Investigating the multifaceted host contribution to COVID-19 disease risk, progression and treatment: an integrative multi-omics network-based approach study

Francis Edem Agamah

Supervised by: Profs. Darren P. Martin, University of Cape Town, Emile R. Chimusa, Northumbria University, and Peter A.C. 't Hoen, Radboud University Medical Center
Co-supervised by: Drs. Michelle Skelton, University of Cape Town, and Thomas H.A. Ederveen, Radboud University Medical Center

February 2024

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the Department of Integrative Biomedical Sciences, Faculty of Health Sciences, University of Cape Town

The research presented in this thesis was conducted at the University of Cape Town, Department of Integrative Biomedical Sciences, Computational Biology Division, Cape Town, South Africa.

Declaration

I, Francis Edem Agamah, hereby declare that the work on which this thesis is based is my original work both in concept and execution (except where acknowledgments indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorize the University of Cape Town to reproduce for research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate Date: February 6, 2024

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publications in my PhD thesis, and my co-authors have agreed that I may include the publications:

1. **Agamah, F.E.**, Bayjanov, J.R., Niehues, A., Njoku, K.F., Skelton, M., Mazandu, G.K., Ederveen, T.H., Mulder, N., Chimusa, E.R. and t' Hoen, P.A., 2022. Computational approaches for network-based integrative multi-omics analysis. *Frontiers in Molecular Biosciences*, *9*, p.1214.
   doi: https://doi.org/10.3389/fmolb.2022.967205

2. **Agamah, F.E.,** Ederveen, T.H., Skelton, M., Martin, D.P., Chimusa, E.R. and 't Hoen, P.A., 2024. Network-based integrative multi-omics approach reveals biosignatures specific to COVID-19 disease phases. *Frontiers in Molecular Biosciences.*. **doi:** 10.3389/fmolb.2024.1393240

3. **Agamah, F.E.,** Ederveen, T.H., Skelton, M., Martin, D.P., Chimusa, E.R. and 't Hoen, P.A., 2024. Network-based multi-omics-disease-drug associations reveal candidate repurposable drugs for COVID-19 disease phases. *Drug Repurposing Journal.* ***doi****:* 10.58647/DRUGREPO.24.1.0007

Signature:  Date: February 6, 2024

Student Name: Francis Edem Agamah Student Number: AGMFRA001

Acknowledgments

I could not have produced this work without the help of many people. First I thank God for seeing me through to this point.

I am very grateful to all those who contributed to my PhD supervision. Heartfelt thanks to Professor Darren Martin, Dr. Michelle Skelton, and Professor Nicola Mulder from the University of Cape Town, Professor Peter A.C. t' Hoen and Dr. Thomas H.A. Ederveen from the Radboud University Medical Center, and Professor Emile R. Chimusa from the Northumbria University. I am grateful for their time, guidance, invaluable advice, and encouragement throughout the PhD fellowship journey.

Special thanks to Professor Denver Hendricks, Head of Department for Integrative Biomedical Sciences, for his efforts particularly during my transition from the Department of Pathology to Integrative Biomedical Sciences.

I would like to take this opportunity to express my gratitude for all the financial support I received from the Trusted World of Corona (TWOC) consortium, Human Heredity, and Health in Africa (H3Africa) Coordinating Center, and Harnessing Data Science for Health and Innovation (DS-I Africa) Coordinating Center throughout my PhD journey.

I am truly grateful to my mentor, Dr. Anita Ghansah, for her insightful academic guidance and constructive suggestions that helped me build the critical thinking skills that benefit my future career.

I also express my gratitude to all those with whom I co-authored the different research articles.

Without data providers, this work would not have been completed, I would like to thank researchers that made their data publicly available for me to use in my research.

I am grateful to all conference organizers who supported my participation in various international conferences and workshops to present parts of this work; I would like to name; EMBL-EBI Mathematics of Life: modelling molecular mechanisms 2022, 14th International Congress of Human Genetics (ICHG 2023), and Deep Learning Indaba 2023.

I am grateful to colleagues from Radboudumc, including Casper de Visser and Junda Huang for their kind contribution to the workflow containerization.

I also acknowledge Ilifu (https://www.ilifu.ac.za/, South Africa) and High-Performance Computing from CHPC (https://www.chpc.ac.za/) for providing cloud computing facilities for computations performed.

Last but not least, my deep heartfelt gratitude goes to my parents, Prosper Quame Agamah and Agnes Blay-Morkeh, my brother, James Dela Agamah, my sister, Francisca Esenam Agamah, and my wife, Ama Atta-Aidoo, whose continuous love and unconditional support encourage me to be myself.

List of Publications

1. Publications included as part of this thesis

The following papers are from the contributions of this thesis.

1. **Chapter 2: Agamah, F.E.**, Bayjanov, J.R., Niehues, A., Njoku, K.F., Skelton, M., Mazandu, G.K., Ederveen, T.H., Mulder, N., Chimusa, E.R. and t' Hoen, P.A., 2022. Computational approaches for network-based integrative multi-omics analysis. *Frontiers in Molecular Biosciences*, *9*, p.1214.
   doi: https://doi.org/10.3389/fmolb.2022.967205

2. **Chapter 3: Agamah, F.E.,** Ederveen, T.H., Skelton, M., Martin, D.P., Chimusa, E.R. and 't Hoen, P.A., 2024. Network-based integrative multi-omics approach reveals biosignatures specific to COVID-19 disease phases. *Frontiers in Molecular Biosciences.* **doi:** 10.3389/fmolb.2024.1393240**doi:** 10.3389/fmol

3. **Chapter 4: Agamah, F.E.,** Ederveen, T.H., Skelton, M., Martin, D.P., Chimusa, E.R. and 't Hoen, P.A., 2024. Network-based multi-omics-disease-drug associations reveal candidate repurposable drugs for COVID-19 disease phases. *Drug Repurposing Journal.* ***doi:*** 10.58647/DRUGREPO.24.1.0007.

2. Publications from collaborative efforts during PhD project

The following papers are from collaborative efforts within the group and networks during this Ph.D study.

1. **Agamah, F.E.**, Damena, D., Skelton, M., Ghansah, A., Mazandu, G.K. and Chimusa, E.R., 2021. Network-driven analysis of human–Plasmodium falciparum interactome: processes for malaria drug discovery and extracting in silico targets. Malaria Journal, 20(1), pp.1-20.
   doi: https://doi.org/10.1186/s12936-021-03955-0

2. Schultz, B., **Agamah, F.E.**, Ewuoso, C., Madden, E.B., Troyer, J., Skelton, M. and Mwaka, E., 2023. Webinar report: stakeholder perspectives on informed

consent for the use of genomic data by commercial entities. *Journal of Medical Ethics*. **doi:** https://doi.org/10.1136/jme-2022-108650

Table of Contents

List of Figures

driven seeds. The graph reveals distinct subnetworks formed by hubs CCL2, CCL4, and NELFCD demonstrating extensive interactions with drug candidates, seed nodes (IL-6 and IL-6R), and other molecular features (B) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), and biological process (pink nodes) as observed from predicting candidate drugs using existing knowledge graphs, moderate disease-state specific omics-graphs, and hypothesis-driven seeds. The graph reveals distinct subnetworks formed by hubs NFKB1, IL-10, and NELFCD demonstrating extensive interactions with drug candidates, seed nodes (IL-6 and IL-6R), and other molecular features (C) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), and biological process (pink nodes) as observed from predicting candidate drugs using existing knowledge graphs, severe disease-state specific omics-graphs, and hypothesis-driven seeds. The graph reveals distinct subnetworks formed by hubs CXCL1, CCL4, and JAK2 demonstrating extensive interactions with drug candidates, seed nodes (IL-6 and IL-6R), and other molecular features. Yellow edges represent drug-protein and drug-transcript pairwise interactions. Red edges represent biological process-protein interactions and biological process-transcripts interactions. Green edges represent protein-protein interactions. Black edges represent transcript-transcript interactions and protein-transcript interactions. Blue edges represent drug-drug interactions. Light blue edges represent biological processes-biological process interactions and biological process-pathway interactions. The graphs were generated by defining filtering criteria based on node degree between 4 and 1633 inclusive from cytoscape. 142

Figure 4.4. (A) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), lipids (blue nodes), and biological process and pathways (pink nodes) as observed from predicting candidate drugs using existing knowledge graphs, mild disease-state specific omics-graphs, and data-driven seeds. The graph reveals distinct subnetworks formed by hub CCL4, demonstrating extensive interactions with drug candidates and other molecular features including seed nodes (STAT1 and SOD2) (B) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), lipids (blue nodes), and biological processes and pathways (pink nodes) as observed from predicting candidate drugs using existing

List of Tables

Glossary

Common terms universal to this thesis are explained in this section, to streamline terminology.

Biosignature	An omics feature, that shows that a particular biological process or condition is present
Biological network	A graph representation of biological entities as nodes and their interactions as edges.
Genomics	The study of the structure, function, evolution, and mapping of genomes.
Lipid	Organic compounds including fats, oils, hormones, and certain components of membranes
Lipidome	The lipidome is the complete lipid profile within a cell, tissue or organism, and is a subset of the metabolome.
Lipidomics	The study or comprehensive analysis of lipidome(s)
Metabolite	The products and intermediates of cellular metabolism
Metabolome	The collection of all metabolite profile(s) in any given organism, cell, tissue, or organ which are the end products of cellular processes.
Metabolomics	The study or comprehensive analysis of metabolome(s)
Multi-omics	Refers to a systems biology investigative approach involving multiple omics data types, such as genomics, epigenomics, transcriptomics, proteomics, and metabolomics.
Omics	Refers to a field of study in biological sciences that ends with -*omics*, such as genomics, transcriptomics, proteomics, or metabolomics
Proteome	The sum of all expressed proteins in a cell, organ, or organism
Proteomics	The study or comprehensive analysis of proteome(s) in relation to their biochemical properties and functional roles, and how their expressions, modifications, and structures change during growth and in response to internal and external stimuli

Transcript	A specific RNA molecule produced from the DNA template of a gene.
Transcriptome	The collection of messenger RNA molecules (mRNA) in one microorganism, cell, or collection of cells
Transcriptomics	The study or comprehensive analysis of transcriptome(s)

Abstract

Coronavirus Disease-2019 (COVID-19) is a contagious respiratory disorder caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), a newly emerged β coronavirus belonging to the *Coronaviridae* family. Since its discovery in Wuhan, China, in December 2019, COVID-19 disease has transformed into a devastating global pandemic that has created disruptions across healthcare, economic, and social systems with approximately seven hundred and seventy million reported cases and close to seven million reported deaths as of January 2024. The clinical presentation of COVID-19 is very heterogeneous, ranging from mild disease states to severe disease states, associated with varying transcription, protein expression, lipid synthesis, and metabolic profiles. As a result of the heterogeneity of COVID-19 disease, the efficacy of drugs used to treat it may vary depending on the disease states when the medicines are administered.

Molecular biology studies investigating COVID-19 clinical heterogeneity and the drugs that might be used to treat the disease have varied in terms of the approaches implemented. These approaches have focused on single features of SARS-CoV-2 infected cells such as gene transcription levels or protein expression levels (so-called “single omics” approaches) to approaches that attempt to simultaneously examine multiple molecular features of infected cells (so-called “multi-omics approaches”). This project adopts the latter approach, implementing a network-based method that integrates multi-omics data and drug-related data to investigate the contribution of host physiology to COVID-19 disease progression and identify promising drug repurposing candidates as potential treatment options. This involved evaluating existing computational integrative multi-omics network-based methods, determining their strengths and limitations, and adapting them to fit the aims of this project: identifying and characterizing biosignatures associated with various COVID-19 disease phases, as well as identifying drug repurposing candidates tailored for mild, moderate, and severe COVID-19 disease phases.

The World Health Organization (WHO) Ordinal Scale (WOS) was used as a disease severity reference to harmonize COVID-19 patient metadata across two studies. A unified COVID-19 knowledge graph was constructed by assembling a disease-specific

interactome from the literature and databases. We leveraged a multi-omics network-based approach to construct disease-state and omics-specific graphs by integrating proteomics, transcriptomics, metabolomics, and lipidomics data with the unified COVID-19 knowledge graph. We used an adapted random walk with restart algorithm, called multiXrank, to explore the disease-state and omics-specific graphs, and drug data to search for and prioritize not only biosignatures associated with COVID-19 disease phases but also drug candidates with the potential for treating mild, moderate, and severe COVID-19.

The network analysis identified critical biosignatures for each COVID-19 phase. Mild cases displayed unique signatures like CCL4 and IRF1, potentially driving chemotaxis and interferon signaling. The moderate phase was characterized by biosignatures like HGF, MMP12, IL-10, and NFKB1, implicating enhanced inflammation, matrix remodeling, and immune regulation. In severe cases, biosignatures such as lysophosphatidylcholines, diglyceride, and sphingomyelin appeared, suggesting profound tissue damage, dysregulated lipid metabolism, and disrupted repair pathways. As expected, the abundance of shared chemokine and cytokine biosignatures in severe and moderate COVID-19 disease phases as compared to either mild vs moderate or mild vs severe disease phases suggests a closer molecular relatedness between these phases. This finding, along with biosignatures that discriminate between the disease states, and interactions between biosignatures that are either common between or associated with COVID-19 disease phases sheds light on the nuanced progression of the illness.

We further investigated the differential influence of interleukin-6 (IL6) and interleukin-6 receptor *(*IL6R*)* on disease progression. We found that IL6 interaction with features of different omics data types increased with disease severity, thus indicating the differential association of IL6 with the different disease states. More specifically, IL6 interaction with proteins (e.g., IFNB, IFIT3), transcripts (e.g., CXCL1, CXCL2, CCL3), and metabolites (e.g., 1-(1-enyl-palmitoyl)-GPC, 1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)) may contribute to its major role in disease severity. We also observed that IL6R interactions mainly with proteins and transcripts increase more clearly with disease severity than do interactions with metabolites and lipids. We present a multilayered visualization tool (hosted at http://cytoscape.h3africa.org, last accessed

on February 6, 2024) to navigate and analyze complex interactions across different biological layers, offering a valuable resource for uncovering key drivers of disease severity. Interestingly, cross-layer interactions between different omics profiles increased with disease severity. These potential association patterns could be useful for providing insights into the underlying molecular causes and consequences of the clinical heterogeneity of COVID-19, enabling early disease diagnosis, and optimal treatment prediction.

The network-based integration of drug data and multi-omics data assisted in drug prediction. Most importantly, we prioritized twenty Food and Drug Administration-approved agents with potential utility for mild, moderate, and severe COVID-19 disease phases. For mild COVID-19, stimulating immune cell recruitment and activation is key. Drugs like histamine, curcumin, and paclitaxel show potential in this regard due to their ability to stimulate immune cell recruitment, potentially mitigating disease progression. Similarly, non-steroidal anti-inflammatory drugs like indomethacin and diclofenac may offer symptomatic relief in mild cases. In mild to moderate COVID-19, drugs like omacetaxine, crizotinib, and vorinostat, known for their antiviral properties, can potentially hinder viral replication and offer additional treatment options. Moreover, glutathione, a potent antioxidant, could be valuable in moderate cases due to its potential to counteract inflammation and potentially prevent the dangerous "cytokine storm" seen in patients with antioxidant deficiencies. In severe COVID-19, the excessive immune response triggers a dangerous "inflammatory cascade." To combat this, drugs with strong anti-inflammatory effects, including anti-inflammatory drugs (sarilumab, tocilizumab), corticosteroids (dexamethasone, hydrocortisone), and immunosuppressives (sirolimus, cyclosporine), emerge as potential candidates for controlling this harmful process.
Moreover, we further explore the interactions involving the drug repurposing candidates and key biosignatures. The findings could be useful for personalized treatment options tailored to individual patients based on their disease severity level.

This project identified both biosignatures of different omics types (proteins, transcripts, metabolites, and lipids) enriched in disease-state pathways and their associated interactions that are either common between, or unique to mild, moderate, and severe COVID-19. These biosignatures include molecular features that underlie the observed

clinical heterogeneity of COVID-19 and emphasize the need for disease-phase-specific treatment strategies. In addition, we explored the potential of multi-omics and drug-related data to predict therapeutics for different phases of COVID-19. This project demonstrates that the integrative analysis of drug data and multi-omics data enables the prioritization of biosignatures and potential drug candidates for COVID-19 disease phases. These findings hold promise for guiding future experimental studies towards potential clinical applications, requiring further investigation to definitively translate into therapeutic advancements.

Chapter 1

General Introduction and Scope

1.1 Overview of Coronavirus Disease-2019 (COVID-19)

Seven human coronaviruses (HCoVs) have been reported: HCoV-NL63, HCoV-229E, HCoV-HKU1, HCoV-OC43, severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and the most recent SARS-CoV-2. Four of these viruses, including HCoV-NL63, -229E, -HKU1, and -OC43, mostly cause mild-to-moderate respiratory diseases with a seasonal pattern [1].

Coronavirus Disease-2019 (COVID-19) is a contagious respiratory disorder caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), a newly emerged β coronavirus belonging to the *Coronaviridae* family [2-4]. Since its discovery in Wuhan, China, in December 2019, COVID-19 has established itself as one of the most devastating global pandemics creating disruptions across health care, economic, and social systems, as compared to Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome-CoV (MERS-CoV) outbreaks in 2002 and 2012 respectively [5, 6]. Healthcare systems in many parts of the world became quickly overwhelmed by high volumes of infected patients between February and July 2020 during the first infection wave and then again in December 2020 and July 2021 during subsequent waves. To date, COVID-19 has rapidly spread across the globe mainly through human-to-human transmission, with over seven hundred and seventy million confirmed cases and nearly seven million deaths [7].

1.2 The virus: An overview of its genetic architecture

SARS-CoV-2 is an enveloped, non-segmented, and positive-stranded RNA virus [8]. The entire SARS-CoV-2 genome encodes a polyprotein of roughly 7096 residues, including both structural and non-structural proteins (NSPs) [9]. Its genome, ranging from 27 to 30kb in length, contains ten protein-coding Open Reading Frames (ORF) which encode non-structural replicase polyproteins (about 20kb of viral genome) and structural proteins (about 10kb of viral genome) [6, 10]. The genome arrangement (**Figure 1.1**) is such that, the first ORF (ORF1a/b) encodes polyprotein 1a, and

polyprotein 1b. These polyproteins are post-translationally processed to produce 16 nonstructural proteins [9-11]. The remaining ORFs encode the structural proteins which comprise the spike protein (S), an envelope protein (E), a membrane protein (M), a nucleocapsid protein (N), and 9 other accessory proteins (ORF3a, −3b, −6, −7a, −7b, −8, −9b, −9c and −10 proteins) [2, 8, 10, 11].

Relative to other RNA viruses which evolve at a rate of approximately 10^{-3} nucleotide substitutions per site per year [12], coronavirus genomes evolve at a considerably lower rate of approximately 10^{-4} substitutions per site per year [13, 14].

In addition to nucleotide substitutions, genetic recombination, insertions, and deletion events across the genome contribute significantly to the genetic diversity of coronaviruses [13]. This genetic diversity is a crucial requirement for efficient viral adaptation: a process that can, among other things, involve alterations in viral virulence, infectivity, transmissibility, and the evasion of host adaptive immune responses [15, 16].

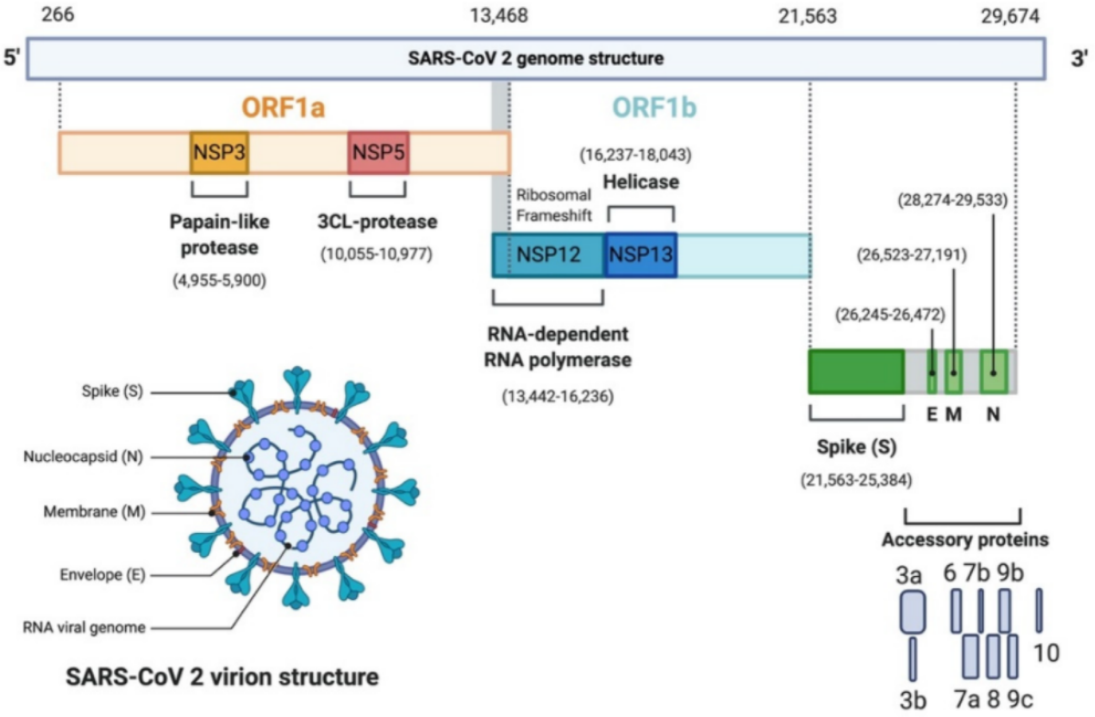


***Figure 1.1**. The genomic organization of SARS-CoV-2. Image retrieved from Alangreh et al.[17].*

1.3 The pathogenesis of COVID-19

The pathogenesis of COVID-19 illness occurs in two distinct phases, an early phase and a late phase [18]. The early phase is characterized by profound viral replication which happens after entry into host cells [18-21]. The mechanisms that underlie the entry of SARS-CoV-2 into human cells are linked with viral S protein which, following priming by host cellular proteases (primarily Transmembrane Serine Protease 2 (TMPRSS2)), binds to host angiotensin-converting enzyme 2 (ACE2) cell surface receptor molecules which, in turn, triggers fusion of the viral envelope with the host-cell membrane. The ACE2 receptors are ubiquitous within the human body, particularly overexpressed on intestinal epithelial cells of the gut, endothelial and smooth cells of the blood vessels, heart (epicardia, adipocytes, fibroblasts, myocytes, coronary arteries), lung (macrophages, bronchial and tracheal epithelial cells, type 2 pneumocytes), brain, testis, and on tubular epithelial cells of the kidney [22]. The late stage of the pathogenesis of COVID-19, which causes severe illness, is characterized not only by a hyperinflammatory state and aggressive inflammation induced by the release of cytokines but also, by collateral tissue damage and systemic failure due to broad organotropism of SARS-CoV-2 [18]. Notable cytokines prominent at the late stage include, but are not limited to, tumor necrosis factor-α (TNF α), granulocyte-macrophage colony-stimulating factor (GM-CSF), Interleukin (IL) 1, IL-6, interferon (IFN)-γ, and activation of the coagulation system resulting in a prothrombotic state.

1.4 COVID-19 disease phases

COVID-19 is characterized by a range of clinical phenotypes that reflect the spectrum of disease severity. Disease phenotypes are broadly classifiable as asymptomatic and symptomatic with about 85% of infected patients characterized as either asymptomatic or showing mild symptoms and about 15% of infected patients suffering from potentially life-threatening complications [23]. The mild to moderate disease phenotype includes disease conditions with few or no infection symptoms, hospitalization with either no oxygen therapy required or with oxygen given by mask or nasal prongs, and no fatalities. In contrast, the severe disease phenotype includes disease conditions that besides death could involve one or a combination of hospitalization, oxygen therapy involving mechanical ventilation, respiratory failure, and significant immune dysregulation (**Table 1.1**).

According to the World Health Organization (WHO) ordinal scale (WOS) for clinical improvement, the state of SARS-CoV-2 infected patient is quantified on a nine-point scale (ranging from 0 to 8) which describes whether a patient has experienced mild, moderate, and severe disease conditions or death.

Other proposed COVID-19 outcome metrics leverage the WOS and/or incorporate additional concepts to classify disease trajectory. Overmyer et al. [21] implemented the concept of Hospital-free days at day 45 (HFD-45), a disease outcome measure that takes into account the length of hospital stays with mortality to describe disease severity. The HFD-45 metric assigns a score of zero to patients who remain admitted longer than 45 days or die while admitted and higher values of HFD-45 to patients with shorter hospitalizations and milder disease severity [21]. Bernardes et al. [19] implemented the concept of disease phase pseudotimes, a scoring metric (ranging from 0 to 7) that takes into account several inflammatory markers (serum c-reactive protein (CRP), serum IL-6, and ferritin) in addition to ventilation needs of patients to describe heterogeneous disease trajectories. The approach implemented by these authors enabled the distinguishing between temporary disease severity and recovery (convalescence) phases. Each pseudotime has a corresponding clinical score which is the sum of WOS and CRP, IL-6, and ferritin-valued-based subscores.

***Table 1.1**. Overview of some general characteristics of COVID-19 disease phases.*

Healthy (WOS = 0)	**Mild (WOS= 1-2)**	**Moderate (WOS= 3-4)**	**Severe/death (WOS =5-8)**	**Reference**
No clinical or virological evidence of infection	• Fever • Non-severe respiratory symptoms	• Hospitalized • No oxygen therapies. • Oxygen by mask or nasal prong • Moderate respiratory symptoms	• Non-invasive ventilation or high-flow oxygen • Incubation and mechanical ventilation • Ventilation and additional organ support • Prolonged deterioration of lymphopenia • High levels of inflammatory cytokines in plasma • Increased neutrophilia • Shortness of breath • Oxygen saturation • viral pneumonia • Fever • Dyspnea • Acute respiratory distress syndrome • Multiple organ failure • Myalgia • Immune cell exhaustion • HLA class II downregulation on monocytes • Elevated interleukin 6 • Abnormal myeloid cells • responses of megakaryocytes, neutrophil, erythroid cells, plasmablasts, platelets • systemic inflammation • drop in blood nutrient • Death	[24-26]

*WOS is the World Health Organization (WHO) ordinal scale

1.5 COVID-19 clinical heterogeneity

The clinical manifestation of COVID-19 varies significantly across infected patients. Within-host diversity of SARS-CoV-2 in infected patients may impact the disease severity [27], its control, and treatment. This phenomenon is explained by the varying prevalence of virus genetic variants within the host, with severely infected patients expressing a higher number of variations in coding and non-coding regions of the viral genome than mild cases [27]. For instance, variants like Delta and Omicron exhibit increased transmissibility and potentially altered virulence, contributing to changes in disease patterns.

The enormous clinical heterogeneity has been reported to be mainly determined by host factors [28-30]. These host factors include but are not limited to, comorbidities, age, gender, and genetics and have a crucial role in disease severity and progression [27, 31]. Specifically, pre-existing conditions like diabetes, obesity, and chronic respiratory diseases exacerbate disease severity and increase complications. Also, elderly populations are disproportionately affected with higher mortality rates, likely due to weakened immune systems and pre-existing comorbidities. Also, males generally exhibit higher hospitalization and mortality rates compared to females, potentially due to hormonal and immune system differences [32-34]. Most importantly, individual genetic variations influencing immune response, viral entry, and inflammatory pathways play a crucial role, however, the mystery behind the differential outcomes of COVID-19 infection is yet to be fully unraveled [35, 36].

Several studies have implicated variants in the Human Leukocyte Antigen (HLA) gene cluster, responsible for immune antigen presentation, with COVID-19 severity [37]. Specifically, alleles like HLA-B57:01, HLA-B*15:01, HLA-A*32, HLA-B*58, and HLA-C12:03 are associated with reduced severity [38, 39], while others like HLA-DRB1*07:01, HLA-DRB1*13:02 and HLA-DRB1*04:01 [40, 41], are linked to increased risk. Also, several studies have suggested a potential association between Toll-like receptor 7 (TLR7) mutations and increased susceptibility to severe outcomes in COVID-19 cases [30]. These mutations may compromise an individual's ability to mount a robust antiviral response, leading to more severe manifestations of the disease.

Also, variations in both an interferon receptor gene (IFNRA2), particularly rs9975538 variant, and a gene region near TYK2, a key player in signalling pathways essential for interferon, interleukin-12 and interleukin-23 are linked with disease severity [42]. These signalling molecules are critical for activating T helper 1 and T helper 17 cells, which are crucial for fighting off virus infections. Additionally, rs1799964 in the promoter of inflammation-related gene TNF, rs1886814 in the intron of FOXP4-AS1, and rs429358 in the Apolipoprotein E (APOE) gene have been associated with COVID-19 severity or hospitalization [43].

In general, systemic inflammation, drops in blood nutrient concentrations, responses of various immune cell types (including megakaryocytes, neutrophil, erythroid cells, plasmablasts, and platelets), distinct metabolic signatures characterizing disease progression, and immune dysregulation (hyperactive immune response) have all been identified as major hallmarks of disease severity [25, 44-47].

1.6 COVID-19 Treatment

An intense global research effort has resulted in substantial advancements that have accelerated the development of a variety of innovative therapeutics and vaccines for managing and improving patient outcomes.

Various vaccines (**Table 1.2**), including those developed (by Pfizer-BioNTech, Moderna, and AstraZeneca (in collaboration with Oxford University) have been approved for use in immunization to prevent COVID-19 by the European Medicines Agency (EMA) and the United States Food and Drug Administration.

Multiple existing drugs have been sort to treat or control infection. For instance, evidence suggests the role of remdesivir (Veklury) (DrugBank: DB14761), an adenosine triphosphate analogue, in shortening the recovery time for adults hospitalized with COVID-19 infection and pneumonia [48]. Dexamethasone (DrugBank: DB01234), a host-directed anti-inflammatory corticosteroid, has also proven effective in reducing mortality among severely infected patients [49]. Additionally, antivirals such as favipiravir (DrugBank: DB12466) and molnupiravir are known to improve patient's health. Further, studies have shown that aspirin

(DrugBank: DB00945) [50], and sarilumab, an IL-6R inhibitor [51], significantly reduce the risk of complications and death by COVID-19 in hospitalized infected patients.

Also, some monoclonal antibodies (**Table 1.3**) have been issued by the Food and Drug Administration (FDA) for emergency use to manage COVID-19. The monoclonal antibodies prevent the viral attachment to the host's ACE2 receptors thus, preventing viral replication in the host. Tocilizumab (DrugBank: DB06273) monotherapy has been authorized by the FDA for emergency use in the treatment of hospitalized adults and peadiatric patients (2 years of age and older) who are receiving systemic corticosteroids and require supplemental oxygen, non-invasive or invasive mechanical ventilation, or extracorporeal membrane oxygenation (ECMO) [52]. Other monoclonal antibodies including bebtelovimab, casirivimab and imdevimab (CAS/IMDEV) combination, Bamlanivimab and Etesevimab (BAM/ETE) combination, and Sotrovimab monotherapy have been issued by FDA for emergency use in the management of COVID-19 among mild to moderate infected adults. Furthermore, a variety of other anti-COVID-19 vaccines, monoclonal antibodies, and drugs are currently under clinical investigation (https://www.pharmgkb.org/page/COVID, and https://clinicaltrials.gov/ct2/results?cond=COVID-19, last accessed in January 2023)

Table 1.2. *Summary of currently approved COVID-19 vaccines for emergency use.*

Vaccine	Vaccine type	Recommended age group	Effectiveness to Disease phase	Side effects/ potential risk	Reference
Pfizer-BioNTech COVID-19 Vaccine (COMIRNATY®(BNT162b2))	mRNA	People 12 years and older	Mild, moderate, and severe	allergic reaction, tiredness, headache, muscle pain, chills, fever, and nausea	[53]
Moderna COVID-19 Vaccine (Spikevax (mRNA-1273))	mRNA	People 12 years and older	Mild, moderate, and severe	myocarditis, pericarditis, allergic reaction, tiredness, headache, muscle pain, chills, fever, nausea	[53]
Johnson & Johnson's Janssen COVID-19 Vaccine (Ad26.COV2.S)	Viral Vector	People 18 years and older	Mild, moderate, and severe	blood clots with low platelets, allergic reaction, tiredness, headache, muscle pain, chills, fever, nausea	[53]
Serum Institute of India Pvt. Ltd COVID-19 Vaccine (Covishield (ChAdOx1_nCoV-19))	Viral Vector	People 18 years and older	Mild, moderate, and severe	Altered sensorium, weakness in limbs, altered coloured urine, enlarged lymph nodes, headache, muscle pain, chills, fever, nausea	[53-55]
Oxford/AstraZeneca COVID-19 vaccine (Vaxzevria/AZD1222 (ChAdOx1 nCoV-19))	Viral Vector	People 18 years and older	Severe (including Persons with COVID-19-related comorbidities)	allergic reaction, tiredness, headache, muscle pain, chills, fever, nausea	[53, 56]

The Gamaleya national center COVID-19 vaccine (Sputnik V)	Viral Vector	People 18 years and older	Mild, moderate, and severe (including persons with COVID-19-related comorbidities)	muscle pain, fatigue, fever, and headache	[53, 56, 57]
CanSino COVID-19 vaccine (Convidecia (Ad5-nCoV))	Viral Vector	People 18 years and older	Mild, moderate, and severe	Induration, muscle pain, fatigue, fever, and headache	[53, 58]
Sinopharm/BIBP COVID-19 vaccine (BBIBP-CorV)	Inactivated virus	People 18 years and older	Mild, moderate, and severe	muscle pain, fatigue, fever, and headache	[53]
Sinopharm/WIBP COVID-19 vaccine (SARS-CoV-2 Vaccine (Vero Cell), Inactivated (InCoV))	Inactivated virus	People 18 years and older	Mild, moderate, and severe	muscle pain, fatigue, fever, and headache	[53]
Sinovac COVID-19 vaccine (CoronaVac)	Inactivated virus	People 18 years and older	Mild, moderate, and severe	Hypersensitivity Reactions, urticaria/angioedema and pruritus	[53, 59]
Bharat Biotech, India COVID-19 vaccine (COVAXIN (BBV152))	Inactivated virus	People 18 years and older	Mild, moderate, and severe	muscle pain, fatigue, fever, and headache	[53, 56]
Valneva COVID-19 vaccine (VLA2001)	Inactivated, adjuvanted virus	People 18 years and older	Mild, moderate, and severe	Immune-mediated disorders, complications associated with COVID-19, muscle pain, fatigue, fever, and headache	[60]
Zhifei Longcom, China COVID-19 vaccine (ZF2001)	Recombinant protein	Children, adolescents aged 3-17 years and people 18 years and older	Mild, moderate, and severe	Acute allergic dermatitis, muscle pain, fatigue, fever, and headache	[53, 61]

BioCubaFarma-Instituto Finlay de Vacunas COVID-19 vaccine (Soberana 02 +Soberana Plus)	Recombinant protein	People 18 years and older	Mild, moderate, and severe	muscle pain, induration and erythema	[53]
Center for Genetic engineering and biothechnology COVID-19 vaccine (Abdala)	Recombinant protein	People 18 years and older	Mild, moderate, and severe	muscle pain, fatigue, fever, and headache	[53, 62]
Novavax COVID-19 vaccine (Covovax (NVX-CoV2373))	Recombinant protein	People 12 years and older	Mild, moderate, and severe	muscle pain, induration and erythema	[53]
Vector State Research Centre of Virology and Biotechnology COVID-19 vaccine (EpiVacCorona)	Peptide antigen	People 18 years and older	Mild, moderate, and severe	muscle pain	[63]
Zydus Cadila COVID-19 vaccine (ZyCoV-D)	DNA based vaccine	People 12 years and older	Mild, moderate, and severe	muscle pain, fatigue, fever, and headache	[64]

Table 1.3*. Summary of monoclonal antibodies issued by FDA for emergency use in managing COVID-19*

Monoclonal antibody	**Recommended age group**	**Effectiveness to disease phases**	**Side effects/ potential risks**	**Reference**
Sotrovimab	Aged ≥12 years and weighing ≥40 kg	Mild to moderate	Infusion-related hypersensitivity reactions, rash, and diarrhoea	[65]
Bamlanivimab and Etesevimab combination	Aged ≥12 years and weighing ≥40 kg	Mild to moderate	Nausea, diarrhoea, infusion-related immediate hypersensitivity reactions manifesting as pruritus, flushing, rash, and facial swelling	[65]
Casirivimab and imdevimab combination	Aged ≥12 years and weighing ≥40 kg	Mild to moderate	Infusion-related or hypersensitivity reactions	[65]
Tocilizumab	Aged ≥ 2 years	Severe	Increase in serum cholesterol	[52]

1.7 Overview of multi-omics COVID-19 studies

Multi-omics studies aim to combine two or more omics datasets including, but not limited to, transcriptomics, proteomics, metabolomics, and lipidomics, to aid in data analysis, visualization, and interpretation to determine the mechanisms underlying specific biological systems. In the context of COVID-19, individual omics datasets provide specific insights into the contributions/manifestations of molecules at that omics level during disease progression. Multi-omics studies present a means to collectively compare multiple omics data levels, from different experiments either on the same samples or across studies, to understand the biochemical underpinnings of COVID-19 outcomes. Multi-omics studies broaden our insights into the aggregative effects of multi-omics data and the flow of biological information at multiple levels during the disease trajectory and have the potential to both yield more nuanced disease subtyping metrics and reveal actual molecular mechanistic causes of disease states [66-68]. Accordingly, COVID-19 has already been the focus of various multi-omics studies (**Table 1.4)**. As outlined in **Table 1.4**, combining data from different omics layers improves disease classification, risk prediction, and identification of potential biomarkers for personalized medicine approaches [44, 46, 69]. Multi-omics studies reveal how individual genetic makeup, immune response profiles, and existing health conditions significantly influence disease severity and progression [26, 70]. Also, multi-omics analyses provide insights into how the virus interacts with host cells, manipulates cellular processes, and evades immune defenses [71].

The COVID-19 pandemic has highlighted the critical role of genomics surveillance in tracking viral evolution, identifying potential threats, and informing public health interventions. While traditional sequencing of viral genomes has been invaluable, multi-omics approaches are rapidly emerging as a powerful tool to deepen our understanding of the virus and its interactions with the host. For instance, by analyzing host responses and viral interactions, multi-omics has contributed to identifying variants with increased virulence or transmissibility [72, 73]. Similarly, multi-omics studies have investigated mutations' influence on viral fitness, transmission dynamics, and potential escape from immune responses [74].

Crucially, these studies have yielded large quantities of publicly available data most of which have not yet been fully utilized to answer some key questions relating to the

clinical heterogeneity of COVID-19 outcomes. Leveraging publicly available multi-omics data drawn from both healthy COVID-19 individuals and coupling this data with clinical measurements in a systematic integrative approach will help to comprehensively assimilate and annotate these datasets to advance our knowledge of the molecular factors behind disease outcomes.

Table 1.4. *Overview of some multi-omics studies on COVID-19.*

Study title	Key findings	Analysed samples	Omics data used	Study population	Reference
Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19	• Significant changes in immune cell compositions between mild and moderate disease phase • Inflammation and a sharp drop in blood nutrient among moderate and severe cases • The emergence of novel immune cell subsets in moderate cases which increases with disease severity	Serial blood from a cohort of 139 patients representing all levels of disease severity and 258 healthy controls	500 proteins and 1000 metabolites from plasma. Whole transcriptome, 192 surface protein data and 32 secreted proteins, T cell and B cell receptor gene sequences retrieved from PBMCs	American	[26]
Large-Scale Multi-omic Analysis of COVID-19 Severity	• Identified 219 diverse biomolecules strongly associated with disease status and severity • Revealed dysregulation of lipid transport process and neutrophil degranulation.	Blood samples from 102 and 26 COVID-19 and non-COVID-19 patients respectively	Leukocyte mRNA expression, plasma protein, metabolite, and lipid	American	[69]
Longitudinal Multi-omics Analyses Identify Response of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19	• COVID-19 disease elicits dynamic changes of circulating cells in the blood • Severe COVID-19 is characterized by increased metabolically active plasmablasts • Elevation of IFN-activated megakaryocytes and erythroid cells in severe COVID-19 • Cell-type-specific expression signatures	Peripheral blood samples from 14 patients	Bulk transcriptome, DNA methylome, single-cell transcriptome	Caucasian Asian	[44]

	influence fatal COVID-19 outcome				
Omics-Driven Systems Interrogation of Metabolic Dysregulation in COVID-19 Pathogenesis	• Exosomes of COVID-19 with an elevating level of disease severity were increasingly enriched in monosialodihexosyl gangliosides • Identified pathologically relevant lipid clusters. • Differential correlation analyses uncovered metabolic dysregulation in COVID-19	Serum lipidome and metabolome from 50 COVID-19 patients and 26 healthy controls	598 lipids and 404 polar small metabolites	China	[75]
Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment	• Revealed alterations in the myeloid cell compartment associated with severe COVID-19 • Provided detailed insights into systemic response to SARS-CoV-2 infection • Mild COVID-19 is marked by inflammatory *HLA-DRhiCD11chi*+ monocytes	Whole blood and peripheral blood mononuclear cells from 53 COVID-19 patients and 56 controls	Single-cell transcriptomics, single-cell proteomics	Germany	[46]
Multi-omics-based identification of SARS-CoV-2 infection biology and candidate drugs against COVID-19	• SARS-CoV-2 shares influenza, tuberculosis, malaria, measles, hepatitis virus, ebola virus, leishmaniasis infection pathways • Identified candidate drugs for COVID-19	Lung epithelial cells upon viral infection, 1067 up-regulated genes from the peripheral blood of COVID-19 patients, 136 up-regulated human proteins of Caco-2 cells upon transfection with SARS-CoV-2	SARS-CoV-2 infected host interactome, proteome, transcriptome, bibliome		[76]
Disease severity-specific neutrophil	• Identified COVID-19 subgroups with distinct transcriptome signatures.	Whole blood samples from 39 COVID-19	RNA-seq of whole blood cell	Netherlands Greece	[45]

signatures in blood transcriptomes stratify COVID-19 patients	• Predicted patient subgroup-specific drug candidates that could counteract immune dysregulations. • Revealed extra molecular subtypes within the immune response of SARS-CoV-2 infected patients that describes disease diversity	patients with community-acquired pneumonia and 10 control donors Granulocyte samples from 16 longitudinally COVID-19 patients	transcriptomes, granulocytes	Germany	
Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV	• Revealed perturbations taking place upon SARS-CoV-2 infection. • Identified molecular mechanisms of SARS coronavirus and vulnerability hotspots of SARS-CoV-2	Virus-host protein-protein interaction, cell line data, transcriptome	Interactome, proteome	Not population-specific	[71]
Comprehensive Meta-Analysis of COVID-19 Global Metabolomics Datasets	• Identify robust metabolic signatures characterizing disease progression and clinical outcome	Secondary multi-omics datasets	Metabolomics datasets	United States China Brazil	[47]

1.8 Problem Statement

COVID-19 has precipitated an unprecedented global health crisis. The burden of this disease is multifaceted, encompassing considerable morbidity and mortality, economic disruption, healthcare system strain, and widespread psychosocial effects. The mode of progression of COVID-19 typically follows a path from exposure of the host cells to the virus, and then, the early phase of the disease, and in some cases, to the late phase of the disease (**as described in sections 1.3 and 1.4**).

Biosignatures are chemical species, features, or processes that provide evidence for the presence of a disease, trait or disease phenotype. To date, only a few biosignatures or biomarkers including IL6, and IL6R have been convincingly demonstrated to be potentially causally linked or discriminate COVID-19 disease severity states. Specifically, some subgroup-specific biosignatures underpinning or driving the observed disease state variations have been identified [77-82]. A major challenge to the current multi-omics approaches to explore COVID-19 biomarkers and clinical heterogeneity relies on multi-omics data generated from the same biological samples in an experiment but not multiple multi-omics data generated from different biological samples and experiments. Another challenge is that limited existing studies have explored the relationships between proteomics, transcriptomics, metabolomics, and lipidomics data and how that could influence COVID-19 clinical heterogeneity. There remains, however, a need to leverage multiple multi-omics data and integrative multi-omics network-based computational approaches to delve deeper into the biological relationships between the biosignatures underlying various COVID-19 disease phases to dissect, and also confirm their contribution to disease severity [83]. Proving that such biosignatures are not merely indirect collateral consequences of the actual molecular drivers of disease states is difficult but is a necessary step along the path to understanding how and why identified biosignatures contribute to disease severity across patient subgroups.

On the other hand, the development and roll-out of vaccines and gradually increasing levels of naturally acquired immunity have both contributed to significantly curtailing the pandemic and have restored societies around the world to a state of normalcy in our social livelihood. However, in addition to issues related to the equitable distribution

of vaccines, there are known side effects to taking the vaccines (e.g., myocarditis, stroke) [84]. Furthermore, there are some concerns surrounding vaccine efficacy against a backdrop of waning immunity [84, 85], emergent viral lineages (including B.1.351 Beta, P.1 Gamma, B.1617.2 Delta, C.37 Lambda, Omicron BA.2.86, XBB.1.5, and other immune evasive variants that have yet to emerge) [53, 86, 87]. Such observation aligns with propositions that COVID-19 is expected to become another endemic human coronavirus, possibly with seasonal epidemic waves [88]. Moreover, it remains unknown whether true herd immunity of the sorts achieved with measles and rubella will ever be achieved for COVID-19 irrespective of how many people become infected or how intensively vaccines are applied. Particularly, herd immunity may not be possible given that the virus changes a lot resulting in different variants and lineages.

While data is currently being collected and analysed to understand how newly evolved SARS-CoV-2 variants might impact the effectiveness of vaccines and the severity of future COVID-19 infection waves, demand for effective host or pathogen-specific drug targets and drug repurposing candidates that could be appropriated for treatments during different disease phases.

Newly identified drug candidates should have the potential to impact the virus's ability to survive within the body and/or dampen the inappropriate/counter-productive immune responses of severely ill patients. Importantly, exploring host-factor-targeted drugs is a strategy that can circumvent issues of virus-acquired drug resistance.
An integrative multi-omics network-based approach coupled with drug data and robust statistical methods could provide additional opportunities to explore data, stratify, predict, rank, and validate disease-related variables that might either have distinctive characteristics with disease progression or be good targets for future therapeutics (whether these be completely uncharacterized small molecules or repurposed already well-characterized drugs and drug candidates).

1.9 Hypothesis

This project is based on the below three hypotheses;

1. Investigating biosignatures (proteins, transcripts, metabolites, and lipids) across different phases of COVID-19 disease will provide insights into the molecular underpinnings of the enormous clinical heterogeneity of COVID-19.

2. Associations between biosignatures within a biological network would permit the prioritization of biosignatures that discriminate between disease states and yield potential drug targets.

3. Associations between biosignatures and drugs within a biological network would enable the prioritization of drug repurposing candidates for use as treatments during different COVID-19 disease phases.

1.10 Aims and Objectives of the thesis

Aim 1: Explore network-based integrative approaches for multi-omics analysis

Objectives

1. Classify network-based multi-omics integrative tools based on methods and/or algorithms implemented.
2. Explore the most appropriate tools and methods for investigating the hypothesis relating to the prioritization of biosignatures within a biological network.
3. Explore the most appropriate tools and methods for investigating the hypothesis relating to the prioritization of potential drug repurposing candidates through associations between biosignatures and drugs within a biological network.

Aim 2: Integrate multi-omics data to construct COVID-19 disease phase biological networks.

Objectives

1. Implement an innovative approach to harmonize different multi-omics data from multiple data sources across different disease severity states.
2. Leverage correlation-based network approach to construct biological networks for mild, moderate, and severe COVID-19 disease phase.

Aim 3: Map biosignatures onto COVID-19 disease phase biological networks.

Objectives

1. Implement an innovative network diffusion algorithm for systematic network analysis to prioritize biosignatures that discriminate COVID-19 disease phases.
2. Implement an innovative network diffusion algorithm for systematic network analysis to prioritize potential drug targets.
3. Evaluate COVID-19 disease phase biological networks to investigate clinical heterogeneity.

Aim 4: Prioritize potential drug repurposing candidates that could be appropriate for mild, moderate, and severe COVID-19 disease phases using a network-based integrative approach that systematically integrates drug-related data and multi-omics data.

Objectives

1. Assemble a drug data interactome using drug data from DrugBank, COVID-19 knowledge graph, drug repurposing knowledge graphs, and disease phase biological networks.
2. Implement an innovative network diffusion algorithm for systematic network analysis to prioritize potential drug repurposing candidates that are FDA-approved.

1.11 Relevance of the proposed study findings to clinical practice in African countries

The impact of COVID-19 in Africa has been substantially lower compared to countries in the Americas, Europe, and Asia [89]. A total of 12,294,580 COVID-19 cases and 257,039 deaths (CFR: 2.1%) have been reported by the 55 African Union (AU) Member States (MS) as of May 2023 (Accessed on June 18, 2024, https://au.int/en). This represents 2% of all cases and 4% of all deaths reported globally.

The study of biosignatures for COVID-19 severity and progression is highly relevant to clinical practice in African countries, as it can potentially improve patient outcomes, optimize resource utilization, and enhance our understanding of the disease in diverse populations. Firstly, understanding the biomarkers associated with the severity and progression of COVID-19 can aid healthcare professionals in identifying high-risk patients early on, allowing for targeted interventions and monitoring. Furthermore, identifying specific biomarkers associated with severe COVID-19 outcomes can help inform treatment decisions and guide therapeutic strategies in African countries.

Computational drug discovery offers a potential solution by rapidly identifying potential drug candidates for COVID-19 treatment, which can help alleviate the strain on healthcare systems in African countries. Additionally, traditional drug discovery methods can be time-consuming and expensive, making it difficult for African countries with limited resources to afford and access new treatments. Computational drug discovery, on the other hand, is a cost-effective and efficient approach that can accelerate the development of new therapies for COVID-19

Overall, this research can potentially contribute to more tailored and effective management of COVID-19 in African settings, thereby helping to mitigate the impact of the pandemic on local healthcare systems and communities.

1.12 Thesis summarizing outline

This thesis has five main chapters.

Chapter 1 covers the general introduction and scope of the work presented in this thesis.

Chapter 2 delves into computational tools and methodologies employed for network-based integrative multi-omics analysis. Aligned with Aim 1 of this thesis, the chapter 2 covers a systematic review of network-based multi-omics/multi-modal integrative analytical approaches. We provide valuable insight into the most suitable methods and tools for addressing pertinent research questions applicable to different diseases. By examining these methods and tools, we aim to equip researchers with the knowledge necessary to select the most appropriate analytical framework for their specific research questions, particularly within the context of various disease studies. To demonstrate the practical application of these methods, we showcased their use in understanding the aetiology and treatment of COVID-19. Furthermore, we acknowledge the challenges associated with integrating diverse data types into a network framework. By highlighting these challenges, we aim to encourage the development of innovative solutions to overcome them. Finally, we presented some recommendations and future directions to advance this field. Chapter 2 of this thesis is published in Frontiers in Molecular Biosciences and accessible at https://doi.org/10.3389/fmolb.2022.967205.

Chapter 3 covers multi-source multi-omics network integrative analysis to identify biosignatures specific to COVID-19 disease phases. In line with the proposed Aims 2 and 3, in this chapter, we test the proposed hypotheses 1 and 2 by presenting a novel concept for multi-omics patient clinical metadata harmonization that enables classifying omics data into distinct COVID-19 disease states (i.e. mild, moderate, and severe). We integrate proteomics, transcriptomics, metabolomics, and lipidomics data from two independent studies together with the existing knowledgebase to construct COVID-19 disease-state specific omics-graphs. Utilizing a random walk with restart algorithm on these COVID-19 disease state omics-specific graphs, we prioritize key biosignatures and identify those that uniquely distinguish between the mild, moderate,

and severe disease states. Furthermore, we uncover pairwise interactions across the different disease states, providing valuable insights into the underlying mechanisms driving disease progression. The analytical findings of chapter 3 reveal a striking correlation between increased cross-layer interactions and disease severity. This observation suggests that the observed clinical heterogeneity in COVID-19 patients may be partially attributed to the intricate interplay between different biological layers within each disease stage. Chapter 3 of this thesis is published in Frontiers in Molecular Biosciences. doi: 10.3389/fmolb.2024.1393240

Chapter 4 covers predicting drug repurposing candidates for COVID-19 disease phases. In line with the proposed Aim 4, in this chapter, we test hypothesis 3 by prioritizing potential candidate drugs that could be repurposed for mild, moderate, and severe COVID-19 disease phases using a network-based integrative approach that systematically integrates drug-related data and multi-omics datasets. The results from the analysis revealed potential FDA-approved drug candidates for the treatment of COVID-19, across mild, moderate, and severe disease stages. These categories include: (1) Antioxidants: These agents, like ghlutathione, neutralize harmful free radicals and may reduce inflammation associated with COVID-19, (2) Immunosuppressives: These drugs, like sirolimus and cyclosporine, suppress the immune system, potentially mitigating the overactive immune response seen in severe COVID-19, (3) Anti-IL-6 receptor monoclonal antibodies: These drugs, like sarilumab and tocilizumab, target the *IL-6* receptor, a key player in the inflammatory cascade of COVID-19, (4) Natural compounds: Natural products like curcumin possess potent anti-inflammatory and antioxidant properties, making them potential candidates for COVID-19 treatment, (5) Corticosteroids: Drugs like dexamethasone exhibit potent anti-inflammatory effects, proving beneficial in managing severe COVID-19 cases, (6) Essential minerals: Maintaining optimal mineral levels like zinc is crucial for immune function and may play a role in COVID-19 recovery, (7) Tyrosine kinase inhibitors: These drugs, like crizotinib, target specific signalling pathways involved in viral replication and inflammation, potentially offering therapeutic benefits for COVID-19, (8) Methyltransferase inhibitors: Drugs like vorinostat inhibit specific enzymes involved in viral replication, presenting a potential avenue for COVID-19 treatment, (9) Protein synthesis inhibitors: Drugs like omacetaxine block viral protein production, potentially halting viral replication and disease progression, (10) Non-steroidal anti-inflammatory

drugs (NSAIDs): Drugs like indomethacin possess anti-inflammatory properties, though their effectiveness in COVID-19 treatment requires further research, and (11) Anticancer drugs: Drugs like crizotinib and omacetaxine originally developed for cancer treatment exhibit antiviral properties, warranting investigation in the context of COVID-19. The chapter 4 is published online in zenodo (doi: https://doi.org/10.5281/zenodo.10568146) and is accepted for publication in Drug Repurposing Journal DOI: 10.58647/DRUGREPO.24.1.0007.

Chapter 5 covers the general discussion and conclusion of the presented results. It offers a concise summary of the entire thesis, encompassing the research objectives, methodologies employed, key findings, and significant contributions. It provides a holistic overview of the work undertaken and its impact on the understanding of network-based multi-omics analysis, particularly in the context of COVID-19. This chapter concludes by gazing into the future, highlighting potential directions for future work in the field of network-based multi-omics analysis.

Chapter 2

Computational approaches for network-based integrative multi-omics analysis

Francis E. Agamah[1,2], Jumamurat R. Bayjanov[3], Anna Niehues[3], Kelechi F. Njoku[1], Michelle Skelton[2], Gaston K. Mazandu[1,2,4], Thomas H.A. Ederveen[3*], Nicola Mulder[2], Emile R. Chimusa[5*], Peter A.C. 't Hoen[3*]

[1]Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
[2]Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, CIDRI-Africa Wellcome Trust Centre, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
[3]Center for Molecular and Biomolecular Informatics (CMBI), Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, Netherlands
[4]African Institute for Mathematical Sciences, Cape Town, South Africa
[5]Department of Applied Sciences, Faculty of Health and Life Sciences, Northumbria University, Newcastle, United Kingdom

doi: https://doi.org/10.3389/fmolb.2022.967205

Abstract

Advances in omics technologies allow for holistic studies into biological systems. These studies rely on integrative data analysis techniques to obtain a comprehensive view of the dynamics of cellular processes, and molecular mechanisms. Network-based integrative approaches have revolutionized multi-omics analysis by providing the framework to represent interactions between multiple different omics-layers in a graph, which may faithfully reflect the molecular wiring in a cell. Here we review network-based multi-omics/multi-modal integrative analytical approaches. We classify these approaches according to the type of omics data supported, the methods and/or algorithms implemented, their node and/or edge weighting components, and their ability to identify key nodes and subnetworks. We show how these approaches can be used to identify biomarkers, disease subtypes, crosstalk, causality, and molecular drivers of physiological and pathological mechanisms. We provide insight into the most appropriate methods and tools for research questions as showcased around the aetiology and treatment of COVID-19 that can be informed by multi-omics data integration. We conclude with an overview of challenges associated with multi-omics network-based analysis, such as reproducibility, heterogeneity, (biological) interpretability of the results, and we highlight some future directions for network-based integration.

Keywords: multi-omics, data integration, multi-modal network, machine learning, network diffusion/propagation, network causal inference

2.1 Introduction

Studies that implement large-scale molecular profiling techniques (-omics technologies) have increased our understanding of disease mechanisms and led to the discovery of new biological pathways, genetic loci underpinning disease progression, biomarkers, and targets for therapeutic development [90-92]. Until recently, these studies have mostly relied on single omics investigations. Dependencies between biological features and the relationships between different molecular layers (for example transcriptome, proteome, metabolome, microbiome, and lipidome) remain mostly elusive. The holistic understanding of the molecular and cellular bases of disease phenotypes and normal physiological processes requires integrated investigations of the contributions and associations between multiple (different but parallel) molecular layers driving the observed outcome. Most importantly, genetic information flows from the genome to traits and involves several molecular layers [66, 91]. Thus, understanding the genetic architecture of complex phenotypes would involve integrating and investigating the interactions between different molecular layers [66, 93-95].

Multi-omics datasets require appropriate computational methods for data integration and analysis. These methods/models implement statistical, network-based, and/or machine learning (ML) techniques on different omics layers to elucidate key omics features associated with diseases at various molecular levels and predict phenotypic traits and outcomes with increased accuracy [96-98].

Based on the hypothesis that molecular features within a system establish functional connections or are part of modules to carry out processes, network-based methods offer a framework to conceptualize the complex interactions in a system as a collection of connected nodes (molecular features). They further suggest possible connections (e.g., genotype to phenotype relationships) and/or subnetworks (e.g., biological pathways) that are informative of an observed phenotype [93]. Therefore, network-based methods are particularly useful for assessing complex interactions within multi-omics datasets and illustrating dependencies among multiple features. In addition, some network-based methods can incorporate prior information to guide the

integrative analysis. For this reason, network-based methods have attracted considerable attention in multi-omics data integration around understanding disease mechanisms and drug discovery [99, 100]. Previous reviews have mostly focused on the network-based analysis of single-omics data [101-103] or different approaches toward multi-omics data integration [104, 105]. Here, we review different integrative network-based approaches and some tools for multi-omics data analysis.

The outline of the review is as follows; we begin with a discussion on integrative multi-omics approaches, where we highlight the approaches for network-based analyses. We then discuss the different classes of methods for multi-modal network analysis. Next, we describe several network-based integrative multi-omics tools. This is followed by a discussion on the application of network-based tools to pertinent biological questions. This section guides the choice of the most appropriate network-based tools to answer a given biological question. As further examples, we show how some tools have been applied to COVID-19 research, which is currently one of the research areas benefiting from multi-omics integration approaches. Finally, we conclude with a discussion on some challenges associated with multi-omics analysis and the possible directions to mitigate such challenges.

2.2 Integrative multi-omics approaches

After initial data selection, processing, and quality assurance, an appropriate data analysis approach needs to be selected. We categorize integrative multi-omics analysis approaches into two main categories, multi-stage and multi-dimensional (multi-modal) analytical approaches (**Figure 2. 1**) [106, 107]. The multi-stage integration involves integrating data from different technologies using a stepwise approach. In this approach, omics layers are analysed separately before investigating statistical correlations between different biological features from the datasets under consideration. This analytical approach puts an initial emphasis on the relationships of features within an omics layer and how they relate to the phenotype of interest [96]. The multi-modal analytical approach involves integrating multiple omics profiles in a simultaneous analysis [90, 96, 106, 108].

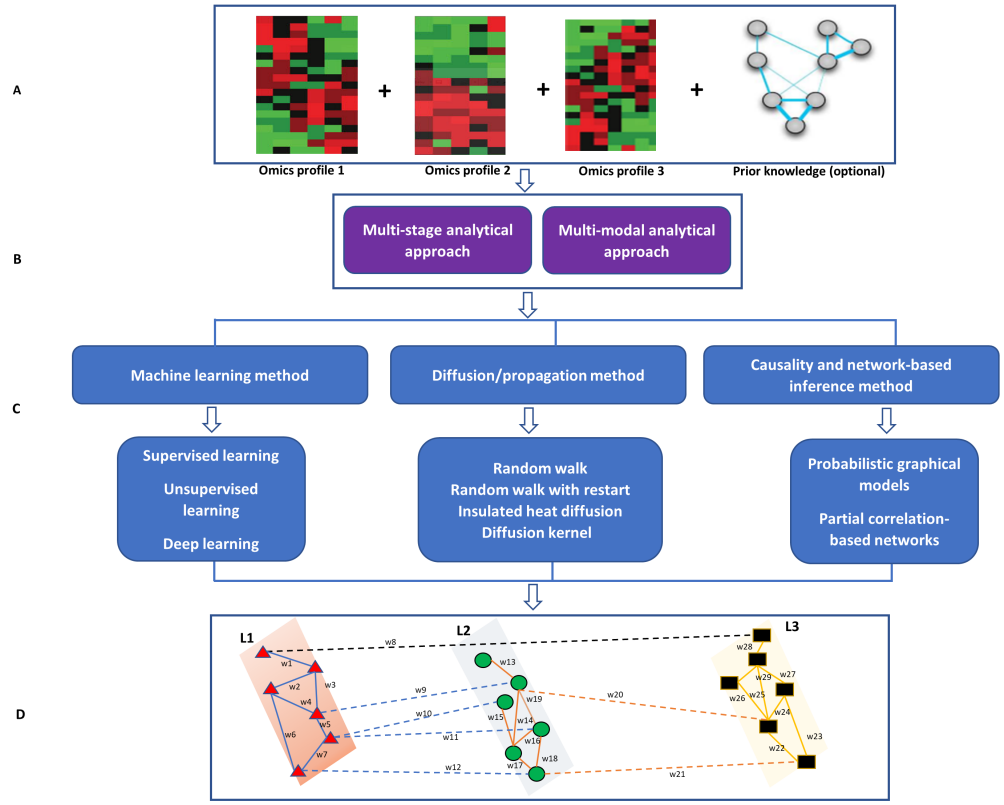


Figure 2. 1*. An overview of the multi-omics integration approach and the methods for network-based integration. (A) Processed omics data and prior knowledge for integrative analysis. (B) An integrative multi-omics approach that could be implemented. (C) Integrative network-based methods (D) Multi-layered network showing intra-layer interaction (solid lines) and crosstalk (dashed lines) across different layers (L1, L2, L3). The nodes are shaped and coloured to represent different omics features within the omics layers they are involved in. The edges are coloured to show different interactions within and between omics layers.*

2.3 Methods for multi-modal network analysis

In this review, we focus on (1) machine learning-driven network-based methods, (2) network-based diffusion/propagation methods, and (3) causality- and network-based inference methods. The selection criteria were based on the fact that these multi-omics/multi-modal network-based methods implement network architectures together with statistical and mathematical models for integrative multi-omics data analysis. Most of these methods can be implemented in both multi-stage and multi-dimensional multi-omics analysis (**Figure 2. 1**).

2.3.1 *Machine learning-driven network-based methods*

Machine learning (ML) is a collection of data-driven techniques for fitting an analytical model to a given dataset. ML methods do not only provide the framework to automatically learn models from large multi-omics data and make accurate predictions

but also implement network architectures to exploit interaction across the different omics layers e.g., for exploring omics-phenotype associations [109]. ML comprises mainly supervised and unsupervised learning methods. Supervised learning uses labelled datasets to train models to yield the desired output and emphasizes predictions by inferring discriminating rules from the data. Supervised learning model training requires comprehensive data and can be time-consuming, while unsupervised learning uses unlabelled data, to find latent structures or patterns in the data.

Classical graph-based ML methods (e.g., label propagation, a method for assigning labels to unlabelled points) can be used for a variety of tasks including generating graph edges, estimating node weights (quantitative measure of node importance) as well as estimating and optimizing edge weights (quantitative measure of the importance of the pairwise interaction between nodes) in a network to exploit the structure of graphs and learn models from the data [110]. Subsequent network optimization techniques introduce perturbations into the network and identify highly perturbed subnetworks to prioritize the most relevant features that correlate with the biological processes under study.

Multiview/multi-modal ML is an emerging method for multi-omics data integration used to exploit information captured in each omics dataset and infer from the associations between the different data types [111]. Multi-view learning implements the alignment-based framework and the factorization-based framework [111]. The alignment-based framework is a method based on the supervised setting for seeking pairwise alignment among different omics data whereas the factorization-based framework is based on an unsupervised setting for seeking a common representation of features across different omics layers. Deep learning methods, an example of multiview/multi-modal learning, have become one of the more promising integration methods not only because of their ability to exploit the structure of graph neural networks/graph or convolutional networks in both supervised and unsupervised settings with high sensitivity, specificity, and efficiency compared to classical ML methods but also, the predictive performance and capability to capture nonlinear and hierarchical representative features [112, 113]. The hierarchical feature processing can capture complex nonlinear associations in a multi-layered manner. The architecture of deep learning models consists of the input layer, hidden layer(s), and output layer. From the

perspective of multi-omics data integration, most deep learning methods follow the steps of (1) feature selection, (2) transforming high dimensional multi-omics data into low-ranked latent variables (3) concatenating multi-omics features into a larger dataset, (4) analysing the data for the desired task such as node ranking, link prediction, node classification and clustering (**Figure 2.2**) [113]. It is worth noting that the deeper the hidden layer, the more it can learn complex patterns in the data. A major challenge for deep learning methods is the problem of overfitting due to large features and the small sample size of multi-omics data. In addition, a large amount of cleaned data is required to train and validate the model, thus influencing how the model is interpreted [113]. Kang et al., in their review discussed several deep learning models including Deep-FS, autoencoder, and multi-omics late integration models that are used in multi-omics integration [113].

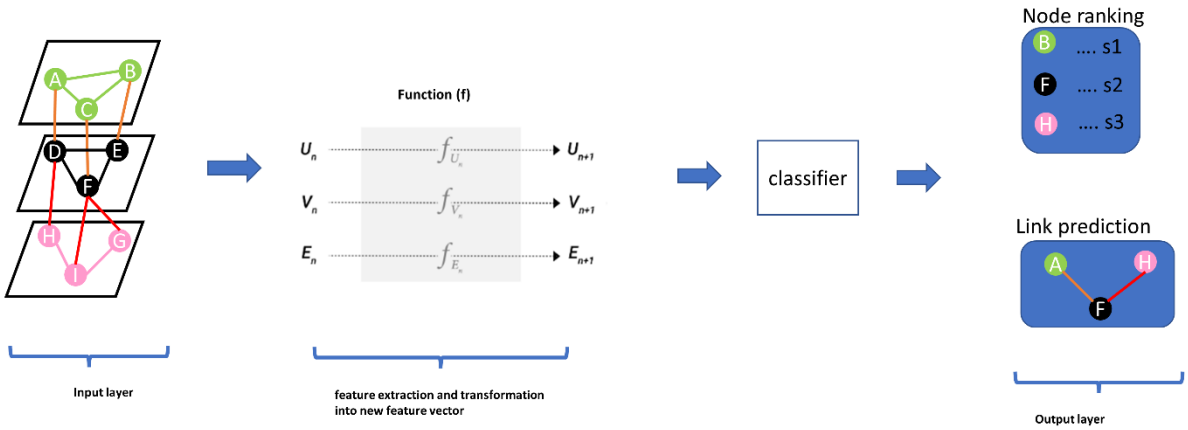


***Figure 2.2**. Graph Neural Networks (GNNs) are a class of deep learning methods designed to perform inference and predictions on graph data by learning embeddings for graph attributes (nodes, edges, global-context). The concept behind the architecture of these methods is such that it accepts graph data as input and produces the same input graph with updated embeddings before making predictions. GNN uses a function (f) on each graph component vector (nodes vector (Vn), edge vector (En), global-context vector (Un)) in the input graph to learn abstract feature representations of the graph to compute a new feature vector for nodes (Vn+1), edges (En+1) and global-context (Un+1)). The output layer could predict nodes ranked according to a particular score (s1, s2, s3) and also predict edges (links) in the input network.*

2.3.2 *Network-based diffusion/propagation methods*

Network-based diffusion/propagation is a technique for detecting the spread of biological information throughout the network along network edges, thanks to its

ability to amplify feature associations based on the hypothesis that node proximity within a network is a measure of their relatedness and contribution to biological processes [114, 115]. The method has been exploited in many network-based analysis pipelines and is suitable for analysing patient-level molecular profiles with different aims including disease subtyping because of its label propagation [114]. Propagation methods, including random walk, random walk with restart, insulated heat diffusion, and diffusion kernel networks, provide a quantitative estimation of proximity between features associated with different data types by considering all possible paths beyond the shortest paths (**Figure 2.3**) [114, 115].

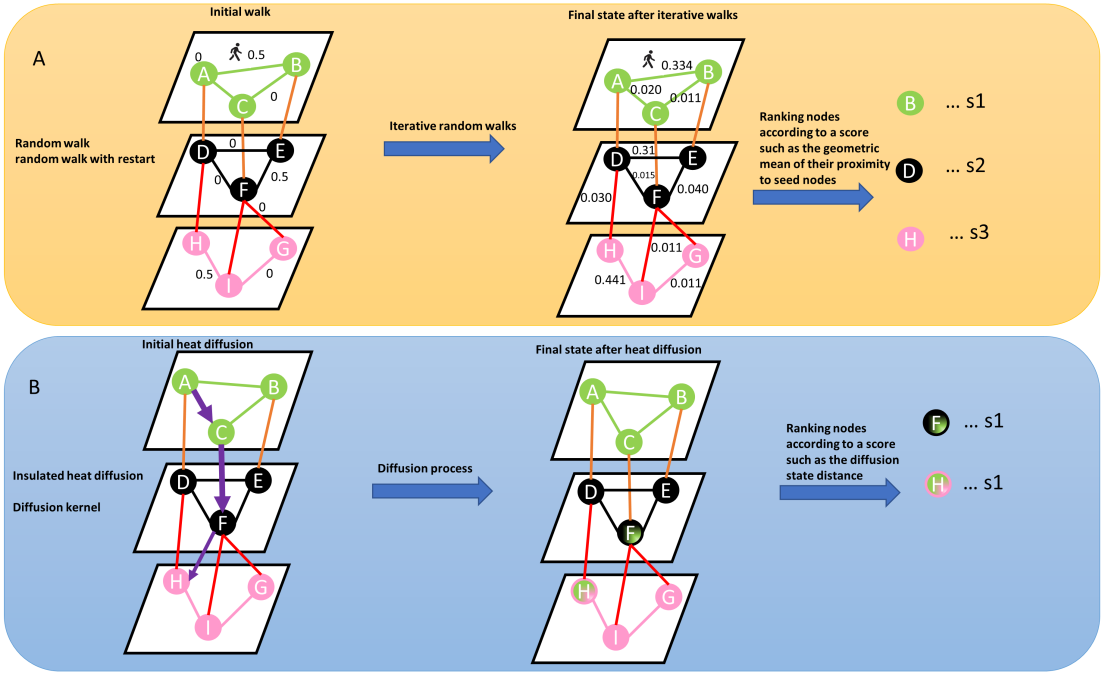


Figure 2.3. (A) *Describes a random walk from the seed node (e.g., node A). The concept behind random walk is a guilt-by-association approach where an imaginary particle explores the network structure from seed nodes. The direction of movement of the particle is completely independent of the previous directions moved. At each step, the particle transitions from any node in the graph with a certain probability (shown on the edges). The probability flow of random walks on a network is used as a proxy for information flows in the network to study the function of features, subnetworks, and prioritize features in the network. After several iterations, we are interested in the distribution of our position (Stationary distribution) in the graph (final state after iterative walks). The stationary probability distribution can be seen as a measure of the proximity between the seed(s) and all the other nodes in the graph. Nodes within the network can be prioritized using a specific metric (s1, s2, s3) such as the geometric mean of their proximity to seed nodes.* ***(B)*** *Describes heat diffusion from a reference query (e.g., node A). The concept behind heat diffusion in biological networks is perturbing nodes and simulating how the disturbance flows across edges within the network. Node disturbance means adding a scalar value (e.g., log fold changes from gene expression experiment, copy number variations) to node(s). Within a biological network, heat diffusion allows for the assessment of connectivity and topology of features which can allow the identification of relevant/dysregulated pathways and/or mutational effects across edges to neighbouring nodes. The purple arrow means diffusion jumps across different layers. The thickness of the purple arrow signifies the effect of query node (A) on nodes (F) and (H) as shown in nodes (F) and (H) in the final state graph after diffusion. Nodes within the network can be prioritized using a specific metric such as diffusion state distance.*

From a data analysis perspective, the network diffusion (ND) methods require omics data and network data. The network data could be obtained from *a priori* knowledge, inferred from omics data, or generated using a mixed approach of *a priori* and novel knowledge [114]. Omics data information, e.g., genetic aberration events underlying differential expression and/or a biological phenotype, are superimposed on the nodes (source nodes) within the network before the information is propagated via the edges until convergence and consensus features are found [114, 115].

ND methods transform input vectors of scores obtained from the omics data into dense vectors to eliminate missing values and ties. This transformation process can be applied before, after, or during the integration step to refine the results based on the molecular network data [114]. In the ND-before integration approach, the diffusion method is applied to a collection of scores (scores obtained from the omics data) that represent the multi-omics data. The ND-after integration approach is implemented when the various multi-omics data have been initially integrated into a unique structure. The ND-during integration approach is implemented in an instance where each layer exchanges information during the diffusion process. **Box 2.1** provides a summary of the equations related to the diffusion methods.

***BOX 2.1**. Summary equations of the network propagation/diffusion methods*

Random Walk

$$x_T = [AD^{-1}]^k . x_o$$

Random Walk with Restart (RWR)

$$x_T = \alpha[I - (1-\alpha)\, AD^{-1}]^{-1} . x_0$$

Insulated Heat Diffusion

$$x_T = \alpha[I - (1-\alpha)\, AD^{-\frac{1}{2}}]^{-1} . x_0$$

Diffusion Kernel

$$x_{T=}\, e^{\alpha(D-A)} . x_o$$

Where
x_T is the final state of the network after the propagation of information throughout the network x_0 is the initial biological information (initial state vector of aberration scores e.g., gene expression scores). A is the adjacency matrix of the network. D is the diagonal matrix of the out-degrees of nodes. AD^{-1} is the normalized adjacency matrix. k is the number of time steps, α is the restart probability and I is an identity matrix

2.3.3 *Causality- and network-based inference methods*

The mechanism of action within a biological system is fundamental to understanding such a system. For this reason, biological network inference and causal learning can

be used to investigate the direct and indirect multi-layer associations and possible causal relations between omics data features in the system [116].

Causal networks are generally graphical representations that demonstrate likely causal relations between nodes by capturing directional interactions and modelling dependencies between biological variables. The method enables researchers to put directionality between features in a network as well as decipher modules (subnetworks) and/or features associated with patient survival, disease processes, or pinpoint sources of perturbations within multi-omics biological network data [117].

Partial correlation-based networks enable the inference of features regulating co-expression or the activities of other features within the network by estimating conditional dependencies (partial correlations) [117]. Partial correlation corrects for spurious associations among features that are mediated by other variables measured in the dataset, thereby reducing the density of the network and enhancing its interpretability [117]. These methods have been implemented to infer mechanistic regulatory interactions or predict markers in biological networks [117].

Alternatively, network-based computational frameworks that implement probabilistic graphical models offer attractive solutions for causal reasoning and inference over multi-omics data [116, 118, 119]. A probabilistic graphical model (PGM) is a graph technique for modelling joint probability distributions and (in)dependencies over a set of random variables [118]. From a data analysis perspective, PGM uses graph-based representation (nodes as features and edges as direct probabilistic interactions between node pairs) as the basis to encode the complex distribution of the data for probabilistic reasoning and inference [118]. The framework of probabilistic graphical models includes a variety of directed and undirected models [118]. Directed models (e.g., Bayesian networks) require pre-defined directionality or capture conditional (in)dependencies to assert an influence on features. Undirected models (e.g., Markov networks) are undirected graphical models that offer a simpler perspective on directed models, especially in instances where the directionality of the interactions between features cannot be determined. Compared to directed models which can be used for causal reasoning and inference, undirected models are limited to inference tasks because they fail to capture the influence of nodes on neighbouring nodes.

In addition to partial correlation and probabilistic graphical models, advanced ML models and frameworks that are more computationally efficient have been explored for inferring causal relationships between multi-modal data [120-123]. Also, new methods that extend Bayesian networks have been developed for causal inference. For instance, Zheng et al. [124] developed a new method to estimate the structure and inference from a Bayesian network by transforming the structure learning problem into a continuous optimization formulation that does not impose any structural assumptions on the graph. In another instance, Lachapelle et al. [125] proposed a novel score-based approach to learning from Bayesian networks via the edge weights of neural networks. The approach developed by the authors adapts the optimization method presented by Zheng et al., [124] to allow for non-linear relationships between variables using neural networks. **Box 2.2** provides a summary of the equations related to the Bayesian and Markov methods. Given that the underlying principles behind network-based approaches for analysis vary, combining such approaches is feasible and may increase prediction accuracy as shown by Zheng et al., [124] and Lacapelle et al., [125].

BOX 2.2. *Summary equations of the Bayesian and Markov network.*

Bayesian network

Each node in a Bayesian network is represented as a probability distribution of ***cause*** given the observed ***evidence*** which is built from the **Bayes theorem** shown below [126].

$$P[Cause\ |Evidence] = P[Evidence\ |Cause]\ .\frac{P[Cause]}{P[Evidence]}$$

Thus, the full probability model for a Bayesian network is obtained by specifying the joint probability distribution (i.e., a series of the conditional probability distribution of the nodes in the network) [126].

Markov network

$P(X = x) = \frac{1}{z}\exp(\sum_i \omega_i f_i(x_{\{i\}}))$Where x is the feature vector, Z is the normalization constant calculated as

$$Z = \sum_{x\epsilon X} \exp\left(\sum_i \omega_i f_i(x_{\{i\}})\right)$$

f_i is the feature function defined as

$$f_i(x_{\{i\}}) = \begin{cases} 1 & F_i(x\{i\}) = true, \\ 0 & otherwise \end{cases}$$

ω_i is the non-negative real-valued weight which reflects constraints on nodes

F_i is the logistic formula.

2.4 Review of network-based integrative multi-omics tools

We systematically reviewed literature primarily published between 2010 to 2022 that report on ML-driven network-based tools, network-based diffusion/propagation tools, and causality- and network-based inference tools. The literature was extracted from PubMed. We further highlight the tool's uniqueness in terms of (1) input data types, (2) method/ algorithm implemented, (3) most important analytical steps, (4) potential node and/or edge weighting, and (5) predicted outcome (crosstalk, disease subtypes, biomarkers, subnetworks, and patient survival). The tools presented in this review (**Table 2.1**) (1) have broad biomedical data applications and are not restricted to specific (disease) research topics only, (2) are implemented as standalone software like R, MATLAB, Python libraries, or as part of a pipeline and, (3) account for the weight of nodes and/or edges within the network.

***Table 2.1**. Network-based multi-omics integrative tools for predicting biomarkers, crosstalk, disease subtypes, and subnetworks/enriched modules.*

Tool	Description	Major steps of the tool	Edge weighting component	Node weighting component	Outcome	Method/ Approach	Input data type	Major limitation	year	Reference
Machine learning-driven network-based tools										
mixOmics	An R toolkit dedicated to the exploration and integration of biological data sets with a specific focus on variable selection. The package contains suite of algorithms and functions. The function network is used for graph visualization.	1. Receives as input multiple matrices each representing a different omics 2. Perform network analysis using the network function	Infer interactions between nodes by using a pairwise association score	Leverages on measurements of variables	Relevance networks	Supervised and unsupervised ML	most omics types (genes, mRNA, metabolites, miRNomics data, proteomics)	it can be challenging to interpret the results, especially for users who are not well-versed in multivariate statistics	2012	[127]
Similarity network fusion	A network-based framework that uses networks of samples as a basis for integration. It fuses individual networks from each omics layer to represent the full spectrum of underlying data.	1. SNF first creates a sample-similarity network for each omics level and then fuses these into one network using a nonlinear combination method	Uses a scaled exponential similarity kernel to determine the edge weight. The weighted edges represent pairwise sample similarities	Nodes represent samples and the node size represents a phenotype like survival.	Identifies disease subtypes, performs survival prediction	Unsupervised ML	most omics types (mRNA, DNA methylation, and microRNA (miRNA) expression data)	SNF assumes equal weights to different -omic datasets, and this may not be biologically appropriate since not all datasets characterize or represent the underlying disease pathology to the same extent	2014	[128]
Lemon-Tree	A multi-omics module network inference software suite that finds co-expressed gene clusters and reconstructs regulatory programs involving other upstream omics data	1. Infer co-expressed gene clusters 2. Build consensus modules using the spectral edge clustering algorithm 3. Build module network 4. Module learning	Computes edge weight which represents the frequency with which pairs of genes belong to the same cluster	Compute the regulator score and considers the number of trees a regulator is assigned to, with what score (posterior probability), and at which level of the tree	Predicts driver genes/ biomarker	Unsupervised ML	expression data, copy number, microRNA, epigenetic profiles	It is computationally complex and requires substantial computational resources when dealing with large datasets	2015	[129]

Multiscale Embedded Gene Co-expression Network Analysis (MEGENA)	An R package co-expression network analysis framework that effectively and efficiently constructs and analyses co-expression networks	1. Constructs fast planar filtered network. 2. Identify multi-scale clustering structures 3. Perform multiscale hub analysis 4. Perform cluster-trait association analysis	Computes a similarity score between node pair	Compute node degree as node weight/size.	Predicts subnetworks, driver hubs	Unsupervised ML	Genes, mRNA, Fast planar filtered network	it can be computationally intensive, particularly for very large genomic datasets	2015	[130]
Omics Integrator	The approach applies advanced network optimization algorithms to a network to find high-confidence, interpretable subnetworks that best explain the data. The software is comprised of the Garnet and Forest tools. Forest provides perturbation strategies for perturbation analyses to determine the robustness of a network	1. Garnet identifies a set of transcriptional factors associated with mRNA expression changes by incorporating epigenetic changes nearby expressed genes. 2. Garnet scans regions proximal to transcribed genes for transcription factor binding sites and then regresses transcription factor affinity scores against gene expression changes 3. Forest identifies a condition-specific functional sub-network from user data and a confidence-weighted interactome 4. The confidence-weighted interactome is integrated with the 'omic' hits using the prize-collecting Steiner forest algorithm, where the data is either connected directly or via intermediate nodes, called 'Steiner nodes'.	Uses least-squares regression to relate the transcription factor affinity scores to mRNA expression changes. Forest converts uniform edge weights to costs using a scoring function	Transcription factors with motifs exhibiting statistically significant regression coefficients are given a weight of– log(p-value). The prize function assigns negative weights to nodes based on the number of connections they have in the interactome	Predicts subnetworks that connect changes observed in omics data.	Supervised ML	most omics types (mRNA, epigenetic changes, proteins, metabolites)	Difficulty in integrating heterogeneous more types of interactions (such as protein-RNA interactions), and multiple sets of prizes derived from patient data	2016	[131]

Weighted Similarity Network Fusion	A method that implements a modified similarity network approach to identify disease subtypes. It accounts for feature weights when clustering patients.	1. Build a regulatory network from the input data 2. Calculating the weight for each feature and ranking the features based on network information and the expression variation of the features. 3. Obtain weighted sample similarity networks from genes (mRNAs, TFs) and miRNAs separately using the weights and expression data of the features 4. Perform network fusion and clustering to find patient groups that imply disease subtypes.	Considers the similarity of two patients by considering the overall difference between the expression levels of all their features and the weight of each feature.	Computes feature weights by first ranking features using a modified PageRank algorithm followed by Integrating feature ranking and feature variation.	Identifies disease subtypes, performs survival prediction	Unsupervised ML	miRNA, mRNA, transcription factors	There is the challenge of determining the weights to assign to each individual network. The process of determining these weights can be subjective and may rely on assumptions or prior knowledge about the quality or reliability of each network	2016	[132]
iOmicsPASS	A method for integrating multi-omics profile over genome-scale biological networks and identifying predictive subnetworks that provides the mechanistic interpretation of a specific phenotype. The tool considers molecular interactions within and between omics data types as a data feature.	1. Integrates quantitative multi-omics data by computing interaction scores for a network. 2. Discover molecular interactions whose joint expression patterns predict phenotypic subnetworks/groups. 3. Report biological pathways enriched in the subnetworks using a modified nearest shrunken centroid algorithm	Computes scores for each molecular interaction. The scores are derived in the context of the type of interactions data (TF regulatory network and protein-protein interaction network with or without DNA copy number)	Utilizes measurement of each molecule in their respective omics data sets as node score.	Predicts phenotypic group-specific subnetworks, feature selection	Supervised ML	Biological network, mRNA, proteomics data, DNA copy number, sample metainformation	Does not provide functionalities for prediction of phenotypic groups in external data sets. Discards molecules that are not represented in the user-provided network data. Hence important markers that are poorly represented in biological networks can be lost in the analysis	2019	[133]
Sparse CRossmodal	A subtype classification model that represents a sparse	1. Biomarker filtering	Estimates connection between nodes.	Compute weight for nodes	Predicts disease subtype	Neural network	DNA methylation, mRNA,	Difficulty in integrating heterogeneous	2020	[134]

Superlayered Neural Network (SCR-SNN)	version of a cross-modal super-layered neural network.	2. Biomarker selection, using a cross-modal, super-layered neural network 3. Integration of selected biomarkers from omics data 4. Prediction model building						data sources with varying quality and formats		
Integrative Network Fusion	A framework for high-throughput omics data integration that leverages machine learning models to extract multi-omics predictive biomarkers.	1. A set of top-ranked features is extracted by juxtaposition by Random Forest (RF) and linear Support Vector Machine (LSVM) classifiers. 2. A feature ranking scheme is computed on similarity network fusion-integrated features 3. A random forest model is trained on the intersection of two sets of top-ranked features from the juxtaposition and feature ranking scheme (rSNF) and provides compact predictive biomarkers.	Uses a scaled exponential Euclidean distance kernel to compute edges weight.	Implements a feature ranking scheme on similarity network fusion integrated features.	Identifies disease subtypes and predictive biomarkers	Supervised ML	mRNA, microRNA expression, protein levels, copy number variants, DNA Methylation	Unable to supports the integration of two dimensional omics layers	2020	[131, 135]
Discovery of active Modules In Networks using Omics (DOMINO)	A network-based active module identification algorithm used for identifying subnetworks that show significant over-representation of accrued activity signal (“active modules”)	1. Receives as input a set of genes flagged as the active genes in a dataset and a network of gene interactions 2. Partition the network into disjoint, highly connected subnetworks 3. Detect relevant subnetworks containing active over-represented genes	Uses the confidence scores of the tissue-specific functional interactions as weights of edges.	Uses gene activity scores	Predicts subnetworks	Unsupervised ML	gene network and transcriptomics data	Difficulty in integrating heterogeneous data sources with varying quality and formats	2021	[136]

		4. Further, refine subnetworks into compartments 5. Repartition's subnetwork compartments in putative modules 6. Reports final modules that are over-represented by active genes								
Multi-source information super network	A network-based framework for constructing a single network from multi-source data	1. Constructs a super network based on the weighted sum of the pairwise weighted edge vectors (for each pair of genes)	Computes edge weights	Computes gene-specific scores based on characteristics and topology of the super network	Predicts subnetworks	Unsupervised ML	Genes, pathway information, CNVs, Drug data, mRNA, miRNA, PPI	The integration methods is not optimised in terms of the optimal weight selection for the integration of multi-source data.	2018	[137]
i-Modern	A deep learning network framework for integrating multi-omics data.	1. Feature extraction using optimized autoencoder 2. Low-dimensional feature extraction via Cox-PH models 3. Patient subgroup classification	Estimate connection between nodes	Implements a randomization approach to explore node weight	predict omics signatures, patient subgroup classification	Neural network	miRNA, somatic mutations, copy number variation (CNV), DNA methylation, proteins	Unabble to predict cross-layer interactions among different omics features.	2022	[138]
OmicsNet 2.0	A network-based multi-omics analysis platform and an R package (OmicsNetR) to easily build, visualize, and analyse multi-omics networks.	1. Accepts different data types as input 2. Search different molecular interaction database 3. Creates multi-omics networks 4. Performs network visual analytics	The methodology does not take edge directionality or weights into account	Uses feature activity scores	Predicts sub-networks, crosstalk	Unsupervised ML	Genes, proteins, transcription factors, miRNAs, metabolites, SNPs, Taxa, lc-ms Peaks	Assumes equal weights to different -omic datasets, and this may not be biologically appropriate since not all datasets characterize or represent the underlying disease pathology to the same extent	2022	[139]
multi-omics data integration	A method for multi-omics data integration that implements	1. Construct an affinity matrix for different omics data based on a	Computes edge weight as a measure of the	Utilizes measurement of each	Predicts disease subtypes	Unsupervised ML	mRNA, miRNA, proteomics	Uses attributes of all omics data to construct the	2022	[140]

for clustering to identify cancer subtypes (MDICC)	affinity matrix and network fusion methods	Gaussian kernel function. 2. Fuse affinity matrices into a new relational matrix with low rank 3. Cluster fused network	Euclidean distance between samples	molecule in their respective omics data.			data, DNA methylation,	affinity matrix as the network between samples, which can only reflect the relationship between samples but not the internal modules in the omics data.		
Network-based diffusion/propagation tools										
Tied Diffusion of Interacting Events (TieDIE)	TieDIE method extends the heat diffusion strategies by leveraging different types of genomic inputs to find relevant genes on a background network with high specificity.	1. Computes scores for each node in the graph 2. Utilizes multiple diffusion processes to predict disease-related genes, subnetworks, and pathways	The diffusion approach is used to describe the edge score between node pairs (1 and -1). Aij= 1 if node i activates node j, Aij = −1 if node i represses or inactivates node j, and 0 otherwise, where A is an adjacency matrix	Scores between -1 and +1 are assigned to the nodes reflecting a positive or negative association with the disease state. A node score of 0 reflects genes not known to be associated with the disease process. Nodes scores could represent experimental measurements	Predicts biomarkers and disease-specific subnetworks	Diffusion-based	genes, proteins, biological pathway features, mRNA, DNA methylation	Difficulty in integrating heterogeneous data sources with varying quality and formats It relies on the assumption that the interactions between different biological entities are accurately captured in the individual input networks. If the input data is noisy, incomplete, or contains errors, it can propagate inaccuracies in the fused network generated. .	2013	[141]
Network-based Integration of Multi-omics Data (NetICS)	A gene prioritization method that is a framework for per-sample network-based integration of diverse data types on a directed functional interaction network.	1. Constructs a directed functional interaction network from input functional interactions.	Compute connectivity scores between node pairs	Compute a ranking score for all genes	Predicts biomarkers	Random walk	miRNA-gene interaction, mRNA, DNA methylation, genetic	NetICS can only examine the effects of genes that are present in the interaction network.	2018	[142]

	NetICS provides insight into how aberration events that are different between samples of the same disease type cause similar expression changes in other genes.	2. Diffuse aberration scores from the aberrant genes following the directionality of the network interactions. 3. Diffuse differential expression scores from differentially expressed genes 4. Predicts how aberration events cause expression changes through gene interaction.					aberrations, protein levels	Moreover, the results may be biased towards highly connected genes as these have a higher chance of having aberrant or differentially expressed genes in their network neighborhood.		
Hierarchical HotNet	An algorithm that simultaneously combines network interactions and vertex scores to construct, identify, and rank statistically significant high-weight altered subnetworks across different omics datasets. It addresses the limitations of HotNet [143], HotNet2 [144] by combating ascertainment bias in data and integrating both network topology and vertex score.	1. Combines network topology and vertex scores 2. Defines a similarity matrix from the network using a random walk-based approach 3. Implements hierarchical clustering to construct a hierarchy of clusters consisting of highly connected components. 4. Assesses the statistical significance of clusters	Defines a similarity measure between node pairs using both network topology and vertex scores	Uses vertex scores in the input network	Predicts a hierarchy of mutated subnetworks	Random walk	Interaction network with vertex scores	Hierarchical HotNet is a modular method; different similarity measures, clustering algorithms or test statistics may be more appropriate for different datasets.	2018	[141, 145]
regNet	regNet R package utilizes gene expression and copy number data to learn regulatory networks to estimate the potential impacts of individual gene expression alterations on clinically relevant signature genes.	1. regNet learns a regulatory network from a large collection of paired gene expression and copy number profiles 2. Uses network propagation to quantify the impacts of altered genes sample-specific gene expression changes on other clinically relevant target genes.	Compute a connectivity table that represents learned links between genes	Compute impact score for regulator genes, describing the contribution to expression changes in another gene.	Predicts driver genes or disease biomarkers.	Diffusion-based	transcription factors, mRNA, copy number data	Difficulty in integrating heterogeneous data sources with varying quality and formats.	2018	[146, 147]

Integrative multi-cohort and multi-omics meta-analysis framework	A multi-omics meta-analysis framework that can identify robust molecular subnetworks and biomarkers for a given disease condition	1. Module (A) takes multiple independent mRNA datasets and performs a leave-one-out meta-analysis to identify reliable differentially expressed genes 2. Module (B) takes multiple independent DNA methylation datasets and identifies differentially methylated genes 3. Module (C) identifies methylation-driven genes 4. Methylation-driven genes are used as inputs in a network propagation algorithm to identify the proposed subnetworks	The confidence score for each protein-protein interaction is obtained from the STRING database.	Utilizes experimental values from differential expression and methylation for omics features.	Predicts biomarkers and subnetworks describing patients' clinical outcome	Diffusion-based	mRNA, DNA methylation, protein-protein interactions	The framework is limited to investigating disease conditions whose mechanisms of actions are known to be triggered by the change in DNA methylation	2019	[148]
Random walk with restart on multiplex and heterogeneous biological networks	A random walk algorithm able to exploit multiple biological interaction sources to integrate multiplex-heterogeneous networks	1. Define adjacency matrix for input networks 2. Compute transition probabilities of the random walk with restart 3. Performs propagation from seed nodes.	Generates weighted or unweighted adjacency matrix	Scores nodes according to their proximity to the seed nodes.	Predicts candidate features and subnetworks	Random walk	Multi-modal data	It is computationaly complex and intensive to perform random walk using all networks nodes as seeds in an iterative manner	2019	[149]
MultiPaths	A Python framework to build customized harmonized multi-omics networks from multiple biological databases. MultiPaths framework contains two independent Python packages: DiffuPy and DiffuPath useful for interpreting and contextualizing results from multi-omics experiments	1. DiffuPy implements four existing network propagation algorithms and five graph kernels and enables propagating user-defined labels, either as lists of entities or lists of entities with their corresponding quantitative values. 2. DiffuPath, wraps the generic diffusion algorithms	The methodology does not take edge directionality or weights into account for propagation	Compute node scores using a function of graph kernel and input scores.	Predicts subnetworks	Diffusion-based	genes, mRNA, metabolites, miRNomics data, biological pathway/processes data	Requires large-scale and integrated multi-layer networks to improve prediction performance, which in turn requires computational power.	2020	[145, 147]

		from DiffuPy and applies them to construct biological networks								
Analytic and integration framework for multi-omics longitudinal datasets	An integrative framework for building multi-omics networks from longitudinal datasets. It consists of multi-omics kinetic clustering and multi-layer network-based analysis. The method is based on the modeling and clustering of expression profiles with similar behaviours using the timeOmics [150] approach.	1. Performs network reconstruction 2. Perform over-representation analysis.	Infers correlations between molecules based on multi-omics data	Uses experimental measurements as node scores.	Identify crosstalk, key biological functions, or mechanisms	Random walk	Metabolites, genes, protein abundance, mRNA	Requires large-scale and integrated multi-layer networks to improve prediction performance, which in turn requires computational power.	2020	[151]
Random Walk with Restart for multi-dimensional data Fusion (RWRF)	The method uses a similarity network of samples as the basis for integration.	1. Construct a similarity network for each data type 2. Fuse similarity networks 3. Performs random walk with restart on the multiplex network. 4. Performs network clustering	Edge weight is estimated by calculating the similarity measure.	Estimate stationary probability distribution which indicates similarity between the seed node and other nodes.	Identify disease subtypes	Random walk with restart	mRNA, DNA methylation, microRNA	It is computationaly complex and intensive to perform random walk using all networks nodes as seeds in an iterative manner	2021	[107]
Causality- and network-based inference tools										
Differential network analysis in genomics (DINGO)	DINGO is a pathway-based model for estimating patient group-specific networks and making inferences on differential network activation between patient-specific groups. DINGO jointly estimates the group-specific conditional dependencies by decomposing them into global and group-specific components.	1. Estimates global component, which represents the relations common to both patient-specific groups. 2. Estimates local group-specific component which represents the differential unique relations in each patient-specific group. 3. Determines significant differential edges.	Constructs differential scores for group-specific edges.	The vertices are ordered by their degree centrality.	Predicts driver genes.	Differential network approach	mRNA, DNA copy number, DNA methylation, microRNA	Limited to analyze data arising from a single platform, and modeling each of the multiple 'omics data independently does not account for the hierarchical structure of the data	2015	[152]

Permutation-based Causal Inference Algorithms with Interventions	The non-parametric algorithm is used to learn directed acyclic graphs comprising both observational and interventional data. An example is the greedy sparsest permutation algorithm	1. Generate an interventional distribution 2. Search for a permutation 3. Learn from interventions	Estimates edge weight	Utilizes experimental measurements of features	Allows for inference of causal graphs	Unsupervised ML	Multi-modal data (omics, clinical data)	Unable to allow for latent confounders.	2017	[153]
iDINGO	iDINGO R package is an expansion of DINGO. The package estimates group-specific dependencies between different omics data and make inferences on the integrative differential networks, considering the biological hierarchy among the omics platforms. It integrates omics data using the chain graph model.	1. Integrate ordered data platforms using the chain graph model 2. Constructs differential scores for group-specific edges to determine the significant differential edges	Constructs differential scores for group-specific edges	The vertices are ordered by degrees (number of connections)	Predicts hub omics features characterized by the number of differential edges	Differential network approach	mRNA, DNA copy number, DNA methylation, microRNA	Difficulty in incorporating unpaired data.	2018	[154]
prior incorporation on Mixed Graphical Model (piMGM)	Can learn with accuracy the structure of probabilistic graphs over mixed data by appropriately incorporating priors from multiple sources. Identifies gene pathways associated with disease subtype	1. Incorporates prior information from multiple sources 2. Score the reliability of prior information by using a weighted scheme. 3. Merge prior information into a single prior distribution for each edge. 4. Learning the structure of probabilistic graphs 5. Uses separate regularization parameters for edges with and without priors. Determine active pathways	Leverage conditional dependencies to estimate the strength of edges.	Utilizes experimental measurements of features	Identify disease subtypes, active pathways in healthy and disease samples.	Probabilistic graphical model	Multi-modal data (omics, clinical data)	Its performance increases with increasing amounts of prior information	2018	[152, 155]
CausalMGM	A method for learning a causal graph over variables of mixed type linked to disease diagnosis and progression.	1. Learn the undirected graph over mixed data types.	Leverage conditional dependencies to estimate the	Leverages on measurements of variables	Identify causal pathways, biomarkers	Probabilistic graphical model	Multi-modal data (omics, clinical data)	Difficulty in recovering edges or directions involving	2019	[156]

		2. Perform local directionality determinations with conditional independence tests	strength of edges.		, and patient stratificatio n			categorical variables in high-dimensional settings.		
Multi-Omic inTegrative Analysis (MOTA)	A network-based method that uses data acquired at multiple layers from the same set of samples to rank candidate disease biomarkers.	1. Builds a differential network 2. Computes partial correlation between node pairs using graphical LASSO 3. Calculates the differential partial correlation to determine intra-omics connections for the network.	The weight of edges represents the partial correlation (above threshold) between node pairs.	Computes an activity score (MOTA Score) for each node based on its p-value and its connected nodes.	Predicts driver genes or disease biomarkers .	Differential network approach	mRNA, metabolite, glycomics data, proteins	MOTA scores under the assumption that strong candidates tend to be differentially expressed and be surrounded by differentially expressed neighbors. This might not be the case for all instances.	2020	[154, 157]
Integrative multi-omics network-based approach (IMNA)	An integrative multi-omics framework for regulatory network analysis.	1. SNP-gene mapping pairs collection 2. Construct SNP-gene bipartite network 3. Construct a functional interaction network 4. Computes signature score for nodes in the network 5. Computes composite score to provide quantitative evidence of node to evaluate the importance of the regulatory function 6. Perform key driver analysis on tissue-specific gene interaction networks	Uses the confidence scores of the tissue-specific functional interactions as edge weight.	Computes signature scores for each node from different networks. Signature scores for a gene from different networks are combined and normalized to get a composite score for each gene.	Identifies tissue-specific gene interaction networks and key nodes	Bayesian network approach	GWAS signals, eQTLs, epigenomic regulatory annotations, mRNA, protein interactome, and chromatin long-range interactions	Because it is a GWAS-based method, other potential important genes without genetic support could be missed	2020	[158]
MRPC	An R package that learns causal graphs and allows for inference.	1. Learning the graph skeleton 2. Orienting edges in the skeleton	Incorporates the principle of Mendelian randomization as constraints	Utilizes experimental measurements of features	Allows for inference of causal graphs	Unsupervised ML	Genomic data, mRNA	It remains highly conservative in its inference of such large graphs and tends to infer	2021	[159]

		3. Simulating continuous and discrete data 4. Assessment of inferred graphs	on edge direction.					fewer edges than there should be		
MIMOSA2	An R package and web application metabolic network-based tool for inferring relationships in microbiome-metabolome data	1. Construct a community metabolic model by linking microbiome data features to reference databases. 2. Compute community metabolic potential (CMP) scores for each taxon, sample, and metabolite. 3. Aggregate CMP scores at the community level. 4. Evaluates the relationship between total CMP scores and metabolites by fitting a linear regression model	The method does not take edge directionality or weights into account	Utilizes CMP score for each feature.	Allows inference of microbe-metabolite relationships and predicts disease-associated features	Unsupervised ML	Metabolite, microbiome data	MIMOSA2 currently uses a relatively simple approach to calculate CMP. Therefore, its ability to detect microbiome effects involving large numbers of taxa may be lower.	2022	[160]

2.5 Research questions explored using integrative multi-omics network approaches

2.5.1 Understanding how crosstalk between omics layers impacts a biological process or disease phenotype

A perturbed biological system is characterized by deviations in the behaviour of the molecules (omics data features) causing changes in crosstalk (**Figure 2. 1**). These changes could become apparent in multiple (connected and dependent) omics levels and may represent a wide range of molecular events responsible for disease phenotype or impaired biological processes.

Network-based diffusion/propagation tools (described in **Table 2.1**) offer a framework to identify aberrant omics features (e.g., gene expression, somatic mutations, copy number variations, molecular subnetworks informative of disease subtype) and how their presence and activities within the network induce possible (downstream) changes that might underpin disease phenotype.

In a study to understand the molecular function of SARS-CoV-2 and SARS-CoV proteins and their interaction with the human host, Stukalov et al. [161] profiled the interactomes of both virus groups and investigated the effect of viral infection on the transcriptome, proteome, ubiquitinome, and phosphoproteome of a lung-derived human cell line. Functional analysis of the various biomolecules within a molecular network revealed crosstalk between the cellular processes during perturbations taking place upon infection at different omics layers and pathway levels. The authors [161] implemented the Hierarchical HotNet ND method to explore host-SARS-CoV-2 protein interactions during viral infection and its impact on omics levels and cell lines to understand how that could influence molecular pathways. Importantly, the group observed that the transforming growth factor beta (TGF-β) signalling pathway, known for its involvement in tissue fibrosis as one of the hallmarks of COVID-19 [162], was specifically dysregulated by SARS-CoV-2 ORF8. Further results revealed that autophagy, one of the mechanisms for controlling SARS-CoV-2 replication and monitoring the progression of viral infection [163], was specifically dysregulated by

SARS-CoV-2 ORF3. These findings highlight the biological relevance of crosstalk and the insights it provides to understanding disease mechanisms.

2.5.2 Identifying modules/subnetworks for disease or disease progression prediction/ prognosis

Modular organizations within a network, characterized by clusters of neighbouring nodes highlight features that are functionally related or involved in similar activities within the system. In contrast to identifying (crosstalk of) features informative of disease mechanism, the focus here is on identifying different omics data features that cluster together to inform molecular transitions that describe disease severity level and/or disease subtypes.

Network-based tools that predict disease subtypes or subnetworks informative of a phenotype or a phenotypic group (described in **Table 2.1**) are useful for answering such questions and can help in e.g., estimating survival rates across different patient groups. Tools that implement ML and ND-based methods are useful to identify clusters in a network (see **Table 2.1**). It is noteworthy that the approach or steps, algorithms, and input data types implemented by such tools to predict subnetworks vary (as described in **Table 2.1**). In a recent application of a network-based method to COVID-19 research, Sun et al., [164] employed MEGENA [130], an unsupervised ML method, to perform protein-metabolite-lipid multi-omics network analysis based on the differential co-expression (correlation between pair of omics features) of these omics data features. The network analysis indicated that tryptophan metabolism and melatonin, a metabolite related to tryptophan metabolism may contribute to molecular transitions in critical COVID-19 patients. Studies have shown that tryptophan and melatonin can improve the immune system and reduce inflammation in COVID-19, suggesting that function disorder may cause impairment to tryptophan metabolism and immune response [165, 166]. Interestingly, activation of tryptophan metabolism has been clinically shown to be selectively enhanced in severe patients [167]. The authors further identified pathologically relevant lipid modules that are being altered among mild COVID-19 patients.

Interestingly, connections between clusters/modules in the omics data may explain the crosstalk of biological features that are specific to the disease state and may serve

as biomarkers for monitoring disease progression, treatment, and management [26, 69, 168].

2.5.3 Identifying candidate drivers of disease mechanisms

The contributory effect of features (nodes) within a system varies and depends on factors including but not limited to the level of feature expression or abundance, the level of interaction with other features, and the (background) state of the system. While some of these omics data features are passive (i.e., have little or no effect on system stability), others may have a significant effect on the observed phenotype.

In many biological disease-related problems, exploring relationships between multi-omics data extends beyond measuring marginal associations between features. Thus, identifying biologically relevant nodes that influence changes within the system could serve as candidate disease-related nodes responsible for an underlying phenotype [142]. Causal and network inference methods described in **Table 2.1** can be implemented to explore likely causal features, potential causal relationships, and infer networks that differentiate severe disease from mild in a multi-modal network. Although causal methods provide insights into likely causal agents, investigating and confirming true causality extends beyond computational analysis to experimental validation in relevant models. Also, ML and diffusion-based methods can be used to explore candidate drivers. We describe in **Table 2.1** some network-based tools that predict candidate disease-related nodes. In a recent COVID-19-related study, Tomazou et al. [169] implemented a network-based multi-omics data integration approach based on a multi-source information super-network scheme (described in **Table 2.1**) to prioritize COVID-19-related genes that could be useful as drug targets. The super network was constructed based on the weighted sum of the pairwise weighted edge vectors (for each pair of features) obtained from different sources. The method then prioritizes genes in the network by calculating a characteristic score known as the Multi-source Information Gain (MIG). Some of the genes identified by the authors include Serum Amyloid A (*SAA1, SAA2, SAA3*) which has been clinically verified as a sensitive biomarker in evaluating the severity and prognosis of COVID-19 [170], C-reactive protein (CRP) clinically shown to be a marker of systemic inflammation associated with adverse outcomes in COVID-19 patients [171], Serine proteinase inhibitor A3 (*SERPINA3*) shown to be a biomarker for COVID-19-related

organ damage (coronary artery disease) and erythropoiesis impairment [172], and vascular cell adhesion molecule (*VCAM1*) shown to be a vascular and inflammatory implicated in the inflammatory response to sever COVID-19 [173].

2.5.4 Drug discovery

Network-based methods that employ systematic integration of disease-specific omics profiles coupled with drug-related data (e.g., FDA-approved, experimental drugs, drug-target interactions) into a heterogeneous network have been shown to provide answers to biological questions related to drug development [174-176]. In this type of network analysis, nodes could represent both omics data features and non-omics data features such as drugs, diseases, and drug targets. The edges represent the functional association between the data types such as pharmacological or phenotypic information.

The network-based view of drug discovery and development may involve multiple methods or tools at different steps. ND and ML methods have been widely implemented in this research area to make predictions [169, 174]. Predictions from such methods present an effective way to complement experimental methods with the aim of, (1) identifying drug targets, (2) understanding the disease-drug relationship, (3) investigating drug-target interactions, (4) identifying potential drug candidates, (5) drug response prediction, (6) drug-drug relations, and (7) predict effective drug combinations. Of note, driver nodes or subnetworks as predicted by tools described in **Table 2.1** might also inform on drug targets. An interesting application of network-based methods for drug discovery is the COVID-19 study by Tomazou et al. [169], whereby some of the predicted candidate compounds including dexamethasone, atorvastatin, beta-estradiol, cyclosporin-A, imatinib, and remdesivir have been found to generate promising results in clinical trials (https://clinicaltrials.gov/). We describe in **Table 2. 2,** some useful integrative multi-modal network-based tools that are specifically for drug discovery.

Table 2. 2. *Useful network-based integrative multi-omics tools for drug discovery.*

Tool/Method	Description	Major steps of tool	Outcome	Method/Approach	Input data type	Major limitation	Year	Reference
DTINet	A computational pipeline focuses on learning a low-dimensional vector representation of features, which accurately explains the topological properties of individual nodes in the heterogeneous network, and then makes prediction based on these representations via a vector space projection scheme	1. Integrates a variety of drug-related information sources to construct a heterogeneous network. 2. Applies a compact feature learning algorithm to obtain a low-dimensional vector representation of the features. 3. Finds the best projection from drug space onto protein space 4. Infers new drug-target interactions	Drug–target interactions	Unsupervised ML	drug-related information protein-protein interactome	Difficulty in integrating heterogeneous data sources with varying quality and formats	2017	[174]
DrugComboExplorer	A tool for identifying driver signalling pathways and inferring the polypharmacy efficacies and synergy mechanisms through drug functional module-induced regulation of target expression analysis	1. Identify the seed (driver) genes 2. Explore networks from the seed genes by integrating the RNA-seq profiles and pathway knowledge 3. Explore networks from the seed genes by integrating the methylation profiles and pathway data 4. Combine the networks generated from the	Prioritize synergistic drug combinations, Uncover potential mechanisms of drug synergy	Unsupervised ML	DNA sequencing, gene copy number, DNA methylation, RNA-seq data	Difficulty in predicting drug combinations among more than two drugs	2019	[177]

		RNA-seq data and the methylation data						
Reciprocal nearest neighbour and contextual information encoding (RNCE)	A network integration approach accounting for network structure by a reciprocal nearest neighbour and contextual information encoding (RNCE) approach	1. Applies the similarity network fusion (SNF) approach to fuse drug networks. 2. Generate contextual information network 3. Compensate for the contextual information network with the initial SNF network	Predicts drug targets, drug mechanism of action	Unsupervised ML	Pharmacogenomic data such as gene expression data under drug perturbation or drug sensitivity data at the cell-line level	The SNF assumes equal weights to different datasets, and this may not be biologically appropriate since not all datasets characterize or represent the underlying disease pathology to the same extent. This may affect the predictions.	2021	[178]

2.6 Current challenges and recommendations

2.6.1 Design of experiment

The choice of a network-based integration method depends not only on the biological question but also on the experimental design. Certain network-based methods can only deal with paired data (i.e., different omics data measurements from the same biological sample), whereas others can also deal with sparse datasets where there is no or only partial overlap between the samples profiled with the different omics layers. Importantly, the scope of the research will inform the type of data that should be generated. For instance, the paired data, herein referring to different omics data measurements from the same biological sample, is preferred when establishing a holistic picture of systems biology underpinning molecular mechanisms linked to disorders, whereas non-paired data (data generated from different biological samples) is more appropriate for comparative (meta)analysis of samples or omics data measurements. It is therefore recommended to consider the scope of research and the network-based methods that fit.

2.6.2 Reproducibility

Researchers routinely expect that results generated by applying network models are reproducible. For network-based methods, the key issues related to reproducibility are non-harmonized data, biased model evaluation, and lack of transferable code or software. First, multi-omics network-based integration involves the use of heterogeneous data, and some sort of data harmonization is required. A promising approach to harmonize multi-omics research is to ensure that the data comply with FAIR data principles (findability, accessibility, interoperability, and reusability). The data FAIRification process ensures that a (meta)data schema/method that captures relations between (omics) measurements, data structure, and concepts are clearly defined and easily interpretable by both humans and computers. The metadata schema provides information about the omics data structure and facilitates easy mapping of measured features onto persistent identifiers and established biological networks to investigate the connection between network elements [179]. Second, confidence in multi-omics network-based methods requires systematic evaluation and validation of both datasets and models as a prerequisite for benchmarking toward

reproducibility [179]. This approach requires harmonized datasets of quality and quantity that provide unbiased ground truth to ensure that the model at least predicts biologically verified features or edges. Given that there is no gold standard metric for validation, it is critical to validate on a variety of data sources and use metrics that are robust to the level of missing data. Third, to replicate results from previous studies, a detailed report of the analysis together with executable analysis code is important to achieve this purpose. The report and code could be hosted in repositories (e.g., GitHub, Bitbucket, GitLab), reproducible scientific workflow management systems (e.g., Nextflow, Galaxy), environment sharing avenues (e.g., Conda, Docker), or packaged as libraries for programming languages [180]. In addition to the key issues, adapting general best practices in the computational analysis will aid reproducibility.

2.6.3 Heterogeneity

Heterogeneity (a measure of variation) of multi-omics datasets, characterized by diverse data sources, data types, and data structure results in computational complexity, analysis bias, and hampers a robust and reproducible integrative network analysis [181]. There is an increasing awareness of controlling heterogeneity across multi-omics integrative analysis, but most of them are focused on paired data rather than non-paired data. In the context of network-based integrative analysis developing models and algorithms that could account for non-uniformity by identifying the most robust signals encompassing data, heterogeneity is important. This could be in the form of variable selection models to identify important covariates with the strength of multiple datasets, and yet maintain the flexibility of variable selection between the datasets to account for the data heterogeneity [182].

2.6.4 (Biological) Interpretation of results

Interpreting results from an integrative multi-omics analysis is a process of disentangling multiple functional relationships. Primarily, the systematic interpretation of results depends on the kind of biological question and the type of omics measurements used for the analysis. Different omics technologies may have different levels of completeness and sensitivity in terms of detecting biological features. This might result in some omics data types containing more information than others as well as impact the results significantly [183]. It is important to consider the inherent

relationship between the omics profiles used during the interpretation of the results. More often functional annotation of features is based on generalized information which allows a less comprehensive understanding of the molecular mechanisms underlying a phenotype. For this reason, incorporating relevant contextualized pathway information (e.g., tissue-specific or cell-specific) in the analysis has been useful in assessing the functional relevance of nodes and subnetworks on the disease/phenotypic landscape, thereby facilitating interpretation.

The capacity to interpret predicted features and interactions of known biological relevance may take the form of deductive reasoning or semantic similarities to support a hypothesis [184]. In the context of algorithms, robust node weighting and edge weighting metrics measured based on known evidence (e.g., text mining, contextualized pathway information) is important to make an inference that is potentially biologically grounded and experimentally confirmable, knowing that the association between omics layers extends from one-to-one and one-to-many to many-to-many.

2.6.5 Sparsity

There is sparsity at the sample level (not all samples have been profiled with the same assays) and at the feature level. The latter is far more prominent in metabolomics and proteomics than in DNA and RNA sequencing. This is mainly due to the selection of peaks (intensities observed in MS1 survey scans) for fragmentation by data-dependent acquisition (DDA) or data-independent acquisition (DIA) tandem mass spectrometry (LC-MS/MS) approaches [185, 186]. Typically, an ideal acquisition mode ought to produce spectra of high quality for as many of the ions present in the sample as possible, however, that is not the case, resulting in sparsity at the feature level. This issue is partly but not completely resolved in the newer DIA and integrated DDA-DIA modes which operate in a less-selective manner and have higher coverage as compared to the older DDA mode [185, 187].

Another contributing factor to sparsity in omics data in tandem with omics technologies is the absence of accumulation of a molecule to a detectable level by omics platforms (evidenced even across platforms of the same omics technology (*e.g.,* next-generation RNA and DNA sequencing). This is partly associated with experimental design, poor biological sample quality, and sample processing.

For computational analysis purposes, imputation can be used to solve missing value problems; however, imputation does not apply to all omics data types [188]. In addition to imputation, sample similarity measurement methods such as the matrix calibration [189] and the Mahalanobis distance approach [190] could be useful to extrapolate for missing values, however, these methods are also limited to specific omics data types. Thus, a feature may have values only in a small percentage of samples leading to sparse matrices, where features may have a wide variety of distributions. Some multi-omics data integration methods can handle sparse data and also feature reduction methods; however, skewed estimates might result in a biased interpretation of results [191]. To address the issue of sparsity in the context of networks, network integration aggregates independent data sources to form a more comprehensive attributed interactome, where the edges are qualified by specific semantic relations or similarity correlation, and the level of confidence in the node pair relationship based on evidence from similarity scores, literature, and graph databases [184]. Also, incorporating autoencoders, a deep learning approach, and its denoising and variational variants autoencoders (e.g., sparse autoencoders) have been used to address this issue in graph neural networks [192]. Autoencoders learn a representation of the data from the input layer, enforce sparsity constraints, and try to reproduce it at the output layer. During this process, the model can learn from incomplete data and generate new plausible values for the imputation [193].

2.7 Future directions

An area of prospect for integrative multi-omics network-based research, which remains an important opportunity, is making efforts to limit the challenges linked with network-based multi-omics integration in the context of heterogeneity, reproducibility, sparsity, and interpretation of results as discussed above. Another area of importance is building hybrid integrative models that are capable of handling paired and non-paired omics data, as well as other biomedical data. Furthermore, efforts to develop a framework tool or metadata schema that standardizes or harmonizes various multi-omics approaches for data integration could be useful. For example, such a framework may leverage an optimized approach to weigh and prioritize genes, pathways, biological processes, drug targets, and relationships between various other biological features from the multi-omics datasets. However, such framework tools will also

require the expertise of domain experts, as well as the detailed and uniform characterization of statistical and technical attributes of the data [179].

2.8 Discussion

Network-based integrative multi-omics analysis offers the opportunity to elucidate interactions that can occur among all classes of molecules in a biological system as well as information flow between and within multiple omics levels. In addition, it potentially provides substantial improvement of biological understanding by helping in the interpretation of results, as compared to single omics analysis, although collecting multi-omics data from different sources does not guarantee that it will be possible to learn about (all of) the relationships present.

Various graph-based multi-omics methods have been developed for network analysis; however, their application is dependent on the scope of the research question of interest and the (omics) data types available. Consequently, this will inform the choice of an integrative analytical approach and tools. The network-based methods discussed use different scoring metrics, algorithms, and data types which together translate into a comprehensive data source/graph to be employed for interpretation into biological knowledge. The overview and description of the tools for network-based integrative analysis (**Table 2.1**) show that different approaches can be implemented in different ways to achieve similar results. Additionally, the classification of tools (**Figure 2.4**) highlights that some tools can be applied to more than one research question. However, due to the difference in approaches of these methods, we recommend the use of multiple analytical and methodological approaches during integrative data analysis, to compare and validate the study results in different ways before interpretation for further downstream tests or follow-up studies.

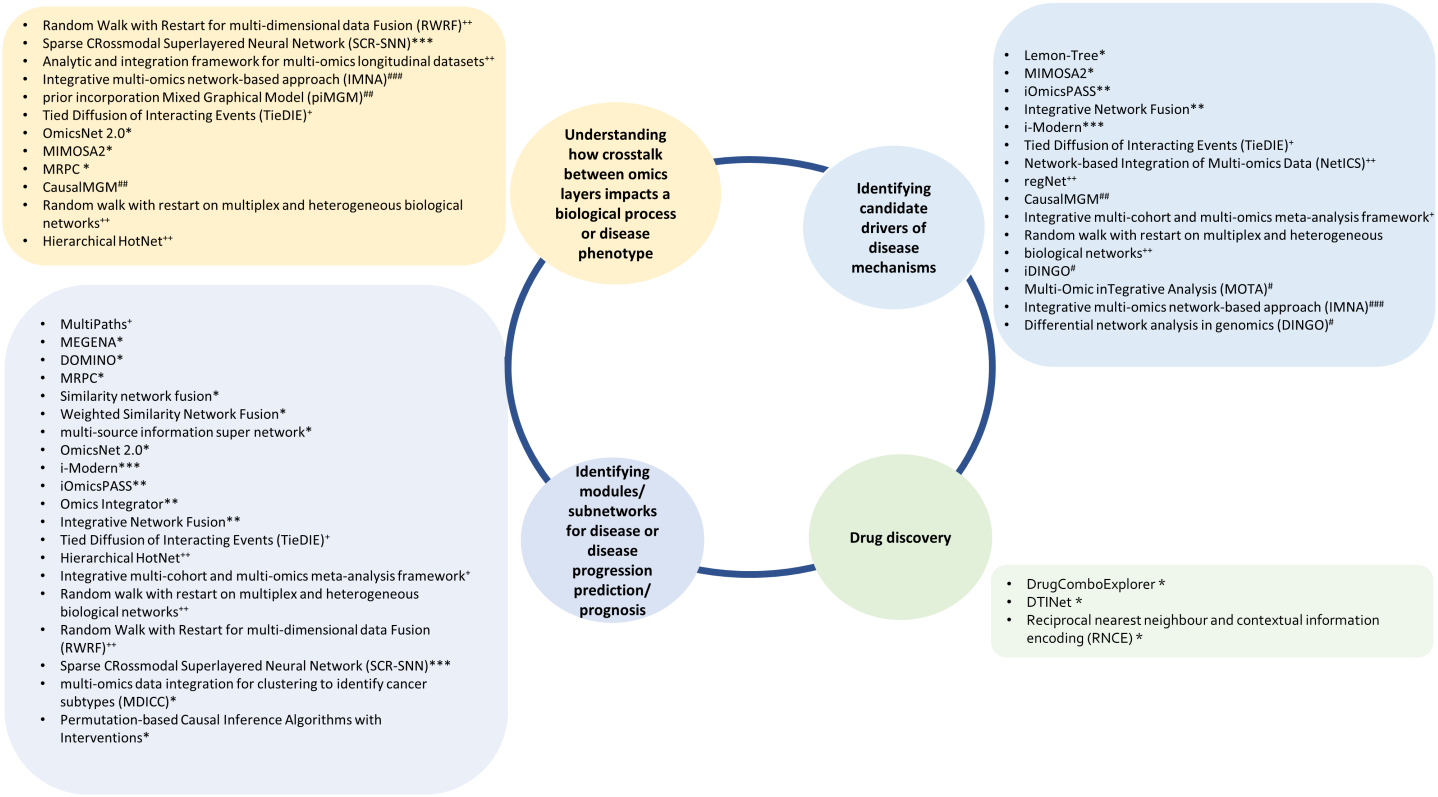


Figure 2.4. *Overview of the discussed network-based multi-omics integrative tools and research questions (in the circle) that they can be applied to. The tools implement different methods including unsupervised machine learning (*), supervised machine learning (**), neural networks (***), diffusion-based (+), random walk (++), differential network (#), probabilistic graphical model (##) and Bayesian methods (###).*

2.9 Contribution to knowledge and science

Network-based integrative approaches have revolutionized multi-omics analysis by providing the framework to conceptualize omics-layered interactions as a graph. Current concepts and approaches for network-based integration are not well-established for multi-omics data analysis. In this chapter, we explain the concepts behind different flavours of network-based integrative approaches and classify them according to the type of omics data supported, the algorithm implemented, and their topological network proprieties. We show how these approaches can be used to identify biomarkers, disease subtypes, crosstalk, causality, and molecular drivers of physiological and pathological mechanisms. This chapter concludes with an overview of challenges associated with multi-omics network-based analysis, such as reproducibility, heterogeneity, and the (biological) interpretability of the results, and we highlight some future directions for network-based integration. Our main contribution to the scientific community is by guiding the readers in selecting the most appropriate

methods and tools and illustrating these with examples from recent studies on the aetiology and treatment of COVID-19 that have been addressed by multi-omics data integration. This chapter remains timely, informative, and insightful and will be of immediate interest to a broad range of readers in the scientific community.

Chapter 3

Network-based integrative multi-omics approach reveals biosignatures specific to COVID-19 disease phases

Francis E. Agamah[1], Thomas H.A. Ederveen[2], Michelle Skelton[1], Darren P. Martin[1], Emile R. Chimusa[3*], Peter A.C. 't Hoen[2*]

[1]Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa.
[2]Department of Medical BioSciences, Radboud University Medical Center Nijmegen, The Netherlands.
[3]Department of Applied Science, Faculty of Health and Life Sciences, Northumbria University, Newcastle, Tyne and Wear, NE1 8ST, UK.

doi: 10.3389/fmolb.2024.1393240.

Abstract

Background: COVID-19 disease is characterized by a spectrum of disease phases (mild, moderate, and severe). Each disease phase is marked by changes in omics profiles with corresponding changes in the expression of features (biosignatures). However, integrative analysis of multiple omics data from different experiments across studies to investigate biosignatures at various disease phases is limited. Exploring an integrative multi-omics profile analysis through a network approach could be used to determine biosignatures associated with specific disease phases and enable the examination of the relationships between the biosignatures.

Aim: To identify and characterize biosignatures underlying various COVID-19 disease phases in an integrative multi-omics data analysis.

Method: We leveraged a multi-omics network-based approach to integrate transcriptomics, metabolomics, proteomics, and lipidomics data. The World Health Organization (WHO) Ordinal Scale (WOS) was used as a disease severity reference to harmonize COVID-19 patient metadata across two studies with independent data. A unified COVID-19 knowledge graph was constructed by assembling a disease-specific interactome from the literature and databases. Disease-state specific omics-graphs were constructed by integrating multi-omics data with the unified COVID-19 knowledge graph. We expanded on the network layers of multiXrank, a random walk with restart on multilayer network algorithm, to explore disease state omics-specific graphs and perform enrichment analysis.

Results: Network analysis revealed the biosignatures involved in inducing chemokines and inflammatory responses as hubs in the severe and moderate disease phases. We observed distinct biosignatures between severe and moderate disease phases as compared to mild-moderate and mild-severe disease phases. Mild COVID-19 cases were characterized by a unique biosignature comprising C-C Motif Chemokine Ligand 4 (*CCL4*), and Interferon Regulatory Factor 1 (*IRF1*). Hepatocyte Growth Factor (HGF), Matrix Metallopeptidase 12 (*MMP12*), Interleukin 10 (*IL-10*), Nuclear Factor Kappa B Subunit 1 (*NFKB1*), and suberoylcarnitine form hubs in the omics network that characterizes the moderate disease state. The severe cases were marked by biosignatures such as *STAT1*, *SOD2*, *HGF,* taurine, lysophosphatidylcholine, diacylglycerol, triglycerides, and sphingomyelin that characterize the disease state.

Conclusion: This study identified both biosignatures of different omics types enriched in disease-related pathways and their associated interactions (such as protein-protein, protein-transcript, protein-metabolite, transcript-metabolite, and lipid-lipid interactions) that are unique to mild, moderate, and severe COVID-19 disease states. These biosignatures include molecular features that underlie the observed clinical heterogeneity of COVID-19 and emphasize the need for disease-phase-specific treatment strategies. The approach implemented here can be used to find associations between transcripts, proteins, lipids, and metabolites in other diseases.

3.1 Background

Coronavirus Disease-2019 (COVID-19) is a contagious respiratory disorder caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), a newly emerged β coronavirus belonging to the *Coronaviridae* family [23]. Since its discovery in Wuhan, China, in December 2019, COVID-19 established itself as a devastating global pandemic that has created disruptions across healthcare, economic, and social systems [5].

COVID-19 is characterized by a range of clinical phenotypes that reflect the spectrum of disease severity (i.e., mild, moderate, and severe herein defined as disease phases). Disease phenotypes are broadly classifiable as asymptomatic and symptomatic with approximately 85% of infected patients (vaccinated and non-vaccinated) showing mild to moderate symptoms and approximately 15% of infected patients suffering from potentially life-threatening complications [194]. The mild to moderate disease phase includes disease conditions with few or no infection symptoms, hospitalization with either no oxygen therapy required or with oxygen given by mask or nasal prongs, and no fatalities. In contrast, the severe disease phase includes disease conditions that, besides death, could involve one or a combination of hospitalization, oxygen therapy involving mechanical ventilation, respiratory failure, and significant immune dysregulation [23].

Each COVID-19 disease phase (i.e., mild, moderate, and severe) is marked by changes in omics profiles with corresponding changes in the expression levels of biosignatures [195-197]. In the context of this research, we define biosignatures as omics features that include proteins, transcripts, lipids, and metabolites. Although some of these biosignatures have a connection to the pathology of COVID-19 illness, not all of them actively impact the expressed disease phenotype when they are dysregulated. Different major dysregulated biosignatures including but not limited to Interferon Alpha 1 (*IFNA1*), Interferon Alpha Inducible Protein 6 (*IFI6*), Toll-Like Receptor 4 (*TLR4*) and interleukin-6 (*IL-6*) are linked to host responses to COVID-19 [44, 46, 68, 70]. Such biosignatures may serve not only as potential biomarkers to stratify patients according to disease severity and/or provide detailed prognostic

information but could also contribute to the development of treatments that are more specifically targeted at particular disease states.

Several individual-omics [77-81, 198-203] and multi-omics COVID-19 investigations [26, 44, 69, 76, 164, 204-207] have identified biosignatures that are associated with disease progression. Individual omics studies provide specific insights into the contributions/manifestations of biosignatures at that omics level during disease progression but have not accounted for the impact(s) of other omics layers. Multi-omics studies present a means to collectively compare multiple omics data from different experiments either on the same samples or across studies, to yield a more holistic understanding of the biochemical underpinnings of COVID-19 outcomes.

However, few of these omics studies [26, 44, 69, 76, 164, 204-206] focusing on identifying the biochemical drivers of COVID-19 clinical heterogeneity have computationally integrated multi-omics data from different study samples with existing biological knowledgebases to explore biosignatures of different omics types and their connections across different disease phases. We suggest that this kind of multi-omics data integration is essential when attempting to explain the molecular dynamics underpinning the heterogeneity of COVID-19 infections while accounting for both prior knowledgebase and data from independent studies.

Network-based integrative approaches have revolutionized multi-omics analyses by providing the framework to build on existing knowledgebases when using new data to infer interactions between multiple different omics profiles within the context of a graph representation [208]. This has been shown to provide the opportunity not only to elucidate interactions that can occur among all classes of biomolecules in a biological system but also, to prioritize biosignatures that could discriminate disease severity. The approach represents the biomolecules that are most indicative of differences between disease states (i.e. the biosignatures) as nodes in the graph and infers relationships between them. For this reason, we hypothesized that (i) investigating biosignatures across different phases of COVID-19 disease will provide insights into the molecular underpinnings of the enormous clinical heterogeneity of COVID-19 and (ii) associations between biosignatures within a biologically meaningful network would

permit the prioritization of biosignatures that discriminate between the disease states and could yield leads for potential drug targets.

Our study implemented a multi-omics network-based approach to identify and characterize biosignatures underlying various COVID-19 disease phases by integrating transcriptomics, metabolomics, proteomics, and lipidomics data (each representing a different data layer) with known interactome data (i.e. our present knowledgebase). We built disease-state specific omics-graphs and applied a network diffusion-based method to predict biosignatures and their associated interactions - both within and between omics layers - that are linked to the various COVID-19 disease phases. This is a form of intermediate-stage multi-omics integration where multi-omics data are integrated simultaneously during the analysis [209].

The major contributions of this work are; (1) a new method of harmonizing patient disease severity metrics by leveraging the WHO Ordinal Scale (WOS) and patient metadata; (2) the assembly of a unified COVID-19 knowledge graph from different curated sources of interactome data (our resent knowledgebase); (3) the construction of a disease-state specific omics-graph by integrating curated transcriptomics, proteomics, lipidomics, and metabolomics datasets from two different multi-omics studies; and (4) identified biosignatures and their associated interactions that are shared and/or unique to mild, moderate, and severe COVID-19 disease-states. Overall, this study identifies biosignatures that discriminate between disease states. For instance, the mild COVID-19 cases were characterized by unique biosignatures comprising C-C Motif Chemokine Ligand 4 (*CCL4*), and Interferon Regulatory Factor 1 (*IRF1*). On the other hand, discriminatory factors such as Hepatocyte Growth Factor (*HGF*), Matrix Metallopeptidase 12 (*MMP12*), Interleukin 10 (*IL-10*), Nuclear Factor Kappa B Subunit 1 (*NFKB1*), and suberoylcarnitine distinguished the moderate disease state. Also in severe cases, biosignatures such as *HGF*, taurine, lysophosphatidylcholine, diacylglycerol, triglycerides, and sphingomyelin. These results suggest that COVID-19 disease severity is influenced by interactions across different omics layers. Importantly, this study's reproducible analysis pipelines offer a valuable tool for identifying biosignatures and interactions not just in COVID-19, but also in other diseases with distinct stages.

3.2 Materials and Methods

3.2.1 Study design and procedures

The approach for this study **(Figure 3.1)** consists of five main steps including: (1) data curation and pre-processing; (2) disease severity harmonization; (3) construction of disease-state specific omics-graphs; (4) multi-layer network-based random walk analysis; and (5) enrichment analysis.

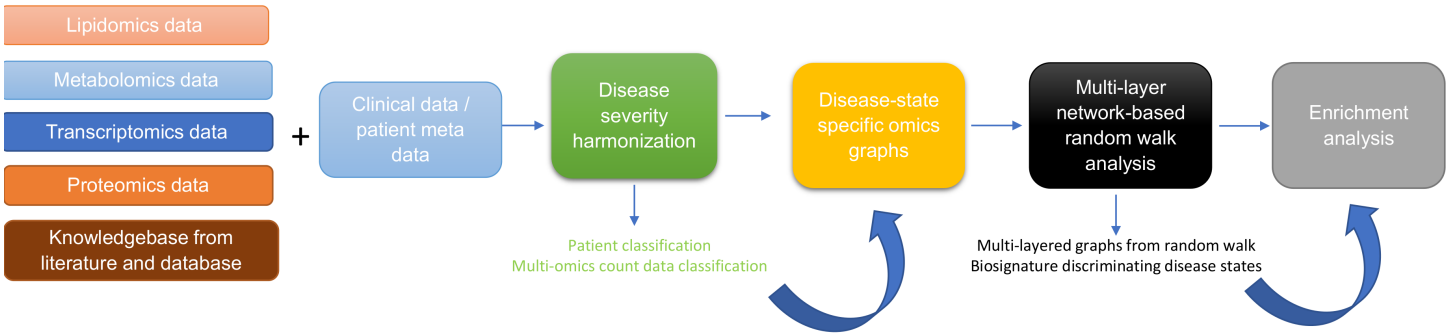


Figure 3.1. *Diagram illustrating the workflow implemented in this study. The workflow begins with curating lipidomics, metabolomics, transcriptomics, proteomics data, and their associated patient metadata and knowledge graph from literature and databases. Next, we leveraged the patient metadata to perform disease severity harmonization. To harmonize the clinical severity of patients, we used the WOS as the reference for classifying disease severity into three disease states, such that: (1) mild disease state represents COVID-19 patients with WOS 1-2, (2) moderate disease state represents COVID-19 patients with WOS 3-4, and (3) severe disease state represents COVID-19 patients with WOS 5-9. We then used the harmonized information to split the omics datasets according to disease severity before constructing co-expression networks and disease-state specific omics-graphs. We then performed random walk analysis on the graphs to predict biosignatures discriminating the various disease states. Finally, we performed enrichment analysis on the proteins, transcripts, metabolites, and lipids.*

3.2.2 Data sources

3.2.2.1 Multi-omics experimental data

Two independent multi-omics experimental datasets were used in this study; quantified blood plasma transcript, protein, and metabolite count data from Su et al., [26] and transcript, protein, metabolite, and lipid count data from Overmyer et al., [69]. For both studies by Su et al., [26] and Overmyer et al., [69], patient recruitment and sample collection occurred in 2020 during the peak of COVID-19 cases. The methodologies utilized for data collection and patient treatment in both studies are different and are outlined in the original studies. For Su et al., [26] the study samples consisted of 139 COVID-19 patients (60 males and 79 females) and 258 healthy

controls (**Supplementary data Table 1**). These enrolled COVID-19 patients had an age range from 18 to 89 years (median = 58). The patients from which blood is drawn were classified as WHO Ordinal Scale (WOS) = 3-4 (n = 83) and WOS = 5-7 (n = 47). The authors used an unpaired Wilcoxon test to determine the statistical difference between WOS = 3-4 and WOS = 5-7, and P values were FDR adjusted. The Overmyer et al., [69] study samples were collected from 128 adult patients of which 102 (64 males and 38 females) were COVID-19-positive and 26 (13 males and 13 females) were negative (**Supplementary data Table 1**). There were no significant differences between the average ages of males and females in either the COVID-19 positive group –(61.3 years for females and 63.1 years for males; (p-value = 0.56; calculated using t-test) or the COVID-19 negative group (59.5 years for females and 67.0 years for males; p-value = 0.25).

3.2.2.2 Protein-protein interactome

We retrieved 1832 names of human genes associated with COVID-19 from the DisGeNET database version 5 [210]. GeneMANIA [211] was used to generate an interactome for 1692 of these genes (140 were unrecognized by the GeneMANIA database). Interactions between genes were based on co-expression, physical interaction, co-localization, shared protein domains, genetic interaction, or predictions from manual curation.

3.2.2.3 Metabolite-metabolite interactome

A co-expression network was constructed from the metabolomics measurement data. Specifically, Pearson's correlations were used to estimate the individual relationships between features for every pre-processed omics measurement using an in-house R script Highly correlated pairwise interaction scores >=0.7 were used to select components of the metabolite-metabolite interactome that was used for downstream analysis. The threshold of 0.7 was determined after carefully evaluating the structural behaviour of the co-expression network and the observed associations at different thresholds.

3.2.2.4 Lipid-lipid interactome

A co-expression network was constructed from the lipidomics measurement data. Similar to the process implemented for metabolite-metabolite interactome above, highly correlated pairwise interaction scores >=0.7 were used to select components of the lipid-lipid interactome that was used for downstream analysis.

3.2.2.5 COVID-19 knowledge graph

We further included a multi-modal, knowledge model of COVID-19 pathophysiology (COVID-19 Knowledge Graph version 0.0.2, https://github.com/covid19kg/covid19kg) [212]. The graph incorporates nodes, covering 10 entity types (e.g. proteins, genes, chemicals, and biological processes) and relationships between the nodes, and we considered only protein, genes, transcripts, lipids, and metabolites node types and their interactions for downstream analysis.

3.2.2.6 Cross-layer interactome

We retrieved protein-transcript, metabolite-protein, and lipid-protein associations from Su et al [26], and Overmyer et al [69], and used these to construct a bipartite graph for network analysis.

3.2.3 Harmonizing the clinical severity of patients

Patient sample metadata from both the Su et al., [26], and Overmyer et al., [69] sources were used for disease severity harmonization. For Su et al., [26] metadata, the disease state linked to the samples was described using the WHO Ordinal Scale (WOS) based on specific categories and characteristics including (1) uninfected – no evidence of infection; (2) infected but ambulatory with no limitation of activities; (3) infected with limitation of activities but still ambulatory; (4) hospitalized with no or mild oxygen therapy; (5) hospitalized with oxygen administered by mask or nasal prongs; (6) hospitalized with non-invasive ventilation or high-flow oxygen; (7) hospitalized with intubation and mechanical ventilation; (8) hospitalized, with intubation and mechanical ventilation together with additional organ support; and (9) death.

In the Overmyer et al., metadata [69], disease severity was quantified using hospital-free days at day 45 (HFD-45) scores: a composite outcome variable that accounts for the length of hospital stay. The utility of the HFD-45 score is derived from the fact that severe COVID-19 patients are those who are admitted to the hospital the longest as they require ventilatory support, while those with the most extreme cases die during hospitalization [69]. The variable assigns a zero value (0-free days) to patients with severe disease who remain admitted longer than 45 days or die due to respiratory deterioration while admitted, and higher values of HFD-45 to patients with shorter hospitalizations and milder disease severity.
To harmonize the clinical severity of patients, we used the WOS as the reference for classifying disease severity into three disease states, such that: (1) mild disease state represents COVID-19 patients with WOS 1-2, (2) moderate disease state represents COVID-19 patients with WOS 3-4, and (3) severe disease state represents COVID-19 patients with WOS 5-9.

The Overmyer et al. [69], sample metadata included the following variables: (1) ICU Status (an indicator variable of the patient's ICU status), (2) HFD-45, (3) Acute Physiologic Assessment and Chronic Health Evaluation II (APACHE II) Score (an indicator variable ranging from 0 (best health) to 71 (worst health) based on physiologic variables, age, and health conditions), and (4) Mechanical Ventilation Status (MVS) (an indicator variable describing the patient's mechanical ventilation status). There was a correlation between the HFD-45 and ICU status with APACHE II and MVS. These variables could feasibly be mapped onto the WOS scale, knowing that the WOS scale is primarily based on respiratory status and oxygen/ventilation support. Accordingly, we leverage this metadata to map characteristics of the Overmyer et al. [69] study patients on the WOS. Specifically, we assigned: a mild disease state to COVID-19 patients with HFD-45 between 29-45 with no time spent in the ICU, a moderate disease state to COVID-19 patients with HFD-45 between 29-45 who spent time in the ICU, or an HFD-45 between 21-28 regardless of time spent in the ICU, and severe disease state to COVID-19 patients with HFD-45 less than 20 regardless of time spent in the ICU.

3.2.4 Data pre-processing

We conducted a two-step data pre-processing operation on the omics experimental data using a custom script. Outlier and missing values were removed, and data were normalized. Samples with more than 20% missing data in a certain data type were excluded. Similarly, biological features such as mRNA expression, with more than 20% of values missing across patients were dropped from the data. Z-score normalization was then applied such that each feature of the data (samples as columns and features as rows) had an average of 0 and a standard deviation of 1 [213].

3.2.5 Feature mapping to unified identifiers.

To ensure that the feature labels were unified, transcript and protein identifiers were mapped to gene-level IDs using the UniProt (https://www.uniprot.org/) and NCBI databases (https://www.ncbi.nlm.nih.gov/gene/). Metabolite and lipid name descriptors were maintained for all analyses other than functional analysis for which KEGG or PubChem IDs were used.

3.2.6 Building a unified knowledge graph.

We assembled a unified knowledge graph by merging the protein-protein interactome, the metabolite-metabolite interactome, the lipid-lipid interactome, and the extracted data from the COVID-19 knowledge graph using a custom script [212].

3.2.7 Building disease-state specific omics-graphs

Protein-protein, transcript-transcript, lipid-lipid, and metabolite-metabolite co-expression networks for the various COVID-19 disease states (i.e, mild, moderate, severe) were constructed based on an integrated unified knowledge graph, and the pre-processed omics data using an R script. From the Su et al., [26] omics data, we constructed three co-expression networks (protein-protein, transcript-transcript, and metabolite-metabolite) for each disease state. Likewise, we constructed four co-expression networks (protein-protein, transcript-transcript, lipid-lipid, and metabolite-metabolite) for each disease state from the Overmyer et al., [69] data. This was achieved by evaluating the correlation between the expression of each linked feature pair where each feature is represented in the unified knowledge graph (i.e., transcript-

transcript, protein-protein, metabolite-metabolite, lipid-lipid feature pairs). Given an edge $e(i,j)$ in the unified knowledge graph (G) which linked feature i and feature j, let $x_i = (x_{i1},\ldots,x_{in})$ and $x_j = (x_{j1},\ldots,x_{jn})$ be the vectors of values in all samples for feature i and j respectively. Pearson's correlations were used to estimate the individual relationships between features for every pre-processed omics measurement using an in-house R script.

The interaction between feature pairs is represented by the Pearson Correlation Coefficient (PCC) ranging between −1 (perfect negative correlation) and 1 (perfect positive correlation), with values of zero representing no correlation. We further rescaled the PCC score for each pairwise interaction by computing the absolute value to attain positive scores ranging between 0 and 1. Additionally, we defined a threshold of 0.1 for edge filtering after careful consideration of the results at different threshold points. We suggest that these steps contributed significantly to addressing the potential for spurious findings and contributed significantly to the reliability of our findings. The co-expression networks constructed from each single omics dataset formed the baseline for the network integration. Specifically, the co-expression networks of the same omics data type constructed from the two independent studies were integrated by merging the networks to construct four omics-specific graphs (one for each omics type) for each of the three disease states.

3.2.8 Random walk network analysis

3.2.8.1 COVID-19 disease state graph exploration by a random walk with restart

We adapted multiXrank [214], a random walk with restart on a multilayer network algorithm to explore the disease-state specific omics-graphs. This algorithm was chosen because it enables random walk with restart on any kind of multilayer network generated from different data sources as compared to other methods that are limited in the combination and heterogeneity of networks that they can handle [208]. We modified the configuration script for the algorithm to accept four disease-state specific omics-graphs for our analysis.

For network exploration on each disease state, the disease-state specific omics-graphs, cross-layer interactome, and seed nodes were used as inputs for the algorithm. Outputs were multi-layered graphs that described the exploration of the

seed nodes across the different disease-state specific omics-graphs and a list of features in each disease-state specific omics-graph ranked according to their proximity to seed nodes.

The parameter values for global restart probability (set to 0.7), and inter-layer jump probability in a given disease-state specific omics-graph (set to 0.5), were maintained. The probability to restart in a specific layer of a specific disease-state specific omics-graph was set to 1: a setting that meant the disease-state specific omics-graph was classifiable as a monoplex network.
The probability of restarting in a specific disease-state specific omics-graph was set to 0. This meant that the random walker stayed within the network within which it began with a probability equal to 1.
To achieve homogeneous exploration, the initial probability of jumping across different disease-state specific omics-graphs was set to 0.25 in consideration of the four disease-state network layers.
Briefly, the first step of the algorithm is to create adjacency matrices for the input graphs, followed by computing different transition probabilities of the random walk with restart on the graphs. The probabilities are estimated based on the concept that an imaginary particle starts a random walk from the seed node to other nodes in the network. These different transition probabilities describe the walks within a graph and the jumps between graphs. A higher probability score (close to 1) suggests a higher likelihood of walking or jumping between graphs.

3.2.8.2 Identifying seed nodes for multi-layered network exploration

To select seed nodes for the analysis, we implemented two approaches; (1) a data-driven approach where we selected, after merging the different co-expression networks, the features with the highest node integrated centrality score in each omics layer as seeds, and (2) a hypothesis-driven approach where we selected seeds based on their impact on disease severity to test the hypothesis of their differential associations with mild, moderate, and severe COVID-19 disease states**.** Hypothesis-driven has the advantage of bringing the question being investigated into focus by designing the model with a specific biological hypothesis in mind and exploring variations across disease phases while the data-driven, enables a more unbiased and

informed model through computationally intensive use of the data [215, 216]. Although hypothesis- and data-driven modeling approaches are not mutually exclusive, it is worth noting that this diversity is beneficial: most model-building tools and models have a specific and clear role, however at the same time, combining hypothesis- and data-driven approaches in an interoperable way, provide an immense impact on our understanding of the disease phases as modelling and integrating data at different biological scales [215, 216].

For the data-driven approach, the features were ranked by leveraging the node degree, closeness, betweenness, and eigenvector centrality metrics to compute an integrated score (**see Supplementary data equations 1 to 3)** [217]. These centrality metrics provide insight into the importance of a node. For instance, the closeness metric measures how close a node is to all other nodes in the network. A lower closeness centrality score indicates the node is on average closer to other nodes, potentially making it a faster "information hub." The degree metric measures the number of edges (connections) a node has with other nodes. A higher degree means the node has more direct connections, suggesting it might be more influential or receive more information flow. The betweenness metric captures how often a node lies on the shortest path between other pairs of nodes. A higher betweenness centrality score suggests the node acts as a crucial bridge for information flow within the network. The eigenvector metric considers not just the number of connections a node has, but also the "importance" of its neighbours. Nodes with high eigenvector centrality are considered influential due to their connections to other influential nodes

3.2.8.3 Ranking candidate multi-omics features for COVID-19 disease states

All the network nodes were scored and ranked by the algorithm according to their proximity to the seed nodes [214]. The computed score was the geometric mean of the node’s proximity to the seeds.

3.2.9 Enrichment Analysis

3.2.9.1 Metabolite Pathway

Metabolite pathway enrichment analysis was performed using the MetaboAnalyst 5.0 online Pathway Analysis tool [218] (accessed 26 January 2023). Metabolite names

were entered as KEGG IDs, and when necessary, metabolite names were automatically adjusted to match the nomenclature recognized by MetaboAnalyst (e.g., *Hydroxypropanoate* as *hydroxypropanoic acid*). Using high-quality KEGG metabolic pathways as the backend knowledgebase, we used the hypergeometric statistical test to examine the overrepresentation of metabolites predefined in the KEGG pathway present in the queried metabolites. This approach determined whether a particular group of compounds was represented more than expected by chance within the user-uploaded compound list.

3.2.9.2 Lipid Pathway

We conducted lipid pathway enrichment analysis using the LIPEA online Pathway Analysis tool [219] (accessed 26 January 2023). Lipid names were entered as a compound list and, when necessary, these names were adjusted to match the nomenclature recognized by LIPEA. The resource computes an unadjusted p-value on overrepresented pathways, and a p-value adjusted with Bonferroni correction to reduce false positives.

3.2.9.3 Gene Ontology Analysis

Protein/transcript enrichment analysis was performed using the online Enrichr Gene Ontology Resource [220] (accessed 26 January 2023). The resource provides gene-set libraries made of a set of related genes that are associated with a functional concept such as a biological pathway or process [220]. Gene ID identifiers were used as input for the enrichment analyses. The resource computes Fisher's exact test followed by a correction based on a mean rank and standard deviation from the expected rank for each term in each gene-set library [220].

3.3 Results

3.3.1 Harmonized clinical severity between patients' metadata

The clinical severity harmonization was a crucial step in integrating and analysing data from multiple sources and platforms to gain a more comprehensive understanding of biological systems and disease mechanisms. There were (**Figure 3.2**) more severe samples than any other in the Overmyer et al. [69] dataset followed by mild and then moderate samples. For the Su et al., [26] dataset, there were similar numbers of mild

and moderate samples and fewer severe samples. The classified samples (**Supplementary data Table 2**) were used to split the omics counts data into disease states before constructing the disease-state specific omics-graphs.

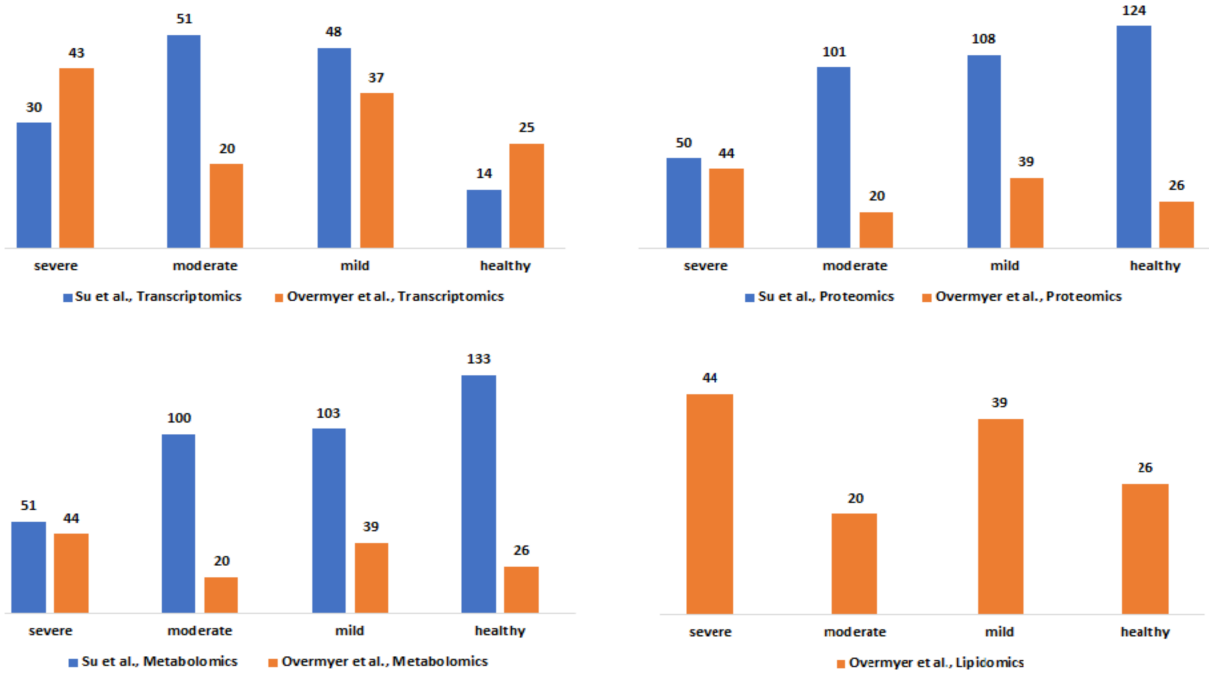


Figure 3.2. *Description of the Su et al., and Overmyer et al., samples based on the omics data type and the disease severity levels after the harmonization process.*

3.3.2 Integrative network-based multi-omics analysis

3.3.2.1 Construction of disease-state specific omics-graphs

We constructed a unified knowledge graph comprising four different edge types (**Figure 3.3A**) by merging the protein-protein interaction data from GeneMANIA, metabolite-metabolite interactome, lipid-lipid interactome, and the extracted data from the COVID-19 knowledge graph. In (**Figure 3.3A**) we provide an overview of the number of edges in the data used to construct the unified knowledge graph. Importantly, the unified knowledge graph formed the basis for integrating the processed multi-omics data (**Figure 3.3B**) and constructing disease-state specific omics-graphs for mild, moderate, and severe COVID-19 disease states (**Supplementary File 3.1**). Additionally, the disease-state specific omics-graphs enabled us to investigate COVID-19 disease states in the context of specific omics data types.

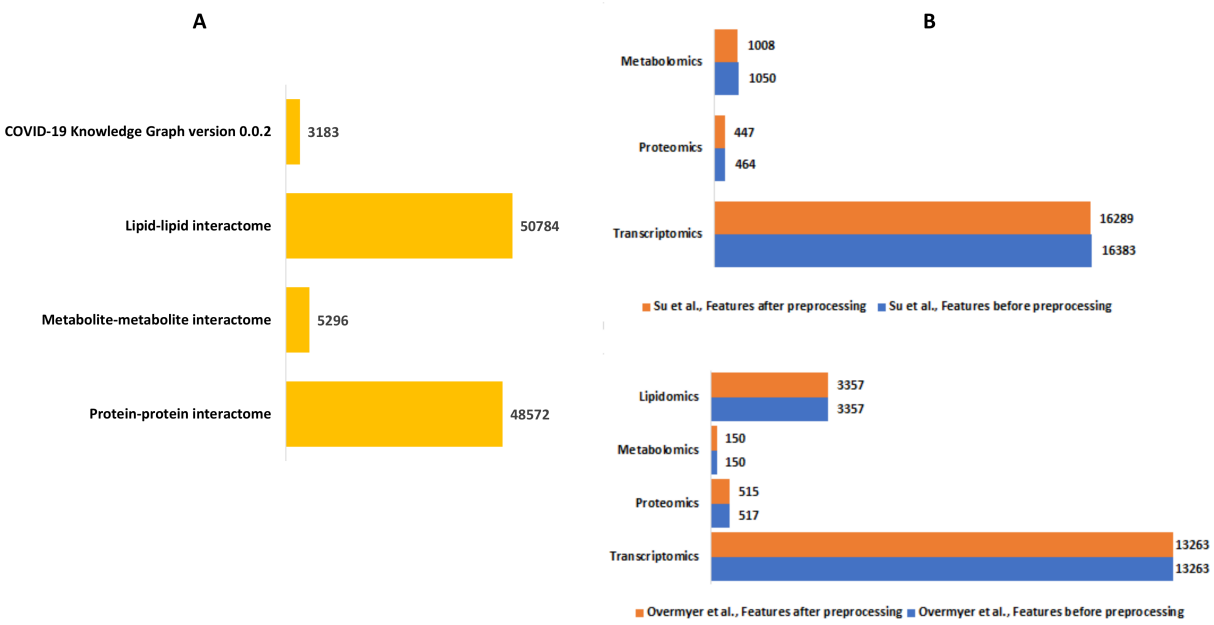


Figure 3.3. *(A) Summary of the edge count in the interactome datasets used to construct the unified knowledge graph. (B) Distribution of features in the omics experimental datasets before and after data processing.*

3.3.2.2 Identified seed nodes for network exploration.

The random walk method is a technique for detecting the flow of biological information throughout networks. The concept behind the random walk method is such that a hypothetical particle exploring the network structure takes random discrete steps (walks) in some direction from a seed node [208]. The walk explores different layers based on the premise that nodes related lie close to each other in the network [149]. We selected the seed nodes at which random walks began (**Table 3.1**) using both data-driven and hypothesis-driven approaches. Whereas for the data-driven approach, we selected a seed primarily based on network topology by computing an integrated node centrality metric score (**Supplementary data**), for the hypothesis-driven approach, we selected nodes based on previously reported associations of molecular features with COVID-19 disease states and to test the differential association these nodes with disease outcomes.

For the data-driven approach, one selected seed node was the Signal Transducer And Activator Of Transcription 1 (*STAT1*) from the transcriptomics layer. *STAT1* is known to be involved in immune responses and antiviral activity [221] and is reported to be upregulated in mild and severe COVID-19 cases, with the phosphorylation of the protein highly enhanced in severe disease states [222]. Another selected seed node

from the proteomics layer was Superoxide Dismutase 2 (*SOD2*), an essential antioxidant enzyme that protects cells from superoxide radical anions which are known to be significantly under-expressed in the plasma [223], and in the lung cells of severely infected COVID-19 patients [224]. From the metabolomics layer, we used 3-hydroxyoctanoate, as a seed: 3-hydroxyoctanoate is a metabolite of medium-chain fatty acid oxidation that has been identified as a marker for primary defects of beta-hydroxy fatty acid metabolism and it is a conjugate acid to 3-hydroxyoctanoate, a biomarker of asymptomatic COVID-19 infection that is involved in important pathways such as the activation of macrophage and platelet aggregation [33]. From the lipidomics layer, we identified unknown_mz_815.61548_+_RT_27.063, an uncharacterized lipid as a seed.

For the hypothesis-driven approach, we selected interleukin-6 *(IL-6)* and interleukin-6 receptor (*IL-6R)* as seeds. Besides the pathologic roles of these molecules in immune-inflammatory diseases such as COVID-19, it has been hypothesized that inhibition of *IL-6* receptors (*IL-6Rs*) by tocilizumab ameliorates the symptoms of severe COVID-19 and reduces mortality [225, 226]. We aimed to find out how *IL-6* and *IL-6R* influence disease severity.

Table 3.1*. Selected seeds for random walk network exploration*

Approach	Seed node	Integrated centrality score	Feature type
Data-driven	STAT1	53529.0403	Transcript
	SOD2	2215.5746	Protein
	3-hydroxyoctanoate	1506.9998	Metabolite
	Unknown_mz_815.61548_+_RT_27.063	9936.9781	Lipid
Hypothesis-driven	IL6R	252.0804	Transcript
	IL6	2208.0288	Protein

3.3.2.3 Random walk analysis on disease-state specific omics-graphs using data-driven seeds.

We used various omics features (**Table 3.1)** as seed nodes for the random walk analysis. Subsequently, features in each disease-state specific omics-graph were ranked by their proximity to the seeds. The generated multi-layered graphs (accessible

at http://cytoscape.h3africa.org), describing the exploration of the seeds during the random walk analysis for each disease state, suggested that cross-layer interactions between the different omics data types influence disease severity. This is evidenced particularly in the protein-metabolite (e.g., *HGF* and 1-palmityl-GPC, *HGF* and 6-bromotryptophan), transcript-metabolite (e.g. *CCL2* and taurine, *CCL2* and 1-(1-enyl-palmitoyl)-2-oleoyl-GPC)), and protein-transcript (e.g. *HGF* and *HLA-B*) interactions. For each omics layer, we identified highly connected features (**Table 3.2**) forming large subnetworks and defined these as "key hubs".

Table 3.2. *Key hubs identified in the disease-state specific omics-graphs upon using seeds from the data-driven approach.*

Disease state	Omics layer	Feature/Hub	Proximity to seeds score
Mild	Transcriptomics	Coagulation Factor XI (*F11*)	0.00555
		C-C Motif Chemokine Ligand 4 (*CCL4*)	0.00548
	Proteomics	Interferon Regulatory Factor 1 (*IRF1*)	0.00115
	Metabolomics	3-hydroxydecanoate	0.02761
		3-hydroxyhexanoate	0.02736
	Lipidomics	Unknown_mz_834.66107_+_RT_26.843	0.00177
		Unknown_mz_838.69214_+_RT_27.059	0.00175
Moderate	Transcriptomics	Hepatocyte Growth Factor (*HGF*)	0.00542
		Matrix Metallopeptidase 12 *(MMP12*)	0.00528
	Proteomics	Interferon Regulatory Factor 1 (*IRF1*)	0.00114
	Metabolomics	3-hydroxydecanoate	0.02761
		3-hydroxyhexanoate	0.02733
	Lipidomics	Unknown_mz_834.66107_+_RT_26.843	0.00177
		Unknown_mz_553.38593_+_RT_19.676	0.00167
Severe	Transcriptomics	Hepatocyte Growth Factor (*HGF*)	0.00589
		Matrix Metallopeptidase 12 (*MMP12*)	0.00568
	Proteomics	Interferon Regulatory Factor 1 (*IRF1*)	0.00118
	Metabolomics	3-hydroxydecanoate	0.02761
		3-hydroxyhexanoate	0.02733
	Lipidomics	Unknown_mz_834.66107_+_RT_26.843	0.00177
		Unknown_mz_553.38593_+_RT_19.676	0.00167
		Unknown_mz_736.64563_+_RT_24.633	0.00167

3.3.2.4 Random walk analysis on disease-state specific omics-graphs using hypothesis-driven seeds.

Like the network exploration using data-driven seeds, we repeated the analysis with hypothesis-driven seeds. We observed from the multi-layer graphs (accessible at http://cytoscape.h3africa.org) that *IL-6* interaction with features of different omics data types increased with disease severity, thus indicating the differential association of *IL-6* with the different disease states. On the other hand, we observed *IL-6R* interactions

mainly with proteins and transcripts to increase with disease severity as compared to interactions with metabolites. The results suggest *IL-6* interaction with proteins (e.g., *IFNB*, *IFIT3*), transcripts (e.g., *CXCL1*, *CXCL2, CCL3*), and metabolites (e.g., 1-(1-enyl-palmitoyl)-GPC, 1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)) may contribute to its major role in disease severity. Similar to analysis using data-driven seeds, we identified key hubs from each omics layer (**Table 3.3**). The hubs with zero proximity scores establish subnetworks with no direct interaction with seed nodes (*IL-6* and *IL-6R*) (as shown in http://cytoscape.h3africa.org).

Table 3.3*. Key hubs identified in the disease-state specific omics-graphs upon using seeds from the hypothesis-driven approach*

Disease state	Omics layer	Feature	Proximity to seeds score
Mild	Transcriptomics	C-C Motif Chemokine Ligand 4 (*CCL4*)	0.00659
		C-C Motif Chemokine Ligand 2 *(CCL2)*	0.005629
	Proteomics	Interleukin 7 Receptor *(IL7R)*	0.00368
	Metabolomics	tricosanoyl sphingomyelin (d18:1/23:0)*	0.0
		suberate (C8-DC)	0.0
		stearoylcholine*	0.0
	Lipidomics	Unknown_mz_765.5752_-_RT_20.235	0.0
		Unknown_mz_765.3349_-_RT_6.822	0.0
		Unknown_mz_765.3349_+_RT_6.842	0.0
Moderate	Transcriptomics	Nuclear Factor Kappa B Subunit 1 *(NFKB1)*	0.00545
		Interleukin 10 *(IL10)*	0.00539
	Proteomics	Interleukin 7 Receptor *(IL7R)*	0.00339
	Metabolomics	sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)*	0.0
		sulfate of piperine metabolite C16H19NO3 (2)*	0.0
		suberoylcarnitine (C8-DC)	0.0
		suberate (C8-DC)	0.0
	Lipidomics	Unknown_mz_768.54962_-_RT_23.304	0.0
		Unknown_mz_763.31421_+_RT_5.276	0.0
Severe	Transcriptomics	C-C Motif Chemokine Ligand 4 *(CCL4)*	0.00612
		C-X-C Motif Chemokine Ligand 1 *(CXCL1)*	0.00549
	Proteomics	Interleukin 7 Receptor *(IL7R)*	0.00358
	Metabolomics	sphingomyelin (d18:2/21:0, d16:2/23:0)*	0.0
		tartronate (hydroxymalonate)	0.0
		sulfate of piperine metabolite C18H21NO3 (1)*	0.0
	Lipidomics	Unknown_mz_765.60468_+_RT_16.383	0.0
		Unknown_mz_766.75885_+_RT_1.234	0.0
		Unknown_mz_766.69232_+_RT_28.305	0.0

3.3.3 Evaluating features and interactions of generated multi-layered graphs

In this section, we dissect different multi-layered graphs generated from random walk analysis to examine common and unique feature interactions related to disease

severity. The multi-layered networks generated for each disease state contain four types of nodes (proteins, transcripts, lipids, and metabolites), and at most six types of edges (transcript-transcript, protein-protein, protein-transcript, protein-metabolite, metabolite-metabolite, lipid-lipid). We did not observe protein-lipid edge types across all the networks. This observation is because the seed exploration prioritizes nodes that have an either direct or indirect connection to the seeds (i.e., related to the seeds), thus no observed protein-lipid edge type is an indication that there was limited bipartite data that captures interactions between nodes and seeds of interest.

3.3.3.1 Evaluating multi-layered graphs generated using data-driven seeds

We evaluated the feature interactions (**Supplementary File 3.2**) present in the multi-layer graphs (accessible at http://cytoscape.h3africa.org) identifying 204 interactions associated with a specific disease state (**Supplementary Figure 3.1A**), of which 79%, 15%, and 6% are transcript-transcript, lipid-lipid, and protein-protein interactions respectively. Of these interactions, most (88%) are associated with the mild disease state (**Supplementary Figure 3.1A**). Additional investigation of 263 interactions associated with only two of the three disease states revealed more pairwise interactions common between the moderate and severe disease states as compared to those in common between the mild-moderate and mild-severe disease state pairs (**Supplementary Figure 3.1B).** Also, an investigation of 397 interactions common across the three disease states revealed a higher proportion of interactions between transcripts associated with various cellular processes and about 1% each for metabolite-metabolite, protein-protein, and protein-metabolite interactions (**Supplementary Figure 3.1C).**

In addition to subnetworks formed by the seed nodes, we observed *CCL4, F11,* and *IRF1* interact directly with seed nodes *SOD2* and *STAT1* and form subnetworks in the multi-layered graph generated for the mild disease state. *CCL4* established interactions with human leucocyte antigen, co-stimulatory molecules (e.g., *CD2, CD4, CD8, CD83, CD53*, *CD3D HLA-DPA1, HLA-DPB1, HLA-DRA*), and other molecules expressed in monocytes and macrophages (e.g., *CCL5, CCL7, CCL8*) to be highly predominant in the mild disease state **(Figure 3.4A).** This observation may suggest a more efficient immune response to the virus among patients with mild disease as

compared to those with moderate and severe disease, thus leading in the latter group to decreased recruitment of immune cells to the site of infection [221].

We identified intra-omics interactions (specifically transcript-transcript and protein-protein) shared between mild-severe and mild-moderate disease state pairs. Notable interactions were among *STAT1,* interferons (e.g., *IFNB1, ISG20*) and *SOD2*. *STAT1* is involved in regulating T-cell activation and differentiation responses, thus regulating the pathogenesis of COVID-19. Therefore, associations between *STAT1* and T-cell receptors (e.g., *CD38, CD40, CD48, CD68*) in the mild-severe disease state, may suggest a more efficient role in the immune response to viral infections during mild disease state and higher *STAT1* activation in severe disease state contributing to the cytokine storm and hyperinflammation that are characteristic of severe COVID-19 [227]. Further, interactions between *STAT1* and interferons highlight the critical role of *STAT1* in the regulation of interferon-stimulated genes, because interferons are key cytokines in the immune response to viral infections. Also, interactions between interferons and other genes (e.g., *TNFRSF25*, *TLR3, MAPK1*) involved in interferon-mediated pathways, highlight the likelihood of innate and adaptive immune stimulatory effects.

A distinctive factor for severe-moderate compared to mild-moderate and mild-severe disease state pairs was the cross-layer interactions between protein-metabolite and protein-transcript. Particularly, Hepatocyte Growth Factor (*HGF*) and Matrix Metallopeptidase 12 (*MMP12*) interactions with metabolites (e.g., 6-bromotryptophan, cortolone glucuronide, 1-palmityl-2-linoleoyl-GPC, 1-(1-enyl-stearoyl)-2-linoleoyl-GPC, sphingomyelin) and proteins implicated in various cellular processes were highly predominant in the moderate disease state **(Figure 3.4B**). On the other hand, *HGF* cross-layer interaction in severe disease state was predominant and not *MMP12* **(Figure 3.4C**) as observed in the moderate disease state.

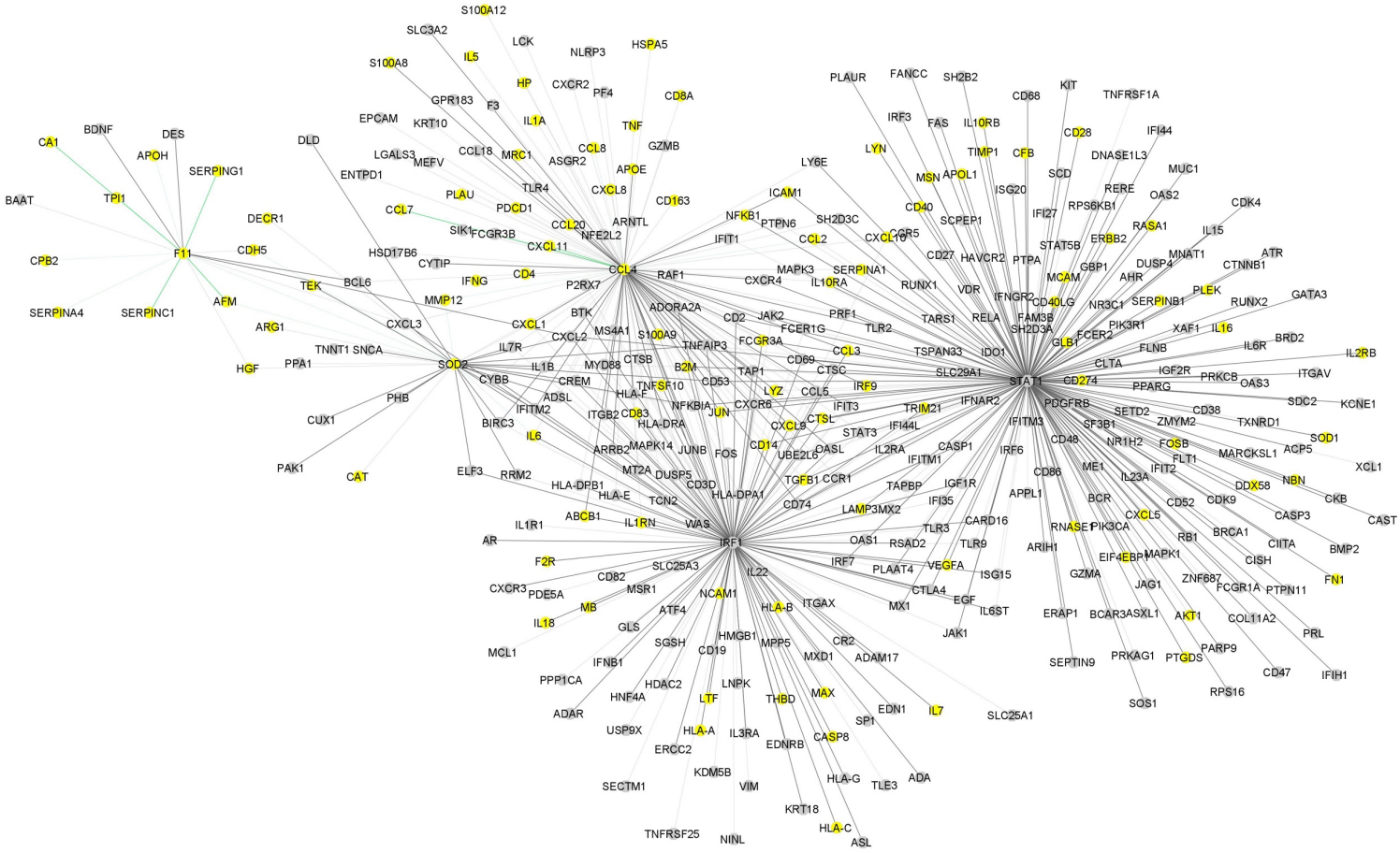
HGF plays a critical role in tissue repair and regeneration, particularly in the liver and lungs, and has been found to have potential therapeutic effects in COVID-19 due to its ability to reduce lung injury and improve pulmonary function. Studies have shown that COVID-19 patients with severe respiratory symptoms have significantly lower levels of *HGF* in their blood and that treatment with *HGF* can reduce lung inflammation

and prevent the progression of COVID-19-related lung injury [228]. *HGF* has also been shown to have antiviral effects against the SARS-CoV-2 virus *in vitro*, suggesting that it may help to inhibit viral replication in infected cells [229, 230]. Thus, the observed interactions may suggest a protective role for *HGF* in ameliorating the progression of COVID-19, as well as a target for drug research [229, 230].

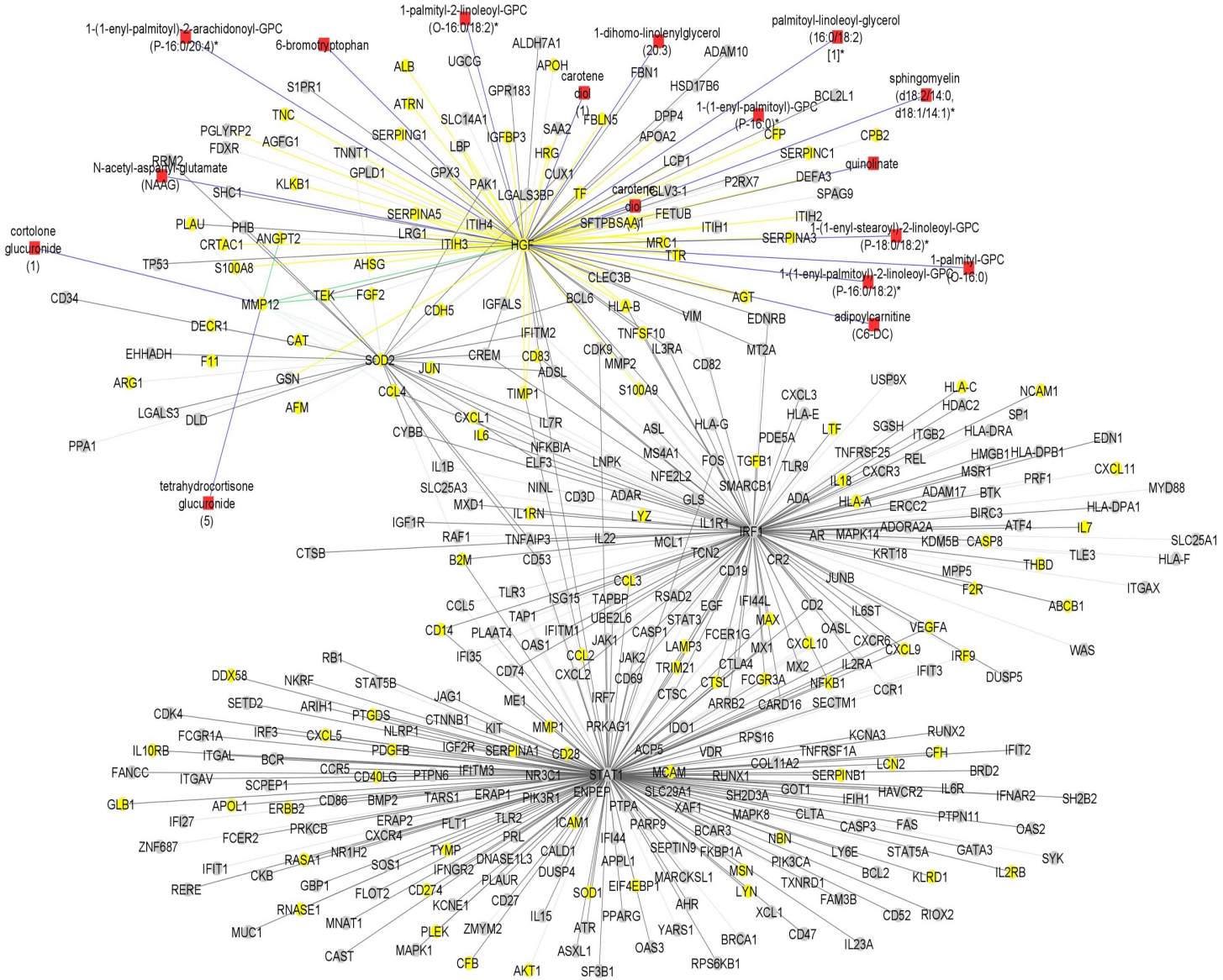
Evidence suggests that both *HGF* and *MMP12* levels are significantly elevated in the lungs of patients with severe disease [231, 232]. Particularly, elevated *MMP12* levels play a role in controlling disease pathogenesis and lung injury, acknowledging that excessively elevated levels can disrupt the balance of the extracellular matrix, resulting in tissue damage [231].

Therefore, the observed cross-layer interactions for *HGF* and *MMP12* in moderate and severe diseases suggest that cross-layer interactions influence clinical heterogeneity, thus influencing the dynamics of disease severity.

A


S100A12
SLC3A2
LCK
NLRP3
HSPA5
S100A8
IL5
HP
CXCR2
PF4
CD8A
GPR183
F3
IL1A
TNF
EPCAM
KRT10
BDNF
DES
CA1
APOH
SERPING1
DLD
LGALS3
MEFV
CCL18
MRC1
ASGR2
CCL8
GZMB
ENTPD1
APOE
TPI1
BAAT
PLAU
TLR4
CXCL8
CD163
DECR1
CCL7
PDCD1
CCL20
ARNTL
SIK1
FCGR3B
NFE2L2
CDH5
CPB2
F11
HSD17B6
CXCL11
CYTIP
CCL4
RAF1
TEK
BCL6
IFNG
CD4
P2RX7
AFM
MMP12
SERPINA4
SERPINC1
ARG1
CXCL3
BTK
ADORA2A
CXCL1
MS4A1
CXCL2
S100A9
TNNT1
SNCA
IL7R
TNFAIP3
HGF
PPA1
IL1B
MYD88
CTSB
B2M
SOD2
CYBB
CREM
TNFSF10
CD53
TAP1
PHB
ADSL
HLA-F
LYZ
CCL5
CUX1
IFITM2
ITGB2
CD83
NFKBIA
JUN
CXCR6
IFIT3
BIRC3
HLA-DRA
CXCL9
CTSL
IL6
ARRB2
MAPK14
JUNB
FOS
CD14
UBE2L6
OASL
PAK1
CAT
ELF3
RRM2
MT2A
DUSP5
CD3D
TGFB1
CCR1
HLA-DPB1
HLA-E
TCN2
HLA-DPA1
CD74
ABCB1
IL1RN
WAS
IL1R1
AR
IRF1
F2R
SLC25A3
IL22
CXCR3
CD82
PDE5A
MSR1
NCAM1
HLA-B
ITGAX
MB
ATF4
IL18
GLS
SGSH
HMGB1
MCL1
CD19
MPP5
CR2
IFNB1
MXD1
ADAM17
PPP1CA
HDAC2
LNPK
HNF4A
LTF
THBD
MAX
ADAR
EDN1
IL7
SP1
USP9X
HLA-A
IL3RA
CASP8
ERCC2
EDNRB
KDM5B
HLA-G
TLE3
ADA
SECTM1
VIM
KRT18
HLA-C
TNFRSF25
NINL
ASL
PLAUR
FANCC
SH2B2
KIT
CD68
TNFRSF1A
IRF3
FAS
IL10RB
CD28
IFI44
LYN
TIMP1
CFB
DNASE1L3
LY6E
MSN
APOL1
SCD
MUC1
ISG20
RERE
OAS2
CD40
IFI27
RPS6KB1
CDK4
SCPEP1
RASA1
IL15
CCR5
CXCL10
ERBB2
CD27
HAVCR2
PTPA
STAT5B
MNAT1
ATR
GBP1
DUSP4
CTNNB1
MCAM
AHR
PLEK
NR3C1
SERPINB1
RUNX2
GATA3
CD40LG
IFNGR2
VDR
RUNX1
SERPINA1
MAPK3
IL10RA
CXCR4
TARS1
RELA
FAM3B
SH2D3A
PIK3R1
XAF1
IL16
BRD2
IL6R
FLNB
IL2RB
GLB1
FCER2
TSPAN33
IDO1
CCL3
FCER1G
TLR2
PRF1
JAK2
CD2
CLTA
IGF2R
PRKCB
ITGAV
SLC29A1
STAT1
CD274
PPARG
OAS3
SDC2
KCNE1
PDGFRB
IFNAR2
IRF9
SETD2
CD38
ZMYM2
TXNRD1
IFITM3
SF3B1
NR1H2
FOSB
SOD1
MARCKSL1
ACP5
CD48
IRF6
IFITM1
IL2RA
CASP1
FLT1
XCL1
IFIT2
IL23A
ME1
CD86
DDX58
NBN
APPL1
IFI35
LAMP3
MX2
IGF1R
TAPBP
CKB
CDK9
CD52
BCR
CASP3
CAST
CXCL5
RNASE1
PIK3CA
CARD16
TLR3
BRCA1
RB1
CIITA
BMP2
ARIH1
TLR9
RSAD2
OAS1
EIF4EBP1
MAPK1
CISH
FN1
GZMA
ISG15
VEGFA
PLAAT4
IRF7
JAG1
ZNF687
FCGR1A
PTPN11
CTLA4
EGF
MX1
IL6ST
ERAP1
BCAR3
ASXL1
AKT1
COL11A2
PRL
JAK1
SEPTIN9
PRKAG1
PTGDS
PARP9
CD47
IFIH1
RPS16
SOS1
SLC25A1

B


1-palmityl-2-linoleoyl-GPC (O-16:0/18:2)*
1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC (P-16:0/20:4)*
6-bromotryptophan
1-dihomo-linolenylglycerol (20:3)
palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*
sphingomyelin (d18:2/14:0, d18:1/14:1)*
1-(1-enyl-palmitoyl)-GPC (P-16:0)*
N-acetyl-aspartyl-glutamate (NAAG)
cortolone glucuronide (1)
quinolinate
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)*
1-palmityl-GPC (O-16:0)
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*
adipoylcarnitine (C6-DC)
tetrahydrocortisone glucuronide (5)
HGF
SOD2
STAT1
IRF1
ALB
UGCG
ALDH7A1
APOH
carotene diol (1)
FBN1
HSD17B6
ADAM10
GPR183
ATRN
SLC14A1
FBLN5
DPP4
BCL2L1
TNC
PGLYRP2
AGFG1
FDXR
SERPING1
IGFBP3
SAA2
APOA2
CFP
CPB2
RRM2
TNNT1
LBP
HRG
SERPINC1
GPLD1
GPX3
CUX1
LCP1
KLKB1
SHC1
PAK1
LGALS3BP
TF
carotene CLV3-1
P2RX7
DEFA3
SPAG9
diol
FETUB
SERPINA5
ITIH4
SFTPBSAA1
ITIH1
ITIH2
PLAU
PHB
LRG1
ITIH3
MRC1
SERPINA3
CRTAC1
ANGPT2
TTR
TP53
S100A8
AHSG
CLEC3B
CD34
MMP12
TEK
FGF2
CDH5
IGFALS
BCL6
HLA-B
AGT
VIM
EDNRB
DECR1
CAT
IFITM2
TNFSF10
CDK9
IL3RA
MT2A
EHHADH
F11
CREM
CD83
MMP2
CD82
ARG1
GSN
JUN
ADSL
CCL4
TIMP1
S100A9
USP9X
CXCL3
HLA-C
NCAM1
HDAC2
LGALS3
DLD
AFM
CXCL1
IL7R
HLA-E
SP1
PPA1
CYBB
IL6
ASL
HLA-G
PDE5A
LTF
SGSH
ITGB2
HLA-DRA
EDN1
NFKBIA
MS4A1
TNFRSF25
REL
HMGB1
HLA-DPB1
IL1B
ELF3
LNPK
FOS
TGFB1
TLR9
CXCR3
MSR1
PRF1
CXCL11
NFE2L2
IL18
SLC25A3
NINL
CD3D
ADAR
SMARCB1
ADA
ADAM17
BTK
MYD88
MXD1
GLS
HLA-A
ERCC2
IL1RN
LYZ
BIRC3
HLA-DPA1
IGF1R
IL1R1
ADORA2A
ATF4
IL7
RAF1
TNFAIP3
IL22
MCL1
AR
MAPK14
KDM5B
CASP8
SLC25A1
CTSB
TCN2
KRT18
TLE3
HLA-F
B2M
CD53
CR2
CD19
THBD
MPP5
JUNB
F2R
ITGAX
TLR3
CCL3
ISG15
TAPBP
RSAD2
EGF
IFI44L
CD2
IL6ST
ABCB1
CCL5
TAP1
UBE2L6
STAT3
MAX
CD14
PLAAT4
IFITM1
CASP1
FCER1G
OASL
VEGFA
OAS1
JAK1
LAMP3
MX1
CXCL10
CXCR6
CXCL9
WAS
IFI35
CCL2
JAK2
CTLA4
MX2
IL2RA
IRF9
RB1
CD74
CXCL2
CD69
TRIM21
IFIT3
DUSP5
DDX58
NKRF
STAT5B
CTSL
FCGR3A
NFKB1
CCR1
IRF7
CTSC
ARRB2
CARD16
SECTM1
SETD2
ARIH1
JAG1
ME1
CDK4
PTGDS
CTNNB1
KIT
MMP1
PRKAG1
IDO1
FCGR1A
IRF3
CXCL5
NLRP1
RPS16
KCNA3
RUNX2
IL10RB
ITGAL
BCR
PDGFB
IGF2R
SERPINA1
CD28
ACP5
VDR
TNFRSF1A
CFH
IFIT2
FANCC
CCR5
IFITM3
NR3C1
MCAM
COL11A2
LCN2
BRD2
ITGAV
CD40LG
PTPN6
RUNX1
SERPINB1
SCPEP1
TARS1
ERAP1
PIK3R1
ENPEP
SLC29A1
SH2D3A
GOT1
HAVCR2
IL6R
IFNAR2
SH2B2
GLB1
APOL1
ERBB2
CD86
BMP2
PTPA
XAF1
IFIH1
IFI27
ERAP2
FLT1
TLR2
ICAM1
MAPK8
CLTA
CASP3
FAS
PTPN11
OAS2
PRKCB
PARP9
BCAR3
FCER2
CXCR4
PRL
NBN
ZNF687
NR1H2
TYMP
DNASE1L3
CALD1
IFI44
SEPTIN9
FKBP1A
LY6E
STAT5A
GATA3
SYK
IFIT1
RASA1
SOS1
IFNGR2
DUSP4
APPL1
PIK3CA
CKB
PLAUR
MARCKSL1
MSN
BCL2
KLRD1
IL2RB
RERE
GBP1
FLOT2
CD274
SOD1
EIF4EBP1
TXNRD1
LYN
FAM3B
RNASE1
KCNE1
CD27
AHR
XCL1
CD52
RIOX2
MUC1
MNAT1
IL15
PPARG
YARS1
CD47
PLEK
ZMYM2
ATR
BRCA1
IL23A
CAST
MAPK1
ASXL1
OAS3
CFB
AKT1
SF3B1
RPS6KB1

C

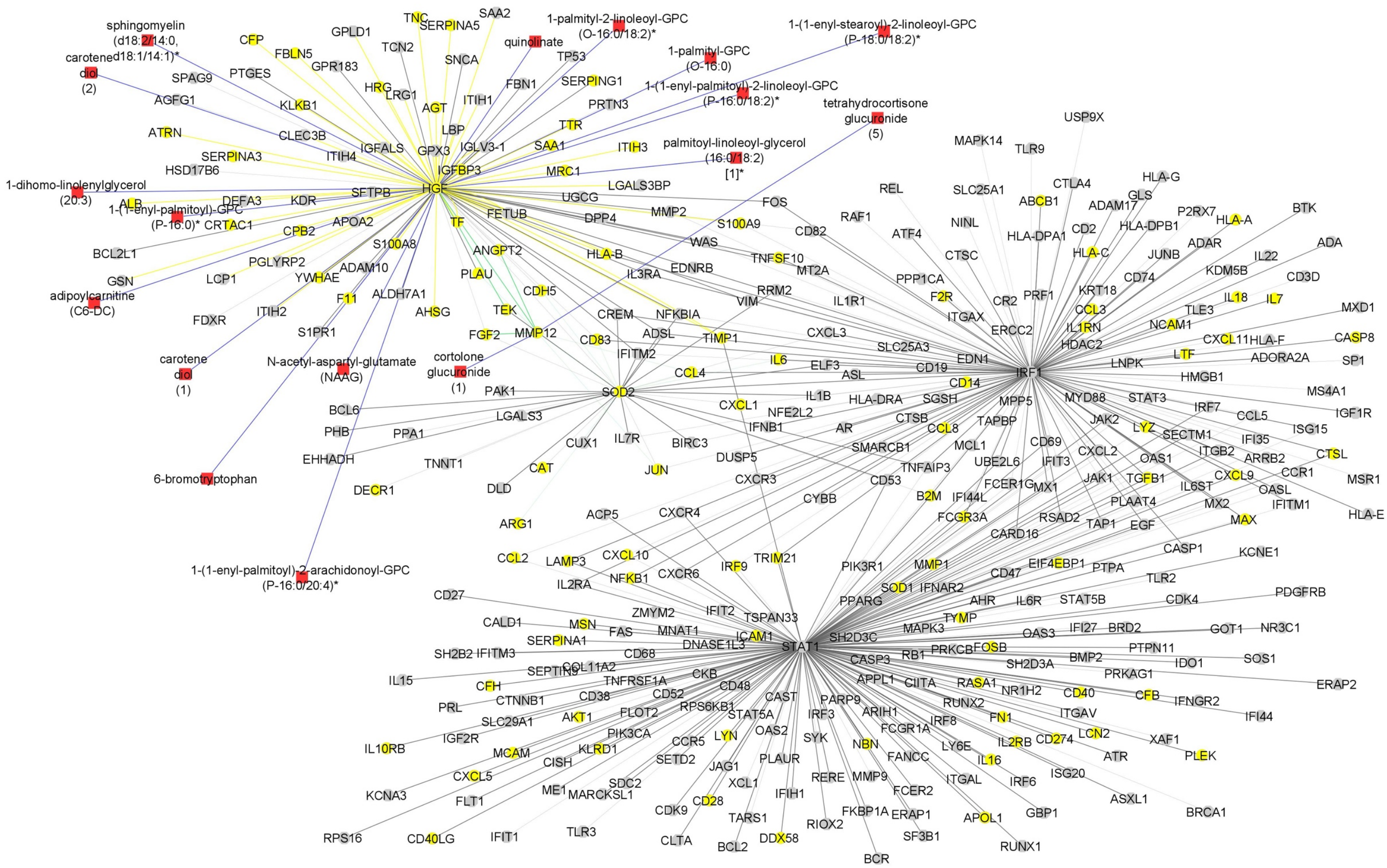


Figure 3.4. *(A) Graph representation of subnetworks formed by hubs CCL4, F11, and IRF1, that establish direct interaction with seed nodes (STAT1 and SOD2) as observed in the multi-layered graph generated for the mild disease state. The graph highlights the interaction of these hubs with other molecular features including proteins (yellow nodes) and transcripts (grey nodes) (B) Graph representation of subnetworks formed by hubs HGF, IRF1, and MMP12, that establish direct interaction with seed nodes (STAT1 and SOD2) as observed in the multi-layered graph generated for the moderate disease state. The graph highlights the interaction of these hubs with other molecular features including proteins (yellow nodes), transcripts (grey nodes), and metabolites (red nodes). (C) Graph representation of subnetworks formed by hubs HGF, and IRF1 that establish direct interaction with seed nodes (STAT1 and SOD2) as observed in the multi-layered graph generated for the severe disease state. The graph highlights the interaction of these hubs with other molecular features including proteins (yellow nodes), transcripts (grey nodes), and metabolites (red nodes). The grey edges represent transcript-transcript interactions, the yellow edges represent protein-protein interactions, the cyan edges represent metabolite-metabolite interactions, the green edges represent protein-transcript interactions and the blue edges represent both protein-metabolite and transcript-metabolite interactions.*

3.3.3.2 Evaluating multi-layered graphs generated using hypothesis-driven seeds

In this section, we evaluated the feature interactions (**Supplementary File 3.2**) in the multi-layered graphs (accessible at http://cytoscape.h3africa.org) generated from hypothesis-driven seeds (**Table 3.1**).

To begin with, we evaluated the topology of the generated multi-layered graphs. Analysis of the generated multi-layered graphs revealed subnetworks established beyond those formed by the seed nodes, *IL-6* and *IL-6R*. Specifically, analysis of the multi-layered graph generated for the mild disease state revealed additional subnetworks, beyond those formed by the seed nodes (*IL-6* and *IL-6R*), involving hubs *CCL2, CCL4*, and *IL-7R* (**Figure 3.5A**). These hubs establish direct interactions with the seed nodes and interactions with other molecular features, including interleukins (e.g., *IL-13, IL-18, IL-1A*), and chemokines (e.g., *CCL3, CCL7, CXCL10)*. Notable cross-layer interactions for the mild disease state were between *IL-6* and *CCL2* with metabolites involved in various metabolic and inflammatory processes. These include but not limited to, taurine, 1-(1-enyl-palmitoyl)-2-oleoyl-GPC, 1-(1-enyl-palmitoyl)-GPC, 1-(1-enyl-palmitoyl)-GPE, 1-(1-enyl-stearoyl)-2-dihomo-linolenoyl-GPE, proline, 1-(1-enyl-palmitoyl)-GPC, 6-bromotryptophan, 1-margaroyl-GPC, and 1-myristoyl-GPC. This observation may further suggest a more efficient immune response to the virus among patients with mild disease. Specifically, 6-bromotryptophan is a biomarker indicative of COVID-19 severity, which is a finding concurrently echoed in the study by Krishan et al [233]. The 6-bromotryptophan possesses antiviral properties that can inhibit viral replication. Moreover, such brominated compounds are understood to exert influence on the immune system, particularly by modulating the kynurenine pathway of tryptophan metabolism, which is involved in regulating immune responses [233-235]

Our analysis of the multi-layered graph generated for the moderate disease state revealed new subnetworks formed by hubs *IL-10*, *IL-7R*, and Nuclear Factor Kappa B Subunit 1 (*NFKB1*). These hubs directly interact with the seed nodes (*IL-6* and *IL-6R*), but also with chemokines (e.g., *CCL4, CXCL5, CXCL8*) and other biosignatures such as *HLA-C, TNF, IFNG*, and *IFNAR2*, indicating a more complex interplay of immune cell recruitment and activation (**Figure 3.5B)**. We also observed cross-layer interactions between *IL-10* and *IL-6*, and their connections to specific metabolites.

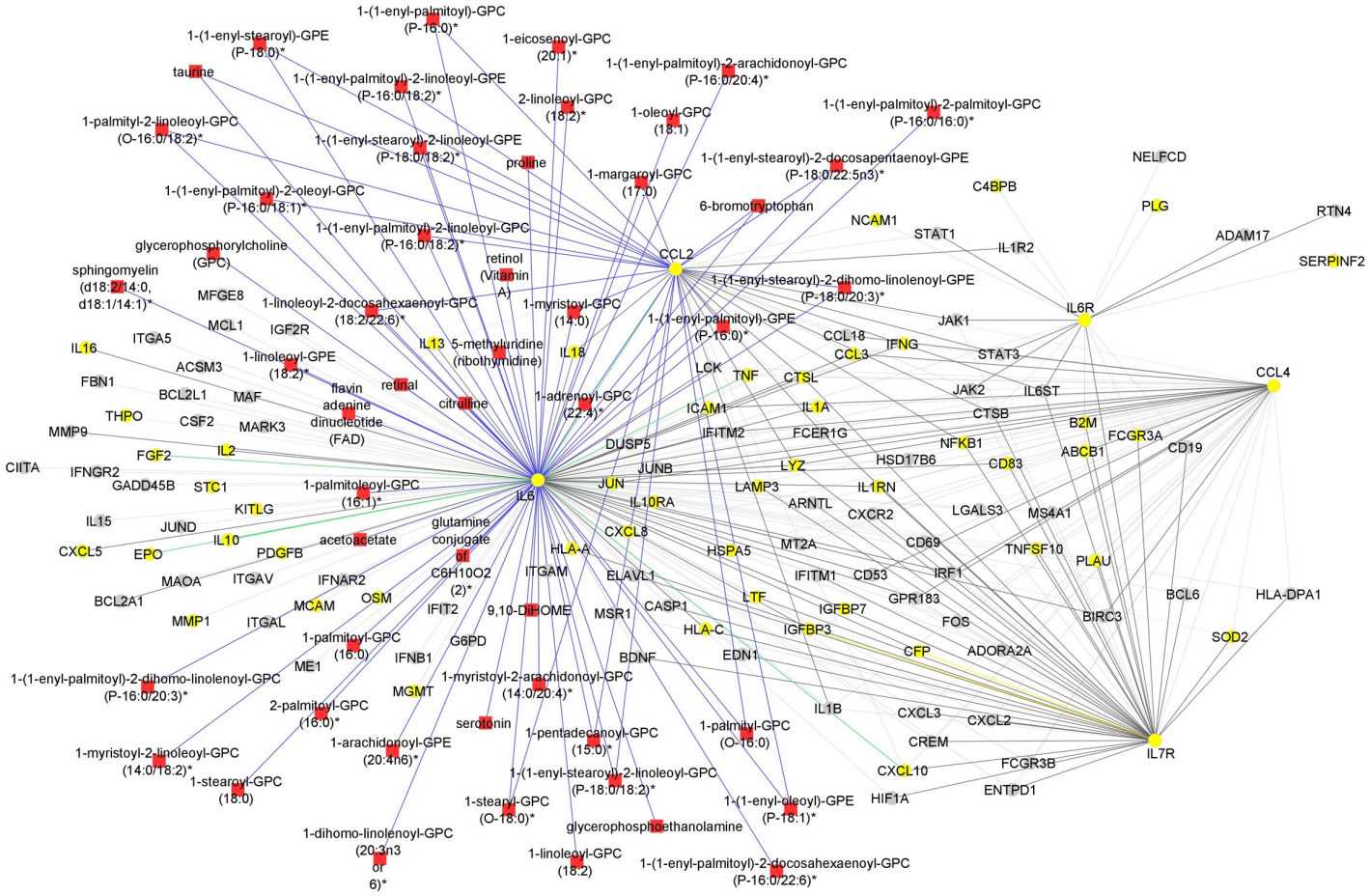
Particularly, the cross-layer interactions involving *IL-10* indicate a potential link between the immune and metabolic responses in moderate disease, suggesting a potential shift in the immune response dynamics in moderate cases as compared to mild cases. These metabolites including phosphatidylethanolamine, taurine, serotonin, and various GPC/GPE molecules such as 1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*, 1-(1-enyl-palmitoyl)-GPC (P-16:0)*, 1-(1-enyl-palmitoyl)-GPE (P-16:0)*, 1-(1-enyl-stearoyl)-2-dihomo-linolenoyl-GPE (P-18:0/20:3)*, and 1-(1-enyl-stearoyl)-GPE (P-18:0)*, play crucial roles in both metabolic and inflammatory processes. Also, the *NFKB1* and *IL-7R* subnetwork primarily interact with other proteins and transcripts. This suggests a focus on regulating gene expression and protein function, possibly contributing to the more severe inflammatory response associated with moderate disease. Compared to the multi-layered graph generated for the mild disease state, analysis of the multi-layered graph generated for the moderate disease state revealed a subnetwork formed by suberoylcarnitine (C8-DC) metabolite that established indirect cross-layer associations with the seed nodes and other hubs. Specifically, the suberoylcarnitine (C8-DC), plays a central role, influencing hubs like *IL-6R, NFKB1, IL-7R,* and *IL-10* indirectly through its connection with *IL-6* via a glutamine conjugate acting as a mediator (**Figure 3.5C)**. This cross-layer pattern suggests a more nuanced control of immune and inflammatory processes compared to the mild disease state.

Analysis of the severe disease state multi-layered graph revealed subnetworks centered on *IL-7R*, *CCL4*, and *CXCL1* (**Figure 3.5D)**. These subnetworks exhibit significant cross-layer interactions with both protein-coding transcripts and metabolites, suggesting a complex interplay between gene expression and metabolic processes in this critical disease state. Similar to the cross-layer pattern observed in the moderate disease state, the severe disease network showcases another example of indirect cross-layer interaction via metabolites. Here, sphingomyelin (d18:2/21:0, d16:2/23:0) takes centre stage, influencing *IL-6* and subsequently impacting *IL-7R, CCL4, IL-6R*, and *CXCL1* (**Figure 3.5E)**. This further emphasizes the shift towards more nuanced, metabolite-driven cross-layer interactions in the severe disease state, potentially indicating immune-metabolic dysregulation in moderate and severe disease states.

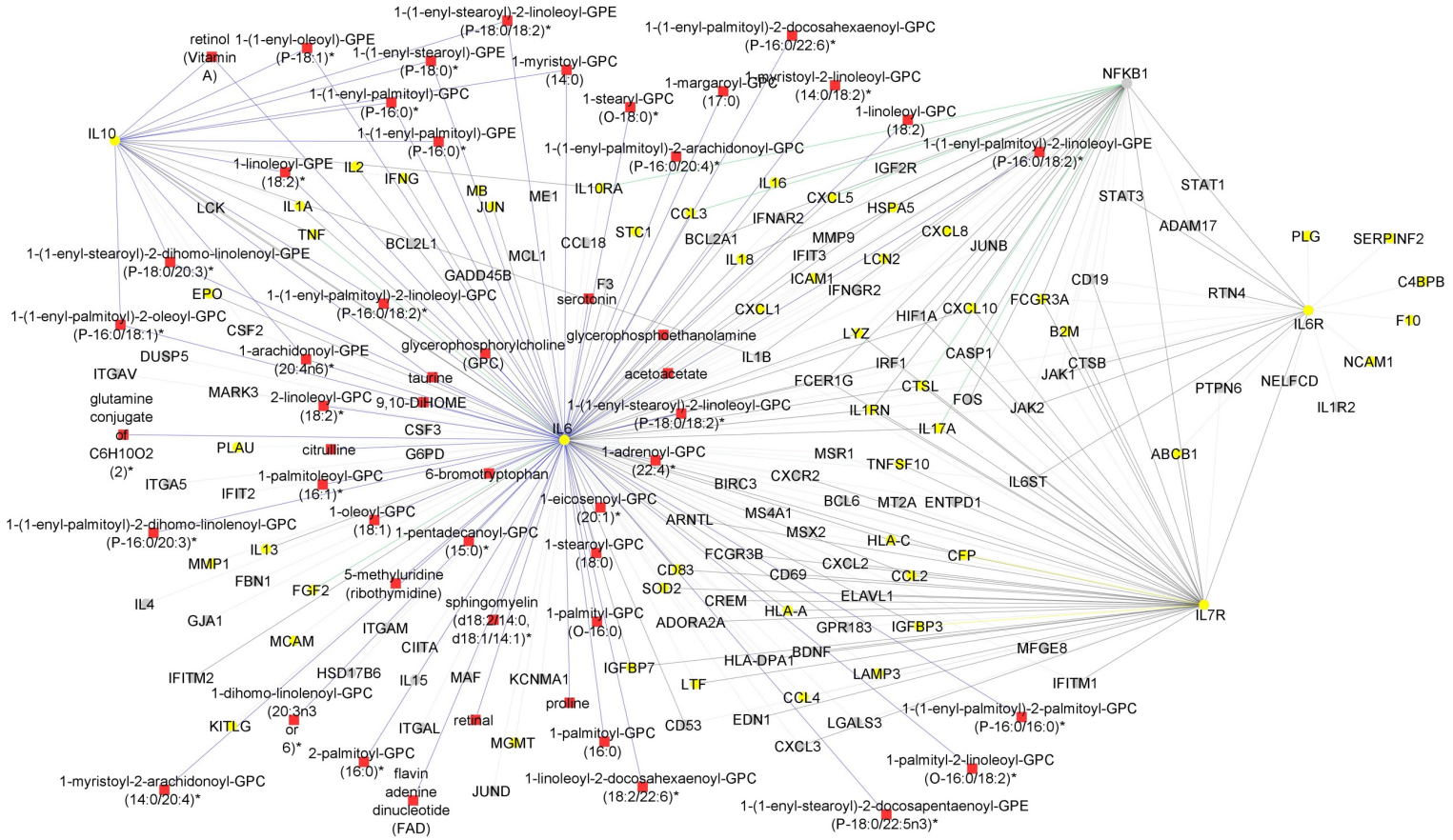
Research suggests a correlation between *IL-6* levels and altered metabolite levels including amino acids, fatty acids, and lipids among severe COVID-19 patients [236, 237]. Similarly, *IL-10* levels show various correlations with altered metabolite levels in infected COVID-19 patients [237]. For instance, the study found that in COVID-19 patients, *IL-10* levels were positively correlated with metabolites involved in glycolysis and the pentose phosphate pathway, such as glucose, fructose, and ribose-5-phosphate. The study also found a negative correlation between *IL-10* levels and metabolites involved in the tricarboxylic acid cycle (TCA cycle) and oxidative phosphorylation, such as citrate, succinate, and ATP. These findings suggest that *IL-10* may be associated with a shift in cellular metabolism towards glycolysis and away from oxidative phosphorylation in COVID-19 patients [237]. The overall results highlight not only the influence of *IL-6* in COVID-19 but also suggest that cross-layer interactions involving *IL-6* influence clinical heterogeneity, thus influencing the dynamics of disease severity.

Furthermore, we evaluated the pairwise interactions (**Supplementary File 2**) and identified 807 interactions associated with one disease state of which approximately 37%, 37%, 2%, 15%, and 9% are transcript-transcript, lipid-lipid, protein-protein, metabolite-metabolite, and protein-metabolite interactions respectively (**Supplementary Figure 3.2A**). Also, of these interactions, approximately 20%, 40%, and 40% are associated with mild, moderate, and severe disease states, respectively. Additionally, the analysis of interactions involved with only two disease states (**Supplementary File 2**) revealed 335 interactions, out of which approximately 6%, 85%, 7%, and 2% are protein-protein, transcript-transcript, lipid-lipid, and metabolite-metabolite interactions, respectively (**Supplementary Figure 3.2B**). Of these interactions, approximately 16%, 10*%,* and 74% are involved with moderate-severe, mild-moderate*,* and mild-severe disease states, respectively.
Finally, we identified 894 interactions (**Supplementary File 2**) common across the three disease states, of which approximately 66%, 3%, 14%, and 17% are transcript-transcript, protein-protein, protein-transcript, and protein-metabolite interactions respectively (**Supplementary Figure 3.2C**).

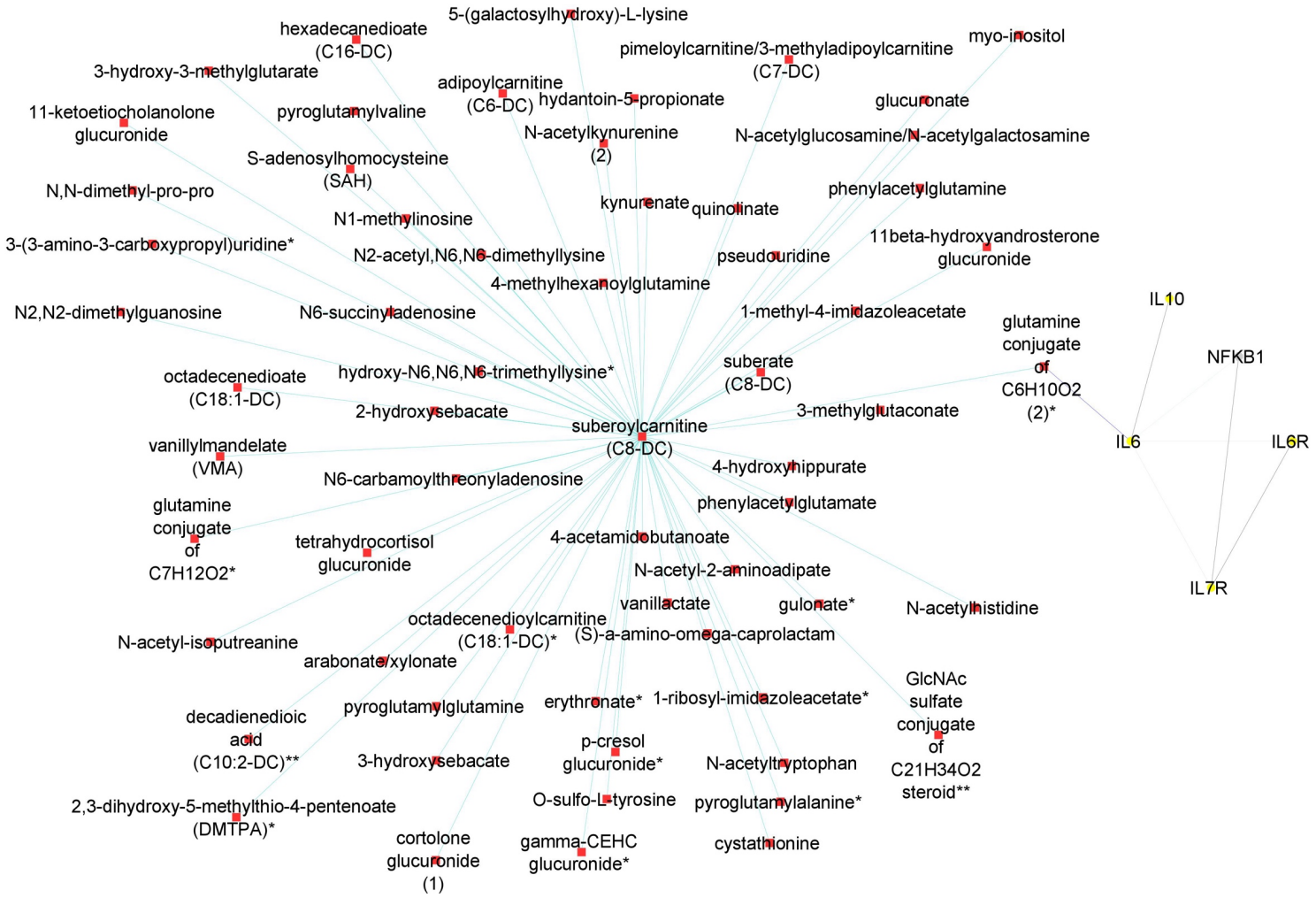
A


1-(1-enyl-palmitoyl)-GPC (P-16:0)*
1-(1-enyl-stearoyl)-GPE (P-18:0)*
1-eicosenoyl-GPC (20:1)*
1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC (P-16:0/20:4)*
taurine
1-(1-enyl-palmitoyl)-2-linoleoyl-GPE (P-16:0/18:2)*
2-linoleoyl-GPC (18:2)*
1-oleoyl-GPC (18:1)
1-(1-enyl-palmitoyl)-2-palmitoyl-GPC (P-16:0/16:0)*
1-palmityl-2-linoleoyl-GPC (O-16:0/18:2)*
1-(1-enyl-stearoyl)-2-linoleoyl-GPE (P-18:0/18:2)*
proline
1-(1-enyl-stearoyl)-2-docosapentaenoyl-GPE (P-18:0/22:5n3)*
1-margaroyl-GPC (17:0)
NELFCD
C4BPB
1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*
6-bromotryptophan
PLG
RTN4
NCAM1
STAT1
ADAM17
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*
IL1R2
glycerophosphorylcholine (GPC)
CCL2
SERPINF2
retinol (Vitamin A)
sphingomyelin (d18:2/14:0, d18:1/14:1)*
1-(1-enyl-stearoyl)-2-dihomo-linolenoyl-GPE (P-18:0/20:3)*
MFGE8
1-linoleoyl-2-docosahexaenoyl-GPC (18:2/22:6)*
1-myristoyl-GPC (14:0)
IL6R
1-(1-enyl-palmitoyl)-GPE (P-16:0)*
JAK1
MCL1
IGF2R
CCL18
IFNG
ITGA5
IL13
5-methyluridine (ribothymidine)
CCL3
STAT3
IL16
IL18
1-linoleoyl-GPE (18:2)*
LCK
TNF
CTSL
CCL4
ACSM3
FBN1
flavin adenine dinucleotide (FAD)
retinal
JAK2
IL6ST
BCL2L1
MAF
citrulline
1-adrenoyl-GPC (22:4)*
ICAM1
IL1A
CTSB
THPO
CSF2
MARK3
IFITM2
FCER1G
B2M
FCGR3A
MMP9
DUSP5
NFKB1
ABCB1
CD19
FGF2
IL2
HSD17B6
CD83
CIITA
IFNGR2
JUNB
LYZ
GADD45B
STC1
1-palmitoleoyl-GPC (16:1)*
JUN
LAMP3
IL1RN
IL6
IL10RA
ARNTL
CXCR2
LGALS3
MS4A1
KITLG
IL15
JUND
glutamine conjugate
CXCL8
IL10
acetoacetate
HLA-A
MT2A
CD69
TNFSF10
CXCL5
EPO
PDGFB
or
HSPA5
PLAU
C6H10O2 (2)*
ITGAM
ELAVL1
IFITM1
CD53
IRF1
MAOA
ITGAV
IFNAR2
GPR183
BCL6
HLA-DPA1
BCL2A1
OSM
MCAM
IFIT2
9,10-DIHOME
MSR1
CASP1
LTF
IGFBP7
BIRC3
MMP1
ITGAL
FOS
1-palmitoyl-GPC (16:0)
G6PD
HLA-C
IGFBP3
SOD2
ME1
IFNB1
BDNF
EDN1
CFP
ADORA2A
1-(1-enyl-palmitoyl)-2-dihomo-linolenoyl-GPC (P-16:0/20:3)*
1-myristoyl-2-arachidonoyl-GPC (14:0/20:4)*
MGMT
2-palmitoyl-GPC (16:0)*
IL1B
CXCL3
CXCL2
serotonin
1-palmityl-GPC (O-16:0)
1-pentadecanoyl-GPC (15:0)*
CREM
IL7R
1-myristoyl-2-linoleoyl-GPC (14:0/18:2)*
1-arachidonoyl-GPE (20:4n6)*
CXCL10
FCGR3B
1-stearoyl-GPC (18:0)
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)*
HIF1A
ENTPD1
1-stearyl-GPC (O-18:0)*
1-(1-enyl-oleoyl)-GPE (P-18:1)*
glycerophosphoethanolamine
1-dihomo-linolenoyl-GPC (20:3n3 or 6)*
1-linoleoyl-GPC (18:2)
1-(1-enyl-palmitoyl)-2-docosahexaenoyl-GPC (P-16:0/22:6)*

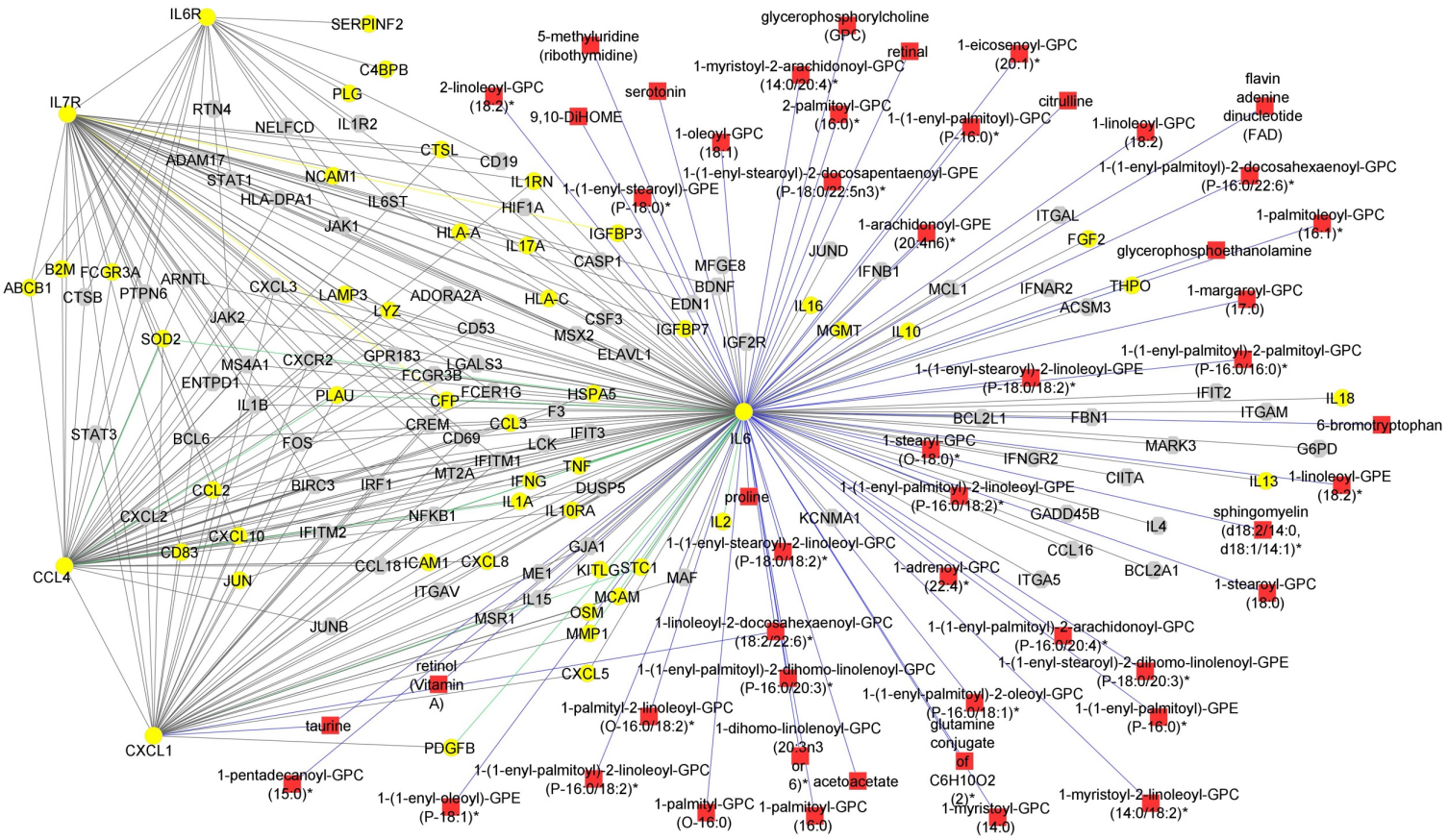
B


1-(1-enyl-stearoyl)-2-linoleoyl-GPE (P-18:0/18:2)*
1-(1-enyl-palmitoyl)-2-docosahexaenoyl-GPC (P-16:0/22:6)*
retinol (Vitamin A)
1-(1-enyl-oleoyl)-GPE (P-18:1)*
1-(1-enyl-stearoyl)-GPE (P-18:0)*
1-myristoyl-GPC (14:0)
1-margaroyl-GPC (17:0)
1-myristoyl-2-linoleoyl-GPC (14:0/18:2)*
NFKB1
1-(1-enyl-palmitoyl)-GPC (P-16:0)*
1-stearyl-GPC (O-18:0)*
1-linoleoyl-GPC (18:2)
IL10
1-(1-enyl-palmitoyl)-GPE (P-16:0)*
1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC (P-16:0/20:4)*
1-(1-enyl-palmitoyl)-2-linoleoyl-GPE (P-16:0/18:2)*
1-linoleoyl-GPE (18:2)*
IL2
IFNG
IGF2R
STAT1
IL16
CXCL5
STAT3
LCK
IL1A
MB
JUN
ME1
IL10RA
CCL3
HSPA5
ADAM17
TNF
BCL2L1
CCL18
STC1
IFNAR2
CXCL8
PLG
SERPINF2
1-(1-enyl-stearoyl)-2-dihomo-linolenoyl-GPE (P-18:0/20:3)*
MCL1
BCL2A1
MMP9
JUNB
IL18
IFIT3
LCN2
GADD45B
ICAM1
CD19
C4BPB
EPO
F3
serotonin
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*
IFNGR2
RTN4
1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*
CXCL1
HIF1A
CXCL10
FCGR3A
IL6R
F10
CSF2
glycerophosphoethanolamine
LYZ
B2M
DUSP5
1-arachidonoyl-GPE (20:4n6)*
glycerophosphorylcholine (GPC)
IL1B
CASP1
NCAM1
ITGAV
IRF1
CTSB
JAK1
taurine
acetoacetate
FCER1G
PTPN6
NELFCD
glutamine conjugate
MARK3
2-linoleoyl-GPC (18:2)*
9,10-DiHOME
CTSL
FOS
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)*
IL1RN
JAK2
IL1R2
IL6
CSF3
IL17A
C6H10O2 (2)*
PLAU
citrulline
G6PD
1-adrenoyl-GPC (22:4)*
MSR1
ABCB1
TNFSF10
ITGA5
1-palmitoleoyl-GPC (16:1)*
6-bromotryptophan
BIRC3
CXCR2
IL6ST
IFIT2
1-eicosenoyl-GPC (20:1)*
BCL6
MT2A
ENTPD1
1-oleoyl-GPC (18:1)
1-pentadecanoyl-GPC (15:0)*
ARNTL
MS4A1
1-(1-enyl-palmitoyl)-2-dihomo-linolenoyl-GPC (P-16:0/20:3)*
MSX2
HLA-C
IL13
1-stearoyl-GPC (18:0)
FCGR3B
CFP
MMP1
CD83
CXCL2
5-methyluridine (ribothymidine)
CD69
CCL2
FBN1
SOD2
ELAVL1
FGF2
IL4
sphingomyelin (d18:2/14:0, d18:1/14:1)*
CREM
HLA-A
IL7R
GJA1
1-palmityl-GPC (O-16:0)
ADORA2A
GPR183
IGFBP3
ITGAM
MCAM
CIITA
BDNF
MFGE8
HLA-DPA1
IGFBP7
IFITM2
HSD17B6
IL15
MAF
KCNMA1
LAMP3
LTF
IFITM1
1-dihomo-linolenoyl-GPC (20:3n3 or 6)*
proline
CCL4
1-(1-enyl-palmitoyl)-2-palmitoyl-GPC (P-16:0/16:0)*
KITLG
ITGAL
retinal
CD53
EDN1
LGALS3
1-palmitoyl-GPC (16:0)
MGMT
CXCL3
2-palmitoyl-GPC (16:0)*
1-palmityl-2-linoleoyl-GPC (O-16:0/18:2)*
flavin adenine dinucleotide (FAD)
1-myristoyl-2-arachidonoyl-GPC (14:0/20:4)*
JUND
1-linoleoyl-2-docosahexaenoyl-GPC (18:2/22:6)*
1-(1-enyl-stearoyl)-2-docosapentaenoyl-GPE (P-18:0/22:5n3)*

C


5-(galactosylhydroxy)-L-lysine
hexadecanedioate (C16-DC)
pimeloylcarnitine/3-methyladipoylcarnitine (C7-DC)
myo-inositol
3-hydroxy-3-methylglutarate
adipoylcarnitine (C6-DC)
hydantoin-5-propionate
glucuronate
11-ketoetiocholanolone glucuronide
pyroglutamylvaline
N-acetylkynurenine (2)
N-acetylglucosamine/N-acetylgalactosamine
S-adenosylhomocysteine (SAH)
N,N-dimethyl-pro-pro
kynurenate
quinolinate
phenylacetylglutamine
N1-methylinosine
3-(3-amino-3-carboxypropyl)uridine*
N2-acetyl,N6,N6-dimethyllysine
pseudouridine
11beta-hydroxyandrosterone glucuronide
4-methylhexanoylglutamine
N2,N2-dimethylguanosine
N6-succinyladenosine
1-methyl-4-imidazoleacetate
glutamine conjugate of C6H10O2 (2)*
IL10
NFKB1
octadecenedioate (C18:1-DC)
hydroxy-N6,N6,N6-trimethyllysine*
suberate (C8-DC)
2-hydroxysebacate
3-methylglutaconate
suberoylcarnitine (C8-DC)
IL6
IL6R
vanillylmandelate (VMA)
4-hydroxyhippurate
N6-carbamoylthreonyladenosine
glutamine conjugate of C7H12O2*
phenylacetylglutamate
tetrahydrocortisol glucuronide
4-acetamidobutanoate
N-acetyl-2-aminoadipate
IL7R
vanillactate
gulonate*
N-acetylhistidine
octadecenedioylcarnitine (C18:1-DC)*
(S)-a-amino-omega-caprolactam
N-acetyl-isoputreanine
arabonate/xylonate
GlcNAc sulfate conjugate of C21H34O2 steroid**
pyroglutamylglutamine
erythronate*
1-ribosyl-imidazoleacetate*
decadienedioic acid (C10:2-DC)**
p-cresol glucuronide*
3-hydroxysebacate
N-acetyltryptophan
2,3-dihydroxy-5-methylthio-4-pentenoate (DMTPA)*
O-sulfo-L-tyrosine
pyroglutamylalanine*
cortolone glucuronide (1)
gamma-CEHC glucuronide*
cystathionine

D


IL6R
SERPINF2
C4BPB
PLG
IL7R
RTN4
NELFCD
IL1R2
ADAM17
CTSL
CD19
STAT1
NCAM1
IL1RN
HLA-DPA1
IL6ST
HIF1A
JAK1
HLA-A
IL17A
IGFBP3
CASP1
B2M
FCGR3A
ARNTL
ABCB1
CTSB
PTPN6
CXCL3
LAMP3
ADORA2A
HLA-C
LYZ
JAK2
CD53
CSF3
SOD2
MSX2
IGFBP7
ELAVL1
MS4A1
CXCR2
GPR183
LGALS3
FCGR3B
ENTPD1
PLAU
CFP
FCER1G
HSPA5
IL1B
F3
STAT3
BCL6
FOS
CREM
CCL3
CD69
LCK
IFIT3
IFITM1
TNF
MT2A
IFNG
DUSP5
CCL2
BIRC3
IRF1
IL1A
IL10RA
CXCL2
NFKB1
CXCL10
IFITM2
CD83
GJA1
CCL4
CCL18
ICAM1
CXCL8
ME1
KITLG
STC1
JUN
ITGAV
MAF
IL15
MCAM
JUNB
MSR1
OSM
MMP1
CXCL5
CXCL1
PDGFB
IL6
IL2
KCNMA1
proline
MFGE8
BDNF
EDN1
IL16
MGMT
IL10
IGF2R
JUND
IFNB1
MCL1
IFNAR2
THPO
ACSM3
FGF2
ITGAL
glycerophosphoethanolamine
1-margaroyl-GPC (17.0)
1-palmitoleoyl-GPC (16.1)*
1-(1-enyl-palmitoyl)-2-docosahexaenoyl-GPC (P-16:0/22:6)*
1-arachidonoyl-GPE (20:4n6)*
1-(1-enyl-stearoyl)-2-docosapentaenoyl-GPE (P-18:0/22:5n3)*
1-(1-enyl-stearoyl)-GPE (P-18:0)*
1-oleoyl-GPC (18:1)
2-palmitoyl-GPC (16:0)*
1-(1-enyl-palmitoyl)-GPC (P-16:0)*
1-linoleoyl-GPC (18:2)
citrulline
flavin adenine dinucleotide (FAD)
glycerophosphorylcholine (GPC)
retinal
1-eicosenoyl-GPC (20:1)*
1-myristoyl-2-arachidonoyl-GPC (14:0/20:4)*
5-methyluridine (ribothymidine)
serotonin
2-linoleoyl-GPC (18:2)*
9,10-DiHOME
1-(1-enyl-palmitoyl)-2-palmitoyl-GPC (P-16:0/16:0)*
1-(1-enyl-stearoyl)-2-linoleoyl-GPE (P-18:0/18:2)*
IFIT2
IL18
ITGAM
6-bromotryptophan
BCL2L1
FBN1
MARK3
G6PD
1-stearoyl-GPC (O-18:0)*
IFNGR2
CIITA
IL13
1-linoleoyl-GPE (18:2)*
1-(1-enyl-palmitoyl)-2-linoleoyl-GPE (P-16:0/18:2)*
GADD45B
IL4
sphingomyelin (d18:2/14:0, d18:1/14:1)*
CCL16
BCL2A1
1-adrenoyl-GPC (22:4)*
ITGA5
1-stearoyl-GPC (18:0)
1-linoleoyl-2-docosahexaenoyl-GPC (18:2/22:6)*
1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC (P-16:0/20:4)*
1-(1-enyl-stearoyl)-2-dihomo-linolenoyl-GPE (P-18:0/20:3)*
1-(1-enyl-palmitoyl)-2-dihomo-linolenoyl-GPC (P-16:0/20:3)*
1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*
1-(1-enyl-palmitoyl)-GPE (P-16:0)*
retinol (Vitamin A)
1-palmityl-2-linoleoyl-GPC (O-16:0/18:2)*
taurine
1-dihomo-linolenoyl-GPC (20:3n3 or 6)*
glutamine conjugate of C6H10O2 (2)*
acetoacetate
1-pentadecanoyl-GPC (15:0)*
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*
1-(1-enyl-oleoyl)-GPE (P-18:1)*
1-palmityl-GPC (O-16:0)
1-palmitoyl-GPC (16:0)
1-myristoyl-GPC (14:0)
1-myristoyl-2-linoleoyl-GPC (14:0/18:2)*

E

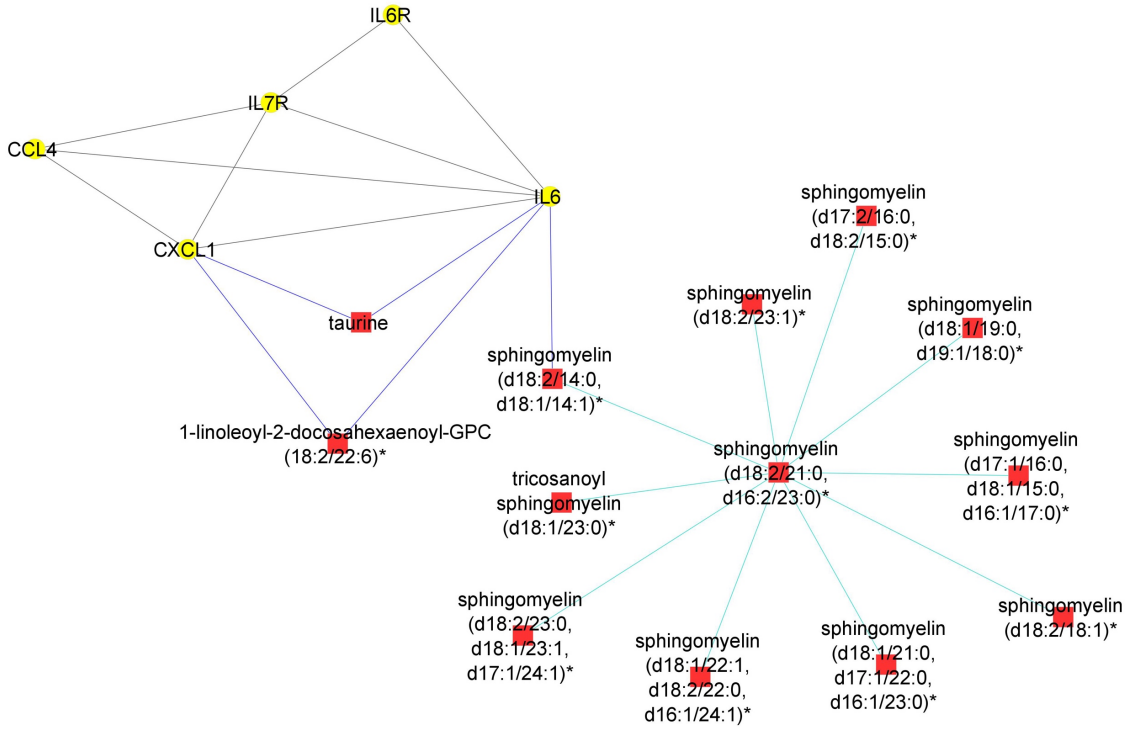


Figure 3.5*. (A) Graph representation of subnetworks formed by hubs CCL2, CCL4, and IL-7R, that establish direct interaction with seed nodes (IL-6 and IL-6R) as observed in the multi-layered graph generated for the mild disease state. The graph highlights the interaction of these hubs with other molecular features including proteins (yellow nodes), transcripts (grey nodes), and metabolites (red nodes). (B) Graph representation of subnetworks formed by hubs IL10, IL-7R, and NFKB1, that establish direct interaction with seed nodes (IL-6 and IL-6R) as observed in the multi-layered graph generated for the moderate disease state. The graph highlights the interaction of these hubs with other molecular features including proteins (yellow nodes), transcripts (grey nodes), and metabolites (red nodes). (C) Graph representation of subnetwork formed by suberoylcarnitine metabolite and the cross-layer interaction with seed nodes (IL-6 and IL-6R), NFKB1, IL-7R, and IL-10 hubs. (D) Graph representation of subnetworks formed by hubs IL-7R, CCL4, and CXCL1, that establish direct interaction with seed nodes (IL-6 and IL-6R) as observed in the multi-layered graph generated for the severe disease state. The blue edges represent protein-metabolite and transcript-metabolite interactions, the green edges represent protein-transcript interactions, the grey edges represent transcript-transcript interactions and, the yellow edges represent protein-protein interaction. (E) Graph representation of subnetwork formed by sphingomyelin (d18:2/21:0, d16:2/23:0) and the cross-layer interaction with seed nodes, IL-6 and IL-6R, and hubs IL-7R, CCL4, IL-6R, and CXCL1 as observed in the multi-layered graph generated for the severe disease state. The grey edges represent transcript-transcript interactions, the yellow edges represent protein-protein interactions, the cyan edges represent metabolite-metabolite interactions, the green edges represent protein-transcript interactions and the blue edges represent both protein-metabolite and transcript-metabolite interactions.*

3.3.4 Characterizing multi-layered graphs.

The observation that cross-layer interactions appear to be a distinctive factor for moderate and severe disease states using both data- and hypothesis-driven methods necessitated the determination of network statistics to further characterize the multi-layered graphs.

According to network density, network heterogeneity, and characteristic path length statistical metrics, graphs with high characteristic route length values have high network heterogeneity and comparatively low network density (**Supplementary data**). The network statistical analysis (**Table 3.4**) further supported the idea that cross-layer interactions could be a factor underlying heterogeneity in disease severity among patients.

Table 3.4. *Results from the network statistical analysis of generated multi-layered graphs*

Network statistical measure	**Data-driven**			**Hypothesis-driven**			**Data-driven validation**			**Hypothesis-driven validation**		
	Mild	Moderate	Severe	Mild	Moderate	Severe	Mild	Moderate	Severe	Mild	Moderate	Severe
Network density	0.008	0.006	0.006	0.010	0.007	0.008	0.102	0.050	0.027	0.049	0.045	0.020
Network heterogeneity	5.133	5.479	5.358	4.475	4.726	4.665	1.844	2.119	3.001	3.083	2.273	3.268
Characteristic path length	2.620	2.845	2.857	2.490	3.027	2.616	1.919	2.353	2.401	1.951	2.452	2.566

3.3.5 Identifying disease states biosignature

3.3.5.1 Biosignatures discriminating between disease states based on data-driven seeds

The results of the multi-layer analysis (**Supplementary File 3.2**) formed the basis for identifying features that discriminate between disease states. Specifically, we explored the pairwise relations associated with one disease state (**Supplementary File 3.2**), as determined using the data-driven approach, and identified 173 discriminatory features (**Table 3. 5**). Of note, the features identified are likely involved in all disease and non-disease phases because they form part of the biological system. However, identifying these features to discriminate between disease states in our analysis may suggest that they are differentially associated (either up or down-regulated) in specific disease states.

Of the 158 discriminatory features that were differentially associated with the mild disease state, approximately 78%, 6%, and 16% were transcripts, proteins, and (uncharacterized) lipids respectively. We identified chemokines (e.g., *CXCL10, CXCL12, CXCL5, CXCL8*, *CCL3, CCL8*), T-cell receptors (e.g., *CD38, CD40, CD48, CD68*), *HLAs* (e.g., *HLA-DPA1, HLA-DPB1, HLA-DRA, HLA-E, HLA-F*), interferons

(e.g., *IFIT1, IFIT3, IFITM2*), and Toll-like receptors (e.g., *TLR2, TLR4, TLR9*) to discriminate the mild disease states. These features are involved in immune responses and play a part in viral entry into host T-cells [221, 238-240]. For instance, *TLR2* activation increased the expression of ACE2, the receptor that SARS-CoV-2 uses to enter cells, suggesting that *TLR2* may play a role in viral entry into host T-cells. *IFIT1* has been shown to have antiviral activity against SARS-CoV-2, and may thus be an important component of the body's immune response against the virus [221]. An elevated level of *CCL18* is associated with inflammation in the lungs of COVID-19 patients through the recruitment and activation of immune cells, including T-cells and dendritic cells in the lungs [241]. Importantly, biosignatures including but not limited to HLA class I alleles, and *CXCL12*, have been validated through sequencing and cohort screening techniques to play a relevant role in immune defense against SARS-CoV-2 [242, 243].

Only 9 and 6 transcripts were associated with moderate and severe disease states respectively: no other omics features were identified to be associated with either of these disease states. There were also no metabolites that were differentially associated with any one of the disease states. This indicated that transcripts strongly differentiate the mild disease state from the moderate and severe disease states.

Table 3. 5. Identified biosignatures that discriminate disease states based on random walk with restart analysis using data-driven seeds.

Feature	Feature type	Disease state	Feature	Feature type	Disease state	Feature	Feature type	Disease state	Feature	Feature	Disease
IRF8	transcript	Severe	CD83	transcript	Mild	IL1B	transcript	Mild	CCL18	transcript	Mild
KDR	transcript	Severe	CD8A	transcript	Mild	IL1RN	transcript	Mild	CCL2	protein	Mild
MMP9	transcript	Severe	CDH5	transcript	Mild	IL5	transcript	Mild	CCL20	transcript	Mild
PRTN3	transcript	Severe	CPB2	protein	Mild	IL6	transcript	Mild	CCL3	transcript	Mild
PTGES	transcript	Severe	CREM	transcript	Mild	IL7R	transcript	Mild	CCL5	transcript	Mild
YWHAE	transcript	Severe	CTSB	transcript	Mild	IRF9	transcript	Mild	CCL7	protein	Mild
CD34	transcript	Moderate	CTSC	transcript	Mild	JAK2	transcript	Mild	CCR1	transcript	Mild
ENPEP	transcript	Moderate	CTSL	transcript	Mild	JUN	transcript	Mild	CD14	transcript	Mild
MAPK8	transcript	Moderate	CXCL1	transcript	Mild	JUNB	transcript	Mild	CD163	transcript	Mild
NKRF	transcript	Moderate	CXCL10	transcript	Mild	KRT10	transcript	Mild	SERPINA4	protein	Mild
NLRP1	transcript	Moderate	CXCL11	protein	Mild	LCK	transcript	Mild	SERPING1	protein	Mild
PDGFB	transcript	Moderate	CXCL2	transcript	Mild	LGALS3	transcript	Mild	SH2D3C	transcript	Mild
SHC1	transcript	Moderate	CXCL8	transcript	Mild	TLR2	transcript	Mild	SIK1	transcript	Mild
SLC14A1	transcript	Moderate	CXCL9	transcript	Mild	TLR4	transcript	Mild	SLC3A2	transcript	Mild
YARS1	transcript	Moderate	CXCR2	transcript	Mild	TLR9	transcript	Mild	SCD	transcript	Mild
ABCB1	transcript	Mild	CXCR4	transcript	Mild	TNF	transcript	Mild	CYTIP	transcript	Mild
ADORA2A	transcript	Mild	CXCR6	transcript	Mild	TNFAIP3	transcript	Mild	DES	transcript	Mild
ADSL	transcript	Mild	CYBB	transcript	Mild	TNFSF10	transcript	Mild	DUSP5	transcript	Mild
AFM	protein	Mild	HNF4A	transcript	Mild	TPI1	protein	Mild	ENTPD1	transcript	Mild
APOE	transcript	Mild	HP	transcript	Mild	UBE2L6	transcript	Mild	EPCAM	transcript	Mild
ARNTL	transcript	Mild	HSD17B6	transcript	Mild	TGFB1	transcript	Mild	F3	transcript	Mild
ARRB2	transcript	Mild	HSPA5	transcript	Mild	STAT3	transcript	Mild	FCER1G	transcript	Mild
ASGR2	transcript	Mild	ICAM1	transcript	Mild	TAP1	transcript	Mild	FCGR3A	transcript	Mild
B2M	transcript	Mild	IFIT1	transcript	Mild	SERPINA1	transcript	Mild	FCGR3B	transcript	Mild
CD2	transcript	Mild	IFIT3	transcript	Mild	BAAT	transcript	Mild	FLNB	transcript	Mild
CD3D	transcript	Mild	IFITM2	transcript	Mild	BCL6	transcript	Mild	GPR183	transcript	Mild
CD4	transcript	Mild	IFNG	transcript	Mild	BDNF	transcript	Mild	GZMA	transcript	Mild
CD53	transcript	Mild	IGF1R	transcript	Mild	BIRC3	transcript	Mild	GZMB	transcript	Mild
CD69	transcript	Mild	IL10RA	transcript	Mild	BTK	transcript	Mild	HLA-DPA1	transcript	Mild
CD74	transcript	Mild	IL1A	transcript	Mild	CA1	protein	Mild	HLA-DPB1	transcript	Mild
HLA-DRA	transcript	Mild	MS4A1	transcript	Mild	Unknown_mz_882.75427_+_RT_27.8	lipid	Mild	Unknown_mz_786.66003_+_RT_26.3	lipid	Mild
HLA-E	transcript	Mild	MT2A	transcript	Mild	Unknown_mz_921.69031_+_RT_29.1	lipid	Mild	Unknown_mz_786.66064_+_RT_27.7	lipid	Mild
HLA-F	transcript	Mild	MYD88	transcript	Mild	Unknown_mz_579.40155_+_RT_20.3	lipid	Mild	Unknown_mz_795.45471_+_RT_25.5	lipid	Mild
S100A12	transcript	Mild	NFKB1	transcript	Mild	Unknown_mz_600.52008_+_RT_23.1	lipid	Mild	Unknown_mz_806.62952_+_RT_25.8	lipid	Mild
S100A8	transcript	Mild	NFKBIA	transcript	Mild	Unknown_mz_650.53461_+_RT_24.0	lipid	Mild	Unknown_mz_810.66071_+_RT_27.0	lipid	Mild
S100A9	transcript	Mild	NLRP3	transcript	Mild	Unknown_mz_652.55096_+_RT_24.6	lipid	Mild	Unknown_mz_812.67621_+_RT_27.8	lipid	Mild
LY6E	transcript	Mild	OASL	transcript	Mild	Unknown_mz_702.56604_+_RT_23.8	lipid	Mild	Unknown_mz_817.63153_+_RT_27.8	lipid	Mild
LYZ	transcript	Mild	PDCD1	transcript	Mild	Unknown_mz_702.56689_+_RT_24.7	lipid	Mild	Unknown_mz_830.72375_+_RT_26.9	lipid	Mild
MAPK14	transcript	Mild	PF4	transcript	Mild	Unknown_mz_728.58246_+_RT_25.1	lipid	Mild	Unknown_mz_832.6452_+_RT_26.26	lipid	Mild
MAPK3	transcript	Mild	PLAU	transcript	Mild	Unknown_mz_736.64563_+_RT_24.6	lipid	Mild	Unknown_mz_836.67639_+_RT_27.4	lipid	Mild
MB	transcript	Mild	PRF1	transcript	Mild	Unknown_mz_750.66119_+_RT_25.5	lipid	Mild	Unknown_mz_838.69214_+_RT_27.0	lipid	Mild
MEFV	transcript	Mild	PTPN6	transcript	Mild	Unknown_mz_756.61407_+_RT_25.8	lipid	Mild	Unknown_mz_838.69214_+_RT_28.1	lipid	Mild
MRC1	transcript	Mild	RAF1	transcript	Mild	Unknown_mz_782.62958_+_RT_26.1	lipid	Mild	Unknown_mz_858.6604_+_RT_26.65	lipid	Mild

3.3.5.2 Biosignatures discriminating between disease states based on hypothesis-driven seeds

We explored the pairwise relations associated with one disease state (**Supplementary File 3.2)** based on analysis using seeds that were selected to test specific hypotheses. The results (**Table 3. 6)** revealed more biosignatures to be differentially associated with moderate and severe disease states than in the mild disease state. Additionally, unlike with the data-driven seed analysis, we observed more proteins and metabolites that discriminated between the moderate disease state than the others. Compounds for which a matching pure standard was not available for confirmation are denoted by adding an asterisk (*) symbol after the name of the compound.

We identified chemokines (*CXCL2, CXCL3, CXCR1, CXCR2, CXCR3, CXCR6),* cytokines *(*e.g., *TNF, TNFRSF1A, TNFSF10)* and other transcripts and proteins (e.g., *ATP6AP2*) that promotes the cytokine storm to discriminate the severe disease state from the other states. Concordantly, several studies have reported elevated levels of these features in COVID-19 patients, particularly those with severe disease [244-246]. Also, some of the biosignatures including but not limited to *TNF, IL-10*, have been verified using multiplex biosensor techniques to be associated with COVID-19, and as such these biosignatures could serve as markers to monitor disease development [239, 247].

The results further revealed lysophosphatidylcholine (LysoPC), diacylglycerol (DG), and triglycerides (TG) to discriminate severe disease states. Several studies have suggested the possible differential association of these lipids with COVID-19 pathogenesis and disease severity [248-251].

We identified metabolites such as kynurenate [252], and gluconate [253], which have been reported to play a role in the cytokine storm and immune response, to discriminate the moderate disease state from the other states.

Comparing the biosignatures discriminating disease states further supports the idea that transcripts, metabolites, lipids, and proteins collectively influence disease progression beyond the mild state. In addition, we identified features (e.g., *SLC14A1, Adipoylcarnitine*) for which no direct roles in influencing disease severity have

previously been reported: further research may provide important insights into their roles in COVID-19 pathogenesis and whether they might be useful targets for therapeutic intervention. Overall, these findings suggest that the discriminatory features play a significant role in the immune response to COVID-19 and that targeting them and/or their associated signalling pathways may be a potential therapeutic approach [244].

***Table 3. 6**. Identified biosignatures that discriminate disease states based on random walk with restart analysis using hypothesis-driven seeds*

Feature	Feature type	Disease state	Feature	Feature type	Disease state
ABCC1	Transcript	Severe	Unknown_mz_742.65503_+_RT_26.11	Lipid	Severe
ACKR3	Transcript	Severe	Unknown_mz_743.61646_+_RT_28.297	Lipid	Severe
ACOD1	Transcript	Severe	Unknown_mz_745.5152_+_RT_14.3	Lipid	Severe
ADRA2B	Transcript	Severe	Unknown_mz_748.39844_-_RT_24.778	Lipid	Severe
ATP6AP2	Transcript	Severe	Unknown_mz_752.67621_+_RT_29.183	Lipid	Severe
BCL6	Transcript	Severe	Unknown_mz_752.67676_+_RT_26.296	Lipid	Severe
BIRC3	Transcript	Severe	Unknown_mz_757.63165_+_RT_28.819	Lipid	Severe
CA1	Transcript	Severe	Unknown_mz_762.66071_+_RT_27.74	Lipid	Severe
CA2	Transcript	Severe	Unknown_mz_764.67676_+_RT_27.522	Lipid	Severe
CCL16	Transcript	Severe	Unknown_mz_765.60468_+_RT_16.383	Lipid	Severe
CD69	Transcript	Severe	Unknown_mz_766.69196_+_RT_26.293	Lipid	Severe
CD83	Transcript	Severe	Unknown_mz_766.69232_+_RT_28.305	Lipid	Severe
CFP	Transcript	Severe	Unknown_mz_769.6311_+_RT_28.523	Lipid	Severe
CREM	Transcript	Severe	Unknown_mz_770.68744_+_RT_27.138	Lipid	Severe
CSF3	Transcript	Severe	Unknown_mz_771.64661_+_RT_29.293	Lipid	Severe
CXCL2	Transcript	Severe	Unknown_mz_776.67676_+_RT_28.157	Lipid	Severe
CXCL3	Transcript	Severe	Unknown_mz_780.69897_+_RT_28.309	Lipid	Severe
CXCR1	Transcript	Severe	Unknown_mz_780.7074_+_RT_29.787	Lipid	Severe
CXCR2	Transcript	Severe	Unknown_mz_780.7077_+_RT_29.279	Lipid	Severe
CXCR3	Transcript	Severe	Unknown_mz_780.70953_+_RT_27.345	Lipid	Severe
CXCR6	Transcript	Severe	Unknown_mz_782.4267_-_RT_24.225	Lipid	Severe
DPP4	Transcript	Severe	Unknown_mz_782.7207_+_RT_29.713	Lipid	Severe
ELF3	Transcript	Severe	Unknown_mz_792.70734_+_RT_28.536	Lipid	Severe
EPX	Transcript	Severe	Unknown_mz_793.63165_+_RT_28.367	Lipid	Severe
GCLC	Transcript	Severe	Unknown_mz_794.72351_+_RT_29.303	Lipid	Severe
GJA1	Transcript	Severe	Unknown_mz_794.7243_+_RT_28.327	Lipid	Severe
IDO1	Transcript	Severe	Unknown_mz_795.64679_+_RT_28.628	Lipid	Severe
IFITM1	Transcript	Severe	Unknown_mz_796.73724_+_RT_30.218	Lipid	Severe
IFITM2	Transcript	Severe	Unknown_mz_798.7207_+_RT_28.176	Lipid	Severe
IL15	Transcript	Severe	Unknown_mz_806.72339_+_RT_29.498	Lipid	Severe
IL33	Transcript	Severe	Unknown_mz_811.62109_+_RT_28.63	Lipid	Severe
IL3RA	Transcript	Severe	Unknown_mz_812.67621_+_RT_28.125	Lipid	Severe
IL7R	Transcript	Severe	Unknown_mz_814.69196_+_RT_28.152	Lipid	Severe
IRF8	Transcript	Severe	Unknown_mz_816.7077_+_RT_29.317	Lipid	Severe
ITGAV	Transcript	Severe	Unknown_mz_820.73932_+_RT_28.552	Lipid	Severe
ITGAX	Transcript	Severe	Unknown_mz_821.66254_+_RT_28.831	Lipid	Severe
ITGB3	Transcript	Severe	Unknown_mz_822.72083_+_RT_27.527	Lipid	Severe
JUN	Transcript	Severe	Unknown_mz_822.75537_+_RT_28.304	Lipid	Severe
KDM5B	Transcript	Severe	Unknown_mz_824.737_+_RT_28.3	Lipid	Severe
KDR	Transcript	Severe	Unknown_mz_826.75671_+_RT_29.2	Lipid	Severe
KIT	Transcript	Severe	Unknown_mz_830.71057_+_RT_27.868	Lipid	Severe
KITLG	Transcript	Severe	Unknown_mz_832.72412_+_RT_28.634	Lipid	Severe
KRT5	Transcript	Severe	Unknown_mz_832.73975_+_RT_27.702	Lipid	Severe
LPAR2	Transcript	Severe	Unknown_mz_836.74402_+_RT_28.521	Lipid	Severe
LRP1	Transcript	Severe	Unknown_mz_836.77008_+_RT_29.296	Lipid	Severe
MAF	Transcript	Severe	Unknown_mz_846.75452_+_RT_27.716	Lipid	Severe
MAPK14	Transcript	Severe	Unknown_mz_846.7561_+_RT_28.63	Lipid	Severe
MCAM	Transcript	Severe	Unknown_mz_850.7865_+_RT_29.293	Lipid	Severe
MDM2	Transcript	Severe	Unknown_mz_851.5166_+_RT_27.535	Lipid	Severe
ME1	Transcript	Severe	Unknown_mz_852.76886_+_RT_29.292	Lipid	Severe
MMP1	Transcript	Severe	Unknown_mz_862.78735_+_RT_29.482	Lipid	Severe
MMP2	Transcript	Severe	Unknown_mz_864.77179_+_RT_28.55	Lipid	Severe
MSN	Transcript	Severe	Unknown_mz_876.77521_+_RT_28.63	Lipid	Severe
MSR1	Transcript	Severe	Unknown_mz_902.78906_+_RT_28.837	Lipid	Severe
MT2A	Transcript	Severe	Unknown_mz_905.5639_+_RT_28.63	Lipid	Severe
OPRD1	Transcript	severe	Unknown_mz_1204.04907_+_RT_24.162	Lipid	severe
OPRM1	Transcript	severe	Unknown_mz_1227.05237_+_RT_24.126	Lipid	severe
OSM	Transcript	severe	Unknown_mz_1237.12878_+_RT_25.831	Lipid	severe
PLAU	Transcript	severe	Unknown_mz_1455.19128_+_RT_24.513	Lipid	severe
PLXNA2	Transcript	severe	Unknown_mz_1468.27539_+_RT_29.197	Lipid	severe
PTGS1	Transcript	severe	Unknown_mz_1494.28638_+_RT_29.248	Lipid	severe
RELA	Transcript	severe	Unknown_mz_1516.27661_+_RT_28.532	Lipid	severe
RNASE1	Transcript	severe	Unknown_mz_1567.38342_+_RT_29.458	Lipid	severe
RTN2	Transcript	severe	Unknown_mz_1572.33655_+_RT_29.463	Lipid	severe
S100A4	Transcript	severe	Unknown_mz_245.22595_+_RT_13.471	Lipid	severe

S100A9	Transcript	severe	Unknown_mz_258.11005_+_RT_1.474	Lipid	severe
S1PR1	Transcript	severe	Unknown_mz_263.23645_+_RT_13.471	Lipid	severe
SDC2	Transcript	severe	Unknown_mz_287.23691_+_RT_13.446	Lipid	severe
SDCBP	Transcript	severe	Unknown_mz_302.10052_-_RT_1.27	Lipid	severe
SELENOP	Transcript	severe	Unknown_mz_311.21869_+_RT_10.981	Lipid	severe
SLC14A1	Transcript	severe	Unknown_mz_312.29016_+_RT_17.443	Lipid	severe
SLC1A5	Transcript	severe	Unknown_mz_313.27353_+_RT_14.323	Lipid	severe
SLC2A1	Transcript	severe	Unknown_mz_317.24741_+_RT_12.26	Lipid	severe
SOD2	Transcript	severe	Unknown_mz_319.22717_-_RT_11.732	Lipid	severe
SP1	Transcript	severe	Unknown_mz_337.27298_+_RT_13.471	Lipid	severe
STT3A	Transcript	severe	Unknown_mz_339.25079_+_RT_13.262	Lipid	severe
TCN1	Transcript	severe	Unknown_mz_726.63995_+_RT_27.125	Lipid	severe
TNFSF10	Transcript	severe	Unknown_mz_733.57343_+_RT_28.179	Lipid	severe
VIPR1	Transcript	severe	Unknown_mz_734.63037_+_RT_25.901	Lipid	severe
ZFP36	Transcript	severe	Unknown_mz_736.64594_+_RT_26.57	Lipid	severe
PGF	Protein	severe	Unknown_mz_738.53009_+_RT_13.467	Lipid	severe
STC1	Protein	severe	Unknown_mz_738.6001_+_RT_14.757	Lipid	severe
1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*	Metabolite	severe	(S)-a-amino-omega-caprolactam	Metabolite	moderate
choline phosphate	Metabolite	Severe	11beta-hydroxyandrosterone glucuronide	Metabolite	moderate
glycerate	Metabolite	Severe	11-ketoetiocholanolone glucuronide	Metabolite	moderate
oxalate (ethanedioate)	Metabolite	Severe	1-arachidonoyl-GPE (20:4n6)*	Metabolite	moderate
sphingomyelin (d17:2/16:0, d18:2/15:0)*	Metabolite	Severe	1-linoleoyl-GPE (18:2)*	Metabolite	moderate
sphingomyelin (d18:2/14:0, d18:1/14:1)*	Metabolite	Severe	1-methyl-4-imidazoleacetate	Metabolite	moderate
sphingomyelin (d18:2/18:1)*	Metabolite	Severe	1-ribosyl-imidazoleacetate*	Metabolite	moderate
sphingomyelin (d18:2/21:0, d16:2/23:0)*	Metabolite	Severe	2,3-dihydroxy-5-methylthio-4-pentenoate (DMTPA)*	Metabolite	moderate
sphingomyelin (d18:2/23:1)*	Metabolite	Severe	3-(3-amino-3-carboxypropyl)uridine*	Metabolite	moderate
sulfate of piperine metabolite C18H21NO3 (1)*	Metabolite	Severe	5-(galactosylhydroxy)-L-lysine	Metabolite	moderate
tartronate (hydroxymalonate)	Metabolite	Severe	adipoylcarnitine (C6-DC)	Metabolite	moderate
threonate	Metabolite	Severe	arabonate/xylonate	Metabolite	moderate
DG 12:0_14:0_RT_21.212	Lipid	Severe	cortolone glucuronide (1)	Metabolite	moderate
LysoPC 18:2_RT_11.045	Lipid	Severe	cystathionine	Metabolite	moderate
LysoPC 18:2_RT_11.623	Lipid	Severe	decadienedioic acid (C10:2-DC)**	Metabolite	moderate
LysoPC 20:2_RT_12.854	Lipid	Severe	erythronate*	Metabolite	moderate
LysoPC 20:3_RT_11.724	Lipid	Severe	gamma-CEHC glucuronide*	Metabolite	moderate
LysoPC 20:3_RT_11.872	Lipid	Severe	GlcNAc sulfate conjugate of C21H34O2 steroid**	Metabolite	moderate
LysoPC 20:3_RT_12.285	Lipid	Severe	glucuronate	Metabolite	moderate
LysoPC 20:4_RT_10.912	Lipid	Severe	glucuronide of piperine metabolite C17H21NO3 (3)*	Metabolite	moderate
LysoPC 20:4_RT_11.056	Lipid	Severe	glucuronide of piperine metabolite C17H21NO3 (4)*	Metabolite	moderate
LysoPC 22:4_RT_12.289	Lipid	Severe	glucuronide of piperine metabolite C17H21NO3 (5)*	Metabolite	moderate
LysoPC 22:4_RT_12.605	Lipid	Severe	glutamine conjugate of C6H10O2 (2)*	Metabolite	moderate
LysoPC 22:5_RT_11.654	Lipid	Severe	gulonate*	Metabolite	moderate
LysoPC 22:5_RT_12.087	Lipid	Severe	hydantoin-5-propionate	Metabolite	moderate
LysoPC 22:6_RT_10.886	Lipid	Severe	hydroxy-N6,N6,N6-trimethyllysine*	Metabolite	moderate
LysoPI 16:0_RT_11.045	Lipid	Severe	hydroxypalmitoyl sphingomyelin (d18:1/16:0(OH))**	Metabolite	moderate
TG 10:0_15:0_16:0_RT_28.707	Lipid	Severe	kynurenate	Metabolite	moderate
TG 12:0_12:0_14:0_RT_27.15	Lipid	Severe	myo-inositol	Metabolite	moderate
TG 12:0_12:0_18:2_RT_27.539	Lipid	Severe	N,N-dimethyl-pro-pro	Metabolite	moderate
TG 12:0_12:0_18:3_RT_26.893	Lipid	Severe	N2-acetyl,N6,N6-dimethyllysine	Metabolite	moderate
TG 12:0_16:1_16:1_RT_28.529	Lipid	Severe	N6-carbamoylthreonyladenosine	Metabolite	moderate
TG 16:0_16:0_24:0_RT_35.146	Lipid	Severe	N6-succinyladenosine	Metabolite	moderate
TG 16:0_18:0_18:0_RT_34.031	Lipid	Severe	N-acetylkynurenine (2)	Metabolite	moderate
TG 42:1_RT_28.302	Lipid	Severe	N-acetyltryptophan	Metabolite	moderate
TG 44:1_RT_29.294	Lipid	Severe	O-sulfo-L-tyrosine	Metabolite	moderate
TG 46:1_RT_30.294	Lipid	Severe	p-cresol glucuronide*	Metabolite	moderate
TG 46:2_RT_29.48	Lipid	Severe	phenylacetylglutamine	Metabolite	moderate
TG 46:3_RT_28.628	Lipid	Severe	phenylalanylphenylalanine	Metabolite	moderate
TG 47:2_RT_29.947	Lipid	Severe	pimeloylcarnitine/3-methyladipoylcarnitine (C7-DC)	Metabolite	moderate
TG 8:0_15:0_16:0_RT_27.751	Lipid	Severe	piperine	Metabolite	moderate
Unknown_mz_1039.67261_+_RT_11.185	Lipid	Severe	pseudouridine	Metabolite	moderate
Unknown_mz_1041.68823_+_RT_11.863	Lipid	Severe	pyroglutamylvaline	Metabolite	moderate
Unknown_mz_1041.68872_+_RT_11.738	Lipid	Severe	quinolinate	Metabolite	moderate
Unknown_mz_1063.672_+_RT_11.885	Lipid	Severe	retinol (Vitamin A)	Metabolite	moderate
Unknown_mz_1063.67212_+_RT_11.077	Lipid	Severe	S-adenosylhomocysteine (SAH)	Metabolite	moderate
Unknown_mz_1063.67285_+_RT_11.193	Lipid	Severe	sphingomyelin (d18:1/17:0, d17:1/18:0, d19:1/16:0)	Metabolite	moderate
Unknown_mz_1087.69226_-_RT_12.293	Lipid	Severe	sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)*	Metabolite	moderate
Unknown_mz_1107.66125_-_RT_11.066	Lipid	Severe	suberoylcarnitine (C8-DC)	Metabolite	moderate
Unknown_mz_1109.65356_+_RT_11.17	Lipid	Severe	sulfate of piperine metabolite C16H19NO3 (2)*	Metabolite	moderate
Unknown_mz_1131.66211_-_RT_11.186	Lipid	Severe	tetrahydrocortisol glucuronide	Metabolite	moderate
Unknown_mz_1155.04663_+_RT_24.774	Lipid	Severe	trans-4-hydroxyproline	Metabolite	moderate
Unknown_mz_1181.06323_+_RT_24.803	Lipid	Severe	*ADAM17*	Transcript	moderate
Unknown_mz_1181.98779_+_RT_23.981	Lipid	Severe	*AR*	Transcript	moderate
Unknown_mz_1193.98499_+_RT_23.882	Lipid	Severe	*ARRB2*	Transcript	moderate
Unknown_mz_339.28937_+_RT_14.753	Lipid	Severe	*B2M*	Transcript	moderate
Unknown_mz_348.07925_+_RT_1.27	Lipid	Severe	*BCL2A1*	Transcript	moderate
Unknown_mz_353.26541_+_RT_14.085	Lipid	Severe	*CASP1*	Transcript	moderate
Unknown_mz_355.28354_+_RT_13.472	Lipid	Severe	*CD19*	Transcript	moderate
Unknown_mz_357.29984_+_RT_14.753	Lipid	Severe	*CD46*	Transcript	moderate
Unknown_mz_360.28915_+_RT_17.001	Lipid	Severe	*CD8B*	Transcript	moderate
Unknown_mz_362.30511_+_RT_17.721	Lipid	Severe	*CSF2*	Transcript	moderate
Unknown_mz_362.3056_+_RT_18.15	Lipid	Severe	*EPCAM*	Transcript	moderate
Unknown_mz_365.24597_-_RT_14.058	Lipid	Severe	*F10*	Transcript	moderate
Unknown_mz_369.23981_+_RT_14.301	Lipid	Severe	*F8*	Transcript	moderate
Unknown_mz_370.29538_+_RT_12.394	Lipid	Severe	*FCGR3A*	Transcript	moderate
Unknown_mz_372.31085_+_RT_13.272	Lipid	Severe	*HIF1A*	Transcript	moderate
Unknown_mz_372.31091_+_RT_13.953	Lipid	Severe	*HMGB1*	Transcript	moderate
Unknown_mz_375.25015_+_RT_12.546	Lipid	Severe	*HSP90AA1*	Transcript	moderate
Unknown_mz_377.26541_+_RT_12.394	Lipid	Severe	*HSPA5*	Transcript	moderate
Unknown_mz_379.28143_+_RT_14.557	Lipid	Severe	*IFIT3*	Transcript	moderate
Unknown_mz_379.28595_+_RT_13.314	Lipid	Severe	*IFNAR2*	Transcript	moderate
Unknown_mz_386.3259_+_RT_14.491	Lipid	Severe	*IFNG*	Transcript	moderate
Unknown_mz_391.24271_-_RT_13.475	Lipid	Severe	*IFNGR2*	Transcript	moderate
Unknown_mz_393.29718_+_RT_15.809	Lipid	Severe	*IGF2R*	Transcript	moderate

Unknown_mz_396.31009_+_RT_13.295	Lipid	Severe	*IKBKG*	Transcript	moderate
Unknown_mz_396.31088_+_RT_13.446	Lipid	Severe	*IL10RB*	Transcript	moderate
Unknown_mz_400.34274_+_RT_15.344	Lipid	Severe	*IL16*	Transcript	moderate
Unknown_mz_405.22116_+_RT_11.052	Lipid	Severe	*IL1A*	Transcript	moderate
Unknown_mz_413.24588_-_RT_13.454	Lipid	Severe	*IL2*	Transcript	moderate
Unknown_mz_420.31012_+_RT_13.354	Lipid	Severe	*IL6R*	Transcript	moderate
Unknown_mz_422.32593_+_RT_13.926	Lipid	Severe	*IRF3*	Transcript	moderate
Unknown_mz_445.2529_+_RT_13.463	Lipid	Severe	*KCNJ2*	Transcript	moderate
Unknown_mz_449.28555_+_RT_16.376	Lipid	Severe	*LCK*	Transcript	moderate
Unknown_mz_453.29675_+_RT_14.74	Lipid	Severe	*MAPK1*	Transcript	moderate
Unknown_mz_496.37521_+_RT_13.592	Lipid	Severe	*MAT2A*	Transcript	moderate
Unknown_mz_508.37686_+_RT_14.624	Lipid	Severe	*MB*	Transcript	moderate
Unknown_mz_532.37421_+_RT_14.589	Lipid	Severe	*MMP9*	Transcript	moderate
Unknown_mz_535.4339_+_RT_22.661	Lipid	Severe	*NBN*	Transcript	moderate
Unknown_mz_552.40222_+_RT_14.023	Lipid	Severe	*NR1H2*	Transcript	moderate
Unknown_mz_561.44934_+_RT_22.704	Lipid	Severe	*NRL*	Transcript	moderate
Unknown_mz_568.33972_+_RT_11.902	Lipid	Severe	*OSBP*	Transcript	moderate
Unknown_mz_570.50861_+_RT_23.416	Lipid	Severe	*PCYT1A*	Transcript	moderate
Unknown_mz_614.3454_-_RT_12.087	Lipid	Severe	*PECAM1*	Transcript	moderate
Unknown_mz_616.36139_-_RT_12.605	Lipid	Severe	*PLEK*	Transcript	moderate
Unknown_mz_621.30347_-_RT_11.079	Lipid	Severe	*PRF1*	Transcript	moderate
Unknown_mz_656.58331_+_RT_25.394	Lipid	Severe	*PTGDR*	Transcript	moderate
Unknown_mz_667.53271_-_RT_14.302	Lipid	Severe	*PTPA*	Transcript	moderate
Unknown_mz_684.61389_+_RT_26.119	Lipid	Severe	*RAB1A*	Transcript	moderate
Unknown_mz_708.54657_+_RT_14.3	Lipid	Severe	*RB1*	Transcript	moderate
Unknown_mz_708.62976_+_RT_27.454	Lipid	Severe	*RIT1*	Transcript	moderate
Unknown_mz_712.6452_+_RT_25.192	Lipid	Severe	*RPS20*	Transcript	moderate
Unknown_mz_715.53271_-_RT_13.482	Lipid	Severe	*SLA*	Transcript	moderate
Unknown_mz_719.56488_-_RT_14.751	Lipid	Severe	*SLC11A1*	Transcript	moderate
Unknown_mz_724.30872_-_RT_23.18	Lipid	Severe	*SOS1*	Transcript	moderate
Unknown_mz_738.65997_+_RT_26.549	Lipid	Severe	*SREBF2*	Transcript	moderate
Unknown_mz_738.66125_+_RT_27.291	Lipid	Severe	*SRF*	Transcript	moderate
Unknown_mz_738.6615_+_RT_25.394	Lipid	Severe	*STAT5A*	Transcript	moderate
Unknown_mz_739.58575_+_RT_26.834	Lipid	Severe	*SYK*	Transcript	moderate
Unknown_mz_741.60077_+_RT_27.532	Lipid	Severe	*TLR3*	Transcript	Moderate
TNF	Transcript	Moderate	Unknown_mz_360.0907_+_RT_5.792	Lipid	Moderate
TNFRSF1A	Transcript	Moderate	Unknown_mz_361.94125_-_RT_5.813	Lipid	Moderate
TXN	Transcript	Moderate	Unknown_mz_361.9415_-_RT_5.724	Lipid	Moderate
TYRO3	Transcript	Moderate	Unknown_mz_368.1171_-_RT_6.349	Lipid	Moderate
UBE3A	Transcript	Moderate	Unknown_mz_369.15408_+_RT_5.026	Lipid	Moderate
UNG	Transcript	Moderate	Unknown_mz_369.15408_+_RT_5.544	Lipid	Moderate
VWF	Transcript	Moderate	Unknown_mz_372.20126_+_RT_7.226	Lipid	Moderate
AREG	Protein	Moderate	Unknown_mz_375.21359_+_RT_6.99	Lipid	Moderate
EPO	Protein	Moderate	Unknown_mz_375.21695_-_RT_7.075	Lipid	Moderate
IL17A	Protein	Moderate	Unknown_mz_383.13434_-_RT_6.069	Lipid	Moderate
NGF	Protein	Moderate	Unknown_mz_385.14993_-_RT_5.135	Lipid	Moderate
THPO	Protein	Moderate	Unknown_mz_385.14999_-_RT_1.473	Lipid	Moderate
TNNI3	Protein	Moderate	Unknown_mz_386.18082_+_RT_2.416	Lipid	Moderate
VEGFA	Protein	Moderate	Unknown_mz_387.16437_+_RT_4.35	Lipid	Moderate
Unknown_mz_404.19128_+_RT_4.73	Lipid	Moderate	Unknown_mz_387.19397_-_RT_6.987	Lipid	Moderate
Unknown_mz_405.13147_-_RT_5.625	Lipid	Moderate	Unknown_mz_388.19608_+_RT_5.18	Lipid	Moderate
Unknown_mz_405.13184_-_RT_1.471	Lipid	Moderate	Unknown_mz_393.15167_+_RT_6.505	Lipid	Moderate
Unknown_mz_407.12833_-_RT_5.279	Lipid	Moderate	Unknown_mz_397.91782_-_RT_5.723	Lipid	Moderate
Unknown_mz_409.14612_+_RT_4.35	Lipid	Moderate	Unknown_mz_397.91791_-_RT_5.811	Lipid	Moderate
Unknown_mz_412.14255_+_RT_6.178	Lipid	Moderate	Unknown_mz_399.12878_-_RT_5.738	Lipid	Moderate
Unknown_mz_413.12015_+_RT_6.629	Lipid	Moderate	Unknown_mz_399.12924_-_RT_5.023	Lipid	Moderate
Unknown_mz_415.12424_-_RT_5.789	Lipid	Moderate	Unknown_mz_399.13174_-_RT_4.738	Lipid	Moderate
Unknown_mz_415.12424_-_RT_6.059	Lipid	Moderate	Unknown_mz_399.91483_-_RT_5.811	Lipid	Moderate
Unknown_mz_415.14279_-_RT_6.442	Lipid	Moderate	Unknown_mz_401.14484_-_RT_3.792	Lipid	Moderate
Unknown_mz_418.17029_+_RT_4.744	Lipid	Moderate	Unknown_mz_402.17587_+_RT_6.068	Lipid	Moderate
Unknown_mz_418.17029_+_RT_5.022	Lipid	Moderate	Unknown_mz_404.19128_+_RT_4.353	Lipid	Moderate
Unknown_mz_419.20432_-_RT_6.987	Lipid	Moderate	diacylglycerol (16:1/18:2 [2], 16:0/18:3 [1])*	Metabolite	Mild
Unknown_mz_431.11887_-_RT_4.204	Lipid	Moderate	dihomo-linolenoyl-choline	Metabolite	Mild
Unknown_mz_432.15057_-_RT_5.265	Lipid	Moderate	docosahexaenoylcholine	Metabolite	Mild
Unknown_mz_433.11655_-_RT_7.027	Lipid	Moderate	dodecenedioate (C12:1-DC)*	Metabolite	Mild
Unknown_mz_437.14285_-_RT_5.276	Lipid	Moderate	hexadecenedioate (C16:1-DC)*	Metabolite	Mild
Unknown_mz_440.13779_-_RT_5.443	Lipid	Moderate	lignoceroyl sphingomyelin (d18:1/24:0)	Metabolite	Mild
Unknown_mz_449.11169_-_RT_4.82	Lipid	Moderate	linoleoylcholine*	Metabolite	Mild
Unknown_mz_453.08536_-_RT_7.227	Lipid	Moderate	oleoylcholine	Metabolite	Mild
Unknown_mz_458.14764_+_RT_5.412	Lipid	Moderate	palmitoylcholine	Metabolite	Mild
Unknown_mz_471.09415_+_RT_5.274	Lipid	Moderate	phosphoethanolamine	Metabolite	Mild
Unknown_mz_476.08188_-_RT_6.986	Lipid	Moderate	sphingomyelin (d18:1/14:0, d16:1/16:0)*	Metabolite	Mild
Unknown_mz_494.01953_-_RT_6.498	Lipid	Moderate	stearoylcholine*	Metabolite	Mild
Unknown_mz_494.02023_-_RT_5.961	Lipid	Moderate	suberate (C8-DC)	Metabolite	Mild
Unknown_mz_496.01746_-_RT_5.962	Lipid	Moderate	tetradecanedioate (C14-DC)	Metabolite	Mild
Unknown_mz_506.16876_+_RT_4.424	Lipid	Moderate	tricosanoyl sphingomyelin (d18:1/23:0)*	Metabolite	Mild
Unknown_mz_511.16394_-_RT_5.692	Lipid	Moderate	Unknown_mz_589.30219_+_RT_9.071	Lipid	Mild
Unknown_mz_512.05823_-_RT_6.993	Lipid	Moderate	Unknown_mz_589.30377_-_RT_8.717	Lipid	Mild
Unknown_mz_514.05322_-_RT_6.984	Lipid	Moderate	Unknown_mz_603.31915_-_RT_7.941	Lipid	Mild
Unknown_mz_529.19226_-_RT_7.004	Lipid	Moderate	Unknown_mz_721.54974_-_RT_20.015	Lipid	Mild
Unknown_mz_529.99689_-_RT_5.25	Lipid	Moderate	Unknown_mz_765.3349_-_RT_6.822	Lipid	Mild
Unknown_mz_548.2337_+_RT_7.003	Lipid	Moderate	Unknown_mz_765.3349_+_RT_6.842	Lipid	Mild
Unknown_mz_551.17395_-_RT_6.21	Lipid	Moderate	Unknown_mz_765.5752_-_RT_20.235	Lipid	Mild
Unknown_mz_553.18915_+_RT_6.211	Lipid	Moderate	Unknown_mz_767.35071_-_RT_6.389	Lipid	Mild
Unknown_mz_557.16913_-_RT_5.138	Lipid	Moderate	Unknown_mz_769.36639_+_RT_6.393	Lipid	Mild
Unknown_mz_573.16394_-_RT_5.023	Lipid	Moderate	Unknown_mz_794.50909_-_RT_19.3	Lipid	Mild
Unknown_mz_643.24255_-_RT_5.275	Lipid	Moderate	Unknown_mz_935.31989_+_RT_6.496	Lipid	Mild
Unknown_mz_643.24255_-_RT_5.626	Lipid	Moderate	*CTLA4*	Transcript	Mild
Unknown_mz_659.23743_-_RT_5.596	Lipid	Moderate	*DDIT3*	Transcript	Mild
Unknown_mz_659.23773_-_RT_5.268	Lipid	Moderate	*ERG*	Transcript	Mild
Unknown_mz_707.32788_-_RT_7.231	Lipid	Moderate	*FLT1*	Transcript	Mild
Unknown_mz_739.31799_-_RT_5.276	Lipid	Moderate	*HAVCR2*	Transcript	Mild
Unknown_mz_753.29736_-_RT_5.638	Lipid	Moderate	*HLA-DRA*	Transcript	Mild

Unknown_mz_760.24139_-_RT_7.23	Lipid	Moderate	*IL1R2*	Transcript	Mild
Unknown_mz_761.30078_-_RT_5.276	Lipid	Moderate	*IL6ST*	Transcript	Mild
Unknown_mz_763.31421_+_RT_5.276	Lipid	Moderate	*IRF9*	Transcript	Mild
Unknown_mz_768.54962_-_RT_23.304	Lipid	Moderate	*ITGB2*	Transcript	Mild
Unknown_mz_785.24414_+_RT_7.219	Lipid	Moderate	*LAMP3*	Transcript	Mild
Unknown_mz_822.26886_-_RT_5.611	Lipid	Moderate	*MAOA*	Transcript	Mild
Unknown_mz_854.29572_-_RT_5.701	Lipid	Moderate	*MAS1*	Transcript	Mild
Unknown_mz_864.41992_-_RT_7.232	Lipid	Moderate	*MUC1*	Transcript	Mild
Unknown_mz_886.25458_-_RT_5.207	Lipid	Moderate	*MX1*	Transcript	Mild
Unknown_mz_904.3288_-_RT_5.061	Lipid	Moderate	*NCAM1*	Transcript	Mild
PC 33:3_RT_22.262	Lipid	Moderate	*NLRP3*	Transcript	Mild
PE 16:0_18:1_RT_22.866	Lipid	Moderate	*P2RX7*	Transcript	Mild
PE 18:1_18:1_RT_20.883	Lipid	Moderate	*PLAAT4*	Transcript	Mild
PE 18:2_18:1_RT_22.266	Lipid	Moderate	*RPGR*	Transcript	Mild
PE 35:1_RT_23.425	Lipid	Moderate	*S100A12*	Transcript	Mild
PE 36:2_RT_23.241	Lipid	Moderate	*S100A8*	Transcript	Mild
PE 38:5_RT_22.233	Lipid	Moderate	*SERPINA1*	Transcript	Mild
Plasmenyl-PE P-18:0_18:2_RT_22.986	Lipid	Moderate	*SH2D3C*	Transcript	Mild
Unknown_mz_1066.38086_-_RT_5.093	Lipid	Moderate	*SIK1*	Transcript	Mild
Unknown_mz_205.08609_+_RT_4.89	Lipid	Moderate	*TGFB1*	Transcript	Mild
Unknown_mz_213.96338_-_RT_4.886	Lipid	Moderate	*TRIM21*	Transcript	Mild
Unknown_mz_223.09703_-_RT_6.088	Lipid	Moderate	*VDR*	Transcript	Mild
Unknown_mz_239.09181_-_RT_4.887	Lipid	Moderate	*CCL2*	Protein	Mild
Unknown_mz_239.09184_-_RT_5.113	Lipid	Moderate	*IL13*	Protein	Mild
Unknown_mz_269.06628_-_RT_4.558	Lipid	Moderate	1-(1-enyl-palmitoyl)-2-linoleoyl-GPE (P-16:0/18:2)*	Metabolite	Mild
Unknown_mz_273.07901_+_RT_5.853	Lipid	Moderate	1-(1-enyl-stearoyl)-2-docosapentaenoyl-GPE (P-18:0/22:5n3)*	Metabolite	Mild
Unknown_mz_287.05914_-_RT_5.49	Lipid	Moderate	1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)*	Metabolite	Mild
Unknown_mz_289.07495_-_RT_4.915	Lipid	Moderate	1-adrenoyl-GPC (22:4)*	Metabolite	Mild
Unknown_mz_303.05392_-_RT_4.963	Lipid	Moderate	1-margaroyl-GPC (17:0)	Metabolite	Mild
Unknown_mz_308.11621_+_RT_4.916	Lipid	Moderate	1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	Metabolite	Mild
Unknown_mz_308.11621_+_RT_5.265	Lipid	Moderate	1-palmityl-2-linoleoyl-GPC (O-16:0/18:2)*	Metabolite	Mild
Unknown_mz_315.19626_-_RT_6.994	Lipid	Moderate	1-palmityl-GPC (O-16:0)	Metabolite	Mild
Unknown_mz_317.13824_+_RT_5.979	Lipid	Moderate	1-pentadecanoyl-GPC (15:0)*	Metabolite	Mild
Unknown_mz_319.0676_-_RT_6.574	Lipid	Moderate	1-stearyl-GPC (O-18:0)*	Metabolite	Mild
Unknown_mz_333.20648_-_RT_6.997	Lipid	Moderate	21-hydroxypregnenolone disulfate	Metabolite	Mild
Unknown_mz_349.95889_-_RT_6.159	Lipid	Moderate	6-bromotryptophan	Metabolite	Mild
Unknown_mz_351.13715_+_RT_7.772	Lipid	Moderate	arachidonoylcholine	Metabolite	Mild
Unknown_mz_352.12076_+_RT_6.347	Lipid	Moderate	behenoyl sphingomyelin (d18:1/22:0)*	Metabolite	Mild
Unknown_mz_360.0907_+_RT_5.792	Lipid	Moderate			

Compounds for which a matching pure standard was not available for confirmation are denoted by adding an asterisk (*) symbol after the name of the compound.

3.3.6 Enrichment analysis reveals enriched processes and pathways

3.3.6.1 Enrichment analysis of biosignatures that discriminate disease states based on data-driven seeds

From the discriminating features identified using the data-driven approach, we performed enrichment analysis (**see Materials and Methods**) based on the disease states they are differentially associated with (**Supplementary File 3.3**). The biological processes associated with proteins that discriminate the mild disease state are given in **Supplementary File 3.3 Table 1.** With a focus on the top 25 enriched biological processes and pathways, proteins were involved with chemotaxis (neutrophil, granulocyte, eosinophil, monocyte), regulating cytokine responses, and cell migration. These findings align with evidence of the role of chemotaxis in the initial response of detecting and destroying infected cells by following a chemical gradient of cytokines, chemokines, and other signalling molecules that are released by these cells [254]. Further, the proteins were enriched in chemokine-mediated and interleukin-mediated pathways (**Supplementary File 3.3 Table 2**). For instance, regulation of the complement cascade pathway is important for controlling the immune response and

preventing tissue damage. Several regulators of the complement system, including complement factor H, complement factor I, and CD59, are downregulated by SARS-CoV-2, which may contribute to complement dysregulation during infection [255, 256].

Transcripts discriminating the moderate disease state were enriched in those involved in cellular responses to organic substances and the mitogen-activated protein kinase (MAPK) cascade (**Supplementary File 3.3 Table 5**) but were especially enriched in transcripts involved in receptor-mediated signalling pathways (**Supplementary File 3.3 Table 6**). The MAPK pathways play a crucial role in regulating a variety of cellular responses, including cell proliferation, differentiation apoptosis, and immune response to COVID-19 [257].

The severe disease state discriminating transcripts were enriched in those involved in regulating ion transmembrane transporter activity, cell differentiation (dendritic, myeloid leukocyte), apoptotic processes, and mitochondrion organization (**Supplementary File 3.3 Table 7**) and enriched in signalling-related and regulatory-related pathways (**Supplementary File 3.3 Table 8**). The identified processes further align with the role of cell differentiation in COVID-19. For instance, dendritic cells act as sentinels to detect and respond to viral infections. They play a critical role in presenting viral antigens to T-cells, which, in turn, activate the immune response. During SARS-CoV-2 infections, dendritic cells can become infected which can lead to impaired antigen presentation and reduced activation of T-cells [258]. Also, myeloid leukocytes, including monocytes and macrophages, are important in the early immune responses to COVID-19. These cells can phagocytose viral particles and present viral antigens to T-cells, activating the immune response [259]. However, during severe infections, excessive activation of myeloid cells can lead to a cytokine storm [260], a dangerous immune response that can cause severe tissue damage and organ failure. In addition to transcripts, some lipids discriminated against the different disease states. We were, however, unable to perform enrichment analysis on these because all of the discriminating lipids are presently uncharacterized.

3.3.6.2 Enrichment analysis of biosignatures that discriminate disease states based on hypothesis-driven seeds

We repeated the enrichment analysis but with discriminatory features identified using hypothesis-driven seed selections (*IL6* and *IL6R*) (**Supplementary File 3.4 Tables 1-17**). The proteins and transcripts discriminating the mild, moderate, and severe disease states, for example, *IL13, CCL2*, *IL1A, SYK, IFNG, IL16, HMGB1, and TLR3* are involved in cytokine-, regulatory-mediated and apoptotic biological processes (**Supplementary File 3.4 Tables 1, 3, 7, 9, 14, and 17**) and were also enriched in pathways including interleukin-mediated signalling, cytokine-mediated signalling, and cellular responses to stimuli (**Supplementary File 3.4 Tables 2, 4, 8, 10, 15 and 17**). The fact that most of the biological processes and pathways are regulatory-, cytokine-, and cellular response-related agrees with other studies on disease severity. In the absence of appropriate regulatory T-cell activity to restrain the immune response to SARS-CoV-2 infections, the over-production of cytokines can ensue leading to a counter-productive cytokine storm [261, 262]. Also, apoptotic biological processes are crucial in preventing severe disease by facilitating the death of infected cells to contain both the sizes of infection foci and the immune responses to the infected cells within these foci [262, 263].

The summary of metabolite pathways (**Supplementary File 3.4 Table 5**) linked to the mild disease states revealed metabolic processes that may play important roles in the pathophysiology of COVID-19 [264-266]. For instance, sphingolipids are important components of cell membranes and are involved in a variety of cellular processes, including inflammation and apoptosis and the metabolism of these lipids has been implicated in the pathogenesis of viral infections, including COVID-19 [264, 267]. Also, the dysregulation of arginine biosynthesis and lysine degradation may play a role in the pathogenesis of COVID-19 by modulating the immune response because arginine and lysine are essential amino acids that are involved in many biological processes including immune function and protein synthesis [266]. We identified Sphingolipid metabolism processes to be common across all disease states. It has been suggested that the virus hijacks sphingolipid metabolism and dysregulates the metabolism activities to promote its replication and to evade the host immune response, aligning with the involvement of these processes in the pathogenesis of severe disease

patients (**Supplementary File 3.4 Tables 5, 6, and 12**) [267]. We also identified other pathways linked with metabolic pathways discriminating moderate and severe disease states including, but not limited to, phenylalanine, tyrosine, and tryptophan biosynthesis pathways, and the pentose phosphate pathway.

Given the inability to perform enrichment analysis for uncharacterized lipids discriminating mild disease states, the summary of lipid pathways (**Supplementary File 3.4 Table 11 and 13**) linked with moderate and severe disease states revealed processes that may play important roles in the pathophysiology of COVID-19. Autophagy is involved in several other biological processes, including antigen presentation, cell death, and immune regulation to maintain or restore homeostasis. Dysregulation of these processes has been implicated in the pathogenesis of various diseases including COVID-19 and has even been presented as a target for therapeutics [268, 269]. Arachidonic acid metabolism is the pathway responsible for the production of various bioactive lipids, including prostaglandins, leukotrienes, and thromboxanes. Dysregulation of arachidonic acid metabolism has been implicated in the pathogenesis of numerous diseases and syndromes, including inflammation, cancer, and cardiovascular disease [266].

Lipids are central components of cell membranes, such that dysregulation of pathways such as glycosylphosphatidylinositol (GPI) anchor biosynthesis and glycerophospholipid metabolism occurred in all disease states (**Supplementary File 3.4 Tables 11 and 13**). For instance, Glycosylphosphatidylinositol (GPI)-anchor biosynthesis - GPI-anchor biosynthesis is the process by which GPI-anchored proteins are synthesized and inserted into the plasma membrane. GPI-anchored proteins play critical roles in cell signalling, immune function, and development. Glycerophospholipid metabolism is the metabolic pathway responsible for the synthesis and degradation of glycerophospholipids, which are also essential components of cell membranes. Abnormalities in glycerophospholipid metabolism have been previously implicated in the pathogenesis of several diseases including COVID-19 [270, 271].

3.3.7 Comparing results from data-driven and hypothesis-driven approach

Although the methodologies differed, the findings from both data-driven and hypothesis-driven approaches have contributed to our understanding of COVID-19 disease progression. The data-driven approach was useful for uncovering unexpected findings and trends, particularly concerning *STAT1* and *SOD2* influence on disease progression thereby sparking new hypotheses and insights. However, the random walk analysis using both the data-driven and hypothesis-driven approaches yielded distinct multi-layered graphs (**Figure 3.4 and Figure 3.5**) characterized by different hubs and interactions, highlighting the unique perspectives offered by each method. With these differences, a consistent finding emerged from both approaches: cross-layer interactions between omics features play a role in disease state. Notably, both approaches revealed transcripts, especially cytokines and inflammatory biosignatures as key contributors to distinguishing disease states in both approaches. Also, both approaches revealed some overlapping biosignatures (as shown in **Table 3.** 5 and **Table 3. 6**).

3.3.8 Validating the integrative network-based multi-omics-driven data approach and replicating results from independent data.

This study performed random walk network analysis on disease-state specific omics-graphs and specifically investigated the behaviour of a new multi-layer graph generated from different datasets from the perspective of network statistical parameters (**Supplementary data**). We retrieved published metabolomics, transcriptomics, and proteomics features reported to be associated with the various COVID-19 disease states. Specifically, transcriptomics features specific to mild, moderate, and severe disease states were retrieved from Alqutami et al., [77]. Proteomics features specific to the moderate disease state were retrieved from Zhong et al., [79]. Features specific to the mild and severe disease states were retrieved from Patel et al., [81] and Suhre et al., [80]. We generated protein-protein interaction networks using GeneMANIA. Metabolites differentially associated with mild, moderate, and severe COVID-19 disease states were retrieved from Jia et al., [201]. For each disease state, by using the metabolite KEGG IDs as inputs we constructed a metabolite-metabolite interactome using MetaboAnalyst 5.0 [218], which is a knowledge-driven multi-omics integration platform (https://www.metaboanalyst.ca/).

We further constructed the metabolite-protein interactome for each disease state from MetaboAnalyst 5.0 by using the metabolite KEGG IDs and gene IDs of the features that were differentially associated with the different disease states. We also included the lipid interactome generated from Overmyer et al., [69] datasets. We repeated the random walk analysis using the data-driven and hypothesis-driven seeds; except that, since hydroxyoctanoate was not a feature in the generated networks, it was excluded from the data-driven seeds. We also performed a statistical network analysis of the multi-layered graphs generated. We observed that network heterogeneity and characteristic path length metrics correlated with disease severity (**Table 3.4)**. Following the analyses that we previously performed on our multi-layered graphs, we observed that network density decreased with disease severity. The statistical analysis also supported the observation during our multi-layered network analyses that crosstalk between features across multiple omics layers (layers containing different feature types) relates to disease severity and could be a distinctive factor underlying the heterogeneity in disease severity among patients.

3.4 Discussion

Different single omics [77-81, 198-203] and multi-omics [26, 44, 69, 76, 164, 204-207] studies have been conducted to provide insights into the aetiology of COVID-19 disease severity. However, computational network-based integrative analysis that considers different omics profiles from multiple studies with existing biological knowledgebases to explore proteomics, transcriptomics, metabolomics, and lipidomics biosignatures and their connections across different disease phases, to help explain clinical heterogeneity is limited.

Here we hypothesized that (i) investigating biosignatures across COVID-19 disease phases would provide insights into the observed clinical heterogeneity and facilitate an understanding of factors associated with disease severity, and (ii) associations between the biosignatures within a biological network would permit the prioritization of those biosignatures that discriminate the disease states, which may, in turn, provide insights into drug research. We performed an integrative multi-omics network analysis by using proteomics, transcriptomics, metabolomics, and lipidomics data from Overmyer et al. [69], and Su et al., [26]. We demonstrate an approach for harmonizing

the clinical severity of COVID-19 patients from independent studies leveraging the WOS and patient clinical metadata. In addition, through our workflow, we leverage both data-driven and hypothesis-driven approaches in an interoperable way at different biological scales, providing an impact on our understanding of the disease phases. We believe that this approach forms the basis of classifying COVID-19 patients from independent multi-omics studies and allows for the grouping of omics experimental data into disease states to perform computational network-based integrative multi-source multi-omics analysis.

From the random walk network analysis on the disease-state specific omics-graphs, we notice some unique patterns in the cross-layer interactions for mild, moderate, and severe disease states from both the hypothesis- and data-driven approaches. Specifically, random walk analysis using the hypothesis-driven seeds resulted in networks with both a greater variety of features (particularly metabolites and lipids) and more interactions between different feature types, than was achieved in networks generated using data-driven seeds. This observation might be partly attributable to the fact that *IL-6* and *IL-6R* (the two hypothesis-driven seeds that were used) have a profound role during the anti-inflammatory response to COVID-19. Compared to interactions established by *IL-6R* as a seed, we observed that *IL-6* as a seed established a network with more interactions between proteins, transcripts, and metabolites related to cell function and immune responses (accessible at http://cytoscape.h3africa.org). Whereas, *IL-6R* as a seed yielded a network that primarily captured protein and transcript interactions. Notable among the metabolites interacting with *IL-6* is taurine, an amino sulfonic acid involved in the regulation of oxidative stress, which is known to play an important role in COVID-19. Taurine levels have been reported to decrease in COVID-19 patients which potentially modulates disease progression via its antiviral, antioxidant, anti-inflammatory, and vascular-related effects [272]. Other metabolites interacting with *IL-6* included serotonin, a neurotransmitter, known to play important roles in the immune system and in regulating inflammatory responses [273].

Similarly, this study observed cross-layer patterns from the random walk analysis using the data-driven approach among *HGF, STAT1,* and *SOD2* (**Figure 3.4**). From the data-driven seeds (**Table 3.1**), we observed that *STAT1* established connections

with hubs *IRF1*, and *CCL4* in mild, and hub *IRF1* in both moderate and severe disease states (**Figure 3.4**). *SOD2* connects with hub *IRF1* in the mild, moderate, and severe disease state networks (**Figure 3.4**). Also, 3-hydroxyoctanoate interacts with 3-hydroxyexanoate and 3-hydroxydecanoate in all disease states. The random walk analysis provided not only an insight into network connectivity but also the results from the network statistical analysis (**Table 3.4**) were consistent and overlapped. Of note, the outcome from the random walk is based on the choice of seeds (**Table 3.1**) used for the random walk analysis. The analysis generated a multi-layered graph for each disease state based on the exploration of the seeds across the disease-state specific omics-graphs.

Despite the heterogeneity of COVID-19 disease outcomes, the individual mild, moderate, and severe disease states seem to have characteristic degrees to which transcript, protein, metabolite, and lipid features associatively interact both with themselves and with one another. Upon evaluating the multi-layered graphs for the various disease states, we identified several associative interactions that were present irrespective of disease state and a number that seemed to be specific for particular disease states (**Supplementary File 3.2**). These associations should be further analysed to better understand the causal effects.

In general, we observed that transcript-transcript interactions were the most commonly detected across all the disease states whereas metabolite-metabolite and lipid-lipid interactions were least commonly detected. This observation could partly be attributed to the fact that between 4 and 16 times more individual transcript features are present in the transcriptomics experimental datasets than are present in the lipidome and metabolome datasets respectively. However, irrespective of these differences, we observed that major distinctions among disease states are a result of cross-layer interactions: with protein-metabolite interactions being particularly notable. Specifically, we observed an overall increase in cross-layer interaction with disease severity using both data-driven and hypothesis-driven seeds for network exploration. We tested network statistics to confirm the network behaviour across disease states in both the original datasets and the validating datasets and found that cross-layer interactions within networks could be a distinctive feature of severe COVID-19.

From the evaluation of interactions associated with the disease states, we identified biosignatures of different omics types that discriminate specific disease states (**Table 3.** 5 **and Table 3. 6**). These biosignatures are differentially associated with the disease states [26, 69]. Further, from the enrichment analysis of these discriminatory biosignatures (**Supplementary Files 3.3 and 3.4**), we notice cytokine-, regulatory-mediated, and cellular responses to infection processes were apparent among disease-state discriminating transcripts and proteins identified from both the data-driven and the hypothesis-driven approach. This gives us a general overview that, despite the heterogeneity of COVID-19 disease outcomes, the biological processes and pathways underlying the disease could be related, but with varying expression levels of the biosignatures involved. As expected, we identified the metabolic processes related to the disease-discriminatory metabolites. However, other metabolite-related pathways relating to the degradation and/or synthesis of essential amino acids were noticeable among moderate and severe disease-state-discriminating metabolites (**Supplementary File 3.4 Tables 5, 6, and 12**). Knowing that these essential amino acid processes contribute to protein synthesis, therefore, suggests that the disturbance of protein synthesis could contribute to the severity of the disease.

This study did not consider the different treatments received by the patients reported in Su et al., [26] and Overmyer et al., [69] as a study variable during the harmonization and analysis processes. This is partly attributed to the different treatment options used and the limited treatment information reported in the study, especially by Su et al., [26], thus making it a difficult study variable to consider. We prioritized the patients' disease states by categorizing them based on the severity of their COVID-19 condition—whether mild, moderate, or severe. This classification, inclusive of patients with potential co-infections or commodifies, was maintained throughout the harmonization process and subsequent downstream analysis.

We acknowledge that both Su et al., [26] and Overmyer et al., [69] utilized different approaches in blood sample collection and processing thereof, which may have impacted the expression levels of the various multi-omics features. Also, the issue of different methods, instruments, or scales used to collect data could have contributed to heterogeneity among the datasets making data merging a non-trivial task. However, we performed data cleaning, implemented the normalization statistical method, and

performed data harmonization on the multi-omics experimental datasets (**as described in the methods section 3.2.4**) before downstream analysis as a measure of controlling heterogeneity and facilitating data merging. In addition, in this study, we combined the co-expression graphs from both studies, as a way of controlling the impact of the between-study methodological differences on the multi-omics feature expression levels as well as any bias introduced during the harmonization process. We only used lipidomics experimental data from a single study and it is likely therefore that this part of our analysis was proportionally underpowered relative to that involving transcripts, proteins, and metabolites. Furthermore, most of the lipids used for the analysis were uncharacterized and we were therefore unable to map to lipid names as well as perform enrichment analyses on them. In the multi-layered networks we created, we also did not discover any protein-lipid edge types. This may be attributable in part to the protein-lipid bipartite data and the seed exploration during random walk analysis as well as limited information on interaction involving annotated features and unannotated features. Furthermore, future investigations could consider incorporating not only characterized (annotated) lipids but also additional data types such as epigenomics, microbiomics, and immunomics. Moreover, from the analysis, we observed more transcript-transcript interactions as compared to other omics features. This observation is at least partly attributable both to the unevenness in the number of features measured in the different omics experiments (with the transcriptomics experiment examining between 5 and 100 times more features than other types of omics experiments) and the fact that more research efforts have been focused on transcriptomics analyses than on those of other omics types. Despite these limitations, our study has some obvious strengths, especially using the hypothesis-driven approach, we provide insight into the specific roles of *IL-6* and *IL-6R* in COVID-19 progression, providing targeted insights into this crucial pathway. On the other hand, the data-driven approach revealed connections between biosignatures, like *STAT1* and *SOD2*, highlighting their previously unknown influence on disease course and generating exciting new avenues for exploration. The results presented revealed biosignatures and their interactions related to disease severity. Having demonstrated that cross-layer interactions could be a distinctive feature, if not a hallmark, of severe COVID-19, it warrants deeper investigation into the potential causal relationships that these cross-layer interactions, have with disease progression: relationships that might illuminate ways to prevent and/or reverse this progression.

IL-6 is a vital innate immune cytokine in protection against other viral infections such as influenza A virus, which can cause pneumonia [274]. An increase in *IL-6* levels has previously been observed in patients with respiratory dysfunction, implying a possible mechanism of cytokine-mediated lung damage caused by COVID-19 infection [274, 275]. Our hypothesis-driven approach revealed that higher COVID-19 disease severity was associated with an increased number of interactions between *IL-6* and other multi-omics layers, therefore suggesting that our approach may discriminate between COVID-19 and other respiratory disorders. However, to confirm we will need future studies on the evaluation of network-based multi-omics approaches for COVID-19 compared to other infectious diseases, including those of viral and bacterial nature.

Our study suggests a deeper understanding of the underlying biological interactions in different phases of COVID-19 disease. The results (Figures 4 and 5, along with Tables 5 and 6) present an extensive analysis of multi-layered graphs generated by complementary approaches including data-driven and hypothesis-driven seeds and shed light on the complex interactions underlying different phases of COVID-19. These findings suggest a nuanced understanding of how various molecular features interact and influence disease severity. However, heterogeneity, different sample sizes, and sensitivity of types of samples used for sequencing across studies might be a potential source of bias. The hypothesis-driven approach offers a reductionist strategy for experimental proof whereas the data-driven approach offers a holistic strategy and hypothesis generation. We recommend considering both data- and hypothesis-driven approaches in studies utilizing multiple source omics datasets. Also, the complexity of disease severity harmonization, identifier mapping, and feature selection might be potential sources of bias in our studies.

The multi-omics harmonization process and integration strategy implemented in this study can be applied to other infectious and complex diseases, thus contributing to aggregating data from multiple sources for downstream analysis. Importantly, the algorithmic framework implemented in this study can be translated to other diseases to investigate biosignatures that underly disease progression, and relevant drug targets, and to understand disease mechanisms from the perspective of different omics layers [101]. For instance, the methods and algorithm from this study can be

used to investigate the underlying biology of complex diseases such as cancer in the context of investigating cancer subtypes and identifying the omics alterations that could help discriminate tumors, thus leading to proper diagnostics and prognosis [276]. It is however important to consider a more comprehensive list of seeds to help interpret and extend the findings.

In the biomedical context, data integration across different omics layers may help detect biosignatures connecting with genomic events to clinical factors (such as response to treatment, mRNA expression levels). This would help to predict the drivers of disease outcomes, eventually leading to better patient stratification that can be translated to better clinical tests, early intervention, and more efficient personalized therapies [276].

3.5 Conclusion

In this work, we delved into the identification and characterization of biosignatures and their specific molecular features that underly various phases of COVID-19 disease, by using an integrative and network-based approach to analyse multi-omics data. We emphasized the critical importance of integrating multi-omics data, to elucidate the molecular dynamics responsible for the wide-ranging clinical presentations of COVID-19. This integration considers both prior knowledgebases and multi-omics data from independent studies.

Our study not only pinpoints biosignatures that distinguish between disease states, but also demonstrates a correlation between the severity of disease states of COVID-19 and cross-layer interactions of proteins, transcripts, metabolites, and lipids.

We are confident that the presented multi-omics data harmonization and network-based analysis approach can also be applied to other diseases. To facilitate replication of our approach, we provide a containerized workflow with an expanded readme file at https://github.com/francis-agamah/Multi-source-multi-omics-network-analysis. All other data and its supplementary information files generated during this study are included in the github repository.

3.6 Contribution to knowledge and science

We believe that the study conducted in this chapter has significant benefits to the scientific community. Firstly, we proposed a new method of harmonizing patient disease severity metrics by leveraging the World Health Organization (WHO) Ordinal

Scale (WOS) and patient metadata. Specifically, we harmonized publicly available independent COVID-19-related multi-omics data from blood (proteomics, transcriptomics, metabolomics, and lipidomics) from Su et al., and Overmyer et al., using the World Health Organization Ordinal Scale (WOS) as disease severity reference. The novelty of the harmonization process is that it has provided the baseline to obtain uniformity across COVID-19 patient (meta)data disease state classification. This will serve as the baseline to which further multi-omics (meta)data can be included to expand the research. Also, The harmonization process enabled the classification of omics feature data into various COVID-19 disease states. Furthermore, the harmonization method presented in this chapter offers the basis to harmonize samples linked with disease severity levels or to harmonize samples linked with diseases with multiple stages by leveraging the clinical metadata information.

The containerized workflow and scripts are available at the GitHub repository, thus contributing to knowledge and science by adhering to FAIR principles and providing a consistent and reproducible way to execute experiments and analyze data. They allow for easier collaboration and sharing of methods and results, leading to more efficient and accurate research.

This study has yielded novel perspectives on the biological underpinnings of COVID-19, shedding light on the role played by proteins, transcripts, metabolites, and lipids in modulating the disease's trajectory through their interplay across the spectrum of mild, moderate, and severe disease phases.
This study also offers essential data regarding biomarkers that distinguish between these stages of the disease, potentially contributing to the prediction of COVID-19 outcomes. Notably, the severity of the disease correlates with the increasing number of predicted interactions among various biological layers. Such insights have the potential to enhance the precision with which healthcare providers can forecast patient prognoses.

Overall, the research presented in this chapter suggests that this kind of multi-omics data integration is essential when attempting to explain the molecular dynamics underpinning the heterogeneity of COVID-19 infections while accounting for both prior knowledgebase and data from independent studies. This study has considerably

enhanced our knowledge and understanding of COVID-19, laying the groundwork for the formulation of more efficacious therapeutic and preventative measures.

Chapter 4

Network-based multi-omics-disease-drug associations reveal drug repurposing candidates for COVID-19 disease phases

Francis E. Agamah[1], Thomas H.A. Ederveen[2], Michelle Skelton[1], Darren P. Martin[1], Emile R. Chimusa[3*], Peter A.C. 't Hoen[2*]

[1]Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa.
[2]Medical BioSciences department, Radboud University Medical Center Nijmegen, The Netherlands.
[3]Department of Applied Science, Faculty of Health and Life Sciences, Northumbria University, Newcastle, Tyne and Wear, NE1 8ST, UK.

doi: https://doi.org/10.5281/zenodo.10568146

Abstract

Background: The development and roll-out of vaccines, and the use of various drugs have contributed to controlling the COVID-19 pandemic. Nevertheless, challenges such as the inequitable distribution of vaccines, the influence of emerging viral lineages and immune evasive variants on vaccine efficacy, and the inadequate immune defense in subgroups of the population continue to motivate the development of new measures to combat the disease.

Aim: In this study, we sought to identify, prioritize, and characterize drug repurposing candidates appropriate for treating mild, moderate, or severe COVID-19 using a network-based integrative approach that systematically integrates drug-related data and multi-omics datasets.

Methods: Here we leveraged drug data, and multi-omics data, and used a random walk restart algorithm called multiXrank to explore an integrated knowledge graph comprised of three sub-graphs: (i) a COVID-19 knowledge graph, (ii) a drug repurposing knowledge graph, and (iii) a COVID-19 disease-state specific omics graph.

Results: We prioritized twenty FDA-approved agents as potential candidate drugs for mild, moderate, and severe COVID-19 disease phases. Specifically, drugs that could stimulate immune cell recruitment and activation including histamine, curcumin, and paclitaxel have potential utility in mild disease states to mitigate disease progression. Similarly, non-steroidal anti-inflammatory drugs including indomethacin and diclofenac offer potential utility for mild cases. Drugs like omacetaxine, crizotinib, and vorinostat that exhibit antiviral properties and have the potential to inhibit viral replication can be considered for mild to moderate COVID-19 disease states. Also, given the association between antioxidant deficiency and high inflammatory factors that trigger cytokine storm, antioxidants like glutathione can be considered for moderate disease states. Drugs that exhibit potent anti-inflammatory effects like (i) anti-inflammatory drugs: sarilumab and tocilizumab, (ii) corticosteroids: dexamethasone and hydrocortisone, (iii) immunosuppressives: sirolimus and cyclosporine are potential candidates for moderate to severe disease states that trigger an inflammatory cascade of COVID-19.

Conclusion: Our study demonstrates that the integrative analysis of drug data and multi-omics data enables prioritizing drug candidates for COVID-19 disease phases,

offering a comprehensive basis for therapeutic strategies that can be brought to market quickly given their established safety profiles.

Keywords: multi-omics, drug repurposing, random walk, COVID-19, networks

4.1 Background

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the highly contagious and virulent coronavirus responsible for the global outbreak of coronavirus disease 2019 (COVID-19). Between 2020 and 2023 COVID-19 imposed an unprecedented burden on global public health systems and by December 2023, had been responsible for at least seven hundred and seventy million reported cases and close to seven million reported deaths [7].

The pandemic's containment and the restoration of societal normalcy have been achieved through two key routes: (1) the development and widespread use of SARS-CoV-2 vaccines, and (2) the gradual increase in natural infection-acquired immunity. While progress has been made, critical hurdles remain, notably the need for equitable vaccine distribution and effective treatment options for those unvaccinated or immunocompromised [84]. Furthermore, there are also some concerns surrounding vaccine efficacy against a backdrop of waning immunity [84, 85] and the emergence of immune evasive viral lineages such as B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), B.1.617.2 (Delta) and Omicron lineages, including BA.2.75, BA.4, BA.4.6, and BA.5 [277, 278]. It is now a widely held view that COVID-19 is likely to transform into another endemic human coronavirus, possibly with seasonal epidemic waves [279]. Irrespective of the number of infections that occur or the intensively with which vaccines are used, it remains uncertain whether true herd immunity of the sorts achieved with measles and rubella will ever be achieved for COVID-19.

While data is currently being collected and analysed to understand how newly evolved SARS-CoV-2 variants might impact the effectiveness of vaccines and the severity of future COVID-19 infection waves, there remains a demand for both host-directed and pathogen-directed drugs that could be utilized to treat the mild, moderate, and severe manifestations of the disease.

In light of this, multiple existing drugs have been sought to treat or control SARS-CoV-2 infection [280]. An example is *ritonavir-boosted nirmatrelvir* (*paxlovid*) (DrugBank: DB16691), a protease inhibitor used for the treatment of mild to moderate COVID-19

in adults who are at high risk of developing severe symptoms [281]. Additionally, *remdesivir* (Veklury) (DrugBank: DB14761), an adenosine triphosphate analogue targeting the conserved viral RNA-dependent RNA polymerase (RdRp), shortens the recovery time for adults hospitalized with COVID-19 infection and pneumonia, while also mitigating disease severity and associated mortality [282-284]. Other virus-directed antiviral drugs like *favipiravir* (DrugBank: DB12466), *molnupiravir* (DrugBank: DB15661), and *ritonavir* (DrugBank: DB00503), may also potentially improve the health outcomes of COVID-19 patients [285].

Furthermore, a host-directed drug such as *dexamethasone* (DrugBank: DB01234), an anti-inflammatory corticosteroid, has demonstrated its effectiveness in reducing mortality among severely infected patients. It achieves this by modulating inflammation-mediated lung injury, preventing in some cases progression to respiratory failure and death [286, 287]. Another host-directed drug, *aspirin* (DrugBank: DB00945) [15], decreases the risk of complications, and mortality in hospitalized COVID-19-infected patients [288-290].

While these drugs are mainly utilized for severe COVID-19 cases, some virus-directed monoclonal antibodies (e.g., *bebtelovimab, casirivimab*) received emergency-use authorization at various stages of the pandemic for managing mild to moderate COVID-19 [291]. Additionally, other monoclonal antibodies including the combination of *bebtelovimab, casirivimab,* and *imdevimab* (CAS/IMDEV), and the combination of *bamlanivimab* and *etesevimab* (BAM/ETE) have been useful for managing mild to moderate COVID-19 in adults [291]. These monoclonal antibodies inhibit viral entry into host cells by preventing viral attachment to human *ACE2* receptors.
Host-targeted monoclonal antibodies such as *tocilizumab* (DrugBank: DB06273) and *sarilumab* (DrugBank: DB11767), [292-295], which modulate aberrant immune responses to infection by binding to the host *IL6* receptor (*IL6R*), have also been granted emergency use authorization. Specifically, these IL6R-binding monoclonals are used for treating hospitalized adults and peadiatric patients (2 years of age and older) who are receiving systemic corticosteroids and require supplemental oxygen, non-invasive or invasive mechanical ventilation, or extracorporeal membrane oxygenation [295].

Although existing drugs have been recommended for managing COVID-19, concerns have arisen about post-hospitalization effects and the appropriateness of using these drugs in different COVID-19 disease phases [296-298]. COVID-19 exhibits a wide spectrum of symptoms and severities, necessitating a nuanced approach to treatment. Personalized medicine, where the right drug is administered to the right patient group at the right disease phase (and time), could revolutionize COVID-19 treatment strategies. Hence, it would be useful to identify effective drugs that are specific to different phases of the disease, and which could also be potentially applicable to combatting any future emergence of other coronaviruses.

Given that the development of new medications is a time-consuming process, the repurposing of existing medications for other indications may prove to be a viable alternative. Applying computational approaches to the repurposing of existing medications is, in fact, a potentially and highly efficient means of drug discovery since the pharmacological properties, formulations, and toxicities of such agents are already known [299-305]. However, most studies that have implemented computational methods to identify drug repurposing candidates for COVID-19 have so far leveraged disease-gene associations, protein-protein interaction, and drug-target data but less consideration is given to the interactions between other biomedical and molecular features specific to different COVID-19 disease phases, such as recorded in large scale multi-omics profiling efforts [300, 301, 304, 305]. In this study, we explore the utility of incorporating disease-state specific omics-graph along with drug-related data to identify drug repurposing candidates for mild, moderate, or severe COVID-19 disease states. We employ multiXrank [214], a random walk with restart (RWR) algorithm that can combine multiple heterogeneous networks and allows for universal multi-layer network exploration. We demonstrate that this integrative multi-omics network-based approach with drug data has the potential to repurpose drugs in different disease states.

4.2 Materials and Methods

4.2.1 Study design and procedures

The methodology employed in this study (**Figure 4.1)** encompasses five main steps: (1) curation and pre-processing of data related to the action of drugs and to the molecular omics profiles associated with the different phases of the disease; (2) multi-layer network-based random walk analysis; (3) predicting drug repurposing candidates; (4) drug prediction robustness analysis; and (5) validation of predicted candidate drugs.

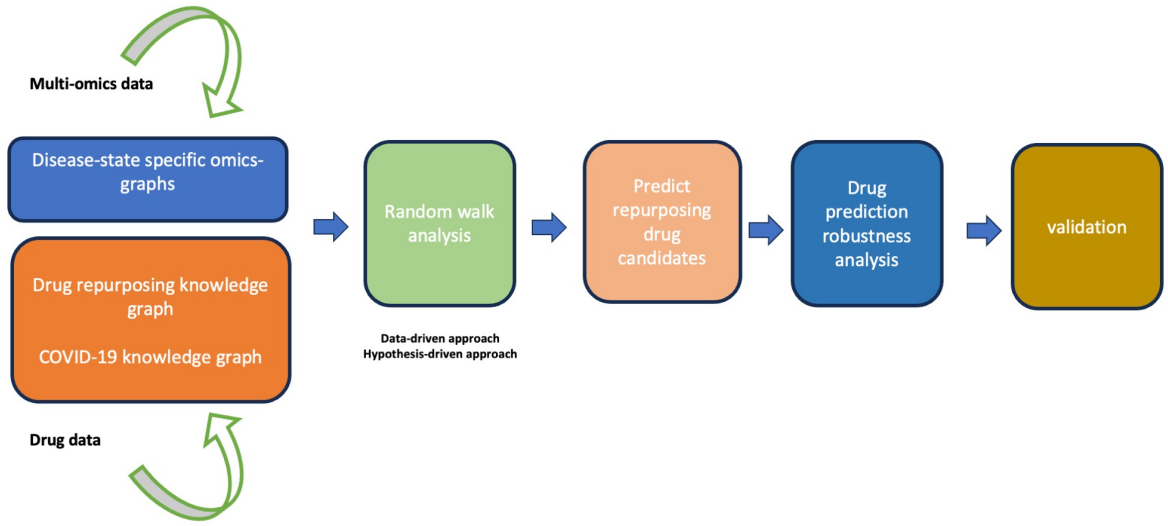


Figure 4.1. *Diagram illustrating the workflow implemented in this study. The workflow begins with curating multi-omics data and drug data followed by a random walk with restart network analysis using both data-driven and hypothesis-driven approaches. Next, we prioritize and characterize candidate drugs followed by drug prediction robustness analysis. Finally, we conclude the analysis by validating the predicted drug candidates.*

4.2.2 Drug repurposing knowledge graph, COVID-19 knowledge graph, and disease-state specific omics-graphs

We utilized an existing drug repurposing knowledge graph constructed by Ioannidis *et al.*, [299]. The Drug Repurposing Knowledge Graph (DRKG) is a biological knowledge graph relating genes, compounds, diseases, biological processes, side effects, and symptoms as of 2020 when it was constructed. The DRKG includes information from six existing databases: (1) Global Network of Biomedical Relationships (GNBR) [306], (2) STRING [307], (3) IntAct [308], (4) Hetionet [309], (5) DrugBank [310] and, (6)

Drug-Gene Interaction database (DGIdb) [311]. The DRKG includes 97,238 entities classified into 13 different node types (**Table 4.1**) and consists of 5,874,261 triplets belonging to 107 edge types (**Supplementary Table 4.1**). We leveraged the gene-pathway and gene-biological process edge types (**Supplementary Table 4.1**) in the DRKG to construct pairwise interaction between biological processes and pathways based on the semantic relation that biological processes and pathways that share similar disease-related genes are indirectly associated. This was achieved by exploring the associations between genes, biological processes, and pathways to investigate the biological processes and pathways enriched among the disease-associated genes. We then paired pathways and biological processes sharing common genes.

Table 4.1. *Description of the node-types in drug repurposing knowledge graph*

Node-type	Number of features
Anatomy	400
Anatomical Therapeutic Chemical (Atc)	4048
Biological Process	11381
Cellular Component	1391
Compound	24313
Disease	5103
Gene	39220
Molecular Function	2884
Pathway	1822
Pharmacologic Class	345
Side Effect	5701
Symptom	415
Taxonomy (Tax)	215

We additionally considered the COVID-19 knowledge graph (COVID-19 KG) built by Hsieh *et al.*, in 2021 [302]. This COVID-KG is notable for its integration of drug data from the Comparative Toxicogenomics Database (CTDbase), specifically data available as of 2021 [312], along with protein-protein interactions involving SARS-CoV-2 and host proteins from a study by Gordon *et al.*, [313]. In addition to the virus-host interaction data, we also included SARS-CoV-2 and host protein interactions from IntAct database [308]. The SARS-CoV-2 and host protein interactions data extracted from IntAct were derived from several studies that examined protein-protein

interactions between SARS-CoV-2 and humans. Following the merge of the COVID-KG and virus-human protein interaction data, the resulting graph represents the interactions between entities belonging to 5 different node types (**Table 4.2**) and consists of 33,621 triplets belonging to 5 edge types (**Supplementary Table 4.2**).

Table 4.2. *Description of the node-types in COVID-19 Knowledge Graph*

Node-type	Number of features
SARS-CoV-2 baits	23
Host genes and drug targets	10959
Pathways	274
Drugs (chemical/compound)	4266
Biological process (Phenotypes)	1893

Furthermore, we considered disease-state specific omics-graphs (DSOG) constructed from chapter 3 for downstream analysis. The DSOG was constructed by integrating harmonized proteomics, transcriptomics, metabolomics, and lipidomics datasets retrieved from Overmyer *et al.*, [82] and Su *et al.*, [314] together with a unified knowledge graph, assembled by merging protein-protein interactome, metabolite-metabolite interactome, transcript-transcript, and lipid-lipid interactome curated from literature and databases [315]. The DSOG consisted of four node types (i.e., protein, transcript, metabolite, lipid) and nine edge-types (i.e., protein-protein, transcript-transcript, metabolite-metabolite, lipid-lipid, protein-transcript, protein-metabolite, transcript-metabolite, protein-lipid, and transcript-lipid).

The process used to construct the graphs is described in Agamah *et al.*, [315]. In summary, the World Health Organization (WHO) Ordinal Scale (WOS) was used as a disease severity reference to harmonize COVID-19 patient metadata across the Overmyer *et al.*, [82] and Su *et al.*, [314] studies. This harmonized metadata was then used to categorize the multi-omics data into mild, moderate, and severe COVID-19 disease phases. Subsequently, a correlation network approach was implemented to construct co-expression networks for proteomics, transcriptomics, metabolomics, and lipidomics data for each disease state. The co-expression networks generated were

integrated/merged based on the disease state and omics data type to construct disease-state specific omics-graphs.

Overall, these above data sources were used to construct a unified and integrated knowledge graph comprised of three sub-graphs including COVID-19 KG, DRKG, and the DSOG from which we conducted an in-depth quality check (see the section below), prioritize, characterize, and repurpose specific drugs for the mild, moderate, and severe state of COVID-19.

4.2.3 Data pre-processing, quality control, and filtering

We observed differences in the gene and drug identifiers across the DRKG and COVID-19 KG. To achieve consistency in the identifiers across the datasets, we mapped gene identifiers to gene symbols using the UniProt database resource [316], and drug identifiers to drug names using the DrugBank database [310]. To identify clinically approved drugs with known safety profiles and pharmacokinetic properties, we filtered/cleaned the drug-related data by maintaining (1) drug-drug interactions between FDA-approved drug candidates, (2) drug-protein/gene interactions between FDA-approved drug candidates and proteins/genes, (3) biological process/pathways-chemical interactions between biological processes/pathways and FDA approved chemicals. Additionally, as our investigation is centered around identifying potential repurposable drugs for COVID-19, we took additional measures to refine our analysis. To prioritize approved drugs with known safety profiles for COVID-19, we specifically removed interactions involving drugs that have been studied (based on literature evidence) and found to lack therapeutic effectiveness in treating COVID-19, as is the case with quinolones like chloroquine and hydroxychloroquine [317, 318]. Moreover, we omitted certain substances including hormones such as progesterone, testosterone, and melatonin, as well as alcohol (ethanol), various compounds like cholesterol, and cocaine, and gases such as oxygen, and hydrogen from our analysis. These substances have limited safety profiles and effectiveness in treating COVID-19.

4.2.4 Random walk with restart network analysis

To perform random walk with restart (RWR) network analyses, we utilized multiXrank [214], a RWR on a multilayer network algorithm, to explore the disease-state specific

omics-graphs, COVID-KG and DRKG. Whereas the multiXrank algorithm enables random walk analysis on multiple large multidimensional datasets in a multilayer network framework, other methods are limited in the combination and heterogeneity of networks that they can handle [208]. We made specific adjustments to the algorithm's configuration script based on the input datasets. While the global restart probability was set at 0.7, the intra-layer jump probability and the probability to restart in a specific layer were both set to 0.5 and 1, respectively. Other parameters such as inter-layer jump probability were fine-tuned to align with the number of input graph layers. As a result of this analysis, we obtained multi-layered graphs that detailed the exploration of seed nodes across various omics layers and drug data, along with a ranked list of features in each graph layer.

4.2.4.1 RWR analysis on DRKG, COVID-19 KG, and DSOG

We performed an initial analysis focused on identifying drugs that can be repurposed for COVID-19 without accounting for omics profiles. Thus, we employed DRKG and COVID-19 KG as the input data sources in the RWR algorithm to predict candidate drugs for COVID-19. The algorithm accepts as layers, monoplex graphs, and/or a combination of monoplex graphs (multiplex). Specifically, in our analysis, edge-types with the same node entities including the drug-drug and gene-gene interactions each served as a monoplex and were interconnected by edge-types with different node entities including the gene-drug, pathway-gene, phenotype (biological process)-gene, phenotype (biological process)-drug, and SARS-CoV-2-host gene interactions (**Supplementary Table 4.1 and Supplementary Table 4.2**).

In subsequent analysis to predict potential drugs for different COVID-19 disease states, we utilized DRKG, COVID-19 KG, and DSOG as input data to predict drug repurposing candidates for COVID-19. Herein, the input data included the proteomics, transcriptomics, metabolomics, and lipidomics disease-state graphs which also served as individual monoplex layers. Similar to the analysis on DRKG and COVID-19, graphs with edges-types of the same node entities were interconnected by graphs with edge-types of different node entities including the protein-transcript, protein-metabolite, transcript-metabolite, protein-lipid, and transcript-lipid interactions from the DSOG which were not utilized in the previous analysis.

Prediction of drugs using the RWR algorithm is based on a network exploration process where simulated particles walk iteratively from one node to one of its neighbours with some probability. In this process, the walk is restricted to restart from seed nodes to prevent the random walker from being trapped in dead-ends [214]. We similarly implement the hypothesis-driven and data-driven approaches for the selection of seeds (**see Section 3.2.8.2**).

4.2.4.2 Selection of seeds for RWR based on a hypothesis-driven approach

Interleukin-6 (*IL-6*) and interleukin 6 receptor **(***IL-6R*) features were used as hypothesis-driven seeds for the random walk analysis because of the evidence for their significant role in the pathology of SARS-CoV-2 and COVID-19 [319-321]. *IL-6* is a cytokine, a type of signalling molecule involved in various inflammatory and immune responses. The inflammatory response plays a critical role in COVID-19, with an excessive inflammatory response leading to a "cytokine storm" increasing the severity of COVID-19. Since *IL-6* interacts with cells via *IL-6R,* it has been hypothesized that inhibition of *IL-6R* might reduce the likelihood of cytokine storms developing, ameliorate the symptoms of severe COVID-19, and reduce mortality [49]. In this context, we, therefore, used *IL-6* and *IL-6R* as hypothesis-driven seeds in a RWR analysis that we refer to as our "hypothesis-driven approach".

4.2.4.3 Selection of seeds for RWR based on a data-driven approach

For a random walk with restart using data-driven seeds, we selected seeds following the approach described in our previous study [315]. Specifically, we selected, after merging the different co-expression networks, the features with the highest integrated centrality score in each omics layer [315]. The features were ranked by leveraging the node degree, closeness, betweenness, and eigenvector centrality metrics to compute an integrated score (see **Supplementary data equation 1**). Herein, RWR analysis using the data-driven seeds is referred to as our "data-driven approach".

4.2.4.4 Ranking candidate drugs for COVID-19 disease states

The RWR approach has the benefit of capturing the global topology of a graph and representing a measure of proximity from all the nodes to the seed(s) based on the graph topology [322]. The measure of proximity between nodes is a relevant measure

quantifying how closely connected a node is with the seeds and can be used to rank nodes [214, 322, 323]. In this study, nodes within each layer were ranked based on their measure of proximity to the seed nodes. The measure of proximity was the geometric mean of the node's proximity to the seeds [214].

4.3 Results

4.3.1 Predicting candidate drugs using existing knowledge graphs and hypothesis-driven seeds.

To predict potential COVID-19 drugs, we systematically analysed the DRKG and COVID-19 KG excluding disease-state specific omics-graphs by implementing a random walk with restart analysis. The random walk with restart is an approach that allows for the exploration of the disease-state specific omics-graphs, COVID-19 KG, and DRKG to identify patterns and prioritize features within the network. The algorithm, multiXrank, conducts multiple random walks over the graphs, each originating from the seed nodes. These walks iteratively traverse from one node to a neighbouring node at random, thus simulating a pattern that results in a multi-layered graph. In our hypothesis-driven approach, we selected, *IL-6* and *IL-6R* as seeds given their established roles as aggravators of the disease. The random walk process restricts the restarts from seed nodes (*IL-6* and *IL-6R)* during network exploration. The RWR analysis revealed a multi-layered graph describing the random walks from the seed nodes (**Figure 4.2**) and a set of potential therapeutics (**Table 4.3**) for the treatment of COVID-19. These included immunosuppressants, vital minerals, anticancer agents, antivirals, antibiotics, angiotensin receptor blockers, and corticosteroids (as detailed in **Table 4.3**). Notable among these are presently recommended drugs for treating the disease such as tocilizumab [226, 295, 324], dexamethasone [286, 287], and losartan [325, 326]. Furthermore, our analysis pinpointed additional potential pharmaceutical options, such as *cannabidol* and *doxorubicin* [327-329]. While these have yet to be definitively endorsed for COVID-19 treatment, prior research indicates their potential candidacy based on their efficacy and merits for further experimental validations [327, 330]. We show in **Figure 4.2**, a multi-layered graph generated from the RWR analysis highlighting interactions involving highly ranked drug candidates and other features.

***Table 4.3**. Top 20 potential COVID-19 drugs, ranked according to their measure of proximity to IL-6 and IL-6R seed nodes as determined through RWR analysis of DRKG and COVID-19 KG. The references point to publications that have reported the drugs' mechanism of action potentially linked with COVID-19.*

Drug name	Drug Category	Mechanism of action potentially linked with COVID-19	Measure of proximity	Reference*
Tocilizumab	Interleukin-6 (IL-6) receptor antagonist	Suppresses immune response by blocking IL-6 signalling	0.0353861	[52]
Zinc	Essential mineral / Nutrient	Interferes with viral RNA synthesis to inhibit replication	0.0007941	[331]
Sirolimus	Immunosuppressive drug	Expresses immunomodulatory and anti-inflammatory properties and inhibits the expression of proinflammatory cytokines.	0.0003784	[332, 333]
Choline	Essential nutrient	Supports cell membrane integrity and neurotransmitter function	0.0003440	[334]
Ivermectin	Antiparasitic drug	Inhibits viral replication and modulate the host immune response	0.0003056	[335]
Dactinomycin	Anticancer	Expresses immune modulatory properties and inhibits viral cellular transcription	0.0002734	[336, 337]
Losartan	Angiotensin receptor blocker	Reduce the activity of the renin-angiotensin system	0.0002327	[338]
Ribavirin	Antiviral	Interferes with viral RNA synthesis and replication	0.0002168	[339, 340]
Azithromycin	Antibiotic	Expresses anti-viral and anti-inflammatory properties	0.0002000	[341]
Tenofovir	Antiviral	Interferes with viral RNA synthesis to inhibit replication	0.0001930	[342]
Acetaminophen	Analgesic	Expresses antipyretic and analgesic effects and inhibit the cyclooxygenase (COX) pathways.	0.0001665	[343, 344]
Dexamethasone	Corticosteroid	Suppresses immune response	0.0001633	[345]
Methotrexate	Immunosuppressive drug	Suppresses immune response	0.0001624	
Cyclosporine	Immunosuppressive drug	Express anti-inflammatory and anti-viral properties	0.0001614	[346, 347]
Cisplatin	Anticancer		0.0001555	
Tacrolimus	Immunosuppressive drug	Mitigate the hyperinflammatory response	0.0001532	[348]
Indomethacin	Non-steroidal anti-inflammatory drug	Expresses anti-inflammatory properties and reduces pain and fever	0.0001505	[330, 349]
Cannabidiol	Cannabinoid	Inhibits viral replication by up-regulating the host inositol-requiring enzyme-1α ribonuclease endoplasmic reticulum stress response and interferon signalling pathways	0.0001488	[327]
Doxorubicin	Anticancer	Expresses antiviral and immunomodulatory properties.	0.0001486	[328]
Diclofenac	Non-steroidal anti-inflammatory drug	Expresses anti-inflammatory properties and reduces pain and fever	0.0001481	[349]

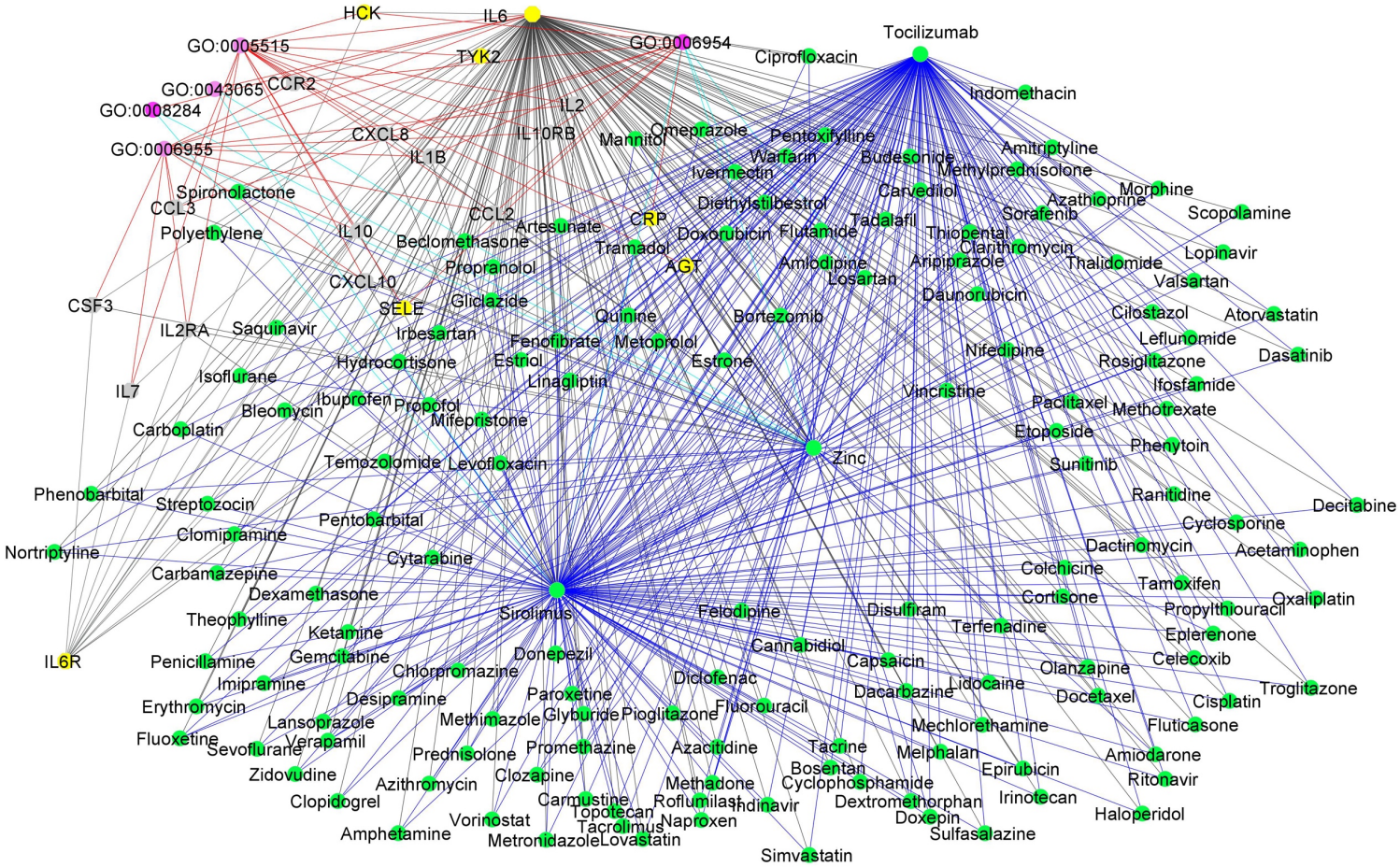


Figure 4.2*. Graph representation of interactions between drugs and other features as observed from predicting candidate drugs using existing knowledge graphs and hypothesis-driven seeds. Blue edges represent interactions between drugs (green nodes). Cyan edges represent interactions between biological processes (pink nodes) and drugs. Red edges represent biological process-proteins (grey nodes) interactions and biological process-transcript (yellow nodes). Black edges represent drug-protein, drug-transcript, protein-transcript, and transcript-transcript interactions. The graph was generated by defining filtering criteria based on node degree between 4 and 1633 inclusive from cytoscape.*

4.3.2 Predicting candidate drugs using existing knowledge graphs, disease-state specific omics-graphs, and hypothesis-driven seeds

To discover potential drug candidates that could be used to specifically treat the disease in either its mild, moderate, or severe phases, we employed the RWR method in a network-based approach utilizing the knowledge graphs (DRKG and COVID-19 KG), alongside the disease-state specific omics-graphs encompassing transcriptomics, proteomics, metabolomics, and lipidomics data. The outcome of the RWR analysis is a ranked list of drugs and a multi-layered graph describing the random walks from the seed nodes (*IL-6* and *IL-6R*). Across different analyses of the

mild, moderate, and severe disease states we consistently identified the same sets of drugs that could potentially be impactful during COVID-19 treatment.

Notably, we found that drugs known to mitigate inflammation by suppressing the immune response (including drugs that promote the production of *IL-6*) were assigned high rankings across all the disease states albeit with slightly varying measures of proximity (**Table 4.4**). Among these drugs are *dexamethasone*, *sirolimus, sarilumab, and tocilizumab*. This observation aligns with expectations, considering our choice of *IL-6* and *IL-6R* as seeds: both are pivotal biomarkers identified in multiple studies as displaying expression levels that are positively associated with severe disease states [239, 350]. These biomarkers are prominently expressed during cytokine storms that are characteristic of severe COVID-19 cases [239]. For instance, anti-IL6R binds to membrane-bound and soluble *IL-6R*, thus blocking the cis- and trans-inflammatory signalling cascade of *IL-6* [350].

Some of the predicted drugs in our analysis are being or have already been, tested in various clinical trials to assess their efficacy and effectiveness in the treatment of COVID-19. Noteworthy among these are corticosteroids such as *dexamethasone* ranked first (**Table 4.4**; with a measure of proximity 0.0033655, 0.0033691, and 0.0033697 for mild, moderate, and severe disease states respectively), and *hydrocortisone* ranked thirteenth (**Table 4.4;** with measures of proximity 0.0001274, 0.0001279, and 0.0001276 for mild, moderate, and severe respectively) which have demonstrated an association with lower 28-day all-cause mortality in critically ill patients with COVID-19 [345]. Others include, *mycophenolic acid* (ranked fifteenth; **Table 4.4**; with measures of proximity 0.0001236, 0.0001236, and 0.0001239 for mild, moderate, and severe respectively) which has been investigated and validated to reduce mortality and hospital stays in patients with moderate to severe COVID-19 [351], *indomethacin* (ranked eighteenth; **Table 4.4**; with measures of proximity 0.0001170, 0.0001165, and 0.0001162) which has been found in clinical trials to be safe and effective for treating mild and moderate COVID-19 cases [330], and the diabetes medication, *metformin*, (ranked twentieth; **Table 4.4**; with measures of proximity 0.0001158, 0.0001158, and 0.0001164 for mild, moderate, and severe respectively) which exhibits potential in reducing prolonged illness by inhibiting virus replication when administered during the acute phase of COVID-19 [352].

We observed differences between the multi-layered graphs generated by the RWR analysis on the mild, moderate, and severe disease states**.** Specifically, these differences were observed in the levels of connectivity between molecular features and the drug repurposing candidates. We therefore performed network topology analysis on the multi-layered graphs generated from the RWR analysis to explore the differences between these graphs and how that can provide more insights into the use of predicted drugs across mild, moderate, and severe disease states. First, we evaluated the degree, betweenness, and closeness measures of the drug repurposing candidates and observed relatively similar scores (**Table 4.5**). The node degree indicates the number of connections a drug has, revealing its involvement in broader disease processes. Higher degrees might suggest broader applicability across multiple diseases. The betweenness centrality measures a drug's "traffic control" role, indicating how often it lies on the shortest paths between other nodes. High betweenness suggests a potential "bridge" molecule connecting different disease pathways. The closeness centrality reflects how quickly information can reach other disease elements or nodes from a candidate drug. High closeness suggests a central position within the disease network, potentially making it a good starting point for treatment. For instance, cyclosporine had a higher degree of 6 in the multi-layered graph generated by RWR analysis on the moderate disease state as compared degree of 5 in the multi-layered graph generated by RWR analysis on the mild disease state and a degree of 4 in the multi-layered graph generated by RWR analysis on the severe disease state with a corresponding high betweenness score of 0.0014 and closeness score of 0.4832 suggesting it might have potential utility for moderate disease state. Similarly, mycophenolic acid had a higher degree of 5 in the multi-layered graph generated by RWR for moderate disease state as compared to a degree of 4 in both multi-layered graph generated by RWR analysis on the mild and multi-layered graph generated by RWR analysis on the severe disease state with a corresponding high betweenness score of 0.009 and closeness score of 0.4739 also suggests it might have potential utility for moderate disease state.

Next, we evaluated the contribution of other molecular features in the networks. To begin with, beyond the central influence of top-ranked drugs (*dexamethasone, tocilizumab,* and *sarilumab*) in the multi-layered graph generated from the RWR

analysis on the mild disease state (**Figure 4.3A**), three key inflammatory-related features: C-C Motif Chemokine Ligand 2 (*CCL2*), C-C Motif Chemokine Ligand 4 (*CCL4*), and Negative Elongation Factor Complex Member C/D (*NELFCD*) formed distinct subnetworks, acting as crucial hubs that connected seed nodes and promising candidates for drug repurposing. In mild COVID-19 cases, *CCL2* helps recruit monocytes and macrophages, which are essential for fighting the virus as compared to excessive immune cell recruitment in severe disease states [353, 354]. Also, *CCL4* levels in mild disease states help recruit necessary immune cells to fight the virus [355]. Thus, using drug repurposing candidates that could influence the recruitment of immune cells in mild disease states could be more appropriate; acknowledging the fact that the development of clinical COVID-19 involves cell activation such as dysfunctional mast cell activation [356]. From the predicted drugs, histamine is a biogenic amine known to attract and activate immune cells, particularly mast cells, basophils, neutrophils, and certain T cells, through specific histamine receptors [357]. Histamine can stimulate mast cell degranulation, leading to the release of *CCL4* and *CCL2* among other inflammatory mediators [358, 359]. This suggests a potential indirect link between histamine and these chemokines in inflammatory processes. Paclitaxel is known to modulate the immune system in various ways, including; (1) promoting the migration of T cells and other immune cells into tumours, (2) enhancing the activity of antigen-presenting cells, vital for activating T cells, and (3) modulating the expression of immune-related genes that influence inflammation and immune responses. Paclitaxel induces the release of cytokines like *TNF*, and *IL-6* and chemokines like *CCL2*, and might, therefore, help control viral infection by stimulating immune cell recruitment and boosting immune responses [360]. Metformin might enhance the activity of certain immune cells like macrophages and natural killer cells potentially aiding in viral clearance [361, 362]. Metformin activates AMPK, a cellular energy sensor that regulates various metabolic and inflammatory processes [363]. AMPK activation can downregulate pro-inflammatory signalling pathways and reduce the production of inflammatory mediators like *TNF* and *IL-6* [364].

The *NELFCD*, as part of NELF, regulates RNA polymerase II pausing, potentially influencing viral RNA synthesis during viral genome replication [365]. This suggests that drugs with the potential to inhibit SARS-CoV-2 replication could be appropriate for repurposing. Such drugs would include, for example, dactinomycin which is, besides having immune-modulatory properties also inhibits viral genome replication [336, 337].

Analysis of the multi-layered graph generated by the RWR analysis during predictions for moderate disease state identified *NELFCD*, Nuclear Factor Kappa B Subunit 1 (*NFKB1*), and interleukin 10 (*IL-10*) as hubs influencing the network based on their high connectivity, forming both direct and indirect pathways between the seed nodes and the top candidates for drug repurposing **(Figure 4.3B)**. *NFKB1* activates genes encoding pro-inflammatory cytokines, chemokines, and adhesion molecules, orchestrating the body's initial response to viral infection and enhancing the severity of COVID-19 symptoms [366, 367]. Thus, moderating *NFKB1* activity could mitigate cytokine storms and improve outcomes. Corticosteroids like *dexamethasone* can be used in severe COVID-19 to suppress *NFKB1* activity and reduce inflammation. It is, however, important to note that while dampening *NFKB1* can be beneficial, completely suppressing it could impair the body's ability to fight the virus, thus, finding the right balance remains crucial.

IL-10 on the other hand is a natural anti-inflammatory cytokine, acting as a brake on the immune response. It helps control excessive inflammation, as disease severity progresses, particularly in moderate disease states, to prevent tissue damage. However, overactive *IL-10* production in moderate COVID-19 cases can dampen the immune system's ability to fight the virus, potentially prolonging the infection and allowing persistent viral replication. From our results (**Figure 4.3B**), drugs with the potential to modulate *IL-10* activity could be beneficial during moderate disease states to balance the suppression of excessive inflammation with optimal immune functioning. For instance, *sirolimus* inhibits the mammalian target of the rapamycin (mTOR) pathway, which can indirectly suppress *IL-10* production by limiting *STAT3* signalling [368, 369]. Also, it can dampen the activation of certain immune cells like T cells, which may indirectly decrease *IL-10* production. Additionally, *sirolimus* promotes the differentiation and expansion of regulatory T cells (Tregs), a subset of T cells that naturally suppress inflammation and can promote *IL-10* production as part of their suppressive function [370].

In the multi-layered graph generated by the RWR analysis during predictions for severe disease state **(Figure 4.3C),** we observed key inflammation-related features like C-X-C Motif Chemokine ligand 1 (*CXCL1*), C-C Motif Chemokine ligands 4

(*CCL4*), and Janus Kinase 2 (*JAK2*) to establish subnetworks which included both direct and indirect interactions with the seed nodes and top-ranked drug candidates **(Figure 4.3C).** *JAK2*, a signalling molecule inside immune cells, expresses both inflammatory and anti-inflammatory effects during COVID-19. For its inflammatory role, *JAK2* activates certain signalling pathways including Janus kinase 2/signal transducer and activator of transcription 3 (*JAK2/STAT3*) pathway that trigger inflammatory responses in the lungs which may help fight viral infections [371]. In its anti-inflammatory role, *JAK2* also activates pathways promoting tissue repair and regeneration. Thus, drug repurposing candidates with the potential to influence *JAK2* signalling, may represent an effective therapeutic strategy for controlling the disease [372]. For instance, *IL-6* binds to soluble and transmembrane *IL-6R* and the resultant complex induces homodimerization of gp130, leading to activation of JAK2 [373]. This suggests that *sarilumab* and *tocilizumab* targeting the *IL-6* receptor, indirectly activate *JAK2* downstream by limiting the IL-6-mediated signalling pathway [373, 374]. *Sirolimus* targeting a mTOR pathway connected to *JAK2* can be appropriate. Increased levels of *CXCL1* and *CCL4* have been associated with severe disease and hyperinflammatory states, suggesting a potential role in COVID-19 disease progression [354, 375]. In general, the drug repurposing candidates (**Table 4.4**) with anti-inflammatory activities, immunomodulatory activities, and viral replication inhibitory activities have the best potential to manage excessive inflammation and limit viral persistence during the severe disease state. We further observed from shortest path analysis how these molecular features act as mediators connecting the drug repurposing candidates to the seed nodes across mild, moderate, and severe disease states (**Supplementary File 4.4**).

Table 4.4. *Top 20 potential drugs for mild, moderate, and severe COVID-19, ranked according to their measure of proximity to IL-6 and IL-6R seed nodes as determined through RWR analysis of COVID-KG, DRKG, and DSOG. The references point to publications that have reported the drugs' mechanism of action potentially linked with COVID-19.*

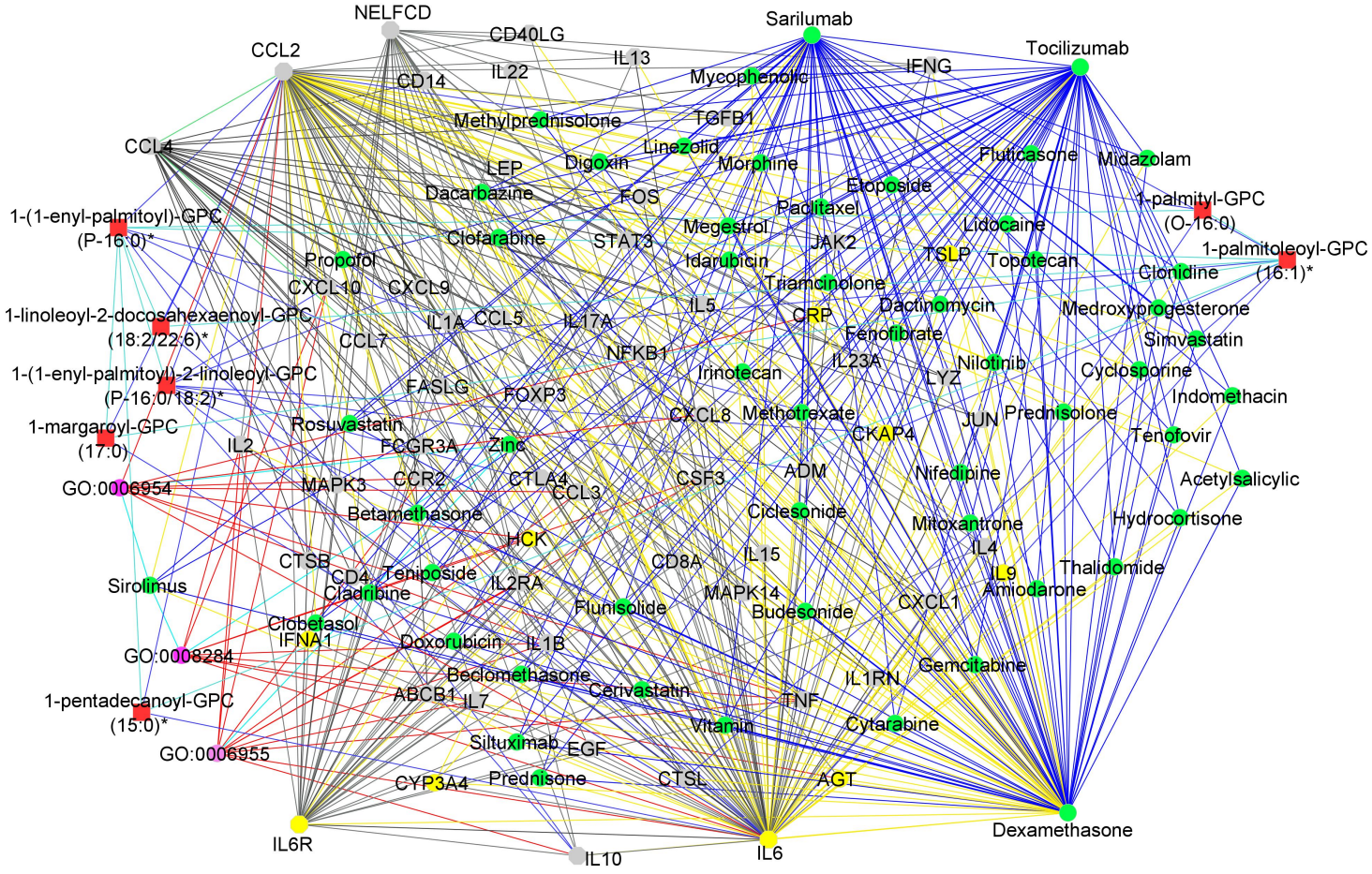
Drug	Drug category	Mechanism of action potentially linked with COVID-19	Measure of proximity in mild	Measure of proximity in moderate	Measure of proximity in severe	Reference*
Dexamethasone	Corticosteroid	Suppresses immune response	0.0033655	0.0033691	0.0033697	[345]
Sarilumab	Anti-interleukin 6 (IL-6) receptor monoclonal antibody	Suppresses immune response by blocking IL-6 signalling	0.0031924	0.0031908	0.0031911	[51]
Tocilizumab	Anti-interleukin 6 (IL-6) receptor monocl onal antibody	Suppresses immune response by blocking IL-6 signalling	0.0031908	0.0031894	0.0031897	[52]
Zinc	Essential mineral / Nutrient	Expresses antiviral properties and interferes with viral RNA synthesis to inhibit replication	0.0002730	0.0002681	0.0002692	[331, 376]
Sirolimus	Immunosuppress ive drug	Expresses immunomodulatory and anti-inflammatory properties and inhibits the expression of proinflammatory cytokines.	0.0001524	0.0001526	0.0001530	[332, 333]
Histamine	Depressor amine	Expresses immunomodulatory and anti-inflammatory properties	0.0001420	0.0001413	0.0001416	[377, 378]
Curcumin	Natural compound	Expresses immune modulatory and anti-inflammatory properties that inhibit severe inflammation and cytokine storm.	0.0001366	0.0001358	0.0001363	[379, 380]
Cyclosporine	Immunosuppress ive drug	Express anti-inflammatory and anti-viral properties	0.0001352	0.0001370	0.0001369	[346, 347]
Doxorubicin	Anticancer	inhibit the protease-mediated viral entry to the host cell	0.0001314	0.0001316	0.0001318	[328]
Morphine	Opioid pain medication	Contribute to improving respiratory failure	0.0001301	0.0001281	0.0001284	[381]
Dactinomycin	Anticancer	Expresses immune modulatory properties and inhibits viral cellular transcription	0.0001292	0.0001290	0.0001288	[336, 337]
Simvastatin	lipid-lowering drug	Expresses anti-inflammatory, immunomodulatory properties and reduce viral replication	0.0001290	0.0001289	0.0001287	[382, 383]
Hydrocortisone	Corticosteroid	Expresses anti-inflammatory and immunomodulatory properties	0.0001274	0.0001279	0.0001276	[345]

Vitamin C	Essential mineral / Nutrient	Express antioxidant properties and improves immune function	0.0001268	0.0001269	0.0001276	[384]
Mycophenolic acid	Immunosuppressive drug	Expresses immunomodulatory properties	0.0001236	0.0001236	0.0001239	[351]
Methotrexate	Immunosuppressive drug	Expresses immunomodulatory properties	0.0001182	0.0001195	0.0001194	
Dopamine	Catecholamine neurotransmitter	Influences the expression of *ACE2*	0.0001178	0.0001187	0.0001197	[385]
Indomethacin	Non-steroidal anti-inflammatory drug	Expresses anti-inflammatory properties and reduces pain and fever	0.0001170	0.0001165	0.0001162	[330]
Paclitaxel	Anticancer	Express anti-inflammatory and anti-viral properties	0.0001159	0.0001159	0.0001161	
Metformin	Biguanide antihyperglycemic	Inhibits viral replication	0.0001158	0.0001158	0.0001164	[386]

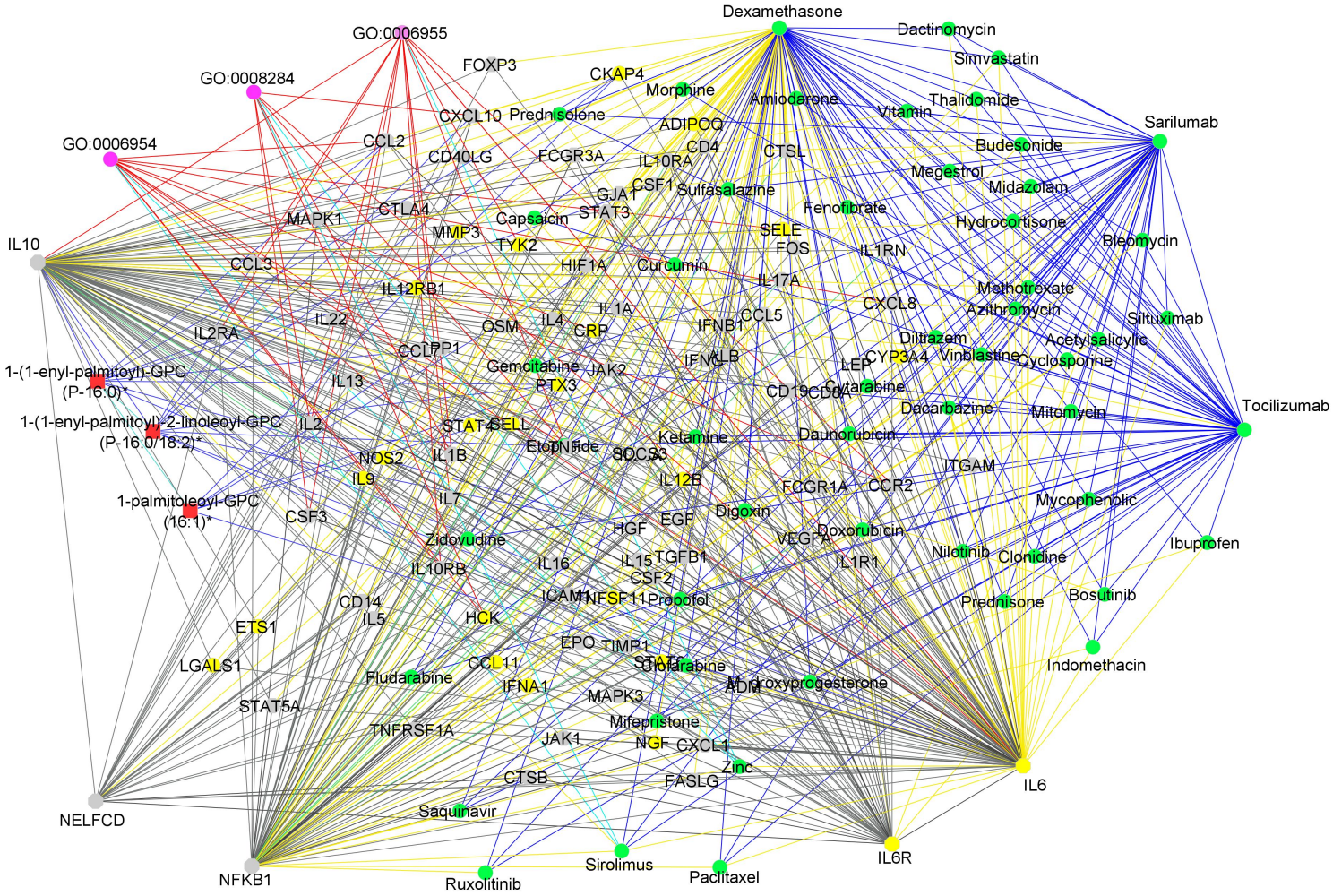
Table 4.5. *Node degree, betweenness, and closeness centrality measures for the drug repurposing candidates predicted using the hypothesis-driven approach.*

Drug name	Degree mild	Degree moderate	Degree severe	Betweenness mild	Betweenness moderate	Betweenness severe	Closeness mild	Closeness moderate	Closeness severe
Dexamethasone	1633	1632	1633	0.6282	0.5881	0.6293	0.6355	0.6233	0.6228
Tocilizumab	830	829	830	0.11436	0.1008	0.1103	0.4643	0.4482	0.4571
Sarilumab	469	469	469	0.0286	0.0246	0.0283	0.4432	0.4289	0.4373
Sirolimus	7	8	6	0.0063	0.0068	0.0054	0.4831	0.4889	0.4512
Vitamin C	6	6	4	0.0047	0.0034	0.0004	0.4860	0.4525	0.4460
Cyclosporine	5	6	4	0.0007	0.0014	0.0004	0.4769	0.4832	0.4460
Dactinomycin	5	4	4	0.0007	0.0003	0.0004	0.4769	0.4367	0.4460
Paclitaxel	5	5	4	0.0007	0.0009	0.0004	0.4769	0.4739	0.4460
Simvastatin	5	5	4	0.0007	0.0009	0.0004	0.4769	0.4739	0.4460
Hydrocortisone	4	5	4	0.0004	0.0007	0.0004.	0.4528	0.4454	0.4460
Zinc	4	5	4	0.0005	0.0011	0.0005.	0.3652	0.3903	0.3637
Indomethacin	4	5	4	0.0004	0.0009	0.0004	0.4528	0.4739	0.4460
Mycophenolic acid	4	5	4	0.0004	0.0009	0.0004	0.4528	0.4739	0.4460
Doxorubicin	4	5	4	0.0004	0.0009	0.0004	0.4528	0.4739	0.4460
Methotrexate	4	5	4	0.0004	0.0007	0.0004	0.4528	0.4454	0.4460
Morphine	4	4	3	0.0005	0.0007	0.0003	0.4768	0.4738	0.4459
Curcumin	3	4	3	0.0	0.0001	0.0004	0.4644	0.4716	0.4483
Metformin	3	3	2	0.0	0.0	0.0	0.4644	0.4355	0.4354
Histamine	3	3	1	0.0002	0.0	0.0	0.3866	0.3705	0.3601
Dopamine	1	1	1	0.0	0.0	0.0	0.3615	0.3584	0.3601

A


NELFCD
Sarilumab
CD40LG
CCL2
IL13
Tocilizumab
IL22
CD14
Mycophenolic
IFNG
Methylprednisolone
TGFB1
CCL4
Linezolid
Fluticasone
Midazolam
LEP
Digoxin
Morphine
Etoposide
Dacarbazine
FOS
Paclitaxel
1-palmityl-GPC
(O-16:0)
1-(1-enyl-palmitoyl)-GPC
(P-16:0)*
Megestrol
Lidocaine
Clofarabine
STAT3
JAK2
1-palmitoleoyl-GPC
(16:1)*
Propofol
Idarubicin
TSLP
Topotecan
Clonidine
CXCL10
CXCL9
Triamcinolone
1-linoleoyl-2-docosahexaenoyl-GPC
(18:2/22:6)*
IL1A
CCL5
IL17A
IL5
CRP
Dactinomycin
Medroxyprogesterone
CCL7
Fenofibrate
Simvastatin
NFKB1
IL23A
Nilotinib
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC
(P-16:0/18:2)*
Irinotecan
LYZ
Cyclosporine
FASLG
FOXP3
Indomethacin
CXCL8
Methotrexate
JUN
Prednisolone
1-margaroyl-GPC
(17:0)
Rosuvastatin
CKAP4
Tenofovir
IL2
FCGR3A
Zinc
ADM
Nifedipine
GO:0006954
MAPK3
CCR2
CTLA4
CCL3
CSF3
Acetylsalicylic
Betamethasone
Ciclesonide
Mitoxantrone
Hydrocortisone
HCK
IL4
CTSB
CD8A
IL15
IL9
Thalidomide
CD4
Teniposide
Sirolimus
Cladribine
IL2RA
MAPK14
Amiodarone
Flunisolide
Budesonide
CXCL1
Clobetasol
IFNA1
Doxorubicin
IL1B
GO:0006284
Gemcitabine
Beclomethasone
IL1RN
1-pentadecanoyl-GPC
(15:0)*
ABCB1
IL7
Cerivastatin
TNF
Vitamin
Cytarabine
Siltuximab
EGF
GO:0006955
CYP3A4
Prednisone
CTSL
AGT
IL6R
IL10
IL6
Dexamethasone

B


Dexamethasone
GO:0006955
GO:0008284
GO:0006954
FOXP3
CKAP4
Morphine
Amiodarone
Dactinomycin
Simvastatin
Thalidomide
Vitamin
Sarilumab
CXCL10
Prednisolone
ADIPOQ
CCL2
CD40LG
FCGR3A
CD4
CTSL
IL10RA
Budesonide
Megestrol
Midazolam
GJA1
CSF1
Sulfasalazine
CTLA4
MAPK1
STAT3
Capsaicin
Fenofibrate
Hydrocortisone
MMP3
TYK2
SELE
FOS
IL1RN
Bleomycin
IL10
CCL3
HIF1A
Curcumin
IL17A
IL12RB1
Methotrexate
IL1A
CCL5
CXCL8
Azithromycin
Siltuximab
IL22
OSM
IL4
CRP
IFNB1
IL2RA
Diltiazem
Acetylsalicylic
CCL7
PP1
IFNL
LB
CYP3A4
Vinblastine
Cyclosporine
1-(1-enyl-palmitoyl)-GPC
(P-16:0)*
Gemcitabine
JAK2
LEP
IL13
PTX3
CD19
CD58
Cytarabine
Dacarbazine
Mitomycin
Tocilizumab
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC
(P-16:0/18:2)*
IL2
STAT4
GELL
Ketamine
Daunorubicin
EtoPNIide
SOCS3
NOS2
IL1B
ITGAM
IL9
IL12B
FCGR1A
CCR2
1-palmitoleoyl-GPC
(16:1)*
IL7
Digoxin
Mycophenolic
CSF3
EGF
HGF
Doxorubicin
VEGFA
Zidovudine
Nilotinib
Clonidine
Ibuprofen
IL10RB
IL16
IL15
TGFB1
IL1R1
CSF2
CD14
ICAM1
TNFSF11
Propofol
Prednisone
Bosutinib
IL5
HCK
ETS1
EPO
TIMP1
LGALS1
CCL11
Clofarabine
Indomethacin
Fludarabine
IFNA1
ADM
Hydroxyprogesterone
MAPK3
STAT5A
Mifepristone
TNFRSF1A
JAK1
NGF
CXCL1
Zinc
CTSB
FASLG
IL6
NELFCD
Saquinavir
IL6R
Sirolimus
NFKB1
Ruxolitinib
Paclitaxel

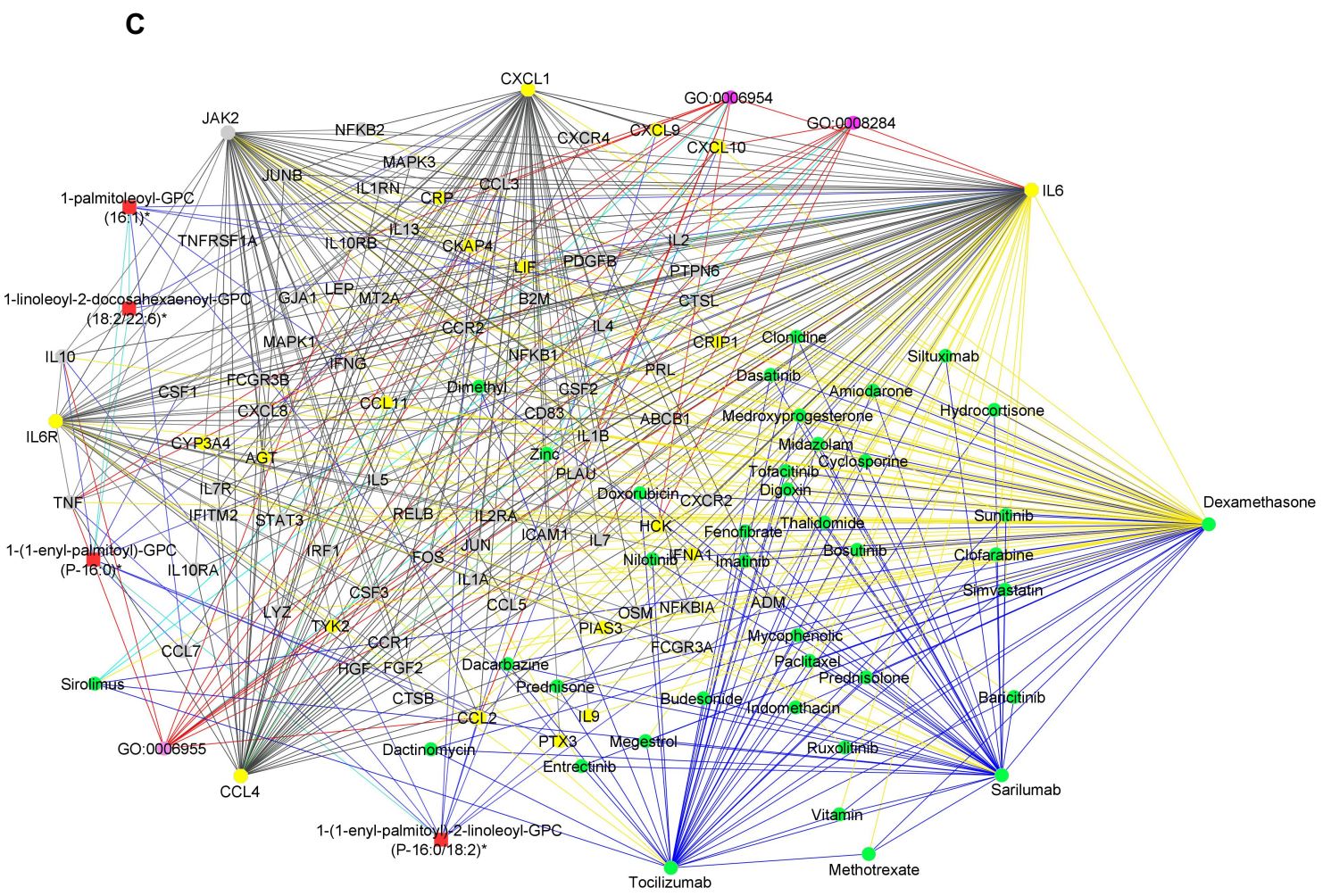


Figure 4.3. *(A) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), and biological process (pink nodes) as observed from predicting candidate drugs using existing knowledge graphs, mild disease-state specific omics-graphs, and hypothesis-driven seeds. The graph reveals distinct subnetworks formed by hubs CCL2, CCL4, and NELFCD demonstrating extensive interactions with drug candidates, seed nodes (IL-6 and IL-6R), and other molecular features (B) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), and biological process (pink nodes) as observed from predicting candidate drugs using existing knowledge graphs, moderate disease-state specific omics-graphs, and hypothesis-driven seeds. The graph reveals distinct subnetworks formed by hubs NFKB1, IL-10, and NELFCD demonstrating extensive interactions with drug candidates, seed nodes (IL-6 and IL-6R), and other molecular features (C) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), and biological process (pink nodes) as observed from predicting candidate drugs using existing knowledge graphs, severe disease-state specific omics-graphs, and hypothesis-driven seeds. The graph reveals distinct subnetworks formed by hubs CXCL1, CCL4, and JAK2 demonstrating extensive interactions with drug candidates, seed nodes (IL-6 and IL-6R), and other molecular features. Yellow edges represent drug-protein and drug-transcript pairwise interactions. Red edges represent biological process-protein interactions and biological process-transcripts interactions. Green edges represent protein-protein interactions. Black edges represent transcript-transcript interactions and protein-transcript interactions. Blue edges represent drug-drug interactions. Light blue edges represent biological processes-biological process interactions and biological process-pathway interactions. The graphs were generated by defining filtering criteria based on node degree between 4 and 1633 inclusive from cytoscape.*

4.3.3 Predicting candidate drugs using existing knowledge graphs, disease-state specific omics-graphs, and data-driven seeds

To investigate the prediction of drugs that might be differentially applicable to treating different COVID-19 disease states, we used a data-driven approach to identify seeds by computing an integrated node centrality metric score leveraging the node degree, closeness, betweenness, and eigenvector centrality metrics (**Supplementary Table 4.3**).

In the transcriptomics layer, we identified Signal Transducer And Activator Of Transcription 1 (*STAT1*) as a seed node. *STAT1* is known to be involved in immune responses and antiviral activity [354] and is reported to be upregulated in mild and severe COVID-19 cases, with the phosphorylation of the gene associated with both the upregulation of *ACE2* expression and the development of severe disease states [222, 387].

In the proteomic layer, Superoxide Dismutase 2 (*SOD2*) was identified as a seed node. *SOD2* is an essential antioxidant enzyme that protects cells from superoxide radical anions which are known to be significantly under-expressed in the plasma [223] and lung cells of severe COVID-19 patients [224].

In the metabolomics layer, 3-hydroxyoctanoate was identified as a seed node. This metabolite is generated during medium-chain fatty acid oxidation and serves as a marker for primary defects in beta-hydroxy fatty acid metabolism. It is also affiliated with essential pathways such as those responsible for macrophage activation and platelet aggregation, with increases in 3-hydroxyoctanoate concentrations being associated with asymptomatic COVID-19 infections [388].

In the lipidomics layer, we identified "*unknown_mz_815.61548_+_RT_27.063"*, an uncharacterized lipid associated with disease severity as a seed node.

Using these seed nodes for RWR analysis across the various disease states, we identified several potential drug repurposing candidates (**Table 4.6**) including "natural compounds" (such as *glutathione* and *curcumin*) and inhibitors of signal transduction

pathways of protein kinases and cell proliferation (tyrosine, histone deacetylase, and methyltransferase). The results revealed the same drug repurposing candidates across the various disease states (**Table 4.6**). However, network topology analysis revealed differences between the multi-layered graphs generated by the RWR analysis (**Figure 4.4A-C**). Specifically, these differences were observed in the connectivity of the drug repurposing candidates with corresponding node degree, betweenness centrality, and closeness centrality scores (**Table 4.7**). This provides more insights into the most appropriate drug repurposing candidates for the different disease states. For instance, *curcumin* had a higher degree of 1076 in the multi-layered graph generated by the RWR analysis on mild disease state as compared to a degree of 4 in the multi-layered graph generated by the RWR analysis on moderate disease state and a degree of 2 in the multi-layered graphs generated by the RWR analysis on severe disease state. *Podofilox* had a higher degree of 478 in the multi-layered graph generated by the RWR analysis on severe disease state as compared to a degree of 2 for both multi-layered graphs generated by the RWR analysis on mild and moderate disease states. Therefore, whereas *curcumin* is an appropriate drug repurposing candidate with potential utility during mild COVID-19, *podofilox* is an appropriate drug repurposing candidate with potential utility during severe COVID-19. Furthermore, *vinblastine's* high degree of 1,905 in the multi-layered graph generated by the RWR analysis on moderate disease state and a degree of 1,378 in the multi-layered graph generated by the RWR analysis on severe disease state as compared to a degree of 4 in the multi-layered graph from the RWR analysis on mild disease state suggests it might be most effective in tackling advanced disease stages. Also, *crizotinib's* high degree of 1,919 in the multi-layered graph generated by the RWR analysis on mild disease state and a degree of 1,947 in the multi-layered graph generated by the RWR analysis on moderate state as compared to a degree of 2 in the multi-layered graph from the RWR analysis on severe disease state indicates it might be most appropriate for treatment of the mild and moderate stages of COVID-19 (**Table 4.7**). Similarly, *glutathione's* high degree of 212 in the multi-layered graph generated by the RWR analysis on mild disease state and a degree of 234 in the multi-layered graph generated by the RWR analysis on moderate disease state as compared to a degree of 118 in the multi-layered graph from the RWR analysis on severe disease state suggests it can be a promising drug repurposing candidate for the mild and moderate disease states. Noticeably, nodes with higher degree scores

(**Table 4.7**) have higher betweenness and closeness scores revealing how often these nodes lie on the shortest paths between other nodes in the network and mediate how quickly information can reach other disease-related features from a candidate drug (**Supplementary File 4.5**). Such nodes have high relevance within biological systems and, besides their specific biological activities, might also facilitate communication and synergy between biological pathways, making them key targets for the management of the disease.

Analysis of the multi-layered graphs generated by the RWR analysis revealed several features that establish subnetworks (**Figure 4.4**). Specifically, we observed the *CCL4* to establish subnetwork in the mild disease state and Hepatocyte Growth Factor (*HGF*) to establish subnetworks in both moderate and severe disease states. *HGF* expresses anti-inflammatory properties and plays a complex and multifaceted role in the battle against COVID-19 [232]. While it initially acts as a crucial player in lung tissue repair following viral damage, its activity can also contribute to excessive inflammation if not properly regulated [232]. *HGF* can activate certain signalling pathways that promote inflammation in moderate to severe cases. On the other hand, up-regulation of *HGF* represents a robust counter-regulatory mechanism employed by the host immune response to counteract pro-inflammatory cytokines.

Similar to the results from the hypothesis-driven approach, some of the drugs predicted to be suitable for repurposing are being, or have already been, tested to assess their efficacy and effectiveness in the treatment of COVID-19. For example, the antihelminthic drug *mebendazole*, enhanced innate immune responses and restored inflammation to normal levels in symptomatic non-hospitalized COVID-19 patients during a recent clinical trial [389]. Also, *etoposide* has been investigated for its potential to treat severe disease, albeit with observed adverse events that warrant further investigation [390].

Table 4.6. *Top 20 potential drugs for mild, moderate, and severe COVID-19, ranked according to their measure of proximity to STAT1, SOD2, 3-hydroxyoctanoate, and unknown_mz_815.61548_+_RT_27.063 seed nodes as determined through RWR analysis of COVID-KG, DRKG, and DSOG. The references point to publications that have reported the drugs' mechanism of action potentially linked with COVID-19.*

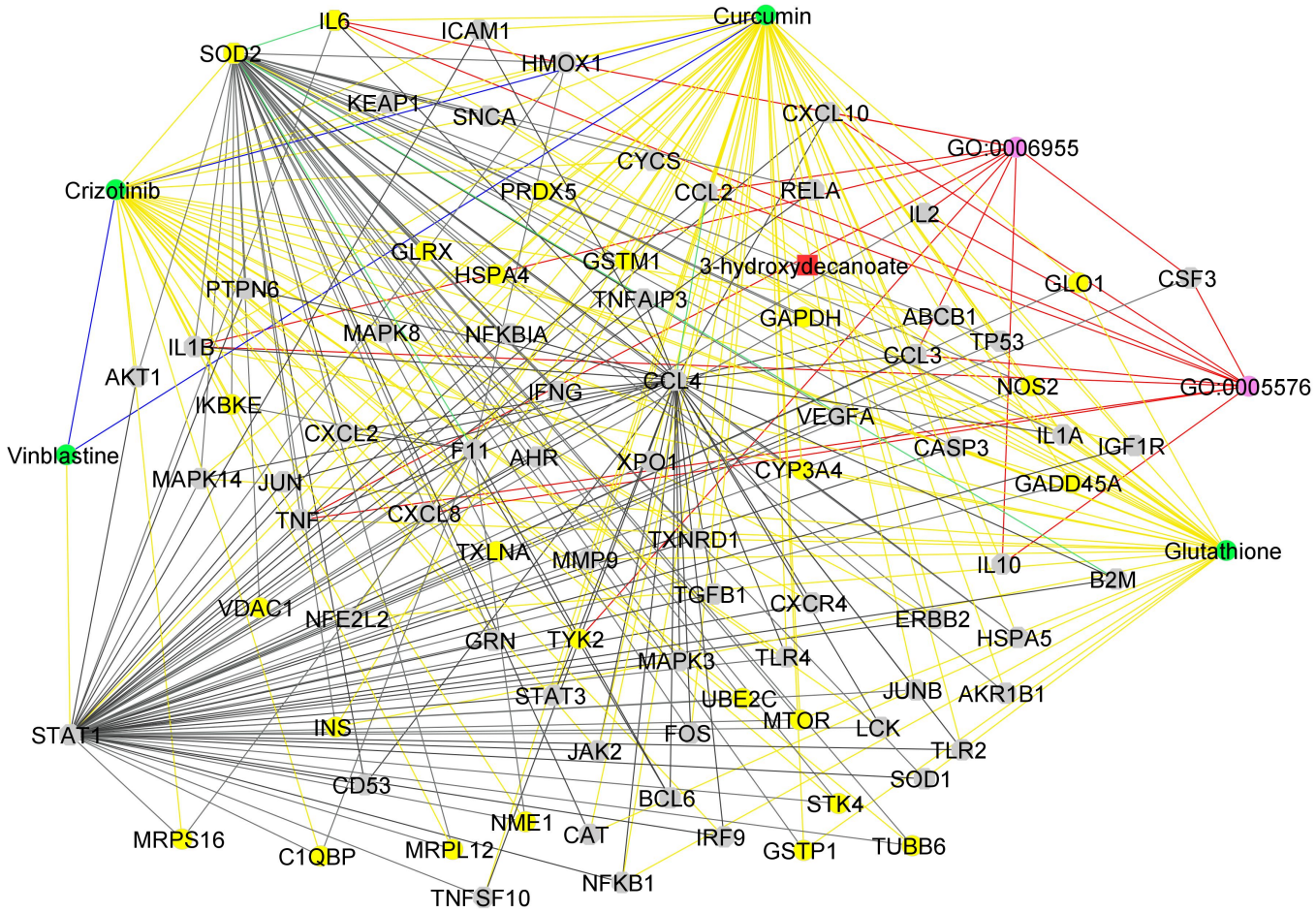
Drug	Drug category	Mechanism of action potentially linked with COVID-19	Measure of proximity in mild	Measure of proximity in moderate	Measure of proximity in severe	Reference*
Glutathione	Antioxidant	Protect cells from damage caused by oxidative stress	0.0007061	0.0007044	0.0007044	[384, 391]
Crizotinib	Tyrosine kinase inhibitor	Express immune modulatory properties and could inhibit receptor tyrosine kinases and affect cellular processes relevant to viral replication	0.0007022	0.0007005	0.0007002	[392]
Curcumin	Natural compound	Expresses antiviral properties, immune modulatory and antioxidant properties that may contribute to inhibit inflammation, oxidative stress, and reduce lung injury	0.0006966	0.0006948	0.0006951	[379]
Vorinostat	Histone deacetylase inhibitor	Express anti-inflammatory and antiviral properties	0.0006943	0.0006952	0.0006954	[393]
Vinblastine	Anticancer	Disrupt microtubule dynamics, leading to mitotic spindle dysfunction and cell cycle arrest.	0.0006920	0.0007029	0.0007022	[394]
Iron	Essential mineral / Nutrient	Expresses anti-inflammatory and immunomodulatory properties	0.0006917	0.0006899	0.0006944	[395]
Mebendazole	Antihelminthic	Express antiviral and immune modulatory properties and may interfere viral replication	0.0006913	0.0006913	0.0006910	[389]
Podofilox	Topical agent	Stimulate the production of interferon-γ, a cytokine that plays a role in the immune response.	0.0006900	0.0007003	0.0007002	[396]
Valine	Essential amino acid	Protein synthesis	0.0006893	0.0006891	0.0006891	[397, 398]
Acetylcysteine	Antioxidant and glutathione inducer	Express antioxidant, immunomodulatory, and mucolytic properties	0.0006878	0.0006877	0.0006878	[399]
Thimerosal	Methyltransferase inhibitor	Induces Th2-type cytokines via influencing cytokine secretion by human dendritic cells	0.0006826	0.0006824	0.0006824	
Fludarabine	Chemotherapy drug	Inhibits type I interferon-induced expression of *ACE2*	0.0006813	0.0006811	0.0006812	[387]
Calcipotriol	Anti-psoriatic	Enhance cell differentiation	0.0006800	0.0006800	0.0006800	

Teniposide	Cytotoxic drug	Inhibits SARS-CoV-2 3-chymotrypsin-like cysteine protease	0.0001087	0.0001091	0.0001096	[400]
Omacetaxine	Cephalotaxine ester and protein synthesis inhibitor	Inhibits protein translation and interfere with viral replication	0.0001015	0.0000955	0.0001009	[401]
Etoposide	topoisomerase II inhibitor	Express anti-inflammatory properties and suppresses cytokine production	0.0001010	0.0000973	0.0001022	[402, 403]
L-Glutamine	Amino acid	Expresses immune modulatory, anti-inflammatory and antioxidant properties	0.0000945	0.0000943	0.0000942	[404, 405]
Carglumic acid	Analog of N-acetylglutamate (NAG)	Express enzyme properties for processing excess nitrogen produced when the body metablizes proteins	0.0000942	0.0000939	0.0000939	
Pregabalin	Anticonvulsant drug	Reduces COVID-19-related pain and cough	0.0000591	0.0000591	0.0000591	[406]
Threonine	Essential amino acid	Expresses immune modulatory, anti-inflammatory and antioxidant properties	0.0000579	0.0000578	0.0000579	[398]

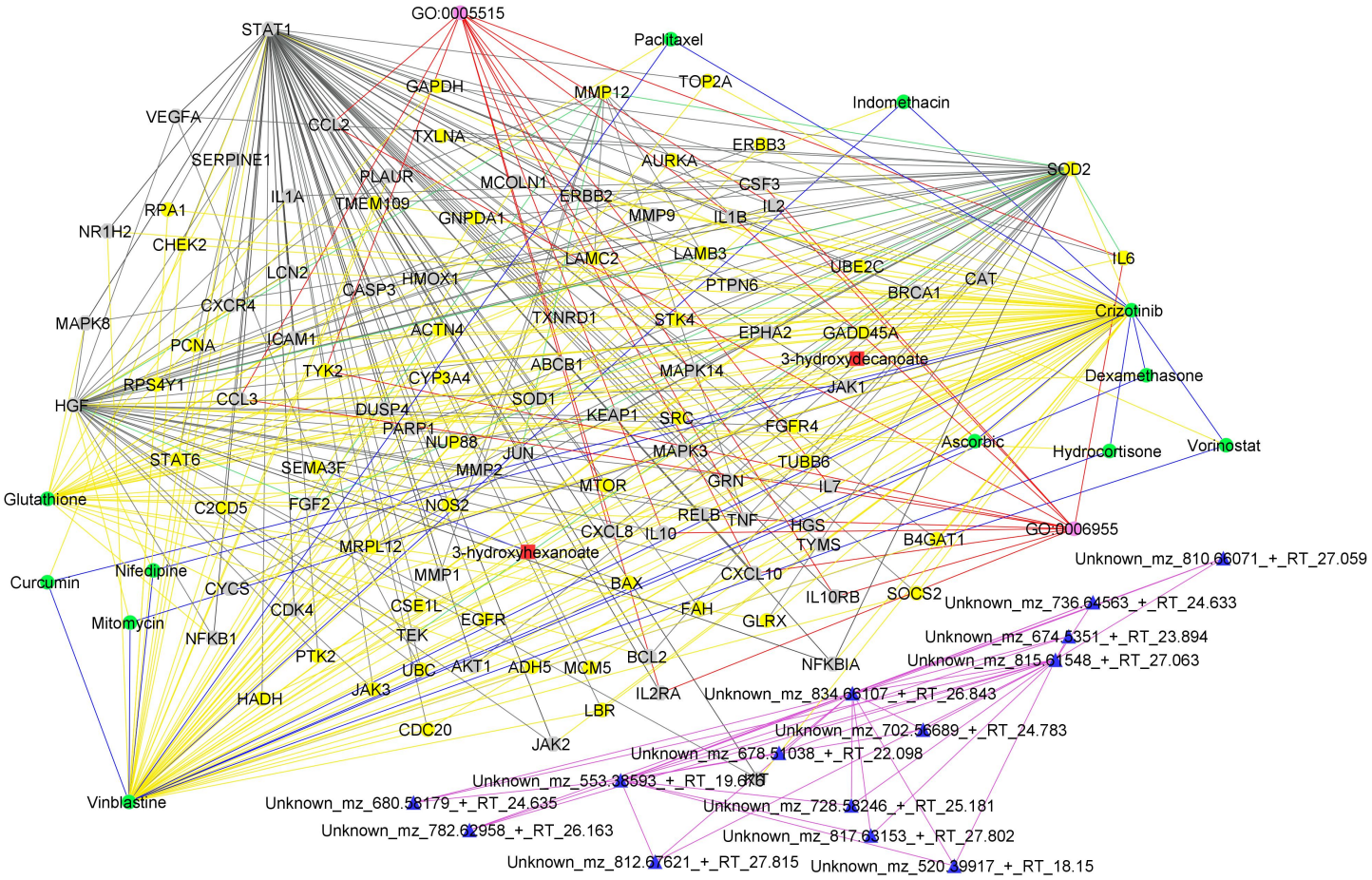
***Table 4.7**. Node degree, betweenness, and closeness centrality measures for the drug repurposing candidates predicted using the data-driven approach.*

Drug name	Degree mild	Degree moderate	Degree severe	Betweenness mild	Betweenness moderate	Betweenness severe	Closeness mild	Closeness moderate	Closeness severe
Crizotinib	1919	1947	2	0.5147	0.4171	0.0025	0.5478	0.5297	0.3987
Curcumin	1076	4	2	0.3281	0.0005	0.0	0.5389	0.4317	0.3983
Glutathione	212	234	118	0.0397	0.0350	0.0439	0.3532	0.3611	0.3657
Vinblastine	4	1905	1378	0.0009	0.4787	0.6440	0.4364	0.5639	0.5556
Acetylcysteine	2	2	1	0.0	0.0	0.0	0.3106	0.3028	0.2743
Calcipotriol	2	2	1	0.0	0.0	0.0	0.3178	0.3186	0.3370
Etoposide	2	2	2	0.0	0.0	0.0001	0.3987	0.3993	0.3715
Fludarabine	2	3	2	0.0	0.0	0.0	0.3178	0.3902	0.3983
Iron	2	2	1	0.0	0.0	0.0	0.3106	0.3028	0.2743
Mebendazole	2	2	1	0.0	0.0	0.0	0.3178	0.3186	0.3370
Omacetaxine	2	2	1	0.0	0.0	0.0	0.2438	0.2466	0.2460
Podofilox	2	2	478	0.0	0.0	0.1438	0.3178	0.31863	0.4092
Teniposide	2	2	2	0.0	0.0	0.0001	0.3987	0.3993	0.3715
Thimerosal	2	2	1	0.0	0.0	0.0	0.3106	0.3028	0.2743
Valine	2	2	1	0.0	0.0	0.0	0.3106	0.3028	0.2743
Vorinostat	2	4	2	0.0009	0.0005	0.0	0.4199	0.4317	0.3983
Carglumic acid	1	1	1	0.0	0.0	0.0	0.2610	0.2653	0.2678
L-Glutamine	1	1	1	0.0	0.0	0.0	0.2610	0.2653	0.2678
Threonine	1	2	1	0.0	0.0	0.0	0.3106	0.3028	0.2743
Pregabalin	1	1	1	0.0	0.0	0.0	0.3539	0.3463	0.3463

A


Curcumin
IL6
ICAM1
SOD2
HMOX1
KEAP1
SNCA
CXCL10
GO:0006955
CYCS
Crizotinib
PRDX5
CCL2
RELA
IL2
GLRX
GSTM1
3-hydroxydecanoate
HSPA4
GLO1
CSF3
PTPN6
TNFAIP3
GAPDH
ABCB1
MAPK8
NFKBIA
TP53
IL1B
CCL3
AKT1
CCL4
NOS2
GO:0005576
IFNG
IKBKE
VEGFA
CXCL2
IL1A
IGF1R
Vinblastine
F11
AHR
XPO1
CASP3
MAPK14
JUN
CYP3A4
GADD45A
TNF
CXCL8
TXNRD1
Glutathione
TXLNA
MMP9
IL10
B2M
TGFB1
CXCR4
VDAC1
NFE2L2
ERBB2
GRN
TYK2
HSPA5
MAPK3
TLR4
STAT3
UBE2C
JUNB
AKR1B1
INS
MTOR
LCK
STAT1
FOS
JAK2
TLR2
CD53
SOD1
BCL6
STK4
NME1
CAT
MRPS16
IRF9
GSTP1
TUBB6
C1QBP
MRPL12
NFKB1
TNFSF10

B


GO:0005515
STAT1
Paclitaxel
GAPDH
TOP2A
MMP12
Indomethacin
VEGFA
CCL2
TXLNA
ERBB3
SERPINE1
AURKA
SOD2
PLAUR
MCOLN1
CSF3
IL1A
TMEM109
ERBB2
IL2
RPA1
GNPDA1
MMP9
IL1B
NR1H2
CHEK2
LAMC2
LAMB3
IL6
LCN2
UBE2C
CAT
HMOX1
CASP3
PTPN6
BRCA1
CXCR4
Crizotinib
MAPK8
TXNRD1
STK4
EPHA2
GADD45A
ACTN4
PCNA
ICAM1
ABCB1
3-hydroxydecanoate
TYK2
CYP3A4
MAPK14
Dexamethasone
RPS4Y1
JAK1
HGF
CCL3
SOD1
DUSP4
KEAP1
SRC
FGFR4
PARP1
NUP88
Ascorbic
Hydrocortisone
Vorinostat
STAT6
JUN
MAPK3
SEMA3F
MMP2
TUBB6
MTOR
GRN
IL7
Glutathione
C2CD5
FGF2
NOS2
RELB
TNF
HGS
CXCL8
IL10
GO:0006955
B4GAT1
MRPL12
TYMS
3-hydroxyhexanoate
Unknown_mz_810.66071_+_RT_27.059
Nifedipine
Curcumin
MMP1
CXCL10
CYCS
BAX
IL10RB
SOCS2
Unknown_mz_736.64563_+_RT_24.633
CSE1L
CDK4
FAH
Mitomycin
EGFR
GLRX
Unknown_mz_674.5351_+_RT_23.894
NFKB1
TEK
BCL2
Unknown_mz_815.61548_+_RT_27.063
PTK2
UBC
AKT1
ADH5
MCM5
NFKBIA
JAK3
IL2RA
Unknown_mz_834.66107_+_RT_26.843
HADH
LBR
CDC20
Unknown_mz_702.56689_+_RT_24.783
JAK2
Unknown_mz_678.51038_+_RT_22.098
Unknown_mz_553.38593_+_RT_19.675
Vinblastine
Unknown_mz_680.58179_+_RT_24.635
Unknown_mz_728.56246_+_RT_25.181
Unknown_mz_782.62958_+_RT_26.163
Unknown_mz_817.63153_+_RT_27.802
Unknown_mz_812.67621_+_RT_27.815
Unknown_mz_520.39917_+_RT_18.15

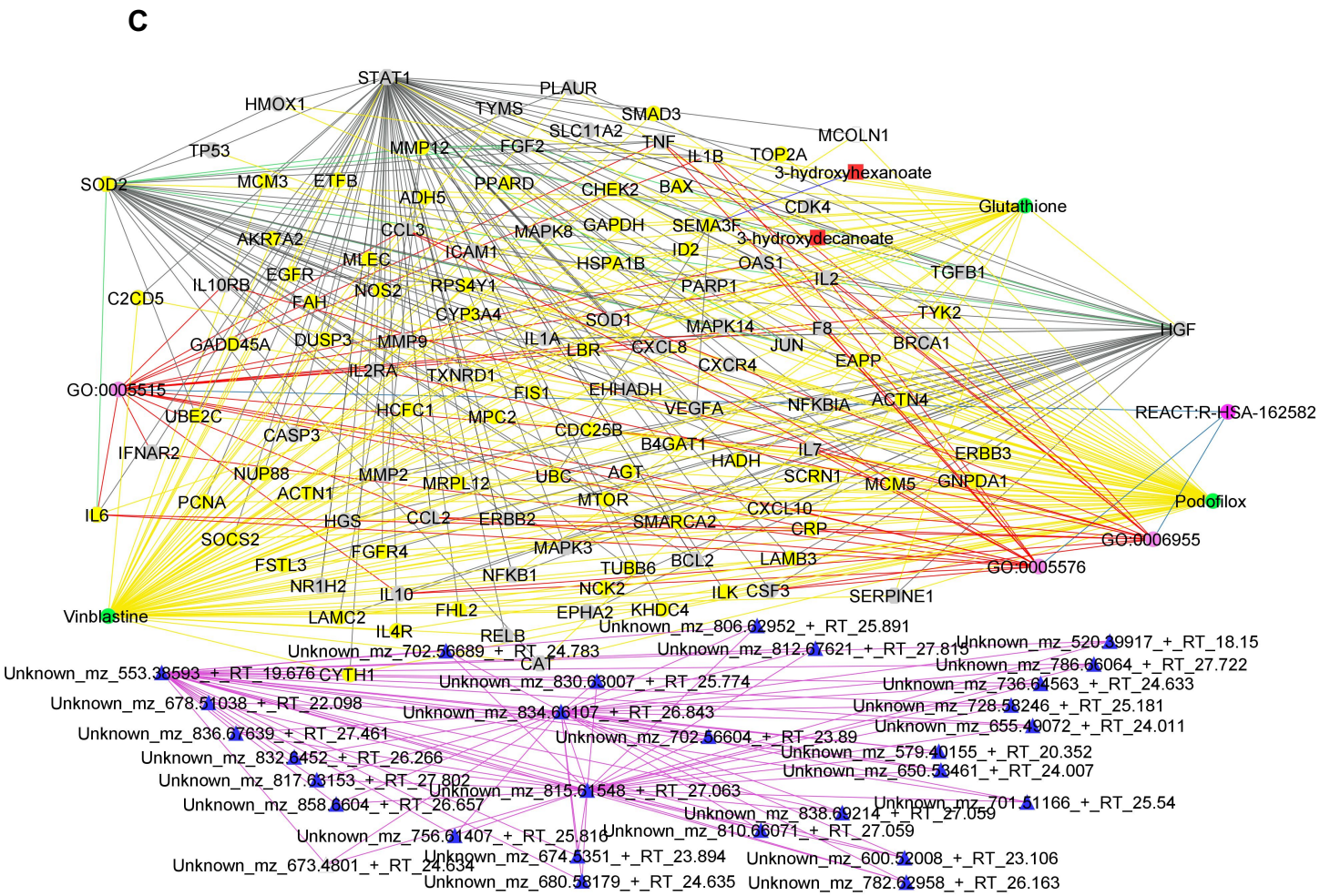


Figure 4.4*. (A) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), lipids (blue nodes), and biological process and pathways (pink nodes) as observed from predicting candidate drugs using existing knowledge graphs, mild disease-state specific omics-graphs, and data-driven seeds. The graph reveals distinct subnetworks formed by hub CCL4, demonstrating extensive interactions with drug candidates and other molecular features including seed nodes (STAT1 and SOD2) (B) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), lipids (blue nodes), and biological processes and pathways (pink nodes) as observed from predicting candidate drugs using existing knowledge graphs, moderate disease-state specific omics-graphs, and data-driven seeds. The graph reveals distinct subnetworks formed by hub HGF, demonstrating extensive interactions with drug candidates and other molecular features including seed nodes (STAT1 and SOD2), as well as a subnetwork formed among lipids. (C) Graph representation of the interaction between drugs (green nodes), proteins (yellow nodes), transcripts (grey nodes), metabolites (red nodes), lipids (blue nodes), and biological processes and pathways (pink nodes) as observed from predicting candidate drugs using existing knowledge graphs, severe disease-state specific omics-graphs, and data-driven seeds. The graph reveals distinct subnetworks formed by hub HGF, demonstrating extensive interactions with drug candidates and other molecular features including seed nodes (STAT1 and SOD2), as well as a subnetwork formed among lipids. Yellow edges represent drug-protein and drug-transcript pairwise interactions. Red edges represent biological process-protein interactions and biological process-transcripts interactions. Pink edges represent pairwise interactions between lipids. Green edges represent protein-protein interactions. Black edges represent transcript-transcript interactions and protein-transcript interactions. Dark blue edges represent drug-drug interactions. Light blue edges represent biological processes-biological process interactions and biological process-pathway interactions. The graphs were generated by defining filtering criteria based on node degree between 4 and 1633 inclusive from cytoscape.*

4.3.4 Drug prediction robustness analysis

The three highest-ranked candidate drugs yielded by both the hypothesis-driven approach (*dexamethasone, sarilumab, tocilizumab*) and the data-driven approach (*glutathione, crizotinib*, and *curcumin*) are all known to be efficacious in controlling moderate to severe COVID-19. However, the high rankings of these drugs (based on measures of proximity) are expected simply because their efficacy during moderate to severe COVID-19 treatment has been comprehensively reported on in the literature as of 2021: reports that are reflected in the COVID-19 KG layers of our networks.
In this section, we removed direct interactions between the top three predicted potential drug candidates identified using the hypothesis-driven and data-driven approaches, and other features (such as drugs, proteins, and transcripts) to assess the influence of these interactions and features on the drug predictions. We assessed the robustness of the drug predictions in both the hypothesis-driven and data-driven approaches by repeating the RWR analysis (as described in the materials and methods) after individually and collectively removing direct interactions associated with the three highest-ranked candidate drugs and examining changes in the drug rankings.

These analyses yielded a relatively consistent trend (**Supplementary Files 4.1 and 4.2**): drugs such as *sirolimus*, *histamine*, *cyclosporine,* and *vorinostat* that initially ranked below the drugs that were removed tended to achieve higher rankings following the removal of the initial top-ranked drug candidates. The measure of proximity of the drugs that attained elevated rankings varied from one drug removal experiment to the next but generally increased relative to the measure of proximity that the drugs obtained in the absence of any exclusions (**Supplementary Files 4.1 and 4.2**).
Further, the drug removal analyses revealed additional candidate drugs that were not apparent in the absence of drug removal. For instance, with the hypothesis-driven approach, when we removed *dexamethasone* and *tocilizumab* simultaneously for each disease state, drugs like *dinoprostone*, and *perhexiline* emerged among the top drug candidates across the mild, moderate, and severe disease states (**Supplementary File 4.1**). Similarly, when we removed sarilumab, we observed *ketamine, acetylsalicylic acid* (aspirin), *menadione*, and *iron* in the top 20. Concurring

with the individual drug removal, when *dexamethasone, tocilizumab*, and *sarilumab* were collectively removed, we observed *dinoprostone*, *perhexiline*, *menadione*, iron, and *ketamine* all entered the top 20 for the disease states.

When we excluded the top-ranked drug candidates that were revealed by the data-driven approach (*glutathione, crizotinib,* and *curcumin*) we observed similar ranking score changes to those seen with the drug candidates identified by the hypothesis-driven approach. For example, when *glutathione* was removed, the measure of proximity of both *l-glutamine* and *carglumic acid* dropped substantially (from 0.0000945 and 0.0000942 to 0.0000022 and 0.0000594) resulting in lower rankings whereas thimerosal disappeared completely (**Supplementary File 4.2**). This observation could be partly because (1) *thimerosal* interacts with nodes that establish a connection with *glutathione* and *crizotinib*, and (2) *thimerosal* establishes direct interaction with *glutathione* and *crizotinib*.
When *glutathione, crizotinib*, and *curcumin* were collectively removed, several notable drug candidates surfaced among the top hits, including *penicillamine, pregabalin, dexamethasone*, *midostaurin,* and *treprostinil* (**Supplementary File 4.2**). Similar to when *crizotinib* and *curcumin* were individually removed, the measure of proximity of *l-glutamine* and *carglumic acid* increased from 0.0000945 and 0.0000942 to 0.0001621 and 0.0001629 for *crizotinib* removal, and from 0.0000945 and 0.0000942 to 0.0001306 and 0.0001308 for *curcumin* removal respectively (**Supplementary File 4.2**).

4.3.5 In silico validation of top hit candidate drugs

To validate the COVID-19 drug predictions, we aimed to investigate how enriched the potential candidate drugs are as anti-COVID drugs in other databases. Specifically, we conducted RWR analyses using hypothesis-driven seeds on drug data extracted from DrugCombDB (version 2.0), a drug resource database [407]. We implemented these analyses to investigate whether we were able to predict known efficacious COVID-19 drugs (**Table 4.4**). Among the top-ranked drugs (**Supplementary File 4.3**) revealed by these analyses were *dexamethasone* (rank 1), *simvastatin* (rank 7), *cyclosporine* (rank 8), *hydrocortisone* (rank 9), *paclitaxel* (rank 11), *indomethacin* (rank 15), and *methotrexate* (rank 16).

4.4 Discussion

In this work, we attempted to employ computational analyses for the prediction of drug repurposing candidates tailored for disease-state-specific COVID-19 treatment. Leveraging a combination of knowledge graphs (DRKG and COVID-19 KG), along with COVID-19 disease-phase specific omics-graphs generated from experimental proteomics, transcriptomics, metabolomics, and lipidomics data, enabled the identification of various drug repurposing candidates that could potentially be useable as treatments for specific COVID-19 disease states. We implemented multiXrank, a random walk algorithm capable of handling multiple multi-layered graphs and integrated drug data to predict candidate drugs for the mild, moderate, and severe COVID-19 disease states. The analysis resulted in the generation of multi-layered graphs that described the exploration of seed nodes across different disease-state specific omics-graphs.

Our initial analysis focused on existing knowledge graphs which yielded a ranked list of potential drug candidates for COVID-19 treatment (**Table 4.3**). Further analyses that combined disease-state specific omics-graph to the knowledge graphs highlighted potential drugs for mild, moderate, and severe COVID-19 (**Table 4.4** and **Table 4.6**). Additionally, the network topology analysis (**Table 4.5** and **Table 4.7**) of the multi-layered graphs generated from the RWR provided more insights into the influence of the molecular features that establish subnetworks, and how they influence the predicted drugs. For instance, the drugs predicted using hypothesis-driven seeds were predominantly immunosuppressive agents (**Table 4.4**) including, but not limited to, *sirolimus*, *tocilizumab*, and *cyclosporine*. In general, immunosuppressive drugs might have a beneficial effect in the moderate to severe phase of COVID-19 because it is in this phase when dysregulated pro-inflammatory immune responses can precipitate tissue damage and result in acute respiratory distress syndrome, organ failure, and mortality [408]. On the other hand, drugs predicted using the data-driven approach either had antioxidant properties (such as glutathione, and curcumin), or were inhibitors of tyrosine kinase, histone deacetylase, methyltransferase, and protein synthesis (**Table 4.6**). The antioxidants can protect immune system cells and those directly targeted by SARS-CoV-2 from oxidative stress. For example, *glutathione* is an antioxidant that is believed to boost immune and other tissue systems by shielding

cells against oxidative stress-induced damage [391]. *Glutathione* is assumed to have a vital role in maintaining the balance of reactive oxygen species (ROS), and aids in diverse cellular processes including immune responses [384, 391]. Notably, oxidative stress reflects an imbalance between ROS generation and antioxidation mechanisms [409] and plays an important role in COVID-19 onset, progression, and severity [410-412], possibly by exacerbating inflammation and precipitating tissue damage [413]. This therefore suggests that *glutathione'*s capability to counteract ROS and diminish oxidative stress holds promise for mitigating some of the adverse effects inflicted by the virus [391]. *Glutathione* and *SOD2* bring unique strengths such that their combined efforts provide a multi-layered defence against oxidative stress and its harmful consequences. By neutralizing superoxide radicals, *SOD2* sets the stage for glutathione to efficiently handle other free radicals and detoxify the cell. Also, given the aggressive inflammatory response and the production of cytokines occurring during severe COVID-19 disease states, some known inhibitors of receptor tyrosine kinases and cell proliferation, such as *crizotinib* and *vorinostat*, have been investigated as COVID-19 treatments [392, 393]. For instance, a recent study has shown that histone deacetylase inhibitors modulate immune responses in stimulated monocytes [393], whereas tyrosine kinase inhibitors have the potential to reverse pulmonary insufficiency because of their anti-inflammatory activities, cytokine suppression activities, or antifibrotic activities [392].

Overall, these prioritized drug repurposing candidates (**Table 4.3, Table 4.4,** and **Table 4.6**) exhibit the potential to target a multitude of specific biological pathways and gene ontology processes that are associated with COVID-19 outcomes. Among these candidates are those that have shown promise in treating other diseases or conditions such as cancer, malaria, viral infections, and obstructive pulmonary disease. For instance, *glutathione* shows activity against HIV, influenza A, and hepatitis C by inhibiting viral replication and modulating immune response [414]. *Curcumin* shows activity against HIV, influenza A, hepatitis C, and Dengue virus by inhibiting viral entry and replication [415]. *Vorinostat* shows activity against HIV by inhibiting viral replication and disrupting HIV-1 latency in patients on antiretroviral therapy [416]. Consequently, there is the possibility of repurposing these to combat COVID-19.

The approach implemented in this study is relevant to identifying drugs that warrant further exploration. It is important to mention that some of the drugs that were highly ranked in our hypothesis-driven and data-driven analyses as potential COVID-19 treatments have not, to our knowledge, been tested before in the context of COVID-19 treatment. These included *podofilox*, *calcipotriol*, *vinblastine*, *etoposide*, and *carglumic acid* identified from the data-driven approach, and *paclitaxel* identified from the hypothesis-driven approach. The shortest path analysis revealed molecular features that lie closer to the drugs (**Supplementary Files 4.4 and 4.5**). These drugs warrant further investigation as potential COVID-19 therapeutics. Considering the drugs generated from the data-driven approach, *podofilox* for example inhibits topoisomerase I by stabilizing the covalent complex formed between the enzyme and a broken DNA strand [417]. This prevents religation, causing DNA damage and eventually cell death [417]. *Podofilox* is known to down-regulate SOD2 expression in cancer cells and indirectly modulate *SOD2* activity, impacting reactive oxygen species levels and influencing cell survival and death. The reactive oxygen species' impact on COVID-19 progression [410-412] suggests that *podofilox* may have a potential role in COVID-19 treatment. Also, *etoposide* possesses an immunosuppressive effect. While suppressing certain immune cells, *etoposide* may also selectively eliminate abnormal or activated T cells involved in the inflammatory process [418]. This can be beneficial in some inflammatory conditions, potentially mitigating immune-mediated damage. Additionally, *etoposide* has the potential to influence the production of certain cytokines and signalling molecules involved in immune communication [419]. In the context of COVID-19, this could have both pro-inflammatory and anti-inflammatory effects. *Vinblastine* can modulate the production of certain cytokines, signalling molecules that orchestrate immune responses [420].

Additionally, our analysis revealed that the HIV protease inhibitor, *ritonavir*, which is used in conjunction with *nirmatrelvir* in the highly effective COVID-19 medication, *paxlovid*, received a lower ranking with a measure of proximity between 0.00000032 and 0.00000034 in our data-driven analyses. This may be attributed to the characteristics of the knowledge graphs that we used in that they contained limited information about the impacts of *ritonavir* on the transcriptomics, proteomics, lipidomics, and metabolomics of human cells. As a result, the topology of the networks that we used was biased in favour of ranking better-researched compounds like

dexamethasone and *tocilizumab*. In our analysis, we also did not identify *nirmatrelvir* among the ranked drug candidates. This observation is partly attributed to the choice of seeds for the RWR analyses and also to the fact we focus here on the omics networks from the host because *nirmatrelvir* targets the viral genome, (polyprotein 1ab), and could, therefore not be captured by the network exploration.

In summary, this study presents a network-based approach for identifying potential drug repurposing candidates for different stages of COVID-19, from mild to severe. This analysis, drawing on diverse datasets, has provided valuable insights that contribute to ongoing efforts to combat endemic COVID-19 and the long-term health consequences of repeated SARS-CoV-2 infections. While some of the identified drugs have been implemented in disease management, several promising candidates are yet to be investigated for COVID-19 disease treatment. The predictions provide a starting point for further experimental validation and clinical investigations. Ensuring the safety and efficacy of new COVID-19 drugs requires rigorous experimental and clinical testing and validation. *In vitro* analyses and clinical trials must be conducted to determine the optimal dosages, administration protocols, and potential interactions with other medications. This approach would ultimately be needed to provide actual proof that many of the less well-studied drug repurposing candidates that we have identified could indeed be used to effectively treat COVID-19.

4.5 Limitations

Considering the limitations of the DRKG and COVID-19 KG data, which predate large-scale drug evaluations, incorporating more recent drug information is crucial for future studies. While this study identified potential drugs for acute COVID-19 treatment, it did not address Long COVID or the impact of comorbidities and disease severity. Thus, future investigations should explicitly explore treatment options for Long COVID Furthermore, our drug prediction analysis did not account for COVID-19 comorbidities and recommends further studies to refine drug prediction analysis specific for mild, moderate, or severe COVID-19-infected patients experiencing other infections. To maximize the potential of our approach, future work should incorporate drug synergy analysis. By systematically evaluating how the therapeutic activities of different drugs

might combine, we can identify and prioritize the most promising combination therapies for further testing and development in the fight against COVID-19.

4.6 Conclusion

This chapter explored an integrative multi-layered network approach to identify drugs for repurposing against COVID-19 disease phases. We analysed multi-omics data (proteomics, transcriptomics, metabolomics, and lipidomics) and drug-related data (drug repurposing knowledge graph and COVID-19 knowledge graph) using RWR technique. Notably, we conducted RWR analyses in both hypothesis-driven and data-driven manners, incorporating information specific to disease severity levels (mild, moderate, severe) via dedicated disease-state specific omics-graphs. Our multi-layered network approach successfully identified potential drug candidates for repurposing against mild, moderate, and severe COVID-19. The incorporation of disease-state specific omics data significantly influenced the predicted drug candidates. Both immune-suppressive and pathway-targeting mechanisms emerged as potential approaches for COVID-19 treatment.

4.7 Contribution to knowledge and science

We believe that this chapter presents a study that has significant benefits to the scientific community. To begin with, we proposed an innovative drug prediction approach that accounts for specific disease states by incorporating disease-state specific omics graphs generated from proteomics, transcriptomics, metabolomics, and lipidomics data to identify targeted treatment options for mild, moderate, and severe COVID-19 cases [315]. Our research illustrates that integrative analysis of drug data and multi-omics data allows for the prioritization of drug candidates across different phases of COVID-19, providing a comprehensive foundation for shaping future therapeutic strategies. Also, we show interactions involving drugs, proteins, transcripts, metabolites, lipids, and biological processes.

The novelty of the methodology is that it has provided the baseline to perform drug prediction analysis for specific disease states and can be applied to other diseases with varying severity levels. Moreover, the method can be readily expanded to include

additional data streams such as synergistic drug combination information. Overall, the work done in this chapter underscores the critical role of this kind of data integration in deciphering potential treatment options for diseases with heterogeneous clinical representation. Building upon prior research, we highlight the value of biosignature-drug associations within biological networks for prioritizing potential drug repurposing candidates across different phases of COVID-19.

To adhere to FAIR principles and facilitate replication of our approach, we provide a containerized workflow with an expanded readme file at https://github.com/francis-agamah/Network-based-multi-omics-disease-drug-associations_drugs-for-COVID-19-disease-phases. All other data and its supplementary files generated during this study are included in the github repository.

Chapter 5

General conclusion and discussion

COVID-19 remains one of the leading health problems globally. The global response to the disease has been multifaceted, encompassing a range of efforts such as vaccinations, testing and isolation, surveillance, and non-pharmaceutical interventions that are aimed at both mitigating the immediate impact of the virus and preparing for its long-term consequences [421]. Despite the success gained, the progress toward understanding disease progression is limited by enormous clinical heterogeneity and its implications for prognosis and treatment [422-424].

In this thesis, we investigated the contribution of host multi-omics biosignatures to COVID-19 susceptibility and progression, applying a multi-omics approach to analyse various biomolecules like proteins, transcripts, metabolites, and lipids. By examining these biosignatures across mild, moderate, and severe COVID-19, we aimed to elucidate the molecular basis of the observed clinical heterogeneity. In addition, we explored potential FDA-approved drug candidates that can be repurposed for COVID-19 treatment. By testing specific hypotheses outlined in **Chapter 1**, we gained a deeper understanding of how changes in biosignatures influence COVID-19 severity.

While a number of multi-omics studies on COVID-19 have been performed [26, 44, 69, 207], an integrative network analysis of proteomics, metabolomics, lipidomics, and transcriptomics data from independent sources to explore biosignatures that discriminate disease states based on both hypothesis-driven and data-driven approaches was still lacking. Network-based integrative approaches have revolutionized multi-omics analysis by providing a graph-based framework for visualizing and analysing interactions across different omics layers (e.g., transcriptomics, proteomics, lipidomics, metabolomics) revealing hidden relationships and pathways that single-omics studies might overlook (**Chapter 2**). This framework serves as a powerful analogy to the intricate interconnectedness of molecular processes within a cell, enabling researchers to uncover novel insights into biological systems and their role in health and disease. However, we observed that various

network approaches have different strengths and weaknesses (**Chapter 2**). Thus in chapter two, we focused on a comprehensive review of current computational approaches for the network-based integration of multi-omics data and their different applications [208]. The chapter introduces the field of integrative multi-omics and classifies the methods into machine learning-based, diffusion/propagation-based, causality-based, and network inference-based methods. In particular, the chapter discusses the application of the computational approaches that implement all these methods to identify biosignatures, categorize disease subtypes, detect crosstalk between biosignatures, infer causality, discover potential drugs, and uncover molecular drivers of physiological and pathological mechanisms in the context of COVID-19. The chapter gives example applications and provides at the end an overview of the challenges and recommendations. Overall, chapter two of this thesis gives guidance to researchers who are interested in the field of network-based integrative multi-omics.

In the field of multi-omics, the main goal is to integrate multiple heterogeneous omics data from different sources and transform them into biological knowledgebases and translatable findings. The heterogeneity of different types of omics data presents difficulties in data integration at the level of both the omics clinical metadata and the experimental feature data. Thus, we observed the need for data harmonization and standardization as one of the main tasks for computational multi-source multi-omics studies. For this reason, we proposed a harmonization approach that formed the basis for integrating patient metadata to help integrate proteomics, transcriptomics, metabolomics, and lipidomics experimental data from independent studies. Also, we standardized the multi-omics data by way of implementing unified and persistent identifiers, thus contributing significantly to studying the complex interplay between these omics layers and COVID-19 disease severity. The harmonization and standardization process described in chapter three is published and accessible at https://doi.org/10.1101/2023.09.29.560110.

In chapter three, we tested two hypotheses: (1) investigating biosignatures across different phases of COVID-19 disease will provide insights into the molecular underpinnings of the enormous clinical heterogeneity of COVID-19, and (2) the association between biosignatures within a biological network would permit prioritizing

biosignatures that discriminate disease states. Specifically, we investigated the differential impact of *IL-6* and *IL-6R* across mild, moderate, and severe COVID-19 states and found that cross-layer interaction among *IL-6* and *IL-6R* is associated with disease severity. Further investigation of other factors that contribute to disease severity using data-driven RWR analysis showed that other cross-layer interactions between omics-derived molecular features may also contribute to COVID-19 disease severity. We reported on biosignatures found to discriminate disease states (**Table 3. 5** and **Table 3. 6**). Finally, we report on an elegant multi-layered system (http://cytoscape.h3africa.org) to present omics data, which interacts at different levels and is potentially very useful for a robust identification of disease severity. Overall, in chapter three, we presented an approach that facilitates data integration and would facilitate integration with more data including some recent studies [207, 425, 426] that we have not been able to incorporate yet in this project. The data and scripts implemented for this chapter together with a containerized workflow is accessible at https://github.com/francis-agamah/Multi-source-multi-omics-network-analysis.

Aiming to predict drug candidates for mild, moderate, and severe COVID-19, in chapter four of this thesis, we tested the hypothesis that associations between biosignatures and drugs within a biological network would enable the prioritization of drug repurposing candidates. We leveraged disease-state specific omics-graphs, a drug repurposing knowledge graph, and a COVID-19 knowledge graph in an integrative network analysis to prioritize potential drug repurposing candidates that could be applied in different COVID-19 disease phases. We used both hypothesis-driven and data-driven RWR approaches to predict drug repurposing candidates. By analysing the network topology and centralities (degree, betweenness, and closeness centrality), we were able to pinpoint crucial drug repurposing candidates for mild, moderate, and severe COVID-19, as well as other molecular features that act as hubs or influencers within the biological network. The molecular features serving as hubs regulate vital cellular processes, providing potential targets for therapeutic interventions or biomarkers for disease detection. Overall, chapter four of this thesis provided significant insights into potential drug repurposing candidates that can be further validated for mild, moderate, and severe COVID-19. This chapter is published online and accessible at https://zenodo.org/records/10568146. Also, the data and scripts implemented for this chapter together with a containerized workflow is

accessible at https://github.com/francis-agamah/Network-based-multi-omics-disease-drug-associations_drugs-for-COVID-19-disease-phases.

In conclusion, we showed that COVID-19 disease severity is influenced by biosignature interactions across different omics-layers of the multi-layered knowledge graphs. Additionally, we showed that integrative multi-omics analysis revealed biosignatures and biosignature interactions associated with COVID-19 disease states. Furthermore, we showed that integrative multi-omics and drug-related data prioritized potential drug candidates that could be repurposed for the treatment of mild, moderate, or severe COVID-19 cases. We suggest that our methodology and findings are useful in understanding host contributions to COVID-19 disease risk, clinical heterogeneity, and disease progression in the context of multi-omics, while at the same time illuminating potential candidate drugs for repurposing. Consequently, the work presented in this thesis can facilitate genetic diagnosis and disease-state-specific treatments.

5.1 Limitations and recommendations

Firstly, in chapter three, we used multi-omics data from Overmyer et al., [69] and Su et al., [26] for the analyses and reported findings. However, this could be extended by incorporating more multi-omics data from different sources as well as including other omics data types such as epigenomics, and microbiomics [427-429]. One key challenge to adding additional multi-omics data from different sources is the bias due to separate experiments and diverse samples, leading to further technical and biological variations. This hinders the effectiveness of the analysis. However, recent efforts like the European Infrastructure for Translational Medicine (EATRIS)-Plus multi-omics toolbox (https://motbx.eatris.eu/) provide several validated resources such as standardized operating procedures, reference protocols, and reference values and therefore help to harmonize and integrate data [430]. The toolbox serves as a knowledge hub to support translational researchers and clinicians in their research and can significantly reduce technical and biological variations, ultimately leading to higher quality multi-omics data and better performance of the algorithms focused on extracting actionable information from these types of data [430].

Secondly, the harmonization approach proposed in chapter three could be extended by factoring in the treatment options received by patients as these may influence the measured levels of the omics features. We do, however, acknowledge the challenge involved in this process given that the different studies may use different treatment options. A promising way to overcome this limitation is to consider common parameters that are measured during treatment trials. These universal parameters could be measures of the therapeutic effects such as the recruitment of immune cells. Another potential way is to consider alternative clinical measures and biomarkers if possible.

Thirdly, the patient metadata upon which the harmonization process is based suffers high heterogeneity because the data come from different experiments and conditions. As a result, the variables in the patient metadata differ between studies, making it impossible to utilize all variables for harmonization. This limitation could potentially be mitigated by utilizing a standardized case report form or patient metadata template. For instance the Virus Outbreak Data Network (VODAN)-Africa has retooled the CEDAR metadata to capture clinical and research data (non-clinical data) [431]. Once such metadata template is widely adopted, it will contribute to streamlining research not only in the context of COVID-19.

Lastly, in chapter four, we recommend expanding the work to include several additional omics data types and drug-related data.

Despite these challenges, the potential of network-based integrative approaches in multi-omics research is immense. By embracing this holistic perspective, scientists can shed light on the intricate workings of life, paving the way for breakthroughs in precision medicine, drug discovery, and our understanding of human health and disease.

5.2 Future perspectives

5.2.1 Implementing FAIR principles in multi-omics research

The FAIR (Findable, Accessible, Interoperable, and Reusable) principles offer a crucial perspective for multi-omics research, promoting an open and collaborative environment that accelerates discovery and unlocks the full potential of the multi-omics field [432, 433]. By embracing FAIR principles, multi-omics research can overcome data silos, foster collaboration, accelerate discovery, and enhance open and transparent research. Currently, there remain many multi-omics studies that are not FAIR yet. Below, we explore how the adoption of the FAIR principles can improve multi-omics research.

Findability: The increasing volume and complexity and size of multi-omics data has emerged as a concern for data hosting and data analytical systems. Applying metadata standards and searchable repositories would ensure that the omics datasets are easily discoverable by researchers regardless of their location or technical expertise. This opens doors for repurposing existing data, building upon previous studies, and avoiding redundant research.

Accessibility: Ensuring that multi-omics data are hosted in repositories capable of hosting multi-dimensional data rather than depositing it in repositories designed to host specific omics data types (such as genomics, metabolomics, proteomics, etc.) and, assay types (such as arrays, sequencing, etc.) could go a long way to enhance data accessibility [434]. Creating infrastructures such as cloud-based hosting and analysis platforms could promote data visiting and the accessibility of multi-omics data. Examples of such platforms include FAIR Data Cube [432] and Lifebit (https://lifebit.ai/).

Interoperability: Implementing standardized data formats and common ontologies for multi-omics features would enable seamless integration of diverse omics datasets, facilitating comprehensive cross-analysis and revealing hidden connections across different biological layers. This would contribute significantly not only to robust network-based analyses but to multi-omics research in general.

Reusability: Good data management and FAIR principles would contribute to clearer insights, more reliable results, and faster scientific progress by promoting knowledge discovery, reproducibility, and data reuse. Well-documented data with detailed

protocols and analysis workflows ensures its reusability for future research [433]. Analysis workflows could be containerized or packaged as FAIR Digital Objects [433]. This transparency allows others to validate findings, reproduce analyses, and build upon existing knowledge, propelling scientific progress forward and thereby enhancing research credibility.

5.2.2 Integrating diverse datasets beyond multi-omics data

The work presented in this thesis lays a stepping stone for future endeavours on several fronts, particularly in the context of the integration of multi-omics data and drug-related data. One exciting direction involves expanding the multi-omics canvas beyond biological data. Integrating phenotypic and environmental exposure data, including crucial social determinants of health (SDoH), into the analysis promises deeper insights into the molecular profiles associated with COVID-19 severity and other diseases [435-437]. The complex interplay of genetic and environmental factors could help understand the patterns underlying genetic susceptibility, environmental factors, and socioeconomic inequities in shaping disease progression [436].

5.2.3 Refining tools for multi-omics integration

Another future pursuit lies in refining the tools of multi-omics integration. In the context of random walks, scaling random walk algorithms such that all nodes within the networks can be used as seeds simultaneously and sequentially. Although computationally intensive, having such a framework would enable testing multiple hypotheses. Besides the network-based approaches, multi-omics data could be integrated by using other computational approaches depending on the purpose of the research problem [208].

While there are varying computational approaches, integrating multi-omics results from these analytical methods poses significant challenges and necessitates careful consideration of various factors to ensure meaningful and accurate interpretation. We explore below the existing needs and challenges associated with the integration of multi-omics results.

Cross-platform integration challenges: Different methods for certain type of analysis have specific underlying algorithms and parameters implemented on which the analysis is performed. Before conceptualizing results integration from different

methods, there is the need to first benchmark the approaches using benchmark datasets to rigorously compare variations in terms of the strength of each method, the accuracy and reliability of the generated results [438]. During benchmarking, it is crucial to consider the benchmarking datasets, parameter setting of algorithms, and software version as this influence the analytical results. Also, it is important to consider standard quantitative performance metrics when evaluating the performance of the methods. The optimal metric for evaluating your results hinges on the type of analysis and data you're using. For instance, in classification tasks with known labels (ground truth), metrics like true positive rate (TPR), false positive rate (FPR), and false discovery rate (FDR) reign supreme. In contrast, clustering tasks often rely on metrics like F1 score, adjusted Rand index, normalized mutual information, precision, and recall to assess accuracy [438].

Biological interpretability and contextualization: The integration of multi-omics results from different methods must not only address technical challenges but also ensure biological relevance. The interpretation of integrated data requires a deep understanding of the underlying biology and the context-specific interactions between different omics layers. While biological interpretability focuses on understanding the biological meaning behind the statistical patterns and associations identified in multi-omics analyses, contextualization, on the other hand, emphasizes the importance of framing multi-omics findings within a relevant biological context [439]. A way to achieve this is through interpretation by experts as well as experimental validation of the integrated results.

5.2.4 Application of methods to other diseases

Besides COVID-19 research, the methods and techniques described in this thesis can be applied in the context of other diseases, particularly those with varying severity levels. For example, the approach described in chapter three of this thesis can be implemented to investigate host multi-omics factors contributions to symptomatic and asymptomatic malaria infections to understand the multi-omics features that discriminate the stages of this disease. Similarly, the method in chapter four of this thesis can be applied to explore drug repurposing candidates for diseases like malaria and cancer.

5.2.5 Causality vs. correlation

Network analysis often reveals associations between biological elements, but differentiating true causal relationships from mere associations remains a significant challenge in network analysis such that establishing causal relationships generally requires further investigation. Integrating experimental approaches and temporal data into network models can strengthen our understanding of cause-and-effect dynamics. Whereas integrating functional assays and experimental validation is crucial for translating network insights into tangible knowledge, unraveling causality often requires robust algorithms and advanced statistical methods to account for confounding factors and network noise [208]. Interestingly, Bayesian networks and network inference methods [440] offer a powerful tool for generating promising hypotheses about cause-and-effect relationships within biological systems as described in chapter 2.

5.2.6 Personalized medicine applications

Network-based multi-omics offers a promising perspective for personalized medicine by paving the way for understanding individual variations, predicting disease risks, and designing personalized therapies [441]. For instance, a network-based approach can be implemented to understand individual variations. This is because everyone’s genome, protein expression, and metabolism form a unique network configuration, contributing to their susceptibility to specific diseases and responses to treatments. Network-based analyses can identify these individual network signatures, providing insights into disease mechanisms and potential therapeutic targets. This can be achieved by clustering based on patient similarity networks utilizing similarity network fusion algorithms [441, 442].

Also, network models can be used to simulate how an individual's biological system might respond to different drugs or environmental factors [443, 444]. Accurately simulating how individual networks respond to interventions requires factoring in temporal changes, environmental influences, and individual variability. Advanced computational models that incorporate real-world (and preferably real-time) data are essential for improving predictive accuracy [445]. This predictive power could allow for personalized (and potentially dynamic) risk assessments, preventive strategies, and tailored treatment plans. Nevertheless, leveraging network-based approaches to

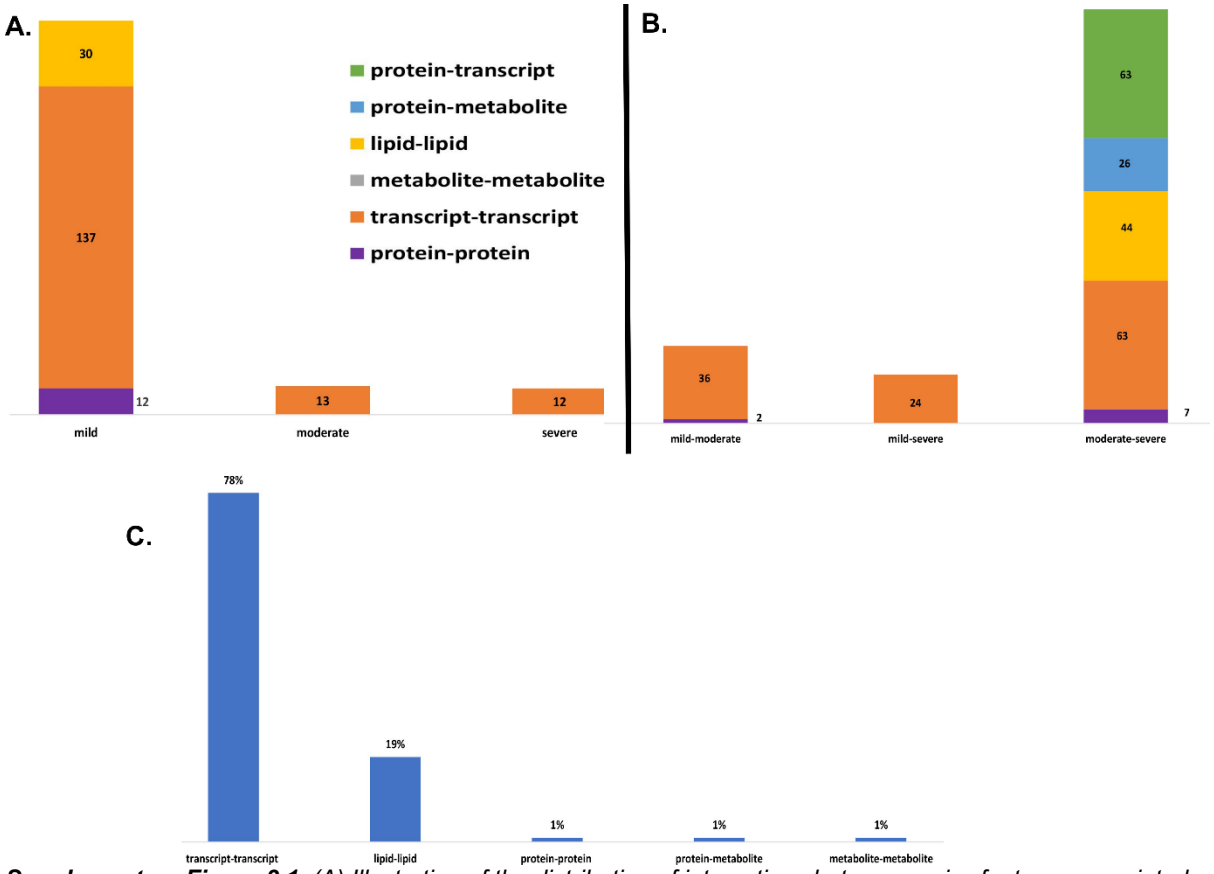
develop personalized medicine strategies for complex diseases remains in its early stages.

Further research is needed to translate network insights into clinically actionable targets and diagnostic tools. Importantly, translating network-based insights into actionable clinical tools requires rigorous validation through (prospective) clinical trials and implementation in healthcare settings. Developing user-friendly platforms and integrating these tools into clinical workflows is key for routine use.

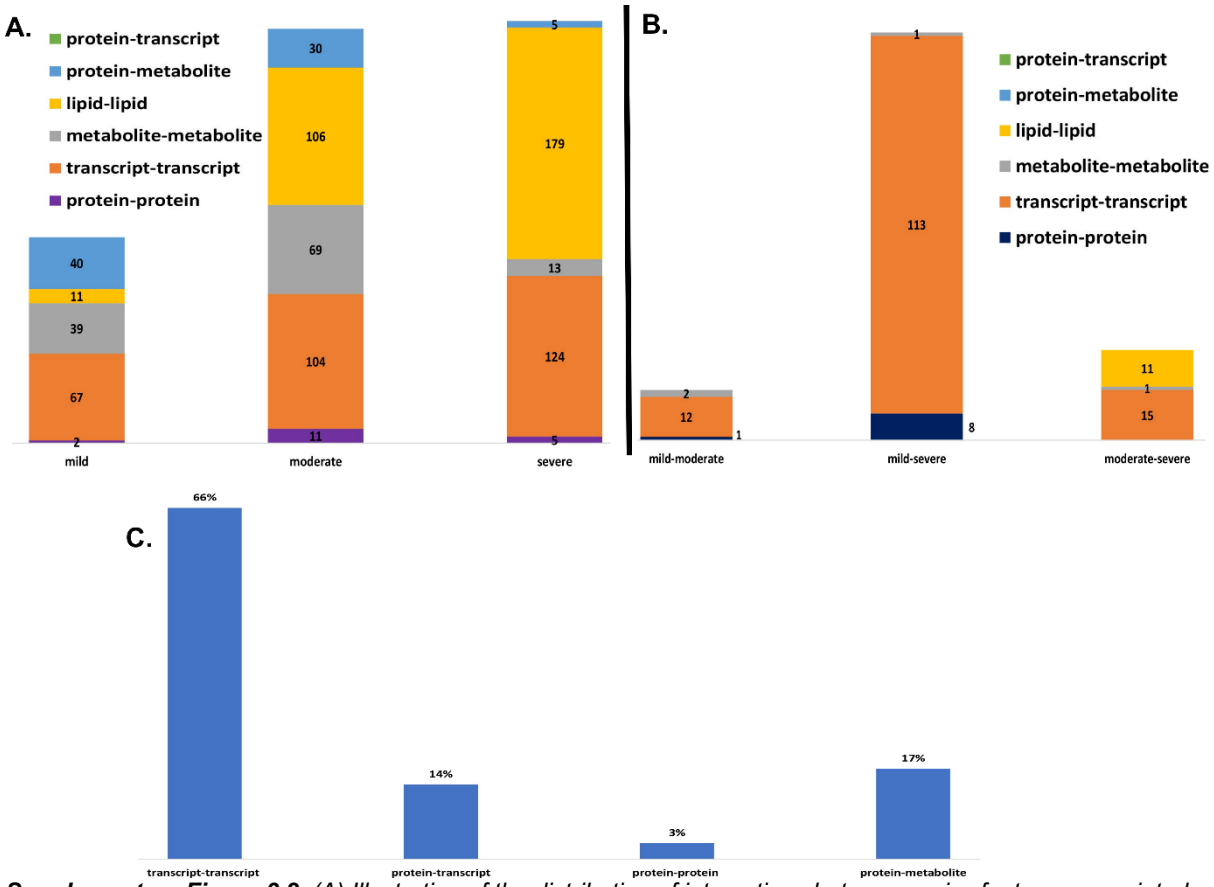
Appendices

The supplementary files for chapter three consisting of supplementary data, supplementary file 3.1, supplementary file 3.2, supplementary file 3.3, and supplementary file 3.4 can be accessed from the GitHub repository, https://github.com/francis-agamah/Multi-source-multi-omics-network-analysis

The supplementary files for chapter four consisting of supplementary data, supplementary file 4.1, supplementary file 4.2, supplementary file 4.3, supplementary file 4.4, and supplementary file 4.5 can be accessed from the GitHub repository https://github.com/francis-agamah/Network-based-multi-omics-disease-drug-associations_drugs-for-COVID-19-disease-phases



Supplementary Figure 3.1. *(A) Illustration of the distribution of interactions between omics features associated with one disease state identified from the data-driven approach. (B) Illustration of the distribution of interactions between omics features associated with two disease states identified from the data-driven approach. (C) Illustration of the distribution of interactions between omics features associated with three disease states identified from the data-driven approach.*



Supplementary Figure 3.2. *(A) Illustration of the distribution of interactions between omics features associated with one disease state identified from the hypothesis-driven approach. (B) Illustration of the distribution of interactions between omics features associated with two disease states identified from the hypothesis-driven approach. (C) Illustration of the distribution of interactions between omics features associated with three disease states identified from the hypothesis-driven approach.*

***Supplementary Table 4.1**. Description of the edge-types in Drug Repurposing Knowledge Graph*

Source	Predicate	Subject	Object	Number of edges
bioarx	DrugHumGen	Compound	Gene	24501
bioarx	DrugVirGen	Compound	Gene	1165
bioarx	Coronavirus_ass_host_gene	Disease	Gene	129
bioarx	Covid2_acc_host_gene	Disease	Gene	332
bioarx	HumGenHumGen	Gene	Gene	58094
bioarx	VirGenHumGen	Gene	Gene	535
DGIDB	ACTIVATOR	Gene	Compound	316
DGIDB	AGONIST	Gene	Compound	3012
DGIDB	ALLOSTERIC MODULATOR	Gene	Compound	317
DGIDB	ANTAGONIST	Gene	Compound	3006
DGIDB	ANTIBODY	Gene	Compound	188
DGIDB	BINDER	Gene	Compound	143
DGIDB	BLOCKER	Gene	Compound	979
DGIDB	CHANNEL BLOCKER	Gene	Compound	352
DGIDB	INHIBITOR	Gene	Compound	5971
DGIDB	MODULATOR	Gene	Compound	243
DGIDB	OTHER	Gene	Compound	11070
DGIDB	PARTIAL AGONIST	Gene	Compound	75
DGIDB	POSITIVE ALLOSTERIC MODULATOR	Gene	Compound	618
DRUGBANK	carrier	Compound	Gene	720
DRUGBANK	ddi-interactor-in	Compound	Compound	1379271
DRUGBANK	enzyme	Compound	Gene	4923
DRUGBANK	target	Compound	Gene	19158
DRUGBANK	treats	Compound	Disease	4968
DRUGBANK	x-atc	Compound	Atc	15750
GNBR	A+	Compound	Gene	1568
GNBR	A-	Compound	Gene	1108
GNBR	B	Compound	Gene	7170
GNBR	C	Compound	Disease	1739
GNBR	E+	Compound	Gene	1970
GNBR	E-	Compound	Gene	2918
GNBR	E	Compound	Gene	32743
GNBR	J	Compound	Disease	1020
GNBR	K	Compound	Gene	12411
GNBR	Mp	Compound	Disease	495
GNBR	N	Compound	Gene	12521
GNBR	O	Compound	Gene	5573
GNBR	Pa	Compound	Disease	2619
GNBR	Pr	Compound	Disease	966
GNBR	Sa	Compound	Disease	16923

GNBR	T	Compound	Disease	54020
GNBR	Z	Compound	Gene	2821
GNBR	B	Gene	Gene	8164
GNBR	D	Gene	Disease	500
GNBR	E+	Gene	Gene	10838
GNBR	E	Gene	Gene	418
GNBR	G	Gene	Disease	2055
GNBR	H	Gene	Gene	2509
GNBR	I	Gene	Gene	5434
GNBR	J	Gene	Disease	30234
GNBR	L	Gene	Disease	48384
GNBR	Md	Gene	Disease	1279
GNBR	Q	Gene	Gene	19372
GNBR	Rg	Gene	Gene	11018
GNBR	Te	Gene	Disease	2836
GNBR	U	Gene	Disease	6432
GNBR	Ud	Gene	Disease	407
GNBR	V+	Gene	Gene	8689
GNBR	W	Gene	Gene	280
GNBR	X	Gene	Disease	1324
GNBR	Y	Gene	Disease	1948
GNBR	in_tax	Gene	Tax	14663
Hetionet	AdG	Anatomy	Gene	102240
Hetionet	AeG	Anatomy	Gene	526407
Hetionet	AuG	Anatomy	Gene	97848
Hetionet	CbG	Compound	Gene	11571
Hetionet	CcSE	Compound	Side Effect	138944
Hetionet	CdG	Compound	Gene	21102
Hetionet	CpD	Compound	Disease	390
Hetionet	CrC	Compound	Compound	6486
Hetionet	CtD	Compound	Disease	755
Hetionet	CuG	Compound	Gene	18756
Hetionet	DaG	Disease	Gene	12623
Hetionet	DdG	Disease	Gene	7623
Hetionet	DlA	Disease	Anatomy	3602
Hetionet	DpS	Disease	Symptom	3357
Hetionet	DrD	Disease	Disease	543
Hetionet	DuG	Disease	Gene	7731
Hetionet	GcG	Gene	Gene	61690
Hetionet	GiG	Gene	Gene	147164
Hetionet	GpBP	Gene	Biological Process	559504
Hetionet	GpCC	Gene	Cellular Component	73566

Hetionet	GpMF	Gene	Molecular Function	97222
Hetionet	GpPW	Gene	Pathway	84372
Hetionet	Gr>G	Gene	Gene	265672
Hetionet	PCiC	Pharmacologic Class	Compound	1029
INTACT	ASSOCIATION	Compound	Gene	1447
INTACT	DIRECT INTERACTION	Compound	Gene	155
INTACT	PHYSICAL ASSOCIATION	Compound	Gene	203
INTACT	ADP RIBOSYLATION REACTION	Gene	Gene	58
INTACT	ASSOCIATION	Gene	Gene	112390
INTACT	CLEAVAGE REACTION	Gene	Gene	93
INTACT	COLOCALIZATION	Gene	Gene	3468
INTACT	DEPHOSPHORYLATION REACTION	Gene	Gene	303
INTACT	DIRECT INTERACTION	Gene	Gene	6950
INTACT	PHOSPHORYLATION REACTION	Gene	Gene	1328
INTACT	PHYSICAL ASSOCIATION	Gene	Gene	129318
INTACT	PROTEIN CLEAVAGE	Gene	Gene	67
INTACT	UBIQUITINATION REACTION	Gene	Gene	371
STRING	ACTIVATION	Gene	Gene	81355
STRING	BINDING	Gene	Gene	315875
STRING	CATALYSIS	Gene	Gene	343533
STRING	EXPRESSION	Gene	Gene	757
STRING	INHIBITION	Gene	Gene	28959
STRING	OTHER	Gene	Gene	310690
STRING	PTMOD	Gene	Gene	15113
STRING	REACTION	Gene	Gene	400426

Supplementary Table 4. 2. *Description of the edge-types in COVID-19 Knowledge Graph*

Edge-type	**Number of interactions**
COVID-19 genes interaction with chemicals	28634
Phenotypes (disease gene-associated biological processes) interactions with COVID-19 drugs (chemical/compound)	1571
Phenotypes (disease gene-associated biological processes) interactions with COVID-19 genes	1610
SARS-CoV-2 baits interaction with host genes	1114
Pathways interaction with COVID-19 genes	692

Supplementary Table 4. 3.*Selected data-driven seeds for random walk network exploration*

Approach	**Seed node**	**Integrated centrality score**	**Feature type**
Data-driven	STAT1	53529.0403	Transcript
	SOD2	2215.5746	Protein
	3-hydroxyoctanoate	1506.9998	Metabolite
	Unknown_mz_815.61548_+_RT_27.063	9936.9781	Lipid

References

1. Kesheh, M.M., et al., *An overview on the seven pathogenic human coronaviruses.* Reviews in Medical Virology, 2022. **32**(2): p. e2282.
2. Maier, H.J., E. Bickerton, and P. Britton, *Coronaviruses: methods and protocols*. 2015: Springer Berlin.
3. Chan, J.F.-W., et al., *A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster.* 2020. **395**(10223): p. 514-523.
4. Lu, R., et al., *Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding.* 2020. **395**(10224): p. 565-574.
5. Chen, N., et al., *Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study.* 2020. **395**(10223): p. 507-513.
6. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* 2020. **579**(7798): p. 270-273.
7. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-19 in real time.* The Lancet infectious diseases, 2020. **20**(5): p. 533-534.
8. Fehr, A.R. and S. Perlman, *Coronaviruses: an overview of their replication and pathogenesis*, in *Coronaviruses*. 2015, Springer. p. 1-23.
9. Naqvi, A.A.T., et al., *Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach.* Biochimica et Biophysica Acta -Molecular Basis of Disease, 2020. **1866**(10): p. 165878.
10. Satarker, S. and M.J.A.o.m.r. Nampoothiri, *Structural proteins in severe acute respiratory syndrome coronavirus-2.* 2020.
11. Silvas, J.A., et al., *Contribution of SARS-CoV-2 accessory proteins to viral pathogenicity in K18 human ACE2 transgenic mice.* Journal of Virology, 2021. **95**(17): p. e00402-21.
12. Holmes, E.C., *Molecular clocks and the puzzle of RNA virus origins.* Journal of virology, 2003. **77**(7): p. 3893-3897.
13. Singh, J., et al., *Evolutionary trajectory of SARS-CoV-2 and emerging variants.* Virology journal, 2021. **18**: p. 1-21.
14. Sánchez, C.M., et al., *Genetic evolution and tropism of transmissible gastroenteritis coronaviruses.* Virology, 1992. **190**(1): p. 92-105.
15. Lucas, M., et al., *Viral escape mechanisms–escapology taught by viruses.* 2001. **82**(5): p. 269-286.
16. Shen, Z., et al., *Genomic diversity of SARS-CoV-2 in Coronavirus Disease 2019 patients.* 2020.
17. Alanagreh, L.a., F. Alzoughool, and M.J.P. Atoum, *The human coronavirus disease COVID-19: its origin, characteristics, and insights into potential drugs and its mechanisms.* 2020. **9**(5): p. 331.
18. Trougakos, I.P., et al., *Insights to SARS-CoV-2 life cycle, pathophysiology, and rationalized treatments that target COVID-19 clinical complications.* Journal of Biomedical Science, 2021. **28**: p. 1-18.
19. Tian, C., et al., *Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections.* Nat Commun, 2017. **8**(1): p. 599.
20. International, H.I.V.C.S., et al., *The major genetic determinants of HIV-1 control affect HLA class I peptide presentation.* Science, 2010. **330**(6010): p. 1551-7.
21. Everitt, A.R., et al., *IFITM3 restricts the morbidity and mortality associated with influenza.* 2012. **484**(7395): p. 519-523.
22. Shirbhate, E., et al., *Understanding the role of ACE-2 receptor in pathogenesis of COVID-19 disease: a potential approach for therapeutic intervention.* Pharmacological Reports, 2021: p. 1-12.

23. Karim, H. and M.S. Khan, *A Systematic Review on Coronavirus Disease 2019 (COVID-19).* 2020.
24. Liu, J., et al., *Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients.* 2020: p. 102763.
25. Hadjadj, J., et al., *Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients.* 2020. **369**(6504): p. 718-724.
26. Su, Y., et al., *Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19.* Cell, 2020.
27. Al Khatib, H.A., et al., *Within-Host Diversity of SARS-CoV-2 in COVID-19 Patients With Variable Disease Severities.* Front Cell Infect Microbiol, 2020. **10**: p. 575613.
28. Khare, K. and R. Pandey, *Cellular heterogeneity in disease severity and clinical outcome: Granular understanding of immune response is key.* Frontiers in Immunology, 2022. **13**: p. 973070.
29. Niemi, M.E., M.J. Daly, and A. Ganna, *The human genetic epidemiology of COVID-19.* Nature Reviews Genetics, 2022. **23**(9): p. 533-546.
30. van der Made, C.I., et al., *Clinical implications of host genetic variation and susceptibility to severe or critical COVID-19.* Genome Medicine, 2022. **14**(1): p. 1-22.
31. Zhang, X., et al., *Viral and host factors related to the clinical outcome of COVID-19.* Nature, 2020. **583**(7816): p. 437-440.
32. Nguyen, N.T., et al., *Male gender is a predictor of higher mortality in hospitalized adults with COVID-19.* PloS one, 2021. **16**(7): p. e0254066.
33. Ciarambino, T., O. Para, and M. Giordano, *Immune system and COVID-19 by sex differences and age.* Women's Health, 2021. **17**: p. 17455065211022262.
34. Ortolan, A., et al., *Does gender influence clinical expression and disease outcomes in COVID-19? A systematic review and meta-analysis.* International Journal of Infectious Diseases, 2020. **99**: p. 496-504.
35. Ji, X.-S., et al., *Human genetic basis of severe or critical illness in COVID-19.* Frontiers in cellular and infection microbiology, 2022. **12**: p. 963239.
36. Salihefendić, L., et al., *Identification of human genetic variants modulating the course of COVID-19 infection with importance in other viral infections.* Frontiers in genetics, 2023. **14**: p. 1240245.
37. Migliorini, F., et al., *Association between HLA genotypes and COVID-19 susceptibility, severity and progression: a comprehensive review of the literature.* European Journal of Medical Research, 2021. **26**(1): p. 1-9.
38. Kousathanas, A., et al., *Whole-genome sequencing reveals host factors underlying critical COVID-19.* Nature, 2022. **607**(7917): p. 97-103.
39. Basir, H.R.G., et al., *Susceptibility and Severity of COVID-19 Are Both Associated With Lower Overall Viral–Peptide Binding Repertoire of HLA Class I Molecules, Especially in Younger People.* Frontiers in Immunology, 2022. **13**: p. 891816.
40. Astbury, S., et al., *HLA-DR polymorphism in SARS-CoV-2 infection and susceptibility to symptomatic COVID-19.* Immunology, 2022. **166**(1): p. 68-77.
41. Langton, D.J., et al., *The influence of HLA genotype on the severity of COVID-19 infection.* Hla, 2021. **98**(1): p. 14-22.
42. Fricke-Galindo, I., et al., *IFNAR2 relevance in the clinical outcome of individuals with severe COVID-19.* Frontiers in Immunology, 2022. **13**: p. 949413.
43. Li, P., et al., *Targeted screening of genetic associations with COVID-19 susceptibility and severity.* Frontiers in Genetics, 2022. **13**: p. 1073880.
44. Bernardes, J.P., et al., *Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19.* Immunity, 2020. **53**(6): p. 1296-1314. e9.

45. Ulas, T., et al., *Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients.* 2020.
46. Schulte-Schrepping, J., et al., *Severe COVID-19 is marked by a dysregulated myeloid cell compartment.* 2020. **182**(6): p. 1419-1440. e23.
47. Pang, Z., et al., *Comprehensive Meta-Analysis of COVID-19 Global Metabolomics Datasets.* 2021. **11**(1): p. 44.
48. Beigel, J.H., K.M. Tomashek, and L.E. Dodd, *Remdesivir for the Treatment of Covid-19 - Preliminary Report. Reply.* N Engl J Med, 2020. **383**(10): p. 994.
49. Horby, P., et al., *Effect of dexamethasone in hospitalized patients with COVID-19: preliminary report. medRxiv 2020.* 2020. **22**.
50. Chow, J.H., et al., *Aspirin use is associated with decreased mechanical ventilation, ICU admission, and in-hospital mortality in hospitalized patients with COVID-19.* 2020.
51. Benucci, M., et al., *COVID-19 pneumonia treated with Sarilumab: A clinical series of eight patients.* 2020. **92**(11): p. 2368-2370.
52. Gupta, S., et al., *Association between early treatment with tocilizumab and mortality among critically ill patients with COVID-19.* 2020.
53. Zhang, Y., et al., *Immune evasive effects of SARS-CoV-2 variants to COVID-19 emergency used vaccines.* Frontiers in Immunology, 2021: p. 4842.
54. Murali, S., et al., *Effectiveness of the ChAdOx1 nCoV-19 Coronavirus Vaccine (CovishieldTM) in Preventing SARS-CoV2 Infection, Chennai, Tamil Nadu, India, 2021.* Vaccines, 2022. **10**(6): p. 970.
55. Kamal, D., et al., *Adverse events following ChAdOx1 nCoV-19 Vaccine (COVISHIELD) amongst health care workers: A prospective observational study.* Medical journal armed forces India, 2021. **77**: p. S283-S288.
56. Zare, H., et al., *Prevalence of COVID-19 vaccines (Sputnik V, AZD-1222, and Covaxin) side effects among healthcare workers in Birjand city, Iran.* International immunopharmacology, 2021. **101**: p. 108351.
57. Jones, I. and P. Roy, *Sputnik V COVID-19 vaccine candidate appears safe and effective.* The Lancet, 2021. **397**(10275): p. 642-643.
58. Jin, P., et al., *Safety and immunogenicity of heterologous boost immunization with an adenovirus type-5-vectored and protein-subunit-based COVID-19 vaccine (Convidecia/ZF2001): A randomized, observer-blinded, placebo-controlled trial.* PLoS Medicine, 2022. **19**(5): p. e1003953.
59. Gumusburun, R., et al., *CoronaVac/sinovac COVID-19 vaccine-related hypersensitivity reactions and second-dose vaccine administration: tertiary allergy center experience.* International Archives of Allergy Immunology

, 2022. **183**(7): p. 778-784.
60. Lazarus, R., et al., *Immunogenicity and safety of an inactivated whole-virus COVID-19 vaccine (VLA2001) compared with the adenoviral vector vaccine ChAdOx1-S in adults in the UK (COV-COMPARE): interim analysis of a randomised, controlled, phase 3, immunobridging trial.* The Lancet Infectious Diseases, 2022. **22**(12): p. 1716-1727.
61. Gao, L., et al., *Safety and immunogenicity of a protein subunit COVID-19 vaccine (ZF2001) in healthy children and adolescents aged 3–17 years in China: a randomised, double-blind, placebo-controlled, phase 1 trial and an open-label, non-randomised, non-inferiority, phase 2 trial.* The Lancet Child Adolescent Health, 2023.
62. Más-Bermejo, P.I., et al., *Cuban Abdala vaccine: Effectiveness in preventing severe disease and death from COVID-19 in Havana, Cuba; A cohort study.* The Lancet Regional Health-Americas, 2022. **16**: p. 100366.
63. Ryzhikov, A., et al., *A single blind, placebo-controlled randomized study of the safety, reactogenicity and immunogenicity of the "EpiVacCorona" Vaccine for the prevention of*

COVID-19, in volunteers aged 18-60 years (phase I-II). Инфекция и иммунитет, 2021. **11**(2): p. 283-296.
64. Khobragade, A., et al., *Efficacy, safety, and immunogenicity of the DNA SARS-CoV-2 vaccine (ZyCoV-D): the interim efficacy results of a phase 3, randomised, double-blind, placebo-controlled study in India.* The Lancet, 2022. **399**(10332): p. 1313-1321.
65. Aleem, A. and A.K. Slenker, *Monoclonal antibody therapy for high-risk coronavirus (COVID 19) patients with mild to moderate disease presentations.* Treasure Island (FL): StatPearls Publishing, 2023.
66. Hasin, Y., M. Seldin, and A. Lusis, *Multi-omics approaches to disease.* Genome biology, 2017. **18**(1): p. 1-15.
67. Subramanian, I., et al., *Multi-omics data integration, interpretation, and its application.* 2020. **14**: p. 1177932219899051.
68. Ovsyannikova, I.G., et al., *The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity.* 2020. **296**(1): p. 205-219.
69. Overmyer, K.A., et al., *Large-scale multi-omic analysis of COVID-19 severity.* Cell systems, 2020.
70. Aschenbrenner, A.C., et al., *Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients.* Genome medicine, 2021. **13**(1): p. 1-25.
71. Stukalov, A., et al., *Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV.* 2020.
72. Hou, J., et al., *Integrated multi-omics analyses identify anti-viral host factors and pathways controlling SARS-CoV-2 infection.* Nature Communications, 2024. **15**(1): p. 109.
73. Pinto, S.M., et al., *Multi-OMICs landscape of SARS-CoV-2-induced host responses in human lung epithelial cells.* Iscience, 2023. **26**(1).
74. Cen, X., et al., *Towards precision medicine: Omics approach for COVID-19.* Biosafety and Health, 2023. **5**(02): p. 78-88.
75. Song, J.-W., et al., *Omics-driven systems interrogation of metabolic dysregulation in COVID-19 pathogenesis.* 2020. **32**(2): p. 188-202. e5.
76. Barh, D., et al., *Multi-omics-based identification of SARS-CoV-2 infection biology and candidate drugs against COVID-19.* 2020. **126**: p. 104051.
77. Alqutami, F., A. Senok, and M. Hachim, *COVID-19 transcriptomic atlas: A comprehensive analysis of COVID-19 related transcriptomics datasets.* Frontiers in Genetics, 2021: p. 2500.
78. Jain, R., et al., *Host transcriptomic profiling of COVID-19 patients with mild, moderate, and severe clinical outcomes.* Computational structural biotechnology journal, 2021. **19**: p. 153-160.
79. Zhong, W., et al., *Next generation plasma proteome profiling of COVID-19 patients with mild to moderate symptoms.* EBioMedicine, 2021. **74**: p. 103723.
80. Suhre, K., et al., *Identification of robust protein associations with COVID-19 disease based on five clinical studies.* Frontiers in immunology, 2022. **12**: p. 5935.
81. Patel, H., et al., *Proteomic blood profiling in mild, severe and critical COVID-19 patients.* Scientific reports, 2021. **11**(1): p. 6357.
82. Overmyer, K.A., et al., *Large-scale multi-omic analysis of COVID-19 severity.* Cell systems, 2021. **12**(1): p. 23-40. e7.
83. Bohn, M.K., et al., *Pathophysiology of COVID-19: mechanisms underlying disease severity and progression.* Physiology, 2020. **35**(5): p. 288-301.
84. Menni, C., et al., *COVID-19 vaccine waning and effectiveness and side-effects of boosters: a prospective community study from the ZOE COVID Study.* The Lancet Infectious Diseases, 2022. **22**(7): p. 1002-1010.
85. Ferdinands, J.M., et al., *Waning of vaccine effectiveness against moderate and severe covid-19 among adults in the US from the VISION network: test negative, case-control study.* bmj, 2022. **379**.

86. Lau, J.J., et al., *Real-world COVID-19 vaccine effectiveness against the Omicron BA. 2 variant in a SARS-CoV-2 infection-naive population.* Nature Medicine, 2023: p. 1-1.
87. Darvishi, M., F. Rahimi, and A.T.B. Abadi, *SARS-CoV-2 Lambda (C. 37): An emerging variant of concern?* Gene Reports, 2021. **25**: p. 101378.
88. Guy, R.K., et al., *Rapid repurposing of drugs for COVID-19.* 2020. **368**(6493): p. 829-830.
89. Adams, J., et al., *The conundrum of low COVID-19 mortality burden in sub-Saharan Africa: myth or reality?* Global Health: Science and Practice, 2021. **9**(3): p. 433-443.
90. Karczewski, K.J. and M.P. Snyder, *Integrative omics for health and disease.* Nature Reviews Genetics, 2018. **19**(5): p. 299.
91. Sun, Y.V. and Y.-J. Hu, *Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases.* Advances in genetics, 2016. **93**: p. 147-190.
92. Horgan, R.P. and L.C. Kenny, *'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics.* The Obstetrician & Gynaecologist, 2011. **13**(3): p. 189-195.
93. Chakravorty, D., K. Banerjee, and S. Saha, *Integrative omics for interactomes*, in *Synthetic Biology*. 2018, Springer. p. 39-49.
94. Zapalska-Sozoniuk, M., et al., *Is it useful to use several "omics" for obtaining valuable results?* Molecular biology reports, 2019. **46**(3): p. 3597-3606.
95. Buescher, J.M. and E.M. Driggers, *Integration of omics: more than the sum of its parts.* Cancer & metabolism, 2016. **4**(1): p. 1-8.
96. Ritchie, M.D., et al., *Methods of integrating data to uncover genotype–phenotype interactions.* Nature Reviews Genetics, 2015. **16**(2): p. 85-97.
97. Bersanelli, M., et al., *Methods for the integration of multi-omics data: mathematical aspects.* BMC bioinformatics, 2016. **17**(2): p. 167-177.
98. Zeng, I.S.L. and T. Lumley, *Review of statistical learning methods in integrated omics studies (an integrated information science).* Bioinformatics and biology insights, 2018. **12**: p. 1177932218759292.
99. Wu, Z., et al., *Network-based methods for prediction of drug-target interactions.* Frontiers in pharmacology, 2018. **9**: p. 1134.
100. Agamah, F.E., et al., *Network-driven analysis of human–Plasmodium falciparum interactome: processes for malaria drug discovery and extracting in silico targets.* Malaria journal, 2021. **20**(1): p. 1-20.
101. Yan, J., et al., *Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data.* Briefings in bioinformatics, 2018. **19**(6): p. 1370-1381.
102. Camacho, D.M., et al., *Next-generation machine learning for biological networks.* Cell, 2018. **173**(7): p. 1581-1592.
103. Zitnik, M., et al., *Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities.* Information Fusion, 2019. **50**: p. 71-91.
104. Cavill, R., et al., *Transcriptomic and metabolomic data integration.* Briefings in bioinformatics, 2016. **17**(5): p. 891-901.
105. Duruflé, H., et al., *A powerful framework for an integrative study with heterogeneous omics data: from univariate statistics to multi-block analysis.* Briefings in Bioinformatics, 2021. **22**(3): p. bbaa166.
106. Holzinger, E.R. and M.D. Ritchie, *Integrating heterogeneous high-throughput data for meta-dimensional pharmacogenomics and disease-related studies.* Pharmacogenomics, 2012. **13**(2): p. 213-222.
107. Wen, Y., et al., *Multi-dimensional data integration algorithm based on random walk with restart.* BMC bioinformatics, 2021. **22**(1): p. 1-22.
108. Ulfenborg, B., *Vertical and horizontal integration of multi-omics data with miodin.* BMC bioinformatics, 2019. **20**(1): p. 1-10.

109. Reel, P.S., et al., *Using machine learning approaches for multi-omics data analysis: A review.* Biotechnology Advances, 2021. **49**: p. 107739.
110. Karasuyama, M. and H. Mamitsuka, *Adaptive edge weighting for graph-based learning algorithms.* Machine Learning, 2017. **106**(2): p. 307-335.
111. Nguyen, N.D. and D. Wang, *Multiview learning for understanding functional multiomics.* PLoS computational biology, 2020. **16**(4): p. e1007677.
112. Martorell-Marugán, J., et al., *Deep learning in omics data analysis and precision medicine.* Exon Publications, 2019: p. 37-53.
113. Kang, M., E. Ko, and T.B. Mersha, *A roadmap for multi-omics data integration using deep learning.* Briefings in Bioinformatics, 2022. **23**(1): p. bbab454.
114. Di Nanni, N., et al., *Network diffusion promotes the integrative analysis of multiple omics.* Frontiers in genetics, 2020. **11**: p. 106.
115. Cowen, L., et al., *Network propagation: a universal amplifier of genetic associations.* Nature Reviews Genetics, 2017. **18**(9): p. 551-562.
116. Griffin, P.J., et al., *Detection of multiple perturbations in multi-omics biological networks.* Biometrics, 2018. **74**(4): p. 1351-1361.
117. Hawe, J.S., F.J. Theis, and M. Heinig, *Inferring interaction networks from multi-omics data.* Frontiers in genetics, 2019. **10**: p. 535.
118. Koller, D. and N. Friedman, *Probabilistic graphical models: principles and techniques*. 2009: MIT press.
119. Friedman, N., *Inferring cellular networks using probabilistic graphical models.* Science, 2004. **303**(5659): p. 799-805.
120. Wein, S., et al., *A graph neural network framework for causal inference in brain networks.* Scientific reports, 2021. **11**(1): p. 1-18.
121. Badsha, M. and A.Q. Fu, *Learning causal biological networks with the principle of Mendelian randomization.* Frontiers in genetics, 2019. **10**: p. 460.
122. Peters, J., D. Janzing, and B. Schölkopf, *Elements of causal inference: foundations and learning algorithms*. 2017: The MIT Press.
123. Luo, Y., J. Peng, and J. Ma, *When causal inference meets deep learning.* Nature Machine Intelligence, 2020. **2**(8): p. 426-427.
124. Zheng, X., et al., *Dags with no tears: Continuous optimization for structure learning.* Advances in Neural Information Processing Systems, 2018.
125. Lachapelle, S., et al., *Gradient-based neural dag learning.* arXiv preprint arXiv:.02226, 2019.
126. Kotiang, S. and A. Eslami, *A probabilistic graphical model for system-wide analysis of gene regulatory networks.* Bioinformatics, 2020. **36**(10): p. 3192-3199.
127. González, I., et al., *Visualising associations between paired 'omics' data sets.* BioData mining, 2012. **5**(1): p. 1-23.
128. Wang, B., et al., *Similarity network fusion for aggregating data types on a genomic scale.* Nature methods, 2014. **11**(3): p. 333.
129. Bonnet, E., L. Calzone, and T. Michoel, *Integrative multi-omics module network inference with Lemon-Tree.* PLoS computational biology, 2015. **11**(2): p. e1003983.
130. Song, W.-M. and B. Zhang, *Multiscale embedded gene co-expression network analysis.* PLoS computational biology, 2015. **11**(11): p. e1004574.
131. Tuncbag, N., et al., *Network-based interpretation of diverse high-throughput datasets through the omics integrator software package.* PLoS computational biology, 2016. **12**(4): p. e1004879.
132. Xu, T., et al., *Identifying cancer subtypes from mirna-tf-mrna regulatory networks and expression data.* PloS one, 2016. **11**(4): p. e0152792.
133. Koh, H.W., et al., *iOmicsPASS: network-based integration of multiomics data for predictive subnetwork discovery.* NPJ systems biology and applications, 2019. **5**(1): p. 1-10.

134. Joshi, P., S. Jeong, and T. Park, *Sparse superlayered neural network-based multi-omics cancer subtype classification.* International Journal of Data Mining and Bioinformatics, 2020. **24**(1): p. 58-73.
135. Chierici, M., et al., *Integrative Network Fusion: a multi-omics approach in molecular profiling.* Frontiers in oncology, 2020. **10**: p. 1065.
136. Levi, H., R. Elkon, and R. Shamir, *DOMINO: a network-based active module identification algorithm with reduced rate of false calls.* Molecular Systems Biology, 2021. **17**(1): p. e9593.
137. Zachariou, M., et al., *Integrating multi-source information on a single network to detect disease-related clusters of molecular mechanisms.* Journal of proteomics, 2018. **188**: p. 15-29.
138. Pan, X., et al., *i-Modern: Integrated multi-omics network model identifies potential therapeutic targets in glioma by deep learning with interpretability.* Computational structural biotechnology journal, 2022. **20**: p. 3511-3521.
139. Zhou, G., et al., *OmicsNet 2.0: a web-based platform for multi-omics integration and network visual analytics.* Nucleic Acids Research, 2022.
140. Yang, Y., et al., *MDICC: novel method for multi-omics data integration and cancer subtype identification.* Briefings in Bioinformatics, 2022. **23**(3): p. bbac132.
141. Paull, E.O., et al., *Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE).* Bioinformatics, 2013. **29**(21): p. 2757-2764.
142. Dimitrakopoulos, C., et al., *Network-based integration of multi-omics data for prioritizing cancer genes.* Bioinformatics, 2018. **34**(14): p. 2441-2448.
143. Vandin, F., et al., *Discovery of mutated subnetworks associated with clinical data in cancer*, in *Biocomputing 2012*. 2012, World Scientific. p. 55-66.
144. Leiserson, M.D., et al., *Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes.* Nature genetics, 2015. **47**(2): p. 106-114.
145. Reyna, M.A., M.D. Leiserson, and B.J. Raphael, *Hierarchical HotNet: identifying hierarchies of altered subnetworks.* Bioinformatics, 2018. **34**(17): p. i972-i980.
146. Seifert, M. and A. Beyer, *regNet: An R package for network-based propagation of gene expression alterations.* Bioinformatics, 2018. **34**(2): p. 308-311.
147. Marín-Llaó, J., et al., *MultiPaths: a Python framework for analyzing multi-layer biological networks using diffusion algorithms.* Bioinformatics, 2020.
148. Shafi, A., et al., *A multi-cohort and multi-omics meta-analysis framework to identify network-based gene signatures.* Frontiers in genetics, 2019. **10**: p. 159.
149. Valdeolivas, A., et al., *Random walk with restart on multiplex and heterogeneous biological networks.* Bioinformatics, 2019. **35**(3): p. 497-505.
150. Bodein, A., et al., *A generic multivariate framework for the integration of microbiome longitudinal studies with other data types.* Frontiers in genetics, 2019. **10**: p. 963.
151. Bodein, A., et al., *Interpretation of network-based integration from multi-omics longitudinal data.* Nucleic acids research, 2020.
152. Ha, M.J., V. Baladandayuthapani, and K.-A. Do, *DINGO: differential network analysis in genomics.* Bioinformatics, 2015. **31**(21): p. 3413-3420.
153. Wang, Y., et al., *Permutation-based causal inference algorithms with interventions.* Advances in Neural Information Processing Systems, 2017.
154. Class, C.A., et al., *iDINGO—integrative differential network analysis in genomics with Shiny application.* Bioinformatics, 2018. **34**(7): p. 1243-1245.
155. Manatakis, D.V., V.K. Raghu, and P.V. Benos, *piMGM: incorporating multi-source priors in mixed graphical models for learning disease networks.* Bioinformatics, 2018. **34**(17): p. i848-i856.

156. Sedgewick, A.J., et al., *Mixed graphical models for integrative causal analysis with application to chronic lung disease diagnosis and prognosis.* Bioinformatics, 2019. **35**(7): p. 1204-1212.
157. Fan, Z., Y. Zhou, and H.W. Ressom, *MOTA: Network-Based Multi-Omic Data Integration for Biomarker Discovery.* Metabolites, 2020. **10**(4): p. 144.
158. Chen, Y.-X., et al., *An integrative multi-omics network-based approach identifies key regulators for breast cancer.* Computational and structural biotechnology journal, 2020. **18**: p. 2826-2835.
159. Badsha, M.B., E.A. Martin, and A.Q. Fu, *MRPC: An R package for inference of causal graphs.* Frontiers in Genetics, 2021. **12**.
160. Noecker, C., et al., *MIMOSA2: a metabolic network-based tool for inferring mechanism-supported relationships in microbiome-metabolome data.* Bioinformatics, 2022. **38**(6): p. 1615-1623.
161. Stukalov, A., et al., *Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV.* Nature, 2021. **594**(7862): p. 246-252.
162. Mo, X., et al., *Abnormal pulmonary function in COVID-19 patients at time of hospital discharge.* European Respiratory Journal, 2020. **55**(6).
163. Sargazi, S., et al., *The role of autophagy in controlling SARS-CoV-2 infection: An overview on virophagy-mediated molecular drug targets.* Cell biology international, 2021. **45**(8): p. 1599-1612.
164. Sun, C., et al., *Longitudinal multi-omics transition associated with fatality in critically ill COVID-19 patients.* Intensive care medicine experimental, 2021. **9**(1): p. 1-14.
165. Essa, M.M., et al., *Possible role of tryptophan and melatonin in COVID-19*. 2020, SAGE Publications Sage UK: London, England. p. 1178646920951832.
166. Shneider, A., A. Kudriavtsev, and A. Vakhrusheva, *Can melatonin reduce the severity of COVID-19 pandemic?* International reviews of immunology, 2020. **39**(4): p. 153-162.
167. Takeshita, H. and K. Yamamoto, *Tryptophan Metabolism and COVID-19-Induced Skeletal Muscle Damage: Is ACE2 a Key Regulator?* Frontiers in Nutrition, 2022. **9**.
168. Yan, W., et al., *Biological networks for cancer candidate biomarkers discovery.* Cancer informatics, 2016. **15**: p. CIN. S39458.
169. Tomazou, M., et al., *Multi-omics data integration and network-based analysis drives a multiplex drug repurposing approach to a shortlist of candidate drugs against COVID-19.* Briefings in bioinformatics, 2021.
170. Li, H., et al., *Serum Amyloid A is a biomarker of severe Coronavirus Disease and poor prognosis.* Journal of Infection, 2020. **80**(6): p. 646-655.
171. Smilowitz, N.R., et al., *C-reactive protein and clinical outcomes in patients with COVID-19.* European heart journal, 2021. **42**(23): p. 2270-2279.
172. Demichev, V., et al., *A time-resolved proteomic and prognostic map of COVID-19.* Cell systems, 2021. **12**(8): p. 780-794. e7.
173. Birnhuber, A., et al., *Between inflammation and thrombosis: endothelial cells in COVID-19.* European Respiratory Journal, 2021. **58**(3).
174. Luo, Y., et al., *A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information.* Nature communications, 2017. **8**(1): p. 1-13.
175. Wang, W., et al., *Drug repositioning by integrating target information through a heterogeneous network model.* Bioinformatics, 2014. **30**(20): p. 2923-2930.
176. Vitali, F., et al., *A network-based data integration approach to support drug repurposing and multi-target therapies in triple negative breast cancer.* PloS one, 2016. **11**(9): p. e0162407.
177. Huang, L., et al., *Driver network as a biomarker: systematic integration and network modeling of multi-omics data to derive driver signaling pathways for drug combination prediction.* Bioinformatics, 2019. **35**(19): p. 3709-3717.

178. Chen, J. and K.-C. Wong, *RNCE: network integration with reciprocal neighbors contextual encoding for multi-modal drug community study on cancer targets.* Briefings in Bioinformatics, 2021. **22**(3): p. bbaa118.
179. Krassowski, M., et al., *State of the field in multi-omics research: From computational needs to data mining and sharing.* Frontiers in Genetics, 2020. **11**.
180. Canzler, S., et al., *Prospects and challenges of multi-omics data integration in toxicology.* Archives of toxicology, 2020. **94**(2): p. 371-388.
181. Lee, D., Y. Park, and S. Kim, *Towards multi-omics characterization of tumor heterogeneity: a comprehensive review of statistical and machine learning approaches.* Briefings in bioinformatics, 2021. **22**(3): p. bbaa188.
182. Lima, E., et al., *Variable selection for inferential models with relatively high-dimensional data: Between method heterogeneity and covariate stability as adjuncts to robust selection.* Scientific reports, 2020. **10**(1): p. 1-11.
183. Jung, G.T., K.-P. Kim, and K. Kim, *How to interpret and integrate multi-omics data at systems level.* Animal Cells and Systems, 2020. **24**(1): p. 1-7.
184. Guo, M.G., D.N. Sosa, and R.B. Altman, *Challenges and opportunities in network-based solutions for biological questions.* Briefings in Bioinformatics, 2022. **23**(1): p. bbab437.
185. Davies, V., et al., *Rapid development of improved data-dependent acquisition strategies.* Analytical chemistry, 2021. **93**(14): p. 5676-5683.
186. Guo, J. and T. Huan, *Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography–mass spectrometry based untargeted metabolomics.* Analytical Chemistry, 2020. **92**(12): p. 8072-8080.
187. Sun, F., et al., *An integrated data-dependent and data-independent acquisition method for hazardous compounds screening in foods using a single UHPLC-Q-Orbitrap run.* Journal of Hazardous Materials, 2021. **401**: p. 123266.
188. Folch-Fortuny, A., et al., *Enabling network inference methods to handle missing data and outliers.* BMC bioinformatics, 2015. **16**(1): p. 1-12.
189. Li, W. *Estimating Jaccard index with missing observations: a matrix calibration approach*. in *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2*. 2015.
190. Sitaram, D., et al. *A measure of similarity of time series containing missing data using the mahalanobis distance*. in *2015 second international conference on advances in computing and communication engineering*. 2015. IEEE.
191. Greenland, S., M.A. Mansournia, and D.G. Altman, *Sparse data bias: a problem hiding in plain sight.* bmj, 2016. **352**.
192. Ng, A., *Sparse autoencoder.* CS294A Lecture notes, 2011. **72**(2011): p. 1-19.
193. Pereira, R.C., et al., *Reviewing autoencoders for missing data imputation: Technical trends, applications and outcomes.* Journal of Artificial Intelligence Research, 2020. **69**: p. 1255-1285.
194. Singhal, T., *A review of coronavirus disease-2019 (COVID-19).* The indian journal of pediatrics, 2020. **87**(4): p. 281-286.
195. Messner, C.B., et al., *Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection.* Cell systems, 2020. **11**(1): p. 11-24. e4.
196. Blanco-Melo, D., et al., *Imbalanced host response to SARS-CoV-2 drives development of COVID-19.* 2020. **181**(5): p. 1036-1045. e9.
197. Sun, J.T., et al., *Lipid profile features and their associations with disease severity and mortality in patients with COVID-19.* Frontiers in cardiovascular medicine, 2020. **7**: p. 584987.
198. Daamen, A.R., et al., *Comprehensive transcriptomic analysis of COVID-19 blood, lung, and airway.* Scientific Reports, 2021. **11**(1): p. 1-19.

199. Roberts, I., et al., *Untargeted metabolomics of COVID-19 patient serum reveals potential prognostic markers of both severity and outcome.* 2022. **18**(1): p. 6.
200. Páez-Franco, J.C., et al., *Metabolomics analysis identifies glutamic acid and cystine imbalances in COVID-19 patients without comorbid conditions. Implications on redox homeostasis and COVID-19 pathophysiology.* Plos one, 2022. **17**(9): p. e0274910.
201. Jia, H., et al., *Metabolomic analyses reveal new stage-specific features of COVID-19.* European Respiratory Journal, 2022. **59**(2).
202. Fraser, D.D., et al., *Metabolomics profiling of critically ill coronavirus disease 2019 patients: identification of diagnostic and prognostic biomarkers.* Critical Care Explorations, 2020. **2**(10).
203. Ciccarelli, M., et al., *Untargeted lipidomics reveals specific lipid profiles in COVID-19 patients with different severity from Campania region (Italy).* Journal of pharmaceutical biomedical analysis, 2022. **217**: p. 114827.
204. Stephenson, E., et al., *Single-cell multi-omics analysis of the immune response in COVID-19.* Nature medicine, 2021. **27**(5): p. 904-916.
205. Chattopadhyay, P., et al., *Single-cell multiomics revealed the dynamics of antigen presentation, immune response and T cell activation in the COVID-19 positive and recovered individuals.* Frontiers in Immunology, 2022. **13**: p. 7319.
206. Suvarna, K., et al., *A multi-omics longitudinal study reveals alteration of the leukocyte activation pathway in COVID-19 patients.* Journal of Proteome Research, 2021. **20**(10): p. 4667-4680.
207. Gygi, J.P., et al., *Integrated longitudinal multi-omics study identifies immune programs associated with COVID-19 severity and mortality in 1152 hospitalized participants.* bioRxiv, 2023: p. 2023.11. 03.565292.
208. Agamah, F.E., et al., *Computational approaches for network-based integrative multi-omics analysis.* Frontiers in Molecular Biosciences, 2022: p. 1214.
209. Adossa, N., et al., *Computational strategies for single-cell multi-omics integration.* Computational and Structural Biotechnology Journal, 2021. **19**: p. 2588-2596.
210. Piñero, J., et al., *The DisGeNET knowledge platform for disease genomics: 2019 update.* Nucleic acids research, 2020. **48**(D1): p. D845-D855.
211. Franz, M., et al., *GeneMANIA update 2018.* Nucleic acids research, 2018. **46**(W1): p. W60-W64.
212. Domingo-Fernández, D., et al., *COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology.* Bioinformatics, 2021. **37**(9): p. 1332-1334.
213. Cheadle, C., et al., *Analysis of microarray data using Z score transformation.* The Journal of molecular diagnostics, 2003. **5**(2): p. 73-81.
214. Baptista, A., A. Gonzalez, and A. Baudot, *Universal multilayer network exploration by random walk with restart.* Communications Physics, 2022. **5**(1): p. 1-9.
215. Arazi, A., et al. *Human systems immunology: hypothesis-based modeling and unbiased data-driven approaches*. in *Seminars in immunology*. 2013. Elsevier.
216. Eriksson, O., et al., *Combining hypothesis-and data-driven neuroscience modeling in FAIR workflows.* Elife, 2022. **11**: p. e69013.
217. Valente, T.W., et al., *How correlated are network centrality measures?* Connections (Toronto, Ont.), 2008. **28**(1): p. 16.
218. Pang, Z., et al., *MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights.* Nucleic acids research, 2021. **49**(W1): p. W388-W396.
219. Acevedo, A., et al., *LIPEA: lipid pathway enrichment analysis.* BioRxiv, 2018: p. 274969.
220. Kuleshov, M.V., et al., *Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.* Nucleic acids research, 2016. **44**(W1): p. W90-W97.

221. Lucas, C., et al., *Longitudinal analyses reveal immunological misfiring in severe COVID-19.* 2020. **584**(7821): p. 463-469.
222. Rincon-Arevalo, H., et al., *Altered increase in STAT1 expression and phosphorylation in severe COVID-19.* European Journal of Immunology, 2022. **52**(1): p. 138-148.
223. Žarković, N., et al., *The Impact of Severe COVID-19 on Plasma Antioxidants.* Molecules, 2022. **27**(16): p. 5323.
224. Zarkovic, N., et al., *Post-mortem findings of inflammatory cells and the association of 4-hydroxynonenal with systemic vascular and oxidative stress in lethal COVID-19.* Cells, 2022. **11**(3): p. 444.
225. Kaye, A.G. and R. Siegel, *The efficacy of IL-6 inhibitor Tocilizumab in reducing severe COVID-19 mortality: a systematic review.* PeerJ, 2020. **8**: p. e10322.
226. Samaee, H., et al., *Tocilizumab for treatment patients with COVID-19: recommended medication for novel disease.* International immunopharmacology, 2020. **89**: p. 107018.
227. Hadjadj, J., et al., *Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients.* Science, 2020. **369**(6504): p. 718-724.
228. Xiao, K., et al., *Mesenchymal stem cells: current clinical progress in ARDS and COVID-19.* Stem Cell Research Therapy, 2020. **11**(1): p. 1-7.
229. Wang, W., et al., *Therapeutic mechanisms of mesenchymal stem cells in acute respiratory distress syndrome reveal potentials for Covid-19 treatment.* Journal of Translational Medicine, 2021. **19**(1): p. 1-13.
230. Quartuccio, L., et al., *Interleukin 6, soluble interleukin 2 receptor alpha (CD25), monocyte colony-stimulating factor, and hepatocyte growth factor linked with systemic hyperinflammation, innate immunity hyperactivation, and organ damage in COVID-19 pneumonia.* Cytokine, 2021. **140**: p. 155438.
231. Salomão, R., et al., *Involvement of matrix metalloproteinases in COVID-19: molecular targets, mechanisms, and insights for therapeutic interventions.* Biology, 2023. **12**(6): p. 843.
232. Perreau, M., et al., *The cytokines HGF and CXCL13 predict the severity and the mortality in COVID-19 patients.* Nature Communications, 2021. **12**(1): p. 4888.
233. Krishnan, S., et al., *Metabolic perturbation associated with COVID-19 disease severity and SARS-CoV-2 replication.* Molecular & Cellular Proteomics, 2021. **20**.
234. Cihan, M., et al., *Kynurenine pathway in Coronavirus disease (COVID-19): potential role in prognosis.* Journal of clinical laboratory analysis, 2022. **36**(3): p. e24257.
235. Takeshita, H. and K. Yamamoto, *Tryptophan metabolism and COVID-19-induced skeletal muscle damage: is ACE2 a key regulator?* Frontiers in Nutrition, 2022. **9**: p. 868845.
236. Thomas, T., et al., *COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status.* JCI insight, 2020. **5**(14).
237. Shen, B., et al., *Proteomic and metabolomic characterization of COVID-19 patient sera.* Cell, 2020. **182**(1): p. 59-72. e15.
238. Shirato, K. and T. Kizaki, *SARS-CoV-2 spike protein S1 subunit induces pro-inflammatory responses via toll-like receptor 4 signaling in murine and human macrophages.* Heliyon, 2021. **7**(2): p. e06187.
239. Del Valle, D.M., et al., *An inflammatory cytokine signature predicts COVID-19 severity and survival.* Nature medicine, 2020. **26**(10): p. 1636-1643.
240. Zayet, S., et al., *Clinical features of COVID-19 and influenza: a comparative study on Nord Franche-Comte cluster.* Microbes infection, 2020. **22**(9): p. 481-488.
241. Liao, M., et al., *Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19.* Nature medicine, 2020. **26**(6): p. 842-844.
242. Weiner, J., et al., *Increased risk of severe clinical course of COVID-19 in carriers of HLA-C* 04: 01.* EClinicalMedicine, 2021. **40**.

243. Martínez-Fleta, P., et al., *A differential signature of circulating miRNAs and cytokines between COVID-19 and community-acquired pneumonia uncovers novel physiopathological mechanisms of COVID-19.* Frontiers in immunology, 2022. **12**: p. 815651.
244. Robinson, P.C., et al., *Accumulating evidence suggests anti-TNF therapy needs to be given trial priority in COVID-19 treatment.* The Lancet Rheumatology, 2020. **2**(11): p. e653-e655.
245. Coperchini, F., et al., *The cytokine storm in COVID-19: An overview of the involvement of the chemokine/chemokine-receptor system.* Cytokine growth factor reviews, 2020. **53**: p. 25-32.
246. Li, X., et al., *Molecular immune pathogenesis and diagnosis of COVID-19.* Journal of pharmaceutical analysis, 2020. **10**(2): p. 102-108.
247. Madhurantakam, S., et al., *Multiplex sensing of IL-10 and CRP towards predicting critical illness in COVID-19 infections.* Biosensors Bioelectronics: X, 2023. **13**: p. 100307.
248. Theken, K.N., et al., *The roles of lipids in SARS-CoV-2 viral replication and the host immune response.* Journal of lipid research, 2021. **62**.
249. Hammoudeh, S.M., et al., *Systems immunology analysis reveals the contribution of pulmonary and extrapulmonary tissues to the immunopathogenesis of severe COVID-19 patients.* Frontiers in Immunology, 2021. **12**: p. 595150.
250. Sheshan, V., et al., *To Correlate Serum Lipid Parameters with Clinical Outcome in COVID-19 Patients.* SAS Journal of Medicine, 2021. **7**(7): p. 295-302.
251. Dai, W., et al., *Hypertriglyceridemia during hospitalization independently associates with mortality in patients with COVID-19.* Journal of Clinical Lipidology, 2021. **15**(5): p. 724-731.
252. Lawler, N.G., et al., *Systemic perturbations in amine and kynurenine metabolism associated with acute SARS-CoV-2 infection and inflammatory cytokine responses.* Journal of Proteome Research, 2021. **20**(5): p. 2796-2811.
253. Toro, A.P., et al., *Vitamin C and COVID-19: An orthomolecular perspective on physiological mechanisms.* Orthomol Med, 2021. **36**(3): p. 1-9.
254. Veras, F.P., et al., *SARS-CoV-2–triggered neutrophil extracellular traps mediate COVID-19 pathology.* Journal of Experimental Medicine, 2020. **217**(12).
255. Risitano, A.M., et al., *Complement as a target in COVID-19?* Nature Reviews Immunology, 2020. **20**(6): p. 343-344.
256. Carvelli, J., et al., *Association of COVID-19 inflammation with activation of the C5a–C5aR1 axis.* Nature, 2020. **588**(7836): p. 146-150.
257. Bouhaddou, M., et al., *The global phosphorylation landscape of SARS-CoV-2 infection.* Cell, 2020. **182**(3): p. 685-712. e19.
258. Law, H.K., et al., *Chemokine up-regulation in SARS-coronavirus–infected, monocyte-derived human dendritic cells.* Blood, 2005. **106**(7): p. 2366-2374.
259. Merad, M. and J.C. Martin, *Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages.* Nature reviews immunology, 2020. **20**(6): p. 355-362.
260. Zhou, Y., et al., *Pathogenic T-cells and inflammatory monocytes incite inflammatory storms in severe COVID-19 patients.* National Science Review, 2020. **7**(6): p. 998-1002.
261. Moore, J.B. and C.H. June, *Cytokine release syndrome in severe COVID-19.* Science, 2020. **368**(6490): p. 473-474.
262. André, S., et al., *T cell apoptosis characterizes severe Covid-19 disease.* Cell Death Differentiation, 2022. **29**(8): p. 1486-1499.
263. Cizmecioglu, A., et al., *Apoptosis-induced T-cell lymphopenia is related to COVID-19 severity.* Journal of medical virology, 2021. **93**(5): p. 2867-2874.
264. Song, J.-W., et al., *Omics-driven systems interrogation of metabolic dysregulation in COVID-19 pathogenesis.* Cell metabolism, 2020. **32**(2): p. 188-202. e5.
265. Barberis, E., et al., *Large-scale plasma analysis revealed new mechanisms and molecules associated with the host response to SARS-CoV-2.* International journal of molecular sciences, 2020. **21**(22): p. 8623.

266. Masoodi, M., et al., *Disturbed lipid and amino acid metabolisms in COVID-19 patients.* Journal of molecular medicine, 2022. **100**(4): p. 555-568.
267. Khan, S.A., K.F. Goliwas, and J.S. Deshane, *Sphingolipids in lung pathology in the coronavirus disease era: A review of sphingolipid involvement in the pathogenesis of lung damage.* Frontiers in Physiology, 2021: p. 1757.
268. García-Pérez, B.E., et al., *Taming the Autophagy as a Strategy for Treating COVID-19.* Cells, 2020. **9**(12): p. 2679.
269. Yang, N. and H.-M. Shen, *Targeting the endocytic pathway and autophagy process as a novel therapeutic strategy in COVID-19.* International journal of biological sciences, 2020. **16**(10): p. 1724.
270. Oliveira, L.B., et al., *Metabolomic profiling of plasma reveals differential disease severity markers in COVID-19 patients.* Frontiers in microbiology, 2022. **13**: p. 1116.
271. Li, Z.-B., et al., *Novel potential metabolic biomarker panel for early detection of severe COVID-19 using full-spectrum metabolome and whole-transcriptome analyses.* Signal transduction targeted therapy, 2022. **7**(1): p. 129.
272. van Eijk, L.E., et al., *The Disease-Modifying Role of Taurine and Its Therapeutic Potential in Coronavirus Disease 2019 (COVID-19)*, in *Taurine 12: A Conditionally Essential Amino Acid*. 2022, Springer. p. 3-21.
273. Eteraf-Oskouei, T. and M. Najafi, *The relationship between the serotonergic system and COVID-19 disease: A review.* Heliyon, 2022: p. e09544.
274. Guirao, J.J., et al., *High serum IL-6 values increase the risk of mortality and the severity of pneumonia in patients diagnosed with COVID-19.* Molecular immunology, 2020. **128**: p. 64-68.
275. Gou, X., et al., *IL-6 during influenza-streptococcus pneumoniae co-infected pneumonia—a protector.* Frontiers in immunology, 2020. **10**: p. 3102.
276. Menyhárt, O. and B. Győrffy, *Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis.* Computational and structural biotechnology journal, 2021. **19**: p. 949-960.
277. Lau, J.J., et al., *Real-world COVID-19 vaccine effectiveness against the Omicron BA. 2 variant in a SARS-CoV-2 infection-naive population.* Nature medicine, 2023. **29**(2): p. 348-357.
278. Zhang, Y., et al., *Immune evasive effects of SARS-CoV-2 variants to COVID-19 emergency used vaccines.* Frontiers in Immunology, 2021. **12**: p. 771242.
279. Telenti, A., et al., *After the pandemic: perspectives on the future trajectory of COVID-19.* Nature, 2021. **596**(7873): p. 495-504.
280. Guy, R.K., et al., *Rapid repurposing of drugs for COVID-19.* Science, 2020. **368**(6493): p. 829-830.
281. Hammond, J., et al., *Oral nirmatrelvir for high-risk, nonhospitalized adults with Covid-19.* New England Journal of Medicine, 2022. **386**(15): p. 1397-1408.
282. Beigel, J.H., et al., *Remdesivir for the treatment of Covid-19.* New England Journal of Medicine, 2020. **383**(19): p. 1813-1826.
283. Diaz, G.A., et al., *Remdesivir and mortality in patients with coronavirus disease 2019.* Clinical Infectious Diseases, 2022. **74**(10): p. 1812-1820.
284. Consortium, W.S.T., *Repurposed antiviral drugs for Covid-19—interim WHO solidarity trial results.* New England journal of medicine, 2021. **384**(6): p. 497-511.
285. Dryden-Peterson, S., et al., *Nirmatrelvir plus ritonavir for early COVID-19 in a large US health system: a population-based cohort study.* Annals of internal medicine, 2023. **176**(1): p. 77-84.
286. Horby, P., et al., *Effect of dexamethasone in hospitalized patients with COVID-19–preliminary report.* MedRxiv, 2020: p. 2020.06. 22.20137273.
287. Group, R.C., *Dexamethasone in hospitalized patients with Covid-19.* New England Journal of Medicine, 2021. **384**(8): p. 693-704.

288. Ma, S., et al., *Does aspirin have an effect on risk of death in patients with COVID-19? A meta-analysis.* European Journal of Clinical Pharmacology, 2022. **78**(9): p. 1403-1420.
289. Group, R.C., *Aspirin in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial.* Lancet (London, England), 2022. **399**(10320): p. 143.
290. Chow, J.H., et al., *Association of early aspirin use with in-hospital mortality in patients with moderate COVID-19.* JAMA network open, 2022. **5**(3): p. e223890-e223890.
291. Liew, M.N.Y., et al., *SARS-CoV-2 neutralizing antibody bebtelovimab–a systematic scoping review and meta-analysis.* Frontiers in Immunology, 2023. **14**.
292. Sivapalasingam, S., et al., *Efficacy and safety of sarilumab in hospitalized patients with coronavirus disease 2019: a randomized clinical trial.* Clinical Infectious Diseases, 2022. **75**(1): p. e380-e388.
293. Benucci, M., et al., *COVID-19 pneumonia treated with sarilumab: a clinical series of eight patients.* Journal of medical virology, 2020. **92**(11): p. 2368.
294. Mushtaq, M.Z., et al., *Tocilizumab in critically ill COVID-19 patients: An observational study.* International Immunopharmacology, 2022. **102**: p. 108384.
295. Gupta, S., et al., *Association between early treatment with tocilizumab and mortality among critically ill patients with COVID-19.* JAMA internal medicine, 2021. **181**(1): p. 41-51.
296. Heskin, J., et al., *Caution required with use of ritonavir-boosted PF-07321332 in COVID-19 management.* The Lancet, 2022. **399**(10319): p. 21-22.
297. Nevalainen, O.P., et al., *Effect of remdesivir post hospitalization for COVID-19 infection from the randomized SOLIDARITY Finland trial.* Nature Communications, 2022. **13**(1): p. 6152.
298. Kocks, J., et al., *A potential harmful effect of dexamethasone in non-severe COVID-19: results from the COPPER-pilot study.* ERJ Open Research, 2022. **8**(2).
299. Ioannidis, V.N., et al., *Drkg-drug repurposing knowledge graph for covid-19.* arXiv preprint arXiv:2010.09600, 2020.
300. Mahdian, S., A. Ebrahim-Habibi, and M. Zarrabi, *Drug repurposing using computational methods to identify therapeutic options for COVID-19.* Journal of Diabetes & Metabolic Disorders, 2020. **19**: p. 691-699.
301. Tomazou, M., et al., *Multi-omics data integration and network-based analysis drives a multiplex drug repurposing approach to a shortlist of candidate drugs against COVID-19.* Briefings in bioinformatics, 2021. **22**(6): p. bbab114.
302. Hsieh, K., et al., *Drug repurposing for COVID-19 using graph neural network and harmonizing multiple evidence.* Scientific reports, 2021. **11**(1): p. 23179.
303. Zhou, Y., et al., *Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2.* Cell discovery, 2020. **6**(1): p. 14.
304. Panda, S., et al., *Computational approaches for drug repositioning and repurposing to combat SARS-CoV-2 infection*, in *Computational Approaches for Novel Therapeutic and Diagnostic Designing to Mitigate SARS-CoV2 Infection*. 2022, Elsevier. p. 247-265.
305. Wang, J., *Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study.* Journal of chemical information and modeling, 2020. **60**(6): p. 3277-3286.
306. Sosa, D.N., et al. *A literature-based knowledge graph embedding method for identifying drug repurposing opportunities in rare diseases*. in *Pacific Symposium on Biocomputing 2020*. 2019. World Scientific.
307. Szklarczyk, D., et al., *The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets.* Nucleic acids research, 2021. **49**(D1): p. D605-D612.
308. Del Toro, N., et al., *The IntAct database: efficient access to fine-grained molecular interaction data.* Nucleic acids research, 2022. **50**(D1): p. D648-D653.

309. Himmelstein, D.S., et al., *Systematic integration of biomedical knowledge prioritizes drugs for repurposing.* Elife, 2017. **6**: p. e26726.
310. Wishart, D.S., et al., *DrugBank 5.0: a major update to the DrugBank database for 2018.* Nucleic acids research, 2018. **46**(D1): p. D1074-D1082.
311. Freshour, S.L., et al., *Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts.* Nucleic acids research, 2021. **49**(D1): p. D1144-D1151.
312. Davis, A.P., et al., *Comparative toxicogenomics database (CTD): update 2021.* Nucleic acids research, 2021. **49**(D1): p. D1138-D1143.
313. Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.* Nature, 2020. **583**(7816): p. 459-468.
314. Su, Y., et al., *Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19.* Cell, 2020. **183**(6): p. 1479-1495. e20.
315. Agamah, F.E., et al., *Network-based integrative multi-omics approach reveals biosignatures specific to COVID-19 disease phases.* bioRxiv, 2023: p. 2023.09. 29.560110.
316. *UniProt: the universal protein knowledgebase in 2021.* Nucleic acids research, 2021. **49**(D1): p. D480-D489.
317. Hinkson, C.R., *COVID-19 Treatment Guidelines.* National Institutes of health, 2022.
318. Bhimraj, A., et al., *Infectious Diseases Society of America Guidelines on the treatment and management of patients with coronavirus disease 2019 (COVID-19).* Clinical Infectious Diseases, 2020: p. ciaa478.
319. Zhou, F., et al., *Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study.* The lancet, 2020. **395**(10229): p. 1054-1062.
320. Chen, L.Y., et al., *Confronting the controversy: interleukin-6 and the COVID-19 cytokine storm syndrome*. 2020, Eur Respiratory Soc.
321. Coomes, E.A. and H. Haghbayan, *Interleukin-6 in COVID-19: a systematic review and meta-analysis.* Reviews in medical virology, 2020. **30**(6): p. 1-9.
322. Lv, B., et al. *Efficient processing node proximity via random walk with restart*. in *Web Technologies and Applications: 16th Asia-Pacific Web Conference, APWeb 2014, Changsha, China, September 5-7, 2014. Proceedings 16*. 2014. Springer.
323. Jung, J., et al., *Random walk with restart on large graphs using block elimination.* ACM Transactions on Database Systems (TODS), 2016. **41**(2): p. 1-43.
324. Group, R.C., *Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial.* Lancet (London, England), 2021. **397**(10285): p. 1637.
325. Bengtson, C.D., et al., *An open label trial to assess safety of losartan for treating worsening respiratory illness in COVID-19.* Frontiers in medicine, 2021. **8**: p. 630209.
326. Puskarich, M.A., et al., *Efficacy of losartan in hospitalized patients with COVID-19–induced lung injury: a randomized clinical trial.* JAMA Network Open, 2022. **5**(3): p. e222735-e222735.
327. Nguyen, L.C., et al., *Cannabidiol inhibits SARS-CoV-2 replication through induction of the host ER stress and innate immune responses.* Science Advances, 2022. **8**(8): p. eabi6110.
328. Sajid Jamal, Q.M., A.H. Alharbi, and V. Ahmad, *Identification of doxorubicin as a potential therapeutic against SARS-CoV-2 (COVID-19) protease: a molecular docking and dynamics simulation studies.* Journal of Biomolecular Structure and Dynamics, 2022. **40**(17): p. 7960-7974.
329. Singh, M.B., et al., *A comparative study of 5-fluorouracil, doxorubicin, methotrexate, paclitaxel for their inhibition ability for Mpro of nCoV: Molecular docking and molecular dynamics simulations.* Journal of the Indian Chemical Society, 2022. **99**(12): p. 100790.
330. Ravichandran, R., et al., *An open label randomized clinical trial of Indomethacin for mild and moderate hospitalised Covid-19 patients.* Scientific reports, 2022. **12**(1): p. 6413.

331. Ben Abdallah, S., et al., *Twice-daily oral zinc in the treatment of patients with coronavirus disease 2019: a randomized double-blind controlled trial.* Clinical Infectious Diseases, 2023. **76**(2): p. 185-191.
332. Patocka, J., et al., *Rapamycin: drug repurposing in SARS-CoV-2 infection.* Pharmaceuticals, 2021. **14**(3): p. 217.
333. Karsulovic, C., et al., *mTORC inhibitor Sirolimus deprograms monocytes in "cytokine storm" in SARS-CoV2 secondary hemophagocytic lymphohistiocytosis-like syndrome.* Clinical Immunology (Orlando, Fla.), 2020. **218**: p. 108539.
334. Chowdhury, P. and P. Pathak, *Neuroprotective immunity by essential nutrient "Choline" for the prevention of SARS CoV2 infections: An in silico study by molecular dynamics approach.* Chemical physics letters, 2020. **761**: p. 138057.
335. Schwartz, E., *Does ivermectin have a place in the treatment of mild Covid-19?* New microbes and new infections, 2022. **46**.
336. El Bairi, K., et al., *Repurposing anticancer drugs for the management of COVID-19.* European Journal of Cancer, 2020. **141**: p. 40-61.
337. Sperry, M.M., et al., *Target-agnostic drug prediction integrated with medical record analysis uncovers differential associations of statins with increased survival in COVID-19 patients.* PLOS Computational Biology, 2023. **19**(5): p. e1011050.
338. Mirjalili, M., et al., *Does Losartan reduce the severity of COVID-19 in hypertensive patients?* BMC Cardiovascular Disorders, 2022. **22**(1): p. 116.
339. Xu, Y., et al., *Ribavirin treatment for critically ill COVID-19 patients: An observational study.* Infection and Drug Resistance, 2021: p. 5287-5291.
340. Khalili, J.S., et al., *Novel coronavirus treatment with ribavirin: groundwork for an evaluation concerning COVID-19.* Journal of medical virology, 2020. **92**(7): p. 740-746.
341. Patel, J., et al., *Azithromycin for mild-to-moderate COVID-19.* The Lancet Respiratory Medicine, 2021. **9**(10): p. e99.
342. Li, G., et al., *Tenofovir disoproxil fumarate and coronavirus disease 2019 outcomes in men with HIV.* Aids, 2022. **36**(12): p. 1689-1696.
343. Kim, J.-W., et al., *Serious Clinical Outcomes of COVID-19 Related to Acetaminophen or NSAIDs from a Nationwide Population-Based Cohort Study.* International Journal of Environmental Research and Public Health, 2023. **20**(5): p. 3832.
344. Manjani, L., et al., *Effects of acetaminophen on outcomes in patients hospitalized with COVID-19.* Chest, 2021. **160**(4): p. A1072.
345. Sterne, J.A., et al., *Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: a meta-analysis.* Jama, 2020. **324**(13): p. 1330-1341.
346. Pathania, Y.S., *Cyclosporine: hope for severe COVID-19?* BMJ Supportive & Palliative Care, 2021.
347. Fenizia, C., et al., *Cyclosporine a inhibits viral infection and release as well as cytokine production in lung cells by three SARS-CoV-2 variants.* Microbiology Spectrum, 2022. **10**(1): p. e01504-21.
348. Belli, L.S., et al., *Protective role of tacrolimus, deleterious role of age and comorbidities in liver transplant recipients with Covid-19: results from the ELITA/ELTR multi-center European study.* Gastroenterology, 2021. **160**(4): p. 1151-1163. e3.
349. Organization, W.H., *The use of non-steroidal anti-inflammatory drugs (NSAIDs) in patients with COVID-19.* 2020.
350. Jones, S.A. and C.A. Hunter, *Is IL-6 a key cytokine target for therapy in COVID-19?* Nature Reviews Immunology, 2021. **21**(6): p. 337-339.
351. Sajgure, A., et al., *Safety and efficacy of mycophenolate in COVID-19: A nonrandomised prospective study in western India.* The Lancet Regional Health-Southeast Asia, 2023. **11**.

352. Bramante, C.T., et al., *Outpatient treatment of COVID-19 and incidence of post-COVID-19 condition over 10 months (COVID-OUT): a multicentre, randomised, quadruple-blind, parallel-group, phase 3 trial.* The Lancet Infectious Diseases, 2023. **23**(10): p. 1119-1129.
353. Giamarellos-Bourboulis, E.J., et al., *Complex immune dysregulation in COVID-19 patients with severe respiratory failure.* Cell host & microbe, 2020. **27**(6): p. 992-1000. e3.
354. Lucas, C., et al., *Longitudinal analyses reveal immunological misfiring in severe COVID-19.* Nature, 2020. **584**(7821): p. 463-469.
355. Xu, Z., et al., *Pathological findings of COVID-19 associated with acute respiratory distress syndrome.* The Lancet respiratory medicine, 2020. **8**(4): p. 420-422.
356. Malone, R.W., et al., *COVID-19: famotidine, histamine, mast cells, and mechanisms.* Frontiers in Pharmacology, 2021. **12**: p. 633680.
357. Ennis, M. and K. Tiligada, *Histamine receptors and COVID-19.* Inflammation Research, 2021. **70**: p. 67-75.
358. Thangam, E.B., et al., *The role of histamine and histamine receptors in mast cell-mediated allergy and inflammation: the hunt for new therapeutic targets.* Frontiers in immunology, 2018. **9**: p. 1873.
359. Carlos, D., et al., *Histamine modulates mast cell degranulation through an indirect mechanism in a model IgE-mediated reaction.* European journal of immunology, 2006. **36**(6): p. 1494-1503.
360. Pashmforosh, M., et al., *Possible Benefits of Paclitaxel Therapy for COVID-19.* Pharmaceutical and Biomedical Research, 2022.
361. Chen, X., et al., *Immunomodulatory and antiviral activity of metformin and its potential implications in treating coronavirus disease 2019 and lung injury.* Frontiers in Immunology, 2020. **11**: p. 2056.
362. Samuel, S.M., E. Varghese, and D. Büsselberg, *Therapeutic potential of metformin in COVID-19: reasoning for its protective role.* Trends in microbiology, 2021. **29**(10): p. 894-907.
363. Zhou, G., et al., *Role of AMP-activated protein kinase in mechanism of metformin action.* The Journal of clinical investigation, 2001. **108**(8): p. 1167-1174.
364. Salt, I.P. and T.M. Palmer, *Exploiting the anti-inflammatory effects of AMP-activated protein kinase activation.* Expert opinion on investigational drugs, 2012. **21**(8): p. 1155-1167.
365. Kim, D., et al., *The architecture of SARS-CoV-2 transcriptome.* Cell, 2020. **181**(4): p. 914-921. e10.
366. Zhou, Q., et al., *The role of SARS-CoV-2-mediated NF-κB activation in COVID-19 patients.* Hypertension Research, 2023: p. 1-10.
367. Davies, D.A., A. Adlimoghaddam, and B.C. Albensi, *The effect of COVID-19 on NF-κB and neurological manifestations of disease.* Molecular Neurobiology, 2021. **58**(8): p. 4178-4187.
368. Jørgensen, P., et al., *Sirolimus interferes with the innate response to bacterial products in human whole blood by attenuation of IL-10 production.* Scandinavian journal of immunology, 2001. **53**(2): p. 184-191.
369. Boor, P.P., et al., *Rapamycin has suppressive and stimulatory effects on human plasmacytoid dendritic cell functions.* Clinical & Experimental Immunology, 2013. **174**(3): p. 389-401.
370. Battaglia, M., et al., *Rapamycin and interleukin-10 treatment induces T regulatory type 1 cells that mediate antigen-specific transplantation tolerance.* Diabetes, 2006. **55**(1): p. 40-49.
371. Ravid, J.D., O. Leiva, and V.C. Chitalia, *Janus kinase signaling pathway and its role in COVID-19 inflammatory, vascular, and thrombotic manifestations.* Cells, 2022. **11**(2): p. 306.
372. Khaledi, M., et al., *COVID-19 and the potential of Janus family kinase (JAK) pathway inhibition: A novel treatment strategy.* Frontiers in medicine, 2022. **9**: p. 961027.
373. Tanaka, T., M. Narazaki, and T. Kishimoto, *Interleukin (IL-6) immunotherapy.* Cold Spring Harbor perspectives in biology, 2018. **10**(8): p. a028456.
374. Gajjela, B.K. and M.-M. Zhou, *Calming the cytokine storm of COVID-19 through inhibition of JAK2/STAT3 signaling.* Drug discovery today, 2022. **27**(2): p. 390-400.

375. Laing, A.G., et al., *A dynamic COVID-19 immune signature includes associations with poor prognosis.* Nature medicine, 2020. **26**(10): p. 1623-1635.
376. Rheingold, S.Z., et al., *Zinc Supplementation Associated With a Decrease in Mortality in COVID-19 Patients: A Meta-Analysis.* Cureus, 2023. **15**(6).
377. Mashauri, H.L., *Covid-19 Histamine theory: Why antihistamines should be incorporated as the basic component in Covid-19 management?* Health Science Reports, 2023. **6**(2).
378. Al-Kuraishy, H.M., et al., *Anti-histamines and Covid-19: hype or hope.* JPMA J Pak Med Assoc, 2021. **71**(12): p. 144-8.
379. Suresh, M.V., et al., *Therapeutic potential of curcumin in ARDS and COVID-19.* Clinical and Experimental Pharmacology and Physiology, 2023. **50**(4): p. 267-276.
380. Vahedian-Azimi, A., et al., *Effectiveness of curcumin on outcomes of hospitalized COVID-19 patients: A systematic review of clinical trials.* Nutrients, 2022. **14**(2): p. 256.
381. Hudzik, B., J. Nowak, and B. Zubelewicz-Szkodzinska, *Consideration of immunomodulatory actions of morphine in COVID-19-Short report.* Eur Rev Med Pharmacol Sci, 2020. **24**(24): p. 13062-13064.
382. Teixeira, L., et al., *Simvastatin downregulates the SARS-CoV-2-induced inflammatory response and impairs viral infection through disruption of lipid rafts.* Frontiers in Immunology, 2022. **13**: p. 820131.
383. Vitiello, A., R. La Porta, and F. Ferrara, *Correlation between the use of statins and COVID-19: what do we know?* BMJ Evidence-Based Medicine, 2022. **27**(2): p. 126-127.
384. DE FLORA, S., R. Balansky, and S. LA MAESTRA, *Antioxidants and COVID-19.* Journal of preventive medicine and hygiene, 2021. **62**(1 Suppl 3): p. E34.
385. Attademo, L. and F. Bernardini, *Are dopamine and serotonin involved in COVID-19 pathophysiology?* The European journal of psychiatry, 2021. **35**(1): p. 62.
386. Bramante, C.T., et al., *Outpatient treatment of COVID-19 and incidence of post-COVID-19 condition over 10 months (COVID-OUT): a multicentre, randomised, quadruple-blind, parallel-group, phase 3 trial.* The Lancet Infectious Diseases, 2023.
387. Xiu, H., et al., *Fludarabine inhibits type I interferon-induced expression of the SARS-CoV-2 receptor angiotensin-converting enzyme 2.* Cellular & Molecular Immunology, 2021. **18**(7): p. 1829-1831.
388. Bustamante, S., et al., *Tryptophan Metabolism 'Hub'Gene Expression Associates with Increased Inflammation and Severe Disease Outcomes in COVID-19 Infection and Inflammatory Bowel Disease.* International Journal of Molecular Sciences, 2022. **23**(23): p. 14776.
389. El-Tanani, M., et al., *Phase II, Double-Blinded, Randomized, Placebo-Controlled Clinical Trial Investigating the Efficacy of Mebendazole in the Management of Symptomatic COVID-19 Patients.* Pharmaceuticals, 2023. **16**(6): p. 799.
390. Halpin, M., et al., *A prospective, single-center, randomized phase 2 trial of etoposide in severe COVID-19 (preprint).* 2023.
391. Nair, A., P. Sharma, and M.K. Tiwary, *Glutathione deficiency in COVID19 illness-does supplementation help?* Saudi Journal of Anaesthesia, 2021. **15**(4): p. 458.
392. Weisberg, E., et al., *Repurposing of kinase inhibitors for treatment of COVID-19.* Pharmaceutical research, 2020. **37**: p. 1-29.
393. Ripamonti, C., et al., *HDAC inhibition as potential therapeutic strategy to restore the deregulated immune response in severe COVID-19.* Frontiers in immunology, 2022. **13**: p. 841716.
394. Salerni, B.L., et al., *Vinblastine induces acute, cell cycle phase–independent apoptosis in some leukemias and lymphomas and can induce acute apoptosis in others when Mcl-1 is suppressed.* Molecular cancer therapeutics, 2010. **9**(4): p. 791-802.
395. Suriawinata, E. and K.J. Mehta, *Iron and iron-related proteins in COVID-19.* Clinical and Experimental Medicine, 2022: p. 1-23.

396. Han, J., et al., *Discovery of podofilox as a potent cGAMP–STING signaling enhancer with antitumor activity.* Cancer Immunology Research, 2023. **11**(5): p. 583-599.
397. Wu, J., et al., *The SARS-CoV-2 induced targeted amino acid profiling in patients at hospitalized and convalescent stage.* Bioscience Reports, 2021. **41**(3): p. BSR20204201.
398. Páez-Franco, J.C., et al., *Metabolomics analysis reveals a modified amino acid metabolism that correlates with altered oxygen homeostasis in COVID-19 patients.* Scientific reports, 2021. **11**(1): p. 6350.
399. Izquierdo-Alonso, J.L., et al., *N-acetylcysteine for prevention and treatment of COVID-19: Current state of evidence and future directions.* Journal of infection and public health, 2022.
400. Bharadwaj, S., et al., *SARS-CoV-2 Mpro inhibitors: identification of anti-SARS-CoV-2 Mpro compounds from FDA approved drugs.* Journal of Biomolecular Structure and Dynamics, 2022. **40**(6): p. 2769-2784.
401. Ma, H., et al., *Homo-harringtonine, highly effective against coronaviruses, is safe in treating COVID-19 by nebulization.* Science China Life Sciences, 2022. **65**(6): p. 1263-1266.
402. Hamizi, K., S. Aouidane, and G. Belaaloui, *Etoposide-based therapy for severe forms of COVID-19.* Medical Hypotheses, 2020. **142**: p. 109826.
403. Aoyagi, T., et al., *Case Report: Successful Treatment of Five Critically Ill Coronavirus Disease 2019 Patients Using Combination Therapy With Etoposide and Corticosteroids.* Frontiers in Medicine, 2021. **8**: p. 718641.
404. Cengiz, M., et al., *Effect of oral l-Glutamine supplementation on Covid-19 treatment.* Clinical nutrition experimental, 2020. **33**: p. 24-31.
405. Durante, W., *Glutamine Deficiency Promotes Immune and Endothelial Cell Dysfunction in COVID-19.* International Journal of Molecular Sciences, 2023. **24**(8): p. 7593.
406. Pektaş, S., C. Gürsoy, and S. Gümüş Demirbilek, *The use of pregabalin in Intensive Care Unit in the treatment of Covid-19-related pain and cough.* 2021.
407. Liu, H., et al., *DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy.* Nucleic acids research, 2020. **48**(D1): p. D871-D881.
408. Schoot, T.S., et al., *Immunosuppressive drugs and COVID-19: a review.* Frontiers in pharmacology, 2020. **11**: p. 1333.
409. Alam, M.S. and D.M. Czajkowsky, *SARS-CoV-2 infection and oxidative stress: Pathophysiological insight into thrombosis and therapeutic opportunities.* Cytokine & Growth Factor Reviews, 2022. **63**: p. 44-57.
410. Cecchini, R. and A.L. Cecchini, *SARS-CoV-2 infection pathogenesis is related to oxidative stress as a response to aggression.* Medical hypotheses, 2020. **143**: p. 110102.
411. Ntyonga-Pono, M.-P., *COVID-19 infection and oxidative stress: an under-explored approach for prevention and treatment?* The Pan African Medical Journal, 2020. **35**(Suppl 2).
412. Delgado-Roche, L. and F. Mesta, *Oxidative stress as key player in severe acute respiratory syndrome coronavirus (SARS-CoV) infection.* Archives of medical research, 2020. **51**(5): p. 384-387.
413. Laforge, M., et al., *Tissue damage from neutrophil-induced oxidative stress in COVID-19.* Nature Reviews Immunology, 2020. **20**(9): p. 515-516.
414. Wróblewska, J., et al., *The Role of Glutathione in Selected Viral Diseases.* Antioxidants, 2023. **12**(7): p. 1325.
415. Jennings, M.R. and R.J. Parks, *Curcumin as an antiviral agent.* Viruses, 2020. **12**(11): p. 1242.
416. Archin, N.M., et al., *Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy.* Nature, 2012. **487**(7408): p. 482-485.
417. Takimoto, C.H., *Anticancer drug development at the US National Cancer Institute.* Cancer chemotherapy and pharmacology, 2003. **52**: p. 29-33.
418. Johnson, T.S., et al., *Etoposide selectively ablates activated T cells to control the immunoregulatory disorder hemophagocytic lymphohistiocytosis.* The Journal of Immunology, 2014. **192**(1): p. 84-91.

419. Bailly, C., *Etoposide: A rider on the cytokine storm.* Cytokine, 2023. **168**: p. 156234.
420. Wang, Y.-N., et al., *Vinblastine resets tumor-associated macrophages toward M1 phenotype and promotes antitumor immune response.* Journal for Immunotherapy of Cancer, 2023. **11**(8).
421. Zhang, S., et al., *COVID-19 containment: China provides important lessons for global response.* Frontiers of Medicine, 2020. **14**: p. 215-219.
422. Paine, S.K., et al., *Multi-faceted dysregulated immune response for COVID-19 infection explaining clinical heterogeneity.* Cytokine, 2024. **174**: p. 156434.
423. Choi, B.Y., A.R. Grace, and J. Tsai, *Heterogeneity of COVID-19 symptoms and associated factors: Longitudinal analysis of laboratory-confirmed COVID-19 cases in San Antonio.* Plos one, 2023. **18**(12): p. e0295418.
424. Sinha, P., et al., *Prevalence of phenotypes of acute respiratory distress syndrome in critically ill patients with COVID-19: a prospective observational study.* The Lancet Respiratory Medicine, 2020. **8**(12): p. 1209-1218.
425. Sameh, M., et al., *Integrated multiomics analysis to infer COVID-19 biological insights.* Scientific Reports, 2023. **13**(1): p. 1802.
426. Zhu, K., et al., *Multi-omics and immune cells' profiling of COVID-19 patients for ICU admission prediction: in silico analysis and an integrated machine learning-based approach in the framework of Predictive, Preventive, and Personalized Medicine.* EPMA Journal, 2023. **14**(1): p. 101-117.
427. Bhat, S., P. Rishi, and V.D. Chadha, *Understanding the epigenetic mechanisms in SARS CoV-2 infection and potential therapeutic approaches.* Virus Research, 2022. **318**: p. 198853.
428. Bradic, M., et al., *DNA methylation predicts the outcome of COVID-19 patients with acute respiratory distress syndrome.* Journal of Translational Medicine, 2022. **20**(1): p. 526.
429. Reuben, R.C., R. Beugnon, and S.D. Jurburg, *COVID-19 alters human microbiomes: a meta-analysis.* Frontiers in Cellular and Infection Microbiology, 2023. **13**.
430. group, E.P.M.-o.w., et al., *Multi-omics Quality Assessment in Personalized Medicine through EATRIS.* bioRxiv, 2023: p. 2023.10. 25.563912.
431. van Reisen, M., et al., *Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research.* Advanced Genetics, 2021. **2**(2): p. e10050.
432. Liao, X., et al., *FAIR Data Cube, a FAIR data infrastructure for integrated multi-omics data analysis.* medRxiv, 2023: p. 2023.04. 23.23289000.
433. Niehues, A., et al., *A Multi-omics Data Analysis Workflow Packaged as a FAIR Digital Object.* bioRxiv, 2023: p. 2023.06. 07.543986.
434. Conesa, A. and S. Beck, *Making multi-omics data accessible to researchers.* Scientific data, 2019. **6**(1): p. 251.
435. Abrams, E.M. and S.J. Szefler, *COVID-19 and the impact of social determinants of health.* The Lancet Respiratory Medicine, 2020. **8**(7): p. 659-661.
436. Edwards, T.M. and J.P. Myers, *Environmental exposures and gene regulation in disease etiology.* Environmental health perspectives, 2007. **115**(9): p. 1264-1270.
437. Singu, S., et al., *Impact of social determinants of health on the emerging COVID-19 pandemic in the United States.* Frontiers in public health, 2020. **8**: p. 406.
438. Weber, L.M., et al., *Essential guidelines for computational method benchmarking.* Genome Biology, 2019. **20**(1): p. 125.
439. Subramanian, I., et al., *Multi-omics data integration, interpretation, and its application.* Bioinformatics and biology insights, 2020. **14**: p. 1177932219899051.
440. Jensen, F.V., *Causal and Bayesian networks*, in *Bayesian networks and decision graphs*. 2001, Springer. p. 3-34.
441. Wang, C., et al., *Network-based integration of multi-omics data for clinical outcome prediction in neuroblastoma.* Scientific Reports, 2022. **12**(1): p. 15425.

442. Mikaeloff, F., et al., *Network-based multi-omics integration reveals metabolic at-risk profile within treated HIV-infection.* Elife, 2023. **12**: p. e82785.
443. Le, D.-H. and V.-H. Pham, *Drug response prediction by globally capturing drug and cell line information in a heterogeneous network.* Journal of molecular biology, 2018. **430**(18): p. 2993-3004.
444. Zhang, F., et al., *A novel heterogeneous network-based method for drug response prediction in cancer cell lines.* Scientific reports, 2018. **8**(1): p. 3355.
445. Haas, M., et al., *Big data to smart data in Alzheimer's disease: Real-world examples of advanced modeling and simulation.* Alzheimer's & Dementia, 2016. **12**(9): p. 1022-1030.