ACE2 Netlas: In-silico functional characterization and drug-gene interactions of *ACE2* gene network to understand its potential involvement in COVID-19 susceptibility

Gita A Pathak, Frank R Wendt, Aranyak Goswami, Flavio De Angelis, COVID-19 Human Genetics Initiative, Renato Polimanti

Yale School of Medicine, Department of Psychiatry, Division of Human Genetics, New Haven, CT
Veteran Affairs Connecticut Healthcare System, West Haven, CT

✉ Corresponding author: Renato Polimanti, PhD
E-mail: renato.polimanti@yale.edu
Address: VA CT 116A2, 950 Campbell Avenue
West Haven, CT, 06516, USA
Phone: +1 (203) 937-5711 *ext. ext.5745*
Fax: +1 (203) 937-3897

Running Title: Characterization of *ACE2*- gene network

Abstract

Angiotensin-converting enzyme-2 (*ACE2*) receptor has been identified as the key adhesion molecule for the transmission of the SARS-CoV-2. However, there is no evidence that human genetic variation in ACE2 is singularly responsible for COVID-19 susceptibility. Therefore, we performed a multi-level characterization of genes that interact with ACE2 (ACE2-gene network) for their over-represented biological properties in the context of COVID-19.

The phenome-wide association of 51 genes including ACE2 with 4,756 traits categorized into 26 phenotype categories, showed enrichment of immunological, respiratory, environmental, skeletal, dermatological, and metabolic domains (p<4e-4). Transcriptomic regulation of ACE2-gene network was enriched for tissue-specificity in kidney, small intestine, and colon (p<4.7e-4). Leveraging the drug-gene interaction database we identified 47 drugs, including dexamethasone and spironolactone, among others.

Considering genetic variants within ± 10 kb of ACE2-network genes we characterized functional consequences (among others) using miRNA binding-site targets. MiRNAs affected by ACE2-network variants revealed statistical over-representation of inflammation, aging, diabetes, and heart conditions. With respect to variants mapped to the ACE2-network, we observed COVID-19 related associations in *RORA*, *SLC12A6* and *SLC6A19* genes.

Overall, functional characterization of ACE2-gene network highlights several potential mechanisms in COVID-19 susceptibility. The data can also be accessed at https://gpwhiz.github.io/ACE2Netlas/

Keywords: ACE2, COVID-19, miRNA, immune response, network

1 Introduction

SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) is the causative agent responsible for recent global spread of COVID-19 (coronavirus disease 2019) [1,2]. Millions of people have been infected with the virus, which caused global lockdowns and heavily restricted interpersonal contact. These measures were taken to reduce viral spread through respiratory droplet exchange between persons.

SARS-CoV-2 is capable of entering the host cells via ACE2 (angiotensin converting enzyme 2) [3]. ACE2 is found on many different cell types, which normally helps regulate blood pressure and inflammation through cleavage of angiotensin II (ANG II) [4]. The virus occupies cell-surface of *ACE2* leading to accumulation of angiotensin (ANGII), inflammation, and cell death [3]. In the lungs, SARS-CoV-2 mediated ANGII accumulation leads to alveolar cell death and a reduction in oxygen uptake [5].

Although ACE2 is the cellular entry point, there is little evidence that genetic variation in *ACE2* is singularly responsible for COVID-19 susceptibility. Indeed, *ACE2* failed to associate with COVID-19 informative phenotype definitions from large genome-wide studies [6–8] . However, due to the functional role of ACE2 in SARS-CoV-2 infection, we hypothesize that genes interacting with ACE2 activity are enriched for molecular pathways relevant for COVID-19 susceptibility. Accordingly, we employed a top-down approach to analyze tissue-specific transcriptomic regulation, drug-gene interactions, and variant prioritization using genetic variants within the ACE2 gene-gene connectome and protein-protein interaction networks. With this approach we identified several biological processes and functional effects of ACE2-gene network relevant for the vast symptoms observed following SARS-CoV-2 infection.

2 Results

A study overview is presented in Supplementary file1 Figure S1.

2.1 The ACE2 gene connectome

A total of 60 ACE2-interacting genes were identified from different network databases (Supplementary file2 Table S1; Figure 1).

2.2 Tissue-specific transcriptomic regulation

Using differential expression data of 54 tissues (GTEx-v8), the genes in the ACE2-gene network were enriched for upregulated expression in small intestine ($p=1.07x10^{-16}$), colon ($p =7.60x10^{-13}$), kidney ($p=1.93x10^{-8}$), and liver ($p=4.63x10^{-4}$) (Figure 2; Supplementary file2 Table S2). No tissue-specific enrichment was observed for down-regulated expression.

2.3 Gene-Drug Interaction and Over-represented Biological Functions

Out of 61 genes, 29 had information about their drug-gene interaction in the drug-gene interaction database (DGIdb)[9] . This assessment resulted in 238 unique drug-gene observations (Supplementary file2 Table S3). Some of the notable drugs observed via this approach were spironolactone, dexamethasone, metformin, and hydrocortisone. To understand the role of these drugs in affecting biological processes, we performed drug-set enrichment analysis. DSEA [10] found gene-ontology mapping for 47 drugs and tested against REACTOME gene ontology database. Although the results did not survive Bonferroni correction, the strongest enrichments were observed for platelet sensitization by low-density lipoprotein cholesterol (p=0.003), IL-7 signaling (p=0.004), glycerophospholipid biosynthesis (p=0.005), and viral messenger RNA synthesis (p=0.011) (Figure 3; Supplementary file2 Table S4).

2.4 Over-representation of phenotypic domains within ACE2 gene network

A phenome-wide association study (PheWAS) was performed for 51 genes leveraging data from the GWASAtlas [11]. The GWASAtlas categorized traits into 26 phenotype domains (Supplementary file1 Figures S2-S52; Supplementary file2 Table S5). Each domain was tested for enrichment of significant traits versus non-significant traits (Supplementary file2 Table S6). Six domains were significant: 'Immunological' ($p=7.62x10^{-25}$), 'Respiratory' ($p=1.30x10^{-8}$), 'Skeletal' ($2.94x10^{-8}$), 'Dermatological' ($p=7.91x10^{-8}$), 'Environmental' ($p=2.21x10^{-7}$), and 'Metabolic' ($4.33x10^{-4}$) (Supplementary

file2 Table S7). *SLC44A4* had the highest number of associated traits across the significant domains (n_{total}= 173) followed by *APOA1* had highest number of traits associations, mostly metabolic (n_{total}= 100; metabolic = 71) (Figure 4). *SLC44A4*, *APOA1*, and *RORA* showed associations across all six enriched domains.

2.5 Characterization of SNPs

We extracted all 957,222 SNPs in the ACE2-network and annotated for allele frequency (Supplementary file3), nearby genes and coordinates (Supplementary file4), Combined Annotation Dependent Depletion (CADD) [12] and DeepSEA [13] scores. There were 98,529 SNPs with CADD score >10, which corresponds to the top 10% pathogenic variants across the human genome (Supplementary file5). To identify their regulatory consequences, variants were annotated with DeepSEA which provides functional probability of the SNPs in serving as gene expression, disease and chromatin regulating variants. There were 12,095 SNPs within the ACE2-gene network which had >50% functional probability (DeepSEA functional score > 0.5) (Supplementary file6). The miRNAs altered by the SNPs were analyzed for over-represented miRNA-family, biological functions, and diseases considering false discovery rate multiple testing correction (FDR p<0.05). There were 4 miRNA clusters that were enriched, miR-302b, miR-181d (p=0.0079), miR-17, and 106a (p= 0.00298). We found 65 biological functions that were significant and the top five significant biological processes were cell death ($p=1.5x10^{-20}$), inflammation ($p=2.57x10^{-20}$), cell cycle ($p=2.09x10^{-18}$), apoptosis ($p=4.15x10^{-18}$), and immune response ($p=3.17x10^{-17}$) (Figure 5). We observed a total of 152 significant diseases of which the most significant were diabetes mellitus type 2 ($p=1.55x10^{-22}$), hepatitis c virus infection ($p=5.56x10^{-21}$), atherosclerosis ($p=3.08x10^{-19}$), heart failure ($p=4.22x10^{-19}$), and Alzheimer's disease ($p=1.35x10^{-17}$) (Supplementary file2 Table S8).

2.6 Neanderthal LA introgression within ACE2 network SNPs

Due to the Neanderthal introgression observed in 3p21 locus as risk to COVID-19 [14], we compared mean probability of Neanderthal LA between the ACE2-network SNP set (mean=0.032) and 1,000 randomly selected SNP sets with comparable genomic features (range of Neanderthal LA means = 0.027-0.036). The ACE2-network SNPs did not show

evidence of Neanderthal LA introgression significantly different from those expected by chance (p=0.663) (Supplementary file1 Figure 55).

2.7 Annotation of network SNPs using the COVID-19 GWAS

We tested ACE2-network SNPs with respect to six COVID-19-related phenotypes (Freeze 3) released by the COVID-19 Host Genetics Initiative [15]. To identify independent variants, the variants were pruned for linkage disequilibrium (LD<0.1 within 250kb genomic size) and clumped for p-value <0.01. Variants surviving multiple testing were annotated for eQTLs, and mQTLs. Three genes – *RORA*, *SLC12A6*, and *SLC6A19* – showed associations with multiple COVID-19 phenotypes (Supplementary file2 Tables S9-S14; Supplementary file6 Figures S56-S61). *RORA* SNPs were associated with COVID-19 positive status (rs17303202, p=2.35E-5), laboratory-confirmed positive COVID-19 status (rs4774377, p=8.25E-5), hospitalized COVID-19 (rs17303202, p=2.76E-05), and COVID-19 with very severe respiratory symptoms (rs341419, p=8.13E-4). The SNPs in *RORA* gene are also associated with gene expression of *RORA* gene (rs12912196; p=3.9E-5) and trans-mQTL (cg00930615 in *ANXA2*). *SLC12A6* associations were observed with respect to COVID-19 (rs145719616, p=1.19E-4), hospitalized COVID-19 (rs192235418, p=4.42E-4), COVID-19 with very severe respiratory (rs2705343, p=1.86E-3), and. *SLC6A19* SNPs were associated with severe COVID-19 phenotype definitions, i.e. COVID-19 with very severe respiratory confirmed (rs76067074, p=2.65E-3) and hospitalized COVID-19(rs76067074, p=2.52E-4).

3 Discussion

ACE2 is expressed in several tissues and plays a key role in host-entry of SARS-CoV-2 [16]. However, the genomic profile of *ACE2* is limited in explaining the vast symptomology observed for COVID-19. Understanding ACE2 associated molecular networks presents several functional insights between genetic targets based on gene expression, topology, and protein and signaling relationships [17]. Due to the well-characterized role of ACE2 in SARS-CoV-2 infection, we generated novel information regarding the molecular and phenotypic characteristics of ACE gene network in the context of their potential

involvement in COVID-19 susceptibility. Our PheWAS-based analysis showed that genetic variation within ACE2 gene network is associated with immunity, respiratory, and metabolic traits. This is in line with known epidemiology of COVID-19 and its comorbidities [18,19].

The expression of ACE2-network genes was enriched for regulatory mechanisms related to small intestine, colon, kidney, and liver. It is hypothesized that furin, a serine protease present in lungs but also highly expressed in small intestine, presents S-spike for attachment of the ACE2 receptor [20]. Patients with kidney disease have higher risk for COVID-19 severe symptoms [21]. Additionally, the inflammation and cytokine storm from COVID-19 is observed to damage kidney tissues [22]. Lastly, modest increase in liver enzymes has been associated with COVID-19, and returning to baseline during the recovery phase [23].

Understanding the genes that interact with *ACE2* receptor has potential to understand drug-targets and molecular processes that might play a role in susceptibility and treatment response of COVID-19. The drug-gene interaction analysis within ACE2 network identified dexamethasone, reported to lower mortality in COVID-19 cases requiring mechanical ventilation [24]. Drugs – spironolactone and hydrocortisone target the androgen system. The androgen receptor has been associated with severe symptomology of COVID-19 [25]. Spironolactone is a diuretic and alleviates respiratory symptoms by reducing fluid from the lungs [26]. The use of spironolactone is currently being tested for acute respiratory distress syndrome in COVID-19 patients [27]. Hydrocortisone is currently under clinical trials for treating COVID-19 related hypoxia symptoms [28]. Among the other compounds identified, metformin, a known drug for treating diabetes, can also affect respiratory outcomes [29]. A recent study reported protective effects of metformin in women with diabetes and obesity who were admitted with COVID-19 diagnosis [30]. Lastly, melatonin has been hypothesized to improve general immunity and lower oxidative stress generated from SARS-CoV-2 infection [31].

The miRNA target sites altered by ACE2-network SNPs identified miR-302b and miR-181d as over-represented miRNA clusters. The downregulated expression of miR-302b has been observed to reduce survival rates in chronic obstructive pulmonary disease (COPD) patients [32]. A meta-analysis showed that COPD diagnosis increased susceptibility to COVID-19 [33]. The miRNA-181 cluster has been associated with regulation of TNF-alpha [34], T-cell aging [35] and emphysema [36]. miRNA-17 and 106 belong to same miRNA family, miRNA-17 is upregulated in bronchoalveolar stem cells to lower SARS-CoV replication [37]. An *in silico* study of miRNA targets for SARS-CoV-2 genomic sequence found miRNA-17 as one of the targets with experimental evidence of its upregulation in H7N9 Influenza virus infection [38]. The top over-represented diseases in miRNA-ACE2-network-SNPs were diabetes, hepatitis C viral infection, heart failure and Alzheimer's disease. COVID-19 in individuals with diabetes has been reported to require hospitalization than non-diabetic individuals [39]. Furthermore, SARS-CoV-2 infection contributes in the development of ketosis in diabetic individuals resulting in longer length of hospitalization stay [40]. Triglyceride and glucose index was associated with severity of COVID-19 [41]. While there are limited studies about hepatitis C in COVID-19 patients [42], heart failure was reported by multiple studies as being associated with COVID-19 severity [43,44]. Alzheimer's disease is another condition associated with COVID-19 susceptibility [45], including *APOE4* carrier status with increased risk of severe COVID-19 [46].

In contrast to specific enrichment of Neanderthal LA in a COVID-19 risk locus on chromosome 3 [47], there is no evidence of increased Neanderthal LA in the ACE2 network investigated here. This suggests that, although some loci conferring risk for COVID-19 severity, such as the one identified on chromosome 3, may have originated from Neanderthal admixture events, this mechanism did not shape the genetic architecture of the ACE2 network responsible for entry of SARS-CoV-2 into host cellular machinery.

Lastly, among ACE2-network-SNPs, potential COVID-19 risk alleles were observed in *RORA* gene with respect to multiple COVID-19 phenotypes. *RORA* protein product is

involved in immune response, cancer and metabolism [48]. *RORA* plays a role in the activation of T helper cells during lung inflammation by regulating tumor necrosis factor and interleukins [49,50] , and its protein product showed multiple regulatory functions in human epithelial cell cultures inoculated with SARS-CoV-1 [51]. The hypothesis-free approach of genome-wide association of hospitalized COVID-19 vs the population highlighted *SLC6A20* with genome-wide significance on chromosome 3 locus. The SLC12 (*SLC12A6*) class is responsible for inorganic ions such as sodium and chloride while the SLC6 class (*SLC6A19*, identified via network approach and SLC6A20, identified via genome-wide approach) are responsible for transport of amino acids such as glutamate and glycine which are important neurotransmitter activity [52]. *SLC6A19* (among other SLC-class genes) serves similar function to *SLC6A20,* both are expressed in the intestinal tissue and contingent upon ACE2 expression [53]. Multiple studies report more than 10% of the COVID-19 confirmed patients exhibit gastrointestinal symptoms[54–56]

Although we provided a wide range of information highlighting the molecular and phenotypic characteristics of *ACE2* gene network and their putative implications with COVID-19 risk, the findings reported have to be considered exploratory. We used appropriate computational methods and statistical approaches to generate reliable evidence useful to open new directions in COVID-19 research. We also highlighted when the results reported did not survive stringent multiple testing correction. This limitation is particularly relevant with respect to the ACE2 network genetic associations. Due to the limited statistical power of the genome-wide data available to date, none of the risk alleles identified as functionally relevant survive genome-wide testing correction. Further analyses will be needed to validate our current findings.

4 Conclusion

ACE2 is one of the few molecular targets recognized to play a key role in the COVID-19 pathogenesis. We conducted a comprehensive analysis leveraging multiple resources (e.g., drug-gene interactions, tissue-specific transcriptomic profile, and phenome-wide

and genome-wide datasets) to expand our understanding of the genomic characteristics of the host *ACE2* gene network. Overall, our findings highlight the potential mechanisms linking *ACE2* systems biology to COVID-19 susceptibility.

5 Methods

5.1 Gene network collection

Information regarding ACE2 gene network was mined from GeneMANIA [57], Stringdb [58], APID [59], GeneNetwork [60], Biogrid[61] and FunctionalNet [62]. Immediate genes connections that were available in each databank with their default settings result in 61 unique genes (60 genes plus *ACE2*) (Supplementary file1 Figure 1; Supplementary file2 Table S1). The genomic coordinates for the genes were annotated using biomart [63], ensemble GRCh37/hg19. The analysis and visualization were performed in R 3.6.

5.2 2.2 Tissue-specific transcriptomic regulation

The tissue specificity was tested for 60 ACE2-interacting genes in FUMA [64]. The input genes were tested for pre-calculated tissue-specific differentially expressed genes from the GTEx v8 [65]. We also considered the t-statistic sign for up and down-regulated genes against protein coding genes as background. Enrichments were performed using hypergeometric tests and significant enrichments were defined according to Bonferroni corrected p-value ≤ 0.05.

5.3 Phenome-wide analysis of ACE2 gene network

A phenome-wide association study (PheWAS) was performed for 51 of 60 genes that were present in GWASAtlas [11] using all traits available per gene. Statistical significance was determined by applying a Bonferroni multiple-testing correction accounting for the number of GWAS traits (4,765 traits) available in the GWASAtlas ($p<1.05 \times 10^{-5}$). Each trait was grouped into a domain (Supplementary file2 Table S5) which was tested for enrichment using one-sided Fisher's exact test for high proportion of significant traits versus all others tested. A significant domain enrichment was defined considering a Bonferroni-corrected threshold accounting for the number of domains tested (p-value < 0.0019; 0.05/26).

5.4 Gene-Drug Interactions and Biological Functions

Information on drugs that interact with ACE2 network genes were extracted from The Drug-Gene Interaction database (DGIdb) [9] followed by drug-set enrichment for over represented biological functions using DSEA (Drug-Set Enrichment Analysis) [10].

5.5 Characterization of SNPs

Single nucleotide polymorphism (SNPs) were extracted based on the genomic coordinates of the genes (± 10kb) for GrCh37; dbSNP153 from the UCSC browser [66] using bigbed utilities [67] , and referred to as 'ACE2-network SNPs.' ACE2-network SNPs were annotated for global allele frequency, Combined Annotation-Dependent Depletion (CADD) score [12], deep learning based algorithm framework (DeepSEA) [13], and target miRNAs using SNPnexus [68]. DeepSEA is a deep learning-based algorithmic framework for predicting the chromatin effects of sequence alterations with single nucleotide sensitivity[13].The identified miRNAs were tested for over-represented miRNA clusters, functions, and diseases using TAM 2.0 [69].

5.6 Neanderthal introgression

Motivated by evidence of a chromosome 3 COVID-19 risk locus enriched of Neanderthal local ancestry (LA) [47], we compared the distribution of probability of Neanderthal LA in our COVID-19 ACE2-network SNP set and 1,000 randomly sampled SNP sets comprised on SNPs across the genome with comparable genomic features. ACE2-network SNPs were mapped using previously-defined Neanderthal LA data [70,71]. A total of 6,822 LD-independent pairwise SNPs (r^2=0.1 and p=0.1 in 250kb window size) were used as standard input for SNPsnap [72]. In SNPsnap, 1,249/6,822 independent ACE2 network SNPs could be matched based on the following genomic features relative to the input SNP list: minor allele frequency within 2%, gene density within 50%, nearest gene within 50%, and number of linkage disequilibrium groups within 50%. SNPsnap was instructed to exclude the ACE2-network SNP list from the pool of eligible feature-matched SNPs. Non-parametric Wilcoxon rank sum tests were used to compare the Neanderthal LA of our ACE2 network SNP list to that of all 1,000 random SNP sets and multiple testing correction was applied to adjust for a false discovery rate of 5%.

5.7 Association statistics of ACE2 network SNPs from the COVID-19 Host Genetics Initiative (HGI)

The ACE2-network SNPs were extracted from association statistics released by the COVID-19 HGI [15] for six phenotypes describing COVID-19 susceptibility. These phenotypes were A2_V2 (very severe respiratory confirmed COVID-19 cases [N=536] vs. population[N=329391]), B1_V2 (hospitalized COVID-19 cases [N=928] vs. not hospitalized COVID-19 cases [N=2028]), B2_V2 (hospitalized COVID-19 cases [N=3199] vs. population [N=897488]), C1_V2 (COVID-19 cases [N=3523] vs. lab/self-reported negative [N=36634]), C2_V2 (COVID-19 cases [N=6696] vs. population [N=1073072]), and D1_V2 (predicted COVID-19 cases from self-reported symptoms [N=1865] vs. predicted or self-reported non-COVID-19 cases [N=29174]). The SNPs of the ACE2 network were extracted and pruned for LD and p-value using plink1.9. The multiple testing correction was applied using Bonferroni p-value < 0.05. These significant SNPs were annotated further for pathogenicity using Combined Annotation Dependent Depletion (CADD) score and their role as quantitative trait loci (QTL) for gene expression using GTEx, and methylation using QTLbase [73].

6 Author Contribution

G.A.P conceptualized the study design, analyzed, and drafted the manuscript. F.R.W. contributed to analysis, and manuscript writing. Authors, A.G., F.D.A. and R.P. contributed to result interpretation, manuscript drafting and revision. R.P. supervised the study and finalized the manuscript.

7 Competing Interests

The authors have no competing interests.

8 Acknowledgements

We would like to acknowledge support from the National Institutes of Health (R21 DC018098, R21 DA047527, R01 DA12690, F32 MH122058), and thank the COVID-19

Host Genetics Initiative (https://www.covid19hg.org/acknowledgements/) for providing open access to genetic association data.

9 Data Availability

The data presented is available in supplementary files and also on https://gpwhiz.github.io/ACE2Netlas/

10 References

[1] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;579:270–273. doi:10.1038/s41586-020-2012-7.

[2] Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579:265–269. doi:10.1038/s41586-020-2008-3.

[3] Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 2020;181:281–292.e6. doi:10.1016/j.cell.2020.02.058.

[4] Hamming I, Cooper ME, Haagmans BL, Hooper NM, Korstanje R, Osterhaus ADME, et al. The emerging role of ACE2 in physiology and disease. J Pathol 2007;212:1–11. doi:10.1002/path.2162.

[5] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020;323:1061–1069. doi:10.1001/jama.2020.1585.

[6] Shelton JF, Shastri AJ, Ye C, Weldon CH, Filshtein-Somnez T, Coker D, et al. Trans-ethnic analysis reveals genetic and non-genetic associations with COVID-19 susceptibility and severity. medRxiv 2020. doi:10.1101/2020.09.04.20188318.

306 [7] Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl
308 J Med 2020. doi:10.1056/NEJMoa2020283.

[8] Pairo-Castineira E, Clohisey S, Klaric L, Bretherick A, Rawlik K, Parkinson N, et al.
310 Genetic mechanisms of critical illness in Covid-19. medRxiv 2020. doi:10.1101/2020.09.24.20200048.

312 [9] Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, et al. DGIdb: mining the druggable genome. Nat Methods 2013;10:1209–1210.
314 doi:10.1038/nmeth.2689.

[10] Napolitano F, Sirci F, Carrella D, di Bernardo D. Drug-set enrichment analysis: a
316 novel tool to investigate drug mode of action. Bioinformatics 2016;32:235–241. doi:10.1093/bioinformatics/btv536.

318 [11] Watanabe K, Stringer S, Frei O, Umićević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat
320 Genet 2019;51:1339–1348. doi:10.1038/s41588-019-0481-0.

[12] Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the
322 deleteriousness of variants throughout the human genome. Nucleic Acids Res 2019;47:D886–D894. doi:10.1093/nar/gky1016.

324 [13] Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods 2015;12:931–934.
326 doi:10.1038/nmeth.3547.

[14] Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited
328 from Neanderthals. Nature 2020. doi:10.1038/s41586-020-2818-3.

[15] COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global
330 initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur J Hum Genet 2020;28:715–718.
332 doi:10.1038/s41431-020-0636-6.

[16] Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 2020;181:271–280.e8. doi:10.1016/j.cell.2020.02.052.

[17] Huang JK, Carlin DE, Yu MK, Zhang W, Kreisberg JF, Tamayo P, et al. Systematic evaluation of molecular networks for discovery of disease genes. Cell Syst 2018;6:484–495.e5. doi:10.1016/j.cels.2018.03.001.

[18] Gardinassi LG, Souza COS, Sales-Campos H, Fonseca SG. Immune and Metabolic Signatures of COVID-19 Revealed by Transcriptomics Data Reuse. Front Immunol 2020;11:1636. doi:10.3389/fimmu.2020.01636.

[19] Ejaz H, Alsrhani A, Zafar A, Javed H, Junaid K, Abdalla AE, et al. COVID-19 and comorbidities: Deleterious impact on infected patients. J Infect Public Health 2020. doi:10.1016/j.jiph.2020.07.014.

[20] Mönkemüller K, Fry L, Rickes S. COVID-19, coronavirus, SARS-CoV-2 and the small bowel. Rev Esp Enferm Dig 2020;112:383–388. doi:10.17235/reed.2020.7137/2020.

[21] Ajaimy M, Melamed ML. COVID-19 in Patients with Kidney Disease. Clin J Am Soc Nephrol 2020;15:1087–1089. doi:10.2215/CJN.09730620.

[22] Gao M, Wang Q, Wei J, Zhu Z, Li H. Severe Coronavirus disease 2019 pneumonia patients showed signs of aggravated renal impairment. J Clin Lab Anal 2020:e23535. doi:10.1002/jcla.23535.

[23] Pawlotsky J-M. COVID-19 and the liver-related deaths to come. Nat Rev Gastroenterol Hepatol 2020. doi:10.1038/s41575-020-0328-2.

[24] RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. N Engl J Med 2020. doi:10.1056/NEJMoa2021436.

[25] Ghazizadeh Z, Majd H, Richter M, Samuel R, Zekavat SM, Asgharian H, et al. Androgen Regulates SARS-CoV-2 Receptor Levels and Is Associated with Severe COVID-19 Symptoms in Men. BioRxiv 2020. doi:10.1101/2020.05.12.091082.

[26] Cadegiani FA, Goren A, Wambier CG. Spironolactone may provide protection from SARS-CoV-2: Targeting androgens, angiotensin converting enzyme 2 (ACE2), and renin-angiotensin-aldosterone system (RAAS). Med Hypotheses 2020;143:110112. doi:10.1016/j.mehy.2020.110112.

[27] Dumanlı GY, Dilken O, Ürkmez S. Use of Spironolactone in SARS-CoV-2 ARDS Patients. Turk J Anaesthesiol Reanim 2020;48:254–255. doi:10.5152/TJAR.2020.569.

[28] Petersen MW, Meyhoff TS, Helleberg M, Kjær M-BN, Granholm A, Hjortsø CJS, et al. Low-dose hydrocortisone in patients with COVID-19 and severe hypoxia (COVID STEROID) trial-Protocol and statistical analysis plan. Acta Anaesthesiol Scand 2020. doi:10.1111/aas.13673.

[29] Yen F-S, Wei JC-C, Yang Y-C, Hsu C-C, Hwu C-M. Respiratory outcomes of metformin use in patients with type 2 diabetes and chronic obstructive pulmonary disease. Sci Rep 2020;10:10298. doi:10.1038/s41598-020-67338-2.

[30] Bramante C, Ingraham N, Murray T, Marmor S, Hoversten S, Gronski J, et al. Observational Study of Metformin and Risk of Mortality in Patients Hospitalized with Covid-19. medRxiv 2020. doi:10.1101/2020.06.19.20135095.

[31] Shneider A, Kudriavtsev A, Vakhrusheva A. Can melatonin reduce the severity of COVID-19 pandemic? Int Rev Immunol 2020;39:153–162. doi:10.1080/08830185.2020.1756284.

[32] Keller A, Ludwig N, Fehlmann T, Kahraman M, Backes C, Kern F, et al. Low miR-150-5p and miR-320b Expression Predicts Reduced Survival of COPD Patients. Cells 2019;8. doi:10.3390/cells8101162.

[33] Lippi G, Henry BM. Chronic obstructive pulmonary disease is associated with severe coronavirus disease 2019 (COVID-19). Respir Med 2020;167:105941. doi:10.1016/j.rmed.2020.105941.

[34] Zhu J, Wang F-L, Wang H-B, Dong N, Zhu X-M, Wu Y, et al. TNF-α mRNA is negatively regulated by microRNA-181a-5p in maturation of dendritic cells induced by high mobility group box-1 protein. Sci Rep 2017;7:12239. doi:10.1038/s41598-017-12492-3.

[35] Ye Z, Li G, Kim C, Hu B, Jadhav RR, Weyand CM, et al. Regulation of miR-181a expression in T cell aging. Nat Commun 2018;9:3060. doi:10.1038/s41467-018-05552-3.

[36] Osei ET, Florez-Sampedro L, Timens W, Postma DS, Heijink IH, Brandsma C-A. Unravelling the complexity of COPD by microRNAs: it's a small world after all. Eur Respir J 2015;46:807–818. doi:10.1183/13993003.02139-2014.

[37] Mallick B, Ghosh Z, Chakrabarti J. MicroRNome analysis unravels the molecular basis of SARS infection in bronchoalveolar stem cells. PLoS One 2009;4:e7837. doi:10.1371/journal.pone.0007837.

[38] Khan MA-A-K, Sany MRU, Islam MS, Mehebub MS, Islam ABMMK. Epigenetic regulator miRNA pattern differences among SARS-CoV, SARS-CoV-2 and SARS-CoV-2 world-wide isolates delineated the mystery behind the epic pathogenicity and distinct clinical characteristics of pandemic COVID-19. BioRxiv 2020. doi:10.1101/2020.05.06.081026.

[39] Apicella M, Campopiano MC, Mantuano M, Mazoni L, Coppelli A, Del Prato S. COVID-19 in people with diabetes: understanding the reasons for worse outcomes. Lancet Diabetes Endocrinol 2020;8:782–792. doi:10.1016/S2213-8587(20)30238-2.

[40] Li J, Wang X, Chen J, Zuo X, Zhang H, Deng A. COVID-19 infection may cause ketosis and ketoacidosis. Diabetes Obes Metab 2020. doi:10.1111/dom.14057.

[41] Ren H, Yang Y, Wang F, Yan Y, Shi X, Dong K, et al. Association of the insulin resistance marker TyG index with the severity and mortality of COVID-19. Cardiovasc Diabetol 2020;19:58. doi:10.1186/s12933-020-01035-2.

[42] Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020;323:2052–2059. doi:10.1001/jama.2020.6775.

[43] Yancy CW, Fonarow GC. Coronavirus Disease 2019 (COVID-19) and the Heart-Is Heart Failure the Next Chapter? JAMA Cardiol 2020. doi:10.1001/jamacardio.2020.3575.

[44] Hanley B, Naresh KN, Roufosse C, Nicholson AG, Weir J, Cooke GS, et al. Histopathological findings and viral tropism in UK patients with severe fatal COVID-19: a post-mortem study. Lancet Microbe 2020. doi:10.1016/S2666-5247(20)30115-4.

[45] Chang TS, Ding Y, Freund MK, Johnson R, Schwarz T, Yabu JM, et al. Prior diagnoses and medications as risk factors for COVID-19 in a Los Angeles Health System. medRxiv 2020. doi:10.1101/2020.07.03.20145581.

[46] Kuo C-L, Pilling LC, Atkins JL, Masoli JAH, Delgado J, Kuchel GA, et al. APOE e4 genotype predicts severe COVID-19 in the UK Biobank community cohort. J Gerontol A, Biol Sci Med Sci 2020. doi:10.1093/gerona/glaa131.

[47] Zeberg H, Paabo S. The major genetic risk factor for severe COVID-19 is inherited from Neandertals. BioRxiv 2020. doi:10.1101/2020.07.03.186296.

[48] Cook DN, Kang HS, Jetten AM. Retinoic Acid-Related Orphan Receptors (RORs): Regulatory Functions in Immunity, Development, Circadian Rhythm, and Metabolism. Nucl Receptor Res 2015;2. doi:10.11131/2015/101185.

[49] Haim-Vilmovsky L, Walker JA, Henriksson J, Miao Z, Natan E, Kar G, et al. *Rora* regulates activated T helper cells during inflammation. BioRxiv 2019. doi:10.1101/709998.

[50] Nejati Moharrami N, Bjørkøy Tande E, Ryan L, Espevik T, Boyartchuk V. RORα controls inflammatory state of human macrophages. PLoS One 2018;13:e0207374. doi:10.1371/journal.pone.0207374.

[51] de Almeida RMC, Thomas GL, Glazier JA. Transcriptogram analysis reveals relationship between viral titer and gene sets responses during Corona-virus infection. BioRxiv 2020. doi:10.1101/2020.06.16.155267.

[52] Lin L, Yee SW, Kim RB, Giacomini KM. SLC transporters as therapeutic targets: emerging opportunities. Nat Rev Drug Discov 2015;14:543–560. doi:10.1038/nrd4626.

[53] Vuille-Dit-Bille RN, Liechty KW, Verrey F, Guglielmetti LC. SARS-CoV-2 receptor ACE2 gene expression in small intestine correlates with age. Amino Acids 2020. doi:10.1007/s00726-020-02870-z.

[54] Lian J, Jin X, Hao S, Jia H, Cai H, Zhang X, et al. Epidemiological, clinical, and virological characteristics of 465 hospitalized cases of coronavirus disease 2019 (COVID-19) from Zhejiang province in China. Influenza Other Respi Viruses 2020. doi:10.1111/irv.12758.

[55] Jin X, Lian J-S, Hu J-H, Gao J, Zheng L, Zhang Y-M, et al. Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. Gut 2020;69:1002–1009. doi:10.1136/gutjnl-2020-320926.

[56] Khan M, Khan H, Khan S, Nawaz M. Epidemiological and clinical characteristics of coronavirus disease (COVID-19) cases at a screening clinic during the early outbreak period: a single-centre study. J Med Microbiol 2020. doi:10.1099/jmm.0.001231.

462 [57] Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, et al. GeneMANIA update 2018. Nucleic Acids Res 2018;46:W60–W64. doi:10.1093/nar/gky311.

464 [58] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made 466 broadly accessible. Nucleic Acids Res 2017;45:D362–D368. doi:10.1093/nar/gkw937.

468 [59] Prieto C, De Las Rivas J. APID: agile protein interaction dataanalyzer. Nucleic Acids Res 2006;34:W298–302. doi:10.1093/nar/gkl128.

470 [60] Deelen P, van Dam S, Herkert JC, Karjalainen JM, Brugge H, Abbott KM, et al. Improving the diagnostic yield of exome- sequencing by predicting gene-phenotype 472 associations using large-scale gene expression analysis. Nat Commun 2019;10:2837. doi:10.1038/s41467-019-10649-4.

474 [61] Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. Nucleic Acids Res 2019;47:D529–D541. 476 doi:10.1093/nar/gky1079.

[62] Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease 478 genes by network-based boosting of genome-wide association data. Genome Res 2011;21:1109–1121. doi:10.1101/gr.118992.110.

480 [63] Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 482 2009;4:1184–1191. doi:10.1038/nprot.2009.97.

[64] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and 484 annotation of genetic associations with FUMA. Nat Commun 2017;8:1826. doi:10.1038/s41467-017-01261-5.

486 [65] Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. BioRxiv 2019. 488 doi:10.1101/787903.

[66] Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2019 update. Nucleic Acids Res 2019;47:D853–D858. doi:10.1093/nar/gky1095.

[67] Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res 2004;32:D493–6. doi:10.1093/nar/gkh103.

[68] Dayem Ullah AZ, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C. SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. Nucleic Acids Res 2018;46:W109–W113. doi:10.1093/nar/gky399.

[69] Li J, Han X, Wan Y, Zhang S, Zhao Y, Fan R, et al. TAM 2.0: tool for MicroRNA set analysis. Nucleic Acids Res 2018;46:W180–W185. doi:10.1093/nar/gky509.

[70] Durvasula A, Sankararaman S. A statistical model for reference-free inference of archaic local ancestry. PLoS Genet 2019;15:e1008175. doi:10.1371/journal.pgen.1008175.

[71] Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 2014;507:354–357. doi:10.1038/nature12961.

[72] Pers TH, Timshel P, Hirschhorn JN. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. Bioinformatics 2015;31:418–420. doi:10.1093/bioinformatics/btu655.

[73] Zheng Z, Huang D, Wang J, Zhao K, Zhou Y, Guo Z, et al. QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. Nucleic Acids Res 2020;48:D983–D991. doi:10.1093/nar/gkz888.

Figures and Supplementary file Legends

Figures

Figure 1: The ACE2-gene network. The genes that connect with ACE2 were extracted from six different gene-network databases and compiled together in one network.

Figure 2: Tissues enriched based on ACE2-network gene expression for GTEx tissues. The genes from the ACE2-network show over-representation of tissues (x-axis) and -log10 p-value (y-axis). The red bars are significant enrichments.

Figure 3: Drug-set enrichment analysis. LEFT: The similarity of drugs based on pathways identified. RIGHT: Biological Processes identified based on drugs that interact with genes from the ACE2-network

Figure 4: Domain distribution of PheWAS of ACE2-network genes. The ACE2 gene network associations are grouped based on domains (y-axis) and gene names (x-axis). The size of the data points reflects number of phenotypes surviving multiple testing correction.

Figure 5: Enrichment of biological functions based on miRNA:SNP annotation. Using miRNAs annotation, over-represented biological processes are shown on y-axis and -log10 pvalue on x-axis.

Supplementary files

Supplementary file1: File containing figures S1-S60

Supplementary file2: Tables S1:S14. Tabular details of the list of genes, tissue enrichment, gene-drug interaction, drug-set enrichment, PheWAS of all genes, Significant traits of PheWAS, domain enrichment, miRNA enrichment, SNPs from the network for six COVID-19 phenotypes from COVID-19 Host Genetics Initiative – Freeze3

Supplementary file3: Text file of all the SNPs within ±10kb of the genes and their genomic coordinates and allele frequency.

- Column Headers:
  - Variation ID: <dbsnp rs#>
  - dbSNP: link to dbSNP, if known
  - Chromosome: Variant mapped chromosome location
  - Position: Variant start position on chromosome
  - REF Allele: Reference allele
  - ALT Allele (IUPAC): Observed allele
  - Minor Allele: Minor allele observed in global population, if known

- Minor Allele Frequency: Minor allele frequency observed in global population, if known
- Contig: Variant mapped contig location
- contigPosition: Variant start position on contig
- Band: SNP cytogenetic location

Supplementary file4: Text file of SNPs with their overlapping and nearest gene annotation using Ensembl GRCh37.

- Column Headers:
  - Variation ID: <dbsnp rs#>
  - Chromosome: Variant mapped chromosome location
  - Position: Variant start position on chromosome
  - Overlapped Gene: Name of the gene (HGNC system) to which the variant is overlapped
  - Type: Gene type, e.g., protein coding, miRNA, non coding, Pseudogene, snoRNA, lincRNA etc.
  - Annotation: Summary of whether the variant overlapped with the coding, intronic or untranslated regions of the various transcript isoforms of the gene, as annotated from Ensembl gene system.
  - Nearest Upstream Gene: If variant is not overlapped with any gene, then the gene whose end position is nearest to the variant on the left (considering the alignment of genes on the positive strand as left-to-right)
  - Type of Nearest Upstream Gene: Gene type, e.g., protein coding, miRNA, non coding, Pseudogene, snoRNA, lincRNA etc.
  - Distance to Nearest Upstream Gene: distance from the end position of the nearest upstream gene.
  - Nearest Downstream Gene: If variant is not overlapped with any gene, then the gene whose start position is nearest to the variant on the right (considering the alignment of genes on the positive strand as left-to-right)
  - Type of Nearest Downstream Gene: Gene type, e.g., protein coding, miRNA, non coding, Pseudogene, snoRNA, lincRNA etc.
  - Distance to Nearest Downstream Gene: distance from the start position of the nearest downstream gene.

Supplementary file5: Text file of all the SNPs with CADD (Combined Annotation Dependent Depletion) scores >10

- Column Headers:

  - Variation ID: <dbsnp rs#>
  - Chromosome: Chromosome name
  - Position: Variant start position on chromosome
  - Variant: <reference allele,"/",observed allele> as reported in the tool's genome-wide score
  - PHRED: PHRED-like (-10*log10(rank/total)) scaled CADD-score ranking a variant relative to all possible substitutions of the human genome. A score≥10 indicates that it is predicted to be in the 10% most deleterious substitutions that you can do to the human genome, a score≥20 indicates the 1% most deleterious and so on.

Supplementary file6: Text file of all the SNPs with DeepSEA (deep learning based algorithm framework) functional scores > 0.5, which represents atleast 50% probability to have regulatory effect

- Column Headers:
  - Variation ID: <dbsnp rs#>
  - Chromosome: Chromosome name
  - Position: Variant start position in the chromosome
  - Variant: <reference allele,"/",observed allele> as reported in the tool's genome-wide score
  - eQTL Probability: The probability of the variant being a eQTL variant given by functional variant prioritization classifier.
  - GWAS Probability: The probability of the variant being a trait-associated (GWAS) variant given by functional variant prioritization classifier.
  - HGMD Probability: The probability of the variant being a inherited disease-associated (HGMD) variant given by functional variant prioritization classifier.
  - Functional Significance Score: A measure in the range [0-1] depicting the significance of magnitude of predicted chromatin effect and evolutionary conservation. Lower score indicates higher likelihood of functional significance of the variant.

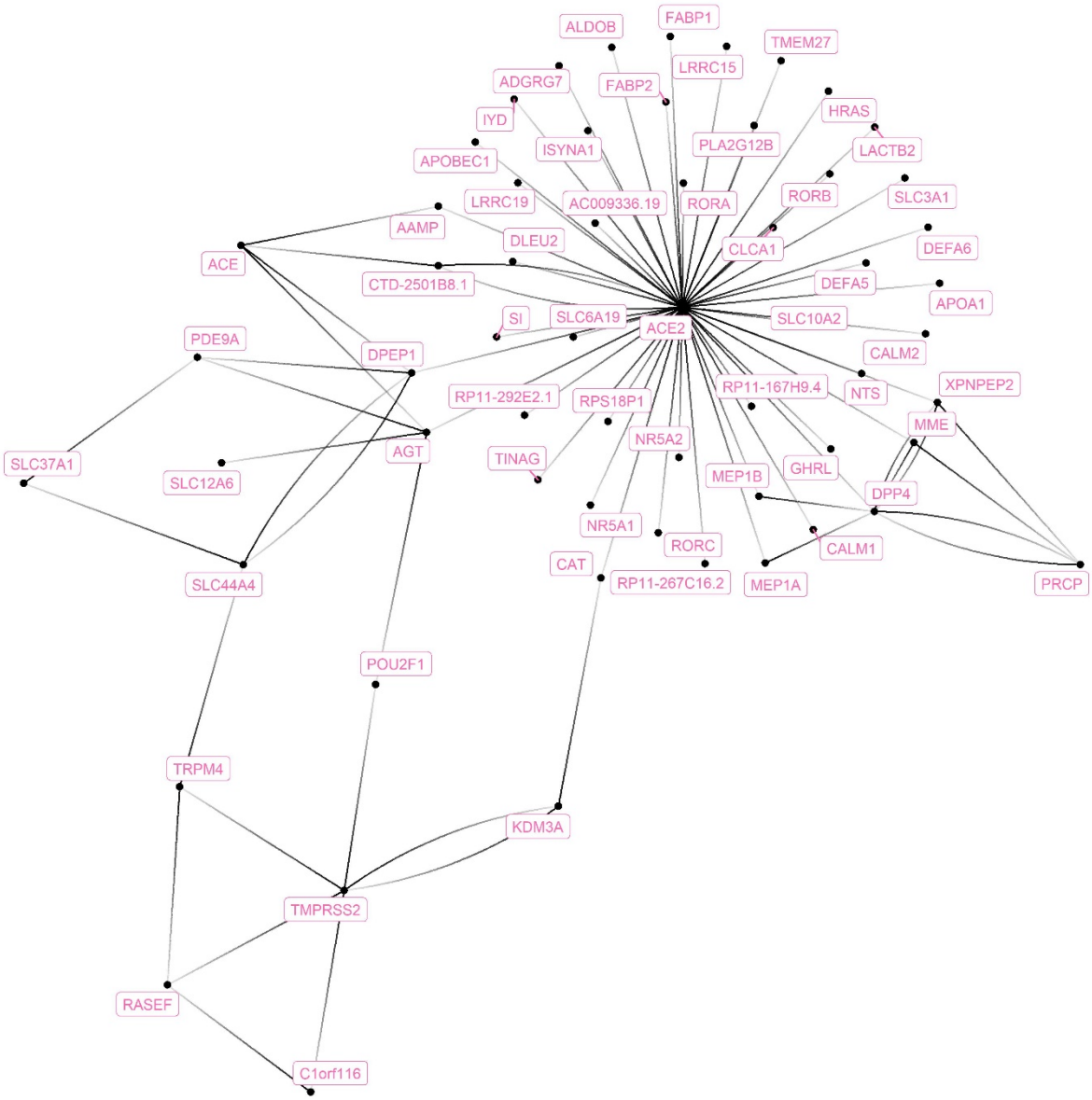

Figure 1: The ACE2-gene network. The genes that connect with ACE2 were extracted from six different gene-network databases and compiled together in one network.

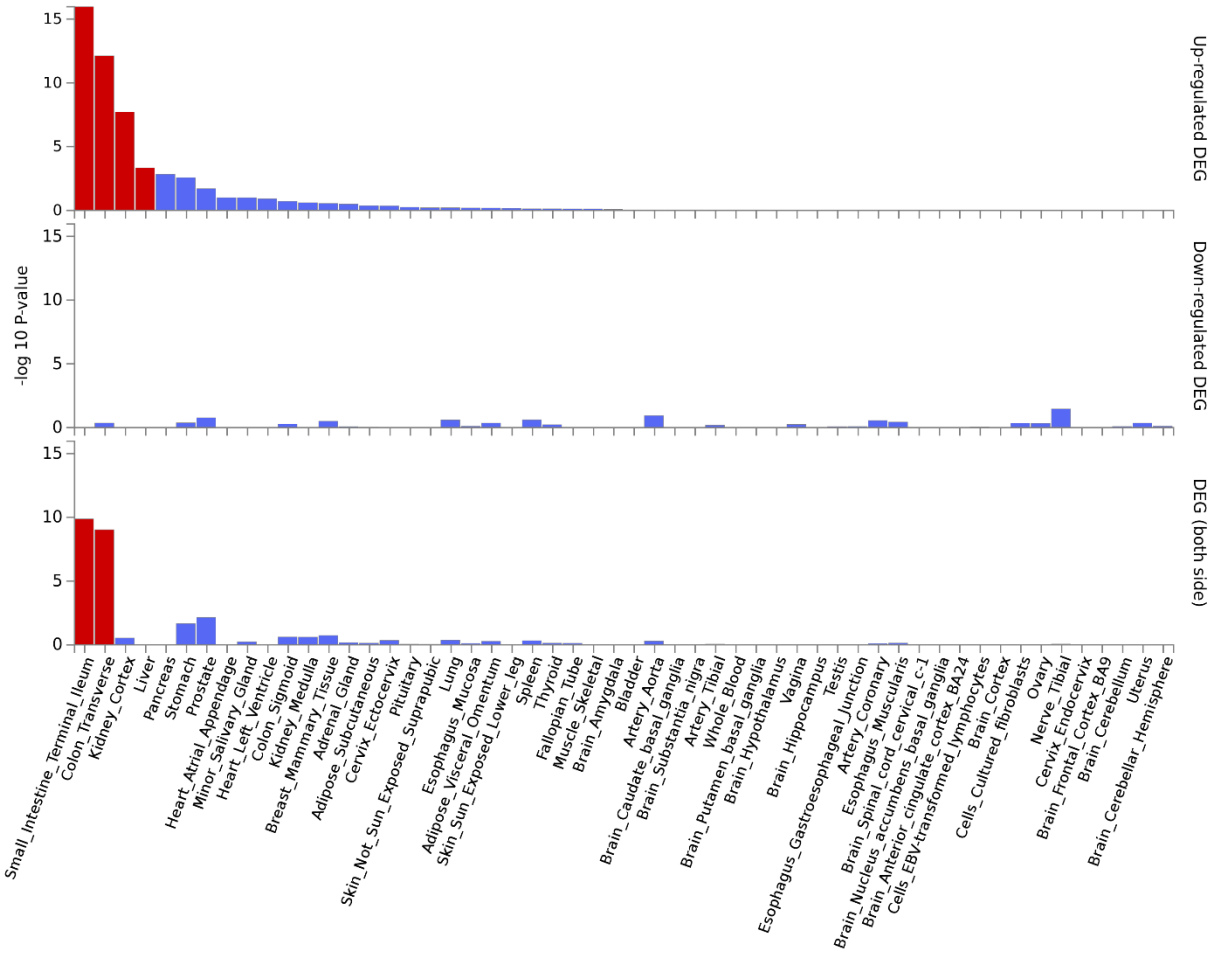

Figure 2: Tissues enriched based on ACE2-network gene expression for GTEx tissues. The genes from the ACE2-network show over-representation of tissues (x-axis) and -log10 p-value (y-axis). The red bars are significant enrichments.

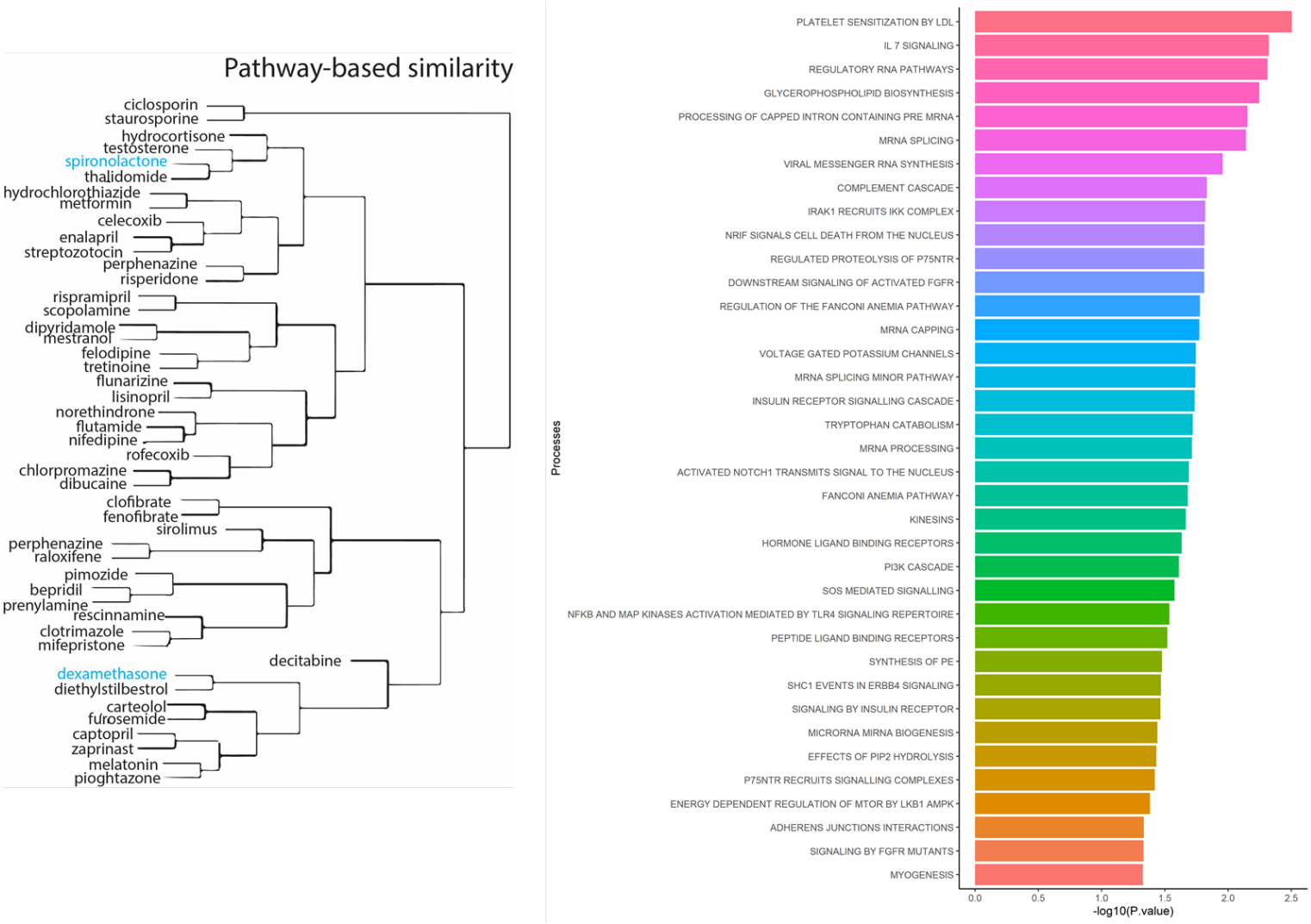

Figure 3: Drug-set enrichment analysis. LEFT: The similarity of drugs based on pathways identified. RIGHT: Biological Processes identified based on drugs that interact with genes from the ACE2-network.

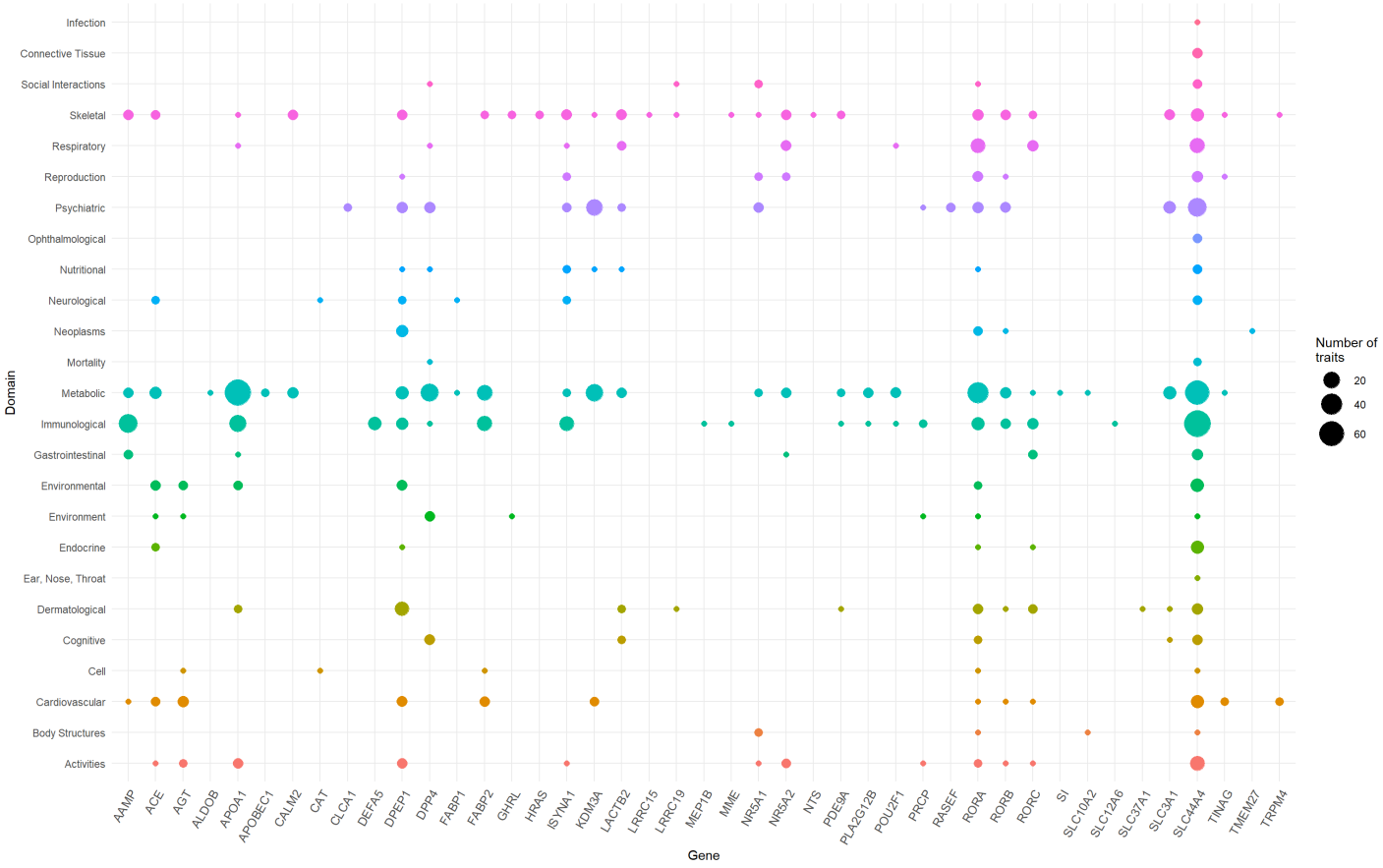

Figure 4: Domain distribution of PheWAS of ACE2-network genes. The ACE2 gene network associations are grouped based on domains (y-axis) and gene names (x-axis). The size of the data points reflects number of phenotypes surviving multiple testing correction.

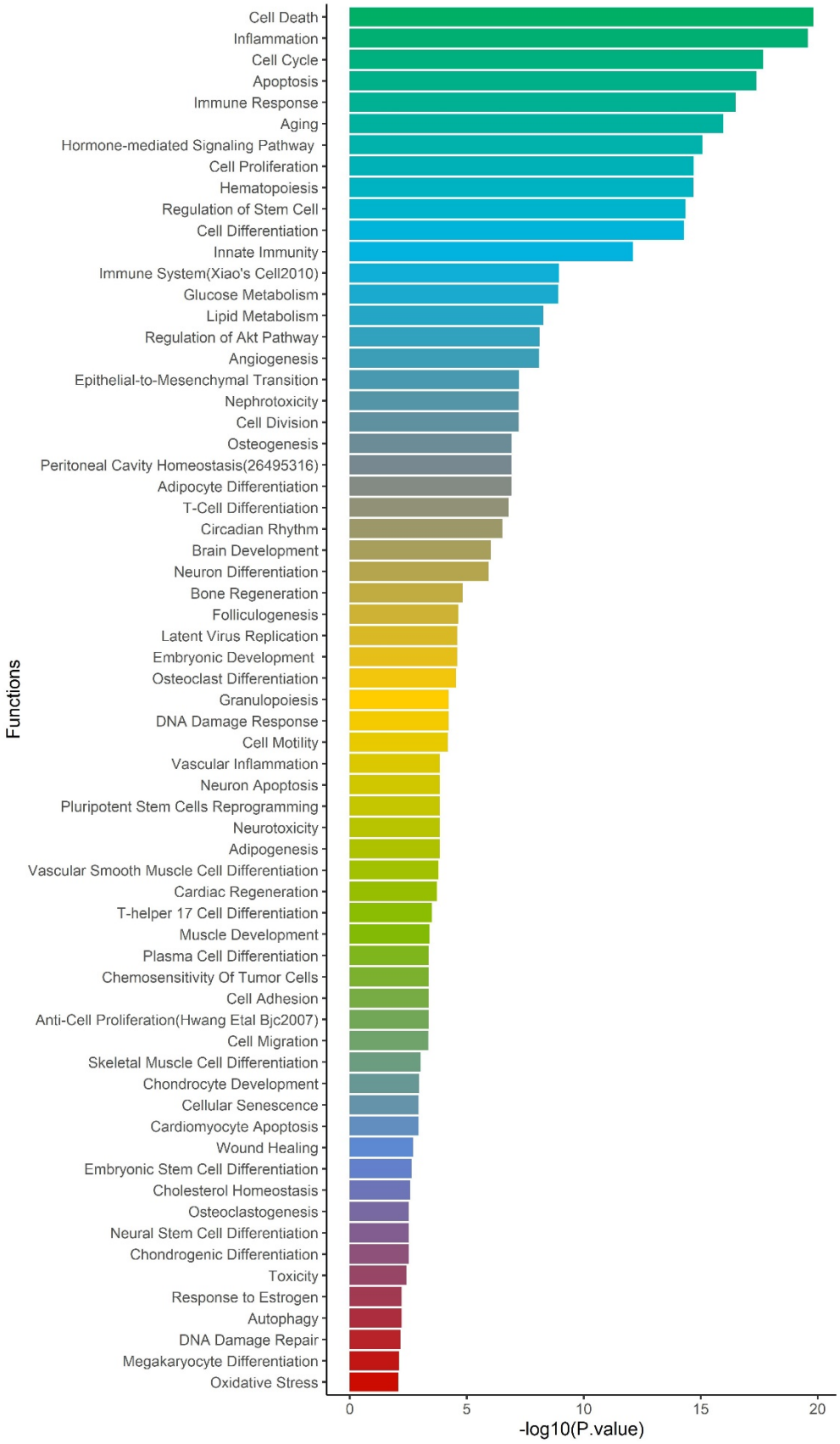

Figure 5: Enrichment of biological functions based on miRNA:SNP annotation. Using miRNAs annotation, over-represented biological processes are shown on y-axis and -log10 pvalue on x-axis.