UNIVERSIDADE D
COIMBRA

Nádia Neves Pereira

# SARS-Cov-2 Characterization – An in Silico Approach

**Dissertação no âmbito do Mestrado em Bioquímica, orientada pela Professora Doutora Irina de Sousa Moreira e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologias da Universidade de Coimbra.**

Outubro de 2021

# SARS-COV-2 CHARACTERIZATION – AN IN SILICO APPROACH

Nádia Neves Pereira

Dissertação no âmbito do Mestrado em Bioquímica, orientada pela Professora Doutora Irina de Sousa Moreira e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologias da Universidade de Coimbra.

Outubro de 2021

# Agradecimentos

Este trabalho não teria sido possível sem a ajuda e o apoio de inúmeras pessoas que, de uma forma ou de outra, acabaram por me marcar e dar outro significado ao ‘ano de tese’.

Primeiro, gostaria de agradecer à minha orientadora, Dr. Irina Moreira, por me ter aceite no grupo e me ter exposto ao mundo da bioinformática. Gostaria de agradecer também a alguns elementos do grupo, à Nícia por ter ajudado na orientação deste e de outros projetos em que estive envolvida, e ao Manuel e à Catarina, por me apoiarem sempre, com toda a paciência do mundo. Sem vocês este trabalho não teria sido possível.

À minha família, por acreditarem sempre em mim mesmo quando eu duvidei e por me apoiarem incondicionalmente. Um agradecimento especial aos meus pais, pelas horas infinitas ao telefone, pela motivação, por nunca desistirem. Ao Tio Mário e Avô Carlos por se preocuparem comigo mesmo durante a vossa recuperação, quando tinham problemas maiores que a minha tese. Ao meu irmão adotivo, por nos alegrar todos os dias e pelas dores de cabeça bem-vindas.

Aos que permaneceram apesar da distância. À Carolina e ao Luís, pelos jantares e pelas viagens para estarem comigo, por me terem surpreendido pela positiva. À Maggie, pelas chamadas aleatórias, pelas histórias e pelo riso constante. Ao Miguel, por me aturar sempre de madrugada. Aos meus afilhados, João e Filipa, por ainda precisarem de mim. À Patrícia, pelos cafés e trocas de histórias.

Aos que Coimbra me deu. A todos os da ´Casa do Chá’: Alice e Bolacha, por serem as pessoas mais amorosas que tive o prazer de conhecer e me acolherem sempre; Ana Rita, por todo o conhecimento partilhado, pelas horas de desabafo e todo o apoio; Tampas, pela paciência e pela porrada amorosa. À MJ por ser o meu símbolo pessoal de superação, por nunca julgar e estar sempre aqui. À Ângela, por me mostrar que nunca é tarde para seguir o que queremos. Por fim, ao meu novo colega de casa, Guilherme, por não nos termos matado um ao outro, por me apoiar nas horas mais inimagináveis, não me deixar desistir e acreditar mais em mim do que eu própria.

# Resumo

O SARS-CoV-2 (Síndrome Respiratório Agudo Grave Coronavírus-2) já infetou mais de 225 milhões de pessoas e foi responsável por mais de 4,64 milhões de mortes em quase dois anos, tornando-se a última pandemia mundial. Ainda há muito que saber sobre este vírus e, tendo em conta a enorme quantidade de dados que surgiram desde a sua descoberta, pensámos numa abordagem que nos permitisse obter diferentes camadas de informação. Usámos *text mining* para obter informações de 179.984 artigos e obtivémos 10.325 genes humanos. Em seguida, usando o *clusterprofiler*, foi possível realizar uma análise de enriquecimento com as databases GO (*Gene Ontology*), KEGG (*Kyoto Encyclopedia of Genes and Genomes*) e MeSH (*Medical Subject Headings*). Os resultados de diferentes databases corresponderam, o que significa que vários termos enriquecidos estavam presentes nas diferentes análises.

**Keywords:** SARS-CoV-2, Análise de Enriquecimento de Genes, GO, KEEG, MeSH

# Abstract

SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2) infected over 225 million people and was responsible for over 4.64 million deaths in almost two years, becoming the last worldwide pandemic. There is still a lot to know about this virus and, considering the huge amount of data that appeared since the virus was discovered, we needed un approach to obtain different layers of information. We used text mining techniques to gather information from 179.984 articles and we were able to retrieve 10.325 human genes. Then, we performed enrichment analysis with GO (Gene Ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes) and MeSH (Medical Subject Headings) databases. The results from the different databases matched, meaning that various enriched terms were present in the different analysis. This research may be continued and taken to a deeper level.

**Keywords:** SARS-CoV-2, Gene Enrichment Analysis, GO, KEEG, MeSH

# Abbreviations

**nm** – nanometer

**RNA** - Ribonucleic Acid

**DNA** - Deoxyribonucleic Acid

**ICTV** – International Committee on Taxonomy of Viruses

**SARS** – Severe Acute Respiratory Syndrome

**SARS-CoV** – Severe Acute Respiratory Syndrome-related Coronavirus

**MuCoV**- Murine Coronavirus

**MHV** – Mouse Hepatitis Virus

**kb** – Kilobyte

**ORFs** – Open Reading Frames

**nsp** – Non-structural Proteins

**HE**- Hemagglutinin Esterase

**S Protein** – Spike Protein

**E Protein** – Envelope Protein

**M Protein** – Membrane Protein

**N Protein** – Nucleocapsid Protein

**TRS** – Transcriptional Regulatory Sequences

**MERS** – Middle East Respiratory Syndrome

**pp** - Polyproteins

**TMPRSS2** - Transmembrane protease, serine 2

**AC2** - Angiotensin-Converting enzyme 2

**COPD** – Chronic Obstructive Pulmonary Disease

**DAD** – Diffuse Alveolar Damage

**OP** – Organizing Pneumonia

**ILD** – Interstitial Lung Disease

**CLD** – Chronic Liver Disease

**CKD** – Chronic Kidney Disease

**NLP** – Natural Language Processing

**COVID-19** – Coronavirus Disease

**TM** – Text Mining

**IR** – Information Retrieval

**NER** – Named Entity Recognition

**NEN** – Named Entity Normalization

**RE** – Relation Extraction

**ML** – Machine Learning

**NIH** – National Institutes of Health

**NLM** – National Library of Medicine

**EA** – Enrichment Analyses

**PPI** – Protein-Protein Interaction

# List of Figures and Tables

## Figures

# Tables

# Index

# I. Introduction

## 1.1. Viruses

### 1.1.1. General Definition

Viruses are known as the smallest infectious agents, with a diameter inferior to 300 nm, and they are able to infect several organisms, such as bacteria, plants, fungi and animals, including humans (Strauss & Strauss, 2008). These infectious agents are parasites on a genetic level, which implies that they need to infect and invade a host cell in order to replicate (Strauss & Strauss, 2008).

Viruses can contain either a RNA (ribonucleic acid) or DNA (desoxyribonucleic acid) genome, however they can have double or single stranded nucleic acids and may contain the enzyme reverse transcriptase (Gelderblom, 2011). Viral genome has the information required for virus replication and it is enfolded in a nucleocapsid protein. However, some viruses possess a lipid envelope for genome protection when the virus is outside the cell. This envelope also facilitates the genome entry in a susceptible host cell. When the encapsulated virus is in an extracellular environment, it is called a virion (Kumar, 2012; Strauss & Strauss, 2008).

### 1.1.2. Viral Classification

Viruses have their own universal classification system categorized in hierarchical levels from Order to Genus and Species, as stipulated by the International Committee on Taxonomy of Viruses (ICTV) (Fauquet, 1999). However, viruses can also be grouped according to different classifications regarding particular characteristics (Kuhn, 2021). Respiratory viruses, in specific, can be labeled both based on their transmission route and their pathogenicity (Leung, 2021).

### 1.1.3. Respiratory Viruses

Respiratory viruses are transmitted from one host to another through the respiratory route. Hosts can be infected via direct contact, when the virus has host-to-host transmission, or indirect contact, when the virus is transmitted through an intermediated object named fomite (Boncristiani et al., 2009; Leung, 2021). Such viral infections that spread through air can happen via droplets (bigger globs of mucus, saliva, and water which fall faster and evaporate slower) or aerosols (similar to droplets, but smaller in size, increasing the necessary time to

fall and decreasing the evaporation time) (Leung, 2021). In both cases, virions shed from the respiratory system and as an infected host sneezes, coughs or talks, viral transmission efficiency increases, since it boosts the number of virions shed (Leung, 2021; Weston & Frieman, 2018). These airborne virions can proceed directly to infect an uninfected host or can contaminate a fomite through contact (Weston & Frieman, 2018). This type of transmission is usually very efficient for human transmission and usually leads to higher transmission pathways (Leung, 2021).

## 1.2. Coronaviruses (SARS-CoV-2)

### 1.2.1. General Coronaviruses Information

Coronaviruses are a family of viruses that belong to the Nidovirus superfamily (Weiss & Leibowitz, 2011). The first virus of this family was discovered in 1930s but the interest in coronaviruses only increased after a first human pandemic caused by Severe Acute Respiratory Syndrome (SARS) (Belouzard et al., 2012). Coronaviruses are 125nm spherical viruses with a shape that resembles a solar corona, hence their name. This characteristic is related to the S Protein (Spike Protein) projected on coronavirus' surface. This family is divided into four genera: Alphacoronaviruses, Betacoronaviruses, Deltacoronaviruses and Gammacoronaviruses (*Taxonomy*, n.d.). Viruses from both alphacoronaviruses and Betacoronaviruses genera are able to infect humans. Initially, Coronavirus were split into three groups, according to their antigenic reactivity, but later were regrouped according to their genome sequence, phylogenetic relationships and genomic structures (Cui et al., 2018; Weiss & Leibowitz, 2011).

Firstly, coronaviruses were mainly of veterinary interest, since they infect a wide variety of mammals and birds (Belouzard et al., 2012). They can cause mild to severe diseases, such as respiratory and enteric diseases, and more rarely, hepatitis and neurologic diseases (Belouzard et al., 2012). For a long time, the most studied Betacoronavirus was the Murine Coronavirus (MuCoV), also known as Mouse Hepatitis Virus (MHV) (Belouzard et al., 2012; Weiss & Leibowitz, 2011). This virus provided a model system that allowed the scientist to study some central nervous system diseases, such as acute hepatitis, encephalitis and multiple sclerosis in murines and mouses (Weiss & Leibowitz, 2011). However, coronaviruses with humans as possible hosts are primarily respiratory pathogens and it is believed that are responsible for 35% of upper respiratory infections in the peak of their viral activity (McIntosh & Perlman, 2015). Most coronavirus that infect humans cause the common cold, being responsible for 15% of these colds and four of these strains responsible for the common cold

circulate globally: HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU1 (Alan Sariol, 2020). Nonetheless, coronaviruses can also cause intestinal infections in humans (Cui et al., 2018).

### 1.2.2. Virology

#### 1.2.2.1. Viral genome and proteins

Coronavirus family contains large (30-32 kb) single-stranded positive-sense RNA viruses, having the largest known RNA viral genomes (Weiss & Leibowitz, 2011) (Alan Sariol, 2020). These genomes are translated into a Polyprotein (Pp) containing proteins involved in RNA replication (Alan Sariol, 2020). Betacoronaviruses' genomes are arranged similarly: the replicase lotus is encoded in the 5' end while the structural proteins are encoded in the 3' end of the genome (Weiss & Leibowitz, 2011). These genomes have between six and ten Open Reading Frames (ORFs) (Belouzard et al., 2012; Weiss & Leibowitz, 2011). Two-thirds of the genome is contained in the first ORF and encodes replicase proteins while structural protein genes are contained in the last third (Belouzard et al., 2012). These replicase proteins are cleaved into 16 nsp (non-structural proteins) and accessory proteins that may have enzymatic properties, such as proteases, RNA modification enzymes, polymerases and helicases (Belouzard et al., 2012; Weiss & Leibowitz, 2011). Structural protein genes are arranged in the following order: HE protein (Hemagglutinin Esterase protein)(HE only exists in some Betacoronaviruses), S Protein, E Protein (Envelope Protein), M Protein (Membrane Protein) and N Protein (Nucleocapsid Protein) (Alan Sariol, 2020; Belouzard et al., 2012; Weiss & Leibowitz, 2011). These proteins' function and some characteristics are briefly described in Table 1 and more detailed bellow. Accessory genes, that are believed to be involved in immune evasion, are encoded in a variable number of ORFs present between structural protein encoding genes (Alan Sariol, 2020; Belouzard et al., 2012; Rimanshee Arya , Mukesh Kumar, 2021).

**Table 1.** Coronaviruses structural proteins, their functions, and characteristics.

| Protein | Function | Characteristics |
|---|---|---|
| Membrane | Gives the virion its shape | Dimer, N-terminal ectodomain and C-terminal endodomain |
| Spike | Facilitates the receptor attachment | Trimer, two functional domains: S1 and S2 |
| Envelope | Helps in the virus assembly and release | Monomer and pentamer, N-terminal ectodomain and C-terminal endodomain |
| Nucleocapsid | Gives beads on a string structure | Monomer, can be divided into 5 domains |
| Hemagglutinin Esterase | Binding to the sialic acids in virion surface | Dimer, consisting of two monomers, each monomer is made of three domains |

HE protein, only found in some Betacoronaviruses, is responsible for binding to the sialic acids that are presented on the glycoproteins in virion surface (Cornelissen et al., 1997; Rabaan et al., 2020). This binding and the esterase activity facilitates viral entry in the host cell and HE protein also improves virus spread through the mucosa (Cornelissen et al., 1997; Klausegger et al., 1999).

S protein is responsible for coronaviruses crown-like structure (Beniac et al., 2006; Cornelissen et al., 1997; Rabaan et al., 2020). This structural protein is a trimeric class I fusion protein, highly N-glycosylated, that helps in the Endoplasmatic Reticulum access and host receptor attachment (Collins et al., 1982; Rabaan et al., 2020). Normally, this protein is cleaved by a protease into two distinct domains: S1 domain, that helps binding with the receptor and S2 domain, that is responsible for S protein stalk's structural support (de Groot et al., 1987; Rabaan et al., 2020).

E protein, barely found in the virion, is a transmembrane protein that has an N-terminal ectodomain and a C-terminal endodomain (Rabaan et al., 2020). This protein is vital for both viral assembly and release. Besides this, E protein has an ion channel activity essential to some coronaviruses pathogenesis, such as SARS-CoV and SARS-CoV-2, two Betacoronaviruses (Rabaan et al., 2020).

M protein, the most abundant protein found in the virion, is responsible for viral shape. It is found in the virion as a dimer and has two main functions: membrane curvature maintenance and facilitating N protein binding (Rabaan et al., 2020). It also has three transmembrane domains, an N-terminal ectodomain and a C-terminal endodomain (Rabaan et al., 2020)

Interactions between M protein and N protein and nsp3 (a non-structural protein, component of replicase complex) improves the encapsulated viral genome packing into the viral particles, since it facilitates the binding to the replicase-transcriptase complex (Rabaan et al., 2020). The N protein is part of the nucleocapsid. Its domains can bind directly with the RNA, since the high phosphorylation of this protein increases the affinity for the viral RNA.

N protein is also responsible for giving beads on a string structure. This protein has two substrates: the genomic packaging signal, where the C-terminal domain of this protein binds to, and the TRSs. (Rabaan et al., 2020)

#### 1.2.2.2. Variants of Concern

Two Betacoronaviruses were responsible for two major outbreaks in the last 20 years: the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) (2002) and the Middle East Respiratory Syndrome (MERS) (2012). In 2019, a novel respiratory virus, from the coronavirus family, SARS-CoV-2 (**Figure 1**), started to spread across the world provoking the current situation of the global pandemic, creating a new disease, COVID-19 (Coronavirus Disease). SARS-CoV-2 is 79% identical to SARS-CoV and 50% with MERS-CoV (R. Lu et al., 2020).

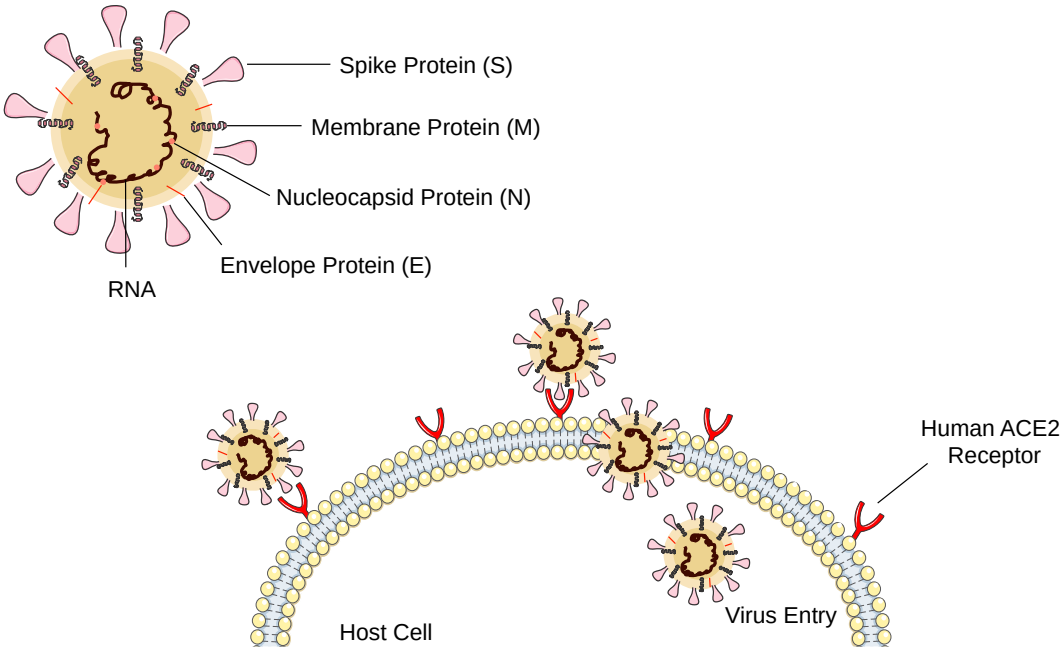


**Fig. 1** - Schematic structure of SARS-CoV-2. SARS-CoV-2' structure is formed by the following structural proteins: spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins. The S, M, and E proteins are in the viral envelope and the N protein interacts with the viral RNA in the core of the virion.

Even though the animal origin of this novel virus is still incomplete and not well understood, it is known that RaTG13 virus, a bat coronavirus, shares the biggest similarity with SARS-CoV-2, with 96.2% (B. Hu et al., 2020; Paraskevis, 2020). This high genetic similarity indicates that SARS-CoV-2 is derived from a bat virus. In addition to RaTG13, scientists also found that RmTG13, another bat coronavirus, and a pangolin coronaviruses group are similar to SARS-CoV and SARS-CoV-2 (Paraskevis, 2020). This denotes that there are other coronaviruses similar to SARS-CoV-2circulating in wildlife and that the reservoir host of SARS-CoV-2 is not clear and could have been transmitted through an animal as an intermediate host to humans (B. Hu et al., 2020). SARS-CoV-2 is also responsible for infecting other domesticated and laboratory animals, such as cats or ferrets (B. Hu et al., 2020).

#### 1.2.2.3. Life Cycle

Although there is a high similarity between SARS-CoV and SARS-CoV-2, SARS-CoV-2 spreads quicker compared to SARS-CoV. This can be explained by the structural differences in S proteins between them (Calabrese et al., 2020; Malik, 2020).

The mechanism of viral entry into host cell (**Figure 2**) using S protein in different Betacoronaviruses is similar. Both viruses use the ACE2 receptor to bind to host cells surface [17,18]. After successful attachment to the host cell, SARS-CoV enters the cell cytosol using proteases such as cathepsin and TM-PRRS2. These proteases perform S protein cleavage which is followed by viral and host cell membrane fusion. This event occurs mainly on endosomes. The presence of this cleavage site is similar to virus that infect furin and facilitates the initial process, increasing the spreading efficiency of SARS-CoV-2 compared to other ORF1a and ORF1b into pps 1a and pp1b, respectively, that are subsequently cleaved into 16 nsps (Trougakos et al., 2021; V'kovski et al., 2021). Viral replicase synthesizes a negative antigenome from the viral RNA genome that serves as template for new viral RNAs genomes. It is known that several structural proteins are subjected to post translational modifications in order to regulate protein function (Trougakos et al., 2021; V'kovski et al., 2021). The newly replicated virus leaves the cell by vesicle-mediated exocytosis.

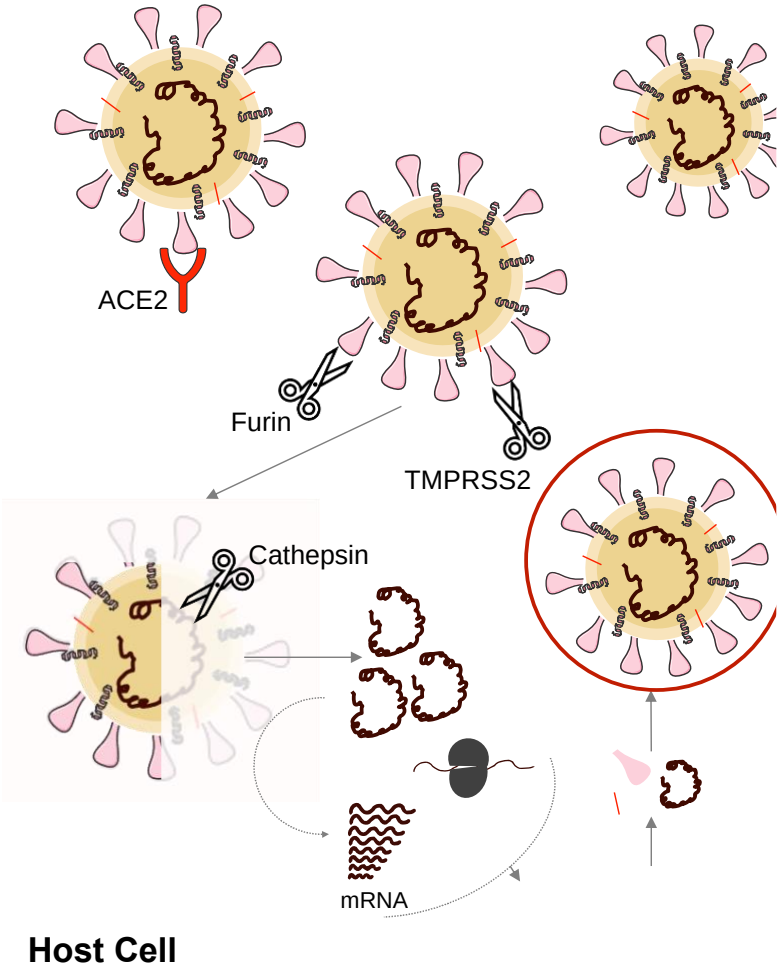


**Fig. 2** - Illustration of SARS-CoV-2 life cycle. First SARS-CoV-2 (extracellular) binds to ACE2, followed by TMPRSS2 or FURIN priming. Then, the virus suffers a clathrin-mediated endocytosis and denotes endosomal compartments during exocytosis. The next steps involve uncoating, genomic RNA release and viral-protein synthesis in free and endoplasmic reticulum-attached ribosomes. Finally, a vesicle-mediated exocytosis and the new replicated virus leaves the cell.

### 1.2.3. Epidemiology

#### 1.2.3.1. Geographic distribution and overall numbers

Recent data reports that on September 15th, 2021, SARS-CoV-2 infected over 225 million people and was responsible for over 4,64 million deaths (*WHO Coronavirus (COVID-19) Dashboard*, n.d.). According to WHO (World Health Organization), the most affected regions are America, with over 86 million confirmed cases, Europe, with over 67 million confirmed cases, and South East Asia, with over 42 million confirmed cases (*WHO Coronavirus (COVID-*

*19) Dashboard*, n.d.). However, according to the same organization, the country that has been more affected was the United States of America, with 41 million cases, followed by India, with more than 33 million cases (*covid19*, n.d.). As of 13 September 2021, a total of 5.534.977.637 vaccine doses have been administered (*WHO Coronavirus (COVID-19) Dashboard*, n.d.).

#### 1.2.3.2. Transmission

SARS-CoV-2, as a respiratory virus, is transmitted through mucus or saliva expelled from the host respiratory tract. The mucus or saliva can be expelled in three different ways: small-droplet aerosol, large-droplet aerosol, and fomites. Aerosols are more effective in the transmission of a respiratory infection at the peak of virus replication (Fa, 2017). Small-droplet aerosols are able to spread through a bigger distance and can quickly create an outbreak (Fa, 2017). Large-droplet aerosols sink faster and, consequently, this type of transmission requires a closer contact (Fa, 2017). Lastly, the transmission through fomites occurs when an object is contaminated with aerosol droplets or respiratory secretions (Fa, 2017). In experimental conditions, scientists discovered that droplets containing SARS-COV-2 could last 3 hours in the air (van Doremalen, Bushmaker, et al., 2020).

Even though respiratory transmission is the most frequent mode of transmission, vertical transmission may also occur, since transplacental transmission (transmission to the newborn via placenta) has already been documented (Courtemanche et al., 2020; Dong et al., 2020; M. Hu et al., 2021; Patanè et al., 2020; Vivanti et al., 2020; Z. Yang & Liu, 2020). SARS-CoV-2 is not transmitted through bloodborne, sexual or fecal-oral routes but several studies have found live viruses from isolations of saliva, semen and blood donations (Chang et al., 2020; J. Gu et al., 2020; J.-M. Kim et al., 2020; D. Li et al., 2020; Parasa et al., 2020; Qiu et al., 2020).

Some animals, like cats, ferrets and minks can be infected with SARS-CoV-2 but only minks can transmit the virus to each other and to humans (Halfmann et al., 2020; Richard et al., 2020; Shi et al., 2020).

Most people showed signs of diseases after an incubation period of 1–14 days (most commonly around 5 days), and developed dyspnea and pneumonia within a median time of 8 days from illness onset (B. Hu et al., 2020; Wu & McGoogan, 2020).

#### 1.2.3.3. Symptoms/Diseases

Humans infected with SARS-CoV-2 exhibit several symptoms, some more frequent like fever, dry cough, and some rarer ones such as fatigue, diarrhea, hemoptysis, sputum production, sore throat, anorexia, chest pain, vomiting, nausea and chills (Eastin & Eastin, 2020; B. Hu et al., 2020).

Besides this, SARS-CoV-2 virus is responsible for several neurological conditions and symptoms. Some patients present disorders of consciousness, delirium, olfactory taste disorders, headache and neuromuscular and cerebrovascular complications (B. Hu et al., 2020; Mukerji SS, 2021). Besides these symptoms, there are other neurological issues provoked by SARS-CoV-2 such as myalgia, rhabdomyolysis, Guillain-Barre syndrome, encephalopathy, and myelopathy and rare occasions of encephalitis (Mukerji SS, 2021).

Most histopathological complications occur in the lungs. Histopathology analysis showed bilateral diffuse alveolar damage, hyaline membrane formation, desquamation of pneumocytes and fibrin deposits in lungs of patients with severe COVID-19 (C. Huang et al., 2020). Exudative inflammation was also shown in some cases. Immunohistochemistry assays detected SARS-CoV-2 antigen in the upper airway, bronchiolar epithelium and submucosal gland epithelium, as well as in type I and type II pneumocytes, alveolar macrophages and hyaline membranes in the lungs (B. Hu et al., 2020; Mehta et al., 2020; Yao et al., 2020). These symptoms can be more severe in people with some risk factors like asthma, COPD, and allergies. It is related to Diffuse Alveolar Damage (DAD), Organizing Pneumonia (OP), reactive type II pneumocytes, and chronic interstitial pneumonia (Calabrese et al., 2020).

#### 1.2.3.4. Risk Factors

These clinical manifestations can be more or less severe according to the patient's age (X. Lu et al., 2020). Overall, children and younger people only develop mild diseases, non-pneumonia or mild pneumonia, or are asymptomatic (X. Lu et al., 2020). Older people, mainly men above 60 years old, with co-morbidities are more prone to severe respiratory diseases that require hospitalization and may even cause fatalities (B. Hu et al., 2020). Other health conditions such as diabetes, cardiovascular disease, or a suppressed immune system can also increase SARS-CoV-2 fatality rate (*Coronavirus 2019*, n.d.) .

Other health conditions such as diabetes, cardiovascular disease, suppressed immune system, pregnancy, arterial hypertension, obesity, allergies, asthma, Chronic Obstructive Pulmonary Disease (COPD), Interstitial Lung Disease (ILD), Chronic Liver Disease (CLD),

Chronic Kidney Diseases (CKD), cancer and chemotherapy can also increase SARS-CoV-2 fatality rate. (*Coronavirus 2019*, n.d.) (Esposito et al., 2020; Gao et al., 2021; Ng et al., 2020; Singh & Khan, 2020; Tian et al., 2020) (Du et al., 2021; S. Huang et al., 2020; Ou et al., 2020)

### 1.2.4. Vaccines

There are three types of SARS-CoV-2 vaccines available and approved worldwide ("COVID-19 vaccines," 2020). The first vaccine type approved was the messenger RNA (mRNA) vaccines, produced by Pfizer-BioNTech and Moderna in the United States of America (USA) and CureVac in Europe ("COVID-19 vaccines," 2020). The second type of vaccines are made with human and primate adenovirus vectors, produced by Janssen-Johnson & Johnson, Astra-Zeneca, Sputnik-V, and CanSino ("COVID-19 vaccines," 2020). Lastly, vaccines are also made with an inactivated whole virus of SARS-CoV-2, this vaccine type is not available in the USA and it's produced by Bharat Biotech, Sinopharm and Sinovac ("COVID-19 vaccines," 2020).

The BNT162b2 vaccine, developed and produced by Pfizer-BioNTech was tested in individuals of 16 years or older (Frenck et al., 2021). BNT162b2 is a lipid nanoparticle-formulated, nucleoside-modified RNA vaccine encoding a prefusion-stabilized, membrane-anchored SARS-CoV-2 full-length S protein (Karikó et al., 2008; S. J. Thomas et al., 2021; Wrapp et al., 2020).

Of the 21,720 participants with two doses 30-µg doses of BNT162b2, 8 of them were infected with Covid-19, showing an effectiveness of 95% (Frenck et al., 2021).

The vaccine produced by Moderna, mRNA-1273 vaccine, is lipid nanoparticle–encapsulated mRNA-based and encodes the prefusion stabilized full-length SARS-CoV-2 S protein (Baden, 2021; Jackson et al., 2020) . This vaccine was administered to 15,210 participants during the trial. After the two doses of 100 µg, only 11 participants had symptomatic SARS-CoV-2 illness, and none had severe symptoms. This shows that the Moderna Vaccine has an effectiveness of 94.1%. (Baden et al., 2021).

The CureVac vaccine, CVnCoV, is a lipid nanoparticle-encapsulated mRNA vaccine that encodes full-length, pre-fusion stabilized SARS-CoV-2 S protein and its made exclusively with naturally occurring nucleotides (Rauch et al., 2021). A study was conducted were 19 783 participants received two 12 µg doses of CVnCoV, this study revealed that this vaccine has a 48.2% effectiveness (Kremsner et al., 2021).

Janssen-Johnson & Johnson vaccine, Ad26.COV2.S COVID vaccine, has a replication-incompetent human adenovirus type 26 (Ad26) vector that expresses a pre-fusion stabilized SARS-CoV-2 S protein, was studied as a single-dose (Alter et al., 2021). This vaccine is a viral vector that express the SARS-CoV-2 S protein and was studied in almost 1.500 participants, over 18 years old, showing an efficacy of 69% (Polinski et al., 2021).

AstraZeneca–University of Oxford vaccine, AZD1222 or ChAdOx1 nCoV-19, is an adenovirus-vectored vaccine encoding the S protein of SARS-CoV-2 (van Doremalen, Lambe, et al., 2020). The vaccine was studied in a diverse adult population of more than 32,000 participants were the participants received two doses of AZD1222, 4 weeks apart. This study showed that the vaccine can prevent symptomatic illness 15 days or more after the second dose with an efficacy of 74% (Falsey et al., 2021).

## 1.3. Methodology Overview

### 1.3.1. Text-Mining as a relevant approach: advantages and disadvantages

The amount of data available from scientific papers, patents, or other sources of information, particularly in the biomedical field, is continuously rising. Often, this data is unstructured, and as such it is not ready for computational interpretation. The use of Text Mining (TM) in the biomedical field has widely increased due to the emergent need to analyze and acquire knowledge from large data sources (Fleuren & Alkema, 2015).

TM provides a set of automated methods that can distill text from heterogeneous sources into actionable data. TM applies NLP (Natural Language Programing) methods to extract and retrieve information from text just like a human reader would. A NLP model should understand the language, semantics and vocabulary to correctly predict token features (Shorten et al., 2021). BioNLP, the application of NLP models in the biomedical field, adds the required knowledge of a specific biological context (Fleuren & Alkema, 2015; Gachloo et al., 2019).

TM comes upon a first step of automated Information Retrieval (IR) to retrieve all the information relevant to a specific problem from dispersed data resources (Fleuren & Alkema, 2015; Zheng et al., 2019).

Biomedical TM, at its core, comprises three stages: Named Entity Recognition (NER), Named Entity Normalization (NEN) and Relation Extraction (RE) (W. Sun et al., 2018).

Part of TM complexity lies in the fact that different sources compile data in different formats, which often requires specific techniques (Fleuren & Alkema, 2015; Gonzalez et al., 2016). These types of data frequently lack common structural frameworks and can have errors like

improper grammar, spelling errors or semantic ambiguities (Fleuren & Alkema, 2015). Text errors increase the complexity of data pre-processing and TM analysis (Fleuren & Alkema, 2015) (Gonzalez et al., 2016) (W. Sun et al., 2018). The recognition and mapping of certain terms in the NER and NEN steps can also be troublesome (Fleuren & Alkema, 2015). In fact, biomedical NER is usually considered more challenging since there are numerous difficulties for biomedical terms automatic identification due to irregularities in how known entities are entitled (Gonzalez et al., 2016; Zhu et al., 2013). Common challenges arise when terms are not a part of the used ontology, as misspellings or ambiguity in the term's designations can occur (Fleuren & Alkema, 2015). Hence, to deal with this issue, choosing the right corpus and/or ontology is crucial (Fleuren & Alkema, 2015). This is particularly true for genes and proteins where nomenclature is frequently messier since proteins and genes can share the same abbreviation and different ontologies may have different spellings (Fleuren & Alkema, 2015; H. Li et al., 2017). However, this type of heterogeneity and ambiguity can also happen in key classes such as drugs or chemicals (H. Li et al., 2017; Zheng et al., 2019). Correctly choosing a corpus to train a TM model and then retrieve relations from text is a complicated task due to the complexity of grammatical construction hindering the machine retrieval of relations from text and, at the same time, incorporation of data from external sources can foster the advances of the RE step (Ghamami & Keyvanpour, n.d.).

Lastly, biological knowledge is complex, and the lack of certain specific information can result in conflicting answers (Saffer & Burnett, 2014). For instance, the same species under different conditions (e.g. age, gender, treatments) may not have the same biological system and what happens in one species may not happen in another (Saffer & Burnett, 2014). These differences, if not noted, may lead to different answers upon TM application (H. Li et al., 2017; Saffer & Burnett, 2014; Zheng et al., 2019).

### 1.3.2. TM Workflow

The text mining workflow is represented in **Figure 3** (adapped from Rosário-Ferreira et al., 2021a). NER, also known as 'entity tagging' or 'concept extraction', is fundamental for automatically extract information from text (Gonzalez et al., 2016). Biomedical NER aims to retrieve relevant biomedical entities such as genes, drugs, diseases, species, proteins, mutations and cell lines existing in natural language documents and tag each word´s location and class (Gonzalez et al., 2016; Zhu et al., 2013). Hence, this step identifies concepts and keywords, categorizing them in user-defined classes. NER tools implement the text

preprocessing stage, where all data is cleaned and tokenized, a step in which typically words are broken down into words or sub-words tokens to build the vocabulary of such text (Perera et al., 2020). Then, all remaining unique tokens are processed through different methods to extract features that represent biomedical classes to be transformed into a suitable representation (Perera et al., 2020).

After accomplishing NER step, NEN algorithms are invoked to semantics and coherence for all the retrieved tokens, solving words that are disambiguous. As such, constitute an essential step in the automated construction of a biomedical database describing and relating concepts, which can be organized either as a hierarchy or as a set of relationships. Abbreviation recognition and synonym recognition are advantageous to unify and normalize biomedical terms (Zhu et al., 2013). Biomedical NEN intends to map entity terms in biomedical text to typical entities in a particular knowledge base, as an example, a database which compiles information about a topical domain in a hierarchical or relationship manner (H. Li et al., 2017). Furthermore, NEN models can exhibit additional steps such as abbreviation resolution, in which acronyms are reformed to the original long words by using the abbreviation dictionary (Cho et al., 2017; H. Li et al., 2017).

Lastly, RE is a task that aims to automatically identify syntactic and semantic relations between the entities originated in the previous TM tasks (Yadav et al., 2020; Y. Zhang & Lu, 2019). Basic RE methods encompassed simple systems based on co-occurrence statistics that evolved to more intricate ones using syntactic analysis and ML/DL models (Muzaffar et al., 2015; Y. Zhang & Lu, 2019). The extracted relations are expressed in a machine understandable format ready for post-TM analysis (Xing et al., 2020). In the biomedical field, relations among entities are pivotal towards understanding complex biological mechanisms by being able to retrieve new relations from previously known ones. The extraction of homogeneous and heterogeneous interactions between chemicals, diseases, genes, proteins, and/or other classes is needed to decipher new knowledge mainly in the fields of regulatory pathways, metabolic processes or adverse drug relations (Yadav et al., 2020; Y. Zhang & Lu, 2019).

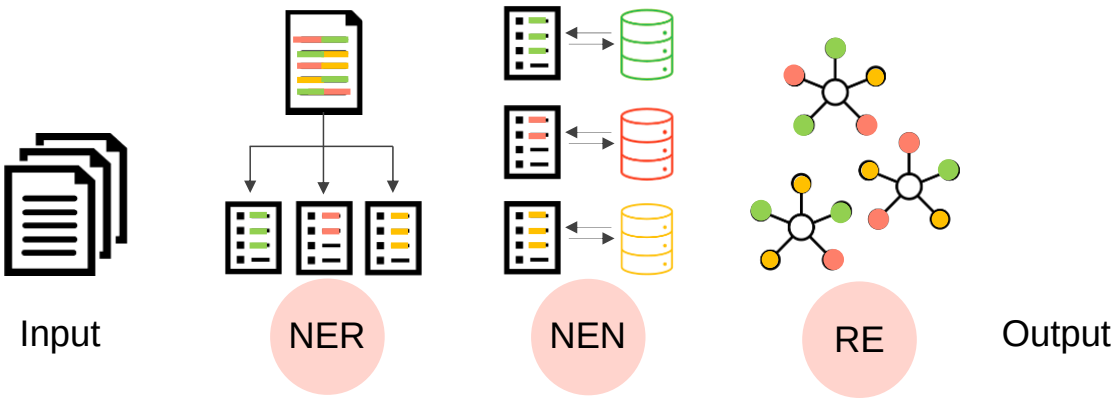


**Fig. 3** - Biomedical text mining workflow: A comprehensive text mining pipeline generally encompasses these 3 steps: named entity recognition (NER), named entity normalization (NEN) and Relation Extraction (RE). These steps can be interdependent and are hierarchical in this form. Still, each step can be dealt with separately in which case the input needed varies (adapped from Rosário-Ferreira et al., 2021a).

### 1.3.3. Biological Meaning: From genes to pathways

The conventional approach that is normally used to study function has been centered in molecular interaction networks. This type of approach have enhanced our understanding of disease, infection, drug pharmacodynamics, and evolution (Barabási et al., 2011; Jiang et al., 2015; Stuart et al., 2003; Suthram et al., 2010).

There are several approaches for this type of study of biological processes using molecular interaction networks. For example, Protein–Protein Interaction (PPI) data is regularly used to create networks, where proteins are presented interacting with functionally related partners. Similar networks are also created using co-expression data, genetic interaction data, and by combining data types (Ames et al., 2013; Costanzo et al., 2010; Stuart et al., 2003). A disadvantage of this method is that it can contain false positive and false negative interactions, which may distort our understanding of functional organization (Snider et al., 2015).

When representing these networks as pathways, the functions of individual genes or gene products are not represented. However, pathway data is considered to be more reliable than molecular data since it is based on a consensus reached by scientists over prolonged time and repeated experimentation (Stoney et al., 2018). Between the most well-known pathways are the ones involved in metabolism, regulation of genes or transmission of signals (Y. V. Sun, 2012).

To develop a general framework for genes and pathway of complex diseases and novel statistics it is needed to test the association of a gene or pathway with the disease. To accomplish this, first it requires to formulate the null hypothesis for testing gene or pathway association with the disease. (J. H. Kim, 2012) Then, it is needed to combine a set of dependent *P*-values into an overall significance level for the genes. The validation of the null distribution and calculate type 1 error rates of the three developed statistics for testing association of the gene or pathway with the disease using extensive simulation studies (J. H. Kim, 2012).

The hypothesis here is that a keyword or pattern can be regarded as the semantic interpretation of the gene group if that certain keyword is significantly over-represented or a meaningful pattern is found among the textual descriptors for a gene group (J. H. Kim, 2012). Considering that these tests are repeated thousands of times, a low p-value means a smaller possibility of false positives and more meaningful results. There are several GO (Gene Ontology) and biological pathway-based tools for gene expression analysis that have been developed and proven to be useful (J. H. Kim, 2012).

### 1.4. Aims and Scope

Over the past year and a half, information regarding SARS-CoV-2 has been constantly growing with hundreds of publications released every day. However, more than 50% of these publications are reviews, letters, notes and editorials instead of research articles (Calabrese et al., 2020; Malik, 2020; Rabaan et al., 2020; Teixeira da Silva et al., 2020). This project aims to collect, extract and organize all the data available about this topic from public databases, in particular PubMed. (*PubMed*, n.d.) We intended to gather, extract, and organize all the information from the public database to determine possible target drug interaction systems and enriched pathways. Then, using an open-source text mining tool (PubTator, via LitCovid), the unstructured information from natural text was organized into structured data with tokens and features well defined, such as genes, diseases, species, mutations, chemicals and cell lines. (Wei et al., 2013) At the end of this task, we had words organized in different classes and several relations between terms to proceed for further analysis. Functional annotation was necessary to restrict the amount of data acquired from a collection of genes through Over-Representation Analysis to obtain Enrichment Analysis (EA), which was used to separate the pathways most likely to be affected due to the list of genes in the NLP protocol, compared to what would be expected in a reference list. This task was done with the resource to G:Profiller and R package Clusterprofiler and the relevant information was compared and studied with data from public databases, OpenTargets (Ochoa et al., 2021), Reactome (A et al., n.d.),

KEGG (Kanehisa & Goto, 2000), MeTeOR (Puppala et al., 2015), CheMBL (Mendez et al., 2019), PubChem (S. Kim et al., 2021), DrugBank (Wishart et al., 2018) and STITCH (Szklarczyk et al., 2016), which allowed the construction and extrapolation of different layers of information. (Raudvere et al., 2019)

# II. Methods

## 2.1. Information Retrieval

All data regarding SARS-CoV-2 used in this project was gathered on 26th of May, 2021. The data was retrieved from the LitCovid website, a website designed by NIH (National Institutes of Health) NLM (National Library of Medicine)(*LitCovid*, n.d.). This is a curated literature hub which tracks all scientific information regarding SARS-CoV-2 available in PubMed. The information is updated daily and gives us access to 179.984 articles. These articles were not downloaded in their unstructured and natural form but with automatic annotations made previously by PubTator. (Wei et al., 2013) An R script was created using the pubtator() function to compare the results with the LitCovid, since both use PubTator' text mining, and the LitCovid data was more accurate. The data downloaded from LitCovid was in BioC format and a python script was developed to access the information. We were able to collect 21.995 different genes IDs from the annotations regarding the accessed articles.

## 2.2. Gene list

Collected genes were not exclusively from humans and, as in this project we aimed to treat only data regarding the human species, gene list needed to be filtered.

Data was analyzed with the help of *g:profiler*, public web server for characterizing and manipulating gene lists. (*g:Profiler – a web server for functional enrichment analysis and conversions of gene lists*, n.d.) This web server has four different types of analyses: *g:GOSt*, *g:Convert*, *g:Orth* and *g:SNPense.* (*g:Profiler – a web server for functional enrichment analysis and conversions of gene lists*, n.d.) In order to obtain only human genes, we used the *g:Converter*, the service from *g:profiler* that translate identifiers (IDs) of genes, proteins and other types of namespaces. (Raudvere et al., 2019) The seamless translation process works on a mixed set of diverse identifiers and maps these through Ensembl gene identifiers as reference. In cases of multiple identifiers, all relevant combinations are highlighted. At least 13 types of IDs are supported for all of the 213 species available in *g:profiler*, and at least 40 types of IDs for more than 50 species. (Raudvere et al., 2019)

The input provided was a whitespace-separated list of 21.995 genes, retrieved from the previous step, and the output was a .txt list of 10.325 human genes. These genes were with Entrez Gene unique integer identifiers.

## 2.3. Statistical Enrichment Analysis

In this particular case, we used the *g:GOSt*, which is the core of the *g:profiler*, to perform statistical enrichment analysis to provide interpretation to user-provided gene lists. This analysis provides data from multiple sources of functional evidence, including GO terms, biological pathways, regulatory motifs of transcription factors and microRNAs, human disease annotations and protein-protein interactions. (Raudvere et al., 2019)

The file with a list of human genes was used as input for *g:GOST*. As default, the *g:profiler* assumes that the list has no relevant order. (Raudvere et al., 2019) The organism selected was *Homo sapiens* (Human), the Statistical domain scope was the option 'Only annotated genes' and a 0.05 and 0.01 *g:SCS* threshold was chosen. Due to the data processing previously described, the numeric IDs were treated as ENTREZGENE_ACC.

In the section 'Gene Ontology' section, options 'GO molecular function', 'GO cellular component' and 'GO biological process' were selected. In the section 'Biological Pathways', the options 'KEGG', 'Reactome' and 'WikiPathways' were selected. In the section 'Regulatory Motifs in DNA' we selected 'TRANSFAC' and 'miRTarBase' and in the 'Protein Databases' we selected 'Human Protein Atlas' and 'CORUM'.

## 2.4. Threshold applied

*g:profiler* automatically applies multiple-testing correction to P values. *g:SCS* algorithm has a method for computing multiple testing correction for p-values gained from GO and pathway enrichment analysis (Raudvere et al., 2019). For example, an experiment-wide threshold of p-value=0.05 means that at least 95% of matches above the threshold are statistically significant. (Raudvere et al., 2019) Other standard multiple testing corrections, such as Bonferroni correction or Benjamini-Hochberg FDR were developed for multiple tests that are independent of each other. GO consists of hierarchically related general and specific terms, so these methods are not correct for the analysis performed in *g:GOSt.* The True Path Rule of GO affirm that genes associated to a given GO term are implicitly associated to all more general parents of this term. The *g:SCS* algorithm considers the set structure underlying gene sets annotated to terms of each organism. According to this, the threshold should be tighter to significant results. So, for this project, we tried a p-value=0.05 and a tighter value, p-value=0.01.

## 2.5. Enriched Analysis

We used g:profiler to obtain a file containing only human genes and to obtain an csv file with all the data from the genes enrichment analysis. The csv was successfully retrieved, and another R script was being developed to visualize the results. At the same time, these results were manually analyzed. Due to some step backs in this development, another alternative was thought.

The package clusterprofiler offered the exact type of analyses we were looking for, therefore was selected among other gene enrichment analysis packages and web servers. From the possible databases for enrichment analysis that the clusterprofiler offers we selected the ones that allowed a better and simpler analysis taking into account the previous results from the csv.

A new script was developed to obtain results from the previous steps. In this script, the input was the file with human genes and the enriched analysis was performed in three ways: GO enrichment analysis, MeSH enrichment analysis and KEGG enrichment analysis.

### 2.5.1. GO enrichment analysis

Gene Ontology (GO) uses defined concepts/classes and the relationships between these concepts in order to describe the gene function. (Ashburner et al., 2000)

GO can classify these functions along three aspects: Molecular Function (MF), molecular activities of gene products, Cellular Component (CC), where gene products are active and, Biological Process (BP), pathways and larger processes made up of the activities of multiple gene products (Ashburner et al., 2000). The R package clusterProfiler has a function, enrichGO(), for gene ontology over-representation test (Yu et al., 2012). For this function, the OrgDb chosen was org.Hs.eg.db and the keyType was ENSEMBL.

Dot Plot

The dotplot() function depicts the enrichment scores (p value), as the dot color, gene count, as the dot size, and gene ratio. This function was repeated in order to have two different graphics, one with all the ontologies (BP, CC and MF) and the other with only the BP ontology. With this plot we intend to observe which terms are the most enriched, their p-value and the number of genes involved in each term.

Gene-Concept Network

The dotplot() function only displays most significant or selected enriched terms. To study the potentially biological complexities in which a gene may belong to multiple annotation categories, we used the category netplot function, cnetplot(), to extract these complex associations. The cnetplot() depicts the linkages of genes and biological concepts as a network. With this network we intend to visualize which terms appear more enriched, which genes are associated with those terms and which genes are associated with each other and with several terms.

### 2.5.2. KEGG enrichment analysis

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database resource that's a computer representation of the biological system, consisting of molecular building blocks of genomic information and other information integrated with the knowledge on molecular wiring diagrams of interaction, reaction and relation networks (systems information). (Kanehisa & Goto, 2000)

The clusterProfiler package has a function (enrichKEGG() function) for pathway over-representation analysis. (Yu et al., 2012)

Dot Plot

The dotplot() function depicts the enrichment scores (p value), as the dot color, gene count, as the dot size, and gene ratio. With this plot we intend to observe which terms are the most enriched, their p-value and the number of genes involved in each term.

### 2.5.3. MeSH enrichment analysis

MeSH (Medical Subject Headings) is the NLM (National Library of Medicine) controlled vocabulary thesaurus used for indexing articles for PubMed. MeSH has 19 categories that contain Anatomy, organisms, diseases, drugs and chemicals, between others. (*Medical subject headings - home page*, 2020)

In this study, we used the function enrichMeSH() and selected the gendoo database (Gene Disease Features Ontology-based Overview) and the category C (the category for diseases).

Dot Plot

The dotplot() function depicts the enrichment scores (p value), as the dot color, gene count, as the dot size, and gene ratio. With this plot we intend to observe which terms are the most enriched, their p-value and the number of genes involved in each term.

# III. Results and Discussion

## 3.1. GO enrichment analysis: dot plot results and discussion

In order to be able to analyze the results from the GO enrichment analysis, a dot plot was created.

**Figure 4** illustrates the dot plot with the three GO categories, BP, CC and MF. In this figure we can analyze the 30 most relevant results for each category. We can observe the gene ratio and gene count, in the x-axis and dot size, respectively. As we can see, as the values in the x-axis get closer to zero, the size of the dots get smaller, meaning that there are less genes found involved in that function, component or process.

As we can observe in both BP and CC, for all processes and components represented, the genes found associated with them are in the same range of p-value (**Figure 4)**. According to this, all sets of genes found to be connected to that specific process or component are biologically relevant since the p-values are the smaller visually possible by the color scale, less than $10^{-9}$.

In the CC part of the dot plot, it is noticeable that the “cell-substrate junction” has a gene ratio bigger than 0.035 and around 300 genes (**Figure 4**). Closer to these results are “focal adhesion” and “neuronal cell body”, both around 0.035 gene ratio and 300 genes (**Figure 4**). The smaller, yet significant, result presented is “rough endoplasmic reticulum”.

Cell-Substrate junction term means a cell junction that forms a connection between a cell and the extracellular matrix. The focal adhesion term means a cell-substrate junction that anchors the cell to the extracellular matrix and forms a point of termination of actin filaments. Other studies, using functional enrichment analysis of downregulated genes of SARS-CoV-2 infected cells, have found that these genes were predominantly enriched in cell adhesion activities such as cell–substrate junction, focal adhesion assembly, between others (G. Li et al., 2020). It is known that some viruses, for example, oncogenic viruses (a virus that can cause cancer) can transform the host cells and trigger a cascade of cellular responses that end up leading to cytomorphological changes and differences in the cell growth characteristics (Morris et al., 2008; Volberg et al., 1991). It is possible that SARS-CoV-2 may affect the cell leading to its cytomorphological change during its entry or replication, disturbing the cell-substrate junction or its focal adhesion.

In the MF part of the dot plot, we can see that the enriched functions have a different p-values range (**Figure 4**). The terms “endopeptidase activity”, “histone deacetylase binding” and “transmembrane receptor protein kinase activity” have a higher p-value than the remain

ones, meaning that, although they are still biological meaningful, they are more likely to be a false positive or less significant (**Figure 4**). The “protein serine/threonine kinase activity” has a close to 300 genes and a gene ratio bigger that 0.03. The terms “endopeptidase activity” and “DNA-binding transcription factor binding” have similar characteristics. The smallest result presented is “chemokine binding” with less than 100 genes and a gene ratio lower than 0.005.

Some studies have found that the SARS-CoV-2 M protein induces apoptosis when it interferes with PDK1-PKB/Akt signaling, since PKB/AKT its import for the cell metabolism, growth, proliferation and survival (Hemmings & Restuccia, 2012; Ren et al., 2021; Tsoi et al., 2014). Both PDK1 and PKB/Akt are serine/threonine kinases (Ren et al., 2021; Tsoi et al., 2014). SARS-CoV-2 also induces apoptosis when interferes with other signaling pathways involving serine/threonine kinases, such as PI3K/Akt signaling pathway, Raf/MEK/ERK signaling pathway, between others (Ghasemnejad-Berenji & Pashapour, 2021; Lokhande & Devarajan, 2021).

The term “endopeptidase activity” means the catalysis of the hydrolysis of internal, alpha-peptide bonds in a polypeptide chain. Even though this term has a higher p-value ($2x10^{-9}$ , when the majority of MF represented terms has a p-value under $1x10^{-9}$), SARS-CoV-2 has endopeptidases and SARS-CoV-2 main protease ($M^{pro}$), which is also known as C30 Endopeptidase, and is one of the most potential drug targets (Bolcato et al., 2020a). There are several studies that found other endopeptidases associated with SARS-CoV-2 and many of them are potential new drug targets (Abdel-Aziz et al., 2021; Bolcato et al., 2020a, 2020b; Pišlar et al., 2020).


**Fig. 4** - GO enrichment analysis dot plot with the three GO categories represented: Biological Process (BP), Cellular Components (CC) and Molecular Function (MF). Dot represents the number of genes in each GO term; p.adjust (adjusted p-value): Yellow < Green < Blue < Purple. GO, Gene Ontology.

The BP dot plot was repeated with a different scale in order to be able to have access to more detailed data. As we can see in **Figure 5** the dots are no longer monochromatic, due to the differences in the p-value scale, that now only comprehends the lower values. The count and ratio of genes scales were also changed for higher values so as to be adequate for a better analysis.
The highest term presented is “calcium ion homeostasis” with around 350 genes and a gene ratio of almost 0.04 (**Figure 5**). The terms “cellular calcium ion homeostasis” and “response to oxidative stress” have also around 350 genes and a ratio above 0.0375. As referred above, some terms now show their true p-value range, such as “cellular response to peptide” and “regulation binding”. The lowest term given is “response do alcohol” (**Figure 5**).

The terms “calcium ion homeostasis”, “cellular calcium ion homeostasis”, “regulation of cytosolic calcium ion concentration” and “positive regulation of cytosolic calcium ion concentration” have a p-value under $1.18 \times 10^{-45}$, which is overpoweringly smaller than what the dot plot of **Figure 4** showed. This means that these terms are biologically meaningful and that there is a smallest chance of a false positive. The term “calcium ion homeostasis” means any process involved in the maintenance of an internal steady state of calcium ions within an organism or cell. Several studies have found that hypocalcemia, the reduced level of serum ionized calcium, is strongly associated with COVID-19 severity (Crespi & Alcock, 2021). Scientists speculate that this happens because SARS-CoV-2 uses calcium ions to orchestrate its entry into host cells, via a fusion peptide derived from the spike protein, similar to the entry mechanism of SARS-CoV and MERS-CoV (Millet & Whittaker, 2018; Straus et al., 2020). So they found that the use of calcium channel blockers (CCBs) can help to reduce the mortality of COVID-19 (Choksi et al., 2021; Crespi & Alcock, 2021; Danta, 2020; L.-K. Zhang et al., 2020). CCBs are used as a drug for hypertension worldwide and are one of the most-commonly prescribed drugs to reduce blood pressure (Wang et al., 2017).

The term “response to oxidative stress” means any process that results in a change in state or activity of a cell or an organism as a result of oxidative stress, a state often resulting from exposure to high levels of reactive oxygen species, or extremely low levels of oxygen. This term has the lowest p-value we can observe with this range of p-values, which means that there is an impact of the response to oxidative stress in SARS-CoV-2 infected patients (**Figure 5**). Similar to this, the term “response to decreased oxygen levels”, “response to oxygen levels”, “response to hypoxia” and “reactive oxygen species metabolic process” have the same range of p-value, so they all are biological relevant and connected (**Figure 5**). There are multiple studies regarding the impact of oxidative stress in patients infected with SARS-CoV-2 (Cekerevac et al., 2021; Chernyak et al., 2020; Fernandes et al., 2020; Forcados et al., 2021). Several new investigations have found elevated levels of oxidative stress markers

(Cecchini & Cecchini, 2020; Muhammad et al., 2021) and found the possibility that increased levels of oxidative stress in COVID-19 patients can be causing DNA oxidation and other downstream effects (Cecchini & Cecchini, 2020). This high levels of oxidative stress can be the cause of oxidation of proteins due to increased apoptosis, necrotic cell debris and pulmonary interstitial fibrosis observed during analysis of postmortem lung sections of fatal COVID-19 patients (Donia & Bokhari, 2021). Some natural antioxidants can counteract altered signaling pathways activated during COVID-19 pathogenesis (Forcados et al., 2021).

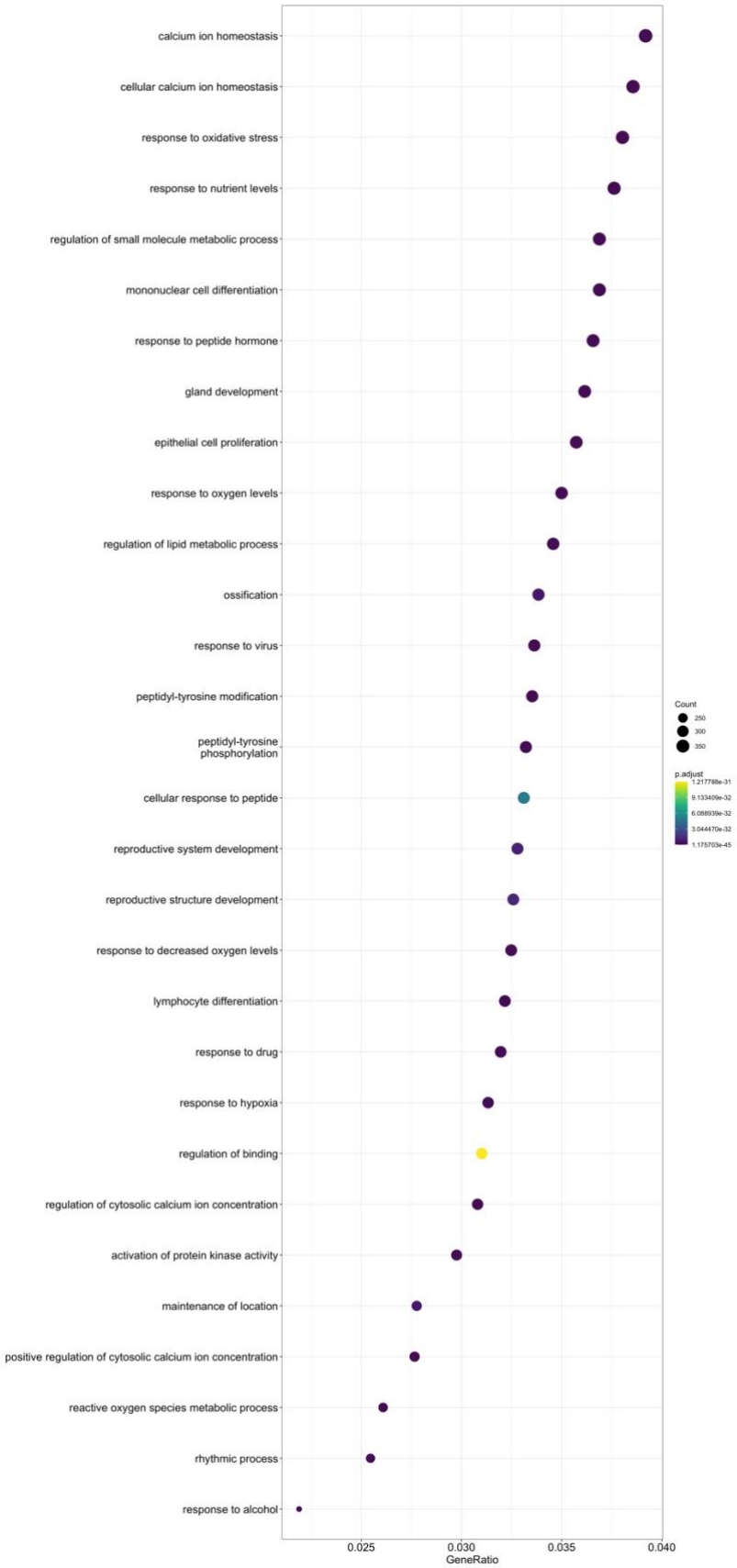


**Fig. 5** - GO enrichment analysis dot plot with the one GO category represented: Biological Process (BP). Dot represents the number of genes in each GO term; p.adjust (adjusted p-value): Yellow < Green < Blue < Purple. GO, Gene Ontology.

### 3.2. KEGG enrichment analysis results

In **Figure 6** the result from the KEGG enrichment analysis is presented. In this figure we can analyze the 30 most relevant results for KEGG pathways/terms.

We can observe the gene ratio and gene count, in the x-axis and dot size, respectively. As we can see, as the values in the x-axis get closer to zero, the size of the dots get smaller, meaning that there are less genes found involved in that pathway/term (**Figure 6**).

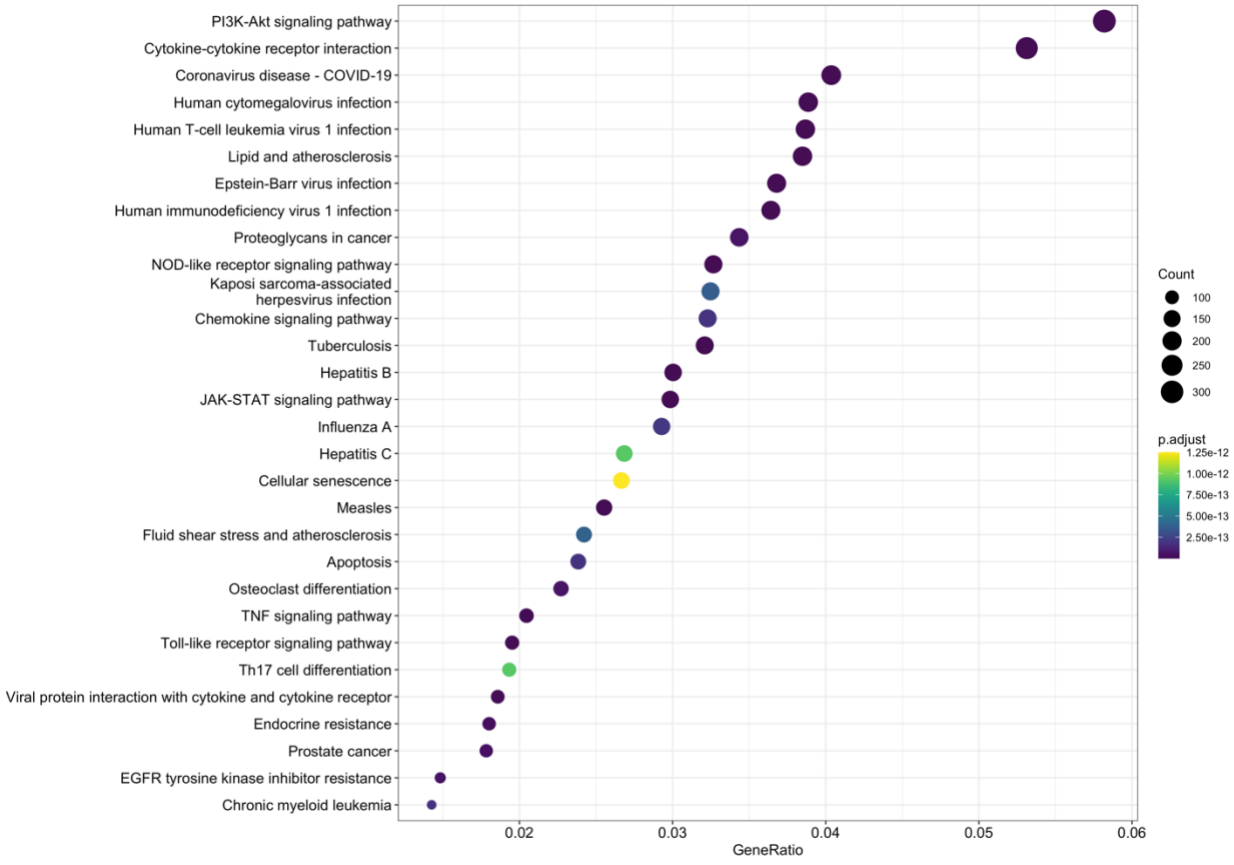


**Fig. 6** - KEGG pathway enrichment analysis. Dot size represents the number of genes in each KEGG pathway; p.adjust (adjusted P-value): Yellow < Green < Blue < Purple.; KEGG, Kyoto Encyclopedia of Genes and Genomes

### 3.3. MeSH enrichment analysis results

In **Figure 7**, the result from the MeSH enrichment analysis is presented. In this figure we can analyze the 30 most relevant results for MeSH terms in the category “Diseases”.

We can observe the gene ratio in the x-axis and gene count in the dot size. As we can see, as the values in the x-axis get closer to zero, the size of the dots get smaller, meaning that there are less genes found involved in that term (**Figure 7**).

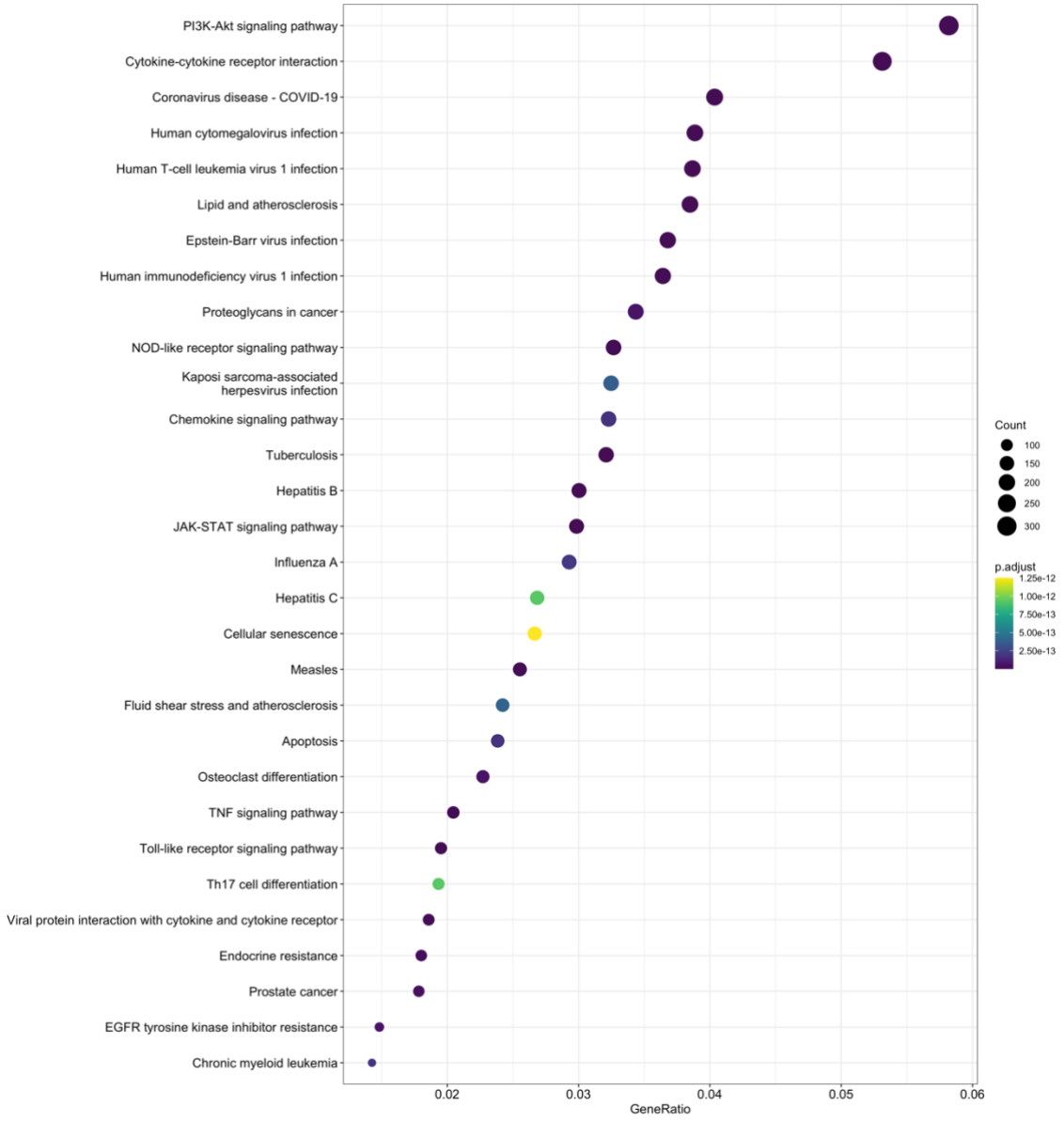


**Fig. 7** - MeSH enrichment analysis dot plot for the category “Diseases”. Dot represents the number of genes in each GO term; p.adjust (adjusted p-value): Yellow < Green < Blue < Purple. MeSH, Medical Subject Haadings

### 3.4 KEGG and MeSH enrichment analysis discussion

Both KEGG and MeSH enrichment analysis dot plots had the exact same results, the only visible difference in the plots is the heigh of the graph. Although the databases used for both plots are different, the results ended up the same (**Figure 6 and 7**). For KEGG results the only source of information was the gene list as input and the KEGG database. However, for the MeSH enrichment analysis, the input was the same gene list, but we used the MeSH "Diseases" category and the gendoo database. Gendoo is a web tool for visualizing disease feature profiles generated from the assignment of MeSH vocabulary for associated drugs, biological phenomena and anatomy to OMIM data. This approach assists in interpreting omic data for its molecular and clinical aspects. With this said, even though we used different databases, the most enriched terms/pathways are the same because the terms in the different databases overlap.

The highest term in both plots is "PI3K/Akt signaling pathway" with 300 genes and a gene ratio of almost 0.06, followed by "Cytokine-cytokine receptor interaction" with 300 genes and a gene ratio of around 0.053 (**Figure 6 and 7**). The next term is "Coronavirus disease - COVID-19" with 250 genes and a gene ratio a little higher than 0.04 (**Figure 6 and 7**). The lowest term is "chronic myeloid leukemia" with less than 100 genes and a gene ration around 0.014 (**Figure 6 and 7**). The terms "Kaposi sarcoma-associated herpesvirus infection", "Influenza A", "Fluid shear stress and atherosclerosis", "Hepatitis C", "Th17 cell differentiation" and "Cellular senescence" show a bigger p-value (**Figure 6 and 7**). The low scale values of both graphs give us certainty that all enriched terms are significant (**Figure 6 and 7**).

"PI3K/Akt signaling pathway" is an intracellular signal transduction pathway that promotes metabolism, proliferation, cell survival, growth and angiogenesis in response to extracellular signals (A, Hemmings, 2012). This term has a p-value under $2.5x10^{-13}$. This pathway is involved in various aspects of the virus entry into the cell and the development of immune responses (Khezri, 2021) There are evidences that SARS-CoV-2 binding to ACE2 and posterior endocytosis occurs through a clathrin-mediated pathway which is regulated by the PI3K/AKT signaling (Lokhande & Devarajan, 2021). Akt is a serine/threonine kinase and as we found above, serine/threonine kinases are an enriched GO term in the Molecular Function category. Other authors have identified, through bioinformatic analysis, that PI3K/Akt is the top-ranked kinase among those who were potentially associated with SARS-CoV-2 (F. Sun et al., 2021). Several other scientist have targeted PI3K/Akt as a potential drug target for COVID-19 patients (Khezri, 2021; Mizutani et al., 2005; Santamaria, 2021; Somanath, 2020).

Cytokines are soluble extracellular proteins or glycoproteins that act as intercellular regulators and mobilizers of cells engaged in innate. They are involved in inflammatory host defenses, cell growth, differentiation, cell death, angiogenesis, and development and repair processes aimed at the restoration of homeostasis as well. There are various cytokines hyperproduced in severe cases of COVID-19, such as IL-1, IL-6, IL-12, IFN-γ, and TNF-α (Turner et al., 2014; Vabret et al., 2020). Scientist are trying to figure how to stop the strong production of these immune mediators (J. S. Kim et al., 2021; Nazerian et al., 2021; L. Yang et al., 2021). Unregulated cytokines are also related to multiple types of cancer (Landskron et al., 2014). In our results, we can observe that some of the most enriched terms relate to cancer, such as “proteoglycans in cancer”, “prostate cancer” and “chronic myeloid leukemia”, and others are related to oncogenic viruses, such as “human T-cell leukemia virus infection” and “Kaposi sarcoma-associated herpesvirus infection” (**Figure 6 and 7**).

There are various viruses refereed in this analysis, in the 30 most enriched terms, 9 of them are diseases or infections caused by other viruses: human cytomegalovirus infection, human T-cell leukemia virus infection, human immunodeficiency virus 1 infection, Kaposi sarcoma-associated herpesvirus infection, hepatitis B, influenza A, hepatitis C and measles (**Figure 6 and 7**).

This can be attributed to the fact that SARS-CoV-2 share the same or similar pathways or processes with other viruses. Therefore, these results reinforce the connection between certain cell activity pathways and different viruses’ entry and replication in human cells.

### 3.5. GO enrichment analysis: gene-concept network results and discussion

The cnetplot depicts the linkages of genes and GO terms as a network. This network is helpful to see which genes are involved in enriched terms and which genes may belong to multiple annotation categories. The results for this network are represented in **Figure 8**. All enriched terms represented have around the same dot size, 300 genes involved in each.

The term “regulation of lipid metabolic process” means any process that modulates the frequency, rate or extent of the chemical reactions and pathways involving lipids. In the dot plot regarding the BP category of the GO enrichment analysis we can see the same term with the same count of genes (**Figure 8**). The lipid composition of cell membranes can influence viral entry by mediating fusion or affecting receptor conformation.

Cell membranes’ lipid composition influence the viral entry by mediating fusion and/or affecting the receptor conformation, therefor lipids play an essential role in the viral life cycle (Theken et al., 2021). Multiple studies have found that lipids have an impact in SARS-CoV-2 entry in the host cell (Ebrahimi & McCullagh, 2021; Luchini et al., 2021; Theken et al., 2021).

The term “response to oxygen levels” may be related with “response to oxidative stress”, and several other terms found relevant in the BP dot plot, since extremely low levels of oxygen can induce oxidative stress, has mentioned above (McGarry et al., 2018). One of the most eminent symptoms of COVID-19 is histopathological complications in the lungs that lead to low oxygen levels and hypoxia in severe cases (Rahman et al., 2021; Shenoy et al., 2020; Tobin et al., 2020).

The term “response to virus” means any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) because of a stimulus from a virus. In the KEGG and MeSH enrichment analysis we saw that out of the 30 most enriched terms, nine of them are diseases or infections caused by other viruses (**Figure 6 and 7**). This term also makes sense since we are analyzing the genes related to SARS-CoV-2, a virus.

The terms “calcium ion homeostasis”, “positive regulation of cytosolic calcium ion concentration”, “calcium ion homeostasis” and “regulation of cytosolic calcium ion concentration” appear together due to the amount of genes they all share, having into

account that this terms all refer to the balance of the calcium ion (**Figure 8**). This balance and its importance have already refereed above, in the BP dot plot discussion.

The terms “peptidyl-tyrosine modification” and “peptidyl-tyrosine phosphorylation” also appear together due to the amount of genes that both terms share (**Figure 8**). Scientist are looking into a way of making these a drug target (Mahoney et al., 2021; McBride & Machamer, 2010).

Finally, the term “gland development” means the process whose specific outcome is the progression of a gland over time, from its formation to the mature structure. A gland is an organ specialized for secretion. Scientist have found that SARS-CoV-2 can infect and replicate in various glands, such as salivary glands, pituitary glands and sweat glands (W. T. Gu et al., 2021; Liu et al., 2020; *SARS-CoV-2 infects and replicates in the salivary glands, study finds*, 2021).

Most of the genes in this network are only related with one set of interrelated terms (terms referred above as similar or related) but some genes are connected with two or more sets (**Figure 8**). Multiple genes are connected between themselves.

For example, the gene Protein-tyrosine kinase 2-beta (PTK2B) is responsible for the cell polarization, cell migration, adhesion, spreading and bone remodeling (*Protein-tyrosine kinase 2-beta*, n.d.). As we saw before, cell adhesion and its related terms appeared various times (**Figure 4**). This gene is also responsible for the osteoclastic bone resorption and when activated in response to stimuli, lead to increased intracellular calcium ion levels (*Protein-tyrosine kinase 2-beta*, n.d.). This activation is not direct and may be mediated by calcium-mediated production of reactive oxygen species, leading to oxidative stress. In the network, this gene is linked to the set of interrelated terms that involve the calcium ion, which makes sense since PTK2B can increase intracellular levels of calcium. PTK2B is also connected with “response to oxygen levels” , which may be related to the production of reactive oxygen species and “peptidyl-tyrosine modification” and “peptidyl-tyrosine phosphorylation”, since the activation of some pathways lead to the phosphorylation of additional tyrosine residues (Lev et al., 1995; Ohya et al., 1999; *Protein-tyrosine kinase 2-beta*, n.d.).

Protein kinase C epsilon type (PRKCE) is a serine/threonine-protein kinase that plays essential roles in cell adhesion, motility, migration and cell cycle, functions in neuron growth and ion channel regulation, and is involved in immune response, cancer cell invasion and regulation of apoptosis (*Protein kinase C epsilon type*, n.d.). PRKCE is connected to the set of terms related to calcium ion, which make sense since it can be

activated by calcium (*PRKCE protein kinase C epsilon [Homo sapiens (human)] - Gene - NCBI*, n.d.), PRKCE has cardioprotective characteristics when one suffers from ischemia (*PRKCE protein kinase C epsilon [Homo sapiens (human)] - Gene - NCBI*, n.d.), that can explain the gene connection with the term “response to oxygen levels” (**Figure 8**). The gene is also connected to “regulation of lipid metabolic process” that can be explained by the new research that found that PKCE in adipose tissue affects liver gene transcription and lipid metabolism (Brandon et al., 2019). This gene is also involved in the regulation of peptidyl-tyrosine phosphorylation regulation of peptidyl-tyrosine phosphorylation (*Protein kinase C epsilon type*, n.d.).

Hypoxia-inducible factor 1-alpha (HIF1A) is a master transcriptional regulator of the adaptive response to hypoxia, when subjected to hypoxia, activates the transcription of over 40 genes to increase oxygen delivery or facilitate metabolic adaptation to hypoxia (*Hypoxia-inducible factor 1-alpha*, n.d.). As predicted, HIF1A is connected to the term “response to oxygen levels” (**Figure 8**). When a human is infected by SARS-CoV-2, HIF1A induces glycolysis in monocytes and a proinflammatory state (Codo et al., 2020). This gene also promotes monocyte inflammatory response, the expression of cytokines and the replication of SARS-CoV-2 (Codo et al., 2020). This alone my explain the connection with the term “response to virus” (**Figure 8**).

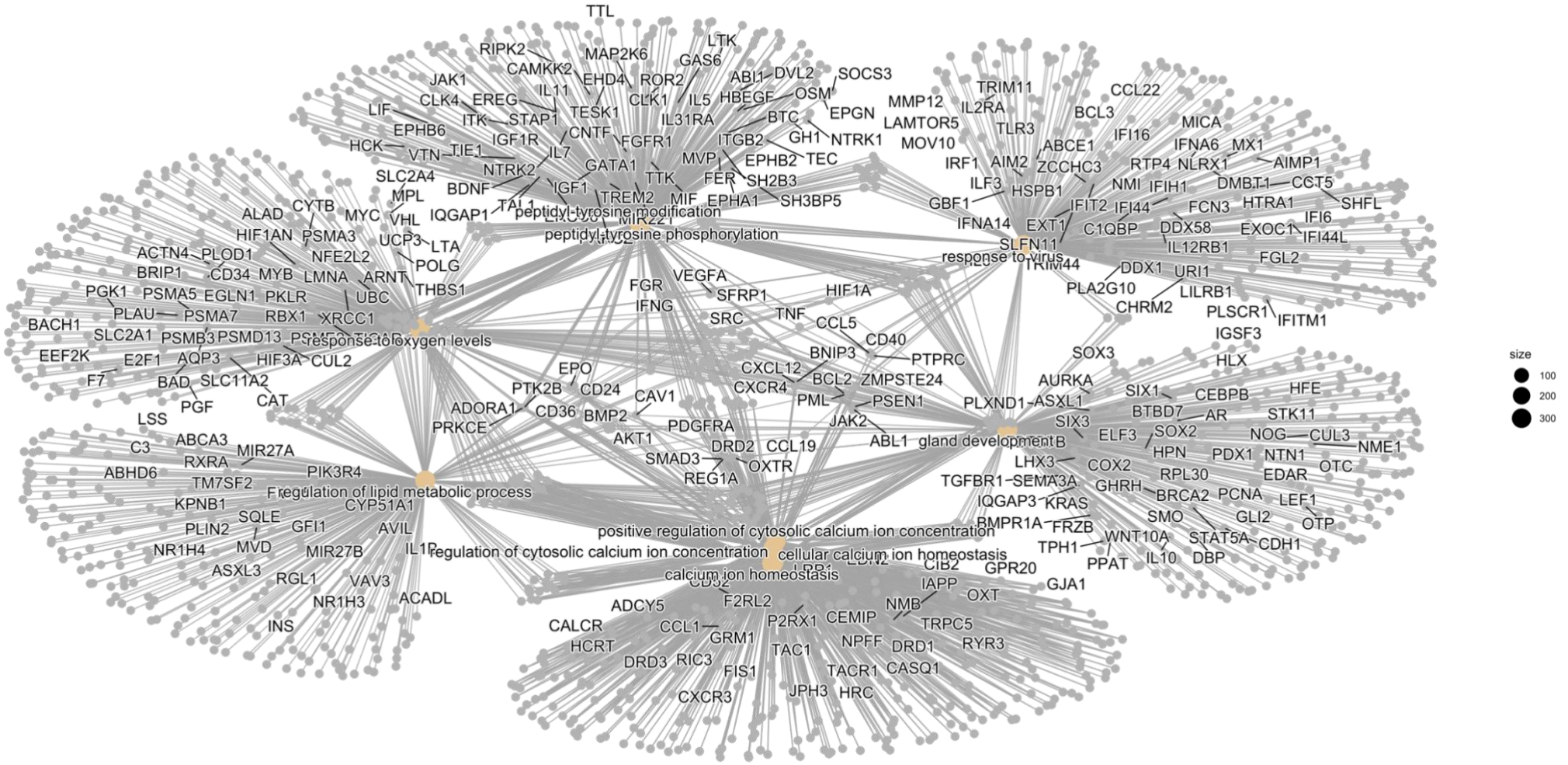


**Fig. 8** - GO enrichment analysis gene-concept network: yellow nodes – GO terms; grey dots: genes; GO: Gene Ontology.

# IV. Supplementary bioinformatic approaches

## 4.1. M protein docking - Methods

There are numerous bioinformatic approaches that can be used to address the SARS-CoV-2 issue. In a side project, we aimed to discover the SARS-CoV-2 M protein structure. With this project we developed an article entitled “SARS-CoV-2 membrane protein: from genomic data to structural new insights”. (Marques-Pereira et al., 2021) This work was split into three main steps: M protein monomer membrane orientation prediction, M protein dimer 3D structure prediction and mutation effect assessment in the homodimer interface. My main participation was in the docking, in the second step.

We selected OPM, TMpred and TMHMM protein monomers from the system equilibration results and subjected to M protein dimer prediction. To guide the protein-protein docking we used known information on SARS-CoV M protein that has a 90.5% sequence identity and 90% homology with SARS-CoV-2 M protein. (S. Thomas, 2020) Two equilibrated M protein monomers from each membrane orientation were used for dimer prediction using the docking tool HADDOCK, version 2.4, a protein quaternary structure predictor based on experimental data. (Zundert, 2016) Since M protein is a membrane protein and most homodimers are symmetric, water docking results were not considered and docking results with TMH2 and TMH3 non-crystallographic symmetry restraints were generated. (Blundell & Srinivasan, 1996) To determine M protein monomer’s active residues, CPORT, a protein-protein residue interaction predictor at an atomic-level, was used and only transmembrane residues predicted by this tool were considered for downstream steps. (Vries & Bonvin, 2011)

For each membrane predictor, 5000 dimer structures were generated in rigid body docking phase and 1000 structures for the semi-flexible refinement phase. Dimer results were examined, according to each monomer membrane orientation prediction through an in-house Python script. Upon the selection of the most 20 promising HADDOCK dimers 3D structures, we extended our work towards interface interacting residues prediction. Protein Interfaces, Surfaces and Assemblies (PISA), a web-based tool that resorts to chemical-physical principles for analyzing and modeling of macromolecular interactions, was used as a first predictor for dimer interface residues on all twenty dimer structures. (Krissinel & Henrick, 2007) Two

dimers were chosen based on PISA results and their comparison with SARS-CoV's M protein dimer experimental results, highlighted homologous SARS-CoV-2 residues W20, W58, P59, W92, Y95, F96 and C159 as important residues for dimer stabilization. Selected structures were further subjected to PRODIGY. (Xue et al., 2016) PRODIGY predicts dimer interacting residues and helps to determine if a protein interface is crystallographic or biological, the latter meaning that the predicted dimer is biologically relevant.

## 4.2. M protein docking - Results and Discussion

Regarding SARS-CoV-2 M protein structure study, OPM, TMpred and TMHMM M monomers were used to model dimer 3D structures using a well-established protein-protein docking software: HADDOCK. From 3000 proposed docking decoys, 1000 for each membrane orientation, 20 dimer structures that respected the membrane orientation prediction were selected: 11 from OPM, 4 from TMpred and 5 from TMHMM. From these 20 dimers, two structures from the TMHMM membrane predictor were chosen based on their similarity with SARS-CoV experimental detected interactions, namely in TMH2 (P59) and TMH3 (W92, L93, F96) regions. From these two TMHMM M protein dimers, the final choice was based on PROtein binDIng enerGY (PRODIGY)'s metrics of biological probability and predicted binding affinity. Hence, the M protein dimer structure chosen for the proceeding studies showed 85.6% biological probability and a predicted binding affinity of -6.3 kcal/mol in comparison to 74.8% biological probability and -5.9 kcal/mol binding affinity results from the other available structure. Regarding the TMHMM monomer membrane prediction that served as template for the final chosen dimer, M protein monomer residues 11-19 were shown to stably belong to N-terminal domain, residues 100-203 to C-terminal domain, residues 20-38 to TMH1, residues 46-70 to TMH2 and residues 76-100 to TMH3 (**Figure 9**).

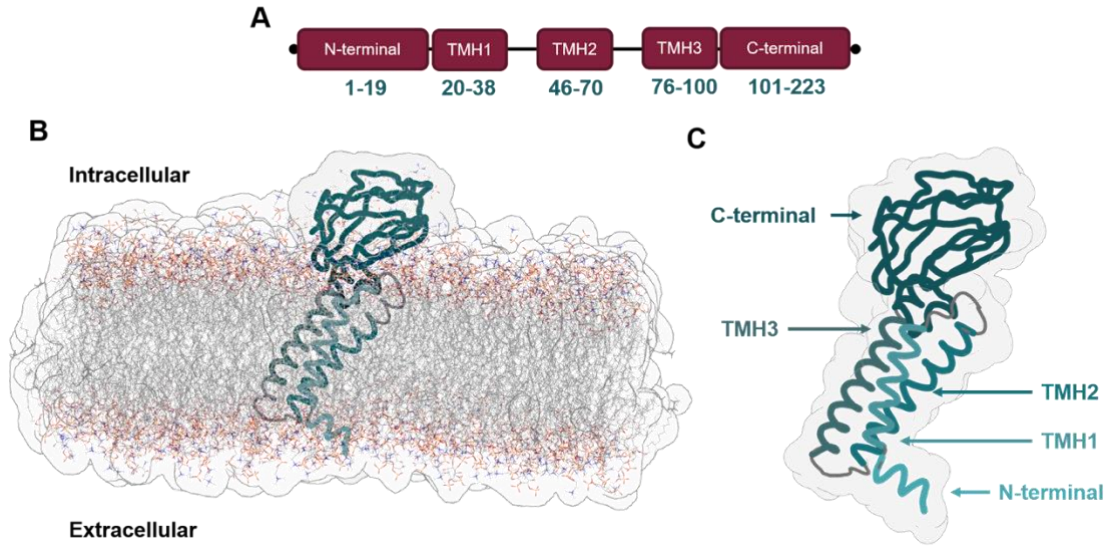


**Fig. 9** - SARS-CoV-2 M protein monomer. a) M protein domains predicted by TMHMM20,21 membrane predictor. b) TMHMM20,21 M protein monomer structure prediction after equilibration in membrane with ER membrane composition. c) M protein structure with domains highlighted.

The final dimer 3D structure (**Figure 10**) was subjected to three independent dimer system molecular dynamic replicas of 0.5 μs. After equilibration, polar contacts between M protein monomer and membrane lipids occurred in M monomer residues K14, Y39, R42, N43, R44, F45, Y71, R72, W75, S94, R101, R107, W110, S173, R174. Transmembrane regions were within membrane lipids throughout the entire equilibration and several M protein residues were able to establish polar contacts with membrane lipids, supporting our transmembrane prediction (**Figure 10**).

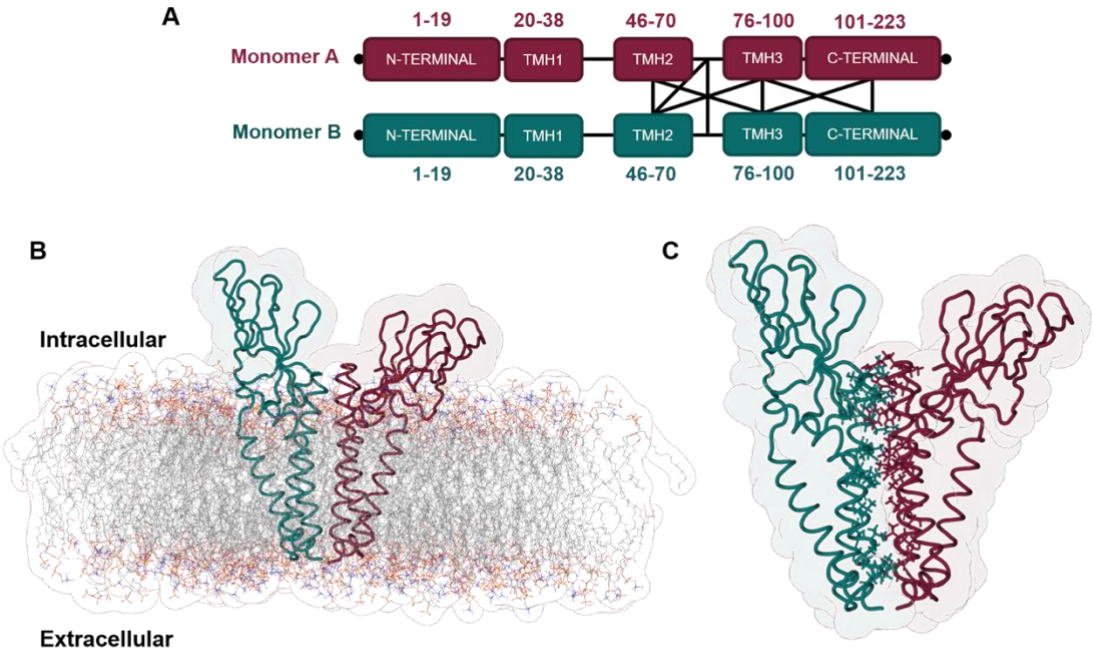


**Fig 10** - SARS-CoV-2 M protein dimer HADDOCK prediction using TMHMM based monomers. **A-** Interaction representation between Monomer A (red) and Monomer B (blue) domains. **B-** M protein dimer within the membrane: Monomer A (red), Monomer B (blue). **C-** M protein dimer with interfacial residues highlighted in a stick representation: Monomer A (red), Monomer B (blue).

For the first time, a reliable SARS-CoV2 M protein membrane orientation was proposed by this work that showed that residues 20-38 belong to TMH1, residues 46-70 to TMH2 and residues 76-100 to TMH3, results in agreement to the above mentioned SARS-CoV experimental results.

# V. Conclusions

In this project we tried several TM and gene enrichment analysis approaches. First, we tried to compare the data retrieved from LitCovid to the one retrieved from the pubtator function and realized that the amount of that retrieved from LitCovid was far superior to the one from the script This confirms that the web tools available for specific biomedical analysis keep improving and that the impact of this pandemic has also been reflected in an explosion of new tools and new studies.

From the joint analysis of the results from clusterprofiler, we have reached some conclusions. Some of the most enriched terms had hundreds of articles concerning COVID-19 or SARS-CoV-2, with data from *in silico*, *in vitro* and *in vivo* research. This bought that the amount of information released regarding this virus came quickly as the pandemic began in 2019. Doing a comprehensive study of so much gene information manually would be tricky. Thus, a bioinformatics analysis was the right choice.

When we analyzed the data regarding KEGG and MeSH results we found out that various viruses share multiple genes with SARS-CoV-2, which was foreseeable since the virus often share characteristics. We also found out that some of these viruses are respiratory virus, such as Influenza A and Measles.

From what we obtained, we can state than the concentration and homeostasis of calcium ion in SARS-CoV-2 patients is essential and being able to control this concentration through drugs would allow better control of the severity of COVID-19. There are some drugs already design to target those mechanisms, redesigning drugs can be the next solution for severe COVID-19.

We can also affirm that the low levels of oxygen and high oxidative stress are getting more research and are a major problem in the development of the disease. Just like HIF1A, more genes can have a direct impact in the virus entry and replication.

This project may lead to future research since we only analyzed a small part of the results and many of the enriched pathways or terms may lead possible drug targets to find out.

# References


A, Hemmings. (2012). Pi3k-pkb/akt pathway[J. *Cold Spring Harbor Perspectives in Biology*, *4*(9), 011189.

A, K, S., G, V., O, F., P, M.-G., V, A., P, D. 'eustachio, L, S., & H, H. (n.d.). Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics. Mar*.

Abdel-Aziz, M. I., Kermani, N. Z., Neerincx, A. H., Vijverberg, S. J. H., Guo, Y., Howarth, P., Dahlen, S.-E., Djukanovic, R., Sterk, P. J., Kraneveld, A. D., Maitland-van der Zee, A. H., Chung, K. F., Adcock, I. M., & On behalf the U-BIOPRED Consortium. (2021). Association of endopeptidases, involved in SARS-CoV-2 infection, with microbial aggravation in sputum of severe asthma. *Allergy*, *76*(6), 1917–1921. https://doi.org/10.1111/all.14731

Alan Sariol, S. P. (2020). Lessons for COVID-19 Immunity from Other Coronavirus Infections. *Immunity*, *53*(2), 248–263. https://doi.org/10.1016/j.immuni.2020.07.005

Alter, G., Yu, J., Liu, J., Chandrashekar, A., Borducchi, E. N., Tostanoski, L. H., McMahan, K., Jacob-Dolan, C., Martinez, D. R., Chang, A., Anioke, T., Lifton, M., Nkolola, J., Stephenson, K. E., Atyeo, C., Shin, S., Fields, P., Kaplan, I., Robins, H., … Barouch, D. H. (2021). Immunogenicity of Ad26.COV2.S vaccine against SARS-CoV-2 variants in humans. *Nature*, *596*(7871), 268–272. https://doi.org/10.1038/s41586-021-03681-2

Ames, R. M., MacPherson, J. I., Pinney, J. W., Lovell, S. C., & Robertson, D. L. (2013). Modular biological function is most effectively captured by combining molecular interaction data types. *PloS One*, *8*(5), e62670. https://doi.org/10.1371/journal.pone.0062670

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Baden, L. R. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *The New England Journal of Medicine*, *384*(5), 403–416. https://doi.org/10.1056/nejmoa2035389

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Rouphael, N., Creech, C. B., McGettigan, J., Khetan, S., Segall, N., Solis, J., Brosz, A., Fierro, C., Schwartz, H., Neuzil, K., Corey, L., … COVE Study Group. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *The New England Journal of Medicine*, *384*(5), 403–416. https://doi.org/10.1056/NEJMoa2035389

Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews. Genetics*, *12*(1), 56–68. https://doi.org/10.1038/nrg2918

Belouzard, S., Millet, J. K., Licitra, B. N., & Whittaker, G. R. (2012). Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses*, *4*(6), 1011–1033. https://doi.org/10.3390/v4061011

Beniac, D. R., Andonov, A., Grudeski, E., & Booth, T. F. (2006). Architecture of the SARS coronavirus prefusion spike. *Nature Structural & Molecular Biology*, *13*(8), 751–752. https://doi.org/10.1038/nsmb1123

Blundell, T. L., & Srinivasan, N. (1996). Symmetry, stability, and dynamics of multidomain and multicomponent protein systems. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(25), 14243–14248. https://doi.org/10.1073/pnas.93.25.14243

Bolcato, G., Bissaro, M., Pavan, M., Sturlese, M., & Moro, S. (2020a). Targeting the Coronavirus SARS-CoV-2: computational insights into the mechanism of action of the protease inhibitors Lopinavir, Ritonavir, and Nelfinavir. In *Research Square*. https://doi.org/10.21203/rs.3.rs-20948/v1

Bolcato, G., Bissaro, M., Pavan, M., Sturlese, M., & Moro, S. (2020b). Targeting the coronavirus SARS-CoV-2: computational insights into the mechanism of action of the protease inhibitors lopinavir, ritonavir and nelfinavir. *Scientific Reports*, *10*(1), 20927. https://doi.org/10.1038/s41598-020-77700-z

Boncristiani, H. F., Criado, M. F., & Arruda, E. (2009). Respiratory Viruses. In *Encyclopedia of Microbiology* (pp. 500–518). Elsevier. https://doi.org/10.1016/b978-012373944-5.00314-x

Brandon, A. E., Liao, B. M., Diakanastasis, B., Parker, B. L., Raddatz, K., McManus, S. A., O'Reilly, L., Kimber, E., van der Kraan, A. G., Hancock, D., Henstridge, D. C., Meikle, P. J., Cooney, G. J., James, D. E., Reibe, S., Febbraio, M. A., Biden, T. J., & Schmitz-Peiffer, C. (2019). Protein kinase C epsilon deletion in adipose tissue, but not in liver, improves glucose tolerance. *Cell Metabolism*, *29*(1), 183-191.e7. https://doi.org/10.1016/j.cmet.2018.09.013

Calabrese, F., Pezzuto, F., Fortarezza, F., Hofman, P., Kern, I., Panizo, A., von der Thüsen, J., Timofeev, S., Gorkiewicz, G., & Lunardi, F. (2020). Pulmonary pathology and COVID-19: lessons from autopsy. The experience of European Pulmonary Pathologists. *Virchows Archiv: An International Journal of Pathology*, *477*(3), 359–372. https://doi.org/10.1007/s00428-020-02886-6

Cecchini, R., & Cecchini, A. L. (2020). SARS-CoV-2 infection pathogenesis is related to oxidative stress as a response to aggression. *Medical Hypotheses*, *143*(110102), 110102. https://doi.org/10.1016/j.mehy.2020.110102

Cekerevac, I., Turnic, T. N., Draginic, N., Andjic, M., Zivkovic, V., Simovic, S., Susa, R., Novkovic, L., Mijailovic, Z., Andjelkovic, M., Vukicevic, V., Vulovic, T., & Jakovljevic, V. (2021). Predicting severity and intrahospital mortality in COVID-19: The place and role of oxidative stress. *Oxidative Medicine and Cellular Longevity*, *2021*, 6615787. https://doi.org/10.1155/2021/6615787

Chang, L., Zhao, L., Gong, H., Wang, L., & Wang, L. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 RNA Detected in Blood Donations. *Emerging Infectious Diseases*, *26*(7), 1631–1633. https://doi.org/10.3201/eid2607.200839

Chernyak, B. V., Popova, E. N., Prikhodko, A. S., Grebenchikov, O. A., Zinovkina, L. A., & Zinovkin, R. A. (2020). COVID-19 and oxidative stress. *Biochemistry. Biokhimiia*, *85*(12), 1543–1553. https://doi.org/10.1134/S0006297920120068

Cho, H., Choi, W., & Lee, H. (2017). A method for named entity normalization in biomedical articles: application to diseases and plants. *BMC Bioinformatics*, *18*(1), 451. https://doi.org/10.1186/s12859-017-1857-8

Choksi, T. T., Zhang, H., Chen, T., & Malhotra, N. (2021). Outcomes of hospitalized COVID-19 patients receiving renin angiotensin system blockers and calcium channel blockers. *American Journal of Nephrology*, *52*(3), 250–260. https://doi.org/10.1159/000515232

Codo, A. C., Davanzo, G. G., Monteiro, L. de B., de Souza, G. F., Muraro, S. P., Virgilio-da-Silva, J. V., Prodonoff, J. S., Carregari, V. C., de Biagi Junior, C. A. O., Crunfli, F., Jimenez Restrepo, J. L., Vendramini, P. H., Reis-de-Oliveira, G., Bispo Dos Santos, K., Toledo-Teixeira, D. A., Parise, P. L., Martini, M. C., Marques, R. E., Carmo, H. R., … Moraes-Vieira, P. M. (2020). Elevated glucose levels favor SARS-CoV-2 infection and monocyte response through a HIF-1α/glycolysis-dependent axis. *Cell Metabolism*, *32*(3), 437-446.e5. https://doi.org/10.1016/j.cmet.2020.07.007

Collins, A. R., Knobler, R. L., Powell, H., & Buchmeier, M. J. (1982). Monoclonal antibodies to murine hepatitis virus-4 (strain JHM) define the viral glycoprotein responsible for attachment and cell--cell fusion. *Virology*, *119*(2), 358–371. https://doi.org/10.1016/0042-6822(82)90095-2

Cornelissen, L. A., Wierda, C. M., van der Meer, F. J., Herrewegh, A. A., Horzinek, M. C., Egberink, H. F., & de Groot, R. J. (1997). Hemagglutinin-esterase, a novel structural protein of torovirus. *Journal of Virology*, *71*(7), 5277–5286. https://doi.org/10.1128/JVI.71.7.5277-5286.1997

*Coronavirus 2019*. (n.d.). CDC. Retrieved June 18, 2021, from https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L. Y., Toufighi, K., Mostafavi, S., Prinz, J., St. Onge, R. P., VanderSluis, B., Makhnevych, T., Vizeacoumar, F. J., Alizadeh, S., Bahr, S., Brost, R. L., Chen, Y., … Boone, C. (2010). The genetic landscape of a cell. *Science (New York, N.Y.)*, *327*(5964), 425–431. https://doi.org/10.1126/science.1180823

Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., & Yelowitz, A. (2020). Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate. *Health Affairs* , *39*(7), 1237–1246. https://doi.org/10.1377/hlthaff.2020.00608

*covid19*. (n.d.). WHO. Retrieved October 24, 2021, from https://covid19.who.int/table

COVID-19 vaccines. (2020). In *Drugs and Lactation Database (LactMed)*. National Library of Medicine. https://www.ncbi.nlm.nih.gov/books/NBK565969/

Crespi, B., & Alcock, J. (2021). Conflicts over calcium and the treatment of COVID-19. *Evolution, Medicine, and Public Health*, *9*(1), 149–156. https://doi.org/10.1093/emph/eoaa046

Cui, J., Li, F., & Shi, Z.-L. (2018). Origin and evolution of pathogenic coronaviruses. *Nature Reviews. Microbiology*, *17*(3), 181–192. https://doi.org/10.1038/s41579-018-0118-9

Danta, C. C. (2020). Calcium channel blockers: A possible potential therapeutic strategy for the treatment of Alzheimer's dementia patients with SARS-CoV-2 infection. *ACS Chemical Neuroscience*, *11*(15), 2145–2148. https://doi.org/10.1021/acschemneuro.0c00391

de Groot, R. J., Luytjes, W., Horzinek, M. C., van der Zeijst, B. A., Spaan, W. J., & Lenstra, J. A. (1987). Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *Journal of Molecular Biology*, *196*(4), 963–966. https://doi.org/10.1016/0022-2836(87)90422-0

Dong, L., Tian, J., He, S., Zhu, C., Wang, J., Liu, C., & Yang, J. (2020). Possible Vertical Transmission of SARS-CoV-2 From an Infected Mother to Her Newborn. *JAMA: The Journal of the American Medical Association*, *323*(18), 1846–1848. https://doi.org/10.1001/jama.2020.4621

Donia, A., & Bokhari, H. (2021). Apoptosis induced by SARS-CoV-2: can we target it? *Apoptosis: An International Journal on Programmed Cell Death*, *26*(1–2), 7–8. https://doi.org/10.1007/s10495-021-01656-2

Du, H., Dong, X., Zhang, J.-J., Cao, Y.-Y., Akdis, M., Huang, P.-Q., Chen, H.-W., Li, Y., Liu, G.-H., Akdis, C. A., Lu, X.-X., & Gao, Y.-D. (2021). Clinical characteristics of 182 pediatric COVID-19 patients with different severities and allergic status. *Allergy*, *76*(2), 510–532. https://doi.org/10.1111/all.14452

Eastin, C., & Eastin, T. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. In *The Journal of Emergency Medicine* (Vol. 58, Issue 4, pp. 711–712). https://doi.org/10.1016/j.jemermed.2020.04.004

Ebrahimi, K. H., & McCullagh, J. S. O. (2021). A lipidomic view of SARS-CoV-2. *Bioscience Reports*, *41*(8). https://doi.org/10.1042/BSR20210953

Esposito, A. J., Menon, A. A., Ghosh, A. J., Putman, R. K., Fredenburgh, L. E., El-Chemaly, S. Y., Goldberg, H. J., Baron, R. M., Hunninghake, G. M., & Doyle, T. J. (2020). Increased Odds of Death for Patients with Interstitial Lung Disease and COVID-19: A Case-Control Study. *American Journal of Respiratory and Critical Care Medicine*, *202*(12), 1710–1713. https://doi.org/10.1164/rccm.202006-2441LE

Fa, B. C. H. C. (2017). *Epidemiology of Viral Infections* (pp. 185–203). Academic Press. https://doi.org/10.1016/B978-0-12-375156-0.00013-8

Falsey, A. R., Sobieszczyk, M. E., Hirsch, I., Sproule, S., Robb, M. L., Corey, L., Neuzil, K. M., Hahn, W., Hunt, J., Mulligan, M. J., McEvoy, C., DeJesus, E., Hassman, M., Little, S. J., Pahud, B. A., Durbin, A., Pickrell, P., Daar, E. S., Bush, L., … AstraZeneca AZD1222 Clinical Study Group. (2021). Phase 3 safety and efficacy of AZD1222 (ChAdOx1 nCoV-19) Covid-19 vaccine. *The New England Journal of Medicine*. https://doi.org/10.1056/NEJMoa2105290

Fauquet, C. M. (1999). Taxonomy, classification and nomenclature of viruses. In *Encyclopedia of Virology* (pp. 1730–1756). Elsevier. https://doi.org/10.1006/rwvi.1999.0277

Fernandes, I. G., de Brito, C. A., Dos Reis, V. M. S., Sato, M. N., & Pereira, N. Z. (2020). SARS-CoV-2 and other respiratory viruses: What does oxidative stress have to do with it? *Oxidative Medicine and Cellular Longevity*, *2020*, 8844280. https://doi.org/10.1155/2020/8844280

Fleuren, W. W. M., & Alkema, W. (2015). Application of text mining in the biomedical domain. *Methods* , *74*, 97–106. https://doi.org/10.1016/j.ymeth.2015.01.015

Forcados, G. E., Muhammad, A., Oladipo, O. O., Makama, S., & Meseko, C. A. (2021). Metabolic implications of oxidative stress and inflammatory process in SARS-CoV-2 pathogenesis: Therapeutic potential of natural antioxidants. *Frontiers in Cellular and Infection Microbiology*, *11*, 654813. https://doi.org/10.3389/fcimb.2021.654813

Frenck, R. W., Jr, Klein, N. P., Kitchin, N., Gurtman, A., Absalon, J., Lockhart, S., Perez, J. L., Walter, E. B., Senders, S., Bailey, R., Swanson, K. A., Ma, H., Xu, X., Koury, K., Kalina, W. V., Cooper, D., Jennings, T., Brandon, D. M., Thomas, S. J., … C4591001 Clinical Trial Group. (2021). Safety, immunogenicity, and efficacy of the BNT162b2 Covid-19 vaccine in adolescents. *The New England Journal of Medicine*, *385*(3), 239–250. https://doi.org/10.1056/NEJMoa2107456

Gachloo, M., Wang, Y., & Xia, J. (2019). A review of drug knowledge discovery using BioNLP and tensor or matrix decomposition. *Genomics & Informatics*, *17*(2), e18. https://doi.org/10.5808/GI.2019.17.2.e18

Gao, Y.-D., Ding, M., Dong, X., Zhang, J.-J., Kursat Azkur, A., Azkur, D., Gan, H., Sun, Y.-L., Fu, W., Li, W., Liang, H.-L., Cao, Y.-Y., Yan, Q., Cao, C., Gao, H.-Y., Brüggen, M.-C., van de Veen, W., Sokolowska, M., Akdis, M., & Akdis, C. A. (2021). Risk factors for severe and critically ill COVID-19 patients: A review. *Allergy*, *76*(2), 428–455. https://doi.org/10.1111/all.14657

Gelderblom, H. R. (2011). Structure and Classification of Viruses. In S. Baron (Ed.), *Medical Microbiology*. University of Texas Medical Branch at Galveston. https://www.ncbi.nlm.nih.gov/books/NBK8174/

Ghamami, F., & Keyvanpour, M. (n.d.). WHY BIOMEDICAL RELATION EXTRACTION IS AN OPEN ISSUE? *ICIC Express Letters*. https://doi.org/10.24507/icicelb.09.08.747

Ghasemnejad-Berenji, M., & Pashapour, S. (2021). SARS-CoV-2 and the possible role of RAF/MEK/ERK pathway in viral survival: Is this a potential therapeutic strategy for COVID-19? *Pharmacology*, *106*(1–2), 119–122. https://doi.org/10.1159/000511280

Gonzalez, G. H., Tahsin, T., Goodale, B. C., Greene, A. C., & Greene, C. S. (2016). Recent Advances and Emerging Applications in Text and Data Mining for Biomedical Discovery. *Briefings in Bioinformatics*, *17*(1), 33–42. https://doi.org/10.1093/bib/bbv087

*g:Profiler – a web server for functional enrichment analysis and conversions of gene lists*. (n.d.). Retrieved October 24, 2021, from https://biit.cs.ut.ee/gprofiler/gost

Gu, J., Han, B., & Wang, J. (2020). COVID-19: Gastrointestinal Manifestations and Potential Fecal-Oral Transmission. *Gastroenterology*, *158*(6), 1518–1519. https://doi.org/10.1053/j.gastro.2020.02.054

Gu, W. T., Zhou, F., Xie, W. Q., Wang, S., Yao, H., Liu, Y. T., Gao, L., & Wu, Z. B. (2021). A potential impact of SARS-CoV-2 on pituitary glands and pituitary neuroendocrine tumors. *Endocrine*, *72*(2), 340–348. https://doi.org/10.1007/s12020-021-02697-y

Halfmann, P. J., Hatta, M., Chiba, S., Maemura, T., Fan, S., Takeda, M., Kinoshita, N., Hattori, S.-I., Sakai-Tagawa, Y., Iwatsuki-Horimoto, K., Imai, M., & Kawaoka, Y. (2020). Transmission of SARS-CoV-2 in Domestic Cats. *The New England Journal of Medicine*, *383*(6), 592–594. https://doi.org/10.1056/NEJMc2013400

Hemmings, B. A., & Restuccia, D. F. (2012). PI3K-PKB/Akt pathway. *Cold Spring Harbor Perspectives in Biology*, *4*(9), a011189. https://doi.org/10.1101/cshperspect.a011189

Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2020). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews. Microbiology*, *19*(3), 141–154. https://doi.org/10.1038/s41579-020-00459-7

Hu, M., Lin, H., Wang, J., Xu, C., Tatem, A. J., Meng, B., Zhang, X., Liu, Y., Wang, P., Wu, G., Xie, H., & Lai, S. (2021). Risk of Coronavirus Disease 2019 Transmission in Train Passengers: an Epidemiological and Modeling Study. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *72*(4), 604–610. https://doi.org/10.1093/cid/ciaa1057

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Huang, S., Wang, J., Liu, F., Liu, J., Cao, G., Yang, C., Liu, W., Tu, C., Zhu, M., & Xiong, B. (2020). COVID-19 patients with hypertension have more severe disease: a multicenter retrospective observational study. *Hypertension Research: Official Journal of the Japanese Society of Hypertension*, *43*(8), 824–831. https://doi.org/10.1038/s41440-020-0485-2

*Hypoxia-inducible factor 1-alpha*. (n.d.). Retrieved October 30, 2021, from https://www.uniprot.org/uniprot/Q16665

Jackson, L. A., Anderson, E. J., Rouphael, N. G., Roberts, P. C., Makhene, M., Coler, R. N., McCullough, M. P., Chappell, J. D., Denison, M. R., Stevens, L. J., Pruijssers, A. J., McDermott, A., Flach, B., Doria-Rose, N. A., Corbett, K. S., Morabito, K. M., O'Dell, S., Schmidt, S. D., Swanson, P. A., 2nd, … mRNA-1273 Study Group. (2020). An mRNA vaccine against SARS-CoV-2 - preliminary report. *The New England Journal of Medicine*, *383*(20), 1920–1931. https://doi.org/10.1056/NEJMoa2022483

Jiang, X., Feyertag, F., Meehan, C. J., McCormack, G. P., Travers, S. A., Craig, C., Westby, M., Lewis, M., & Robertson, D. L. (2015). Characterizing the diverse mutational pathways associated with R5-tropic maraviroc resistance: HIV-1 that uses the drug-bound CCR5 coreceptor. *Journal of Virology*, *89*(22), 11457–11472. https://doi.org/10.1128/JVI.01384-15

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Karikó, K., Muramatsu, H., Welsh, F. A., Ludwig, J., Kato, H., Akira, S., & Weissman, D. (2008). Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *16*(11), 1833–1840. https://doi.org/10.1038/mt.2008.200

Khezri, M. R. (2021). PI3K/AKT signaling pathway: a possible target for adjuvant therapy in COVID-19. *Human Cell*, *34*(2), 700–701. https://doi.org/10.1007/s13577-021-00484-5

Kim, J. H. (2012). Chapter 8: Biological knowledge assembly and interpretation. *PLoS Computational Biology*, *8*(12), e1002858. https://doi.org/10.1371/journal.pcbi.1002858

Kim, J. S., Lee, J. Y., Yang, J. W., Lee, K. H., Effenberger, M., Szpirt, W., Kronbichler, A., & Shin, J. I. (2021). Immunopathogenesis and treatment of cytokine storm in COVID-19. *Theranostics*, *11*(1), 316–329. https://doi.org/10.7150/thno.49713

Kim, J.-M., Kim, H. M., Lee, E. J., Jo, H. J., Yoon, Y., Lee, N.-J., Son, J., Lee, Y.-J., Kim, M. S., Lee, Y.-P., Chae, S.-J., Park, K. R., Cho, S.-R., Park, S., Kim, S. J., Wang, E., Woo, S., Lim, A., Park, S.-J., … Yoo, C. K. (2020). Detection and Isolation of SARS-CoV-2 in Serum, Urine, and Stool

Specimens of COVID-19 Patients from the Republic of Korea. *Osong Public Health and Research Perspectives*, *11*(3), 112–117. https://doi.org/10.24171/j.phrp.2020.11.3.02

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research*, *49*(D1), D1388–D1395. https://doi.org/10.1093/nar/gkaa971

Klausegger, A., Strobl, B., Regl, G., Kaser, A., Luytjes, W., & Vlasak, R. (1999). Identification of a coronavirus hemagglutinin-esterase with a substrate specificity different from those of influenza C virus and bovine coronavirus. *Journal of Virology*, *73*(5), 3737–3743. https://doi.org/10.1128/JVI.73.5.3737-3743.1999

Kremsner, P. G., Guerrero, R. A. A., Arana, E., Aroca Martinez, G. J., Bonten, M. J. M., Chandler, R., Corral, G., De Block, E. J. L., Ecker, L., Gabor, J. J., Garcia Lopez, C. A., Gonzales, L., Granados González, M. A., Gorini, N., Grobusch, M. P., Hrabar, A. D., Junker, H., Kimura, A., Lanata, C. F., … HERALD Study Group. (2021). Efficacy and safety of the CVnCoV SARS-CoV-2 mRNA vaccine candidate: Results from herald, a phase 2b/3, randomised, observer-blinded, placebo-controlled clinical trial in ten countries in Europe and Latin America. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3911826

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, *372*(3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022

Kuhn, J. H. (2021). Virus Taxonomy. In *Encyclopedia of Virology* (pp. 28–37). Elsevier. https://doi.org/10.1016/b978-0-12-809633-8.21231-4

Kumar, S. (2012). General properties of viruses. In *Textbook of Microbiology* (pp. 501–501). Jaypee Brothers Medical Publishers (P) Ltd. https://doi.org/10.5005/jp/books/11668_53

Landskron, G., De la Fuente, M., Thuwajit, P., Thuwajit, C., & Hermoso, M. A. (2014). Chronic inflammation and cytokines in the tumor microenvironment. *Journal of Immunology Research*, *2014*, 149185. https://doi.org/10.1155/2014/149185

Leung, N. H. L. (2021). Transmissibility and transmission of respiratory viruses. *Nature Reviews. Microbiology*, *19*(8), 528–545. https://doi.org/10.1038/s41579-021-00535-6

Lev, S., Moreno, H., Martinez, R., Canoll, P., Peles, E., Musacchio, J. M., Plowman, G. D., Rudy, B., & Schlessinger, J. (1995). Protein tyrosine kinase PYK2 involved in Ca(2+)-induced regulation of ion channel and MAP kinase functions. *Nature*, *376*(6543), 737–745. https://doi.org/10.1038/376737a0

Li, D., Jin, M., Bao, P., Zhao, W., & Zhang, S. (2020). Clinical Characteristics and Results of Semen Tests Among Men With Coronavirus Disease 2019. *JAMA Network Open*, *3*(5), e208292. https://doi.org/10.1001/jamanetworkopen.2020.8292

Li, G., Wang, J., He, X., Zhang, L., Ran, Q., Xiong, A., Wu, D., Hu, L., Song, Q., & Zhu, D. (2020). An integrative analysis identifying transcriptional features and key genes involved in COVID-19. *Epigenomics*, *12*(22), 1969–1981. https://doi.org/10.2217/epi-2020-0168

Li, H., Chen, Q., Tang, B., Wang, X., Xu, H., Wang, B., & Huang, D. (2017). CNN-based ranking for biomedical entity normalization. *BMC Bioinformatics*, *18*(Suppl 11), 385. https://doi.org/10.1186/s12859-017-1805-7

*LitCovid*. (n.d.). Retrieved October 24, 2021, from https://www.ncbi.nlm.nih.gov/research/coronavirus/,

Liu, J., Li, Y., Liu, L., Hu, X., Wang, X., Hu, H., Hu, Z., Zhou, Y., & Wang, M. (2020). Infection of human sweat glands by SARS-CoV-2. *Cell Discovery*, *6*(1), 84. https://doi.org/10.1038/s41421-020-00229-y

Lokhande, A. S., & Devarajan, P. V. (2021). A review on possible mechanistic insights of Nitazoxanide for repurposing in COVID-19. *European Journal of Pharmacology*, *891*(173748), 173748. https://doi.org/10.1016/j.ejphar.2020.173748

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., … Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, *395*(10224), 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

Lu, X., Zhang, L., Du, H., Zhang, J., Li, Y. Y., Qu, J., Zhang, W., Wang, Y., Bao, S., Li, Y., Wu, C., Liu, H., Liu, D., Shao, J., Peng, X., Yang, Y., Liu, Z., Xiang, Y., Zhang, F., … Chinese Pediatric Novel Coronavirus Study Team. (2020). SARS-CoV-2 Infection in Children. *The New England Journal of Medicine*, *382*(17), 1663–1665. https://doi.org/10.1056/NEJMc2005073

Luchini, A., Micciulla, S., Corucci, G., Batchu, K. C., Santamaria, A., Laux, V., Darwish, T., Russell, R. A., Thepaut, M., Bally, I., Fieschi, F., & Fragneto, G. (2021). Lipid bilayer degradation induced

by SARS-CoV-2 spike protein as revealed by neutron reflectometry. *Scientific Reports*, *11*(1), 14867. https://doi.org/10.1038/s41598-021-93996-x

Mahoney, M., Damalanka, V. C., Tartell, M. A., Chung, D. H., Lourenço, A. L., Pwee, D., Mayer Bridwell, A. E., Hoffmann, M., Voss, J., Karmakar, P., Azouz, N. P., Klingler, A. M., Rothlauf, P. W., Thompson, C. E., Lee, M., Klampfer, L., Stallings, C. L., Rothenberg, M. E., Pöhlmann, S., … Janetka, J. W. (2021). A novel class of TMPRSS2 inhibitors potently block SARS-CoV-2 and MERS-CoV viral entry and protect human epithelial lung cells. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(43), e2108728118. https://doi.org/10.1073/pnas.2108728118

Malik, Y. A. (2020). Properties of Coronavirus and SARS-CoV-2. *The Malaysian Journal of Pathology*, *42*(1), 3–11. https://www.ncbi.nlm.nih.gov/pubmed/32342926

Marques-Pereira, C., Pires, M., Gouveia, R., Pereira, N., Caniceiro, A., Rosário-Ferreira, N., & Moreira, I. (2021). --SARS-CoV-2 membrane protein: from genomic data to structural new insights. In *Research Square*. Research Square. https://doi.org/10.21203/rs.3.rs-702792/v1

McBride, C. E., & Machamer, C. E. (2010). A single tyrosine in the severe acute respiratory syndrome coronavirus membrane protein cytoplasmic tail is important for efficient interaction with spike protein. *Journal of Virology*, *84*(4), 1891–1901. https://doi.org/10.1128/JVI.02458-09

McGarry, T., Biniecka, M., Veale, D. J., & Fearon, U. (2018). Hypoxia, oxidative stress and inflammation. *Free Radical Biology & Medicine*, *125*, 15–24. https://doi.org/10.1016/j.freeradbiomed.2018.03.042

McIntosh, K., & Perlman, S. (2015). Coronaviruses, including severe acute respiratory syndrome (SARS) and middle east respiratory syndrome (MERS). In *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases* (pp. 1928-1936.e2). Elsevier. https://doi.org/10.1016/b978-1-4557-4801-3.00157-0

*Medical subject headings - home page*. (2020). https://www.nlm.nih.gov/mesh/meshhome.html

Mehta, P., McAuley, D. F., Brown, M., Sanchez, E., Tattersall, R. S., Manson, J. J., & HLH Across Speciality Collaboration, UK. (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. *The Lancet*, *395*(10229), 1033–1034. https://doi.org/10.1016/S0140-6736(20)30628-0

Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Félix, E., Magariños, M. P., Mosquera, J. F., Mutowo, P., Nowotka, M., Gordillo-Marañón, M., Hunter, F., Junco, L., Mugumbate, G., Rodriguez-Lopez, M., Atkinson, F., Bosc, N., Radoux, C. J., Segura-Cabrera, A., … Leach, A. R. (2019). ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research*, *47*(D1), D930–D940. https://doi.org/10.1093/nar/gky1075

Millet, J. K., & Whittaker, G. R. (2018). Physiological and molecular triggers for SARS- CoV membrane fusion and entry into host cells. *Virol*, *517*, 3–8.

Mizutani, T., Fukushi, S., Saijo, M., Kurane, I., & Morikawa, S. (2005). JNK and PI3k/Akt signaling pathways are required for establishing persistent SARS-CoV infection in Vero E6 cells. *Biochimica et Biophysica Acta*, *1741*(1–2), 4–10. https://doi.org/10.1016/j.bbadis.2005.04.004

Morris, M. A., Young, L. S., & Dawson, C. W. (2008). DNA tumour viruses promote tumour cell invasion and metastasis by deregulating the normal processes of cell adhesion and motility. *European Journal of Cell Biology*, *87*(8–9), 677–697. https://doi.org/10.1016/j.ejcb.2008.03.005

Muhammad, Y., Kani, Y. A., Iliya, S., Muhammad, J. B., Binji, A., El-Fulaty Ahmad, A., Kabir, M. B., Umar Bindawa, K., & Ahmed, A. (2021). Deficiency of antioxidants and increased oxidative stress in COVID-19 patients: A cross-sectional comparative study in Jigawa, Northwestern Nigeria. *SAGE Open Medicine*, *9*, 2050312121991246. https://doi.org/10.1177/2050312121991246

Mukerji SS, S. I. H. (2021). What can we learn from brain autopsies in COVID-19? *Neuroscience Letters*, *742*, 135528. https://doi.org/10.1016/j.neulet.2020.135528

Muzaffar, A. W., Azam, F., & Qamar, U. (2015). A Relation Extraction Framework for Biomedical Text Using Hybrid Feature Set. *Computational and Mathematical Methods in Medicine*, *2015*, 910423. https://doi.org/10.1155/2015/910423

Nazerian, Y., Vakili, K., Ebrahimi, A., & Niknejad, H. (2021). Developing cytokine storm-sensitive therapeutic strategy in COVID-19 using 8P9R chimeric peptide and soluble ACE2. *Frontiers in Cell and Developmental Biology*, *9*, 717587. https://doi.org/10.3389/fcell.2021.717587

Ng, J. H., Hirsch, J. S., Wanchoo, R., Sachdeva, M., Sakhiya, V., Hong, S., Jhaveri, K. D., Fishbane, S., & Northwell COVID-19 Research Consortium and the Northwell Nephrology COVID-19 Research Consortium. (2020). Outcomes of patients with end-stage kidney disease hospitalized with COVID-19. *Kidney International*, *98*(6), 1530–1539. https://doi.org/10.1016/j.kint.2020.07.030

Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Uriarte, A., Malangone, C., Miranda, A., Fumis, L., Carvalho-Silva, D., Spitzer, M., Baker, J., Ferrer, J., Raies, A., Razuvayevskaya, O.,

Faulconbridge, A., Petsalaki, E., Mutowo, P., Machlitt-Northen, S., Peat, G., … McDonagh, E. M. (2021). Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Research*, *49*(D1), D1302–D1310. https://doi.org/10.1093/nar/gkaa1027

Ohya, K., Kajigaya, S., Kitanaka, A., Yoshida, K., Miyazato, A., Yamashita, Y., Yamanaka, T., Ikeda, U., Shimada, K., Ozawa, K., & Mano, H. (1999). Molecular cloning of a docking protein, BRDG1, that acts downstream of the Tec tyrosine kinase. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(21), 11976–11981. https://doi.org/10.1073/pnas.96.21.11976

Ou, M., Zhu, J., Ji, P., Li, H., Zhong, Z., Li, B., Pang, J., Zhang, J., & Zheng, X. (2020). Risk factors of severe cases with COVID-19: a meta-analysis. *Epidemiology and Infection*, *148*, e175. https://doi.org/10.1017/S095026882000179X

Parasa, S., Desai, M., Thoguluva Chandrasekar, V., Patel, H. K., Kennedy, K. F., Roesch, T., Spadaccini, M., Colombo, M., Gabbiadini, R., Artifon, E. L. A., Repici, A., & Sharma, P. (2020). Prevalence of Gastrointestinal Symptoms and Fecal Viral Shedding in Patients With Coronavirus Disease 2019: A Systematic Review and Meta-analysis. *JAMA Network Open*, *3*(6), e2011335. https://doi.org/10.1001/jamanetworkopen.2020.11335

Paraskevis, E. al. (2020). Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *79*, 104212. https://doi.org/10.1016/j.meegid.2020.104212

Patanè, L., Morotti, D., Giunta, M. R., Sigismondi, C., Piccoli, M. G., Frigerio, L., Mangili, G., Arosio, M., & Cornolti, G. (2020). Vertical transmission of coronavirus disease 2019: severe acute respiratory syndrome coronavirus 2 RNA on the fetal side of the placenta in pregnancies with coronavirus disease 2019-positive mothers and neonates at birth. *American Journal of Obstetrics & Gynecology MFM*, *2*(3), 100145. https://doi.org/10.1016/j.ajogmf.2020.100145

Perera, N., Dehmer, M., & Emmert-Streib, F. (2020). Named Entity Recognition and Relation Detection for Biomedical Information Extraction. *Frontiers in Cell and Developmental Biology*, *8*, 673. https://doi.org/10.3389/fcell.2020.00673

Pišlar, A., Mitrović, A., Sabotič, J., Pečar Fonović, U., Perišić Nanut, M., Jakoš, T., Senjor, E., & Kos, J. (2020). The role of cysteine peptidases in coronavirus cell entry and replication: The therapeutic potential of cathepsin inhibitors. *PLoS Pathogens*, *16*(11), e1009013. https://doi.org/10.1371/journal.ppat.1009013

Polinski, J. M., Weckstein, A. R., Batech, M., Kabelac, C., Kamath, T., Harvey, R., Jain, S., Rassen, J. A., Khan, N., & Schneeweiss, S. (2021). Effectiveness of the single-dose Ad26.COV2.S COVID vaccine. In *bioRxiv*. https://doi.org/10.1101/2021.09.10.21263385

Preto, A. J., Marques-Pereira, C., Baptista, S. J., Bueschbell, B., Barreto, C. A. V., Gaspar, A. T., Pinheiro, I., Pereira, N., Pires, M., Ramalhão, D., Silvério, D., Rosário-Ferreira, N., Melo, R., Mourão, J., & Moreira, I. S. (2021). Targeting GPCRs Via Multi-Platforms Arrays and AI. In *Reference Module in Biomedical Sciences*. Elsevier. https://doi.org/10.1016/b978-0-12-820472-6.00048-7

*PRKCE protein kinase C epsilon [Homo sapiens (human)] - Gene - NCBI*. (n.d.). Retrieved October 30, 2021, from https://www.ncbi.nlm.nih.gov/gene/5581

*Protein kinase C epsilon type*. (n.d.). Retrieved October 30, 2021, from https://www.uniprot.org/uniprot/Q02156

*Protein-tyrosine kinase 2-beta*. (n.d.). Retrieved October 30, 2021, from https://www.uniprot.org/uniprot/Q14289

*PubMed*. (n.d.). Retrieved October 24, 2021, from https://pubmed.ncbi.nlm.nih.gov

Puppala, M., He, T., Chen, S., Ogunti, R., Yu, X., Li, F., Jackson, R., & Wong, S. T. C. (2015). METEOR: An enterprise health informatics environment to support evidence-based medicine. *IEEE Transactions on Bio-Medical Engineering*, *62*(12), 2776–2786. https://doi.org/10.1109/TBME.2015.2450181

Qiu, L., Liu, X., Xiao, M., Xie, J., Cao, W., Liu, Z., Morse, A., Xie, Y., Li, T., & Zhu, L. (2020). SARS-CoV-2 Is Not Detectable in the Vaginal Fluid of Women With Severe COVID-19 Infection. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *71*(15), 813–817. https://doi.org/10.1093/cid/ciaa375

Rabaan, A. A., Al-Ahmed, S. H., Haque, S., Sah, R., Tiwari, R., Malik, Y. S., Dhama, K., Yatoo, M. I., Bonilla-Aldana, D. K., & Rodriguez-Morales, A. J. (2020). SARS-CoV-2, SARS-CoV, and MERS-COV: A comparative overview. *Le Infezioni in Medicina: Rivista Periodica Di Eziologia, Epidemiologia, Diagnostica, Clinica e Terapia Delle Patologie Infettive*, *28*(2), 174–184. https://www.ncbi.nlm.nih.gov/pubmed/32275259

Rahman, A., Tabassum, T., Araf, Y., Al Nahid, A., Ullah, M. A., & Hosen, M. J. (2021). Silent hypoxia in COVID-19: pathomechanism and possible management strategy. *Molecular Biology Reports*, *48*(4), 3863–3869. https://doi.org/10.1007/s11033-021-06358-1

Rauch, S., Roth, N., Schwendt, K., Fotin-Mleczek, M., Mueller, S. O., & Petsch, B. (2021). mRNA-based SARS-CoV-2 vaccine candidate CVnCoV induces high levels of virus-neutralising antibodies and mediates protection in rodents. *Npj Vaccines*, *6*(1), 57. https://doi.org/10.1038/s41541-021-00311-w

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198. https://doi.org/10.1093/nar/gkz369

Ren, Y., Wang, A., Fang, Y., Shu, T., Wu, D., Wang, C., Huang, M., Min, J., Jin, L., Zhou, W., Qiu, Y., & Zhou, X. (2021). SARS-CoV-2 membrane glycoprotein M triggers apoptosis with the assistance of nucleocapsid protein N in cells. *Frontiers in Cellular and Infection Microbiology*, *11*, 706252. https://doi.org/10.3389/fcimb.2021.706252

Richard, M., Kok, A., de Meulder, D., Bestebroer, T. M., Lamers, M. M., Okba, N. M. A., Fentener van Vlissingen, M., Rockx, B., Haagmans, B. L., Koopmans, M. P. G., Fouchier, R. A. M., & Herfst, S. (2020). SARS-CoV-2 is transmitted via contact and via the air between ferrets. *Nature Communications*, *11*(1), 3496. https://doi.org/10.1038/s41467-020-17367-2

Rimanshee Arya , Mukesh Kumar. (2021). Structural insights into SARS-CoV-2 proteins. *Journal of Molecular Biology*, *433*(2), 166725. https://doi.org/10.1016/j.jmb.2020.11.024

Rosário-Ferreira, N., Marques-Pereira, C., Pires, M., Ramalhão, D., Pereira, N., Guimarães, V., Santos Costa, V., & Moreira, I. S. (2021a). The treasury chest of text mining: Piling available resources for powerful biomedical text mining. *BioChem*, *1*(2), 60–80. https://doi.org/10.3390/biochem1020007

Rosário-Ferreira, N., Marques-Pereira, C., Pires, M., Ramalhão, D., Pereira, N., Guimarães, V., Santos Costa, V., & Moreira, I. S. (2021b). The treasury chest of text mining: Piling available resources for powerful biomedical text mining. *BioChem*, *1*(2), 60–80. https://doi.org/10.3390/biochem1020007

Saffer, J. D., & Burnett, V. L. (2014). Introduction to biomedical literature text mining: context and objectives. *Methods in Molecular Biology* , *1159*, 1–7. https://doi.org/10.1007/978-1-4939-0709-0_1

Santamaria, S. (2021). Targeting the PI3K/AKT pathway: a potential new weapon in the global fight against SARS-CoV-2? [Review of *Targeting the PI3K/AKT pathway: a potential new weapon in the global fight against SARS-CoV-2?*]. *International Journal of Biological Sciences*, *17*(11), 2770–2771. Ivyspring International Publisher. https://doi.org/10.7150/ijbs.63969

*SARS-CoV-2 infects and replicates in the salivary glands, study finds*. (2021, July 6). https://www.news-medical.net/news/20210706/SARS-CoV-2-infects-and-replicates-in-the-salivary-glands-study-finds.aspx

Shenoy, N., Luchtel, R., & Gulani, P. (2020). Considerations for target oxygen saturation in COVID-19 patients: are we under-shooting? *BMC Medicine*, *18*(1), 260. https://doi.org/10.1186/s12916-020-01735-2

Shi, J., Wen, Z., Zhong, G., Yang, H., Wang, C., Huang, B., Liu, R., He, X., Shuai, L., Sun, Z., Zhao, Y., Liu, P., Liang, L., Cui, P., Wang, J., Zhang, X., Guan, Y., Tan, W., Wu, G., … Bu, Z. (2020). Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science*, *368*(6494), 1016–1020. https://doi.org/10.1126/science.abb7015

Shorten, C., Khoshgoftaar, T. M., & Furht, B. (2021). Deep Learning applications for COVID-19. *Journal of Big Data*, *8*(1), 18. https://doi.org/10.1186/s40537-020-00392-9

Singh, S., & Khan, A. (2020). Clinical Characteristics and Outcomes of Coronavirus Disease 2019 Among Patients With Preexisting Liver Disease in the United States: A Multicenter Research Network Study. *Gastroenterology*, *159*(2), 768-771.e3. https://doi.org/10.1053/j.gastro.2020.04.064

Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I., & Stagljar, I. (2015). Fundamentals of protein interaction network mapping. *Molecular Systems Biology*, *11*(12), 848. https://doi.org/10.15252/msb.20156351

Somanath, P. R. (2020). Is targeting Akt a viable option to treat advanced-stage COVID-19 patients? *American Journal of Physiology. Lung Cellular and Molecular Physiology*, *319*(1), L45–L47. https://doi.org/10.1152/ajplung.00124.2020

Stoney, R., Robertson, D. L., Nenadic, G., & Schwartz, J.-M. (2018). Mapping biological process relationships and disease perturbations within a pathway network. *Npj Systems Biology and Applications*, *4*(1), 22. https://doi.org/10.1038/s41540-018-0055-2

Straus, M. R., Tang, T., & Al, L. (2020). Ca2þ ions promote fusion of Middle East Respiratory Syndrome coronavirus with host cells and increase in- fectivity. *J Virol*, *94:e00426–20*.

Strauss, J. H., & Strauss, E. G. (2008). Overview of viruses and virus infection. In *Viruses and Human Disease* (pp. 1–33). Elsevier. https://doi.org/10.1016/b978-0-12-373741-0.50004-0

Stuart, J. M., Segal, E., Koller, D., & Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science (New York, N.Y.)*, *302*(5643), 249–255. https://doi.org/10.1126/science.1087447

Sun, F., Mu, C., Kwok, H. F., Xu, J., Wu, Y., Liu, W., Sabatier, J.-M., Annweiler, C., Li, X., Cao, Z., & Xie, Y. (2021). Capivasertib restricts SARS-CoV-2 cellular entry: a potential clinical application for COVID-19. *International Journal of Biological Sciences*, *17*(9), 2348–2355. https://doi.org/10.7150/ijbs.57810

Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., & Wang, G. (2018). Data Processing and Text Mining Technologies on Electronic Medical Records: A Review. *Journal of Healthcare Engineering*, *2018*, 4302425. https://doi.org/10.1155/2018/4302425

Sun, Y. V. (2012). Integration of biological networks and pathways with genetic association studies. *Human Genetics*, *131*(10), 1677–1686. https://doi.org/10.1007/s00439-012-1198-7

Suthram, S., Dudley, J. T., Chiang, A. P., Chen, R., Hastie, T. J., & Butte, A. J. (2010). Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Computational Biology*, *6*(2), e1000662. https://doi.org/10.1371/journal.pcbi.1000662

Szklarczyk, D., Santos, A., von Mering, C., Jensen, L. J., Bork, P., & Kuhn, M. (2016). STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Research*, *44*(D1), D380-4. https://doi.org/10.1093/nar/gkv1277

*Taxonomy*. (n.d.). Retrieved May 29, 2021, from https://talk.ictvonline.org/taxonomy/

Teixeira da Silva, J. A., Tsigaris, P., & Erfanmanesh, M. (2020). Publishing volumes in major databases related to Covid-19. *Scientometrics*, 1–12. https://doi.org/10.1007/s11192-020-03675-3

Theken, K. N., Tang, S. Y., Sengupta, S., & FitzGerald, G. A. (2021). The roles of lipids in SARS-CoV-2 viral replication and the host immune response. *The Journal of Lipid Research*, *62*(100129), 100129. https://doi.org/10.1016/j.jlr.2021.100129

Thomas, S. (2020). The structure of the membrane protein of SARS-CoV-2 resembles the sugar transporter semiSWEET. In *Preprints*. https://doi.org/10.20944/preprints202004.0512.v6

Thomas, S. J., Moreira, E. D., Jr, Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Pérez Marc, G., Polack, F. P., Zerbini, C., Bailey, R., Swanson, K. A., Xu, X., Roychoudhury, S., Koury, K., Bouguermouh, S., Kalina, W. V., Cooper, D., Frenck, R. W., Jr, … C4591001 Clinical Trial Group. (2021). Six month safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. In *bioRxiv*. medRxiv. https://doi.org/10.1101/2021.07.28.21261159

Tian, J., Yuan, X., Xiao, J., Zhong, Q., Yang, C., Liu, B., Cai, Y., Lu, Z., Wang, J., Wang, Y., Liu, S., Cheng, B., Wang, J., Zhang, M., Wang, L., Niu, S., Yao, Z., Deng, X., Zhou, F., … Wang, Z. (2020). Clinical characteristics and risk factors associated with COVID-19 disease severity in patients with cancer in Wuhan, China: a multicentre, retrospective, cohort study. *The Lancet Oncology*, *21*(7), 893–903. https://doi.org/10.1016/S1470-2045(20)30309-0

Tobin, M. J., Laghi, F., & Jubran, A. (2020). Why COVID-19 silent hypoxemia is baffling to physicians. *American Journal of Respiratory and Critical Care Medicine*, *202*(3), 356–360. https://doi.org/10.1164/rccm.202006-2157CP

Trougakos, I. P., Stamatelopoulos, K., Terpos, E., Tsitsilonis, O. E., Aivalioti, E., Paraskevis, D., Kastritis, E., Pavlakis, G. N., & Dimopoulos, M. A. (2021). Insights to SARS-CoV-2 life cycle, pathophysiology, and rationalized treatments that target COVID-19 clinical complications. *Journal of Biomedical Science*, *28*(1), 1–18. https://doi.org/10.1186/s12929-020-00703-5

Tsoi, H., Li, L., Chen, Z. S., Lau, K.-F., Tsui, S. K. W., & Chan, H. Y. E. (2014). The SARS-coronavirus membrane protein induces apoptosis via interfering with PDK1-PKB/Akt signalling. *The Biochemical Journal*, *464*(3), 439–447. https://doi.org/10.1042/BJ20131461

Turner, M. D., Nedjai, B., Hurst, T., & Pennington, D. J. (2014). Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. *Biochimica et Biophysica Acta*, *1843*(11), 2563–2582. https://doi.org/10.1016/j.bbamcr.2014.05.014

Vabret, N., Britton, G. J., Gruber, C., Hegde, S., Kim, J., Kuksin, M., Levantovsky, R., Malle, L., Moreira, A., Park, M. D., Pia, L., Risson, E., Saffern, M., Salomé, B., Esai Selvan, M., Spindler, M. P., Tan, J., van der Heide, V., Gregory, J. K., … Sinai Immunology Review Project. (2020). Immunology of COVID-19: Current state of the science. *Immunity*, *52*(6), 910–941. https://doi.org/10.1016/j.immuni.2020.05.002

van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., Tamin, A., Harcourt, J. L., Thornburg, N. J., Gerber, S. I., Lloyd-Smith, J. O., de Wit, E., & Munster, V. J. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *The New England Journal of Medicine*, *382*(16), 1564–1567. https://doi.org/10.1056/NEJMc2004973

van Doremalen, N., Lambe, T., Spencer, A., Belij-Rammerstorfer, S., Purushotham, J. N., Port, J. R., Avanzato, V., Bushmaker, T., Flaxman, A., Ulaszewska, M., Feldmann, F., Allen, E. R., Sharpe, H., Schulz, J., Holbrook, M., Okumura, A., Meade-White, K., Pérez-Pérez, L., Bissett, C., … Munster, V. J. (2020). ChAdOx1 nCoV-19 vaccination prevents SARS-CoV-2 pneumonia in rhesus macaques. *BioRxiv.Org: The Preprint Server for Biology*. https://doi.org/10.1101/2020.05.13.093195

Vivanti, A. J., Vauloup-Fellous, C., Prevot, S., Zupan, V., Suffee, C., Do Cao, J., Benachi, A., & De Luca, D. (2020). Transplacental transmission of SARS-CoV-2 infection. *Nature Communications*, *11*(1), 3572. https://doi.org/10.1038/s41467-020-17436-6

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews. Microbiology*, *19*(3), 155–170. https://doi.org/10.1038/s41579-020-00468-6

Volberg, T., Geiger, B., Dror, R., & Zick, Y. (1991). Modulation of intercellular adherens-type junctions and tyrosine phosphorylation of their components in RSV-transformed cultured chick lens cells. *Cell Regulation*, *2*(2), 105–120. https://doi.org/10.1091/mbc.2.2.105

Vries, S. J., & Bonvin, A. M. J. J. J. J. (2011). Cport: A consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS One*, *6*, 17695.

Wang, A. L., Iadecola, C., & Wang, G. (2017). New generations of dihydropyridines for treatment of hypertension. *Journal of Geriatric Cardiology: JGC*, *14*(1), 67–72. https://doi.org/10.11909/j.issn.1671-5411.2017.01.006

Wei, C.-H., Kao, H.-Y., & Lu, Z. (2013). PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Research*, *41*(Web Server issue), W518-22. https://doi.org/10.1093/nar/gkt441

Weiss, S. R., & Leibowitz, J. L. (2011). Coronavirus pathogenesis. *Advances in Virus Research*, *81*, 85–164. https://doi.org/10.1016/B978-0-12-385885-6.00009-2

Weston, S., & Frieman, M. B. (2018). Respiratory Viruses. In *Reference Module in Biomedical Sciences* (pp. 85–101). Elsevier. https://doi.org/10.1016/b978-0-12-801238-3.66161-5

*WHO Coronavirus (COVID-19) Dashboard*. (n.d.). Retrieved October 24, 2021, from https://covid19.who.int

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., … Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, *46*(D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.Y.)*, *367*(6483), 1260–1263. https://doi.org/10.1126/science.abb2507

Wu, Z., & McGoogan, J. M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA: The Journal of the American Medical Association*, *323*(13), 1239–1242. https://doi.org/10.1001/jama.2020.2648

Xing, R., Luo, J., & Song, T. (2020). BioRel: towards large-scale biomedical relation extraction. *BMC Bioinformatics*, *21*(Suppl 16), 543. https://doi.org/10.1186/s12859-020-03889-5

Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics (Oxford, England)*, *32*(23), 3676–3678. https://doi.org/10.1093/bioinformatics/btw514

Yadav, S., Ramesh, S., Saha, S., & Ekbal, A. (2020). Relation Extraction from Biomedical and Clinical Text: Unified Multitask Learning Framework. *IEEE/ACM Transactions on Computational Biology and Bioinformatics / IEEE, ACM*, *PP*. https://doi.org/10.1109/TCBB.2020.3020016

Yang, L., Xie, X., Tu, Z., Fu, J., Xu, D., & Zhou, Y. (2021). The signal pathways and treatment of cytokine storm in COVID-19. *Signal Transduction and Targeted Therapy*, *6*(1), 255. https://doi.org/10.1038/s41392-021-00679-0

Yang, Z., & Liu, Y. (2020). Vertical Transmission of Severe Acute Respiratory Syndrome Coronavirus 2: A Systematic Review. *American Journal of Perinatology*, *37*(10), 1055–1060. https://doi.org/10.1055/s-0040-1712161

Yao X. H., Li T. Y., He Z. C., Ping Y. F., Liu H. W., Yu S. C., Mou H. M., Wang L. H., Zhang H. R., Fu W. J., Luo T., Liu F., Guo Q. N., Chen C., Xiao H. L., Guo H. T., Lin S., Xiang D. F., Shi Y., …

Bian X. W. (2020). [A pathological report of three COVID-19 cases by minimal invasive autopsies]. *Zhonghua bing li xue za zhi Chinese journal of pathology*, *49*(5), 411–417. https://doi.org/10.3760/cma.j.cn112151-20200312-00193

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Zhang, L.-K., Sun, Y., Zeng, H., Wang, Q., Jiang, X., Shang, W.-J., Wu, Y., Li, S., Zhang, Y.-L., Hao, Z.-N., Chen, H., Jin, R., Liu, W., Li, H., Peng, K., & Xiao, G. (2020). Calcium channel blocker amlodipine besylate therapy is associated with reduced case fatality rate of COVID-19 patients with hypertension. *Cell Discovery*, *6*(1), 96. https://doi.org/10.1038/s41421-020-00235-0

Zhang, Y., & Lu, Z. (2019). Exploring semi-supervised variational autoencoders for biomedical relation extraction. *Methods* , *166*, 112–119. https://doi.org/10.1016/j.ymeth.2019.02.021

Zheng, S., Dharssi, S., Wu, M., Li, J., & Lu, Z. (2019). Text Mining for Drug Discovery. *Methods in Molecular Biology* , *1939*, 231–252. https://doi.org/10.1007/978-1-4939-9089-4_13

Zhu, F., Patumcharoenpol, P., Zhang, C., Yang, Y., Chan, J., Meechai, A., Vongsangnak, W., & Shen, B. (2013). Biomedical text mining and its applications in cancer research. *Journal of Biomedical Informatics*, *46*(2), 200–211. https://doi.org/10.1016/j.jbi.2012.10.007

Zundert, G. C. P. P. (2016). The HADDOCK2. 2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol*, *428*, 720–725.