

# MOATAI-VIR - an AI algorithm that predicts severe adverse events and molecular features for COVID-19's complications

Courtney Astore<sup>1</sup>, Hongyi Zhou<sup>1</sup>, Joshy Jacob<sup>2,3,4</sup>, and Jeffrey Skolnick<sup>1\*</sup>

\* Corresponding author: Jeffrey Skolnick

Email: [skolnick@gatech.edu](mailto:skolnick@gatech.edu)

<sup>1</sup>Center for the Study of Systems Biology, School of Biological Sciences, Georgia Institute of Technology, 950 Atlantic Drive, N.W., Atlanta, GA 30332, USA

<sup>2</sup>Emory Vaccine Center, Emory University, Atlanta, GA30329, USA

<sup>3</sup>Yerkes National Primate Research Center, Emory University, Atlanta, GA30329, USA

<sup>4</sup>Department of Microbiology and Immunology, Emory Vaccine Center, School of Medicine, Emory University, Atlanta, GA 30329, USA

## Abstract

Following SARS-CoV-2 infection, some COVID-19 patients experience severe adverse events caused by pathogenic host responses. To treat these complications, their underlying etiology must be identified. Thus, a novel AI-based methodology, **MOATAI-VIR**, which predicts disease-protein-pathway relationships for 22 clinical manifestations attributed to COVID-19 was developed. SARS-CoV-2 interacting human proteins and GWAS identified respiratory failure associated risk genes provide the input from which the mode-of-action (MOA) proteins/pathways of the resulting disease comorbidities are predicted. These comorbidities are then mapped to their clinical manifestations. Three uncharacterized manifestation categories are found: neoplasms, mental and behavioral disorders, and congenital malformations, deformations, and chromosomal abnormalities. The prevalence of neoplasms suggests a possible association between COVID-19 and cancer, whether by shared molecular mechanisms between oncogenesis and viral replication, or perhaps, SARS-CoV-2 is an oncovirus. To assess the molecular basis of each manifestation, the proteins shared across each group of comorbidities were prioritized and subject to global pathway analysis. From these most frequent pathways, the molecular features associated with hallmark COVID-19 phenotypes, such as loss of sense of smell/taste, unusual neurological symptoms, cytokine storm, and blood clots were explored. Results of **MOATAI-VIR** are available for academic users at: <http://pwp.gatech.edu/cssb/MOATAI-VIR/>.

## Introduction

The COVID-19 pandemic is caused by SARS-CoV-2, a positive-sense, single-stranded, rapidly mutating RNA coronavirus(1). The societal impact of COVID-19 is amplified by the minority of individuals experiencing significant complications/death following infection. Some victims experience acute respiratory distress syndrome(2), while others have clotting issues, cytokine storms, hypoxemia, low white blood cell counts, and bone marrow failure(3-6). Despite the development of COVID-19 vaccines, until they are widely administered, there will be new cases of COVID-19 with complications including loss of smell/taste and/or unusual neurological symptoms (7, 8). Thus, systematic approaches are desperately needed to combat COVID-19's catastrophic long-term effects(9).

The primary objective of this work is to provide the molecular mechanisms for each clinical manifestation attributed to COVID-19. We define clinical manifestations as in Ref.(10) which provides 30 respiratory and non-respiratory COVID-19 in-hospital clinical complications (excluding the "other" category). To assist in the effort to further define these complications, a systematic method to identify comorbid diseases and the molecular features of these complications is preferred over a random, anecdotal approach.

To address this pressing need, we describe a new algorithm, **MOATAI-VIR**, **Mode Of Action proteins & Targeted therapeutic discovery driven by Artificial Intelligence for VIRuses** designed to predict the mode-of-action (MOA) proteins of COVID-19s severe patient responses based on predicted COVID-19 disease comorbidities. To accomplish this, we independently input either the experimentally determined human-SARS-CoV-2 interactome or COVID-19 GWAS survival-associated risk genes in a MOA indication profile(5). These profiles are used to determine the disease comorbidities associated with the MOA proteins presumably causing a particular complication. In practice, each COVID-19 comorbid disease is mapped to its respective clinical manifestation group. Then, the top comorbidity enriched MOA proteins are subject to pathway analysis to identify the molecular processes underlying the etiology of each complication. There were also comorbidities for both the interactome and GWAS sets that did not map to a characterized COVID-19 clinical manifestation group; one example is cancers. These "uncharacterized" complications were grouped by their ICD-10 main classification(11). We next determined the enriched MOA proteins and performed global pathway analysis.

## Results

### Prediction of COVID-19's complications and underlying molecular mechanisms

An overview of the **MOATAI-VIR** approach is shown in Figure 1, with a more detailed flowchart in Figure S1. The goal of **MOATAI-VIR** is to identify severe adverse responses associated with SARS-CoV-2 and their corresponding human mode-of-action proteins. To accomplish these objectives, we input either the experimentally determined human proteins from the human-SARS-CoV-2 interactome(6) or COVID-19 GWAS survival associated risk genes(2, 5) as a MOA profile in **MEDICASCY**. We then employ our recently developed **LeMeDISCO** algorithm which predicts disease comorbidity. **LeMeDISCO** then outputs indications that likely cause the adverse events due to SARS-

CoV-2 infection. The set of top identified proteins is then passed to CoPathway and the pathways are ranked by their statistical significance in  $LIST_{CoPathway}$ .

## Large Scale Benchmarking of MEDICASCY MOA Predictions

The first step uses **MEDICASCY**(12) to predict MOA protein targets of all diseases. For large scale benchmarking for MOA prediction, we mapped all drugs in our indication library to DrugBank drugs(13) (v5.09) and obtained their respective human protein targets. These are combined with those from the Therapeutic Target Database(14). Using the drug-indication relationships in our training library of **MEDICASCY**, we compiled indication-protein target relationships of 145,722 pairs for 3,539 indications (with an average ~41 proteins/indication) for benchmarking. In benchmarking, any drug in the training library having a Tanimoto coefficient,  $TC(15) \geq 0.8$  to the given drug whose indications are predicted is excluded from training. We define a MOA prediction for an indication when its p-value  $< 0.001$  using the upper tailed null hypothesis. 61.0% of predicted protein-indication prediction are correctly predicted. If a training  $TC=1$  cutoff is used, this increases to 76.9%.

## Large Scale Benchmarking of LeMeDISCO

As shown and discussed in SI, Table S0, large scale benchmarking of **LeMeDISCO** on different clinical data sets shows high comorbid disease coverage and accuracy compared to alternative methods(16-19). It is also superior to alternatives that rely solely on symptom data, which lack molecular mechanism-based associations and cannot provide information of confidence-ranked putative protein targets. **LeMeDISCO**'s recall rate is close to 70% for a representative set of 2,630 disease pairs.

## COVID-19's Clinical Manifestations

We first predicted COVID-19 comorbidities using the 332 high confidence human proteins that interact with SARS-CoV-2(6). There are 697 significant comorbidities (with a Z-score cutoff  $> 1.65$  corresponding to a p-value of 0.05 using the upper tailed null hypothesis). The top five disease comorbidities ranked by their Z-score that mapped to a COVID-19 clinical manifestation group are shown in Table 1, with an expanded list in Table S1. Also provided are comorbidity enriched protein targets. Without extrinsic information or training, these results recapitulate many key COVID-19 phenotypes such as myelosuppression, immunodeficiency, neurotoxicity, blood indications, myocardial infarctions, stroke and cytokine storm symptoms(2, 6, 20). The ICD-10 code of the comorbid diseases was used to map them to the 30 complications listed in Ref.(10). After ICD-10 mapping, we manually examine the unmapped ones and search for literature evidence of a possible match. For example, we do not automated have an ICD-10 match of Fallot's tetralogy (ICD-10:Q21) to the Cardiovascular complication(ICD-10: I00-I99). Since it is a congenital heart disease, it is manually assigned to this complication. Manually mapped indications are ~10% of the mapped ones. Thus, even if there are few errors, it will not significantly affect the final results. In practice, indications are mapped into 17 of the 30 COVID-19 complications (see Table 1). Our library of 3,608 indications(21) does not have any matches to 4/30 complications (Dialysis initiation, Intracranial hemorrhage, Hypertensive crisis, Cardiogenic shock). Effectively,

the complication recall rate is 17/26~65%. These mapped indications are then used to prioritize MOA proteins, pathways.

The 6 human genes near the 3p21.31 locus of human genome identified in a GWAS study as being strongly associated with respiratory failure in COVID-19 patients(5) (odds ratio 1.77) were next used to predict comorbidities. We determined 380 significant comorbidities having a Z-score >1.65. Strikingly, as shown in Table 2, many severe clinical complications associated with COVID-19 are in the top-ranked comorbidity predictions including respiratory complications, myocardial infarction and cytokine storms. Table S2 provides an expanded list that includes myocardial infarction, stroke, neurological manifestations, hearing disorders, hypoxemia, lung, cardiovascular and diabetic risk factors(2, 20, 22). Excluding 4 indications not in our library, with GWAS risk gene input, the recall rate of the COVID-19 complications is 19/26~73%. Since the 332 proteins of the SARS-CoV-2 human interactome and the 6 GWAS COVID-19 survival risk genes that do not overlap except for FYCO1, their comorbidity predictions are complementary. Combining both the human interactome and GWAS complications provides a prediction of 22/26 (84%) of COVID-19's complications.

### **Diseases comorbid with COVID-19**

From the significant comorbidities found from the interactome data, there were 297 neurologic, 41 hematologic, 34 ocular, 26 endocrine, 14 gastrointestinal, 13 cardiovascular, 11 respiratory, 9 renal failure, 9 sepsis, 6 cerebral ischemia, 6 dermatological, 5 bacteremia, 5 hepatocellular, 2 acute myocardial infarction, 1 diabetic ketoacidosis, 1 Disseminated intravascular coagulation (DIC), and 1 pulmonary embolism diseases. From the GWAS data, there were 48 neurologic, 47 cardiovascular, 28 ocular, 27 respiratory, 26 endocrine, 17 gastrointestinal, 13 cerebral ischemia, 10 hepatocellular injury, 9 renal failure, 8 acute myocardial infarction, 7 dermatological, 6 bacteremia, 5 hematologic, 4 sepsis, 4 asthma exacerbation, 3 chronic obstructive pulmonary disease (COPD), 1 acute congestive heart failure (CHF), 1 acute respiratory distress syndrome (ARDS), and 1 respiratory failure diseases.

### **Pathway analysis**

The top 100 comorbidity enriched MOA proteins for each complication with  $n > 1$  diseases were used as input into global pathway analysis. If there was only 1 disease in a complication, all of the **MEDICASCY** significantly predicted MOA proteins for that disease were used as inputs into the global pathway analysis. A top important pathway, the interactome and GWAS as input results are shown in Tables 1 and 2, with full lists in Tables S1 and S2 respectively.

The top 20 most frequent significant pathways across the clinical manifestations calculated from the interactome and GWAS inputs are shown in Tables 3 and 4, respectively. Combining this with the hierarchically ranked pathways for each clinical manifestation allowed us to further identify hallmark pathways attributed with unusual COVID-19 symptoms such as loss of sense of smell, cytokine storms, blood clots, and unusual neurological symptoms.

### **Loss of sense of smell**

The olfactory signaling pathway is associated with 18/19 clinical manifestation groups from the GWAS results. This pathway involves enzymatic activity that initiates the

binding of an odorant molecule to a receptor resulting in an electrical signal that is transmitted to the brain(23). The olfactory signaling pathway was the top pathway for the asthma exacerbation, hematologic, respiratory failure, and gastrointestinal GWAS clinical manifestations. There is clinical evidence that some individuals infected with SARS-CoV-2 experience a loss of smell and taste(23). Furthermore, it has been found that olfactory receptors may be an alternative SARS-CoV-2 entry into the local host cells, which may lead to its spread into the central nervous system(24).

### **Neurological symptoms**

There have been a number of unusual COVID-19-related neurological symptoms reported, such as stroke, confusion, and as previously mentioned, loss of sense of smell and taste(25). The most frequent clinical manifestation group for the interactome results was neurologic. The top neurologic pathway was glutathione conjugation associated with facilitating the metabolism of xenobiotics. Dysregulation of glutathione plays a role in a wide variety of diseases including neurodegenerative diseases and cancer(26, 27). Decreased glutathione levels can lead to oxidative stress which may result in Parkinson's and Alzheimer's disease. Moreover, an imbalance in glutathione levels can impact the immune system(27). A decreased glutathione concentration is highly associated with serious manifestations causing increased COVID-19 mortality, which may be due to the increased susceptibility to uncontrolled viral replication(28).

The most frequent clinical manifestation group from the GWAS results was also neurological. The top pathway was Na<sup>+</sup>/Cl<sup>-</sup> dependent neurotransmitter sodium symporters, which use sodium and chloride electrochemical gradients to import/export a number of substrates. They are associated with Parkinson's disease, orthostatic intolerance, and depression(29). A meta-analysis found that low blood sodium increases the risk and severity of COVID-19(30). Thus, neurotransmitter transporters that depend on Na<sup>+</sup> could be dysregulated due to decreased levels of blood sodium.

### **Blood clotting**

Increased blood clots have been seen in individuals infected with COVID-19. A meta-analysis found that individuals infected by COVID-19 with blood clots have increased mortality(31). The second most frequent clinical manifestation group for the interactome results was hematologic. The top hematologic pathway was RAB geranylgeranylation, the eighth most frequent pathway across all clinical manifestations from the interactome. RAB geranylgeranylation is a post-translational modification that allows RABs to connect with intracellular membranes where they regulate vesicle transport pathways(32). Dysregulation of RAB geranylgeranylation transferase function has been linked to abnormal blood clotting.

### **Cytokine storm**

There have been a number of COVID-19 patients that face ARDS, which may be due to a cytokine storm elicited the nuclear factor kappa B (NF-κB) pathway(33). The second most frequent clinical manifestation group for the GWAS results was cardiovascular. The top pathway was G alpha (s) signaling events, which covered all GWAS clinical manifestation groups. This pathway's primary function is to activate adenylate cyclase producing cAMP. G-protein receptors are associated with heart disease. Among the proteins in this pathway, C5aR1 is a G-protein-coupled receptor. The C5a-C5aR1 complex is involved in COVID-19 progression and is part of a potential therapeutic

strategy(34). This complex is associated with the innate immune response, with C5 a key driver in complement-mediated inflammation(34). Other cytokine storm-related pathways covering a majority of clinical manifestation groups for the interactome results include interleukin receptor SHC signaling, Interleukin-4 and Interleukin-13 signaling(35). ADORA2B mediated anti-inflammatory cytokines production and chemokine receptors bind chemokines are also prevalent pathways.

## Uncharacterized Manifestations

There was a total of 234 significant comorbidities that were not mapped to a known COVID-19 clinical manifestation group from the interactome results. To further understand their effects, we grouped them by their main ICD-10 classification and performed **CoPathway** analysis on the top two groups to further understand their molecular bases. The top two uncharacterized groups were Neoplasms and Congenital malformations, deformations, and chromosomal abnormalities. The comorbidities, comorbidity enriched MOA proteins and pathways for the interactome results are shown in SI, Table S3.

A total of 155 diseases were not mapped to a clinical manifestation group from the GWAS results. The comorbidities, comorbidity enriched MOA proteins and pathways for the GWAS results are shown in Table S4. Again, to further understand these potentially important COVID-19 adverse events, we grouped them by their main ICD-10 classification and performed **CoPathway** analysis on the top two groups. The top two uncharacterized groups were Neoplasms and Mental and behavioral disorders. There were 37 pathways with a p-value < 0.05. The top pathway was amine-ligand binding receptors, class A GPCRs that bind to biogenic amine ligands. These amines can act as neurotransmitters. A study found that specific trace amines are highly prevalent in patients with mental disorders(36). To the best of our knowledge, there has been no research investigating the relationship between amine-ligand binding receptors and COVID-19 as a suitable biomarker for therapeutic intervention; however, there are numerous COVID-19 associated mental and behavioral occurrences(37).

## Neoplasms

From interactome, there were 90 pathways with a p-value < 0.05 involving neoplasms, with many involving hormonal regulation. The top pathway involves activation of AMPK downstream of NMDARs, that is associated with the neuronal system. AMPK is an enzyme that regulates cellular energy and homeostasis via activating catabolic pathways while switching off cellular growth and proliferation(38). AMPK has been targeted for cancer treatment because its activation can reduce cancer incidence. NMDARs are vital in controlling synaptic plasticity and memory. Increased expression of NMDARs occurs in a variety of cancers such as neuroblastoma, breast, small-cell lung, and ovarian cancer(39). Anti-NMDAR encephalitis, characterized by abnormal neurological and behavioral symptoms, has been reported in both COVID-19 and HSV(40). Notably, HSV can lead to an increased cervical cancer risk(41, 42).

From GWAS, there were 29 pathways for Neoplasms with a p-value < 0.05. The top pathway was the nuclear receptor transcription pathway. Nuclear receptors are DNA-binding transcription factors capable of binding hormones, vitamins, small molecules, and other ligands. There are a number of underlying disease mechanisms associated with dysregulation of nuclear receptors. This can yield a wide range of conditions such as cancer, diabetes, and hormone-related conditions. Nuclear receptors have been targeted by cancer therapeutics as they are key players in gene regulatory networks(43). There has not been substantial research on the relationship between nuclear receptors and COVID-19; but we note that some viruses target nuclear receptors as part of the viral replication process(44).

## Is SARS-CoV-2 an oncovirus?

As indicated above, there were a large number of disease comorbidities associated with neoplasms from both the interactome and GWAS results. Neoplasms cause abnormal tissue growth, a significant cancer characteristic. Perhaps, SARS-CoV-2 hijacks the human host replication machinery or proliferation pathways(6). Indeed, viruses can initiate signal transduction pathways leading to cytokine and chemokine expression. They also dysregulate signaling pathways to promote viral infection and cellular transformations(45) that elicit a proinflammatory response similar to that in cancer(46). A salient example is Human Papillomavirus (HPV). Most cervical cancers(47) are caused by the cytokine flux associated with inflammation post-HPV infection(42). Furthermore, the second most significant COVID-19 comorbid disease T-cell leukemia is linked to the human T-cell lymphotropic virus (HTLV-I), an RNA retrovirus. More generally, a number of oncoviruses cause cancer(48). Certain viruses transform human cells causing loss of ability to regulate cell division.

Although we still do not know the long-term consequences post-COVID-19 infection, these results raise the possibility that SARS-CoV-2 is an oncovirus. To assess this potential relationship, we screened our comorbidity enriched MOA proteins associated with Neoplasms from both the interactome and GWAS sets against the COSMIC(49) database gene set containing 723 oncogenes. We found that 47% (n=340) and 29% (n=208) of the MOA proteins from the interactome and GWAS Neoplasms are oncogenes in the COSMIC database(42) respectively; see SI Tables S5-6. For further validation, we compared the differential gene expression set from Table S4 of (50), which demonstrated the differential gene expression analysis of COVID-19 patients (n=1,918 differentially expressed genes with an adjusted p-value < 0.05), to the COSMIC database(49). From this, 11% (n=82) of the genes overlapped with the oncogenes in the COSMIC database (see SI Table S7). We next performed a 3-way merge between the interactome/GWAS Neoplasms comorbidity enriched MOA proteins, the SARS-CoV-2 differentially expressed genes, and the COSMIC database oncogenes. We found 40 and 23 overlapping genes from the interactome and GWAS 3-way merge, respectively. The pathway analysis on the overlapping COVID-19 differentially expressed genes and the COSMIC database indicates homogeneity of viral replication mechanisms and oncogenesis. A number of the identified pathways such as interferon-gamma signaling(51, 52), immunoregulatory interactions between lymphoid and non-lymphoid cell(53), and antigen processing-cross presentation(54, 55) are related to viral replication and oncogenesis. Possibly the machinery of viral replication and oncogenesis are similar. Clearly, additional investigation is needed to explore the possibility that SARS-CoV-2 is an oncovirus.

## MOATAI-VIR Website

The resulting data from **MOATAI-VIR** is provided to academic users at: <http://pwp.gatech.edu/cssb/MOATAI-VIR/>. Provided on the webpage are downloadable files for the significant diseases, proteins and pathways for the highlighted symptoms (e.g., loss of sense of smell) and clinical manifestations for interactome and/or GWAS. Furthermore, the significant diseases, proteins, and pathways for the summary and uncharacterized manifestation results for interactome and/or GWAS are also downloadable.



## Discussion

**MOATAI-VIR** has been shown to identify possible molecular mechanisms responsible for COVID-19's severe adverse consequences. Most of COVID-19's severe symptoms successfully predicted. These predicted comorbidities and the human MOA proteins possibly responsible for COVID-19's complications will be validated in future work in cell lines and animal models. They will also be combined with antiviral drugs that directly target SARS-CoV-2 proteins to kill the virus. The goal is to mitigate both COVID-19 infection and subsequent adverse complications to improve clinical outcomes. Thus, **MOATAI-VIR** provides a series of logical, systematic choices designed to suggest treatments to COVID-19's adverse reactions. Equally important, **MOATAI-VIR** is a general methodology can be applied to understand the possible etiology of human host response new outbreaks of other novel viral infections as they emerge. Finally, we note that an extension of **MOATAI-VIR** can be used to suggest repurposed drugs that modulate the host response to COVID-19's complications.

## Materials and Methods

Additional details concerning **MOATAI-VIR** are provided in SI. Here, we note that **CoPathway** determines significant pathways associated with the most frequent comorbidity enriched MOA proteins. It first assesses the frequency of MOA proteins across the comorbidities for a desired group and then processing a top number (typically 100) of those MOA proteins through the Reactome(56) for global pathway analysis. Pathways with a p-value < 0.05 are then extracted and deemed the significant pathways portrayed in LIST<sub>CoPathway</sub>.

## Acknowledgments

This project was funded by R35GM118039 of the Division of General Medical Sciences of the NIH.

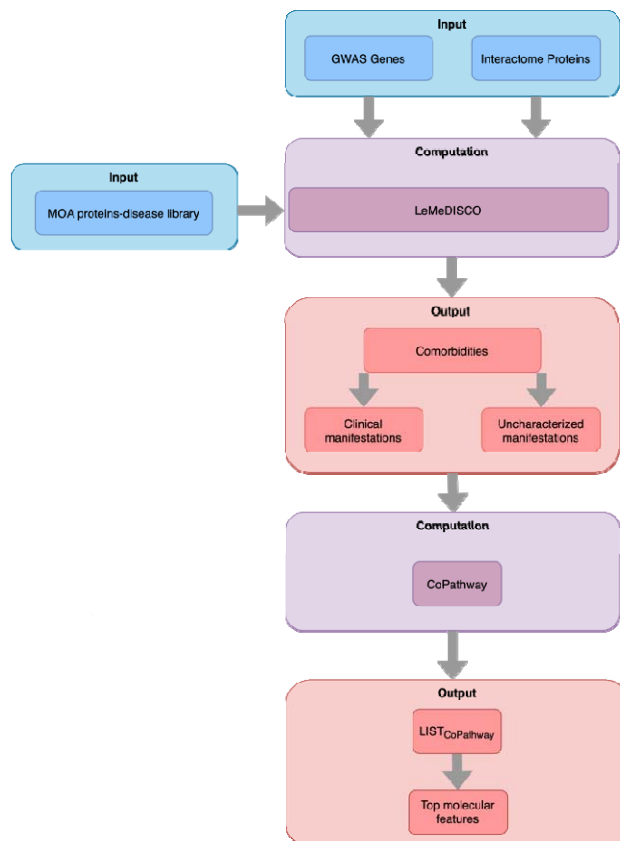
## References

1. P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology* 10.1038/s41579-020-00468-6 (2020).
2. C. Wu *et al.*, Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* 10.1001/jamainternmed.2020.0994, doi:10.1001/jamainternmed.2020.0994 (2020).
3. W. J. Guan *et al.*, Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* **382**, 1708-1720 (2020).
4. R. J. Jose, A. Manuel, COVID-19 cytokine storm: the interplay between inflammation and coagulation. *The Lancet Respiratory Medicine* **8**, e46-e47 (2020).
5. D. Ellinghaus, *et al.*, The Severe Covid-19 GWAS Group, Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.*, doi: 10.1056/NEJMoa2020283 (2020).
6. D. E. Gordon *et al.*, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 10.1038/s41586-020-2286-9, doi:10.1038/s41586-41020-42286-41589 (2020).
7. J. H. Beigel *et al.*, Remdesivir for the Treatment of Covid-19 — Preliminary Report. *New England Journal of Medicine* 10.1056/NEJMoa2007764 (2020).
8. L. Mao *et al.*, Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China. *JAMA Neurol* **77**, 683-690 (2020).
9. D. Yelin *et al.*, Long-term consequences of COVID-19: research needs. *The Lancet Infectious Diseases* **20**, 1115-1117 (2020).
10. L.-O. C. Cates J, Dahl RM, *et al.*, Risk for In-Hospital Complications Associated with COVID-19 and Influenza — Veterans Health Administration, United States, October 1, 2018–May 31, 2020. *MMWR Morb Mortal Wkly Rep* **2020** **69**, 1528–1534 (2020).
11. World Health Organization., "ICD-10 : international statistical classification of diseases and related health problems : tenth revision". (World Health Organization, Geneva, 2004).
12. H. Zhou *et al.*, MEDICASCY: A Machine Learning Approach for Predicting Small Molecule Drug Side Effects, Indications, Efficacy and Mode of Action. *Mol Pharm* **17**, 1558-1574 (2020).
13. D. S. Wishart *et al.*, DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* **46**, D1074-D1082 (2018).
14. Y. H. Li *et al.*, Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.* **46**, D1121-D1127 (2018).
15. T. T. Tanimoto (1958) An elementary mathematical theory of classification and prediction. in *IBM Internal Report*.
16. Y. Ko, M. Cho, J.-S. Lee, J. Kim, Identification of disease comorbidity through hidden molecular mechanisms. *Scientific Reports* **6**, 39433 (2016).
17. J. Menche *et al.*, Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
18. X. Zhou, J. Menche, A.-L. Barabási, A. Sharma, Human symptoms–disease network. *Nature Communications* **5**, 4212 (2014).

19. C. A. Hidalgo, N. Blumm, A. L. Barabasi, N. A. Christakis, A dynamic network approach for the study of human phenotypes. *PLoS computational biology* **5**, e1000353 (2009).
20. M. R. Mehra, et al., Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. *N Engl J Med*, DOI: 10.1056/NEJMoa2007621 (2020).
21. Y. Zhou, F. Wang, J. Tang, R. Nussinov, F. Cheng, Artificial intelligence in COVID-19 drug repurposing. *The Lancet Digital Health* **2**, e667-e676 (2020).
22. R. T. Gandhi, J. B. Lynch, C. Del Rio, Mild or Moderate Covid-19. *N Engl J Med* 10.1056/NEJMcp2009249, DOI: 10.1056/NEJMcp2009249 (2020).
23. A. S. Bhattacharjee, S. V. Joshi, S. Naik, S. Sangle, N. M. Abraham, Quantitative assessment of olfactory dysfunction accurately detects asymptomatic COVID-19 carriers. *EClinicalMedicine* **28** (2020).
24. A. Gori *et al.*, COVID-19-Related Anosmia: The Olfactory Pathway Hypothesis and Early Intervention. *Frontiers in Neurology* **11** (2020).
25. F. Ftiha, M. Shalom, H. Jradeh, Neurological symptoms due to Coronavirus disease 2019. *Neurol Int* **12**, 8639-8639 (2020).
26. A. J. L. Cooper, M. H. Hanigan, Metabolism of Glutathione S-Conjugates: Multiple Pathways. *Comprehensive Toxicology* 10.1016/B978-0-12-801238-3.01973-5, 363-406 (2018).
27. N. Ballatori *et al.*, Glutathione dysregulation and the etiology and progression of human diseases. *Biol Chem* **390**, 191-214 (2009).
28. A. Polonikov, Endogenous Deficiency of Glutathione as the Most Likely Cause of Serious Manifestations and Death in COVID-19 Patients. *ACS Infect Dis* **6**, 1558-1562 (2020).
29. A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, E. Gouaux, Crystal structure of a bacterial homologue of Na<sup>+</sup>/Cl<sup>-</sup>-dependent neurotransmitter transporters. *Nature* **437**, 215 (2005).
30. Y. Luo, Y. Li, J. Dai, Low blood sodium increases risk and severity of COVID-19: a systematic review, meta-analysis and retrospective cohort study. *medRxiv* 10.1101/2020.05.18.20102509, 2020.2005.2018.20102509 (2020).
31. M. B. Malas *et al.*, Thromboembolism risk of COVID-19 is high and associated with a higher risk of mortality: A systematic review and meta-analysis. *EClinicalMedicine* **29** (2020).
32. M. C. Seabra, Nucleotide dependence of Rab geranylgeranylation. Rab escort protein interacts preferentially with GDP-bound Rab. *J Biol Chem* **271**, 14398-14404 (1996).
33. S. Hojyo *et al.*, How COVID-19 induces cytokine storm with high mortality. *Inflammation and Regeneration* **40**, 37 (2020).
34. T. M. Woodruff, A. K. Shukla, The Complement C5a-C5aR1 GPCR Axis in COVID-19 Therapeutics. *Trends in Immunology* **41**, 965-967 (2020).
35. T. Kanai *et al.*, Regulatory effect of interleukin-4 and interleukin-13 on colon cancer cell adhesion. *British Journal of Cancer* **82**, 1717-1723 (2000).
36. B. Borowsky *et al.*, Trace amines: identification of a family of mammalian G protein-coupled receptors. *Proc Natl Acad Sci U S A* **98**, 8966-8971 (2001).
37. N. M. Simon, G. N. Saxe, C. R. Marmar, Mental Health Disorders Related to COVID-19-Related Deaths. *JAMA* **324**, 1493-1494 (2020).
38. S. Herzig, R. J. Shaw, AMPK: guardian of metabolism and mitochondrial homeostasis. *Nature Reviews Molecular Cell Biology* **19**, 121-135 (2018).

39. A. R. Banreti, P. Meier, Myc supercompetitor cells exploit the NMDA receptor to subdue their wild-type neighbours via cell competition. *bioRxiv* 10.1101/2020.02.11.943498, 2020.2002.2011.943498 (2020).
40. T. Burr, C. Barton, E. Doll, A. Lakhotia, M. Sweeney, N-Methyl-d-Aspartate Receptor Encephalitis Associated With COVID-19 Infection in a Toddler. *Pediatric Neurology* **114**, 75-76 (2021).
41. K. A. Rafferty, Jr., Herpes viruses and cancer. *Sci Am* **229**, 26-33 (1973).
42. N. Hemmat, H. Bannazadeh Baghi, Association of human papillomavirus infection and inflammation in cervical cancer. *Pathog Dis* **77** (2019).
43. L. Zhao, S. Zhou, J.-Å. Gustafsson, Nuclear Receptors: Recent Drug Discovery for Cancer Therapies. *Endocrine Reviews* **40**, 1207-1249 (2019).
44. M. S. Miller, J. S. Mymryk, An unhealthy relationship: viral manipulation of the nuclear receptor superfamily. *Future Microbiol* **6**, 999-1019 (2011).
45. N. A. Krump, J. You, Molecular mechanisms of viral oncogenesis in humans. *Nature Reviews Microbiology* **16**, 684-698 (2018).
46. T. H. Mogensen, S. R. Paludan, Molecular pathways in virus-induced cytokine production. *Microbiol Mol Biol Rev* **65**, 131-150 (2001).
47. C. de Martel, M. Plummer, J. Vignat, S. Franceschi, Worldwide burden of cancer attributable to HPV by site, country and HPV type. *Int J Cancer* **141**, 664-670 (2017).
48. E. Matutes, Adult T-cell leukaemia/lymphoma. *J Clin Pathol* **60**, 1373-1377 (2007).
49. J. G. Tate *et al.*, COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research* **47**, D941-D947 (2019).
50. D. Blanco-Melo *et al.*, Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045.e1039 (2020).
51. S. Kang, H. M. Brown, S. Hwang, Direct Antiviral Mechanisms of Interferon-Gamma. *Immune Netw* **18**, e33-e33 (2018).
52. L. Wang, Y. Wang, Z. Song, J. Chu, X. Qu, Deficiency of Interferon-Gamma or Its Receptor Promotes Colorectal Cancer Development. *Journal of Interferon & Cytokine Research* **35**, 273-280 (2014).
53. S. Nedvetzki *et al.*, Reciprocal regulation of human natural killer cells and macrophages associated with distinct immune synapses. *Blood* **109**, 3776-3785 (2007).
54. E. L. Heipertz, M. L. Davies, E. Lin, C. C. Norbury, Prolonged Antigen Presentation following an Acute Virus Infection Requires Direct and Then Cross-Presentation. *The Journal of Immunology* **193**, 4169 (2014).
55. J. Bandola-Simon, P. A. Roche, Dysfunction of antigen processing and presentation by dendritic cells in cancer. *Mol Immunol* **113**, 31-37 (2019).
56. B. Jassal *et al.*, The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-d503 (2020).

## Figures and Tables



**Figure 1.** Overview of the **MOATAI-VIR** approach that predicts comorbid human diseases, their MOA proteins and molecular features associated with COVID-19's severe secondary adverse events. Blue are **MOATAI-VIR** inputs, purple are computational algorithms and pink are output predictions.

**Table 1.** Top 5 comorbidities, top 5 comorbidity enriched MOA proteins, and topmost relevant pathway for each COVID-19 severe adverse clinical manifestations predicted using the SARS-CoV-2 interactome as input

Clinical manifestation	Comorbidities	Comorbidity enriched MOA proteins	Topmost relevant pathway
Respiratory	pleural disease idiopathic pulmonary fibrosis idiopathic interstitial pneumonia viral pneumonia severe acute respiratory syndrome	KCNA10 OSBPL5 OSBPL8 ITGB1 ORC3	TNFs bind their physiological receptors
Cardiovascular/Arrhythmia	coronary stenosis Wolff-Parkinson-White syndrome atrioventricular block carotid artery disease lymphatic system disease	DDRKG1 RPS19 CCT6B FKBP10 FKBP1A	Interleukin receptor SHC signaling
Acute myocardial infarction/Unstable angina	coronary thrombosis intermediate coronary syndrome	FKBP4 GSN CILP TLE3 CDC20B	Signaling by activated point mutants of FGFR1
Hematologic	severe combined immunodeficiency MHC class II deficiency myelofibrosis Diamond-Blackfan anemia paroxysmal nocturnal hemoglobinuria	CCT6B GART HENMT1 METTL21B PAN2	RAB geranylgeranylation
Pulmonary embolism	pulmonary embolism and infarction	LOC390956 GTF2H2C WDR13 OTC LANCL3	Glutathione conjugation

Disseminated intravascular coagulation (DIC)	purpura fulminans	BCL2L1 RPS19 TCOF1 DDRGK1 CTSA	Defective gamma-carboxylation of F9
Neurologic	aseptic meningitis relapsing-remitting multiple sclerosis toxic encephalopathy Bell's palsy encephalitis	PARS2 MMP27 SV2B MMP28 MMP23B	Glutathione conjugation
Cerebral ischemia/infarction	carotid artery disease lymphatic system disease brain stem infarction peripheral vascular disease	FKBP11 FKBP10 FKBP2 IL7R FKBP3	Events associated with phagocytolytic activity of PMN cells
Endocrine	ovarian disease lysosomal storage disease lipomatosis Chediak-Higashi syndrome metachromatic leukodystrophy	RDH13 RDH14 RDH12 RETSAT PLEKHG4	Olfactory Signaling Pathway
Diabetic ketoacidosis/ Hyperglycemia and ketosis	diabetic retinopathy diabetic	GRP APH1B OSBPL8 OSBPL5 NR4A3	Chemokine receptors bind chemokines
Gastrointestinal symptoms	hemoglobinuria oral hairy leukoplakia exanthem esophagitis chronic fatigue syndrome	WRAP53 OR4K5 DYNC112 WSB2 DYNC111	Mitotic Prometaphase
Hepatocellular injury/ hepatitis/ failure	exocrine pancreatic insufficiency biliary tract disease cholestasis pancreas disease	PRMT5 RHOT2 EIF5B ARL15 RHOD	RAB geranylgeranylation

	liver disease		
Renal/Acute kidney failure or injury	Fanconi syndrome prostate disease vasculogenic impotence renal artery obstruction IgA glomerulonephritis	GSN VPS41 WDR13 FKBP14 WDR33	Interleukin receptor SHC signaling
Sepsis	hepatitis D hepatitis A hepatitis E genital herpes pulmonary syndrome	TTLL10 RAB40C C16orf13 RHOT2 CHTF18	RAB geranylgeranylation
Bacteremia	hemoglobinuria exanthema chronic fatigue syndrome papilloma pulmonary coin lesion	RAB27A RAB27B AAGAB EHD4 RRAS	RAB geranylgeranylation
Dermatologic complications/pressure ulcer	pemphigus neurodermatitis diffuse scleroderma lichen planus acrodermatitis	DGKE DCAF8L2 WDR89 TLE1 CDC20	Calcineurin activates NFAT
Ocular symptoms	scleritis macular retinal edema cystoid macular edema pars planitis intermediate uveitis	ITGB2 ITGB3 ITGB1 ITGB6 ITGB7	Interleukin receptor SHC signaling



**Table 2.** Top 5 comorbidities, top 5 comorbidity enriched MOA proteins, and topmost relevant pathway for each COVID-19 severe adverse clinical manifestations predicted using GWAS input

Clinical manifestation	Comorbidities	Comorbidity enriched MOA proteins	Topmost relevant pathway
Respiratory	upper respiratory tract disease lower respiratory tract disease asthma adult respiratory distress syndrome allergic asthma	RGR GPR146 OR5AK2 DRD5 GPR119	Olfactory Signaling Pathway
Respiratory failure	respiratory failure respiratory	VEGFA LOC651959 LOC648044 PRRG3 FGF13	Olfactory Signaling Pathway
Acute respiratory distress syndrome (ARDS)	Adult respiratory distress syndrome	GRP PIGQ VDR ANXA1 TSPAN13	Olfactory Signaling Pathway
Asthma exacerbation	Asthma Allergic asthma Intrinsic asthma Status asthmaticus	GPR19 SLC30A6 BDNF OR8S1 GPR146	ADORA2B mediated anti-inflammatory cytokines production
Chronic obstructive pulmonary disease (COPD) exacerbation/Acute coronary syndromes	obstructive lung disease Dressler's syndrome chronic obstructive pulmonary disease	ADORA2B GPR50 TAAR6 GPR119 TAAR8	Olfactory Signaling Pathway
Cardiovascular/Arrhythmia	acute anterolateral myocardial infarction acute inferolateral myocardial infarction	F2RL2 OR51B6 LGR6 GPR25 APLNR	Olfactory Signaling Pathway receptors)

	acute inferoposterior infarction subendocardial infarction acute myocardial infarction strictly posterior acute myocardial infarction		
Acute myocardial infarction/Unstable angina	acute anterolateral myocardial infarction acute inferolateral myocardial infarction acute inferoposterior infarction subendocardial infarction acute myocardial infarction strictly posterior acute myocardial infarction	LGR6 GPR146 OR52D1 RXFP2 CCR9	ADORA2B mediated anti-inflammatory cytokines production
Acute congestive heart failure (CHF)	Congestive heart failure	TACR1 NPBWR1 NPBWR2 OR52B4 TACR3	Glutathione conjugation
Hematologic	methemoglobinemia sickle cell anemia myofascial pain syndrome hemorrhagic disease autoimmune lymphoproliferative syndrome	ACKR3 CCR9 OR2A1 ACKR2 LTB4R	Olfactory Signaling Pathway
Neurologic	sleep disorder congenital myasthenic syndrome facial neuralgia multiple system atrophy autonomic nervous system disease	ACKR2 XCR1 CCR9 GPR65 CCR6	Na <sup>+</sup> /Cl <sup>-</sup> dependent neurotransmitter transporters

Cerebral ischemia/infarction	arteriosclerosis obliterans cerebrovascular disease aortic aneurysm renal artery disease brain ischemia	F2RL2 AGTR1 GHSR NPBWR1 SUCNR1	G alpha (s) signaling events
Endocrine	obsolete Cushing's syndrome diabetes mellitus diabetic neuropathy hyperinsulinemic hypoglycemia nodular goiter	OR13A1 GPR4 PROKR2 F2RL2 NPBWR2	Peptide ligand-binding receptors
Gastrointestinal symptoms	alexia esophageal disease sexual dysfunction intestinal obstruction	GPR65 GPR132 OR51F2 GPR4 HCAR1	Olfactory Signaling Pathway
Hepatocellular injury/ Acute hepatitis/ liver failure	coronary artery disease hepatorenal syndrome acute pancreatitis alcoholic liver cirrhosis drug-induced hepatitis short bowel syndrome pancreatitis liver disease	GPR35 CCR9 CXCR2 NMBR GPR25	Peptide ligand-binding receptors
Renal/Acute kidney failure or injury	enterocele nephrosclerosis nephrolithiasis ureterolithiasis chronic interstitial cystitis	ACKR4 CXCR3 OR13C8 GPR174 FFAR2	Peptide ligand-binding receptors
Sepsis	bacterial sepsis fungal meningitis tinea unguium respiratory syncytial virus infectious disease	SLC6A15 SLC6A20 SLC6A18 SLC6A5 SLC6A19	Na <sup>+</sup> /Cl <sup>-</sup> -dependent neurotransmitter transporters

Bacteremia	alexia sexual dysfunction disease chronic fatigue syndrome papilloma	CCR9 CCR6 HCRTR2 NMBR HOMER3	Na <sup>+</sup> /Cl <sup>-</sup> - dependent neurotransmitt er transporters
Dermatologic complications/pres sure ulcer	carbuncle skin hemangioma dermatographia capillary hemangioma hemorrhoid	LGR6 OR52D1 OR52N2 PTGDR2 TAS2R41	Olfactory Signaling Pathway
Ocular symptoms	ptosis auditory system disease Marfan syndrome ocular hyperemia amblyopia	GPR50 OR52I2 CRHR1 AGTR1 PTGDR2	Olfactory Signaling Pathways

**Table 3.** Top 20 most frequent pathways across the interactome clinical manifestations.

<b>Pathway Top pathway</b>	<b>Frequency of clinical manifestations</b>
Association of TriC/CCT with target proteins during biosynthesis Metabolism of proteins	9
Interleukin receptor SHC signaling Immune System	7
TGFBR1 LBD Mutants in Cancer Disease	6
Synthesis of very long-chain fatty acyl-CoAs Metabolism	6
RHO GTPases Activate Formins Signal Transduction	6
Nuclear Receptor transcription pathway Gene expression	6
Elastic fibre formation Extracellular matrix organization	6
RAB geranylgeranylation Metabolism of proteins	5
N-Glycan antennae elongation Metabolism of proteins	5
Molecules associated with elastic fibres Extracellular matrix organization	5
Keratan sulfate/keratin metabolism Metabolism	5
Keratan sulfate biosynthesis Metabolism	5
Interleukin-4 and Interleukin-13 signaling Immune System	5
HSP90 chaperone cycle for steroid hormone receptors (SHR) Cellular responses to external stimuli	5
HSF1-dependent transactivation Cellular responses to external stimuli	5
Attenuation phase Cellular responses to external stimuli	5
Acyl chain remodeling of PS Metabolism	5
The NLRP1 inflammasome Immune system	4
Sodium/Calcium exchangers Transport of small molecules	4
Reduction of cytosolic Ca <sup>++</sup> levels Hemostasis	4

**Table 4.** Top 20 most frequent pathways across the GWAS clinical manifestations.

<b>Pathway Top pathway</b>	<b>Frequency of clinical manifestations</b>
Signaling by GPCR Signal Transduction	19
Peptide ligand-binding receptors Signal Transduction	19
GPCR downstream signaling Signal Transduction	19
G alpha (s) signaling events Signal Transduction	19
Class A/1 (Rhodopsin-like receptors) Signal Transduction	19
Olfactory Signaling Pathway Signal Transduction	18
GPCR ligand binding Signal Transduction	18
ADORA2B mediated anti-inflammatory cytokines production Disease	18
G alpha (q) signaling events Signal Transduction	16
Signal Transduction Signal Transduction	14
Leishmania parasite growth and survival Disease	14
G alpha (i) signaling events Signal Transduction	14
Anti-inflammatory response favoring Leishmania parasite infection Disease	14
Leishmania infection Disease	12
Amine ligand-binding receptors Signal Transduction	12
Adrenoceptors Signal Transduction	12
Relaxin receptors Signal Transduction	11
Chemokine receptors bind chemokines Signal Transduction	11
Tachykinin receptors bind tachykinins Signal Transduction	10
Glucagon-type ligand receptors Signal Transduction	10