10.1161/CIRCULATIONAHA.121.057888

Genetic Landscape of the ACE2 Coronavirus Receptor

Running title: *Yang et al.; Genetic landscape of the ACE2 coronavirus receptor*

Zhijian Yang, BSc[1,3*]; Erin Macdonald-Dunlop, PhD[4*]; Jiantao Chen, BSc[1,3]; Ranran Zhai, BSc[1,3]; Ting Li, BSc[1,3]; Anne Richmond, MSc[5]; Lucija Klaric, PhD[5]; Nicola Pirastu, PhD[4,6]; Zheng Ning, PhD[1,7]; Chenqing Zheng, MSc[1]; Yipeng Wang, BSc[1]; Tingting Huang, MD, PhD[7]; Yazhou He, MD, PhD[4,8]; Huiming Guo, MD[9]; Kejun Ying, BSc[10,11]; Stefan Gustafsson, PhD[39]; Bram Prins, PhD[13,14]; Anna Ramisch, PhD[15]; Emmanouil T. Dermitzakis, PhD[15]; Grace Png, MPhil[16,17]; Niclas Eriksson, PhD[18]; Jeffrey Haessler, MSc[19]; Xiaowei Hu, PhD[20]; Daniela Zanetti, PhD[21,22]; Thibaud Boutin, PhD[5]; Shih-Jen Hwang, PhD[22,35]; Eleanor Wheeler, PhD[23]; Maik Pietzner, PhD[23,46]; Laura M. Raffield, PhD[24]; Anette Kalnapenkis, MSc[25,26]; James E. Peters, MBChB, MPhil, PhD[12,13,14]; Ana Viñuela, PhD[15,27]; Arthur Gilly, PhD[16,29]; Sölve Elmståhl, MD, PhD[30]; George Dedoussis, PhD[31]; John R. Petrie, MD, PhD[32]; Ozren Polašek, PhD[33,37]; Lasse Folkersen, PhD[34]; Yan Chen, PhD[7]; Chen Yao, PhD[22,35]; Urmo Võsa, PhD[25]; Erola Pairo-Castineira, PhD[5,36]; Sara Clohisey, PhD[36]; Andrew D. Bretherick, PhD[5]; Konrad Rawlik, PhD[36]; GenOMICC Consortium[‡]; IMI-DIRECT Consortium[‡]; Tõnu Esko, PhD[25]; Stefan Enroth, PhD[38]; Åsa Johansson, PhD[38]; Ulf Gyllensten, PhD[38]; Claudia Langenberg, PhD[23,46]; Daniel Levy, MD[22,35]; Caroline Hayward, PhD[5]; Themistocles L. Assimes, PhD[21,22]; Charles Kooperberg, PhD[19]; Ani W. Manichaikul, PhD[20]; Agneta Siegbahn, PhD[39]; Lars Wallentin, PhD[39]; Lars Lind, PhD[39]; Eleftheria Zeggini, PhD[16,29,40]; Jochen M. Schwenk, PhD[41]; Adam S. Butterworth, PhD[13,14,42,43]; Karl Michaëlsson, MD, PhD[44]; Yudi Pawitan, PhD[7]; Peter K. Joshi, PhD[4]; J. Kenneth Baillie, MD, PhD[36,45]; Anders Mälarstig, PhD[7,28]; Alexander P. Reiner, MD[19]; James F. Wilson, DPhil[4,5†]; Xia Shen, PhD[1,2,3,4,7†]

[*]Z. Yang and E. Macdonald-Dunlop contributed equally.

[†]J. F. Wilson and X. Shen contributed equally.

[‡]Two lists of the GenOMICC Consortium members and IMI-DIRECT Consortium members are provided in the Supplemental Material.

[1]Biostatistics Group, School of Life Sciences, Sun Yat-sen University, Guangzhou, China; [2]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China; [3]Center for Intelligent Medicine Research, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China; [4]Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, UK; [5]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, UK; [6]Human Technopole Viale Rita Levi-Montalcini, 1 Area MIND, Milano, Italy; [7]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [8]West China School of Public Health, West China Fourth Hospital, Sichuan University, Chengdu, China; [9]Department of Cardiac Surgery, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital Guangdong Academy of Medical Sciences, Guangzhou, China; [10]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; [11]T. H. Chan School of Public Health, Harvard University, Boston, MA; [12]Department of Immunology and Inflammation, Imperial College London, London, UK; [13]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; [14]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK; [15]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; [16]Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany; [17]Technical University of Munich (TUM), School of Medicine, Munich, Germany; [18]Uppsala Clinical Research Center, Sweden; [19]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA; [20]Center for Public Health Genomics, University of Virginia, Charlottesville, VA; [21]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; [22]Stanford Cardiovascular Institute, Stanford University, Stanford, CA; [23]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK; [24]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; [25]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia; [26]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; [27]Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle, UK; [28]Pfizer Worldwide Research, Development and Medical, Stockholm, Sweden; [29]Wellcome Sanger

Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; [30]Faculty of Medicine, Lund University, Lund, Sweden; [31]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University of Athens, Greece; [32]Institute of Cardiovascular & Medical Sciences, University of Glasgow, United Kingdom; [33]University of Split School of Medicine, Split, Croatia; [34]Danish National Genome Center, Copenhagen, Denmark; [35]Population Sciences Branch, Division of Intramural Research, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD; [36]Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, UK; [37]Algebra University College, Ilica, Zagreb, Croatia; [38]Department of Immunology, Genetics and Pathology, Uppsala Universitet, Science for Life Laboratory, Uppsala, Sweden; [39]Department of Medical Sciences, Uppsala University, Uppsala, Sweden; [40]Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine, Munich, Germany; [41]Affinity Proteomics, Science for Life Laboratory, KTH Royal Institute of Technology, Solna, Sweden; [42]British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK; [43]National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK; [44]Department of Surgical Sciences, Uppsala University, Uppsala, Sweden; [45]Intenstive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, UK; [46]Computational Medicine, Berlin Institute of Health at Charité, Universitätsmedizin Berlin, Germany

Address for Correspondence:
Xia Shen, PhD
Center for Intelligent Medicine Research,
Greater Bay Area Institute of Precision Medicine (Guangzhou),
Fudan University,
2nd Nanjiang Road, Nansha District,
511458 Guangzhou, P. R. China
Email: xia.shen@ed.ac.uk

James F. Wilson, DPhil
Centre for Global Health Research, Usher Institute,
University of Edinburgh, Old Medical School, Teviot Place,
EH8 9AG Edinburgh, UK
Email: jim.wilson@ed.ac.uk

Abstract

Background: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the etiologic agent of COVID-19, enters human cells using the angiotensin-converting enzyme 2 (ACE2) protein as a receptor. ACE2 is thus key to the infection and treatment of the coronavirus. ACE2 is highly expressed in the heart, respiratory and gastrointestinal tracts, playing important regulatory roles in the cardiovascular and other biologic systems. However, the genetic basis of the ACE2 protein levels is not well understood.

Methods: We conduct so far the largest genome-wide association meta-analysis of plasma ACE2 levels in over 28,000 individuals of the SCALLOP Consortium. We summarize the cross-sectional epidemiologic correlates of circulating ACE2. Using the summary-statistics-based high-definition likelihood method, we estimate relevant genetic correlations with cardiometabolic phenotypes, COVID-19, and other human complex traits and diseases. We perform causal inference of soluble ACE2 on vascular disease outcomes and COVID-19 disease severity using Mendelian randomization. We also perform in silico functional analysis by integrating with other types of omics data.

Results: We identified ten loci, including eight novel, capturing 30% of the protein's heritability. We detected that plasma ACE2 was genetically correlated with vascular diseases, severe COVID-19, and a wide range of human complex diseases and medications. An X-chromosome cis-pQTL-based Mendelian randomization analysis suggested a causal effect of elevated ACE2 levels on COVID-19 severity (odds ratio (OR), 1.63; 95% CI, 1.10 to 2.42; $P = 0.01$), hospitalization (OR, 1.52; 95% CI, 1.05 to 2.21; $P = 0.03$), and infection (OR, 1.60; 95% CI, 1.08 to 2.37; $P = 0.02$). Tissue- and cell-type-specific transcriptomic and epigenomic analysis revealed that the ACE2 regulatory variants were enriched for DNA methylation sites in blood immune cells.

Conclusions: Human plasma ACE2 shares a genetic basis with cardiovascular disease, COVID-19, and other related diseases. The genetic architecture of the ACE2 protein is mapped, providing a useful resource for further biological and clinical studies on this coronavirus receptor.

Key Words: ACE2, COVID-19, SARS-CoV-2, Cardiovascular disease, Complex disease

Non-standard Abbreviations and Acronyms

ARB	Angiotensin receptor blockers
BMI	Body mass index
COVID-19	Coronavirus disease 2019
CRP	C-reactive protein
CTSL1	Cathepsin L1
CVD	Cardiovascular disease
eQTL	Expression quantitative trait loci
GI	Gastrointestinal
GWAS	Genome-wide association study
GWAMA	Genome-wide association meta-analysis
HWE	Hardy-Weinberg equilibrium
LD	Linkage disequilibrium
MAF	Minor allele frequency
MR	Mendelian randomization
PCR	Polymerase chain reaction
PEA	Proximity Extension Assay
pQTL	Protein quantitative trait loci
RAS	Renin-angiotensin system
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SBP	Systolic blood pressure
SNP	Single nucleotide polymorphism
TF	Transcription factor
TG	Triglycerides

Clinical Perspective

What is new?

- The overall heritability of ACE2 level is 16%, of which 30% can be explained by ten protein quantitative trait loci (pQTL) identified in this study.
- ACE2 level is genetically correlated with both COVID-19 and cardiovascular disease.
- Elevated ACE2 levels show a causal relationship with COVID-19 severity, hospitalization, and infection, as shown by a cis-pQTL-based Mendelian randomization analysis.
- ACE2 regulatory variants are enriched on DNA methylation sites in blood immune cells.

What are the clinical implications?

- The causal evidence for ACE2 suggests that pharmacological inhibition of circulating ACE2 may be a promising approach for treating COVID-19 or its comorbidities.
- Transcription factors such as HNF1A and HNF4A play essential roles in ACE2 regulation and could provide alternative paths to pharmacological modulation of ACE2 plasma levels.
- The genetic correlations between ACE2 levels and both COVID-19 and cardiovascular disease risk imply that the cardiovascular complications seen in COVID-19 patients may be intrinsic to the disease and mechanistically driven by ACE2.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the etiologic agent of COVID-19, enters human cells using the ACE2 protein as a receptor[1–5]. ACE2 is highly expressed in the heart, respiratory, and gastrointestinal (GI) tracts, and it plays important regulatory roles in the cardiovascular and other biologic systems[6]. ACE2 proteolytically degrades angiotensin II (a potent vasoconstrictive, pro-inflammatory, and pro-thrombotic mediator) into angiotensin (1-7) thereby regulating blood pressure, salt and water balance, glucose homeostasis, and amino acid absorption in the kidney and small intestine.

Shedding of a soluble form of ACE2 from the cell surface is regulated by membrane-bound enzymes such as TMPRSS2 and ADAM17[7,8]. Enzymatic cleavage of the ACE2 extracellular domain by TMPRSS2 following binding of the spike protein of SARS-CoV-2 to ACE2 also plays a role in SARS-CoV-2 cell entry and infection. To date, several large, observational cohort studies including either patients with heart failure or atrial fibrillation or healthy children and adults have demonstrated that circulating ACE2 antigen levels[9–12] are higher in men than women, increase with age, and correlate with cardiovascular outcomes and cardiometabolic and inflammatory biomarkers, but not with the use of ACE inhibitors or angiotensin receptor blockers (ARB)[13,14]. Therefore, further understanding the epidemiologic relationships and genetic regulation of the soluble ACE2 protein could have important implications for risk of SARS-CoV-2 infection or disease severity, cardiovascular disease risk in general, and also motivate further assessment of the causal role of ACE2 in these diseases.

A recent genome-wide analysis of soluble ACE2 measured in plasma of 3,442 heart failure patients identified three genome-wide significant protein quantitative trait loci (pQTL): a cis-pQTL on chromosome X (located near the cognate ACE2 gene) and two trans-loci on

chromosomes 12 and 21 encompassing the genes encoding transcription factors HNF1A, and ERG, respectively[15]. Current antibody-based ACE2 pQTL studies have been insufficient to capture enough heritability of the protein or estimate downstream genetic connections between ACE2 and diseases such as Cardiovascular disease (CVD) and COVID-19.

Here, by performing so far the largest genome-wide association meta-analysis (GWAMA) of plasma ACE2 (N = 28,204), we summarize the cross-sectional epidemiologic correlates of circulating ACE2, report ten pQTL including eight novel for ACE2, estimate relevant genetic correlations with other cardiometabolic phenotypes, conduct causal inference of soluble ACE2 on vascular disease outcomes and COVID-19 disease severity using Mendelian randomization (MR).

Methods

Transparency and Openness Promotion The summary statistics data of the plasma ACE2 GWAS meta-analysis are publicly available at: https://doi.org/10.6084/m9.figshare.19189307. Other summary-level data used in this study are also available as stated in Data availability in Supplemental Material. The individual-level human phenotypic and genetic data and biological samples used in this study will not be made available to other researchers for the purposes of reproducing the results or replicating the procedure.

Data collection Across the participating cohorts, plasma ACE2 abundance was measured using the Proximity Extension Assay (PEA) technology of Olink©Proteomics on the Multiplex CVDII targeted 96-protein panel. PEA has high readout specificity and sensitivity (see also Supplemental Material) and consumes a minimal amount of sample. In PEA, matched pairs of oligonucleotide-labeled antibodies bind to the target ACE2 antigen, and upon antibody binding,

the matched oligonucleotides are brought into proximity[16]. A polymerase chain reaction (PCR) target sequence corresponding to the ACE2 protein is then created, amplified, and quantified by quantitative PCR.

Genome-wide association summary statistics of plasma ACE2 protein were obtained from 14 cohorts of European ancestry. The cohorts' details are given in Table S1 and Supplemental Methods. The institutional review committees for each cohort approved the study methodology, and all study participants provided informed consent for their clinical and genetic data to be used for research. The maximum sample size per SNP is 28,204 individuals. Each cohort provided data imputed to 1000 Genomes Project phase 3 reference or to the Haplotype Reference Consortium (HRC) reference, which resulted in 17,166,011 autosomal SNPs and 3,922,856 X chromosome SNPs. We tested 8,682,405 and 730,046 genetic variants on the autosomes and X chromosome, respectively, with a minor allele frequency (MAF) > 0.01. Each cohort applied quality control measures for call rate filters, sex mismatch, population outliers, heterozygosity, and cryptic relatedness as documented in the Supplemental Methods. Before running the genetic analyses, protein values (on the log2 scale) were inverse-Gaussian transformed to zero mean and unit variance. No detection threshold was applied to the raw Olink measurement values.

The summary association statistics for severe COVID-19 and other complex traits and diseases were obtained from publicly available established resources (see Data Availability in Supplemental Material). The ACE protein serum concentration GWAS in 4,147 individuals from the Outcome Reduction with Initial Glargine INtervention (ORIGIN) trial[17] were obtained from the authors (see also Acknowledgements). The aptamer-based plasma ACE levels GWAS made use of the recent study in 35,559 Icelanders[18]. For severe COVID-19, we considered the

GenOMICC Consortium GWAS[19]. For COVID-19 hospitalization and infection GWAS we used COVID-19 Host Genetics Initiative (HGI) GWAS meta-analysis round 4 (October 20, 2020)[20]. To provide a proper replication of causal effect inference, we subtracted GenOMICC GWAS from the HGI meta-analysis using the R package MetaSubtract.

Genome-wide association meta-analysis Genetic analyses were conducted using additive model regressions, with adjustment for population structure and study-specific parameters (details in Supplemental Methods). For the X chromosome, the analysis was performed separately for males and females, with 0-1 genotype coding for males. Each contributing cohort uploaded the result summary statistics in a standardized format to a secure computational cluster at the University of Edinburgh. The meta-analysis was performed using METAL (2011-03-25)[21] with the inverse-variance weighted approach (STDERR option).

Heritability and genetic correlation analysis For the autosomes, we estimated the genome-wide SNP-based heritability and genetic correlation values for plasma ACE2 and other complex traits using high-definition likelihood (HDL)[22], an approach based on summary association statistics. We used the default reference panel containing linkage disequilibrium (LD) information of 1,029,876 quality-controlled imputed SNPs provided in the HDL software. The genetic variance captured by the cis-pQTL of ACE2 on chromosome X was estimated based on the top SNP assuming Hardy-Weinberg equilibrium (HWE), i.e., $2f(1-f)\hat{\beta}^2$, where f and $\hat{\beta}$ are the MAF and the genetic effect estimate, respectively. The genetic variance captured by each autosomal trans-pQTL was also estimated based on the top SNP assuming HWE. Since the phenotypic variance of ACE2 was standardized to one via inverse-Gaussian transformation, the sum of $2f(1-f)\hat{\beta}^2$ across all the pQTL yields the proportion of phenotypic variance explained by the discovered pQTL. Thus, the ratio of the genetic variance captured by the pQTL to the

estimated heritability gives the proportion of heritability explained.

Mendelian randomization For cis-pQTL-based Mendelian randomization (MR) analysis, we used SNP genotypes within the X chromosomal ACE2 gene region as instrumental variables to perform a standard inverse-variance-weighted (IVW) causal effect estimation using the TwoSampleMR package[23]. The LD r^2 clumping threshold was set to 0.001 to ensure independently associated genetic instruments. For autosomal genome-wide multi-instrument MR analysis, we conducted the analysis using the generalized summary-data-based Mendelian randomization (GSMR)[24] module in the GCTA software. The genome-wide significance threshold of p-values was set to 5×10^{-8}. The LD r^2 clumping threshold was set to 0.05.

cis-eQTL analysis To analyze the effect of any identified plasma ACE2 pQTL on nearby gene expression, we used two well-established, publicly available genetics of gene expression databases: (1) the large (N > 31,000 individuals) blood gene expression eQTLGen and (2) the GTEx consortium which is smaller (N ~ 1000 individuals) but contains a much broader tissue representation (54 non-diseased tissue sites) in order to detect tissue-specific expression quantitative trait loci (eQTL). For each discovered trans-pQTL of plasma ACE2, we extracted the eQTL association summary statistics for the genes within a 1Mb window of the lead ACE2-associated variant. The analysis was performed to evaluate candidate genes for the discovered ACE2 pQTL, excluding the MHC region on chromosome 6.

Chromatin states enrichment analysis We extracted 127 consolidated epigenomes from the Roadmap Epigenomics Project[25], which covers 15 chromatin states quantified across 28 human tissues and cell types. We used a SNP-based logistic regression to test for the enrichment of LD-corrected genome-wide significant ACE2 association signals on each chromatin state in each tissue or cell type:

$$\log\left(\frac{E[Pr(P_j<5\times10^{-8})]}{E[Pr(P_j\leq5\times10^{-8})]}\right)=\mu+\delta\ell_j+\beta C_j \quad (1)$$

where ℓ_j is the LD score of the *j*-th SNP, pre-calculated by the ldsc software for the HapMap3 SNPs (MHC region excluded); C_j takes a value of zero or one, as an indicator for whether the SNP is annotated to be within the particular chromatin state; β is the parameter of interest, i.e., the log odds ratio of significant ACE2 associations in the chromatin state compared to that in the other SNPs. The higher enrichment of the ACE2 association signals at the annotated SNPs, the higher β would be.

Statistical analysis The genetic effects from different cohorts were inverse-variance weighted in the meta-analysis using the METAL software. Subtraction of cohorts from GWAS summary statistics was done using the R package MetaSubtract. The heritability and genetic correlation parameters were estimated using the HDL and ldsc software. MR analysis was performed using the R package TwoSampleMR and the GCTA software. Linear and generalized linear models were fitted using the lm() and glm() procedures in R. The displayed parameters estimates are shown with standard errors or 95% confidence intervals.

Results

Epidemiologic correlates of plasma ACE2 levels Across our SCALLOP cohorts, plasma ACE2 levels were higher in men than women (Table S2). Significant positive associations were found between higher plasma ACE2 and CTSL1 (Cathepsin L1), BMI (Body mass index), triglycerides (TG), liver fat, hypertension, CVD, blood pressure, and diabetes (Figure S1, Table S3). Among commonly used anti-hypertensive drug classes, calcium channel blockers were positively associated with ACE2 levels, but there was no significant correlation between ACE2 levels and either ACE inhibitor use or angiotensin receptor blocker use in the overall sample. Stratifying the

sample by sex, ARB use was significantly associated with higher ACE2 levels in men only. We also identified significant sex heterogeneity for the associations with CSTL1, BMI, TG, and diabetes: the ACE2 associations with BMI and TG were stronger in men, whereas the associations with CSTL1 and diabetes appeared to be stronger in women.

GWAS identified ten genome-wide significant loci for plasma ACE2 We conducted a GWAMA in up to 28,204 individuals from 14 cohorts in the SCALLOP consortium for plasma ACE2 protein measured using the ©Olink platform. Our analysis included both the autosomes and the X chromosome, where the *ACE2* gene is encoded. Genotypes in each cohort were either obtained from whole-genome sequencing or imputed to the 1000 Genomes Phase 3 or Haplotype Reference Consortium reference panels (see Supplemental Methods). The ACE2 measurements were inverse-Gaussian transformed prior to analysis, adjusted for age, sex, population structure, and cohort-specific covariates. We tested 8,682,405 and 730,046 genetic variants on the autosomes and the X chromosome, respectively, with minor allele frequencies (MAFs) > 0.01.

The GWAMA discovered ten genome-wide significant loci for plasma ACE2 (Table 1, Figure 1, Figures S2 – S12). From conditional and joint multi-SNP analysis[26] on these discovered loci, we identified a secondary genome-wide significant association at the *EXOC3L4* locus on chromosome 14 (lead variant rs73356643, GWAS $P = 1.3 \times 10^{-6}$, conditional $P = 1.0 \times 10^{-8}$). Using the high-definition likelihood method[22], we estimated the autosomal SNP-based heritability of plasma ACE2 to be 16.1% (s.e. 2.5%). Adding the X chromosome heritability of 0.5% for male and 1.3% for female (Figure S13), we obtained an estimated genome-wide heritability of 16.6% and 17.4% for men and women, respectively. Based on an additive genetic effects model under Hardy-Weinberg equilibrium, the ten lead variants together explain 4.1% of the phenotypic variance of plasma ACE2 equivalent to about 30% of the heritability.

Except for the *ACE2* cis-pQTL (lead variant rs1849863, $P = 1.1 \times 10^{-85}$) and *ACE2*'s transcription factor *HNF1A* trans-pQTL (lead variant rs1169288, $P = 4.5 \times 10^{-78}$), the other eight trans-pQTL have not been previously associated with plasma ACE2 (Table 1, Table S4). In contrast to ACE2 and HNF1A pQTL, we were unable to replicate another previously reported locus on chromosome 21 (rs2186346, $P = 0.41$)[15].

To further characterize the potential functional, biologic, and clinical impact of our newly discovered ACE2-associated loci and prioritize plausible causal genes at these loci, we cross-referenced each of our index SNPs using (1) two available gene expression and eQTL resources: the eQTLGen consortium (Figure S14A, Table S5) and the GTEx consortium (Figure S14B, Table S6); (2) epigenomic information from the Roadmap Epigenomics Project[25], which covers 15 chromatin states quantified across various human tissues and cell types (Figure S15, Table S7); (3) genotype-phenotype association database records from PhenoScanner[27] (Table S8). We excluded the HLA locus on chromosome 6 because of the complicated genomic and LD structure in this region.

Tissue, epigenomic, and regulatory characterization of the ACE2-associated genomic loci

Across the eight autosomal ACE2-associated genomic loci, analysis of blood gene expression associations using eQTLGen data showed that plausible candidate genes generally ranked high among the genes underlying each trans-pQTL. However, multi-tissue analysis of the GTEx eQTL associations of the prioritized genes in the discovered pQTL at each of the eight loci did not point to clear specific tissues of ACE2 regulation. In GTEx, *ACE2* cis-eQTL associations were also found in multiple tissues, including brain, tibial nerve, tibial artery, and pituitary (Figure S16), though *ACE2* tends to have specifically lower expression in these tissues (Figure S17). Across diverse tissues and cell types, chromatin states such as weak transcription and

heterochromatin harbored enriched plasma ACE2 genetic associations (Figure S15). Notably, the enrichment was more tissue-specific for the strong transcription chromatin state, where the ACE2 genetic associations were particularly enriched in T cells, natural killer cells, hematopoietic cells, and blood lymphoblastoid cells. This suggests a shared genetic regulation between plasma ACE2 and hemocyte immune functions.

Functional and phenotypic annotation of the ACE2-associated loci Notably, the majority of the autosomal trans-pQTL for ACE2 have been pleiotropically associated through prior GWAS with various cardiovascular, metabolic, inflammatory/immune, and pulmonary traits (Table S8). These loci are detailed further below.

Two loci involve coding variants from the same family of hepatic/pancreatic transcription factor genes. These include missense variants of *HNF1A* (lead variant rs1169288 or p.Ile27Leu, $P = 4.5 \times 10^{-78}$) and hepatocyte nuclear factor *HNF4A* (lead variant rs1800961, p.Thr139Ile, $P = 1.3 \times 10^{-9}$). Both of these *HNF1A* and *HNF4A* missense variants have been previously associated with HDL, LDL and total cholesterol, type 2 diabetes, CHD, C-reactive protein (CRP), fibrinogen, and coagulation factor VII, and other hepatically-synthesized enzymes or enzyme inhibitors such as GGT or alpha1-antitrypsin levels. Especially, p.Ile27Leu is a well-known variant with the Leu allele (corresponds to the C allele of rs1169288), increasing the risk of type 2 diabetes; meanwhile, it reduces plasma ACE2 concentration ($\beta = -0.175$, *s.e.* = 0.009) in our study. Rare loss-of-function variants of *HNF1A* and *HNF4A* can cause the Mendelian disorder mature onset diabetes of the young (MODY). The transcription factors HNF1A and HNF4A have been shown to regulate ACE2 expression in a pancreatic and ileal-specific manner, respectively[28,29]. Interestingly, both HNF1A and HNF4A also have a crucial role in protein fucosylation[30]. Given that the SARS-CoV-2 spike (S) protein is heavily fucosylated with host-

derived glycans[31], in addition to regulating the expression of ACE2, these two loci could potentially influence S protein glycosylation and thereby influence the virus' entry to the host cell[3].

In addition to the two HNF transcription factors, two other loci contain genes involved in hepatic gene transcriptional regulation. *RORA* encodes retinoic acid receptor-alpha, another hepatic transcriptional activator involved in the regulation of circadian rhythm, metabolism, and immune function[32]. The intronic *RORA* variant (lead variant rs340005, $P = 4.7\times10^{-9}$) is located in an ENCODE distal enhancer region and has been previously associated with CRP and GGT levels. The lead SNP rs2954021 ($P = 4.6 \times 10^{-14}$) located upstream of the pleiotropic *TRIB1* gene has been associated with various metabolic traits, liver enzymes, plasma lipids, kidney function, blood cell traits, hepatic steatosis, and CHD (examples of colocalization of the ACE2 associations with HDL, LDL, BMI, and waist circumference are given in Figure S18). The protein product of *TRIB1*, Tribbles-1, post-translationally regulates the degradation of CCAAT/enhancer-binding protein α (C/EBPα), which contributes to the dysregulation of hepatic lipid-related gene expression[33] and is also a positive regulator of HNF4A[34].

The lead variant at the *SERPINA1* locus, rs28929474 ($P = 1.0 \times 10^{-27}$), encodes the canonical European Pi*Z allele p.Glu342Lys causing recessively inherited alpha-1-antitrypsin deficiency (AATD), a disease that affects the lung and liver[35]. The Lys allele leads to lower serum concentrations of AAT[36] and higher plasma ACE2 levels ($\beta = 0.312$, *s.e.* = 0.029). In population-based genetic studies, rs28929474 has been associated with various metabolic phenotypes (height, bone mineral density, systolic blood pressure, fat-free mass, gallstones, lipids), pulmonary function, alcohol consumption, and plasma levels of various hepatic enzymes (alkaline phosphatase, alanine aminotransferase) and acute-phase proteins synthesized in the

liver (e.g., CRP, AFP). Importantly, the alpha-1-antitrypsin serine protease inhibitor additionally possesses anti-viral and anti-inflammatory properties, including inhibition of TMPRSS2 and ADAM17 enzyme activities, and therefore has been proposed as a possible host protective factor against COVID-19[37,38]. The class 1 MHC locus (lead variant rs3094087, $P = 6.3 \times 10^{-9}$) has been associated with a variety of autoimmune, GI, blood cell counts, anthropometric, and pulmonary traits.

Several of the newly identified pQTL have less clearly understood biological relationships to plasma ACE2 and/or are less clearly mapped with respect to causal genes. The *MICAL3* intronic lead variant rs5992134 ($P = 2.5 \times 10^{-10}$) is a cis-eQTL for *MICAL3* and located within a predicted distal enhancer region and has been previously associated with liver enzymes, pulmonary function, and SBP. MICAL-3 is an NADPH-dependent oxidoreductase enzyme that participates in actin cytoskeleton reorganization[39]. The pQTL on chromosome 16 is located in a gene desert (closest gene *SALL1* ~300 kb), but has been previously associated with CRP and red blood cell count. The lead variant rs17616063 is located within an ENCODE distal enhancer and overlaps a transcription factor (TF) binding site (Ensembl regulatory feature id ENSR00000085879), with the nearest experimentally verified motif ENSM00156191351 binding with protein HNF4A (encoded on the chromosome 20 locus) in the HepG2 cell line (Figure 1, Figure S19). The chromosome 14 locus (*EXOC3L4* intronic lead variant rs2274685, $P = 3.7 \times 10^{-16}$) has been associated with liver enzymes and platelet quantitative traits; and with gene expression of *CDC42BPB*, whereas another gene nearby, *TNFAIP2*, is abundantly expressed in immune cells and has been implicated in inflammation and infectious disease[40]. From conditional and joint multi-SNP analysis on these discovered loci, we identified a secondary genome-wide significant association at the *EXOC3L4* locus on chromosome 14 (lead

variant rs73356643, GWAS $P = 1.3\times10^{-6}$, conditional $P = 1.0 \times 10^{-8}$).

At the cis-pQTL located near the *ACE2* gene, the lead variant rs1849863 was strongly associated with plasma ACE2 levels ($P = 1.1 \times10^{-85}$). With a threshold of LD $r^2 < 0.001$, we obtained another two independent cis-instruments rs143380244 ($P = 8.8 \times 10^{-9}$) and rs73202884 ($P = 2.8 \times 10^{-13}$) for plasma ACE2. The three cis-variants together capture 1.4% of the phenotypic variance of ACE2, equivalent to 8.8% of ACE2's heritability. Notably, no established clinical or phenotypic association for the ACE2 cis-regulatory locus was found, possibly due to the lack of association studies on the X chromosome (Figure S20). Our sentinel SNP at the cis-pQTL rs1849863 is in strong LD with the previously reported top SNP rs12558179 by Nelson et al.[15] (R^2 = 0.9489 and D' = 0.9741 in 1000 Genomes European individuals). A recent report identified a cis-regulatory rare variant rs190509934 for *ACE2* gene expression that influences COVID-19 risk[41], for which the rare allele C only forms a haplotype with the minor, C, allele of our detected top SNP rs1849863. However, the rare allele of rs190509934 was reported by Horowitz et al. to reduce the ACE2 mRNA expression, whereas we found that the minor allele of rs1849863 increases the plasma ACE2 level.

Genetic and causal relationships between plasma ACE2 and cardiovascular diseases We explored the shared genetic architecture between plasma ACE2 and cardiovascular and metabolic risk factors and CVD outcomes. We estimated bivariate ACE2-trait genetic correlations using two different databases of GWAS summary statistics[42]. First, we assessed genetic correlations between plasma ACE2 and various human lifestyle/behavioral, psychosocial, health-related and biomarker traits with available GWAS summary statistics using the high-definition likelihood method[22] (Table S9). Among these, plasma ACE2 showed significant positive genetic correlations with cigarette smoking, blood pressure, cholesterol

levels, CRP, and anthropometric traits (false discovery rate < 5%, Figure 2, Table S9). Next, in order to examine the shared genetic basis between plasma ACE2 level and specific vascular diseases or diagnoses, we specifically extracted the GWAS summary-level data for 48 vascular-disease-related phenotypes from UK Biobank (see Methods, Table S10), categorized into 18 for heart disease, 15 for blood pressure, 9 for stroke, and the other 6 for blood lipids. We assessed genetic correlations and found that plasma ACE2 levels had positive genetic correlations with most of the vascular disease phenotypes, except for HDL cholesterol (Figure 3, Table S11).

Next, we conducted bidirectional causal inference using generalized summary-level Mendelian randomization (GSMR) to identify potential causal effects of ACE2 on these traits (Figure 3, Table S12). In the GSMR analysis, the instrumental variables used to assess causality are restricted to the autosomal trans-pQTL for ACE2. Paradoxically, GSMR suggested that genetically elevated plasma ACE2 level is associated with reduced risk of heart disease, HDL, LDL, and total cholesterol. However, the results of MR analyses can be biased if some of the genetic instruments are invalid because they additionally influence the outcome through pathways other than ACE2 (so-called “horizontal pleiotropy”).

Since several of the ACE2-associated autosomal loci are pleiotropically associated with various cardiometabolic and inflammation-related phenotypes, we further conducted the MR analysis for vascular diseases using only the X chromosome cis-pQTL *ACE2* locus. No significant causal effect was detected (Figure 3, Table S13) for any of the phenotypes with summary association statistics of the X chromosome. It should be noted that while on the one hand using only cis-pQTL as the MR instrument may avoid the issue of bias due to horizontal pleiotropy, on the other hand, there may be reduced power to detect a true causal effect using

only the cis-pQTL due to the smaller number of SNPs considered.

Genetic correlation and causal inference between ACE2 and COVID-19 Based on the latest genome-wide association summary statistics for severe COVID-19[19], we obtained an estimated autosomal genetic correlation of 0.476 ($P = 9.4 \times 10^{-3}$) with plasma ACE2, indicating an increased risk of severe COVID-19 for people that have genetically raised levels of circulating ACE2 protein. At the cis-pQTL located near the *ACE2* gene, the lead variant rs4830984 was nominally significantly associated with severe COVID-19 ($P = 0.026$)[19]. GSMR analysis based on all the discovered ACE2 pQTL as instruments revealed no significant causal effect. To avoid horizontal pleiotropy, we considered in our MR analysis only the independent cis-variants as genetic instruments for causal inference (Figure 4A). Based on inverse-variance weighted (IVW) Mendelian randomization (MR), we estimated an odds ratio of 1.63 (95% CI, 1.10 to 2.42; $P = 0.01$) for ACE2 on COVID-19 severity (Figure 4B). Reverse MR analysis instrumenting on the severe COVID-19 loci did not reveal a significant estimate (GSMR $P = 0.95$), suggesting the absence of a causal effect of COVID-19 on ACE2 levels. As validation, in the HGI COVID-19 hospitalization data (GWAS B2, GenOMICC subtracted), we estimated the causal effect of ACE2 with an OR of 1.52 (95% CI, 1.05 to 2.21; $P = 0.03$), and in the HGI COVID-19 infection data (GWAS C2, GenOMICC subtracted), the estimated OR was 1.60 (95% CI, 1.08 to 2.37; $P = 0.02$).

Discussion

We provide a detailed characterization of the phenotypic and genetic correlates of plasma ACE2 using data from several large cohort studies, including the identification of eight novel genetic loci, which together explain 30% of the protein's heritability. We further demonstrate that ACE2

levels are genetically correlated with both cardiovascular disease and COVID-19 clinical outcomes. Moreover, we found genetic evidence that elevated ACE2 levels may be causally related to COVID-19 severity, hospitalization, and infection. These data add further evidence that the cardiovascular complications of COVID-19 disease patients may be mechanistically related to ACE2.

Genetics of plasma ACE2 levels Previous studies have measured the heritability of ACE2 levels[43,44], with the SNP heritability estimated to be 33% in the PURE study[44] using the same Olink PEA-based ACE2 antigen assay. The heritability was estimated to be 16% among Europeans (with somewhat higher estimates for Latinos and Persians) in our study. Differences in heritability estimation methods or geographic or demographic differences between populations may account for different heritabilities across populations. Despite these differences, we were able to confirm two recently identified genetic loci associated with ACE2, the cis-pQTL on chromosome X and *HNF1A*. The peak signal on the X chromosome is located around 100 kb upstream of *ACE2* and therefore likely reflects regulatory effects on *ACE2* expression. These variants are not in LD with several common missense variants of *ACE2* that are predicted to impact ACE2 protein stability[45] or SARS-CoV-2-binding[46,47]. The presence of a strong X-linked locus may explain in part the observed sex differences in *ACE2* expression across various tissues[48] as well as the higher circulating levels of ACE2 in men.

Along with the *HNF1A* ACE2-associated autosomal locus, the eight newly identified genetic loci associated with ACE2 may help to shed additional light on mechanisms by which cellular or plasma ACE2 levels are regulated under physiologic and pathologic conditions[48]. The transcription factors HNF1A and HNF4A regulate *ACE2* expression in the pancreas and GI tract[28,29]. Besides *ACE2* expression itself, we also identified another discovered novel locus

being the TF binding site of HNF4A. The *HNF1A, HNF4A, TRIB1* loci are also important determinants of cardiovascular disease, diabetes, lipids, and adiposity-related traits, and therefore likely contribute to the genetic correlations we observed between these cardiometabolic phenotypes as well as the findings from Mendelian randomization analysis between ACE2 for BMI and diabetes in the PURE study. While the *ACE2* locus on chromosome X and several of the newly identified autosomal genetic loci likely influence ACE2 levels through regulation of cellular expression or transcription of *ACE2* in various tissues, other loci may influence the amount of ACE2 shedding from the cell membrane. In this regard, the protease inhibitor alpha-1-antitrypsin is capable of inhibiting the enzymatic activities of two proteases TMPRSS2 and ADAM17, both of which are involved in ACE2 shedding[37,38].

Relationship of ACE2 and RAS to CVD In our analysis, soluble ACE2 levels were positively correlated with several traditional CVD risk factors, which is consistent with its important role in counter-regulation of the renin-angiotensin system (RAS). Higher ACE2 has additionally been associated with poorer prognosis in patients with pre-existing CVD and recently was found to prospectively predict incident CVD, mortality, diabetes, and heart failure in previously unaffected individuals[9–12], independently of traditional risk factors[44]. In particular, ACE2 was ranked higher as a predictor of overall mortality compared to smoking, diabetes, blood pressure, BMI, and lipids. We also found that higher plasma ACE2 level is genetically correlated with a higher risk of vascular diseases, including coronary heart disease, hypertension, stroke, and heart failure. Nonetheless, Mendelian randomization analysis restricted to the cis-pQTL ACE2 locus was unable to demonstrate a causal relationship between ACE2 and CVD-related outcomes. We speculate that the apparent but paradoxical protective effect of ACE2 on CVD by GSMR using the additional autosomal trans-pQTL associations may reflect that

extensive pleiotropy exhibited by most of the autosomal trans-pQTL on blood lipids and other CVD risk factors.

Both ACE and ACE2 have connections with cardiovascular mechanisms. Although the sparse genetic architecture of both proteins did not result in significant genetic correlation, a suggestive causal effect of ACE2 on ACE was detected (see also Supplemental Material). We have shown above that ACE2 is genetically correlated with a series of vascular traits while its causal role was detected only for CVD-related phenotypes; in contrast, the hypertension target ACE showed a clear causal effect on blood pressure. Our results thus suggest distinct downstream functions of the two homologous proteins.

Relationship of ACE2 to COVID-19 In addition to its role in the regulation of the RAS, cellular ACE2 is an important receptor for SARS-CoV-2 and other coronaviruses. Our findings of a genetic correlation of soluble ACE2 and additional evidence that genetically determined sACE2 is associated with increased risk of COVID-19 severity are consistent with recent in vitro data that the secreted form of ACE2 plays a direct role in cell entry of SARS-CoV-2 via receptor-mediated endocytosis[49]. Our results are also consistent with human genetic studies indicating genetic variation in soluble ACE2 influences COVID-19 risk[41,50].

Based on animal studies, ACE inhibitors or ARB may up-regulate *ACE2* gene expression in cardiac cells[51,52], which might increase COVID-19 susceptibility. However, human observational studies have not found a robust relationship between higher ACE2 levels and ACE-inhibitors or ARB use[13,14]. Consistent with these results, we did not find a significant association of ACE inhibitors or ARB on plasma ACE2.

Study limitations While our study adds important information to the regulation of ACE2 and the genetic relationship between ACE2 and other phenotypes, several limitations should be

highlighted.

1) The genetic loci identified herein are associated with the soluble form of ACE2 found in human plasma, which is relatively easily obtainable and can be studied with sufficiently large sample sizes. The relationship of the plasma ACE2 pQTL or other genetic loci to tissue levels of the full-length cellular ACE2 receptor, the precise genetic regulatory mechanisms underlying these associations, as well as the relevant tissue sources of soluble ACE2, all remain to be determined. 2) Despite the large sample size, we were still only able to account for 30% of ACE2's heritability, leaving the remaining loci (e.g., common genetic variants with smaller effect sizes or rare genetic variants of large effect) to be discovered; 3) The Olink PEA assay quantifies plasma ACE2 concentration, rather than the enzymatic activity of the protein. ACE2's activity as an enzyme may or may not have the same biological basis as its general abundance. An earlier family-based study of plasma ACE2 *activity* levels in healthy individuals[43] estimated that genetic factors accounted for 67% of the phenotypic variance in ACE2 but only 7% of individuals had detectable ACE2 levels using a fluorogenic activity assay. Using a more sensitive fluorogenic assay, readouts from an ACE2 enzyme activity assay tend to have a strong correlation with protein levels of ACE2[53]. Given the specificity of the Olink ACE2 assay (Supplemental Material), one can expect that the effects on Olink and enzyme activity assay measurements would likely be directionally concordant.

In summary, our findings suggest that ACE2 may play an important role not only in susceptibility to cardiovascular, metabolic, and pulmonary disorders but also in susceptibility to COVID-19 disease severity. These findings have potential therapeutic implications for the counter-regulatory ACE2/Ang-(1-7) axis in modulating COVID-19 disease severity[54]. In particular, identification of additional genetic factors involved in the regulation of ACE2 levels

may help disentangle the potential causal roles of soluble vs. cellular ACE2 in the regulation of chronic diseases, COVID-19 infection, disease severity, and immunity.

Acknowledgements

We thank Dr. Jie Zheng at the University of Bristol for sharing the harmonized GWAS summary statistics used in LD-Hub and IEU Open GWAS Project. We thank Marie Pigeyre and Guillaume Paré at McMaster University for sharing the serum ACE GWAS summary statistics from the ORIGIN trial. We thank Yuanyuan Sun and Ao Lan at Sun Yat-sen University for useful comments on the manuscript. We thank the Edinburgh Compute and Data Facility at the University of Edinburgh for providing high-performance computing resources. We thank the members of the cited consortia of genome-wide association studies for making their data available. Cohort-specific acknowledgements are given in the Supplemental Material.

Author contributions

X.S. and J.F.W. initiated and coordinated the study. E.M.-D. and P.K.J. contributed to autosomal meta-analysis. Z.Y., X.S., A. Richmond, and L.K. contributed to X chromosome meta-analysis. Z.Y., N.P., T.H., Y.H., and H.G. contributed to the MR analysis. Z.Y., J.C., R.Z., T.L., Y.W., and K.Y. contributed to functional analysis. Z.Y., Z.N., and C.Z. contributed to heritability and genetic correlation analysis. S.G., B.P., A. Ramisch, E.T.D., G.P., N.E., J.H., X.H., D.Z., T.B., S.-J.H., E.W., M.P., L.M.R., A.K., J. Peters, A.V., L.L., S. Elmståhl, G.D., J. Petrie, O.P., L.F. Y.C., C.Y., U.V., T.E., S. Enroth, Å.J., U.G., C.L., D.L., C.H., T.L.A., C.K., A.W.M., A.S., L.W., A.G., E.Z., J.M.S., A.B., K.M., and A.M. contributed to the cohort-level analysis. L.R., E.P.-C., S.C., A.D.B., K.R., and J.K.B. contributed to GenOMICC severe COVID-19 genomic analysis. X.S. drafted the manuscript. L.K., Y.P., A.P.R., and J.F.W. contributed to manuscript

writing. All authors gave final approval to publish.

Sources of Funding

X.S. was in receipt of a Swedish Research Council Starting Grant (No. 2017-02543) and National Natural Science Foundation of China grant (No. 12171495). J.F.W. and C.H. acknowledge support from the Medical Research Council Human Genetics Unit program grant "Quantitative Traits in Health and Disease" (U. MC_UU_00007/10). The work of L.K. was supported by an RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1). A.D.B. would like to acknowledge funding from the Wellcome PhD training fellowship for clinicians (204979/Z/16/Z), the Edinburgh Clinical Academic Track (ECAT) programme. A.P.R was supported by grants R01HL146500 and R01HL132947.

Disclosures

A.S.B. has received grants outside of this work from AstraZeneca, Bayer, Biogen, Bioverativ, Merck, Novartis, Pfizer and Sanofi. E.T.D. is currently an employee of GSK, and the work was performed before he joined GSK.

10.1161/CIRCULATIONAHA.121.057888

Supplemental Material

Supplemental Methods

Supplemental Results & Discussion

Figures *S1 - S24*

Tables *S1 - S21*

Supplemental Information

References *55 - 57*

References

1. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S. 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.
2. Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, Lu G, Qiao C, Hu Y, Yuen KY, Wang Q, Zhou H, Yan J, Qi J. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell*. 2020;181:894-904.e9.
3. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology*. 2020;5:562–569.
4. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, Li F. Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America*. 2020;117:11727–11734.
5. Sungnak W, Huang N, Bécavin C, Berg M, Queen R, Litvinukova M, Talavera-López C, Maatz H, Reichart D, Sampaziotis F, Worlock KB, Yoshida M, Barnes JL, Banovich NE, Barbry P, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature Medicine*. 2020;26:681–687.
6. Turner AJ, Hiscox JA, Hooper NM. ACE2: From vasopeptidase to SARS virus receptor. *Trends in Pharmacological Sciences*. 2004;25:291–294.
7. Lambert DW, Yarski M, Warner FJ, Thornhill P, Parkin ET, Smith AI, Hooper NM, Turner AJ. Tumor necrosis factor-α convertase (ADAM17) mediates regulated ectodomain shedding of the severe-acute respiratory syndrome-coronavirus (SARS-CoV) receptor, angiotensin-converting enzyme-2 (ACE2). *Journal of Biological Chemistry*. 2005;280:30113–30119.
8. Heurich A, Hofmann-Winkler H, Gierer S, Liepold T, Jahn O, Pohlmann S. TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. *Journal of Virology*. 2014;88:1293–1307.
9. Kornilov SA, Lucas I, Jade K, Dai CL, Lovejoy JC, Magis AT. Plasma levels of soluble ACE2are associated with sex, Metabolic Syndrome, and its biomarkers in a large cohort, pointing to a possible mechanism for increased severity in COVID-19. *Critical Care*. 2020;24:452.
10. Swärd P, Edsfeldt A, Reepalu A, Jehpsson L, Rosengren BE, Karlsson MK. Age and sex differences in soluble ACE2 may give insights for COVID-19. *Critical Care*. 2020;24:221.
11. Wallentin L, Lindbäck J, Eriksson N, Hijazi Z, Eikelboom JW, Ezekowitz MD, Granger CB, Lopes RD, Yusuf S, Oldgren J, Siegbahn A. Angiotensin-converting enzyme 2 (ACE2) levels in relation to risk factors for COVID-19 in two large cohorts of patients with atrial fibrillation. *European Heart Journal*. 2020;41:4037–4046.
12. Chirinos JA, Cohen JB, Zhao L, Hanff T, Sweitzer N, Fang J, Corrales-Medina V, Anmar R, Morley M, Zamani P, Bhattacharya P, Brandimarto J, Jia Y, Basso MD, Wang Z, et al. Clinical and proteomic correlates of plasma ACE2 (angiotensin-converting enzyme 2) in human heart failure. *Hypertension*. 2020;76:1526–1536.
13. Sama IE, Ravera A, Santema BT, Van Goor H, Ter Maaten JM, Cleland JGF, Rienstra M, Friedrich AW, Samani NJ, Ng LL, Dickstein K, Lang CC, Filippatos G, Anker SD, Ponikowski P, et al. Circulating plasma concentrations of angiotensin-converting enzyme 2

inmen and women with heart failure and effects of renin-angiotensin-aldosterone inhibitors. *European Heart Journal*. 2020;41:1810–1817.
14. Zimmermann T, Walter JE, Lopez-Ayala P, Strebel I, Amrein M, Koechlin M, Honegger U, Mueller C, Albus M, Zellweger M, Liu YC, Glarner N, Boeddinghaus J, Freese M, Haaf P, et al. Influence of renin-angiotensin-aldosterone system inhibitors on plasma levels of angiotensin-converting enzyme 2. *ESC Heart Failure*. 2021:1717–1721.
15. Nelson CP, Sama IE, Codd V, Webb TR, Ye S, Lang CC, Voors AA, Ng LL, Samani NJ. Genetic Associations with Plasma Angiotensin Converting Enzyme 2 Concentration: Potential Relevance to COVID-19 Risk. *Circulation*. 2020;142:1117–1119.
16. Wik L, Nordberg N, Broberg J, Björkesten J, Assarsson E, Henriksson S, Grundberg I, Pettersson E, Westerberg C, Liljeroth E, Falck A, Lundberg M. Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. *Molecular and Cellular Proteomics*. 2021;20:100168.
17. Pigeyre M, Sjaarda J, Chong M, Hess S, Bosch J, Yusuf S, Gerstein H, Paré G. ACE and type 2 diabetes risk: A mendelian randomization study. *Diabetes Care*. 2020;43:835–842.
18. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, Gunnarsdottir K, Helgason A, Oddsson A, Halldorsson B V., Jensson BO, Zink F, Halldorsson GH, Masson G, Arnadottir GA, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nature Genetics*. 2021;53:1712–1721.
19. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S, Parkinson N, Fourman MH, Russell CD, Furniss J, Richmond A, Gountouna E, Wrobel N, Harrison D, et al. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2021;591:92–98.
20. The COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *European Journal of Human Genetics*. 2020;28:715–718.
21. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191.
22. Ning Z, Pawitan Y, Shen X. High-definition likelihood inference of genetic correlations across human complex traits. *Nature Genetics*. 2020;52:859–864.
23. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, et al. The MR-base platform supports systematic causal inference across the human phenome. *eLife*. 2018;7:0.
24. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, Robinson MR, McGrath JJ, Visscher PM, Wray NR, Yang J. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications*. 2018;9:224.
25. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–329.
26. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, Frayling TM, McCarthy MI, Hirschhorn JN, Goddard ME, Visscher PM. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*. 2012;44:369–375.
27. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, Paul DS, Freitag D, Burgess S, Danesh J, Young R, Butterworth AS. PhenoScanner: A database of human genotype-

phenotype associations. *Bioinformatics*. 2016;32:3207–3209.
28. Pedersen KB, Chhabra KH, Nguyen VK, Xia H, Lazartigues E. The transcription factor HNF1α induces expression of angiotensin-converting enzyme 2 (ACE2) in pancreatic islets from evolutionarily conserved promoter motifs. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. 2013;1829:1225–1235.
29. Verstockt B, Verstockt S, Abdu Rahiman S, Ke BJ, Arnauts K, Cleynen I, Sabino J, Ferrante M, Matteoli G, Vermeire S. Intestinal Receptor of SARS-CoV-2 in Inflamed IBD Tissue Seems Downregulated by HNF4A in Ileum and Upregulated by Interferon Regulating Factors in Colon. *Journal of Crohn's and Colitis*. 2021;15:485–498.
30. Lauc G, Essafi A, Huffman JE, Hayward C, Knežević A, Kattla JJ, Polašek O, Gornik O, Vitart V, Abrahams JL, Pučić M, Novokmet M, Redžić I, Campbell S, Wild SH, et al. Genomics meets glycomics-the first gwas study of human N-glycome identifies HNF1A as a master regulator of plasma protein fucosylation. Gibson G, ed. *PLoS Genetics*. 2010;6:1–14.
31. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*. 2020;369:330–333.
32. Kojetin DJ, Burris TP. REV-ERB and ROR nuclear receptors as drug targets. *Nature Reviews Drug Discovery*. 2014;13:197–216.
33. Bauer RC, Sasaki M, Cohen DM, Cui J, Smith MA, Yenilmez BO, Steger DJ, Rader DJ. Tribbles-1 regulates hepatic lipogenesis through posttranscriptional regulation of C/EBPα. *Journal of Clinical Investigation*. 2015;125:3809–3818.
34. Soubeyrand S, Martinuk A, McPherson R. TRIB1 is a positive regulator of hepatocyte nuclear factor 4-Alpha. *Scientific Reports*. 2017;7:5574.
35. Brebner JA, Stockley RA. Recent advances in α-1-antitrypsin deficiency-related lung disease. *Expert Review of Respiratory Medicine*. 2013;7:213–230.
36. Seixas S, Marques PI. Known mutations at the cause of Alpha-1 antitrypsin deficiency an updated overview of SERPINA1 Variation spectrum. *Application of Clinical Genetics*. 2021;14:173–194.
37. de Loyola MB, dos Reis TTA, de Oliveira GXLM, da Fonseca Palmeira J, Argañaraz GA, Argañaraz ER. Alpha-1-antitrypsin: A possible host protective factor against Covid-19. *Reviews in Medical Virology*. 2021;31:e2157.
38. Shapira G, Shomron N, Gurwitz D. Ethnic differences in alpha-1 antitrypsin deficiency allele frequencies may partially explain national differences in COVID-19 fatality rates. *FASEB Journal*. 2020;34:14160–14165.
39. Wu H, Yesilyurt HG, Yoon J, Terman JR. The MICALs are a Family of F-actin Dismantling Oxidoreductases Conserved from Drosophila to Humans. *Scientific Reports*. 2018;8:937.
40. Jia L, Shi Y, Wen Y, Li W, Feng J, Chen C. The roles of TNFAIP2 in cancers and infectious diseases. *Journal of Cellular and Molecular Medicine*. 2018;22:5188–5195.
41. Horowitz JE, Kosmicki JA, Damask A, Sharma D, Roberts GHL, Justice AE, Banerjee N, Coignet M V, Yadav A, Leader JB, Marcketta A, Park DS, Lanche R, Maxwell E, Knight SC, et al. Genome-wide analysis in 756,646 individuals provides first genetic evidence that ACE2 expression influences COVID-19 risk and yields genetic risk scores predictive of severe disease. *medRxiv*. 2021:2020.12.14.20248176.
42. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, Hemani G, Tansey K, Laurin C, Pourcain BS, Warrington NM, Finucane HK, Price AL, Bulik-Sullivan BK, Anttila V, et al. LD Hub: A centralized database and web interface to perform LD score

regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*. 2017;33:272–279.
43. Rice GI, Jones AL, Grant PJ, Carter AM, Turner AJ, Hooper NM. Circulating activities of angiotensin-converting enzyme, its homolog, angiotensin-converting enzyme 2, and neprilysin in a family study. *Hypertension*. 2006;48:914–920.
44. Narula S, Yusuf S, Chong M, Ramasundarahettige C, Rangarajan S, Bangdiwala SI, van Eikels M, Leineweber K, Wu A, Pigeyre M, Paré G. Plasma ACE2 and risk of death or cardiometabolic diseases: a case-cohort analysis. *The Lancet*. 2020;396:968–976.
45. Benetti E, Tita R, Spiga O, Ciolfi A, Birolo G, Bruselles A, Doddato G, Giliberti A, Marconi C, Musacchia F, Pippucci T, Torella A, Trezza A, Valentino F, Baldassarri M, et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. *European Journal of Human Genetics*. 2020;28:1602–1614.
46. Hussain M, Jabeen N, Raza F, Shabbir S, Baig AA, Amanullah A, Aziz B. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *Journal of Medical Virology*. 2020;92:1580–1586.
47. Cruz J de O, Conceição IMCA, Sousa SMB, Luizon MR. Functional prediction and frequency of coding variants in human ACE2 at binding sites with SARS-CoV-2 spike protein on different populations. *Journal of Medical Virology*. 2021;93:71–73.
48. Li Y, Xu Q, Ma L, Wu D, Gao J, Chen G, Li H. Systematic profiling of ACE2 expression in diverse physiological and pathological conditions for COVID-19/SARS-CoV-2. *Journal of Cellular and Molecular Medicine*. 2020;24:9478–9482.
49. Yeung ML, Teng JLL, Jia L, Zhang C, Huang C, Cai JP, Zhou R, Chan KH, Zhao H, Zhu L, Siu KL, Fung SY, Yung S, Chan TM, To KKW, et al. Soluble ACE2-mediated cell entry of SARS-CoV-2 via interaction with proteins related to the renin-angiotensin system. *Cell*. 2021;184:2212-2228.e12.
50. Gaziano L, Giambartolomei C, Pereira AC, Gaulton A, Posner DC, Swanson SA, Ho YL, Iyengar SK, Kosik NM, Vujkovic M, Gagnon DR, Bento AP, Barrio-Hernandez I, Rönnblom L, Hagberg N, et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nature Medicine*. 2021;27:668–676.
51. Ishiyama Y, Gallagher PE, Averill DB, Tallant EA, Brosnihan KB, Ferrario CM. Upregulation of Angiotensin-Converting Enzyme 2 after Myocardial Infarction by Blockade of Angiotensin II Receptors. *Hypertension*. 2004;43:970–976.
52. Ferrario CM, Jessup J, Chappell MC, Averill DB, Brosnihan KB, Tallant EA, Diz DI, Gallagher PE. Effect of angiotensin-converting enzyme inhibition and angiotensin II receptor blockers on cardiac angiotensin-converting enzyme 2. *Circulation*. 2005;111:2605–2610.
53. Xiao F, Burns KD. Measurement of angiotensin converting enzyme 2 activity in biological fluid (ACE2). *Methods in Molecular Biology*. 2017;1527:101–115.
54. Monteil V, Kwon H, Prado P, Hagelkrüys A, Wimmer RA, Stahl M, Leopoldi A, Garreta E, Hurtado del Pozo C, Prosper F, Romero JP, Wirnsberger G, Zhang H, Slutsky AS, Conder R, et al. Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell*. 2020;181:905-913.e7.
55. Anderson G, Cummings S, Freedman LS, Furberg C, Henderson M, Johnson SR, Kuller L, Manson J, Oberman A, Prentice RL, Rossouw. JE, Finnegan L, Hiatt R, Pottern L, McGowan J, et al. Design of the Women's Health Initiative clinical trial and observational study. *Controlled Clinical Trials*. 1998;19:61–109.
56. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and

annotation of genetic associations with FUMA. *Nature Communications*. 2017;8:1826.
57. Di Angelantonio E, Thompson SG, Kaptoge S, Moore C, Walker M, Armitage J, Ouwehand WH, Roberts DJ, Danesh J, Donovan J, Ford I, Henry R, Hunt BJ, Le Huray B, Mehenny S, et al. Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. *The Lancet*. 2017;390:2360–2371.

Table 1. Summary of Ten Genome-wide Significant Loci for Plasma ACE2.

Lead SNP	Chr.	Position (bp)	R.A.	E.A.	Freq.	β (s.e.)	*P*	Nearest Coding Gene
rs3094087	**6**	31061561	**C**	**T**	0.853	0.073 (0.013)	6.3×10^{-9}	*HLA-B*
rs2954021	**8**	126482077	**G**	**A**	0.473	0.067 (0.009)	4.6×10^{-14}	*TRIB1*
rs1169288*	**12**	121416650	**A**	**C**	0.334	-0.175 (0.009)	4.5×10^{-78}	*HNF1A*
rs28929474†	**14**	94844947	**C**	**T**	0.026	0.312 (0.029)	1.0×10^{-27}	*SERPINA1*
rs2274685	**14**	103575070	**G**	**A**	0.550	-0.074 (0.009)	3.7×10^{-16}	*EXOC3L4*
rs340005	**15**	60878030	**G**	**A**	0.581	0.054 (0.009)	4.7×10^{-9}	*RORA*
rs17616063	**16**	51436882	**G**	**A**	0.915	0.115 (0.017)	5.2×10^{-11}	*SALL1*
rs1800961	**20**	43042364	**C**	**T**	0.049	0.130 (0.021)	1.3×10^{-9}	*HNF4A*
rs5992134	**22**	18433994	**G**	**T**	0.229	-0.067 (0.011)	2.5×10^{-10}	*MICAL3*
rs1849863	**23**	15736245	**T**	**C**	0.245	0.164 (0.008)	1.1×10^{-85}	*ACE2*

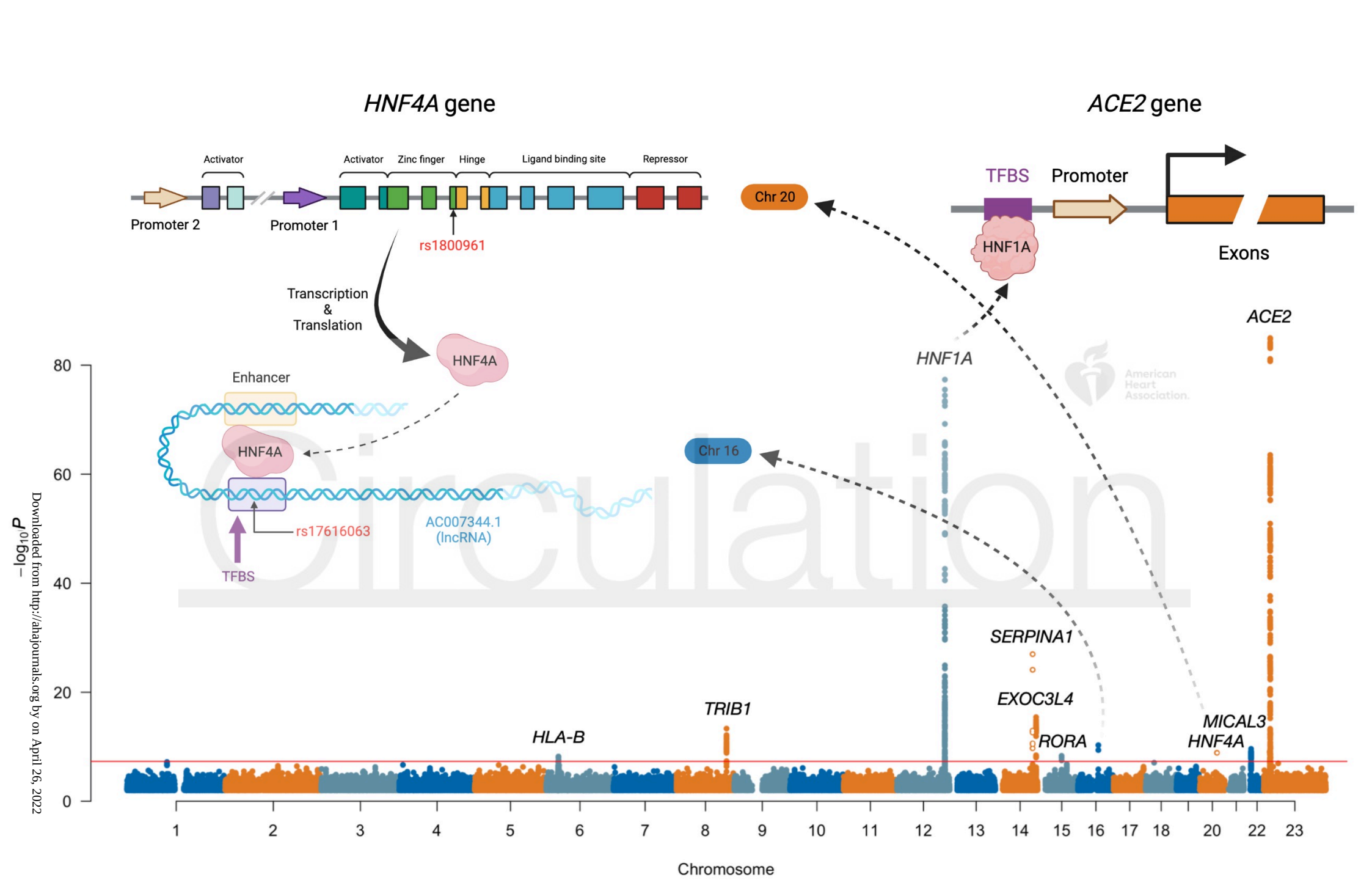
Chr.: Chromosome. R.A.: Reference allele. E.A.: Effective allele. Freq.: Frequency of the effective allele. s.e.: Standard error.
*: also known as p.Ile27Leu, where the Leu allele corresponds to the C allele.
†: also known as p.Glu342Lys, where the Lys allele corresponds to the T allele.

Figure 1. Genomic meta-analysis scan of plasma ACE2. Mapped genes are labeled at genome-wide significant loci ($P < 5 \times 10^{-8}$). Genome-wide significant variants with minor allele frequency < 0.05 are marked as circles instead of solid dots. Illustrations are provided for the interactions between two pairs of mapped loci, where the locus on chromosome 16 is a transcription binding site for the transcription factor HNF4A mapped on chromosome 20, and HNF1A acts as the transcription factor for the *ACE2* gene. TFBS: transcription factor binding site.

Figure 2. Genetic correlations between plasma ACE2 and human complex traits and diseases. A, Statistically significant (false discovery rate < 5%) genetic correlations with ACE2 are shown, where severe COVID-19, C-reactive protein (CRP), as well as other representative traits are labeled. The error bars represent standard errors. The colors label different groups of phenotypes. SBP: systolic blood pressure, BMI: body mass index, FVC: forced vital capacity. Detailed explanations of the annotated phenotypes are in the Supplemental Material. B, Enrichment of genetic correlations with ACE2 within each group of phenotypes. The circles are the quantile-quantile (QQ) plots of the genetic correlations test statistics against the null, whereas the solid dots are the QQ plots of the test statistics within each phenotype group against all the analyzed phenotypes.

Figure 3. Genetic and causal relationships between plasma ACE2 and vascular diseases. Estimates significantly different from zero are highlighted in filled circles. The first two forest plots show the bidirectional generalized summary-level Mendelian randomization (GSMR) analysis results between plasma ACE2 and 48 vascular-disease-related traits. The third forest plot gives the corresponding genetic correlations estimates between plasma ACE2 and these phenotypes. The last forest plot shows the estimated MR effects based on cis-pQTL only. OR: odds ratio.

Figure 4. Causal inference between plasma ACE2 and COVID-19 based on the *ACE2* cis-pQTL. A, The regional GWAS Z-scores across four traits are compared, where the alleles are coded so that the estimated SNP effects on ACE2 are all positive. Genome-wide significant SNPs for ACE2 ($P < 5 \times 10^{-8}$) are highlighted in yellow. The three SNPs representing independent significant associations after LD clumping ($r^2 < 0.001$) are marked in red. B, Inference of ACE2's causal effect on COVID-19 via Mendelian randomization (MR). The MR was performed using an inverse-variance weighted causal effect estimator, based on multiple genome-wide significant cis-regulatory SNPs. A threshold of $R^2 < 0.001$ was applied to prune out SNPs in LD. OR: odds ratio. The whiskers represent 95% confidence intervals.



A

Genetic correlation

0.75
0.5
0.25
0
−0.25
−0.5
−0.75
−1

Wheeze or whistling in the chest
CRP
Severe COVID-19
Number of treatments/ medications taken
Waist hip ratio
Waist circumference
Major depressive disorder
Current tobacco smoking
Education/work
SBP
Hypertension
Blood pressure medication
Amlodipine
Leg fat
BMI
Arm fat
None vascular/heart problems diagnosed by doctor
Parents' age at death
None blood pressure medication
FVC
Usual walking pace
Bipolar disorder
Red wine (weekly)
Years of schooling
Fluid intelligence
Alcohol with meals
Qualifications
Mental
Cognitive function
Physical measures/activities
Lifestyle/behavior
Health conditions
Medication/treatment

B

Observed χ^2
Theoretical χ^2

Physical measures/activities
Medication/treatment
Health conditions

Education/work
Lifestyle/behavior
Mental
Cognitive function

Metabolite
Blood lipids
Other