAP1
			NFE2L2
			CTNNB1
Hepatocellular carcinoma	3.63E-04	1.36E-02	KEAP1
			NFE2L2
			CTNNB1
Pathways in cancer	8.88E-04	2.22E-02	NOS2
			KEAP1
			NFE2L2
			CTNNB1
Adipocytokine signaling pathway	1.54E-03	0.025871965	LEP
			ADIPOQ
p53 signaling pathway	1.72E-03	0.025871965	CHEK1
			SIVA1
Peroxisome	2.17E-03	0.027123205	NOS2
			CAT
Longevity regulating pathway	3.33E-03	0.035571645	CAT
			ADIPOQ
HIF-1 signaling pathway	3.79E-03	0.035571645	NOS2
			FLT1
AMPK signaling pathway	4.58E-03	0.038145847	LEP
			ADIPOQ
Pathways of neurodegeneration	7.06E-03	0.05126798	CTNNB1
			NOS2
			CAT

Table 3: GO Biological Process Identification for Differential Gene Expression with Enrichr

Name	P-value	Adjusted P-value	Genes
Upregulated			
Negative Regulation Of Blood Coagulation (GO:0030195)	3.62E-07	0.0002856	PROC
			PLAU
			KRT18
			PLAT
Fibrinolysis (GO:0042730)	0.00000161	0.0006352	PLAU
			KRT18
			PLAT
Negative Regulation Of Defense Response (GO:0031348)	0.000002833	0.000745	PROC
			FNDC4
			ADIPOQ
			TNFRSF1A
			ATG5
Regulation Of Plasminogen Activation (GO:0010755)	0.000004076	0.0008039	PLAT
			PLAU
			HPN
Response To Organic Cyclic Compound (GO:0014070)	0.000005382	0.0008493	GH1
			CLU
			CCL2
			CYP1A1
Negative Regulation Of Response To External Stimulus (GO:0032102)	0.00001037	0.001363	PROC
			FNDC4
			ADIPOQ
			TNFRSF1A
			ATG5
Negative Regulation Of Plasminogen Activation (GO:0010757)	0.0000582	0.00656	PLAT
			PLAU

Negative Regulation Of Inflammatory Response (GO:0050728)	0.0000704	0.006577	PROC
			FNDC4
			ADIPOQ
			TNFRSF1A
Regulation Of T-helper 2 Cell Differentiation (GO:0045628)	0.00008138	0.006577	CD86
			RARA
Negative Regulation Of Multicellular Organismal Process (GO:0051241)	0.00009314	0.006577	ADIPOQ
			PROC
			HPN
			ID1
			TIMP1
Downregulated			
Cellular Response To Laminar Fluid Shear Stress (GO:0071499)	0.00001018	0.002428	SREBF2
			NFE2L2
Regulation Of Cellular Response To Oxidative Stress (GO:1900407)	0.00002439	0.002428	DHFR
			NFE2L2
Regulation Of Leukocyte Mediated Cytotoxicity (GO:0001910)	0.00002439	0.002428	NOS2
			LEP
Response To Laminar Fluid Shear Stress (GO:0034616)	0.00002439	0.002428	SREBF2
			NFE2L2
Prostaglandin Transport (GO:0015732)	0.00003048	0.002428	NOS2
			LEP
Signal Release (GO:0023061)	0.00003048	0.002428	NOS2
			LEP
Cellular Response To Fluid Shear Stress (GO:0071498)	0.00003723	0.002542	SREBF2
			NFE2L2
Icosanoid Secretion (GO:0032309)	0.00004466	0.002668	NOS2
			LEP
Positive Regulation Of Transferase Activity (GO:0051347)	0.0002261	0.01162	FLT1
			ADIPOQ
			CTNNB1
Regulation Of Nitric-Oxide Synthase Activity (GO:0050999)	0.0002537	0.01162	DHFR
			LEP

Table 4: GO Cellular Component Identification for Differential Gene Expression with Enrichr

Name	P-value	Adjusted P-value	Genes
Upregulated			
Collagen-Containing Extracellular Matrix (GO:0062023)	0.000008138	0.0006348	GH1
			SERPINA3
			PLAT
			ADIPOQ
			KRT1
			CLU
			INHBE
Serine Protease Inhibitor Complex (GO:0097180)	0.00003885	0.001515	PLAT
			PLAU
Platelet Alpha Granule Lumen (GO:0031093)	0.0003108	0.007631	SERPINA3
			CLU
			TIMP1
Secretory Granule Lumen (GO:0034774)	0.0003988	0.007631	SERPINA3
			CLU
			TIMP1
			PRDX6
			PRDX4
Ficolin-1-Rich Granule (GO:0101002)	0.0004892	0.007631	PRDX4
			KRT1
			CLU
			RAC1
Platelet Alpha Granule (GO:0031091)	0.0007471	0.009712	SERPINA3

			CLU
			TIMP1
Secretory Granule Membrane (GO:0030667)	0.002282	0.02543	PLAU
			RAC1
			CLU
			CD14
Keratin Filament (GO:0045095)	0.002758	0.02689	KRT1
			KRT18
Tertiary Granule (GO:0070820)	0.004137	0.03586	PLAU
			RAC1
			CLU
Endoplasmic Reticulum Membrane (GO:0005789)	0.004856	0.03788	ABCA1
			CYP1A2
			CYP1A1
			HPN
			HMOX1
			RAC1
Downregulated			
Peroxisomal Matrix (GO:0005782)	0.0007811	0.01601	CAT
			NOS2
Microbody Lumen (GO:0031907)	0.0007811	0.01601	CAT
			NOS2
Nuclear Chromosome (GO:0000228)	0.002898	0.0347	TOP2A
			CHEK1
Focal Adhesion (GO:0005925)	0.003996	0.0347	FLT1
			CAT
			CTNNB1
Cell-Substrate Junction (GO:0030055)	0.004231	0.0347	FLT1
			CAT
			CTNNB1
Peroxisome (GO:0005777)	0.005269	0.03601	CAT
			NOS2
beta-catenin-TCF Complex (GO:1990907)	0.01016	0.05948	CTNNB1
Catenin Complex (GO:0016342)	0.02354	0.1073	CTNNB1
Condensed Nuclear Chromosome (GO:0000794)	0.02354	0.1073	CHEK1
ul3-RING Ubiquitin Ligase Complex (GO:0031463)	0.02935	0.1203	KEAP1

Table 5: GO Molecular Function Identification for Differential Gene Expression with Enrichr

Name	P-value	Adjusted P-value	Genes
Upregulated			
Serine-Type Endopeptidase Activity (GO:0004252)	0.0001116	0.009764	PROC
			PLAU
			HPN
			PLAT
Serine-Type Peptidase Activity (GO:0008236)	0.0001825	0.009764	PROC
			PLAU
			HPN
			PLAT
Receptor Ligand Activity (GO:0048018)	0.0004164	0.01043	CD86
			NPPB
			GH1
			TIMP1
			INHBE
Hormone Activity (GO:0005179)	0.0004354	0.01043	NPPB
			GH1
			ADIPOQ
Endopeptidase Activity (GO:0004175)	0.0005204	0.01043	PROC
			PLAU
			HPN
			REN

			PLAT
Hormone Receptor Binding (GO:0051427)	0.0005847	0.01043	NPPB
			GH1
Heme Binding (GO:0020037)	0.0006992	0.01069	CYP1A2
			HMOX1
			NFE2L2
Steroid Hydroxylase Activity (GO:0008395)	0.0009608	0.01285	CYP1A1
			CYP1A2
Oxidoreductase Activity, Acting On Paired Donors, With Incorporation Or Reduction Of Molecular Oxygen, Reduced Flavin Or Flavoprotein As One Donor, And Incorporation Of One Atom Of Oxygen (GO:0016712)	0.001136	0.01351	CYP1A2
			HMOX1
Transcription Coactivator Binding (GO:0001223)	0.002101	0.02248	RARA
			EPAS1
Downregulated			
Adenyl Nucleotide Binding (GO:0030554)	0.000002259	0.0001514	DHFR
			CAT
			NOS2
Heme Binding (GO:0020037)	0.00005173	0.001733	CAT
			NOS2
			NFE2L2
NADP Binding (GO:0050661)	0.000398	0.008888	DHFR
			NOS2
RNA Polymerase II-specific DNA-binding Transcription Factor Binding (GO:0061629)	0.0008835	0.0148	KEAP1
			CTNNB1
			NFE2L2
DNA-binding Transcription Factor Binding (GO:0140297)	0.001629	0.01921	KEAP1
			CTNNB1
			NFE2L2
Hormone Activity (GO:0005179)	0.001772	0.01921	LEP
			ADIPOQ
Protein Homodimerization Activity (GO:0042803)	0.002007	0.01921	CAT
			NOS2
			TOP2A
			ADIPOQ
Arginine Binding (GO:0034618)	0.005936	0.04605	NOS2
FMN Binding (GO:0010181)	0.008469	0.04605	NOS2
Estrogen Response Element Binding (GO:0034056)	0.008469	0.04605	NR3C1

Table 6: Top 10 CytoHubba-Selected Gene Interaction with COVID-19 Gene Set via Enrichr

Name	P-value	Adjusted P-value	Genes
Top 500 upregulated genes for SARS-CoV-2 in human colon organoid from GSE148696	0.000001323	0.000217	PLAU
			CCL2
			CTNNB1
			HMOX1
			PLAT
SARS Perturbation 357 Up Genes from GEN3VA Mouse Lung; Accession: GSE68820 Platform: GPL7202 Entry 4	0.0005675	0.01455	PLAU
			CCL2
			TIMP1
Top 500 up genes for SARS-CoV-2 infection in Mesocricetus auratus hamster lung Day 5 from GSE162208	0.0007226	0.01455	HMOX1
			CCL2
			TIMP1
Top 500 up genes from control vs. Ad5-hACE2 for SARS-CoV-2 infection in mouse lung from GSE158069	0.0008033	0.01455	PLAU
			CCL2
			HMOX1
500 genes up-regulated by MHV-A59 in murine liver cells from GSE146074 5d	0.0008214	0.01455	HMOX1
			CCL2
			TIMP1

500 top upregulated genes from SARS-CoV-2 infection at 24 HPI from GSE157852	0.001051	0.01455	PLAU CCL2 HMOX1
500 genes down-regulated by SARS-CoV-2 in mouse Lung cells at 7 dpi from GSE162113	0.001073	0.01455	ADIPOQ LEP CTNNB1
Top 500 downregulated genes in mouse lung with SARS-CoV-2 infection (Day 7) from GEO GSE162113	0.001073	0.01455	ADIPOQ LEP CTNNB1
Top 500 upregulated genes in human nasal epithelial cells with SARS-CoV-2 infection (WT, 8 hpi) from GEO GSE162131	0.001222	0.01455	PLAT REN TIMP1
Top 500 upregulated genes in human nasal epithelial cells with SARS-CoV-2 infection (Mutant, 8 hpi) from GEO GSE162131	0.001238	0.01455	PLAT REN TIMP1

Table 7: Enrichr-Based Virus-Host Interactions from COVID-19 Gene Set Analysis

Name	P-value	Adjusted P-value	Genes
Human cosavirus D 3C	0.000005272	0.0097	PLAU CTNNB1 PLAT
Vaccinia virus Ankara Haemagglutinin	0.00004606	0.02083	PLAU CCL2 CTNNB1 PLAT
Astrovirus VA1 non-structural protein 1a	0.00006709	0.02083	PLAU CTNNB1
Astrovirus SG nonstructural protein 1a	0.00007266	0.02083	PLAU CTNNB1
Horsepox virus HSPV177 (Immunoglobulin like)	0.00009557	0.02083	PLAU CCL2 CTNNB1 PLAT
Astrovirus SG nonstructural protein 1ab	0.00009715	0.02083	PLAU CTNNB1
Chikungunya virus excised_polyprotein 1..748	0.00009715	0.02083	CTNNB1 PLAT
Astrovirus VA1 non-structural protein 1ab	0.0001038	0.02083	CTNNB1 PLAT
Ectromelia virus ERPV chemokine binding protein	0.0001038	0.02083	CCL2 CTNNB1
Chikungunya virus full_polyprotein 1..1248	0.0001178	0.02083	CTNNB1 PLAT

Table 8: Top 10 COVID-19 genes along with their corresponding SMILES, ChEMBL IDs and drug-likeness assessments.

Genes	SMILES	CHEMBL ID	Druglikeness
PLAU	CC1OC(OC2=C(Oc3cc(O)cc(O)c3C2=O)c2ccc(O)c(O)c2)C(O)C(O)C1O	CHEMBL82242	
	CC1OC(OC2=C(Oc3cc(O)cc(O)c3C2=O)c2ccc(O)c(O)c2)C(O)C(O)C1O	CHEMBL82242	
	Cc1ccc(cc1)C(=O)c1cc(O)c(O)c(c1)N(=O)=O	CHEMBL1324	Yes
PLAT	Oc1ccc2cc(Oc3ccccc3C(=O)Nc3nc4cc(O)c(O)cc4s 3)ccc2c1	CHEMBL4440832	
TIMP1	CC1OC(OC2=C(Oc3cc(O)cc(O)c3C2=O)c2ccc(O)c(O)c2)C(O)C(O)C1O	CHEMBL82242	
CTNNB1	CC1OC(OC2=C(Oc3cc(O)cc(O)c3C2=O)c2ccc(O)c(O)c2)C(O)C(O)C1O	CHEMBL82242	
	Cc1ccc(cc1)C(=O)c1cc(O)c(O)c(c1)N(=O)=O	CHEMBL1324	
NFE2L2	CC1OC(OC2=C(Oc3cc(O)cc(O)c3C2=O)c2ccc(O)c(O)c2)C(O)C(O)C1O	CHEMBL82242	
CCL2	Cc1ccc(cc1)C(=O)c1cc(O)c(O)c(c1)N(=O)=O	CHEMBL1324	Yes
REN	CCCCCOCc1c(Nc2ccc(cc2)S(N)(=O)=O)nc(nc1C(=O)NCCOc1ccccc1)N1CCOCC1	CHEMBL4454990	
LEP	OCc1c(C[N-][N+]#N)cnc(C(=O)NO)c1O	CHEMBL4439416	
	Cc1ccc(cc1)C(=O)c1cc(O)c(O)c(c1)N(=O)=O	CHEMBL1324	
	Cc1ccc(cc1)C1=NC(=O)C(S1)=Cc1cc(O)c(O)c(Br)c1	CHEMBL1458891	Yes
	CCC(CC)OCC1CC(=CC(N)C1NC(C)=O)C(O)=O	CHEMBL3818159	Yes
ADIPOQ	OCc1c(C[N-][N+]#N)cnc(C(=O)NO)c1O	CHEMBL4439416	
HMOX1	CC1OC(OC2=C(Oc3cc(O)cc(O)c3C2=O)c2ccc(O)c(O)c2)C(O)C(O)C1O	CHEMBL82242	

FIGURES

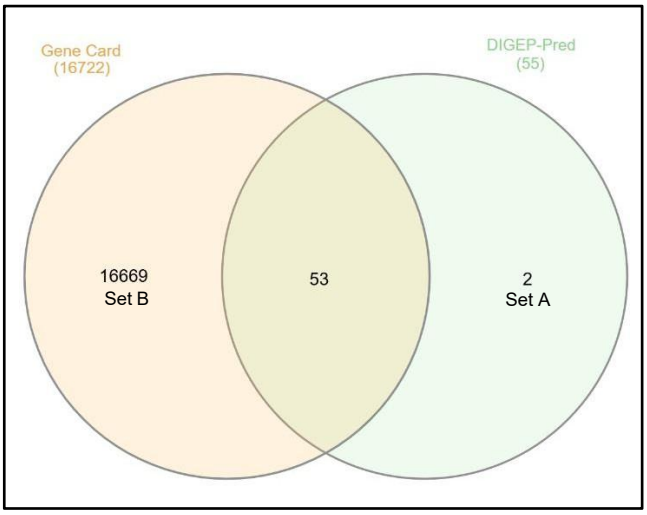

Figure 1: The overlap of upregulated and downregulated genes with Covid-19 genes.

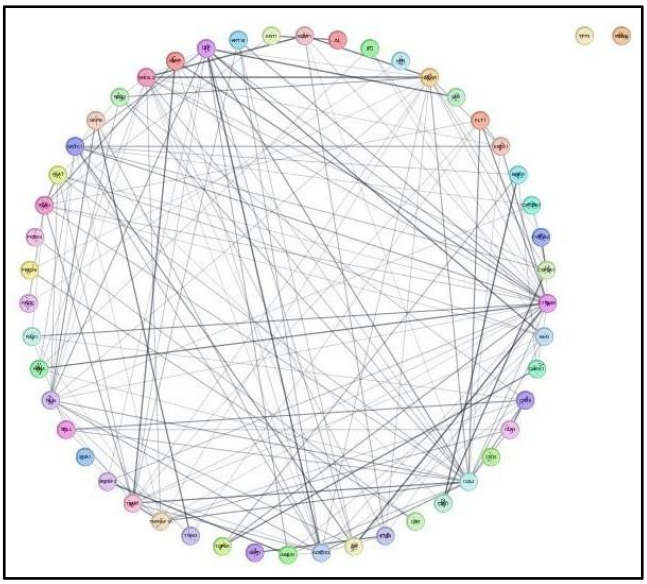

Figure 2: Protein-protein interaction network overview for upregulated and downregulated genes built using STRING in Cytoscape.

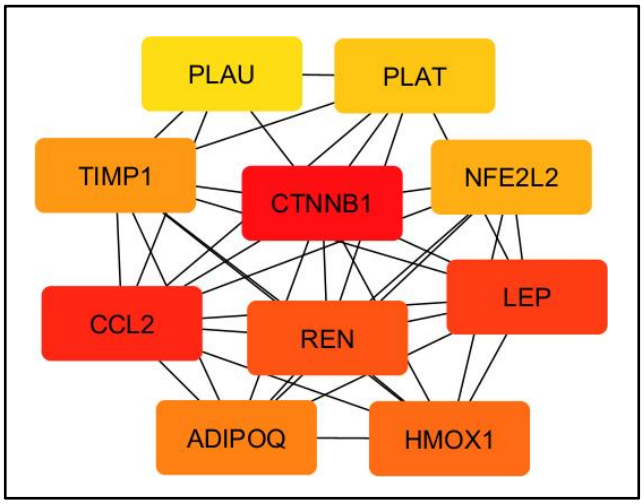

Figure 3: Ranking of the top 10 nodes based on MCC with Cytohubba plugin.

MANAGEMENT AND SCIENCE UNIVERSITY
FINAL PROJECT -RUBRICS FOR WRITTEN REPORT

NAME:_____ **ID:** _____

Criteria	5	4	3	2	1	Marks
Clarity and Organization (____) X 10m 5	• Purely own words OR REPHRASED from the referred sources. • Constructed the words independently for the entire assignment. • Good usage of language with no grammatical error. 10	• Only 20% of assignment is copy paste from website. • 80% of own words or from books, and notes. 8	• 50% of assignment is copy paste from website. • 50% of own words or from books, and notes. 6	• 70% of assignment is copy paste directly from website. • 30% of own words. 4	• **Entire** assignment is copy paste directly from website with evidence e.g. underline, italic fonts, color, page, etc. not removed. 2	
Formatting & writing skills (____) X 20m 5	• No error- subject name, code, lecturer's name, title all included. • Content format given by lecturer e.g. Header footer, name, id number, subject, code. • Page number included. • Tidy and organized. • Do not exit 10 pages. • Include table of content. 20	• Tidy and clean. • Appropriate format but not following the one given by lecturer. • Exit 10 pages. 16	• Tidy. • Spelling mistake on lecturer's name, subject, code, title. • Information incomplete, no id , etc. • Can still be improved. • No table of contents. 12	• Irregular size of fonts. • 70% of the assignment is untidy. 8	• Not using any format. • Untidy, unorganized. 4	
Creativity and Innovation in Solutions (____) X 40m 5	• Complete contents exactly required by lecturer. • With extra information from own research. • Included new discoveries on the topic given. 40	• Complete contents exactly required by lecturer. • NO extra information. 32	• 20% of the required contents not included. 24	• 50% of the contents not discussed. • Some information is not answering the title of assignment. 16	• Most of the content not elaborated/discussed. • Some statements are not true. 8	
Critical Analysis & Problem Solving (____) X 20m 5	• Insightful critical analysis of bioinformatics data • identifying key issues and assumptions. • proposes highly innovative and feasible solutions 20	• Shows good analysis with minor gaps and errors. • offers creative solutions with good feasibility and minor issues • Can still be improved. 16	• Provides basic analysis with some significant gaps • somewhat creative solutions with significant feasibility issues. 12	• Shows limited critical analysis. • proposes unoriginal or impractical solutions. 8	• Fails to critically analyze bioinformatics data and propose creative or feasible solutions. • Not attractive. 4	
References (____) X 10m 5	• Complete references from website, book, journal, notes according to APA style 10	• 4 out of 5 references are according to APA style 8	• 3 out of 5 references are according to APA style 6	• 1 out of 5 references are according to APA style not included. 4	• All the references are not according to APA style • '0' **MARKS** given if reference is not included. 2	
Assessor1: Assessor2:			---------- X 20% = 100		**TOTAL MARKS**	/100

Date: