Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Personalized Medicine

Investigating underlying human immunity genes, implicated diseases and their relationship to COVID-19

Zeeshan Ahmed*[1,2], Eduard Gibert Renart[1] & Saman Zeeshan[3]

[1]Rutgers Institute for Health, Health Care Policy & Aging Research, Rutgers University, 112 Paterson Street, New Brunswick, NJ 08901, USA

[2]Department of Medicine, Robert Wood Johnson Medical School, Rutgers Biomedical & Health Sciences, 125 Paterson Street, New Brunswick, NJ 08901, USA

[3]Rutgers Cancer Institute of New Jersey, Rutgers University, 195 Little Albany St, New Brunswick, NJ 08901, USA

*Author for correspondence: zahmed@ifh.rutgers.edu

Aim: A human immunogenetics variation study was conducted in samples collected from diverse COVID-19 populations. **Materials & methods:** Whole-genome and whole-exome sequencing (WGS/WES), data processing, analysis and visualization pipeline were applied to identify variants associated with genes of interest. **Results:** A total of 2886 mutations were found across the entire set of 13 genomes. Functional annotation of the gene variants revealed mutation type and protein change. Many variants were found to be biologically implicated in COVID-19. The involvement of these genes was also found in multiple other diseases. **Conclusion:** The analysis determined that *ACE2*, *TMPRSS4*, *TMPRSS2*, *SLC6A20* and *FYCOI* had functional implications and *TMPRSS4* was the gene most altered in virally infected patients.

Plain language summary: The quest to establish an understanding of the genetics underlying COVID-19 is a central focus of life sciences today. COVID-19 is triggered by SARS-CoV-2, a single-stranded RNA respiratory virus. Several clinical-genomics studies have emerged positing different human gene mutations occurring due to COVID-19. A global analysis of these genes was conducted targeting major components of the immune system to identify possible variations likely to be involved in COVID-19 predisposition. Gene-variant analysis was performed on whole-genome sequencing samples collected from diverse populations. *ACE2*, *TMPRSS4*, *TMPRSS2*, *SLC6A20* and *FYCOI* were found to have functional implications and *TMPRSS4* may have a role in the severity of clinical manifestations of COVID-19.

First draft submitted: 5 October 2021; Accepted for publication: 8 February 2022; Published online: 9 March 2022

Keywords: COVID-19 • disease • gene • immunity • phenotypes • variant

Viruses, bacteria, fungi and parasites elicit the emergence of infectious diseases that have the potential to significantly impact the world [1]. These re-emerging infectious diseases include the plague, syphilis, tuberculosis, smallpox, yellow fever, cholera, Spanish flu, Ebola, acute hemorrhagic conjunctivitis, HIV/AIDS, severe acute respiratory syndrome (SARS), H1N1, chikungunya and zika [2]. Currently, over 6 million cases have been confirmed, over 200,000 deaths have been reported in the United States and over 30 million people have been affected worldwide due to the most recent fatal infectious disease (COVID-19) [3]. COVID-19 is a respiratory disease triggered by SARS-CoV-2, a single-stranded RNA (ranging from 26 to 32 kilobases [4]) coupled with a nucleoprotein within a capsid consisting of matrix protein [5]. It is an unsegmented respiratory virus (diameter of 80–120 nm [6]) belonging to the subfamily of *Coronavirinae*, divided into *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* [7–10]. SARS-CoV-2 can cause anything from a symptom-free infection to death, with many outcomes in between [11]. The symptomatology of this virus is categorized as mild, severe and critical disease [11,12], which includes but is not limited to fever, headache, fatigue, dry cough, sore throat, pneumonia, dyspnea, stroke, taste and smell impairment, congestion or runny nose, nausea or vomiting, diarrhea, shortness of breath, dyspnea, respiratory failure, septic shock and multiple organ failure [13–20]. However, due to severe cytopathic effects and induced acute

immune-inflammatory response, the lung is one of the organs most affected by SARS-CoV-2 infection [21,22], especially among the elderly and those with specific comorbidities (e.g., diabetes and obesity) [23].

Investigating COVID-19-disease-causing variants among highly expressed genes supports finding the root causes of uncertainties in patient care [24]. Informative exposure signatures can help to assess the association of the COVID-19 transcriptome, offering new insights into the biological and pathological underpinnings of health disparities. Investigation of susceptibility, severity and protection against infectious diseases and vaccine production requires comprehension of the genetic and phenotypic structure of COVID-19 in pathogenesis [25]. Population genomics can provide insights into COVID-19 genetic elements, immunity genes and their variants, and gene-disease networks and pathways [26]. Studying individual and population genetic variations will establish an understanding of COVID-19 disease transmission and pathogenesis [27]. Studies have used a candidate gene approach and identified patients with severe COVID-19 who have mutations in genes involved in the regulation of type I and III interferon (IFN) immunity [28]. These mutations may have a crucial role in the causes and mechanisms of critical COVID-19 complications. The scope of the current work includes known and reported immunity genes, implicated diseases and their sequel relationships with COVID-19. The current state of COVID-19 related genetics were analyzed for the identification of variants, alleles and haplotypes by processing sequence data of variable lengths [29]. In this study, the genetic susceptibility and variations among the DNA of healthy and COVID-19 diseased individuals were investigated.

Presently, variability in COVID-19 disease has been explored mainly in the context of immunity. Zhang *et al.* reported that SARS-CoV-2's RNA genome could integrate into the DNA of human chromosomes, explaining the reports of positive tests for SARS-CoV-2 after recovery [30]. These findings were the first published results from the COVID Human Genetic Effort. Evidence also suggests genetics may play a role in the severity of infection. However, these results do not imply that SARS-CoV-2 establishes permanent genetic residence in human cells to replicate.

Material & methods

COVID-19 disease is a complicated multisystem infectious disease [31]. International efforts (e.g., epidemiology and basic science, etc.) have been made to understand its complexities and adverse effects on the human species [32]. Along with other clinical and epidemiological elements, it is important to examine the genetic factors involved in mutations in the genome of individuals diagnosed with COVID-19 [33]. This study was designed to identify and investigate known immunity genes, possible variants, associated common and variable diseases and their relationship with COVID-19. Furthermore, the authors aimed to create networks and describe the possible gene-disease associations (direct and indirect) to support clinical and translational research.

Immunity genes

An extensive literature search and global analysis of key genes targeting major components of immune systems prognostic of COVID-19 were conducted to identify possible variations. From this search, a list of immune genes was created (Table 1), comprising *ACE2*, *TMPRSS4*, *TMPRSS2*, *CCR9*, *IFNG*, *CD147*, *CXCR6*, *FYCO1*, *IL6*, *LZTFL1*, *MIF*, *SLC6A20*, *XCR1*, *IFNAR1*, *IFNAR2*, *IRF3*, *IRF7*, *IRF9*, *STAT1*, *STAT2*, *TBK1*, *TICAM1/TRIF*, *TLR3*, *TRAF3*, *UNC93B1*, *RAG1* and *RAG2* [23,28,34–38]. In addition to their link to COVID-19, these genes have also been known for mutations leading to other disorders.

Angiotensin-converting enzyme 2 (ACE2; ENSG00000130234) is an enzyme connected to the membranes of lung, artery, heart, kidney and intestinal cells [39–41]. It has been reported to be involved in different disorders, which include but are not limited to posterior urethral valves, myocardial infarction, hypertension essential, intracranial aneurysm, SARS, Hartnup disease, internal hemorrhoid, tetanus neonatorum, neurogenic hypertension and COVID-19 [3,21,29,34,36,39–41]. Transmembrane serine protease 4 (TMPRSS4; ENSG00000137648) is a single-pass type II membrane protein that can cause a variety of oncological and nononcological human diseases and disorders due to its involvement in a variety of biological processes [23,36,42,43]. Transmembrane serine protease 2 (TMPRSS2; ENSG00000184012) is a protein belonging to the serine protease family, involved in many physiological and pathological processes [23]. It is associated with several diseases, which include but are not limited to prostate cancer, prostatic acinar adenocarcinoma, autosomal recessive deafness [25,26], small cell carcinoma, influenza, suppression of tumorigenicity 12, SARS, male reproductive organ cancer and COVID-19 [23,34,36,41,44].

C–C motif chemokine receptor 9 (CCR9; ENSG00000173585) is involved in the pathological progression of inflammatory diseases by its effect on the chemotaxis of inflammatory cells [45,46]. Emerging evidence reported

Table 1. Immunity genes reported to be affected by COVID-19.

Gene	ENSG ID	Species
ACE2	ENSG00000130234	Human
TMPRSS4	ENSG00000137648	Human
TMPRSS2	ENSG00000184012	Human
CCR9	ENSG00000173585	Human
IFNG	ENSG00000111537	Human
CD147/BSG	ENSG00000172270	Human
CXCR6	ENSG00000172215	Human
FYCO1	ENSG00000163820	Human
IL6	ENSG00000136244	Human
LZTFL1	ENSG00000163818	Human
MIF	ENSG00000240972	Human
SLC6A20	ENSG00000163817	Human
XCR1	ENSG00000173578	Human
IFNAR1	ENSG00000142166	Human
IFNAR2	ENSG00000159110	Human
IRF3	ENSG00000126456	Human
IRF7	ENSG00000185507	Human
IRF9	ENSG00000213928	Human
STAT1	ENSG00000115415	Human
STAT2	ENSG00000170581	Human
TBK1	ENSG00000183735	Human
TICAM1/TRIF	ENSG00000127666	Human
TLR3	ENSG00000164342	Human
TRAF3	ENSG00000131323	Human
UNC93B1	ENSG00000110057	Human
RAG1	ENSG00000166349	Human
RAG2	ENSG00000175097	Human

major diseases associated with it, including precursor T-cell acute lymphoblastic leukemia, cholangitis primary sclerosing, CD45 deficiency, celiac disease 1, rheumatoid arthritis and COVID-19 [34,37,47–50]. Interferon gamma (IFNG; ENSG00000111537) is a soluble cytokine that is a member of the type II interferon class [51] and is important for tumor control and immunity against intracellular pathogens [52]. Diseases that can be caused due to IFNG include asthma, mouth disease, autoimmune disease, Behçet syndrome, colorectal cancer, Rasmussen encephalitis, poliomyelitis, dermatitis atopic, cytomegalovirus infection, multiple sclerosis and COVID-19 [51,52]. CD147 (CD147; ENSG00000172270) is a protein proven to be an important receptor on red blood cells and a key molecule involved in the interaction between hepatocellular carcinoma cells and hematopoietic stem cells. CD147 is involved in a variety of biological processes and diseases including breast cancer, thyroid carcinoma, ameloblastoma, ascending cholangitis, grade III astrocytoma, Newcastle disease, necrotizing ulcerative gingivitis, mucolipidosis-III alpha/beta, accommodative esotropia, tongue squamous cell carcinoma, tumor invasion, metastasis, angiogenesis, energy metabolism, multidrug resistance and COVID-19 [53,54].

C–X–C motif chemokine receptor type 6 (CXCR6; ENSG00000172215), a coreceptor for simian immunodeficiency virus (SIV) and HIV [55], can cause disease xanthogranulomatous cholecystitis, sarcoidosis-1, diabetes mellitus and COVID-19 [34,37,56]. FYVE and coiled-coil domain autophagy adaptor 1 (FYCO1; ENSG00000163820) encodes a Rab7 adapter protein involved in the microtubule transport of autophagosomes [57]. Mutations in FYCO1 are due to autosomal recessive congenital cataract-2 (CATC2), which can lead to diseases including cataract, cataract-44, cataract-18, early-onset nuclear cataract, inclusion body myositis and COVID-19 [34,37]. IL-6 (ENSG00000136244) is involved in a variety of clinical and biological features related to the production of acute-phase proteins, a proinflammatory cytokine and an anti-inflammatory myokine [58,59]. Disorders in the *IL-6* gene can cause extramedullary plasmacytoma, plasmacytoma, asthma, hemorrhagic fever, extrinsic cardiomy-

opathy, mouth disease, idiopathic neutropenia, glucose metabolism disease, arteries anomalies of arteriovenous malformations of the brain and COVID-19.

LZTFL1 (*LZTFL1*; ENSG00000163818) is a protein-coding gene and diseases associated with *LZTFL1* include Bardet–Biedl syndrome, Bardet–Biedl syndrome-1, situs inversus, Bardet–Biedl syndrome 17 and COVID-19 [34,37,60]. Macrophage migration inhibitory factor (MIF; ENSG00000240972) is an important regulator of innate immunity that promotes the proinflammatory functions of immune cells [61]. MIF can lead to disease including macular holes, inflammatory bowel disease, rheumatoid arthritis, photokeratitis, malaria, diffuse cutaneous systemic sclerosis, lepromatous leprosy, sarcoidosis-1, cysticercosis, pulmonary hemosiderosis and COVID-19. Solute carrier family 6 member 20 (SLC6A20; ENSG00000163817) is a transporter for neurotransmitters, proteinogenic amino acids, betaine, taurine and creatine [62]. Mutations in *SLC6A20* can lead to hyperglycinuria, Hartnup disorder, iminoglycinuria and COVID-19 [34,37]. X–C motif chemokine receptor 1 (XCR1; ENSG00000173578) is the receptor for XCL1 and XCL2 and diseases associated with XCR1 include Leber plus disease, Parkinson's and COVID-19 [34,37,63].

IFNAR is a common receptor consisting of two polypeptide subunits: IFN-α and -β receptor subunit 1 (IFNAR1; ENSG00000142166) and IFN-α and -β receptor subunit 2 (IFNAR2; ENSG00000159110) [64]. IFNAR1 and IFNAR2 encode type I membrane proteins that form chains of receptors for IFN-α and -β [28,64,65]. IFNAR1 disorder can produce multiple sclerosis, viral infectious disease, malaria, hepatitis C, septicemic plague, papilloma type I and COVID-19. IFNAR2 can trigger hepatitis, hepatitis B, measles, primary immunodeficiency with post measles-mumps-rubella vaccine viral infection type I, hepatitis C, coronary aneurysm, mumps, heart aneurysm, rubella and COVID-19 [28]. Interferon regulatory factor 3 (IRF3; ENSG00000126456), interferon regulatory factor 7 (IRF7; ENSG00000185507), and interferon regulatory factor 9 (IRF9; ENSG00000213928) encode members of the IRF family and play an essential part in the innate immune response to DNA and RNA viruses [28,66–68]. Disorders in IRF3 can affect mouth disease, pediatric lymphoma, hepatitis, influenza, Kaposi sarcoma, eczema herpeticum, viral infectious disease, HIV type 1, hepatitis C virus and COVID-19 [28]. IRF7 is linked to several diseases including pediatric lymphoma, Kaposi sarcoma, viral infectious disease, yellow fever, hepatitis C virus, vaccinia, immunodeficiency-39, Newcastle disease, hepatitis C, Venezuelan equine encephalitis and COVID-19. IRF9 is associated with skin papilloma and COVID-19 [28].

Signal transducer and activator of transcription 1 (STAT1; ENSG00000115415) and signal transducer and activator of transcription 2 (STAT2; ENSG00000170581) are vital elements of the cellular antiviral response and adaptive immunity and are important arbitrators of type I and type III IFN signaling [28,69,70]. Genetic disorders in STAT1 can cause mouth disease, breast cancer, colorectal cancer, chronic lymphocytic leukemia, Ewing sarcoma, ulceroglandular tularemia, acute promyelocytic leukemia, Fanconi anemia complementation group A, immunodeficiency 31B, fibrosarcoma and COVID-19. STAT2 has been reported to be involved in different disorders, which include but are not limited to avian influenza, dengue hemorrhagic fever, mumps, rabies, microphthalmia with limb anomalies, primary immunodeficiency with post measles-mumps-rubella vaccine viral infection, immunodeficiency 44, dengue virus, skin squamous cell carcinoma and COVID-19 [28].

TANK binding kinase 1 (TBK1; ENSG00000183735) inhibits IκB proteins and performs crucial functions in the signaling pathway of immunoreceptors (TLRs, RLRs and STING-mediated sensing of cytosolic DNA) [71]. TBK1 disorders can result in diseases such as amyotrophic lateral sclerosis-1, lateral sclerosis, viral infectious disease, glaucoma 1 open-angle p, retinitis pigmentosa-33, frontotemporal dementia, amyotrophic lateral sclerosis 4, low tension glaucoma, dementia, open-angle glaucoma, Bartter syndrome types 3 and 2 and COVID-19 [28]. Toll-like receptor adaptor molecule 1 (TICAM1/TRIF; ENSG00000127666) encodes an adaptor protein including toll/IL-1 receptor (TIR) and reconciles the expression of many genes [72,73]. Mutations in TICAM1 are associated with acute infection-induced encephalopathy 6, pertussis, encephalitis, herpes simplex encephalitis, herpes simplex and COVID-19 [28].

Toll-like receptor 3 (TLR3; ENSG00000164342) is a receptor for dsRNA, which includes an extracellular leucine-rich repeat (LRR) motif, a transmembrane (TM) domain and intracellular toll and IL-1R (TIR) domains. It is generated during most viral infections [74]. Alterations in TLR3 can cause herpes simplex encephalitis, age-related macular degeneration 1, measles, HIV type 1, retinal vasculitis, hepatitis C, encephalopathy acute infection induced-2, Vogt–Koyanagi–Harada disease, rabies, allergic conjunctivitis and COVID-19 [28]. TNF receptor associated factor 3 (TRAF3; ENSG00000131323) is an enigmatic part of the TRAF family that is involved in substantial physiological and cellular functions in multiple organs [75,76]. Diseases caused by TRAF3 mutations include herpes simplex, splenic marginal zone lymphoma, herpes simplex encephalitis, encephalopathy acute infection-induced

5, hereditary fructose intolerance and COVID-19 [28]. Unc-93 homolog B1 (UNC93B1; ENSG00000110057) encodes a transmembrane protein that regulates the movement of TLRs from the endoplasmic reticulum [77,78]. It is associated with several disorders, including encephalopathy acute infection-induced 1, encephalitis, Melkersson–Rosenthal syndrome, herpes simplex, herpes simplex encephalitis, lymphoid interstitial pneumonia, mite infestation and COVID-19 [28].

Recombination activating 1 (RAG1; ENSG00000166349) and recombination activating 2 (RAG2; ENSG00000175097) begin V(D)J recombination [79,80]. RAG2 mutations can cause common variable immunodeficiency, baylisascariasis, combined immunodeficiency x-linked, immune deficiency disease, combined cellular and humoral immune defects with granulomas, salivary gland disease, sialadenitis, malignant histiocytosis, immunodeficiency with hyper-IgM type 3, severe combined immunodeficiency and COVID-19. Whereas, RAG2 disorder can lead to severe combined immunodeficiency (autosomal recessive T-cell negative B-cell positive NK-cell negative), Omenn syndrome, Lig4 syndrome, severe combined immunodeficiency, malignant histiocytosis, recombinase activating gene-1 deficiency, combined cellular and humoral immune defects with granulomas, immune deficiency disease, combined immunodeficiency x-linked, gastroduodenitis and COVID-19 [3].

Data processing

Whole-genome sequencing (WGS) is widely applied to sequence the entirety of the genome and whole-exome sequencing (WES) sequences mainly the protein-coding structures. To realize the clinical impact of this study and perform gene-variant analysis at the listed known immunity genes, a prospective dataset was created that includes total of 752 WGS samples. The sample population selection criteria included short read sequencing (paired), diversity, open-access availability through authenticated resources, reproducibility and success rate. The genomics pipeline was applied to the complete dataset and most of the samples were with very low sequencing quality and were unable to produce expected results. Therefore, only those results which were reproduced from good-quality WGS samples (n = 13; SRR12474733, SRR12486921, SRR12328890, ERR4387385, ERR4387386, ERR4387388, SRR12336742, SRR12336753, SRR12336755, SRR12336756, SRR12336761, SRR12336765 and SRR12336766) were selected and analyzed. All samples were collected from public resources (COVID-19 WGS data-related information is shown in Supplementary Table 1). Collected, sequenced WGS samples were from variable COVID-19 populations and belonging to different regions and sizes and were produced using Illumina sequencing technology. All the samples were downloaded from public repositories and analyzed in accordance with relevant guidelines and regulations and protocols were approved by the Institutional Review Board (IRB), Rutgers University.

An in-house developed gene-variant analysis pipeline (JWES) was applied for the whole genome and exome data preprocessing, modeling and downstream analysis (Figure 1). JWES is mainly based on processing the raw sequence data, converting raw signals into base calling, identifying regions of interest in the genome, aligning, and assembling contigs and scaffolds and variant detection [81]. Its overall operations are divided into three modules: data preprocessing, storage and management, and visualization. JWES is a cross-platform and user-friendly Java-based application that integrates multiple open-source command-line tools for sequence data processing and analysis, consisting of FASTQC (quality assessment) [82], Burrows–Wheeler Aligner software (BWA for short read alignment to reference human genome) [83,84], MarkDuplicates (removes redundant reads) [85], SAMtools (sorting and indexing) [86] and Genome Analysis Tool Kit (GATK for finding SNPs and indels) [87–89]. JWES is freely available for download and use by the community [81].

Results

Gene-variant analysis was performed at the sequenced WGS and WES samples on COVID-19 populations from diverse backgrounds and regions across the world. Using published sequence data from SARS-CoV-2-infected individuals, the likelihood that genetics may have a role in the risk of severity of COVID-19 through the components of the immune system was examined. The gene variants were annotated for mutation type and protein change using known algorithms and tools for interpreting mutations. To better understand the clinical impact of these variants, an ontology of gene-disease organizations, progressions and networks was created.

Gene-variant analysis

The gene-variant analysis was performed using the JWES pipeline to identify abundantly mutated core immune genes in COVID-19 patients with WGS samples. Inborn errors in the immune genes can tip the delicate balance

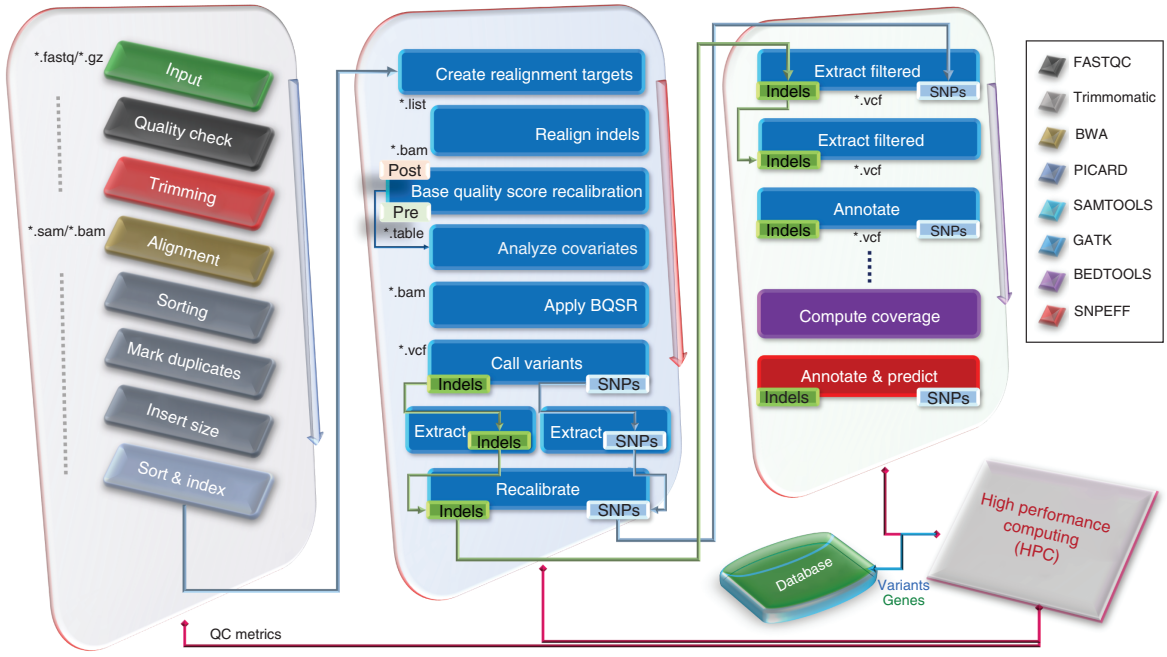


Figure 1. Whole-genome and whole-exome sequencing data processing, quality checking, modeling, analysis and visualization pipeline for the identification and annotation of variants associated with genes of interest.
BQSR: Base quality score recalibration; QC: Quality control; SNP: Single-nucleotide polymorphism.

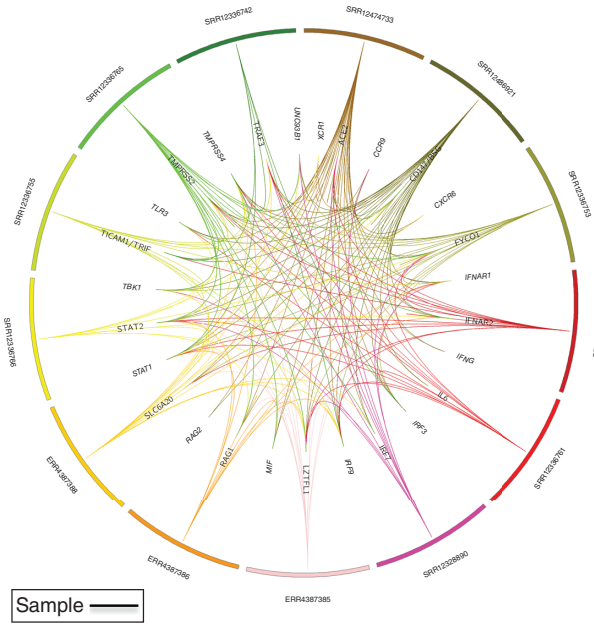
of protective immunity leading to deleterious, innate and adaptive immune responses. For each gene, the details of the mutations found and the parent genes in the chromosome are illustrated in Circos plots (Figure 2). Figure 2A represents the COVID-19 core genes and associated genes across the dataset. The outer circle is composed of patients and the inner circle consists of COVID-19 genes linking to the patients. The ribbons represent the interaction between the parent genes and patients. The highest number of mutations were found in *XCR1* (n = 307) and the lowest number of mutations were found in *IFNG* (n = 10). Other genes also had mutations: *ACE2* (n = 40), *TMPRSS2* (n = 186), *TMPRSS4* (n = 255), *CCR9* (n = 41), *CD147/BSG* (n = 90), *CXCR6* (n = 29), *FYCO1* (n = 234), *IL-6* (n = 68), *LZTFL1* (n = 153), *MIF* (n = 21), *SLC6A20* (n = 99), *IFNAR1* (n = 140), *IFNAR2* (n = 200), *IRF3* (n = 28), *IRF7* (n = 20), *IRF9* (n = 19), *STAT1* (n = 128), *STAT2* (n = 60), *TBK1* (n = 94), *TICAM1/TRIF* (n = 106), *TLR3* (n = 178), *TRAF3* (n = 202), *UNC93B1* [50], *RAG1* (n = 109) and *RAG2* (n = 22).

Figure 2B shows the Circos plot of COVID-19 core genes with known and experimentally deleterious variants and their parent genes in the chromosome. The outer circle is composed of the chromosome ideogram linking the location of the parent gene in the chromosomes. The inner circle shows a histogram of the number of mutations reported in the genes. The most prevalent gene was *TMPRSS4*, which occurred in all patients, while the least prevalent genes included *RAG2*, *IFNG*, *CCR9* and *IL6* (Figure 2C). Further investigation revealed all earlier reported human transcripts for each gene among the list (Table 1 & Figure 2C). Except for *ACE2*, *IFNAR1*, *IRF3*, *IRF7*, *STAT1*, *STAT2* and *TBK1*, all reported transcripts (ENST) were found for *TMPRSS4*, *TMPRSS2*, *CCR9*, *IFNG*, *CD147/BSG*, *CXCR6*, *FYCO1*, *IL6*, *LZTFL1*, *MIF*, *SLC6A20*, *XCR1*, *IFNAR2*, *IFNAR2*, *IRF9*, *TICAM1/TRIF*, *TLR3*, *TRAF3*, *UNC93B1*, *RAG1* and *RAG2* (Table 2). Gene-variant analysis details are shown in Supplementary Table 2.

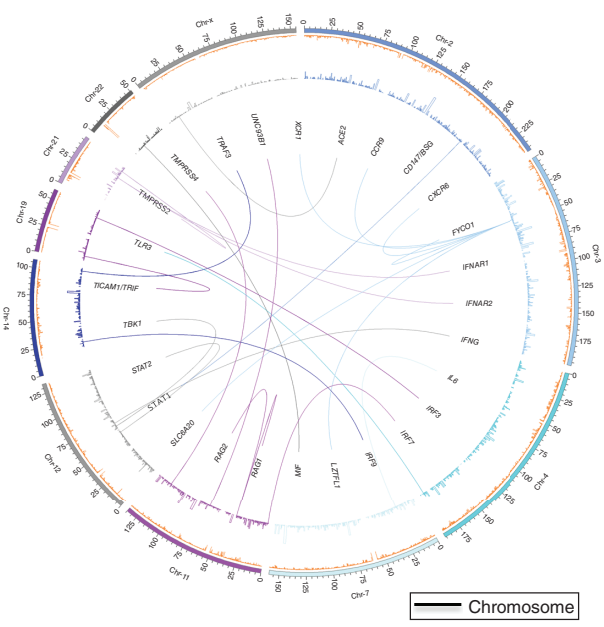
Gene-variant-protein analysis

The mutations in genes implicated in COVID-19 were analyzed as coding proteins with labeled recurrent hotspots. Lollipop plots were produced representing the number of mutations per gene. The variant data was annotated for biological and functional implications with three computational algorithms: Scale-Invariant Feature Transform (SIFT) [90–92], Polymorphism Phenotyping v2 (PolyPhen-2) [93], and MutationAssessor [94]. SIFT was used to predict the effect of coding variants on protein function [91], as the challenge is to identify causative variants for the phenotypes linked to COVID-19. Originally, SIFT was proposed to predict the impact of amino acid substitutions

Ⓐ **Sample, gene-variant analysis.**



Ⓑ **Protein coding and non-coding gene-variant analysis.**



Ⓒ **Sample, gene-variant analysis table.**

Gene name	Samples ID/ ENSG ID	Variants	SRR12474733_1	SRR12486921_1	SRR12336753_1	SRR12336756_1	SRR12336761_1	SRR12328890_1	ERR4387385_1	ERR4387386_1	ERR4387388_1	SRR12336766_1	SRR12336755_1	SRR12336765_1	SRR12336742_1
ACE2	ENSG00000130234	40	26	4	2	3	2	3	0	0	0	0	0	0	0
TMPRSS4	ENSG00000137648	255	76	73	13	13	10	4	4	8	7	7	7	31	2
TMPRSS2	ENSG00000184012	186	86	67	2	9	4	0	3	0	6	0	2	6	1
CCR9	ENSG00000173585	41	23	17	0	0	0	1	0	0	0	0	0	0	0
IFNG	ENSG00000111537	10	8	2	0	0	0	0	0	0	0	0	0	0	0
CD147/BSG	ENSG00000172270	90	60	25	1	0	0	0	0	0	2	0	1	1	0
CXCR6	ENSG00000172215	29	20	8	0	0	0	0	0	0	1	0	0	0	0
FYCO1	ENSG00000163820	234	125	80	11	1	3	3	3	0	2	4	0	2	0
IL6	ENSG00000136244	68	40	28	0	0	0	0	0	0	0	0	0	0	0
LZTFL1	ENSG00000163818	153	79	59	6	1	2	1	0	1	2	0	1	1	0
MIF	ENSG00000240972	21	13	5	0	0	0	0	0	2	0	0	0	1	0
SLC6A20	ENSG00000163817	99	54	41	1	1	0	0	1	0	1	0	0	0	0
XCR1	ENSG00000173578	307	136	168	0	0	0	0	1	0	0	2	0	0	0
IFNAR1	ENSG00000142166	140	53	59	18	1	0	2	2	3	0	0	2	0	0
IFNAR2	ENSG00000159110	200	122	68	1	2	0	3	0	1	0	2	0	1	0
IRF3	ENSG00000126456	28	10	9	4	0	1	0	0	0	0	0	0	3	1
IRF7	ENSG00000185507	20	11	5	0	0	0	0	0	0	2	0	0	2	0
IRF9	ENSG00000213928	19	6	1	2	1	1	0	2	1	2	1	1	0	1
STAT1	ENSG00000115415	128	50	22	7	2	10	0	10	8	3	3	8	3	2
STAT2	ENSG00000170581	60	4	11	10	1	11	1	3	2	1	0	2	14	0
TBK1	ENSG00000183735	94	32	30	13	0	0	7	2	0	2	0	7	1	0
TICAM1/TRIF	ENSG00000127666	106	54	42	0	4	0	0	0	0	0	0	0	1	5
TLR3	ENSG00000164342	178	125	47	0	0	1	0	1	0	0	2	0	2	0
TRAF3	ENSG00000131323	202	130	57	4	1	2	2	0	0	1	0	5	0	0
UNC93B1	ENSG00000110057	47	35	6	3	0	0	1	0	0	1	0	0	1	0
RAG1	ENSG00000166349	109	55	44	0	0	0	0	1	1	0	6	0	2	0
RAG2	ENSG00000175097	22	5	17	0	0	0	0	0	0	0	0	0	0	0
Total variants		**2886**	**1438**	**995**	**98**	**40**	**47**	**28**	**33**	**27**	**33**	**27**	**36**	**72**	**12**

Figure 2. Gene-variant analysis. (A) Circos plot of sample-based gene-variant analysis. **(B)** Circos plot of protein-coding and noncoding gene-variant analysis. **(C)** Gene-variant analysis table including details about genes, samples and observed variants.

Table 2. Gene-variant human transcript analysis.

Gene	ENSG	Present human transcripts (ENST)	Missing ENST
ACE2	ENSG00000130234	ENST00000252519, ENST00000471548, ENST00000473851, ENST00000427411	ENST00000484756
TMPRSS4	ENSG00000137648	ENST00000437212, ENST00000517483, ENST00000517544, ENST00000518413, ENST00000518610, ENST00000519126, ENST00000519236, ENST00000519813, ENST00000520063, ENST00000522151, ENST00000522307, ENST00000522462, ENST00000522824, ENST00000523251, ENST00000523770, ENST00000524218, ENST00000528118, ENST00000534111, ENST00000616579, ENST00000618855	NA
TMPRSS2	ENSG00000184012	ENST00000332149, ENST00000398585, ENST00000424093, ENST00000454499, ENST00000455813, ENST00000458356, ENST00000463138, ENST00000469395, ENST00000488556, ENST00000489201	NA
CCR9	ENSG00000173585	ENST00000357632, ENST00000395963, ENST00000463197, ENST00000422395, ENST00000355983	NA
IFNG	ENSG00000111537	ENST00000229135	NA
CD147/BSG	ENSG00000172270	ENST00000545507, ENST00000346916, ENST00000333511, ENST00000576925, ENST00000573216, ENST00000353555, ENST00000572899, ENST00000590218, ENST00000618112, ENST00000613627, ENST00000618006, ENST00000614867, ENST00000573784, ENST00000574970, ENST00000576984, ENST00000571735	NA
CXCR6	ENSG00000172215	ENST00000438735, ENST00000304552, ENST00000458629, ENST00000457814	NA
FYCO1	ENSG00000163820	NST00000296137, ENST00000433878, ENST00000438446, ENST00000471739, ENST00000535325	NA
IL6	ENSG00000136244	ENST00000258743, ENST00000401630, ENST00000401651, ENST00000404625, ENST00000406575, ENST00000407492, ENST00000426291, ENST00000464710, ENST00000485300	NA
LZTFL1	ENSG00000163818	ENST00000296135, ENST00000411866, ENST00000418700, ENST00000440576, ENST00000445698, ENST00000448111, ENST00000469874, ENST00000472635, ENST00000478551, ENST00000480156, ENST00000483279, ENST00000490463, ENST00000492333, ENST00000495864, ENST00000536047, ENST00000539217	NA
MIF	ENSG00000240972	ENST00000215754, ENST00000465752, ENST00000498385	NA
SLC6A20	ENSG00000163817	ENST00000353278, ENST00000358525, ENST00000413781, ENST00000456124, ENST00000470226, ENST00000473146, ENST00000493980	NA
XCR1	ENSG00000173578	ENST00000309285, ENST00000395946	NA
IFNAR1	ENSG00000142166	ENST00000270139, ENST00000442071	ENST00000651609, ENST00000652450, ENST00000652513, ENST00000652601, ENST00000652654
IFNAR2	ENSG00000159110	ENST00000342101, ENST00000342136, ENST00000382238, ENST00000382264, ENST00000404220, ENST00000413881, ENST00000417007, ENST00000420068, ENST00000443073, ENST00000447980	NA
IRF3	ENSG00000126456	ENST00000309877, ENST00000377135, ENST00000377139, ENST00000442265, ENST00000593337, ENST00000593818, ENST00000593922, ENST00000594387, ENST00000595034, ENST00000595240, ENST00000596644, ENST00000596756, ENST00000596765, ENST00000596788, ENST00000596822, ENST00000597180, ENST00000597198, ENST00000597369, ENST00000597636, ENST00000598108, ENST00000598808, ENST00000599144, ENST00000599223, ENST00000599680, ENST00000600022, ENST00000600453, ENST00000600911, ENST00000601291, ENST00000601373, ENST00000601809, ENST00000602190	
IRF7	ENSG00000185507	ENST00000330243, ENST00000348655, ENST00000397566, ENST00000397570, ENST00000469048, ENST00000525445, ENST00000525750, ENST00000527160, ENST00000528413, ENST00000531912, ENST00000532096, ENST00000532326, ENST00000532788, ENST00000533182, ENST00000533190	ENST00000647801, ENST00000649187
IRF9	ENSG00000213928	ENST00000324076, ENST00000396864, ENST00000557894, ENST00000559229, ENST00000559284, ENST00000559863, ENST00000560275, ENST00000560311, ENST00000560365, ENST00000560542, ENST00000560852, ENST00000561009, ENST00000561342, ENST00000561412, ENST00000561415	NA

NA: Not applicable.

Table 2. Gene-variant human transcript analysis (cont.).

Gene	ENSG	Present human transcripts (ENST)	Missing ENST
STAT1	ENSG00000115415	ENST00000361099, ENST00000392322, ENST00000392323, ENST00000409465, ENST00000415035, ENST00000423282, ENST00000424722, ENST00000432058, ENST00000452281, ENST00000454414, ENST00000464072, ENST00000540176	ENST00000673638, ENST00000673734, ENST00000673762, ENST00000673777, ENST00000673816, ENST00000673832, ENST00000673841, ENST00000673847, ENST00000673858, ENST00000673859, ENST00000673863, ENST00000673885, ENST00000673942, ENST00000673952, ENST00000674028, ENST00000674080, ENST00000674081, ENST00000674153
STAT2	ENSG00000170581	ENST00000314128, ENST00000418572, ENST00000555519, ENST00000555646, ENST00000556140, ENST00000557156, ENST00000557235, ENST00000557417	ENST00000555488, ENST00000556539, ENST00000557199, ENST00000557252, ENST00000650805, ENST00000651078, ENST00000651301, ENST00000651339, ENST00000651805, ENST00000651915, ENST00000651934, ENST00000651967, ENST00000652091, ENST00000652398, ENST00000652624, ENST00000652741
TBK1	ENSG00000183735	ENST00000331710, ENST00000536906, ENST00000538890, ENST00000539810, ENST00000540417, ENST00000541805, ENST00000545025, ENST00000545392	ENST00000650708, ENST00000650762, ENST00000650786, ENST00000650790, ENST00000650997, ENST00000651014, ENST00000651262, ENST00000651878, ENST00000651889, ENST00000651947, ENST00000652389, ENST00000652537, ENST00000652657
TICAM1/TRIF	ENSG00000127666	ENST00000248244, ENST00000621756	NA
TLR3	ENSG00000164342	ENST00000296795, ENST00000504367, ENST00000508051, ENST00000512264, ENST00000513189	NA
TRAF3	ENSG00000131323	ENST00000347662, ENST00000351691, ENST00000392745, ENST00000539721, ENST00000558700, ENST00000558880, ENST00000559734, ENST00000560371, ENST00000560463	NA
UNC93B1	ENSG00000110057	ENST00000227471, ENST00000524455, ENST00000525368, ENST00000528096, ENST00000528423, ENST00000530138, ENST00000531152, ENST00000533424, ENST00000610659, ENST00000622364	NA
RAG1	ENSG00000166349	ENST00000299440, ENST00000529126, ENST00000534663	NA
RAG2	ENSG00000175097	NA	NA

NA: Not applicable.

on protein function [95]. PolyPhen-2 is an automatic web tool used for the extraction of sequences and structure-based features of the substitution site. Single-nucleotide polymorphisms (SNPs) were analyzed in a batch mode, predicted for the functional impact and searched in a database of precomputed predictions for WES data.

MutationAssessor was applied to differentiate between conserved patterns using conservation and specificity scores to account for functional shifts between subfamilies of proteins [94]. MutationMapper [96] was used to generate lollipop plots for all genes [97]. It displays the highly recurrent mutations (amino acid alterations) but does not annotate mutations with low frequency (Table 3). *ACE2*, *TMPRSS4*, *TMPRSS2*, *SLC6A20* and *FYCOI* were found to have mutations with functional implications (Figure 3). Three functional missense mutations were found in *ACE2* (V404A, G405W and S409P) that were computationally annotated to have adverse effects. Three functional missense mutations were also found in *SLC6A20* (A9G, V591G and V591M) but none were convincingly deleterious. *TMPRSS4* had four missense mutations (V208G, K257M, N357Y and G402V) reported. Only G402V was predicted to be deleterious. *TMPRSS2* had three functional missense mutations (V160M, W267R and Q431H) with the first two displaying evidence of adverse effects. *FYCO1* was found to have the greatest number of functional missense mutations (12) but all were predicted to be benign. The remaining genes had no functional impact reported. This may be because the functional databases used, SIFT, PolyPhen-2 and MutationAssessor, are not updated (Figure 3). Additional gene-variant-protein analysis details are shown in Supplementary Table 2.

Gene-variant-disease analysis

A gene-variant-disease network analysis was also performed to identify which immune genes are associated with other disease phenotypes (Supplementary Figure 1 & Table 4). To elaborate the clinical associations between the

Table 3. Gene-variant functional impact analysis.

Gene	Chromosome	Start position	End position	Reference allele	Variant allele	Protein change	Variant type	Mutation type	Functional impact
ACE2	chrX	15578175	15578175	A	G	V404A	SNP	Missense_Mutation	MutationAssessor: impact: high, score: 3.57; SIFT: impact: deleterious, score: 0; Polyphen-2: impact: probably_damaging, score: 0.956
	chrX	15578173	15578173	C	A	G405W	SNP	Missense_Mutation	MutationAssessor: impact: high, score: 3.97; SIFT: impact: deleterious, score: 0; Polyphen-2: impact: probably_damaging, score: 1
	chrX	15578161	15578161	A	G	S409P	SNP	Missense_Mutation	MutationAssessor: impact: medium, score: 3.455; SIFT: impact: deleterious, score: 0; Polyphen-2: impact: probably_damaging, score: 0.983
TMPRSS4	chr11	118111780	118111780	T	G	V208G	SNP	Missense_Mutation	MutationAssessor: impact: neutral, score: -3.15; SIFT: impact: tolerated, score: 1; Polyphen-2: impact: benign, score: 0
	chr11	118113295	118113295	A	T	K257M	SNP	Missense_Mutation	MutationAssessor: impact: low, score: 1.45; SIFT: impact: deleterious, score: 0; Polyphen-2: impact: possibly_damaging, score: 0.905
	chr11	118115197	118115197	A	T	N357Y	SNP	Missense_Mutation	MutationAssessor: impact: low, score: 1.675; SIFT: impact: deleterious, score: 0; Polyphen-2: impact: probably_damaging, score: 0.992
	chr11	118117357	118117357	G	T	G402V	SNP	Missense_Mutation	MutationAssessor: impact: high, score: 4.33; SIFT: impact: deleterious, score: 0; Polyphen-2: impact: probably_damaging, score: 1
TMPRSS2	chr21	41480570	41480570	C	T	V160M	SNP	Missense_Mutation	MutationAssessor: NA; SIFT: impact: deleterious, score: 0.01; Polyphen-2: impact: probably_damaging, score: 0.937
	chr21	41473425	41473425	A	G	W267R	SNP	Missense_Mutation	MutationAssessor: NA; SIFT: impact: deleterious, score: 0; Polyphen-2: impact: probably_damaging, score: 0.977
	chr21	41468417	41468417	C	A	Q431H	SNP	Missense_Mutation	MutationAssessor: NA; SIFT: impact: tolerated, score: 0.11; Polyphen-2: impact: benign, score: 0.039
FYCO1	chr3	45968585	45968585	C	T	R250Q	SNP	Missense_Mutation	MutationAssessor: impact: neutral, score: -2.67; SIFT: impact: tolerated, score: 0.82; Polyphen-2: impact: benign, score: 0
	chr3	45968372	45968372	C	G	G321A	SNP	Missense_Mutation	MutationAssessor: impact: neutral, score: -1.78; SIFT: impact: tolerated, score: 1; Polyphen-2: impact: benign, score: 0
	chr3	45967995	45967995	G	A	R447C	SNP	Missense_Mutation	MutationAssessor: impact: low, score: 1.175; SIFT: impact: tolerated, score: 0.11; Polyphen-2: impact: benign, score: 0.28
	chr3	45967298	45967298	G	A	A679V	SNP	Missense_Mutation	MutationAssessor: impact: medium, score: 2.3; SIFT: impact: tolerated, score: 0.08; Polyphen-2: impact: benign, score: 0.356
	chr3	45967211	45967211	T	A	E708V	SNP	Missense_Mutation	MutationAssessor: impact: medium, score: 2.19; SIFT: impact: tolerated, score: 0.05; Polyphen-2: impact: benign, score: 0.079

SIFT: Scale-Invariant Feature Transform; SNP: Single-nucleotide polymorphism.

Table 3. Gene-variant functional impact analysis (cont.).

Gene	Chromosome	Start position	End position	Reference allele	Variant allele	Protein change	Variant type	Mutation type	Functional impact
	chr3	45967208	45967208	C	A	C709F	SNP	Missense_Mutation	MutationAssessor: impact: medium, score: 2.43; SIFT: impact: deleterious, score: 0.01; Polyphen-2: impact: probably_damaging, score: 0.938
	chr3	45967206	45967206	G	C	Q710E	SNP	Missense_Mutation	MutationAssessor: impact: low, score: 1.67; SIFT: impact: tolerated, score: 0.87; Polyphen-2: impact: benign, score: 0.003
	chr3	45967205	45967205	T	C	Q710R	SNP	Missense_Mutation	MutationAssessor: impact: neutral, score: 0.77; SIFT: impact: tolerated, score: 0.51; Polyphen-2: impact: benign, score: 0
	chr3	45967203	45967203	G	C	Q711E	SNP	Missense_Mutation	MutationAssessor: impact: medium, score: 2.25; SIFT: impact: tolerated, score: 0.2; Polyphen-2: impact: benign, score: 0.001
	chr3	45966333	45966333	T	C	N1001D	SNP	Missense_Mutation	MutationAssessor: impact: neutral, score: -1.14; SIFT: impact: tolerated, score: 1; Polyphen-2: impact: benign, score: 0
	chr3	45966331	45966331	G	T	N1001K	SNP	Missense_Mutation	MutationAssessor: impact: low, score: 1.39; SIFT: impact: tolerated, score: 0.09; Polyphen-2: impact: benign, score: 0.007
	chr3	45964418	45964418	T	A	T1063S	SNP	Missense_Mutation	MutationAssessor: impact: low, score: 1.85; SIFT: impact: tolerated, score: 0.21; Polyphen-2: impact: benign, score: 0.012
SLC6A20	chr3	45796394	45796394	G	C	A9G	SNP	Missense_Mutation	MutationAssessor: impact: neutral, score: -0.585; SIFT: impact: tolerated, score: 0.08; Polyphen-2: impact: benign, score: 0
	chr3	45758985	45758985	A	C	V591G	SNP	Missense_Mutation	MutationAssessor: impact: medium, score: 2.14; SIFT: impact: deleterious, score: 0; Polyphen-2: impact: benign, score: 0.143
	chr3	45758986	45758986	C	T	V591M	SNP	Missense_Mutation	MutationAssessor: impact: low, score: 1.245; SIFT: impact: deleterious, score: 0.05; Polyphen-2: impact: benign, score: 0.021

SIFT: Scale-Invariant Feature Transform; SNP: Single-nucleotide polymorphism.

10.2217/pme-2021-0132

Table 4. Gene-disease associations.

Genes	Diseases									
ACE2	Posterior urethral valves	Myocardial infarction	Hypertension essential	Intracranial aneurysm	Severe acute respiratory syndrome	Hartnup disorder	Internal hemorrhoid	Tetanus neonatorum	Neurogenic hypertension	
TMPRSS4	Deafness autosomal dominant 6	Deafness autosomal recessive 8	Deafness autosomal dominant 13	Deafness autosomal recessive 16	Nonsyndromic deafness	Deafness autosomal recessive 24-2	Deafness autosomal recessive 25	Deafness autosomal recessive 85	Deafness autosomal recessive 83	Deafness autosomal dominant 2a
TMPRSS2	Prostate cancer	Prostatic acinar adenocarcinoma	Deafness autosomal recessive 25	Deafness autosomal recessive 24-2	Small cell carcinoma	Influenza	Suppression of tumorigenicity 12	Severe acute respiratory syndrome	Male reproductive organ cancer	
CCR9	Precursor T cell acute lymphoblastic leukemia	Cholangitis primary sclerosing	Celiac disease 1	CD45 deficiency						
IFNG	Asthma	Mouth disease	Autoimmune disease	Behçet syndrome	Colorectal cancer	Rasmussen encephalitis	Poliomyelitis	Dermatitis atopic	Cytomegalovirus infection	Multiple sclerosis
CD147/BSG	Breast cancer	Differentiated thyroid carcinoma	Ameloblastoma	Ascending cholangitis	Grade III astrocytoma	Newcastle disease	Necrotizing ulcerative gingivitis	Mucolipidosis III alpha-beta	Accommodative esotropia	Tongue squamous cell carcinoma
CXCR6	Xanthogranulomatous cholecystitis	Sarcoidosis 1								
FYCO1	Cataract	Cataract 44	Cataract 18	Early-onset nuclear cataract	Inclusion body myositis					
IL6	Extramedullary plasmacytoma	Plasmacytoma	Asthma	Hemorrhagic fever	Extrinsic cardiomyopathy	Mouth disease	Idiopathic neutropenia	Glucose metabolism disease	Arteries anomalies	Arteriovenous malformations of the brain
LZTFL1	Bardet–Biedl syndrome	Bardet–Biedl syndrome 1	Situs inversus	Bardet–Biedl syndrome 17						
MIF	Macular holes	Inflammatory bowel disease	Rheumatoid arthritis	Photokeratitis	Malaria	Diffuse cutaneous systemic sclerosis	Lepromatous leprosy	Sarcoidosis 1	Cysticercosis	Pulmonary hemosiderosis
SLC6A20	Hyperglycinuria	Hartnup disorder	Iminoglycinuria							
XCR1	Leber plus	Parkinson								
IFNAR1	Multiple sclerosis	Viral infectious disease	Malaria	Hepatitis C	Septicemic plague	Papilloma	Type I			
IFNAR2	Hepatitis	Hepatitis B	Measles	Primary immunodeficiency with post measles-mumps-rubella vaccine viral infection	Type I	Hepatitis C	Coronary aneurysm	Mumps	Heart aneurysm	Rubella
IRF3	Mouth disease	Pediatric lymphoma	Hepatitis	Influenza	Kaposi sarcoma	Eczema herpeticum	Viral infectious disease	HIV type 1	Hepatitis C virus	

Table 4. Gene-disease associations (cont.).

Genes	Diseases									
ACE2	Posterior urethral valves	Myocardial infarction	Hypertension essential	Intracranial aneurysm	Severe acute respiratory syndrome	Hartnup disorder	Internal hemorrhoid	Tetanus neonatorum	Neurogenic hypertension	
IRF7	Pediatric lymphoma	Kaposi sarcoma	Viral infectious disease	Yellow fever	Hepatitis C	Vaccinia	Immunodeficiency 39	Newcastle disease	Hepatitis C	Venezuelan equine encephalitis
IRF9	Skin papilloma									
STAT1	Mouth disease	Breast cancer	Colorectal cancer	Leukemia chronic lymphocytic	Ewing sarcoma	Ulceroglandular tularemia	Acute promyelocytic leukemia	Fanconi anemia complementation group A	Immunodeficiency 31b	Fibrosarcoma
STAT2	Avian influenza	Dengue hemorrhagic fever	Mumps	Rabies	Microphthalmia with limb anomalies	Primary immunodeficiency with post measles-mumps-rubella vaccine viral infection	Immunodeficiency 44	Dengue virus	Skin squamous cell carcinoma	
TBK1	Amyotrophic lateral sclerosis 1	Lateral sclerosis	Viral infectious disease	Glaucoma 1 open-angle p	Retinitis pigmentosa 33	Frontotemporal dementia and/or amyotrophic lateral sclerosis 4	Low tension glaucoma	Dementia	Open-angle glaucoma	Bartter syndrome type 3 2
TICAM1/TRIF	Encephalopathy acute infection-induced 6	Pertussis	Encephalitis	Herpes simplex encephalitis	Herpes simplex					
TLR3	Herpes simplex encephalitis	Macular degeneration age-related 1	Measles	HIV type 1	Retinal vasculitis	Hepatitis C	Encephalopathy acute infection-induced 2	Vogt Koyanagi Harada disease	Rabies	Allergic conjunctivitis
TRAF3	Herpes simplex	Splenic marginal zone lymphoma	Herpes simplex encephalitis	Encephalopathy acute infection-induced 5	Fructose intolerance hereditary					
UNC93B1	Encephalopathy acute infection-induced 1	Encephalitis	Melkersson-Rosenthal syndrome	Herpes simplex	Herpes simplex encephalitis	Lymphoid interstitial pneumonia	Mite infestation			
RAG1	Common variable immunodeficiency	Baylisascariasis	Combined immunodeficiency x-linked	Immune deficiency disease	Combined cellular and humoral immune defects with granulomas	Salivary gland disease	Sialadenitis	Malignant histiocytosis	Immunodeficiency with hyper igm type 3	Severe combined immunodeficiency
RAG2	Severe combined immunodeficiency autosomal recessive T-cell negative B-cell positive NK-cell negative	Omenn syndrome	Lig4 syndrome	Severe combined immunodeficiency	Malignant histiocytosis	Recombinase-activating gene 1 deficiency	Combined cellular and humoral immune defects with granulomas	Immune deficiency disease	Combined immunodeficiency x-linked	Gastroduodenitis

10.2217/pme-2021-0132

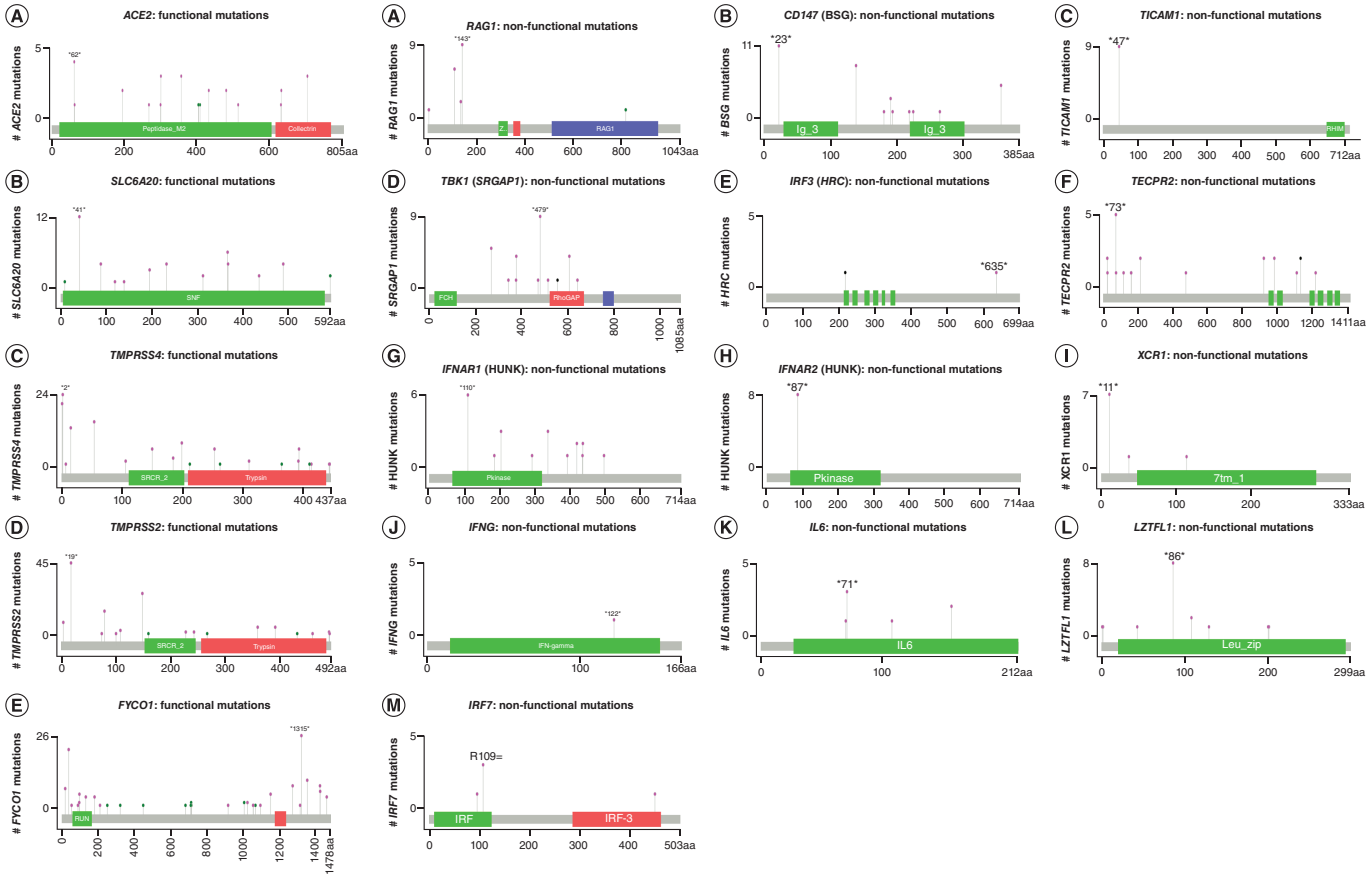


Figure 3. Genes with functional and nonfunctional mutations. Functional gene mutation analysis of *ACE2, SLC6A20, TMPRSS4, TMPRSS2* and *FYCO1*. Nonfunctional gene mutation analysis of *RAG1, CD147 (BSG), TBK1 (SRGAP1), IRF3 (HRC), IFNAR1 (HUNK), INFAR2 (HUNK), IFNG, IL6, IRF7, LZTFL1, TICAM1, TECPR2* and *XCR1*.

genes utilized, the authors' previously designed the gene-SNP-disease-drug smart database [97] which was used to draw the network of diseases linking to the genes (Figure 4). Furthermore, a comparative analysis of gene-associated diseases investigated in this study and reported in the recently published literature to be included among the symptomatology of COVID-19 was performed (Figure 5). To annotate the genes with diseases, the authors used an in-house developed gene-annotation database [95,98,99]. Many diseases were found shared between the genes. They included deafness autosomal recessive-24-2, leukemia, sarcoidosis-1, Hartnup disorder, papilloma, viral infectious disease, SARS, malaria, influenza, asthma, fever, mouth disease, herpes simplex, acute infection-induced encephalopathy, encephalitis, mumps, rabies, hepatitis, combined immunodeficiency x-linked, immune deficiency disease, combined cellular and humoral immune defects with granulomas, severe combined immunodeficiency, malignant histiocytosis, measles, Kaposi sarcoma, pediatric lymphoma, HIV type-1 and COVID-19. Five diseases had more than two associated genes. Additional gene-variant-disease analysis details are shown in Supplementary Table 3.

Discussion

SARS-CoV-2 infection is a complex multisystem disorder with a wide spectrum of clinical manifestations [100]. The clinical presentation of COVID-19 varies from patient to patient [101]. To precisely diagnose and treat patients infected with COVID-19, it is important to understand the functional impact of variations in the patient's DNA [102,103]. Furthermore, to establish a deeper understanding of novel changes that may increase susceptivity to COVID-19, it is essential to combine phenotypically similar probands to distinguish connected clusters of rare genes [104,105], analyze the rare mutations and phenotypes and identify substantial gene-disease associations [106].

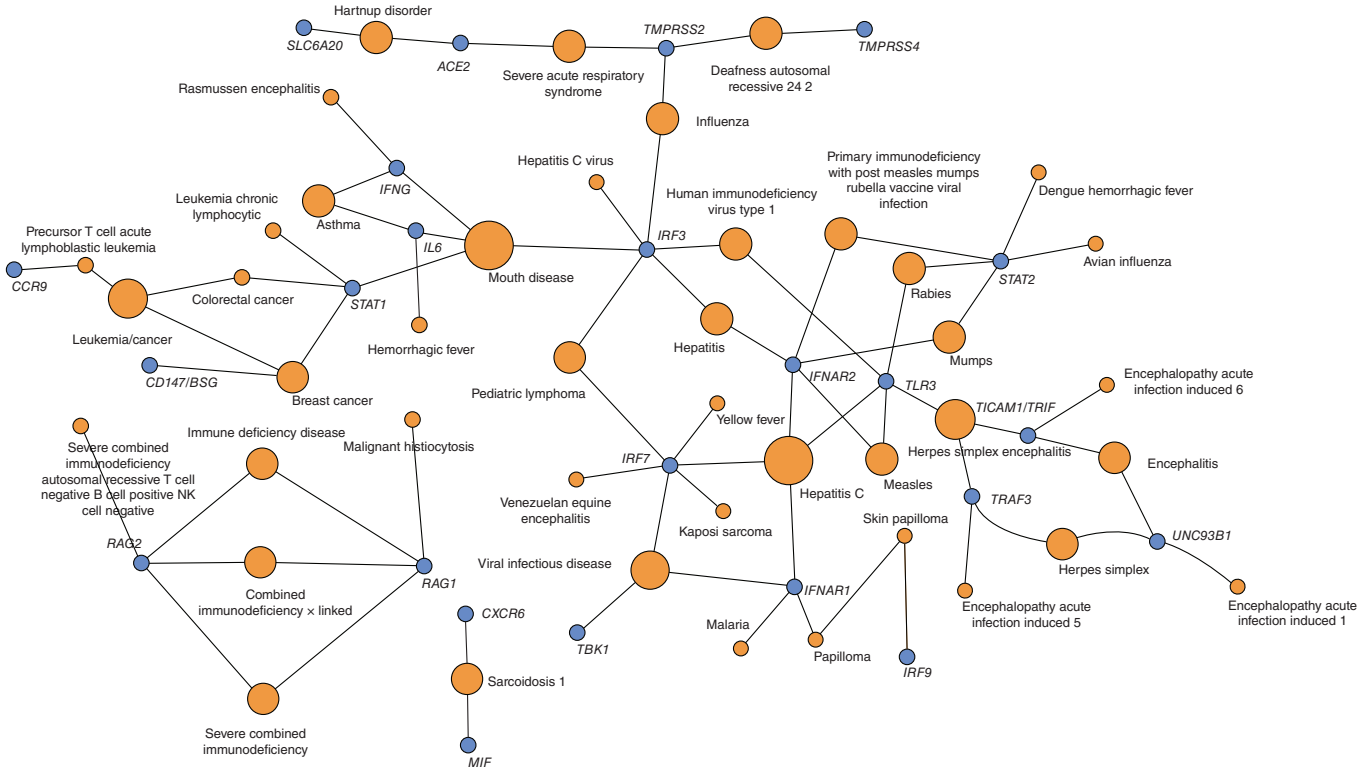


Figure 4. Phenotyping, network and pathway analysis of genes linked with common diseases. Orange circles represent diseases and blue circles represent genes.

Diseases/Genes	TMPRSS4	CD147/BSG	CXCR6	CCR9	SLC6A20	IRF9	TBK1	ACE2	MIF	TMPRSS2	IL6	TRAF3	IFNG	UNC93B1	TICAM1/TRIF	STAT1	STAT2	IFNAR1	RAG1	RAG2	IFNAR2	TLR3	IRF7	IRF3
Deafness autosomal recessive 24 2	X									X														
Leukemia		X		X									X			X								
Sarcoidosis 1			X						X															
Hartnup disorder					X			X																
Papilloma						X												X						
Viral infectious disease							X											X					X	X
Severe acute respiratory syndrome								X		X														
Malaria									X									X						
Influenza										X							X							X
Asthma											X		X											
Fever											X						X						X	
Mouth disease											X		X			X								X
Herpes simplex												X		X	X							X		
Encephalopathy acute infection induced												X		X	X							X		
Encephalitis													X	X	X								X	
Mumps																X								
Rabies																	X							
Hepatitis																		X				X	X	X
Combined immunodeficiency × linked																			X	X				
Immune deficiency disease																			X	X				
Combined cellular and humoral immune defects with granulomas																			X	X				
Severe combined immunodeficiency																			X	X				
Malignant histiocytosis																			X	X				
Measles																					X	X		
Kaposi sarcoma																							X	X
Pediatric lymphoma																							X	X
Human immunodeficiency virus type 1																						X		X
COVID-19	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Figure 5. Gene-disease comparative analysis of variants among genes with COVID-19 symptomatology.

10.2217/pme-2021-0132

The genomic and transcriptomic basis of the complications associated with COVID-19 has not yet been fully determined [107]. We speculate that COVID-19 patients with inborn errors of *IFNAR1*, *IFNAR2*, *TLR3*, *IRF7* and *IRF3* may also be indicative of viral hepatitis. However, genes with identified pathogenic mutations have also been reported to show hepatic expression of *ACE2*, *TMPRSS4* and *TMPRSS2* and have been implicated in chronic liver damage. Various studies have reported varying degrees of liver damage with high levels of alanine aminotransferase (ALT) or aspartate aminotransferase (AST) in patients infected with SARS-CoV-2 [19,108]. Genetic testing can be helpful for accurate clinical decision-making for the diagnosis and prediction of symptomatic individuals with COVID-19 [109]. It can support the definition of pharmacogenetic profiles, guiding treatment and reproductive genetic counseling [110]. Substantial cosegregation of gene-variants with COVID-19 disease can provide solid genetic relationships to support pathogenicity [111]. Most of the symptomologies of COVID-19 are based on common signs and symptom-driven diagnosis and treatment [112]. Therefore, it is important that physicians and patients understand and correctly interpret the relevant findings of clinical-genomics research to COVID-19 disease. The clinical significance of a known immunity gene in association with COVID-19 can be guided by functional change (gain or loss) and clinical conditions [113]. Variant interpretation guidelines are also standardized by the American College of Medical Genetics and Genomics (ACMG) [114–116].

With current technological advancements and reduced cost in high-throughput sequencing, high-quality next-generation sequencing (NGS) data can be quickly produced, processed and analyzed [14]. However, the field of genomics is struggling with the successful implementation of gold standard machine learning (ML) algorithms for clinically proven reproducible computational predictions [38]. It is necessary to automate the process of gene-variant data annotation, expression and simulation to produce timely presentable results [117]. Current limitations in this context imply gaps among clinics and fundamental basic and applied research; difficulties in getting exigent approvals and timeliness of data availability; levels of granularity in clinical information; and application of appropriate modeling strategies that allow learning in the data continuum [118]. Robust scientific solutions are needed in everyday clinical and public health practices for intelligent and integrative analysis of clinical and multiomics data with the application of artificial intelligence (AI) [112]. Layering ML techniques to quantify and annotate the leading indicators in population data will enhance the public health response to bring resources and prepare regions for what is to come [112]. Gene expression analysis is planned as a follow-up to this work to identify enrichment of these immune genes in COVID-19 patients and validate that genetics may determine the clinical course of the infection. This will improve the identification of genetic alterations facilitating COVID-19 disease mechanisms, leading to new predictive models, pinpointing major causes of morbidity and mortality, building personalized therapies and reducing medical costs.

A generation of genomic surveillance data with deep immune genotyping is the need of the hour, as most COVID-19 diagnostics are stranded on clinical assays [112]. Hyper-inflammatory factors and their response to drugs that can target them and block inflammation require investigation [119,120]. It is conceivable that even after clearing the virus, the residual viral genetic material and particles persistently elicit an immune response with variable degrees of symptomology that lead to fatigue, myalgia and neurological issues. The potential of monoclonal antibodies, gamma-globulins and convalescent plasma usage is significant in bringing mortality down, but its administration is late in treatment. A sophisticated analysis is needed based on deep phenotyping of inflammatory markers along with gene expression at different time points. Furthermore, it is important to anticipate the evolution of more COVID-19 mutants with variations in lineages. Comprehensive genome and transcriptome sequencing of large cohorts is needed to determine the prevalence of these factors [121].

Conclusion

To implement effective precision medicine with enhanced ability to positively impact COVID-19 patient outcomes and provide real-time decision support, it is important to examine disease-causing variants and genotype and phenotype associations. The need to investigate disease-causing variants and genotype and phenotype associations among the COVID-19 population data to find the root cause of uncertainties in patient care, such as genetic variants possibly associated with phenotypic manifestations, has become critical. In this study, the authors investigated underlying immunity genes, including their implications in complex diseases and their sequel relationships with COVID-19. Sequence alignment data analysis was performed for known immunity genes revealing inborn mutation errors that may be responsible for complications in COVID-19. The analysis shows the separation of subsets of COVID-19 patients with significantly variable expression for a cluster of genes. The clinical significance of these known immunity genes includes their implications in other complex diseases and their possible sequel relationships

with COVID-19. To better understand the gene-disease organization, progression and network, an ontology was also created. Gene-variants were annotated for mutation type and protein change using algorithms and tools for interpreting mutations.

Summary points

- Investigating COVID-19 disease-causing variants among highly expressed genes enables the determination of root causes of uncertainties in patient care.
- Informative exposure signatures can help in assessing the associations of the COVID-19 transcriptome, offering new insights into the biological and pathological underpinnings of health disparities.
- The current analysis shows subsets of COVID-19 patients with significantly variable expression for a cluster of genes.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/pme-2021-0132

Author contributions

Z Ahmed proposed, designed and lead the study. Z Ahmed and S Zeeshan designed the WGS/WES data processing and analysis pipeline and Z Ahmed and EG Renart implemented the pipeline. Z Ahmed, S Zeeshan and EG Renart performed data analysis and visualization. Z Ahmed drafted the manuscript and S Zeeshan and EG Renart participated in writing and review. All authors approved the paper.

Acknowledgments

The authors appreciate the great support from the Rutgers Institute for Health, Health Care Policy and Aging Research (IFH); Department of Medicine, Rutgers Robert Wood Johnson Medical School (RWJMS); and Rutgers Biomedical and Health Sciences (RBHS), at Rutgers, The State University of New Jersey. The authors also thank members and collaborators of the Ahmed lab at Rutgers (IFH, RWJMS, RBHS) for their support, participation and contribution to this study.

The authors acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey for providing access to the Amarel cluster and associated research computing resources that have contributed to the results reported here.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants, or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

All project data were completely anonymous and collected from public resources. The study was conducted in accordance with relevant guidelines and regulations, and protocols approved by Institutional Review Board (IRB), Rutgers University.

References

1. Lederberg J. Infectious history. *Science* 288(5464), 287–293 (2000).
2. Brachman PS. Infectious diseases–past, present, and future. *Int. J. Epidemiol.* 32(5), 684–686 (2003).
3. Busnadiego I, Fernbach S, Pohl MO *et al.* Antiviral activity of type I, II, and III interferons counterbalances*ACE2* inducibility and restricts SARS-CoV-2. *mBio* 11(5), e01928–20 (2020).
4. Mousavizadeh L, Ghasemi S. Genotype and phenotype of COVID-19: their roles in pathogenesis. *J. Microbiol. Immunol. Infect.* 54(2), 159–163 (2020).

5. Zhu N, Zhang D, Wang W *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382(8), 727–733 (2020).
6. Wang L, Wang Y, Ye D, Liu Q. Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence. *Int. J. Antimicrob. Agents* 55(6), 105948 (2020).
7. Wang H, Li X, Li T *et al.* The genetic sequence, origin, and diagnosis of SARS-CoV-2. *Eur. J. Clin. Microbiol. Infect. Dis.* 39(9), 1629–1635 (2020).
8. Weiss SR, Leibowitz JL. Coronavirus pathogenesis. *Adv. Virus Res.* 81, 85–164 (2011).
9. Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* 3(1), 237–261 (2016).
10. Helmy YA, Fawzy M, Elaswad A *et al.* The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *J. Clin. Med.* 9(4), 1225 (2020).
11. Zhu N, Zhang D, Wang W *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382(8), 727–733 (2020).
12. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 323(13), 1239–1242 (2020).
13. Ahmed Z, Zeeshan S, Foran DJ *et al.* Integrative clinical, genomics and metabolomics data analysis for mainstream precision medicine to investigate COVID-19. *BMJ Innov.* 7, 6–10 (2021).
14. Ahmed Z. Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis. *Hum. Genomics* 14(1), 35 (2020).
15. Zhai P, Ding Y, Wu X *et al.* The epidemiology, diagnosis and treatment of COVID-19. *Int. J. Antimicrob. Agents* 55(5), 105955 (2020).
16. Helms J, Kremer S, Merdji H *et al.* Neurologic features in severe SARS-CoV-2 infection. *N. Engl. J. Med.* 382(23), 2268–2270 (2020).
17. Chen G, Wu D, Guo W *et al.* Clinical and immunological features of severe and moderate coronavirus disease 2019. *J. Clin. Invest.* 130(5), 2620–2629 (2020).
18. Qiu H, Wu J, Hong L *et al.* Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study. *Lancet Infect. Dis.* 20(6), 689–696 (2020).
19. Chen N, Zhou M, Dong X *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 395(10223), 507–513 (2020).
20. Tian S, Hu N, Lou J *et al.* Characteristics of COVID-19 infection in Beijing. *J. Infect.* 80(4), 401–406 (2020).
21. Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med.* 46(4), 586–590 (2020).
22. Baig AM, Khaleeq A, Ali U, Syeda H. Evidence of the COVID-19 virus targeting the CNS: tissue distribution, host-virus interaction, and proposed neurotropic mechanisms. *ACS Chem. Neurosci.* 11(7), 995–998 (2020).
23. Zang R, Gomez Castro MF, McCune BT *et al. TMPRSS2* and *TMPRSS4* promote SARS-CoV-2 infection of human small intestinal enterocytes. *Sci. Immunol.* 5(47), eabc3582 (2020).
24. Choudhary S, Sreenivasulu K, Mitra P, Misra S, Sharma P. Role of genetic variants and gene expression in the susceptibility and severity of COVID-19. *Ann. Lab. Med.* 41(2), 129–138 (2021).
25. Anastassopoulou C, Gkizarioti Z, Patrinos GP, Tsakris A. Human genetic factors associated with susceptibility to SARS-CoV-2 infection and COVID-19 disease severity. *Hum. Genomics* 14(1), 40 (2020).
26. Debnath M, Banerjee M, Berk M. Genetic gateways to COVID-19 infection: implications for risk, severity, and outcomes. *FASEB J.* 34(7), 8787–8795 (2020).
27. Sironi M, Hasnain SE, Rosenthal B *et al.* SARS-CoV-2 and COVID-19: a genetic, epidemiological, and evolutionary perspective. *Infect. Genet. Evol.* 84, 104384 (2020).
28. Zhang Q, Bastard P, Liu Z *et al.* Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* 370(6515), eabd457 (2020).
29. Busnadiego I, Fernbach S, Pohl MO *et al.* Antiviral activity of type I, II, and III interferons counterbalances *ACE2* inducibility and restricts SARS-CoV-2. *mBio* 11(5), e01928–20 (2020).
30. Bastard P, Rosen LB, Zhang Q *et al.* Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* 370(6515), eabd4585 (2020).
31. Roberts CM, Levi M, McKee M, Schilling R, Lim WS, Grocott MPW. COVID-19: a complex multisystem disorder. *Br. J. Anaesth.* 125(3), 238–242 (2020).
32. Morens DM, Fauci AS. Emerging pandemic diseases: how we got to COVID-19. *Cell* 182(5), 1077–1092 (2020).
33. de Sousa E, Ligeiro D, Lérias JR *et al.* Mortality in COVID-19 disease patients: correlating the association of major histocompatibility complex (MHC) with severe acute respiratory syndrome 2 (SARS-CoV-2) variants. *Int. J. Infect. Dis.* 98, 454–459 (2020).

34. Hou Y, Zhao J, Martin W *et al.* New insights into genetic susceptibility of COVID-19: an *ACE2* and *TMPRSS2* polymorphism analysis. *BMC Med.* 18(1), 216 (2020).
35. Sardar R, Satish D, Birla S, Gupta D. Integrative analyses of SARS-CoV-2 genomes from different geographical locations reveal unique features potentially consequential to host-virus interaction, pathogenesis and clues for novel therapies. *Heliyon* 6(9), e04658 (2020).
36. Voinsky I, Gurwitz D. Smoking and COVID-19: Similar bronchial *ACE2* and *TMPRSS2* expression and higher *TMPRSS4* expression in current versus never smokers. *Drug Dev. Res.* (2020) (Epub ahead of print).
37. Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F *et al.* Genomewide association study of severe COVID-19 with respiratory failure. *N. Engl. J. Med.* 383(16), 1522–1534 (2020).
38. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database* 2020, baaa010 (2020).
39. Donoghue M, Hsieh F, Baronas E *et al.* A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9. *Circ. Res.* 87(5), E1–E9 (2000).
40. Lan J, Ge J, Yu J *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581(7807), 215–220 (2020).
41. Hoffmann M, Kleine-Weber H, Schroeder S *et al.* SARS-CoV-2 cell entry depends on *ACE2* and *TMPRSS2* and is blocked by a clinically proven protease inhibitor. *Cell* 181(2), 271–280.e8 (2020).
42. de Aberasturi AL, Calvo A. TMPRSS4: an emerging potential therapeutic target in cancer. *Br. J. Cancer* 112(1), 4–8 (2015).
43. Valero-Jiménez A, Zúñiga J, Cisneros J *et al.* Transmembrane protease, serine 4 (TMPRSS4) is upregulated in IPF lungs and increases the fibrotic response in bleomycin-induced lung injury. *PLoS ONE* 13(3), e0192963 (2018).
44. Paoloni-Giacobino A, Chen H, Peitsch MC, Rossier C, Antonarakis SE. Cloning of the *TMPRSS2* gene, which encodes a novel serine protease with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3. *Genomics* 44(3), 309–320 (1997).
45. Maho A, Bensimon A, Vassart G, Parmentier M. Mapping of the *CCXCR1*, *CX3CR1*, *CCBP2* and *CCR9* genes to the CCR cluster within the 3p21.3 region of the human genome. *Cytogenet. Cell Genet.* 87(3–4), 265–268 (1999).
46. Zaballos A, Gutiérrez J, Varona R, Ardavín C, Márquez G. Cutting edge: identification of the orphan chemokine receptor GPR-9-6 as CCR9, the receptor for the chemokine TECK. *J. Immunol.* 162(10), 5671–5675 (1999).
47. Carramolino L, Zaballos A, Kremer L *et al.* Expression of CCR9 beta-chemokine receptor is modulated in thymocyte differentiation and is selectively maintained in CD8(+) T cells from secondary lymphoid organs. *Blood* 97(4), 850–857 (2001).
48. Schmutz C, Cartwright A, Williams H *et al.* Monocytes/macrophages express chemokine receptor CCR9 in rheumatoid arthritis and CCL25 stimulates their differentiation. *Arthritis. Res. Ther.* 12(4), R161 (2010).
49. Tu Z, Xiao R, Xiong J *et al. CCR9* in cancer: oncogenic role and therapeutic targeting. *J. Hematol. Oncol.* 9, 10 (2016).
50. Huang Y, Ding HS, Song T *et al.* Abrogation of CC chemokine receptor 9 ameliorates ventricular electrical remodeling in mice after myocardial infarction. *Front. Cardiovasc. Med.* 8, 716219 (2021).
51. Schroder K, Hertzog PJ, Ravasi T, Hume DA. Interferon-gamma: an overview of signals, mechanisms and functions. *J. Leukoc. Biol.* 75(2), 163–189 (2004).
52. Schoenborn JR, Wilson CB. Regulation of interferon-gamma during innate and adaptive immune responses. *Adva. Immunol.* 96, 41–101 (2007).
53. Kuang Y, Wang S, Tang L, Hai J, Yan G, Liao L. Cluster of differentiation 147 mediates chemoresistance in breast cancer by affecting vacuolar H+-ATPase expression and activity. *Oncol. Lett.* 15(5), 7279–7290 (2018).
54. Zhao Y, Zhang CL, Zeng BF, Wu XS, Gao TT, Oda Y. Enhanced chemosensitivity of drug-resistant osteosarcoma cells by lentivirus-mediated Bcl-2 silencing. *Biochem. Biophys. Res. Commun.* 390(3), 642–647 (2009).
55. Ashhurst AS, Flórido M, Lin LCW *et al.* CXCR6-deficiency improves the control of pulmonary mycobacterium tuberculosis and influenza infection independent of T-lymphocyte recruitment to the lungs. *Front. Immunol.* 10, 339 (2019).
56. Kim CH, Kunkel EJ, Boisvert J *et al.* Bonzo/CXCR6 expression defines type 1-polarized T-cell subsets with extralymphoid tissue homing potential. *J. Clin. Invest.* 107(5), 595–601 (2001).
57. Pankiv S, Alemu EA, Brech A *et al.* FYCO1 is a Rab7 effector that binds to LC3 and PI3P to mediate microtubule plus end-directed vesicle transport. *J. Cell Biol.* 188(2), 253–269 (2010).
58. Tanaka T, Narazaki M, Kishimoto T. IL-6 in inflammation, immunity, and disease. *Cold Spring Harb. Perspect. Biol.* 6(10), a016295 (2014).
59. Gabay C. Interleukin-6 and chronic inflammation. *Arthritis. Res. Ther.* 8(Suppl. 2), S3 (2006).
60. Seo S, Zhang Q, Bugge K *et al.* A novel protein LZTFL1 regulates ciliary trafficking of the BBSome and Smoothened. *PLoS Genet.* 7(11), e1002358 (2011).
61. Calandra T, Froidevaux C, Martin C, Roger T. Macrophage migration inhibitory factor and host innate immune defenses against bacterial sepsis. *J. Infect. Dis.* 187(Suppl. 2), S385–S390 (2003).

62. Bröer S, Gether U. The solute carrier 6 family of transporters. *Br. J. Pharmacol.* 167(2), 256–278 (2012).
63. Lei Y, Takahama Y. XCL1 and XCR1 in the immune system. *Microbes Infect.* 14(3), 262–267 (2012).
64. Rosenfeld CS, Han CS, Alexenko AP, Spencer TE, Roberts RM. Expression of interferon receptor subunits, IFNAR1 and IFNAR2, in the ovine uterus. *Biol. Reprod.* 67(3), 847–853 (2002).
65. Qian J, Zheng H, Huangfu WC *et al.* Pathogen recognition receptor signaling accelerates phosphorylation-dependent degradation of IFNAR1. *PLoS Pathog.* 7(6), e1002065 (2011).
66. DeFilippis VR, Robinson B, Keck TM, Hansen SG, Nelson JA, Früh KJ. Interferon regulatory factor 3 is necessary for induction of antiviral genes during human cytomegalovirus infection. *J. Virol.* 80(2), 1032–1037 (2006).
67. Ning S, Pagano JS, Barber GN. IRF7: activation, regulation, modification and function. *Genes Immun.* 12(6), 399–414 (2011).
68. Fink K, Grandvaux N. STAT2 and IRF9: Beyond ISGF3. *JAKSTAT* 2(4), e27521 (2013).
69. Au-Yeung N, Mandhana R, Horvath CM. Transcriptional regulation by STAT1 and STAT2 in the interferon JAK-STAT pathway. *JAKSTAT* 2(3), e23931 (2013).
70. Schubert C, Allhoff M, Tillmann S *et al.* Differential roles of STAT1 and STAT2 in the sensitivity of JAK2V617F- vs. BCR-ABL-positive cells to interferon alpha. *J. Hematol. Oncol.* 12(1), 36 (2019).
71. Shu C, Sankaran B, Chaton CT *et al.* Structural insights into the functions of TBK1 in innate antimicrobial immunity. *Structure* 21(7), 1137–1148 (2013).
72. Seya T, Oshiumi H, Sasai M, Akazawa T, Matsumoto M. TICAM-1 and TICAM-2: toll-like receptor adapters that participate in induction of type 1 interferons. *Int. J. Biochem. Cell Biol.* 37(3), 524–529 (2005).
73. Oshiumi H, Matsumoto M, Funami K, Akazawa T, Seya T. TICAM-1, an adaptor molecule that participates in Toll-like receptor 3-mediated interferon-beta induction. *Nat. Immunol.* 4(2), 161–167 (2003).
74. Zhang SY, Herman M, Ciancanelli MJ *et al.* TLR3 immunity to infection in mice and humans. *Curr. Opin. Immunol.* 25(1), 19–33 (2013).
75. He JQ, Oganesyan G, Saha SK, Zarnegar B, Cheng G. TRAF3 and its biological function. *Adv. Exp. Med. Biol.* 597, 48–59 (2007).
76. Bishop GA, Stunz LL, Hostager BS. TRAF3 as a multifaceted regulator of B lymphocyte survival and activation. *Front. Immunol.* 9, 2161 (2018).
77. Lee BL, Moon JE, Shu JH *et al.* UNC93B1 mediates differential trafficking of endosomal TLRs. *Elife* 2, e00291 (2013).
78. Kim YM, Brinkmann MM, Paquet ME, Ploegh HL. UNC93B1 delivers nucleotide-sensing toll-like receptors to endolysosomes. *Nature* 452(7184), 234–238 (2008).
79. Akamatsu Y, Oettinger MA. Distinct roles of RAG1 and RAG2 in binding the V(D)J recombination signal sequences. *Mol. Cell. Biol.* 18(8), 4670–4678 (1998).
80. Landree MA, Wibbenmeyer JA, Roth DB. Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes Dev.* 13(23), 3059–3069 (1999).
81. Ahmed Z, Renart EG, Mishra D, Zeeshan S. JWES: a new pipeline for whole genome/exome sequence data processing, management, and gene-variant discovery, annotation, prediction, and genotyping. *FEBS Open Bio.* 11(9), 2441–2452 (2021).
82. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front. Genet.* 4, 288 (2013).
83. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14), 1754–1760 (2009).
84. Jo H, Koh G. Faster single-end alignment generation utilizing multi-thread for BWA. *Biomed. Mater. Eng.* 26(Suppl. 1), S1791–S1796 (2015).
85. Ebbert MT, Wadsworth ME, Staley LA *et al.* Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics.* 17(Suppl. 7), 239 (2016).
86. Li H, Handsaker B, Wysoker A *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16), 2078–2079 (2009).
87. McKenna A, Hanna M, Banks E *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9), 1297–1303 (2010).
88. Brouard JS, Schenkel F, Marete A, Bissonnette N. The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. *J. Anim. Sci. Biotechnol.* 10, 44 (2019).
89. Franke KR, Crowgey EL. Accelerating next generation sequencing data analysis: an evaluation of optimized best practices for genome analysis toolkit algorithms. *Genomics Inform.* 18(1), e10 (2020).
90. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31(13), 3812–3814 (2003).
91. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 40(Web Server issue), W452–W457 (2012).

92. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4(7), 1073–1081 (2009).
93. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* Chapter 7, Unit 7.20 (2013).
94. Gnad F, Baucom A, Mukhyala K, Manning G, Zhang Z. Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics* 14(Suppl. 3), S7 (2013).
95. Ahmed Z, Zeeshan S, Mendhe D, Dong X. Human gene and disease associations for clinical-genomics and precision medicine research. *Clin. Transl. Med.* 10(1), 297–318 (2020).
96. Vohra S, Biggin PC. Mutationmapper: a tool to aid the mapping of protein mutation data. *PLoS One* 8(8), e71711 (2013).
97. Zhang W, Wang C, Zhang X. Mutplot: an easy-to-use online tool for plotting complex mutation data with flexibility. *PLoS One* 14(5), e0215838 (2019).
98. Ahmed Z, Zeeshan S, Xiong R, Liang BT. Debutant iOS app and gene-disease complexities in clinical genomics and precision medicine. *Clin. Transl. Med.* 8(1), 26 (2019).
99. Zeeshan S, Xiong R, Liang BT, Ahmed Z. 100 Years of evolving gene-disease complexities and scientific debutants. *Brief. Bioinform.* 21(3), 885–905 (2020).
100. Patel P, DeCuir J, Abrams J, Campbell AP, Godfred-Cato S, Belay ED. Clinical characteristics of multisystem inflammatory syndrome in adults: a systematic review. *JAMA Netw. Open* 4(9), e2126456 (2021).
101. Mehta OP, Bhandari P, Raut A, Kacimi SEO, Huy NT. Coronavirus disease (COVID-19): comprehensive review of clinical presentation. *Front. Public Health* 8, 582932 (2021).
102. Vandenberg O, Martiny D, Rochas O, van Belkum A, Kozlakidis Z. Considerations for diagnostic COVID-19 tests. *Nat. Rev. Microbiol.* 19(3), 171–183 (2021).
103. Udugama B, Kadhiresan P, Kozlowski HN *et al.* Diagnosing COVID-19: the disease and tools for detection. *ACS Nano* 14(4), 3822–3835 (2020).
104. Liu P, Meng L, Normand EA *et al.* Reanalysis of clinical exome sequencing data. *N. Engl. J. Med.* 380(25), 2478–2480 (2019).
105. Bick D, Jones M, Taylor SL, Taft RJ, Belmont J. Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases. *J. Med. Genet.* 56(12), 783–791 (2019).
106. Cirulli ET, White S, Read RW *et al.* Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat. Commun.* 11(1), 542 (2020).
107. Gagliardi S, Poloni ET, Pandini C *et al.* Detection of SARS-CoV-2 genome and whole transcriptome sequencing in frontal cortex of COVID-19 patients. *Brain Behav. Immun.* 97, 13–21 (2021).
108. Wu J, Song S, Cao HC, Li LJ. Liver diseases in COVID-19: etiology, treatment and prognosis. *World J. Gastroenterol.* 26(19), 2286–2293 (2020).
109. Xu M, Ouyang L, Han L *et al.* Accurately differentiating between patients with COVID-19, patients with other viral infections, and healthy individuals: multimodal late fusion learning approach. *J. Med. Internet Res.* 23(1), e25535 (2021).
110. Wang LY, Cui JJ, OuYang QY *et al.* Complex analysis of the personalized pharmacotherapy in the management of COVID-19 patients and suggestions for applications of predictive, preventive, and personalized medicine attitude. *EPMA J.* 12(3), 1–18 (2021).
111. D'Antonio M, Nguyen JP, Arthur TD *et al.* SARS-CoV-2 susceptibility and COVID-19 disease severity are associated with genetic variants affecting gene expression in a variety of tissues. *Cell Rep.* 37(7), 110020 (2021).
112. Ahmed Z. Intelligent health system for the investigation of consenting COVID-19 patients and precision medicine. *Per. Med.* 18(6), 573–582 (2021).
113. Duzkale H, Shen J, McLaughlin H *et al.* A systematic approach to assessing the clinical significance of genetic variants. *Clin. Genet.* 84(5), 453–463 (2013).
114. Richards CS, Bale S, Bellissimo DB *et al.* ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet. Med.* 10(4), 294–300 (2008).
115. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet. Med.* 13(7), 680–685 (2011).
116. Rehm HL, Bale SJ, Bayrak-Toydemir P *et al.* ACMG clinical laboratory standards for next-generation sequencing. *Genet. Med.* 15(9), 733–747 (2013).
117. Chen L, Li Z, Zeng T *et al.* Identifying COVID-19-specific transcriptomic biomarkers with machine learning methods. *Biomed. Res. Int.* 2021, 9939134 (2021).
118. Ahmed Z, Renart EG, Zeeshan S. Genomics pipelines to investigate susceptibility in whole genome and exome sequenced data for variant discovery, annotation, prediction and genotyping. *Peer J.* 9, e11724 (2021).

119. Gustine JN, Jones D. Immunopathology of hyperinflammation in COVID-19. *Am. J. Pathol.* 191(1), 4–17 (2021).

120. Tan LY, Komarasamy TV, Rmt Balasubramaniam V. Hyperinflammatory immune response and COVID-19: a double edged sword. *Front. Immunol.* 12, 742941 (2021).

121. Chen AT, Altschuler K, Zhan SH, Chan YA, Deverman BE. COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest. *Elife* 10, e63409 (2021).