

Comorbidities and Susceptibility to COVID-19: A Generalized Gene Set Meta-Analysis Approach

Beckman MF, Igba CK, Mougeot FB*, Mougeot J-LC*

Department of Oral Medicine, Carolinas Medical Center, Atrium Health, Charlotte, NC

Jean-Luc.Mougeot@atriumhealth.org

*Corresponding authors

ABSTRACT

Background

The COVID-19 pandemic has led to over 820,000 deaths for almost 24 million confirmed cases worldwide, as of August 27th, 2020, per WHO report. Risk factors include pre-existing conditions such as cancer, cardiovascular disease, diabetes, obesity, and cancer. There are currently no effective treatments. Our objective was to complete a meta-analysis to identify comorbidity-associated single nucleotide polymorphisms (SNPs), potentially conferring increased susceptibility to SARS-CoV-2 infection using a computational approach.

Results

SNP datasets were downloaded from publicly available GWAS catalog for 141 of 258 candidate COVID-19 comorbidities. Gene-level SNP analysis was performed to identify significant pathways by using MAGMA program. SNP annotation program was used to analyze MAGMA-identified genes. COVID-19 comorbidities from six disease categories were found to have significant associated pathways, which were validated by Q-Q plots ($p < 0.05$). The top 250 human mRNA gene expressions for SNP-affected pathways, extracted from publicly accessible gene expression profiles, were evaluated for significant pathways. Protein-protein interactions of identified differentially expressed genes, visualized with STRING program, were significant ($p < 0.05$). Gene interaction networks were found to be relevant to SARS and influenza pathogenesis.

Conclusion

Pathways potentially affected by or affecting SARS-CoV-2 infection were identified in underlying medical conditions likely to confer susceptibility and/or severity to COVID-19. Our findings have implications in COVID-19 treatment development.

Keywords: SARS-CoV-2, COVID-19, comorbidity, susceptibility, severity

INTRODUCTION

COVID-19 pandemic became prominent in Wuhan, China in December of 2019 [1]. As of August 27th, 2020, there have been over 24 million confirmed COVID-19 cases affecting over 200 countries [2]. This staggering number of cases includes more than 820,000 deaths, with the U.S. representing roughly one fourth of cases and deaths. WHO estimated the mortality rate of COVID-19 to be 3.4% in March 2020, which is significantly higher than reported average mortality rate of 0.1% for influenza [2].

COVID-19 is caused by severe acute respiratory syndrome coronavirus-2, SARS-CoV-2. This highly pathogenic coronavirus can cause severe respiratory illness and is highly contagious. The incubation period for SARS-CoV-2 can last up to 14 days with a median range of 4 to 5 days from exposure to onset of symptoms [3]. Transmission of the infection is due to inhalation of droplets and contact with contaminated surfaces. Symptoms include but are not limited to fever, cough, shortness of breath, fatigue, and body aches [4]. Current treatment for severe SARS-CoV-2 illness involves the management of complications associated with the disease and supportive care [1]. There is currently no treatment proven to be sufficiently effective for COVID-19 patients, although many drugs are being evaluated for effectiveness in reducing disease progression, severity, or mortality [5].

Multiple studies have shown that comorbidities influence severity, quality of life, or 1-year mortality, or altogether, in patients subject to viral infections [6, 7]. Risk factors for severity of SARS-CoV-2 infection include age 65 or above, and/or having a pre-existing condition [8, 9]. Thus, patients with pre-existing cardiovascular disease, diabetes, kidney dysfunction, obesity, and pulmonary diseases may have worse clinical outcomes when infected with SARS-CoV-2 [10].

Understanding pathways which could determine COVID-19 degree of susceptibility and severity is critical for drug development. Alternative drug combinations might act in synergy to complement future vaccination strategies, because no vaccination provides absolute protection [11]. Indeed, protection from vaccination against viral diseases may range from 9 to 90% and up to 60% of people vaccinated against influenza, for example, still fall ill due to the virus [12, 13]. In addition, vaccination is not systematically mandatory across countries.

Computational approaches may be utilized to identify candidate target molecules for the development of tailored drug treatments [14-16]. Insofar, drug targets can be predicted based on analysis of gene expression or genetic polymorphism profiles accessible through publicly available databases such as Gene Expression Omnibus (GEO) [17, 18]. Computational tools may (i) help elucidate or compare the mechanism(s) of action of drugs, (ii) assist in identification and characterization of interactions between a drug and its target, and (iii) provide a better understanding of mechanistic cellular reactions occurring at the molecular level in a drug response [19].

Computational approaches may also serve to investigate viral susceptibility and contribute to the development of vaccines [20-22]. For instance, machine learning has been used in the development of viral vaccines (*e.g.*, influenza) and in the investigation of genetic adaptation of the virus to the host [23]. Indeed, the genetic variability of SARS-CoV-2 will likely impact vaccine development in future outbreaks. Additionally, an appropriate computational model that accounts for the complexities of molecular interactions in COVID-19 patients affected by comorbidities, would allow for a predictive and better assessment of viral mechanisms and patients' response to drug or vaccine treatment.

The aim of this study was to complete a meta-analysis to identify genes associated with comorbidities/underlying medical conditions, potentially conferring increased susceptibility to SARS-CoV-2 infection or leading to the manifestation of more severe viral symptoms. To this end, we conducted a generalized gene set analysis using single nucleotide polymorphisms (SNPs) data from genome-wide association studies (GWAS) of a comprehensive list of possible comorbidities using Multi-marker Analysis of GenoMic Annotation (MAGMA)[24]. This analysis was complemented with (i) the investigation of predicted effects from the significant SNPs identified by MAGMA and (ii) the determination of differential human gene expression, most likely relevant to the pathogenesis of the viral respiratory illnesses, severe acute respiratory syndrome (SARS) and influenza.

METHODS

Multi-marker Analysis of GenoMic Annotation (MAGMA)

GWAS catalog and gene mapping

From an initial list of 258 mostly chronic diseases (data not shown), possibly representing comorbidities/underlying medical conditions associated with increased SARS-CoV-2 infectivity or disease severity, SNP datasets from the online GWAS catalog database [25] were identified using disease name, then downloaded and parsed. SNPs were mapped to genes using MAGMAv1.07b with the publicly available gene reference file NCBI37.3.gene.loc (<https://ctg.cncr.nl/software/magma>) [24].

Determination of multiple SNPs significance

The significance of SNPs (p -values) and derived sample sizes pertaining to genetic studies of comorbidities were extracted from the GWAS catalog datasets to compute correlations between neighboring genes and gene-level metrics *via* MAGMAv1.07b. To this end, the publicly available 1000 Genomes datasets (<https://ctg.cncr.nl/software/magma>) were used as reference files, considering the ethnicities associated with the possible comorbidity tested (European, East Asian, African, South American) [24]. To perform the gene set SNP analysis in MAGMAv1.07b, the 'ncol' flag was set to the sample size column in the SNP p -value file where each sample size corresponds to a p -value for the GWAS study completed. The flag for "multi=snp-wise" model was used to perform the 'mean' and 'top1' model analysis.

Pathway analysis using Enrichment Map and MAGMAv1.07b programs

Reactome analysis

To conduct the pathway analysis, Reactome[26] human pathways were downloaded from Enrichment Map program [27] using Entrez gene IDs [28-30]. Computationally derived Gene Ontology (GO) biological terms and "No Data" were excluded. Based on the significant SNPs in comorbidity gene sets per MAGMAv1.07b analysis, significant Reactome gene ontology defined pathways were identified. The significant pathways also underwent a per-gene analysis with the MAGMAv1.07b model flag set to 'alpha=0.05' [24].

Interaction networks

Visualization of protein-protein interaction networks was completed using STRINGv11.0 [31] program by testing different confidence levels to identify ontologies of biological significance for the significant pathways associated with comorbidities.

Quality control

Possible comorbidity significant associated gene sets/pathways were checked for quality control by generating Quantile-Quantile (Q-Q) plots using observed quantiles and residual Z-scores of genes within the gene set, based on the MAGMAv1.07b publicly available Rv3.6.2 script (posthoc_qc_107a.r) [32, 33].

Prediction of SNP effects

Ensembl's Variant Effect Predictor program (VEP) [34] was used to analyze MAGMAv1.07b annotation files for each gene set associated with comorbidities [35]. MAGMAv1.07b annotation files were converted into VEP format using a bash script. All converted annotation files were uploaded into VEP online tool separately. VEP summary statistics and analysis tables were downloaded for the comorbidities' associated genes and pathways found significant by MAGMAv1.07b. Corresponding tables were merged *via* Pythonv3.8.2 and SNPs containing a Sorting Intolerant from Tolerant (SIFT) score of 0 and a Polymorphism Phenotyping2 (PolyPhen2) score of 1, were removed (**Supplemental Data File 1**). Human Genome Organization (HUGO) gene symbols were extracted from the table with remaining SIFT and PolyPhen2 scores. Duplicate HUGO gene symbols were removed using Rv4.0.2. The most recently updated Affymetrix HG-U133A/B Human Genome Files [36, 37] containing annotated gene symbols and Entrez gene identifiers for all human genes were used to retrieve missing gene identification [38]. These tabular (.csv) files were merged and loaded into Rv4.0.2. Entrez gene IDs were matched to gene symbols from VEP analysis files to identify Affymetrix gene symbols. Genes and their corresponding Entrez ID's were then matched to significant genes' Entrez IDs found through combined MAGMAv1.07b - STRING analysis.

Transcriptional gene expression analysis

GEO2R [39] was used to test the top 250 human mRNA gene expressions for each comorbidity based on available human data using NCBI GEO[39], by only including comorbidities that had significant pathways identified by MAGMAv1.07b and VEP STRING analyses. Human mRNA expression datasets comparing disease group to healthy controls since 2010 were searched. If no datasets were available post-2010 the latest dataset was downloaded using characteristics described for prior datasets. For diseases with no publicly available datasets comparing healthy controls to disease type, the newest, most relevant dataset was used. Tissue types used for analysis included: (i) peripheral blood mononuclear cells (PBMCs), (ii) cancer tissues, (iii) adipose tissue, (iv) pulmonary tissue, (v) post-mortem brain tissue, (vi) cardiovascular tissue, and (vii) blood stem cells.

For each comorbidity, human mRNA gene expression data corresponding to average log-fold change (aLFC) were formatted for clustering of genes identified by MAGMAv1.07b and VEP and subsequently matched to STRING protein-protein interactions. Gene weights were added manually to account for duplicate genes in the dataset. Genes were mean centered and normalized. Hierarchical clustering was completed using a similarity metric of Manhattan city-block distance for genes and arrays with average linkage *via* Cluster3.0v1.59. Clusters were visualized using heatmaps created using JavaTreeViewv1.1.6r4 [40]. Clustered groups of genes for MAGMAv1.07b and VEP genes were run separately through GeneCodisv4.0 online tool [41] for identification of possible biological processes or pathways involved in viral infection [42].

Gene involvement in influenza and/or SARS

Significant genes (n=119) were investigated to determine their roles in relation to influenza and/or SARS respiratory viral infections. Genes were cross referenced using Pubmed [43] literature searches, DisGeNETv6 [44], Influenza Research Database[45] and conventional Google searches including HUGO gene symbol and either “influenza” or “SARS” [46, 47]. Risk of bias was assessed according to “Cochrane’s Handbook for Systematic Reviews of Interventions” [48]. Human tissue expression relevant to COVID-19 for genes with direct involvement was validated using Ensembl Expression Atlas [49, 50]. Genes not generally expressed in central nervous, cardiovascular, or pulmonary systems were removed from the dataset. Visualization of protein-protein interaction network of genes directly involved with influenza and SARS (caused by SARS-CoV-1) was completed using STRINGv11.0 using an interaction score of 0.400 [31].

RESULTS

The overall computational analytical design and associated primary results are presented in **Figure 1**.

MAGMA analysis of multiple SNPs associated with candidate COVID-19 comorbidities

To conduct generalized gene set analysis, we retrieved publicly available GWAS catalog datasets for 141 out of 258 COVID-19 possible comorbidities/underlying medical conditions. The 141 comorbidities were grouped into 8 categories by disease type based on organ most affected (**Table S1**). Following our MAGMA analysis (**Figure 1**: Flowchart section A), gene set and Reactome gene level analyses yielded 69 pathways representing 119 significant genes ($p < 0.05$). These pathways were significant for 22 COVID-19 comorbidities representing 6 disease categories, namely, cancer ($n=9$); cardiovascular ($n=4$); neurologic/mental ($n=3$); respiratory ($n=2$); skin/musculoskeletal ($n=1$); autoimmune/endocrine/metabolic ($n=3$). Reactome significant pathways and genes obtained through MAGMAv1.07b gene-level analysis from Enrichment Map are shown in **Tables 1a and b**.

Using STRINGv11.0 program with the highest confidence interaction score (CIS) of 0.9, processing of the 119 genes yielded a protein-protein interaction network of 70 genes, which was found to be highly significant based on hypergeometric test with Benjamini-Hochberg correction ($p = 4.36 \times 10^{-11}$) (**Figure 2a**). The top Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, identified by using STRINGv11.0, corresponded to Epstein-Barr virus infection with a false discovery rate of 6.72×10^{-9} .

Verification of significant pathways using Q-Q plots showed a high association between genes and their relative gene ontology defined pathways, since all plots show a distribution of residual z-scores deviating from the diagonal early on. There were no Q-Q plots with any ambiguous feature. Significant genes had high levels of association with each pathway. Q-Q plots of more than five genes, representing the pathways ontologies “post-translational protein modification”,

“translocation of ZAP-70 to immunological synapse”, “metabolism” and “cell cycle” and associated possible COVID-19 comorbidities (including asthma), are described in **Figure S1**.

VEP analysis of MAGMA-identified COVID-19 comorbidity-associated genes

Annotation files were converted for 134 of the 141 comorbidities with GWAS catalog datasets available (**Figure 1**: Flowchart section B). Of 3704 HUGO gene symbols extracted from VEP, 2996 corresponding Entrez gene IDs were identified using Affymetrix human genome annotation file. Of these gene IDs, 50 were matched with the 119 significant genes identified by MAGMAv1.07b for the 22 comorbidities with significant pathways (**Table 2**). Of the 50 genes, all were included in a protein-protein interaction network of 55 genes using a low CIS in STRINGv11.0 (**Figure 2b**). The top KEGG pathway identified using STRINGv11.0 was HTLV-1 infection with a false discovery rate of 4.38×10^{-7} using hypergeometric test with Benjamini-Hochberg correction.

Transcriptional gene expression analysis of MAGMA- and VEP-identified genes

GEO human mRNA expression datasets were retrieved for 19 of 22 comorbidities. A description of GEO datasets is presented in **Table S2**. Using 119 MAGMAv1.07b identified genes (**Figure 1**: Flowchart section A), JavaTreeViewv1.1.6r4 clustered 4 of 9 cancer types in the heatmap (partial view in **Figure 3a**, full heatmap in **Supplementary Image**). Also, interstitial lung disease, multiple sclerosis, asthma, obesity, and heart failure were clustered (**Figure 3a**). VEP STRING matched genes (n=50) also clustered 4 of 9 cancer types and clustered interstitial lung disease, multiple sclerosis, asthma, obesity, and heart failure together (**Figure 3b**). In both heatmaps Nucleoporin 160 (NUP160), Nucleoporin 153 (NUP153), Fibroblast Growth Factor Receptor 2 (FGFR2), and Karyopherin Subunit Beta 1 (KPNB1) showed lower aLFC expression in 12, 10, 6, and 10 out of 19 comorbidities, respectively (**Figures 3a and b**). GeneCodisv4.0 was able to confirm the four above mentioned genes, as well as others to be involved in lung or viral biological processes.

The 119 genes analyzed for gene expression were also investigated for their possible role in influenza and SARS-Cov-1 infection, as these might be relevant to SARS-Cov-2 infection. We identified three genes with a primary role in influenza infection: FGFR2, KPNB1 and NUP153

[51-54]. We also identified three genes KPNB1, Signal Transducer and Activator of Transcription 3 (STAT3), and Interleukin 2 Receptor Subunit Alpha (IL2RA) shown to play a significant role in SARS [55-57]. Genes identified as being possibly directly associated with influenza and/or SARS are shown in **Table S3**. STRING protein-protein interaction network yielded 38/46 (82.6%) genes involved in influenza and 15/17 (88.2%) genes involved in SARS, using an interaction score of 0.4 (**Figures 4a and b**). No GWAS study was found for SARS-CoV-1 infection to identify possible susceptibility genes within the 119 genes. Additionally, no studies were found to be at high risk for bias (**Table S4**).

DISCUSSION

This is the first study conducting generalized gene set analysis on a broad spectrum of possible COVID-19 comorbidities, with the prospect of identifying comorbidity-specific genes that could impact infection by SARS-Cov-2.

Starting with a list of 258 diseases, our MAGMA pipeline was able to identify 69 significant Reactome pathways with a total of 119 significant genes corresponding to 22 comorbidities that might have implications in predicting the severity of SARS-CoV-2 infection (**Figure 1, Table 1, Table S1**). Of the 22 comorbidities, we were able to validate pathways associated with cardiovascular disease, diabetes, obesity, and pulmonary diseases. Cardiovascular diseases identified included heart failure, atherosclerosis, Kawasaki's disease, and hypertension. Pulmonary diseases included asthma and interstitial lung disease. Cancer has been reported as a possible risk factor for COVID-19 [9]. We were able to identify nine cancers with GWAS data and significant associated pathways including acute myeloid leukemia, renal cell cancer, small cell lung cancer, and lung cancer. Furthermore, the known COVID-19 comorbidities, hypertension, obesity and diabetes had significant pathways and genes.

While Q-Q plots indicated validity of our findings, caution for interpretation of Q-Q plots must be used as these plots are normally used for pathways containing many genes. To a certain degree, these allow us to convey a certain level of confidence that there is a true association between gene and pathway [33]. In our analysis, however, less genes identified allowed us to narrow possible gene targets and pathways. Indeed, certain genes identified in our study may have significant biological relevance to infection by SARS-COV-2. For instance, sialyl

transferase ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3 (ST6GALNAC3) was found significant in the post-translational protein modification pathway (**Figure S1**). Another sialyl transferase, ST6GALNAC1, has been previously investigated as a drug target against infection of smooth airways epithelial cells by influenza virus [58]. It remains, however, to be determined whether ST6GALNAC3, generally expressed at high levels in renal cell cancer[59], plays a significant role in COVID-19 pathogenesis. Interestingly, aLFC gene expression of ST6GALNAC3 was positive in our analysis for 8 of 19 comorbidities (including renal cell cancer), namely, in tumor, pulmonary, brain, adipose tissues and PBMCs (**Figure 3a&b**).

STRINGv11.0 analysis produced significant enrichment for both MAGMAv1.07b genes and VEP matched genes containing SNPs that had characteristics of deleterious effects (**Table 2**). Therefore, we believe the interactions among genes from significant pathways from MAGMA and matched VEP genes are likely not due to chance and that these genes are biologically connected. Furthermore, STRINGv11.0 analyses identified top KEGG pathways including, Epstein-Barr virus pathway (MAGMA genes), and HTLV-1 pathway (VEP matched genes). STRING was able to cluster 70 genes into four functional groups among the 119 MAGMA significant genes: cell regulation and immune response, cell transport and nervous tissue function, protein homeostasis and gene expression, transcriptional regulation and RNA-mediated silencing (**Figure 2a**). Additionally, NUP160, NUP153, and KPNB1 clustered tightly together in the cell transport and nervous tissue function group.

STRINGv11.0 analysis of the 50 VEP matched genes with a lower confidence interval of 0.150 was required to obtain sufficient network connections for interpretation. Although network analysis may be subjective and is dependent on established knowledge, it is important to note that the enriched protein-protein interaction *p*-value was statistically significant. For the VEP matched gene STRINGv11.0 analysis, there were four distinct biological groupings recognized within the mapped network based on the closeness of protein interactions (**Figure 2b**). Those groupings were (i) antigen specific immune response, (ii) cell division and molecule formation/development, (iii) cell growth, survival, proliferation, motility, and morphology, (iv) and voltage gated ion channel transmembrane proteins. Notably, one of the comorbidities with significant associated pathways, breast cancer, contained SNPs affecting Solute Carrier Family 4 Member 7 (SLC4A7) and Solute Carrier Family 24 Member 3 (SLC24A3) genes. These genes are involved with sodium, calcium, and potassium ion transport and play a role in the malignant progression of breast cancer [60]. In addition, Euchromatic Histone Lysine Methyltransferase 2

(EHMT2) was mapped within close protein interactions. EHMT2 is involved with post-translational histone modification and epigenetic transcriptional repression. The orthologous gene (G9A) in drosophila is related to viral infection and susceptibility [61]. EHMT2 has been associated with the asthma comorbidity [62]. These results could indicate that EHMT2 deregulation possibly confers susceptibility to COVID-19.

The heatmap of MAGMA and VEP matched STRINGv11.0 genes shows that in 6 out of 19 comorbidities (**Figure 1: Flowchart section C**), FGFR2 had low aLFC (**Figure 3a and b**). Epithelial signaling by fibroblast growth factors is required for effective recovery from lung injuries resulting from influenza infection [51]. Our analysis coincides with previous findings linking induced inactivation of FGFR2 with increased mortality and influenza-induced lung injury [51]. KPNB1 (**Figure 2a**), NUP153, and NUP160 (**Figures 2a and 2b**) clustered in our heatmap analysis and were observed to have negative aLFC in majority (>50%) of possible comorbidities (**Figure 3a and b**). Downregulation of these factors under the condition of a comorbidity/underlying medical condition might impair cellular function in the presence of an active replicating virus.

Indeed, it has been previously established that a lower concentration of KPNB1 (interacting with NUP153, **Figure 2a**) in epithelial colorectal adenocarcinoma cells results in diminished Signal Transducer and Activator of Transcription 1 (STAT1) nuclear translocation activity [55]. Such regulation is critical for STAT1-activated genes to determine the level of anti-viral response and disease severity in an infection by SARS-CoV-1 [55]. Furthermore, KPNB1 is involved in the early stage of influenza virus replication *via* nuclear trafficking, by way of, nuclear import of viral cDNA or viral/host proteins into the host chromosome [52, 53].

Based on previous studies, the interaction between NUP153 and KPNB1 has been investigated in relation to nuclear transport [63]. The degradation of NUP153 in influenza virus A infected cells, such as Madin-Darby canine kidney II and human lung epithelial cells, results in an enlargement and widening of nuclear pores [54]. This disease process allows viral ribonucleoprotein complexes to be exported from the nucleus to the plasma membrane[54]. Additionally, NUP160 has been shown to work in conjunction with NUP153 to mediate nuclear import and export [64]. Therefore, degradation of one or both can prevent the import of signal transducers and activators of transcription, reducing effectiveness of the anti-viral interferon

response [65]. Our results support the interactions between these genes and viral respiratory diseases, such as influenza and SARS.

NUP153 showed negative aLFC of three cancer types (tumor tissues), three cardiovascular diseases (PBMCs and cardiovascular tissue), two respiratory diseases (pulmonary tissues), and two autoimmune/endocrine/metabolic diseases (PBMCs and adipose tissue). NUP160 showed lower aLFC in four cancer types (tumor tissues), two cardiovascular disease (PBMCs and cardiovascular tissues), two respiratory diseases (pulmonary tissues), two autoimmune /endocrine/metabolic diseases (PBMCs and adipose tissue), one mental health disease (post-mortem brain tissue), and one skin/musculoskeletal (PBMCs). KPNB1 showed low aLFC in four cancer types (tumor tissues), two respiratory diseases (pulmonary tissues), two mental health diseases (post-mortem brain tissue), one cardiovascular disease (PBMCs), and one autoimmune/endocrine/metabolic disease (adipose tissue).

Moreover, FGFR2 showed lower aLFC in four cancer types (tumor tissue), one respiratory disease (pulmonary tissues), and one cardiovascular disease (PBMCs) (**Figure 3a and b**). Notably, the possible comorbidities, asthma and prostate cancer, were identified as having all four genes (*i.e.*, FGFR2, KPNB1, NUP153, and NUP160) downregulated. Three of the four genes were downregulated in obesity (KPNB1, NUP153, NUP160), heart failure (FGFR2, KPNB1, NUP153), and interstitial lung disease (KPNB1, NUP153, NUP160) (**Figure 3a**). Only NUP153, and NUP160 were downregulated in hypertension and type 1 diabetes mellitus (**Figure 3a and b**).

GeneCodis was able to identify FGFR2 as being involved in mesenchymal cell differentiation involved in lung development while NUP153 and NUP160 are involved in viral replication process and intracellular transport of viruses. STAT3 was identified as being involved in primary miRNA binding and viral process and has been observed to be downregulated in SARS-CoV-1 infected Vero E6 kidney epithelial cells extracted from an African green monkey [56]. Additionally, IL2RA has been recently identified as significantly upregulated in the plasma of patients with severe COVID-19 [57] (**Figures 2a and 2b**). In our analysis, GeneCodis also identified Transporter 2, ATP binding Cassette Subfamily B Member (TAP2), Major Histocompatibility Complex, Class II, DR Beta 1 (HLA-DRB1), and Major Histocompatibility Complex, Class II, DQ Beta 1 (HLA-DQB1) as being involved in Epstein-Barr virus infection.

Further research is needed to confirm these genes (or associated regulations) as possible drug targets for SARS-CoV-2 infection.

Limitations

While there is no shortage of publicly available data, not all diseases have the same level of dedicated research. Therefore, not all possible comorbidities had publicly available SNP datasets from GWAS catalog or human mRNA gene expression datasets from NCBI's GEO datasets database. This resulted in a large decrease from 258 possible comorbidities to 141. Additionally, we were only able to use 19 of 22 significant comorbidities for GEO2R analysis and heatmap visualization. Another caveat is that GEO2R mRNA expression datasets have been generated through different independent studies using different genomic platforms and analysis pipelines, so that optimal normalization of raw data cannot be implemented. Little is still known about COVID-19 pathogenesis, although research on the matter has increased greatly since the beginning of the pandemic.

Conclusions

Significant pathways were identified associated with comorbidities/underlying medical conditions conferring susceptibility and/or severity to SARS-CoV-2 infection, which have been reported in conjunction with decreased clinical outcomes. Our findings may have implications in development of COVID-19 therapies.

SUPPLEMENTAL MATERIALS

Supplemental File 1: Figure S1 – COVID Q-Q plots

Supplemental File 2: Table S1 – COVID Comorbidities

Supplemental File 3: Table S2 – GEO dataset description

Supplemental File 4: Table S3 – Influenza/SARS involved genes

Supplemental File 5: Table S4 – Risk of bias assessment

Supplemental File 6: Supplementary Image – Full Heatmap Figure 3a

Supplemental File 7: Supplementary Data File 1.csv

ABBREVIATIONS

ADAMTS9: a disintegrin-like and metalloprotease with thrombospondin type 1 motif, 9; AHABE: adenocarcinoma human alveolar basal epithelial; aLFC: average log-fold change; A-MuLV: Abelson murine leukemia virus; C127: mouse fibroblast; CIS: confidence interaction score; COVID-19: coronavirus disease of 2019; CTL: cytotoxic T-lymphocyte; EHMT2: euchromatic histone lysine methyltransferase 2; FGFR2: fibroblast growth factor receptor 2; GEO: Gene Expression Omnibus; GIDEON: Global Infectious Disease and Epidemiology Network database; GO: Gene Ontology; GWAS: genome-wide association studies; HBE: human bronchial epithelial; HEK: human embryonic kidney; HLA-DRB1: major histocompatibility complex, class II, DR Beta 1; HLA-DRB5: major histocompatibility complex, class II, DR Beta 5; HLA-DQA2: major histocompatibility complex, class II, DQ Beta 1; HLA-DQB1: major histocompatibility complex, class II, DQ Beta 1; HTL: helper T-lymphocyte; HUGO: Human Genome Organization; IFI30: IFI30 lysosomal thiol reductase ; IL2RA: interleukin 2 receptor subunit alpha; KEGG: Kyoto Encyclopedia of Genes and Genomes; KPNB1: karyopherin subunit beta 1; MAGMA: Multi-marker Analysis of GenoMic Annotation; MDCK: madin-darby canine kidney; miRDB: microRNA Target Prediction Database; ND: No Data; NUP153: nucleoporin 153; NUP160: nucleoporin 160; PBMCs: peripheral blood mononuclear cells; PIK3R2: phosphoinositide-3-kinase regulatory subunit 2; PM: primary macrophages; PPARGC1A: peroxisome proliferator-activated receptor gamma coactivator 1 alpha; PolyPhen2: Polymorphism Phenotyping 2; Q-Q: quantile-quantile; SARS: severe acute respiratory syndrome; SARS-CoV-1: severe acute respiratory syndrome coronavirus-1; SARS-CoV-2: severe acute respiratory syndrome coronavirus-2; SIFT: Sorting Intolerant from Tolerant; SLC4A7: solute carrier family 4 member 7; SLC24A3: solute carrier family 24 member 3; SNPs: single nucleotide polymorphisms; ST6GALNAC1: ST6-acetylgalactosaminide alpha-2,6-sialyltransferase 1; ST6GALNAC3: ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3; STAT1: signal transducer and activator transcription 1; STAT3: signal transducer and activator of transcription 3; TAP2: transporter 2, ATP binding cassette subfamily B member; VEP: Variant Effect Predictor; Vero: epithelial kidney cells derived from African green monkey (*Chlorocebus spp.*); WHO: World Health Organization; ZAP-70: zeta-chain associated protein kinase 70

AUTHORS' CONTRIBUTIONS

Dr. Jean-Luc Mougeot and Dr. Farah Bahrani Mougeot conceived the study, contributed to the design of the analytical strategy, data interpretation and verification. Micaela Beckman designed the overall analytical strategy, conducted most computational analyses and data interpretation, and wrote the manuscript draft. Chica Igba contributed to the design of analytical strategy, conducted analyses and participated into data interpretation. All authors had significant contributions in writing and revisions of main manuscript, tables and figures.

ACKNOWLEDGEMENTS

We thank Kathleen Sullivan for her editorial expertise in preparation of this manuscript. This work was supported by the Atrium Health Foundation Research Fund (internal).

SOURCE OF FUNDING

N/A

AVAILABILITY OF DATA

Publicly available data were used. Analyzed data available upon request.

COMPETING INTERESTS

The authors declare having no competing interests to disclose.

CONSENT FOR PUBLICATION

All authors have approved the final version of this manuscript.

ETHICS APPROVAL

N/A

REFERENCES

References

1. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med.* 2020;382 18:1708-20.
doi:10.1056/NEJMoa2002032.
2. World Health Organization: Q&A: Influenza and COVID-19 - Similarities and Differences. www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-similarities-and-differences-covid-19-and-influenza#:~:text=Mortality for COVID-19, quality of health care. Accessed 1 Aug 2020.
3. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med.* 2020;172 9:577-82.
doi:10.7326/M20-0504.
4. Wang L, Wang Y, Ye D and Liu Q. Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence. *Int J Antimicrob Agents.* 2020;55 6:105948.
doi:10.1016/j.ijantimicag.2020.105948.
5. Cascella M, Rajnik M, Cuomo A, Dulebohn SC and Di Napoli R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). *StatPearls.* Treasure Island (FL); 2020.
6. Diaz E, Rodriguez A, Martin-Loeches I, Lorente L, Del Mar Martin M, Pozo JC, et al. Impact of obesity in patients infected with 2009 influenza A(H1N1). *Chest.* 2011;139 2:382-6. doi:10.1378/chest.10-1160.

7. Renu K, Prasanna PL and Valsala Gopalakrishnan A. Coronaviruses pathogenesis, comorbidities and multi-organ damage - A review. *Life Sci.* 2020;255:117839. doi:10.1016/j.lfs.2020.117839.
8. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J Infect Dis.* 2020;94:91-5. doi:10.1016/j.ijid.2020.03.017.
9. Robilotti EV, Babady NE, Mead PA, Rolling T, Perez-Johnston R, Bernardes M, et al. Determinants of Severity in Cancer Patients with COVID-19 Illness. *medRxiv.* 2020; doi:10.1101/2020.05.04.20086322.
10. Ji W, Huh K, Kang M, Hong J, Bae GH, Lee R, et al. Effect of Underlying Comorbidities on the Infection and Severity of COVID-19 in Korea: a Nationwide Case-Control Study. *J Korean Med Sci.* 2020;35 25:e237. doi:10.3346/jkms.2020.35.e237.
11. Vetter V, Denizer G, Friedland LR, Krishnan J and Shapiro M. Understanding modern-day vaccines: what you need to know. *Ann Med.* 2018;50 2:110-20. doi:10.1080/07853890.2017.1407035.
12. Centers for Disease Control and Prevention: Vaccine Effectiveness: How Well Do the Flu Vaccines Work? <https://www.cdc.gov/flu/vaccines-work/vaccineeffect.htm> (2020). Accessed 1 Jul 2020.
13. Ohmit SE, Victor JC, Teich ER, Truscon RK, Rotthoff JR, Newton DW, et al. Prevention of symptomatic seasonal influenza in 2005-2006 by inactivated and live attenuated vaccines. *J Infect Dis.* 2008;198 3:312-7. doi:10.1086/589885.
14. Villaveces JM, Koti P and Habermann BH. Tools for visualization and analysis of molecular networks, pathways, and -omics data. *Adv Appl Bioinform Chem.* 2015;8:11-22. doi:10.2147/AABC.S63534.

15. Hwang WC, Zhang A and Ramanathan M. Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery. *Clin Pharmacol Ther.* 2008;84 5:563-72. doi:10.1038/clpt.2008.129.
16. Zhao S and Li S. Network-based relating pharmacological and genomic spaces for drug target identification. *PLoS One.* 2010;5 7:e11764. doi:10.1371/journal.pone.0011764.
17. Kim J, Manivannan A, Kim DS, Lee ES and Lee HE. Transcriptome sequencing assisted discovery and computational analysis of novel SNPs associated with flowering in *Raphanus sativus* in-bred lines for marker-assisted backcross breeding. *Hortic Res.* 2019;6:120. doi:10.1038/s41438-019-0200-0.
18. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science.* 2006;313 5795:1929-35. doi:10.1126/science.1132939.
19. Dai YF and Zhao XM. A survey on the computational approaches to identify drug targets in the postgenomic era. *Biomed Res Int.* 2015;2015:239654. doi:10.1155/2015/239654.
20. Durmus S, Cakir T, Ozgur A and Guthke R. A review on computational systems biology of pathogen-host interactions. *Front Microbiol.* 2015;6:235. doi:10.3389/fmicb.2015.00235.
21. Brusica V and Petrovsky N. Immunoinformatics--the new kid in town. *Novartis Found Symp.* 2003;254:3-13; discussion -22, 98-101, 250-2.
22. Petrovsky N, Silva D and Brusica V. The future for computational modelling and prediction systems in clinical immunology. *Novartis Found Symp.* 2003;254:23-32; discussion 3-42, 98-101, 250-2.
23. Qiu X, Duvvuri VR and Bahl J. Computational Approaches and Challenges to Developing Universal Influenza Vaccines. *Vaccines (Basel).* 2019;7 2 doi:10.3390/vaccines7020045.

24. de Leeuw CA, Mooij JM, Heskes T and Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11 4:e1004219.
doi:10.1371/journal.pcbi.1004219.
25. GWAS catalog database <https://www.ebi.ac.uk/gwas/>. Accessed 11 May 2020.
26. Reactome Database <https://reactome.org>. Accessed 22 Jul 2020.
27. Enrichment map program, <http://baderlab.org/Software/EnrichmentMap> Accessed 22 Jul 2020.
28. Entrez gene IDs <http://www.ncbi.nlm.nih.gov/gene>. Accessed 20 Jun 2020.
29. Merico D, Isserlin R, Stueker O, Emili A and Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*. 2010;5 11:e13984. doi:10.1371/journal.pone.0013984.
30. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48 D1:D498-D503.
doi:10.1093/nar/gkz1031.
31. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47 D1:D607-D13. doi:10.1093/nar/gky1131.
32. Multi-marker analysis of genomic annotation, <https://ctg.cncr.nl/software/magma>
Accessed 20 Mar 2020.
33. de Leeuw CA, Stringer S, Dekkers IA, Heskes T and Posthuma D. Conditional and interaction gene-set analysis reveals novel functional pathways for blood pressure. *Nat Commun*. 2018;9 1:3768. doi:10.1038/s41467-018-06022-6.
34. Ensembl's variant effect predictor, Ensembl's Variant Effect Predictor.
<https://useast.ensembl.org/Tools/VEP>. Accessed 1 Aug 2020.

35. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17 1:122. doi:10.1186/s13059-016-0974-4.
36. Affymetrix HG-U133A Human Genome.
<http://www.affymetrix.com/Auth/analysis/downloads/na35/ivt/HG-U133A.na35.annot.csv.zip>.
37. Affymetrix HG-U133B Human Genome.
<http://www.affymetrix.com/Auth/analysis/downloads/na35/ivt/HG-U133B.na35.annot.csv.zip> .
38. Huang RS, Chen P, Wisel S, Duan S, Zhang W, Cook EH, et al. Population-specific GSTM1 copy number variation. *Hum Mol Genet.* 2009;18 2:366-72.
doi:10.1093/hmg/ddn345.
39. GEO2R, <https://www.ncbi.nlm.nih.gov/geo/geo2r> Accessed 7 May 2020.
40. Saldanha AJ. Java Treeview--extensible visualization of microarray data. *Bioinformatics.* 2004;20 17:3246-8. doi:10.1093/bioinformatics/bth349.
41. GeneCodisv4.0, <https://genecodis.genyo.es/> Accessed 24 Aug 2020.
42. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM and Pascual-Montano A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol.* 2007;8 1:R3. doi:10.1186/gb-2007-8-1-r3.
43. Pubmed <https://pubmed.ncbi.nlm.nih.gov/>. Accessed 1 Aug 2020.
44. DisGeNETv6 <https://www.disgenet.org>. Accessed 1 Aug 2020.
45. Influenza Research Database <https://www.fludb.org/brc/home.spg?decorator=influenza>
Accessed 1 Aug 2020.
46. Pinero J, Ramirez-Angueta JM, Sauch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 2020;48 D1:D845-D55. doi:10.1093/nar/gkz1021.

47. Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, et al. Influenza Research Database: An integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res.* 2017;45 D1:D466-D74. doi:10.1093/nar/gkw857.
48. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
49. Ensembl Expression Atlas, <https://www.ebi.ac.uk/gxa/home> Accessed 1 Aug 2020.
50. Papatheodorou I, Fonseca NA, Keays M, Tang YA, Barrera E, Bazant W, et al. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* 2018;46 D1:D246-D51. doi:10.1093/nar/gkx1158.
51. Dorry SJ, Ansbro BO, Ornitz DM, Mutlu GM and Guzy RD. FGFR2 Is Required for AEC2 Homeostasis and Survival after Bleomycin-induced Lung Injury. *Am J Respir Cell Mol Biol.* 2020;62 5:608-21. doi:10.1165/rcmb.2019-0079OC.
52. Konig R, Stertz S, Zhou Y, Inoue A, Hoffmann HH, Bhattacharyya S, et al. Human host factors required for influenza virus replication. *Nature.* 2010;463 7282:813-7. doi:10.1038/nature08699.
53. Mehle A and Doudna JA. A host of factors regulating influenza virus replication. *Viruses.* 2010;2 2:566-73. doi:10.3390/v2020566.
54. Muhlbauer D, Dzieciolowski J, Hardt M, Hocke A, Schierhorn KL, Mostafa A, et al. Influenza virus-induced caspase-dependent enlargement of nuclear pores promotes nuclear export of viral ribonucleoprotein complexes. *J Virol.* 2015;89 11:6009-21. doi:10.1128/JVI.03531-14.
55. Frieman M, Yount B, Heise M, Kopecky-Bromberg SA, Palese P and Baric RS. Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. *J Virol.* 2007;81 18:9812-24. doi:10.1128/JVI.01012-07.

56. Mizutani T, Fukushi S, Murakami M, Hirano T, Saijo M, Kurane I, et al. Tyrosine dephosphorylation of STAT3 in SARS coronavirus-infected Vero E6 cells. *FEBS Lett.* 2004;577 1-2:187-92. doi:10.1016/j.febslet.2004.10.005.
57. Jiang W, Wragg KM, Tan HX, Kelly HG, Wheatley AK, Kent SJ, et al. Identification of murine antigen-specific T follicular helper cells using an activation-induced marker assay. *J Immunol Methods.* 2019;467:48-57. doi:10.1016/j.jim.2019.02.008.
58. Wu D, Huang W, Wang Y, Guan W, Li R, Yang Z, et al. Gene silencing of beta-galactosamide alpha-2,6-sialyltransferase 1 inhibits human influenza virus infection of airway epithelial cells. *BMC Microbiol.* 2014;14:78. doi:10.1186/1471-2180-14-78.
59. Senda M, Ito A, Tsuchida A, Hagiwara T, Kaneda T, Nakamura Y, et al. Identification and expression of a sialyltransferase responsible for the synthesis of disialylgalactosylgloboside in normal and malignant kidney cells: downregulation of ST6GalNAc VI in renal cancers. *Biochem J.* 2007;402 3:459-70. doi:10.1042/BJ20061118.
60. Litan A and Langhans SA. Cancer as a channelopathy: ion channels and pumps in tumor development and progression. *Front Cell Neurosci.* 2015;9:86. doi:10.3389/fncel.2015.00086.
61. Merklings SH and van Rij RP. Analysis of resistance and tolerance to virus infection in *Drosophila*. *Nat Protoc.* 2015;10 7:1084-97. doi:10.1038/nprot.2015.071.
62. Stefanowicz D, Ullah J, Lee K, Shaheen F, Olumese E, Fishbane N, et al. Epigenetic modifying enzyme expression in asthmatic airway epithelial cells and fibroblasts. *BMC Pulm Med.* 2017;17 1:24. doi:10.1186/s12890-017-0371-0.
63. Ando Y, Tomaru Y, Morinaga A, Burroughs AM, Kawaji H, Kubosaki A, et al. Nuclear pore complex protein mediated nuclear localization of dicer protein in human cells. *PLoS One.* 2011;6 8:e23385. doi:10.1371/journal.pone.0023385.

64. Vasu S, Shah S, Orjalo A, Park M, Fischer WH and Forbes DJ. Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA export. *J Cell Biol.* 2001;155:339-54. doi:10.1083/jcb.200108007.
65. Le Sage V and Moulard AJ. Viral subversion of the nuclear pore complex. *Viruses.* 2013;5 8:2019-42. doi:10.3390/v5082019.

TABLES

Table 1. Reactome significant COVID-19 comorbidity-associated pathways and genes from EnrichmentMap program *via* MAGMAv1.07b

a. Gene-level analysis of significant pathways

Comorbidity ^a	R-HSA pathway ID ^b	Reactome pathways ^c	p-value min; max ^d	p-value median ^e
Acute myeloid leukemia ^f	597592	Post-translational protein modification	1.2E-4	1.2E-4
Asthma ^g	1222499; 75105; 881907; 202433; 389948; 75876; 202430	Fatty acid metabolism; Fatty acyl-CoA biosynthesis and synthesis of very long-chain fatty acyl-CoAs; Gastrin-CREB signaling pathway via PKC and MAPK; Generation of second messenger molecules; PD-1 signaling; Translocation of ZAP-70 to immunological synapse	2.72E-10; 6.93E-6	2.27E-7
Atherosclerosis ^h	112310; 181430; 425407; 112315; 425366; 382551	Neurotransmitter release cycle & Norepinephrine neurotransmitter release cycle; SLC-mediated transmembrane transport; Transmission across chemical synapses; transport of bile salts and organic acids, metal ions and amine compounds, transport of small molecules	6.32E-9; 4.18E-4	4.36E-6
Bipolar disorder ⁱ	983231	Factors involved in megakaryocyte development and platelet production	1.8E-6	1.8E-6
Breast cancer ^f	176814; 174048; 176409; 174143; 179419; 174048; 113507; 5687128; 176412; 176408; 453276; 425407; 425393	Activation of APC C and APC C: Cdc20 mediated degradation of mitotic proteins; Cyclin B; mitotic proteins; cell cycle proteins; cell cycle protein prior to satisfaction of cell cycle checkpoint; Phospho-APC C mediated degradation of Cyclin A; Phosphorylation and regulation of APC C between G1 S and early anaphase; E2F enabled inhibition of pre-replication complex formation; MAPK MAPK4 signaling; Regulation of mitotic cell cycle; SLC-mediated transmembrane transport; Transport of inorganic cations anions and amino acids oligopeptides	3.57E-11; 3.32E-5	1.34E-5
Colorectal cancer ^f	8875878; 6806834; 9006934	MET promotes cell motility; Signaling by MET and receptor tyrosine kinases	1.52E-4; 5.67E-4	3.6E-4
Heart failure ^h	382551	Transport of small molecules	4.77E-5	4.77E-5
Hypertension ^h	5576891; 397014; 6802957; 6802952	Cardiac conduction; Muscle contraction; Oncogenic MAPK signaling	1.67E-6; 3.03E-4	3.70E-5
Hypothyroidism ^j	1430728	Metabolism	3.08E-4	3.08E-4
Interstitial lung disease ^g	168249	Innate immune system	6.06E-6	6.06E-6
Kawasaki's disease ^h	1280218; 983169; 168256	Adaptive immune system & immune system; Class I MHC mediated antigen processing & presentation	8.02E-5; 5.06E-4	2.93E-4
Lung cancer ^f	1483257	Phospholipid metabolism	1.06E-4	1.06E-4
Multiple sclerosis ^k	1280215	Cytokine signaling in immune system	4.86E-5	4.86E-5
Obesity ^l	422475; 204998; 73887; 1266738; 416482; 9675108; 193648; 193704; 194840; 194315	Axon guidance; Cell death signaling via NRAGE, NRIF, and NADE; Death receptor signaling; Developmental biology; G alpha (12 13) signaling events; Nervous system development; NRAGE signals death through JNK; P75 NTR receptor-mediated signaling; Rho GTPase cycle; Signaling by Rho GTPases	5.78E-7; 1.42E-4	6.44E-7
Ovarian cancer ^f	1483257	Phospholipid metabolism	9.76E-7	9.76E-7
Pancreatic cancer ^f	1226099	Signaling by FGFR in disease	2.4E-4	2.4E-4
Prostate cancer ^f	556833; 1483255; 1660516	Metabolism of lipids, PI; Synthesis of PIPs at the early endosome membrane	1.78E-5; 8.64E-5	5.21E-5
Renal cell cancer ^f	109582	Hemostasis	1.71E-3	1.71E-3
Schizophrenia ⁱ	2559583; 2559586	Cellular senescence; DNA damage telomere stress induced senescence	1.16E-6; 2.07E-6	1.61E-6
Small cell lung cancer ^f	8953897; 2262752	Cellular responses to external stimuli & stress	1.05E-3	1.05E-3
Type 1 diabetes mellitus ^l	4086398; 9607240; 5683057; 5673001; 8878171	ERK1 ERK2 pathway; FLT3 signaling, MAPK family signaling cascades; RAF MAP kinase cascade; Transcriptional regulation by RUNX1	2.77E-4	2.77E-4
Unipolar depression ^l	1640170	Cell cycle	9.12E-5	9.12E-5

b. Gene-level analysis of significant genes

Comorbidity ^a	Entrez gene ID/s ^b	Gene symbol ^c	p-value min; max ^d	p-value median ^e
Acute myeloid leukemia ^f	3065; 256435; 51377; 6670; 5468; 57599; 56999; 10891; 4306; 9972; 1780; 55958; 23287; 1075; 84259; 50863; 4287; 123624; 641; 9491; 7109	HDAC1; ST6GALNAC3; UCHL5; SP3; PPARG; WDR48; PPARGC1A; NR3C2 ; NUP153 ; KLHL9; AGTPBP1 ; CTSC; DCUN1D5; NTM ; ATXN3; AGLB1; BLM; PSMF1; TRAPPC10	1.53E-22; 3.00E-6	1.70E-10
Asthma ^g	55289; 2181; 79993; 47 2181; 79993; 47 2520; 5578; 6196 3122; 3127; 3117; 3119; 3120 2181; 79993	ACOXL ; ACSL3; ELOVL7; ACLY ACSL3; ELOVL7; ACLY GAST; PRKCA ; RPS6KA2 HLA-DRA; HLA-DRB5 ; HLA-DQA1 ; HLA-DQB1; HLA-DQB2 ACSL3; ELOVL7	6.26E-41; 4.00E-7 1E-50; 6.41E-9 1E-50; 6.41E-9 1E-50; 6.41E-9	3.94E-24 1.69E-10 3.21E-9 4.72E-24
Atherosclerosis ^h	6580; 6857 6580; 6564; 23446 3773; 5577; 6580; 6857 6580; 23446 6580; 23446; 23457; 5577; 5664 11311; 23046; 26153; 25970	SLC22A1 ; SYT1 SLC22A1 ; SLC15A1; SLC44A1 KCNJ16; PRKAR2B; SLC22A1 ; SYT1 SLC44A1; SLC22A1 SLC44A1; SLC22A1 ; ABCB9; PRKAR2B; SLC15A1 VPS45 ; KIF21B ; KIF26A; SH2B1	2E-43; 1E-9 2E-43; 2.67E-6 2E-43; 1E-9 2E-43; 2.67E-6 2E-43; 2.67E-6	5E-10 7E-7 1.2E-10 1.33E-6 4E-7
Bipolar disorder ⁱ	64682; 983 983; 23112 64682; 983 23446; 57419; 9497 57419; 9497	ANAPC1; CDK1 CDK1; TNRC6B ANAPC1; CDK1 SLC44A1; SLC24A3 ; SLC4A7 SLC24A3 ; SLC4A7	1E-24; 2E-6 1E-50; 3E-8 1E-50; 9.99E-35 1E-50; 2E-9	1E-6 1.5E-8 5E-35 1E-9
Breast cancer ^f	23446; 57419; 9497 57419; 9497	SLC44A1; SLC24A3 ; SLC4A7 SLC24A3 ; SLC4A7	9.58E-51; 9E-6 3E-45; 9E-6	3E-45 4.5E-6
Colorectal cancer ^f	3915; 64759 3915; 8936; 64759	LAMC1 ; TNS3 LAMC1 ; WASF1; TNS3	6.36E-14; 2E-11 6.36E-14; 1E-6	1E-11 2E-11
Heart failure ^h	6570; 366	SLC18A1; AQP9	2E-44; 2E-35	1E-35
Hypertension ^h	57085; 27044 3752; 776; 783; 4633 3752; 766; 783 57085; 200734; 27044	AGTRAP; SND1 KCND3 ; CACNA1D ; CACNB2 ; MYL2 KCND3 ; CACNA1D ; CACNB2 AGTRAP; SPRED2; SND1	4E-34; 5E-7 1E-21; 7E-12 1E-21; 1.31E-15 4E-34; 5E-7	2.5E-7 6.59E-16 1.08E-17 2E-7
Hypothyroidism ^l	1213; 26275; 2131; 960; 8898; 113; 5296	CLTC; HIBCH; EXT1 ; CD44; MTMR2; ADCY7 ; PIK3R2	3E-39; 3E-10	2E-17
Interstitial lung disease ^g	4583; 54472	MUC2 ; TOLLIP	7E-34; 4.45E-13	2.23E-13
Kawasaki's disease ^h	55521; 6891; 208	TRIM36; TAP2 ; AKT2	5E-11; 2E-8	4E-10
Lung Cancer ^f	55521; 6981	TRIM36; TAP2	5E-11; 2E-8	1E-8
Multiple sclerosis ^k	374986; 79888; 22876 942; 5602; 3575; 3123; 3119	MIGA1; LPCAT1 ; INPP5F CD86 ; MAPK10 ; IL7R ; HLA-DRB1 ; HLA-DQB1	8E-35; 9E-6 6.08E-24; 1E-11	4E-7 5E-20
Obesity ^l	25791; 5924 8648; 25791; 5915; 57698 25791; 57698 25791; 1062; 10788; 5924	NGEF ; RASGRF2 NCOA1 ; NGEF ; RARB ; SHTN1 NGEF ; SHTN1 NGET; CENPE; IQGAP2 ; RASGRF2	1E-50; 5E-6 1E-50; 8E-6 1E-50; 8E-6 1E-50; 8E-6	2.5E-6 2E-6 4E-6 2.5E-6
Ovarian cancer ^f	114884; 22876	OSBPL10 ; INPP5F	8E-35; 2E-6	1E-6
Pancreatic cancer ^f	2263; 6776; 6774	FGFR2 ; STAT5A; STAT3	1E-50; 7E-6	1E-6
Prostate cancer ^f	22876; 55697 2629; 22876; 55697; 83394; 8714	INPP5F; VAC14 GBA; INPP5F; VAC14 ; PITPNM3; ABCC3	8E-35; 2E-8 1E-50; 2E-6	1E-8 2E-8
Renal cell cancer ^f	5581; 8793	PRKCE ; TNFRSF10D	1.5E-25; 6E-9	3E-9
Schizophrenia ^l	8294; 8341; 3009 192669; 8294; 8341; 3009; 10919	HIST1H4I; HIST1H2BN; HIST1H1B AGO3; HIST1H4I; HIST1H2BN; HIST1H1B; EHMT2	5E-27; 2E-21 5E-27; 2E-6	9E-27 3.51E-19
Small cell lung cancer ^f	10919; 55466	EHMT2 ; DNAJA4	5E-21; 5E-6	2.5E-6
Type 1 diabetes mellitus ^l	10213; 3559	PSMD14; IL2RA	3.71E-31; 4E-18	2E-18
Unipolar depression ^l	64326; 5500; 8347; 8294; 8348; 23345; 8379; 11064; 8945; 5702; 23279; 8655; 91750; 3837	RFWD2; PPP1CB; HIST1H2BC; HIST1H4I; HIST1H2BO; SYNE1 ; MAD1L1 ; CNTRL; BTRC; PSMC3; NUP160 ; DYNLL1; LIN52; KPNB1	4E-25; 7E-6	6.75E-11

Footnote:

a. Gene-level analysis of significant pathways

^bReactome pathway unique identifier

^cReactome pathway name

b. Gene-level analysis of significant genes

^bEntrez gene unique identifiers found to be significant in Reactome pathways are shown

^cGene symbol identified using DAVIDv6.8 (<https://david.ncifcrf.gov/tools.jsp>) from Entrez gene ID

(**a,b**). ^aSARS-CoV-2 infection-related possible comorbidity name having significant pathways and genes analyzed using MAGMAv1.07b (n=22)

^dp-value minimum and maximum

^ep-value median

Comorbidities include: ^fCancer (non-head and neck cancer) group (n=9), ^gRespiratory group (n=2), ^hCardiovascular/blood group (n=4), ⁱNeurologic/mental group (n=3), ^jAutoimmune/metabolic/ endocrine group (n=3), and ^kSkin/musculoskeletal group (n=1)

***Note:** Bolded gene symbols (n=50) represent matched genes from VEP analysis to significant genes found through MAGMAv1.07b analysis; Synonymous gene symbols include HLA-DRB1 with HLA-DRB5; HLA-DQA1 with HLA-DQA2; ADAMTS9 with PPARGC1A; IFI30 with PIK3R2.

Table 2. VEP genes and SNPs matched to significant MAGMAv1.07b COVID-19 comorbidity-associated genes

Comorbidity ^a	Entrez gene ID ^b	Gene symbol ^c	Variant ID (rs#) ^d	Consequence ^e
Acute myeloid leukemia	2263; 5602; 55289; 56999; 9972; 50863	FGFR2; MAPK10; ACOXL; ADAMTS9; NUP153; NTM	7090018, 2912759; 6838659; 4640633; 17524344; 4849120; 4849121; 13395354; 9868005; 13095235; 4371513; 4605539; 11714364; 9851598; 4716165; 4716167; 10949435; 2274136; 9383307; 6906499; 9350055; 9396787; 10949436; 1006066; 11753865; 16879902; 12199222; 11222631; 11222631; 11222647; 12278021; 7107326; 11222652; 11222653; 992564; 12419920; 12575010; 4937627 12622534; 281508; 7717955; 6881270; 114798579; 146668528; 9272105; 3104369; 3104367; 9272346; 9270911; 2760995; 7760841; 4713555; 3997868;	IV; NMD; NC-T DGV; NC-TV; 3prime; MS
Asthma Atherosclerosis	5581; 3575; 3117; 3123; 6891; 3118; 10919; 56999 114884; 50863	PRKCE; IL7R; HLA-DQA1; HLA-DRB1; TAP2; HLA-DQA2; EHMT2; ADAMTS9; OSBPL10; NTM	151027268; 3104369; 3104367; 9272346; 41267086; 9866261 1902341; 11827555 778353; 2592118; 7071123; 3888190; 1203233; 17082664; 9371601; 7747960; 4523096; 4236274; 10275045; 4332037; 12668848; 3931398; 4721295; 1107592; 9895770; 2297909	IV; IV,NC-TV; DGV; UGV; IV,NMD; 3prime; NC-EV IV; IV,NC-TV
Bipolar disorder	25791; 783; 25970; 23345; 8379; 5578; 23046	NGEF; CACNB2; SH2B1; SYNE1; MAD1L1; PRKCA; KIF21B	4973768; 7619833; 113118767; 77674461 17550549; 6678517; 4546885; 10911251; 3801081; 113118767; 11200014	IV; IV,NC-TV; IV,NMD; DGV; UGV 3prime; DGV; IV,NC- TV; IV; IV,NMD
Breast cancer	9497; 57419; 23287	SLC4A7; SLC24A3; AGTPBP1	113118767; 11200014	IV; IV,NC-TV; DGV; IV,NMD
Colorectal cancer	4633; 3915; 64759; 57419; 2263	MYL2; LAMC1; TNS3; SLC24A3; FGFR2	192154334	IV; DGV
Heart failure	64759	TNS3	3774427; 12715461; 9814480; 12258967; 4815879	IV; IV,NC-TV; IV,NMD IV; DGV; MS,NMD; NC-EV; UGV
Hypertension	776; 783; 84515	CACNA1D; CACNB2; MCM8	78534766	IV; IV,NC-TV; UGV; IV,NMD
Hypothyroidism	113	ADCY7	5743894; 5743890 4406174; 62140840; 11902506; 6710503	IV; IV,NMD; IV,NC-TV
Interstitial lung disease	54472	TOLLIP	11554159; 2293152; 6897932; 10063294; 6881706; 2681424; 3104373; 2040406; 72665771; 3104373; 2040406; 2104286; 3118470; 12722489	DGV; IV; UGV; MS; NC-EV; IV,NMD; 3prime; IV,NC-TV DGV; MS; NC-EV IV,NC-TV; UGV; IV,NC-TV; IV
Lung cancer	79888; 8648	LPCAT1; NCOA1 PIK3R2; STAT3; IL7R; CD86; HLA-DQA1; MAPK10; HLA- DQA2; IL2RA	875858; 56232506; 2788612 651164; 4646284; 527510716; 11704416; 9623117; 58133635; 12628051; 4821941; 11200014 778371; 778353; 2944591 6897932; 9272346; 9272346	IV; IV,NC-TV; IV,NMD; DGV; UGV DGV; IV; UGV MS; IV; UGV; NC-EV; NMD; NC-TV
Multiple sclerosis	5296; 6774; 3575; 942; 3117; 5602; 3118; 3559	PIK3R2; IFI30; SH2B1; RARB	11554159; 7498665; 1435703 28568660 73328514	IV; IV,NC-TV; IV IV; IV,NC-TV; DGV IV
Obesity	5296; 10437; 25970; 5915	PIK3R2; IFI30; SH2B1; RARB	11554159; 7498665; 1435703 28568660 73328514	IV; IV,NC-TV; IV IV; IV,NC-TV; DGV IV
Ovarian cancer	114884	OSBPL10	28568660	IV; IV,NC-TV; DGV IV
Pancreatic cancer	64759	TNS3	73328514	IV; IV,NC-TV; IV IV; IV,NC-TV; DGV IV
Prostate cancer	55697; 64759; 3752; 6580; 8379; 23112; 2263	VAC14; TNS3; KCND3; SLC22A1; MAD1L1; TNRC6B; FGFR2	875858; 56232506; 2788612 651164; 4646284; 527510716; 11704416; 9623117; 58133635; 12628051; 4821941; 11200014 778371; 778353; 2944591 6897932; 9272346; 9272346	IV; IV,NC-TV; IV,NMD; DGV; UGV DGV; IV; UGV MS; IV; UGV; NC-EV; NMD; NC-TV
Schizophrenia	25791	NGEF	778371; 778353; 2944591 6897932; 9272346; 9272346	DGV; IV; UGV MS; IV; UGV; NC-EV; NMD; NC-TV
Type 1 diabetes mellitus	3575; 3117; 3118; 3559	IL7R; HLA-DQA1; HLA- DQA2; IL2RA	61839660; 12722495; 706778; 10795791	IV; IV,NC-TV; IV,NMD; DGV; UGV DGV; IV; UGV MS; IV; UGV; NC-EV; NMD; NC-TV
Unipolar depression	25791; 783; 3123; 23345; 2131; 23279; 8379; 23046	NGEF; CACNB2; HLA-DRB1; SYNE1; EXT1; NUP160; MAD1L1; KIF21B;	778353; 2799573; 7071123; 535777; 17082664; 9371601; 17506336; 11039409; 12668848; 1107592; 11514731; 2056477; 56072378; 3823624; 2297909	IV; NMD; NC-TV; UGV; DGV

Footnote:

Total of 54 gene IDs from STRINGv11.0 analysis (50 genes after synonyms processed) were matched with significant genes from 22 comorbidities with significant pathways identified by MAGMAv1.07b.

^aVariant effect predictor (VEP) analysis of significant MAGMAv1.07b comorbidities

^bEntrez gene unique identifiers found to be significant from STRINGv11.0 matched gene analysis

^cHuman Genome Organisation (HUGO) gene symbol identified using most recently updated Affymetrix HG-U133A/B Human Genome Files

(<http://www.affymetrix.com/Auth/analysis/downloads/na35/ivt/HG-U133A.na35.annot.csv.zip>,

<http://www.affymetrix.com/Auth/analysis/downloads/na35/ivt/HG-U133B.na35.annot.csv.zip>)

^dSingle nucleotide polymorphism (SNP) identifiers of upload variant numbers from MAGMAv1.07b and VEP analysis (rs number)

^eConsequence of SNP variants on sequence; **IV** – intron variant, **NMD** – nonsense-mediated decay transcript variant, **NC-T** – noncoding transcript, **NC-TV** – noncoding transcript variant, **3prime** – 3 prime untranslated region variant, **DGV** – downstream gene variant, **UGV** – upstream gene variant, **NC-EV** – noncoding exon variant, **IV-NC** – intron variant, non-coding transcript, **MS** – missense variant

***Note:** Comorbidities without significantly matched genes include mucocutaneous lymph node syndrome, renal cell cancer, and small cell lung cancer.

FIGURE LEGENDS

Figure 1. Computational analytical design for determination of genes/pathways associated with comorbidities, possibly contributing to COVID-19 severity/infectivity

Legend.

Flowchart section A.

A list of candidate comorbidities (n=258) possibly associated with increased severity/infectivity of COVID-19 were curated. SNPs associated with comorbidities with available GWAS catalogdata (n=141) were analyzed. Multi-marker Analysis of GenoMic Annotation (MAGMA) was performed. SNPs were annotated to genes using NCBI gene reference file (NCBI37.3.gene.loc). In MAGMAv1.07b, gene set/pathway analysis was performed for which each SNP that was identified, using the “multi-mean=snp-wise” model-generated results, taking into account ethnicities associated with a possible comorbidity. Gene-level analysis was completed using Reactome pathways retrieved from Enrichment Map program. STRINGv11.0 protein-protein interaction program was used to visualize the network of 119 significant genes. Quantile-quantile (Q-Q) plots in Rv3.4.2 for 69 significant pathways were used for quality control. NCBI-gene expression omnibus (GEO) human mRNA differential expression datasets were downloaded *via* GEO2R for each comorbidity with associated genes/ pathways (n=19 of 22). Human mRNA expression was visualized with a heatmap of the 119 significant genes using Cluster3.0v1.59 and JavaTreeViewv1.1.6r4. Tissue expression relevance to SARS and influenza was determined using DisGeNETv6 and Ensembl Expression Atlas databases.

Flowchart section B.

MAGMAv1.07b annotation files were converted for Ensembl Variant Effect Predictor (VEP) format (n=134 of the 141 GWAS datasets). Gene symbols (n=3704) were extracted for VEP analysis from 22 significant comorbidity-associated genes/pathways per MAGMAv1.07b analysis. Entrez gene IDs (n=2996) were matched to gene symbols using Affymetrix gene symbols annotation files (HG-U133A/B Human Genome Files). STRINGv1.0 protein-protein interaction program was used to visualize the network (n=50 genes).

Flowchart section C.

NCBI-Gene expression omnibus (GEO) human mRNA differential expression datasets were downloaded *via* GEO2R for each of the 22 comorbidities with associated genes/pathways (n=19 of 22). VEP genes were matched to network genes and formatted for Cluster3.0v1.59. Human mRNA expression was visualized with a heatmap using average log-fold changes (aLFC) in JavaTreeViewv1.1.6r4 (*i.e.*, 50 VEP identified genes matched to 119 MAGMA identified genes).

Figure 2. STRING protein-protein interaction network for significant COVID-19 comorbidity-associated genes

- a.** *MAGMA significant genes (CIS=0.9)*
- b.** *Matched MAGMA and VEP significant genes (CIS=0.15)*

Legend.

a. Significant genes were identified via MAGMAv1.07b gene-level analysis (<https://ctg.cncr.nl/software/magma>) using Reactome human pathways (<https://reactome.org>) obtained from Enrichment Map (<http://baderlab.org/Software/EnrichmentMap>) for 22 COVID-19 possible comorbidities with 119 significant genes. In STRINGv11.0 program, the confidence interaction score (CIS) was set to the maximum of 0.9, resulting in a network of 70 connected genes. Clustering by biological functions is represented by outlines. **Green (...):** cell regulation and immune response. **Red (___):** cell transport and nervous tissue function. **Blue (--):** protein homeostasis and gene expression. **Orange (-.-):** transcriptional regulation and RNA-mediated silencing.

b. Significant genes (n=50) were identified for 22 COVID-19 possible comorbidities in both VEP SNP and MAGMAv1.07b gene-level analyses. CIS was set to low value of 0.150, which resulted in integration of all 50 genes in a network of 55 genes. Clustering by biological functions is outlined. **Red (___):** antigen specific immune response. **Blue (--):** cell division and molecule formation/development. **Green (--.-):** cell growth, survival, proliferation, motility, and morphology. **Purple (...):** voltage gated ion channel transmembrane proteins.

Note: Three gene symbols had other synonymous gene symbols (HLA-DRB1 and HLA-DRB5, HLA-DQA1 and HLA-DQA2, PIK3R2 and IFI30), which were also entered into STRINGv11.0 program for verification.

Figure 3. Human gene expression heatmap of MAGMA and VEP genes of COVID-19 associated comorbidities

- a.** Heatmap of MAGMAv1.07b significant genes (partial view)
- b.** Heatmap of VEP matched genes

Legend.

Heatmap from GEO2R human gene expression datasets using average log-fold changes (aLFC) from **(a)** 119 significant genes ($p < 0.05$) identified *via* MAGMAv1.07b and **(b)** 50 genes matched to MAGMAv1.07b 120 significant genes from Ensembl Variant Effect Predictor (VEP) with 19 of 22 significant comorbidities identified *via* MAGMAv1.07b as arrays **(a, b)**.

Genes were filtered by removing those with less than 60% values present, mean centered and normalized. Hierarchical clustering was completed using weights for duplicate/synonymous genes with a similarity metric of city-block (Manhattan) distance for genes and arrays with average linkage through Cluster3.0v1.59 software. Heatmaps were created using JavaTreeView1.1.6r4. Yellow depicts positive aLFC, blue depicts negative aLFC, black depicts missing data and values of zero.

Figure 4. STRING protein-protein interaction of COVID-19 comorbidity-associated genes with involvement in influenza and/or SARS

- a.** Influenza (CIS=0.4)
- b.** SARS (CIS=0.4)

Legend.

STRING protein-protein interactions of MAGMAv1.07b identified genes with direct involvement with **(a)** influenza (n=46) and/ or **(b)** SARS (n=17) are shown. Level of stringency in STRINGv11.0 program was set to a medium confidence interaction score (CIS) of 0.4 in both influenza and SARS related molecular networks **(a, b)**, resulting in a cluster of 38/46 (82.6%) and 15/17 (88.2%) genes, respectively.













