Title

Human genetic diversity possibly determines human's compliance to COVID-19

Yiqiang Zhao[1*], Yongji Liu[2], Sa Li[1], Yuzhan Wang[1], Lina Bu[1], Qi Liu[1], Hongyi Lv[2], Fang Wan[2], Lida Wu[2], Zeming Ning[3], Yuchun Gu[2,4*]

1. Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing, China
2. Allife Medicine Ltd., Beijing, China
3. The Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK
4. Regenerative Medicine Center, Aston University, Birmingham, UK

Corresponding Author
Prof. Yiqiang Zhao
Beijing Advanced Innovation Center for Food Nutrition and Human Health
China Agricultural University
Beijing, China
Email: yiqiangz@cau.edu.cn

And
Prof. Yuchun Gu
The regenerative Medicine Center
Aston University
Birmingham, UK
&
Allife Medicine Co.Ltd
Beijing, China
Email: y.gu1@aston.ac.uk

Abstract

The rapid spread of the coronavirus disease 2019 (COVID-19) is a serious threat to public health systems globally and is subsequently, a cause of anxiety and panic within human society. Understanding the mechanisms and reducing the chances of having severe symptoms from COVID-19 will play an essential role in treating the disease, and become an urgent task to calm the panic. However, the COVID-19 test developed to identify virus carriers is unable to predict symptom development in individuals upon infection. Experiences from other plagues in human history and COVID-19 statistics suggest that genetic factors may determine the compliance with the virus, i.e., severe, mild, and asymptomatic. Here, a hypothesis is put forward based on the epidemiological characteristics and traits of COVID-19, and our gene expression analysis. It proposes that COVID-19 inactivation in the blood by blocking virus entry into other internal organs for reproduction through the blood circulation after lung cell invasion prevents severe symptoms. Additionally, we investigated a genetic connection between candidate genes and severe COVID-19 symptoms through the utilization of strategies combining hypothesis and data-driven approaches. A list of genes and important SNPs that require further investigation to aid the screening of individuals who may suffer severe illness if exposed to the virus is present. Those individuals should be intensively safeguarded and prioritised for treatment. Concurrently to further research on the COVID-19 pathogenesis, our results also offer a new research strategy for pandemic prevention and health maintenance.

Introduction

Within a few months of the first report, COVID-19, a virus causing serious respiratory illnesses such as pneumonia and lung failure, had spread worldwide and reached a pandemic level [1-3]. As no specific therapeutics and vaccines are available for COVID-19 disease control, the virus has triggered enormous human casualties, serious global economic loss and the emergence of panic worldwide. A COVID-19 test has been rapidly developed and applied publicly within communities and clinics to identify COVID-19 carriers. Identified individuals are required to self-quarantine or accept treatment in hospitals despite most being asymptomatic. However, panic induced by COVID-19 is still present and worsening as it is certain that those lacking appropriate protection will become carriers of the virus. Furthermore, virus detection does not predict symptom development of individuals upon infection.

A proportion of virus infected individuals exhibit a progression of dyspnea, high respiratory frequency, low oxygen saturation and severe lung infiltration, leading to some entering critical condition (respiratory failure, septic shock and / or multiple organ dysfunction / failure) [4-6]. However, over millions of people only possess light symptoms such as a dry cough, nasal congestion, a runny nose, a sore throat, myalgia and diarrhea, despite carrying considerable titers of COVID-19 [7, 8]. Although mild symptomatic and asymptomatic individuals will not proceed to severe conditions, it is still difficult to distinguish the definite ascription of infected people at this stage.

It reminds us that human beings have experienced several major pandemic diseases such as the historical black fever and Spanish flu by coincidence. Whilst studying and recognizing these deadly pathogens, human beings should notice that the human body itself instinctively resists the invasion and dissemination of these pathogens, manifested within those asymptomatic or light symptomatic individuals who are, therefore, able to survive the plagues. This natural ability is determined by genetic composition, but due to human genetic diversity, does not apply to everyone. The diversity of human genes determines the compliance with the virus, i.e. severe, mild and asymptomatic. This kind of individual selection occurs every time an epidemic/pandemic breaks out, thus, in order to exploit the genetic variance of COVID-19, it is necessary to review human genetic diversity to seek the pathogen resistance genetic keys through the use of genetic algorithms and gene database.

There are many additional risk factors for the progression in individuals to severe and critical condition including obesity [9], COPD[10-12], IPF, chronic heavy smoking, cardiovascular diseases [13-17] such as hypertension, cerebral infraction, chronic kidney diseases [18-20] and so on. These additional complications may cause confusion in initial proteomic analyzing of samples. Genetic algorithms, therefore, become the first step in distinguishing the candidates which request validation in sample studies.

In this study, a hypothesis is proposed that COVID-19 inactivation in the blood reduces severe symptoms. Focusing on severe COVID-19 symptoms that may be caused by genetic factors, we scanned the genome in two human populations, and reported candidate genes and

variations that may be associated with severe COVID-19 symptoms. Individuals who may suffer severe illness should be safeguarded and prioritised for treatment. Our results suggest different genetic factors are possibly contributing to the severe COVID-19 symptoms.

Result and Discussion

1. RNA-seq analysis suggests tissues potentially being infected by COVID-19

Virus entry pathways are largely defined by the interactions between virus particles and their receptors at the cell surface [21]. ACE2 [22], BSG and TMPRSS2 [23] expressed in the host cells are essential for COVID-19 invasion, with higher expression possibly indicating higher risk. By collecting 299 RNA-seq data from NCBI SRA database, including major human tissues under normal physiological conditions (Supplementary Table 1), we quantified the expression level of each gene in the genome using standard Hisat2 and Stringtie pipeline. ANOVA showed that these three genes were significantly deferentially expressed across tissues (Table 1). Contrary to expectations, ACE2 was not highly expressed in lung tissue but was relatively highly expressed in the kidneys, the testis, and the heart. TMPRSS2 was highly expressed in the testes and lungs, but not in most other tissues. The overall expression level of the BSG gene was the highest. It was highly expressed in tissues other than the liver and muscle tissues (Table 1).

Tissues with high expressions of these genes might facilitate the entry of COVID-19 and, in turn, cause virus replication and tissue damage. In support, clinical symptoms regarding the heart, kidneys, and testicles have been reported. Since the BSG gene is highly expressed in multiple tissues, we should also pay attention to possible damage to tissues such as the ovaries, fat, the brain, and the spleen. In addition, Levene's tests showed that the expression of BSG and ACE2 varied differently in different tissues (Table 1), suggesting considerable variations in the degree of tissue damages may occur across individuals. Interestingly, all three genes are highly expressed in the testis. This may also explain why the proportion of male patients is higher than that of females, due to the high expression of ACE2 in the testis possibly facilitating viral entry, replication, and accumulation of a large amount of virus.

2. Screening of mutations potentially associated with severe COVID-19 symptoms

A piece of important information given by current clinical reports is that most patients with COVID-19 have mild to moderate symptoms and can recover without treatment. The ratio of individuals suffering severe symptoms, requiring admission to intensive care units, is about 20-30% of the population. Severe symptoms are often reported in older adults, but there is still a low proportion of young people who become seriously ill. We, thus, believe that the incidence of severe symptoms may be connected to genetic (inherited) mutations, DNA methylation-induced somatic mutations, and dysregulation of gene expression due to aging or other factors. The severe symptoms of young may be more attributed to genetic factors, that is, germline defects of coding or regulatory sequence in the genome. In contrast, the severe symptoms of the elderly are more likely related to protein or gene dysfunction during aging. Patients who meet either of these two conditions may develop severe symptoms after

COVID-19 infection, and these two effects may be additive in the elderly. Screening individuals who may develop severe symptoms in advance has important guiding significance for targeted prevention and effective treatment.

At present, there is no literature so far reporting the proportions of severe symptoms in young and elderly separately. A recent review paper on the epidemiological characteristics of COVID-19 in China announced a death toll of 44,672 confirmed patients at different ages [24]. The overall mortality rate of COVID-19 in China is 2.3% whilst the mortality rate for people under 40 is 0.2%. Since we already know that the ratio of patients experiencing severe symptoms is about 20-30% of the population, and assuming that a fixed percentage of the patients with severe symptoms will eventually die, it can be estimated that the severe symptoms caused by genetic factors (i.e., severe symptoms for young people under 40) is about 2-3%. This information reminds us that we can further screen for genes that are related to severe symptoms by the frequency of 2-3% in each variant locus in the population.

Two population data sets were used to explore the frequency of each locus in the genome. One was the 11,670 whole-genome sequences of the female Han Chinese from the CONVERGE project. The other one was the widely used 1000 genome project phase 3 data, which includes 2,504 individuals with different genetic backgrounds. For these two sets, we calculated the allele frequency and genotype frequency of each site recorded in the VCF, in the respective populations. Based on the estimation that the severe symptoms caused by genetic factors is 2-3%, we first scanned the genome for loci whose allele frequency or genotype frequency ranged between 1-5%. Considering nucleotide alterations that change amino acids are more functionally meaningful, we limited loci to those which are also nonsynonymous. It produced 13,712 loci from 7,291 genes in the CONVERGE project and 42,221 loci from 12,815 genes in the 1000 genome project. If we narrowed down the candidates by restricting the frequency ranges between 2-3%, it produced 4,327 loci from 3,109 genes in the CONVERGE project and 13,545 loci from 7,321 genes in the 1000 genome project.

It is worth noting that although the frequency screening is for severe symptoms caused by genetic factors, it also contributes to the understanding of how aging causes these severe symptoms. This is because genes with genetic alternations leading to functional defects in the young may also be inactivated via methylation-induced somatic mutations during the aging. We checked whether the nonsynonymous loci with the expected frequency could also be potential methylation sites. Using the Infinium MethylationEPIC human 850k BeadChip annotations, we found these variations ranged from dozens to a few hundred in the two populations (Table 2, Supplementary Table 2).

3. The hypothesis of COVID-19 inactivation in the blood to prevent severe symptoms

COVID-19 enters the body through the mouth, eyes or nose, moves to the throat and reaches the lungs. Due to the expression of ACE2, known as the main receptor of COVID-19, not being high in the lungs, pneumonia caused by COVID-19 may rely on virus replication in other tissues with high expressions of ACE2. This subsequently leads to the release of a large

quantity of the virus, which again invade lung tissue and bring about reduced lung function and severe hypoxia. Since most internal organs are not directly exposed to the virus, the accumulation and transportation of large amounts of virus from other tissues to the lungs require blood as a transport medium for COVID-19, i.e., the viruses pass from tissues into blood and vice versa.

It has been reported that the envelope protein of some viruses may be inactivated by human serum complements [25, 26]. If this is also the case for COVID-19, the "replication-after-transmission" process of the virus can be prevented, and in turn, the symptoms can be minimized. Conversely, if the inactivation of the COVID-19 envelope protein is not successful in the blood, the virus may be more easily transmitted to other organs through the blood, and damage the kidneys, testis, heart, etc. Once a high level of virus replication occurs in these tissues, the virus would be released and passed from the blood into the lungs again, resulting in an aggravation of lung symptoms (Figure 1). Therefore, a hypothesis is made here that the failure to inactivate the COVID-19 envelope protein in the blood may be one of the possible causes of the development of severe symptoms. Under this mechanism, the appearance of asymptomatic carriers in the population can be explained; the viruses they carry are not enriched in the lungs and other organs. However, the virus may return to infect the lungs at any time, making the condition suddenly worsen in respect with the cases of high titer.

Intuitively, the substance that inactivates the COVID-19 envelope protein in the blood may be immune-associated proteins. This is consistent with the fact that there are significantly more severe symptoms and death in the elderly, since the immune function gradually decreases with age, irrespective of underlying medical conditions. However, there is currently no research on this topic yet.

4. Screening of candidate proteins for COVID-19 inactivation or aging marker proteins in plasma

Based on the above hypothesis and the observation of a higher proportion of severe COVID-19 symptoms in the elderly, we next explored the proteins in the blood that are relatively stable in the young but change significantly in the elderly.

Lehallier et al. [27] studied the changes in human plasma proteome across a lifespan. The author reported lists of proteins that changes significantly in expression in three age intervals: 34 ± 10, 60 ± 10, and 78 ± 10. Based on these three lists, we generated a list of unique proteins that are changed significantly in ages 78 ± 10 by subtracting proteins found in the other two lists. The author also released quantitative expressions of 1305 plasma proteins for 171 individuals of different ages from 21 to 107 years old. Amongst these 1305 proteins, we found that 790 proteins decrease in expression with increasing age between 60 and 85 (Spearman rho<0, regardless of the p-value), 439 of these 790 down-regulated proteins did not show a tendency to decrease before age 45 (Spearman rho>0, irrespective of the p-value). If we set a loose p-value for the elderly, such as 0.1, 142 proteins were obtained that are not down-regulated before age 45 but do decrease between 60 and 85 (Table 3). In the opposite

case, the protein expression is up-regulated between 60 and 85 whilst showing no tendency to increase before age 45. The numbers of proteins become 515, 212, and 57 respectively, which are much smaller in size (Table 3). It seems that proteins decreasing in expression between 60 to 85, but not before 45 is more biologically relevant.

Due to the higher proportion of severe COVID-19 symptom appearing in the elderly, we speculate that the protein list with a significant change in the 78±10 age interval in addition to the protein we identified (142 + 57) may itself contain functional proteins related to the inactivation of the viral envelope protein (or inhibitors of the functional proteins), and therefore deserve further investigation (Figure 2). Interestingly, the protein expression of ACE2 and BSG is not significantly associated with age (R-square 0.069 and 0.036, respectively) which suggests that the higher proportion of severe COVID-19 symptoms in the elderly cannot be simply explained by the increased expression of COVID-19 receptors.

On the other hand, even functional proteins directly responsible for the inactivation of the COVID-19 envelope protein are not present in our list since the data available here is far from complete. However, we can still use the proteins that are relatively stable in the young but change significantly in the elderly identified here as markers for aging, and help to screen people who are likely to develop severe symptoms (Supplementary Table 3 and 5).

5. Screening of candidate genes for COVID-19 infection, morbidity, and clearance

It has been reported that ACE2, BSG, and TMPRSS2 are essential for COVID-19 infection. Conversely, among clinical manifestations chest tightness, pulmonary effusion, and increased phlegm are listed, which are similar to the symptoms of cystic fibrosis. We thus speculated that CFTR and ENaC channel associated genes might be related to the occurrence and development of these symptoms. Using ACE2, BSG, DPP4, TMPRSS2, CFTR, SCNN1A, SCNN1B, SCNN1G, SCNN1D, ACCN1, ACCN2, ACCN3, and ACCN4 as seed genes, we searched for other genes associated to the seed genes using GeneMANIA, with respect to genetic interactions, pathways, co-expression, co-localization and protein domain similarity. To retrieve more comprehensive information, we used the genes outputted by GeneMANIA as seeds again to extend the candidate gene list further. After three iterations, a total of 74 candidate genes were obtained, of which we believe are possibly related to COVID-19 infection and morbidity.

Considering that previous studies have shown that immune-related genes such as IFITM-, IL-, TLR-, IFNA-family genes, and complements are related to the viral infection and inactivation, we manually collected genes from the above-mentioned immune-related gene family. These genes were also used as seed input of GeneMANIA to retrieve functionally similar genes. We believe that these genes might constitute genes related to COVID-19 clearance and inhibition.

Finally, we merged the above two gene lists to obtain 173 non-redundant genes (Supplementary Table 4 and 5). Although limited by our knowledge, it is not possible to cover all candidate genes, we could expand candidates for COVID-19 infection, morbidity and clearance at any time according to the updates of our knowledge.

6. Functional summaries of candidate genes containing candidate variations

For all the candidate genes we collected above, we extracted variations within the genes and the regulatory eQTL for the genes. Variations within the genes were extracted according to gene coordinates and refSNP coordinates. For the eQTLs corresponding to these genes, we used the eQTL information recorded in the SCAN and ExSNP databases and extracted the eQTL records matching the candidate genes (see Materials and Methods). Due to the large volume of nonsynonymous loci with the expected frequency, we kept those that are annotated as deleterious by at least one of the SIFT or FATHMM predictions (Table 4 and 5, Supplementary Table 6 and 7).

By reviewing the function of each candidate variation, we found that the corresponding genes could be categorized into several groups: serum complement including C2; histocompatibility complex including HLA-A, HLA-DPB1, MICA; immune response molecule including SIGGIRR, TLR3, TLR5, SECTM1, MST1R, CRTAM, LILRB1, CD58, NCR3, GLIPR1, CD40; cytokines including IL12B, IL1B, IL17RC, LTBR, CCL23, CCL16; inflammation factors including SERPIND1; adhesive glycoproteins that mediate cell-to-cell and cell-to-matrix interactions and extracellular matrix association including THBS4, TNXB, CDH15, CD38, MMP3, MMP10, SCIN; transmembrane receptor proteins including TAP1, STAB1; ion channels including TRPM8, SCN7A, BEST3; endoplasmic reticulum including FAF2, ERO1L; DNA/RNA enzymes and binding motif including PR1M1, RMB19, RNABP17, OAS1, BUB1; mitochondria association including SIRT3, NQ01; ACE related including AGT, AGTR2.

The COVID-19 envelope protein is highly glycosylated. The virus enters the respiratory tract, makes contact with epithelial cells, and enters cells via ACE2, BSG, and TMPRSS2 on the cell membrane by endocytosis. THBS4 and TNXB may be highly associated with these events. RNABP17 is also suspected of playing a role as the transport of large RNAs through the nuclear pore complexes. As mentioned above, serum complement C2 is one of the candidates to inactivate the envelop protein of the virus, as well as the typical innate immune response to the viral invasion including SIGGIRR, TLR, and MST1R. In coincidence with the hypoesthesia of smell of COVID-19 patients [28], OR10AD1, which associates with smell, were selected due to nonsynonymous eQTL mutation, as well as ALPK3, which associates with hypertrophic heart cardiomyopathy [29, 30].

Conclusion

In scenario, genetic diversity plays a prime role in the progression of severe symptoms. We propose a hypothesis that the inactivation of COVID-19 in the blood of some people helps these individuals to avoid severe symptoms. Examples of how to find candidate genes and loci related to severe COVID-19 symptoms caused by genetic factors are also given. On the basis of more accurate statistical data of patients with severe symptoms of different ages and genders, further generic algorithms can more accurately predict SNPs which determine whether individuals will suffer from severe symptoms. In addition to vaccines and antiviral compounds, it may be more direct and effective to find countermeasures to cope with the threat of COVID-19 from the body itself, and to fight against the virus by completing and

improving the function of defense genes. After obtaining these predicted SNPs, they can be fully verified in samples of severe, mild and asymptomatic patients. These SNPs can be used to identify defective individuals in the population. This may be a better method of eliminating fear, anxiety and panic, as virus testing cannot predict the progress of infection. At the same time, the defective proteins that target the invasion of COVID-19 can be used as a new strategy to solve the serious diseases and casualties caused by COVID-19.

Acknowledgements
The project was supported by the National Special Foundation for Transgenic Species of China (2018ZX08007001) and Key-Area Research and Development Program of Guangdong Province(2018B020203001). We thank the support of the GENEGOALS platform and the high-performance computing platform of the State Key Laboratory of Agrobiotechnology. We thank Mrs. Yin Cao and Miss Lauren Gu for their effort in editing.

Materials and Methods

RNA-seq data

RNA-seq sequencing data for humans were downloaded from the NCBI SRA database. It comprises 299 data sets from 10 tissues (adipose, brain, heart, kidney, liver, lung, muscle, ovary, spleen, testis) under normal physiological conditions.

To estimate the expression level of genes in the genome, we firstly converted SRA to FASTQ sequences using the fastq-dump command. TrimGalore-0.6.2 was used to perform the quality filtering of the FASTQ sequences with default settings. The expression level for each gene was then estimated as FPKM value, using the standard procedure of Hisat2 ver2.1.0 [31] + Stringtie ver1.3.5 [32] pipeline. Genes with expression values less than 0.1 and outliers with more than three standard deviations away from the mean according to the first three principal components were removed. The ComBat method was finally used to remove the batch effect to obtain corrected expression values.

Protein expression data in human plasma

The protein expression data in human plasma came from the supplementary tables of the Nature Medicine paper by Lehallier et al. [27]. The list of proteins with significant changes in expression in 34±10, 60±10, and 78±10 age interval was obtained supplementary table 14. Only proteins with a p-value of less than 0.05 were kept. The quantitation of 1,305 proteins expressed in plasma from 171 individuals of different ages was directly obtained from supplementary table 16, without additional processing.

eQTL data

Human eQTL records were downloaded from SCAN (http://www.scandb.org/newinterface/downloadannots.html) and ExSNP [33] (http://www.exsnp.org/Download) databases. All eQTL information downloaded were

merged to produce a total of 2,239,338 non-redundant paired records, corresponding to 1,074,528 eQTL sites from 23,065 genes.

Gene and variant annotation

The genomic coordinates of candidate genes were retrieved from the UCSC Genome Browser Table Browser [34]. For each gene, we extended 2kb to the 5 'end and 500bp to the 3' end according to the exon boundaries to include potential regulatory regions. ANNOVAR [35] was used to perform gene-based and filter-based annotation for variants, including Func.refGene, Gene.refGene, ExonicFunc.refGene, SIFT_pred, FATHMM_pred. R package clusterProfiler [36] was used for functional annotations for genes and GO enrichment analysis for the gene list.

Allele and genotype frequency for variants

The 1000genome project phase 3 data [37] in VCF format was downloaded from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/, which includes 2,504 individuals sequenced at 30x coverage. We also downloaded the phased variant data in VCF format from the CONVERGE project [38] from https://www.ebi.ac.uk/ena/data/view/PRJNA289433, which comprises 11,670 female Han Chinese sequenced at 1.7x coverage. Both data sets are provided in phased VCF format. For each biallelic site in 1-22 and X chromosome, we calculated the two alleles and three genotypes frequencies in the two populations.

GeneMANIA analysis

GeneMANIA [39] finds other genes that are related to a set of input genes using an extensive collection of functional association data, including protein and genetic interactions, pathways, co-expression, co-localization, and protein domain similarity. We used the GeneMANIA web service https://genemania.org/ to extend the candidate gene list with functionally similar genes.

Reference

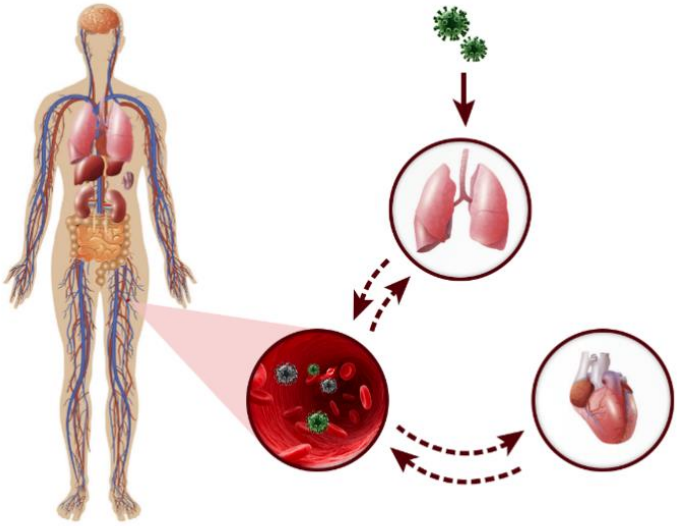
1. Lurie, N., et al., *Developing Covid-19 Vaccines at Pandemic Speed.* N Engl J Med, 2020.
2. Hua, J. and R. Shaw, *Corona Virus (COVID-19) "Infodemic" and Emerging Issues through a Data Lens: The Case of China.* Int J Environ Res Public Health, 2020. **17**(7).
3. Keesara, S., A. Jonas, and K. Schulman, *Covid-19 and Health Care's Digital Revolution.* N Engl J Med, 2020.
4. Henry, B.M. and J. Vikse, *Clinical Characteristics of Covid-19 in China.* N Engl J Med, 2020. **382**.
5. Chen, A.T.C., G.B. Coura-Filho, and M.H.H. Rehder, *Clinical Characteristics of Covid-19 in China.* N Engl J Med, 2020. **382**.
6. Zavascki, A.P. and D.R. Falci, *Clinical Characteristics of Covid-19 in China.* N Engl J Med, 2020. **382**.
7. Sivakorn, C., et al., *Case Report: Walking Pneumonia in Novel Coronavirus Disease (COVID-19): Mild Symptoms with Marked Abnormalities on Chest Imaging.* Am J Trop Med Hyg, 2020.

8. Mash, B., *Primary care management of the coronavirus (COVID-19).* S Afr Fam Pract (2004), 2020. **62**(1): p. e1-e4.
9. Muniyappa, R. and S. Gubbi, *COVID-19 Pandemic, Corona Viruses, and Diabetes Mellitus.* Am J Physiol Endocrinol Metab, 2020.
10. Lian, J., et al., *Analysis of Epidemiological and Clinical features in older patients with Corona Virus Disease 2019 (COVID-19) out of Wuhan.* Clin Infect Dis, 2020.
11. Emami, A., et al., *Prevalence of Underlying Diseases in Hospitalized Patients with COVID-19: a Systematic Review and Meta-Analysis.* Arch Acad Emerg Med, 2020. **8**(1): p. e35.
12. Brake, S.J., et al., *Smoking Upregulates Angiotensin-Converting Enzyme-2 Receptor: A Potential Adhesion Site for Novel Coronavirus SARS-CoV-2 (Covid-19).* J Clin Med, 2020. **9**(3).
13. Yang, C. and Z. Jin, *An Acute Respiratory Infection Runs Into the Most Common Noncommunicable Epidemic-COVID-19 and Cardiovascular Diseases.* JAMA Cardiol, 2020.
14. Tan, W. and J. Aboulhosn, *The cardiovascular burden of coronavirus disease 2019 (COVID-19) with a focus on congenital heart disease.* Int J Cardiol, 2020.
15. Bansal, M., *Cardiovascular disease and COVID-19.* Diabetes Metab Syndr, 2020. **14**(3): p. 247-250.
16. Guo, T., et al., *Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19).* JAMA Cardiol, 2020.
17. Hulot, J.S., *COVID-19 in patients with cardiovascular diseases.* Arch Cardiovasc Dis, 2020.
18. Henry, B.M. and G. Lippi, *Chronic kidney disease is associated with severe coronavirus disease 2019 (COVID-19) infection.* Int Urol Nephrol, 2020.
19. Chen, T., et al., *Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study.* BMJ, 2020. **368**: p. m1091.
20. South, A.M., D. Diz, and M.C. Chappell, *COVID-19, ACE2 and the Cardiovascular Consequences.* Am J Physiol Heart Circ Physiol, 2020.
21. Grove, J. and M. Marsh, *The cell biology of receptor-mediated virus entry.* J Cell Biol, 2011. **195**(7): p. 1071-82.
22. Ou, X., et al., *Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV.* Nat Commun, 2020. **11**(1): p. 1620.
23. Lukassen, S., et al., *SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells.* EMBO J, 2020.
24. Novel Coronavirus Pneumonia Emergency Response Epidemiology, T., *[The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China].* Zhonghua Liu Xing Bing Xue Za Zhi, 2020. **41**(2): p. 145-151.
25. Schauber-Plewa, C., et al., *Complement regulatory proteins are incorporated into lentiviral vectors and protect particles against complement inactivation.* Gene Ther, 2005. **12**(3): p. 238-45.
26. DePolo, N.J., et al., *VSV-G pseudotyped lentiviral vector particles produced in human cells are inactivated by human serum.* Mol Ther, 2000. **2**(3): p. 218-22.

27. Lehallier, B., et al., *Undulating changes in human plasma proteome profiles across the lifespan.* Nat Med, 2019. **25**(12): p. 1843-1850.
28. Gautier, J.F. and Y. Ravussin, *A New Symptom of COVID-19: Loss of Taste and Smell.* Obesity (Silver Spring), 2020.
29. Aung, N., et al., *Genome-Wide Analysis of Left Ventricular Image-Derived Phenotypes Identifies Fourteen Loci Associated With Cardiac Morphogenesis and Heart Failure Development.* Circulation, 2019. **140**(16): p. 1318-1330.
30. Al Senaidi, K., et al., *Phenotypic spectrum of ALPK3-related cardiomyopathy.* Am J Med Genet A, 2019. **179**(7): p. 1235-1240.
31. Kim, D., B. Langmead, and S.L. Salzberg, *HISAT: a fast spliced aligner with low memory requirements.* Nat Methods, 2015. **12**(4): p. 357-60.
32. Pertea, M., et al., *StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.* Nat Biotechnol, 2015. **33**(3): p. 290-5.
33. Yu, C.H., L.R. Pal, and J. Moult, *Consensus Genome-Wide Expression Quantitative Trait Loci and Their Relationship with Human Complex Trait Disease.* OMICS, 2016. **20**(7): p. 400-14.
34. Karolchik, D., et al., *The UCSC Table Browser data retrieval tool.* Nucleic Acids Res, 2004. **32**(Database issue): p. D493-6.
35. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.
36. Yu, G., et al., *clusterProfiler: an R package for comparing biological themes among gene clusters.* OMICS, 2012. **16**(5): p. 284-7.
37. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
38. Cai, N., et al., *11,670 whole-genome sequences representative of the Han Chinese population from the CONVERGE project.* Sci Data, 2017. **4**: p. 170011.
39. Warde-Farley, D., et al., *The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W214-20.

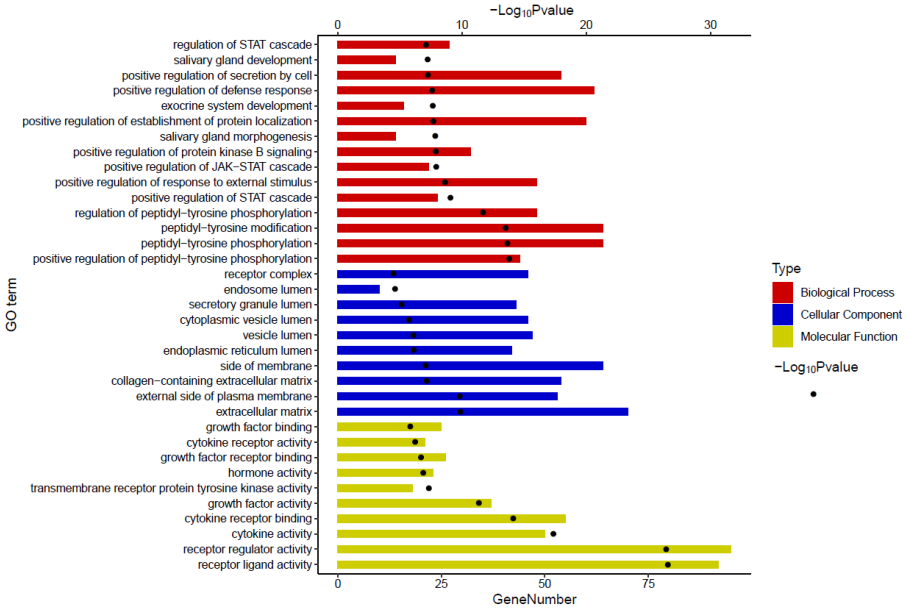
Figure and Figure Legends

Figure 1. A model of COVID-19 passes from tissues into the blood and passes from the blood into tissues



The dash lines present the secondary transmissions of COVID-19 between organs through the blood. Particles in gray indicate inactivated COVID-19 in plasma, while particles in green indicate intact COVID-19.

Figure 2. Functional enrichment of aging-associated proteins in plasma



Tables

Table 1. Expression of ACE2, BSG and TMPRSS2 across tissues

Gene	Statistics	Adipose	Brain	Heart	Kidney	Liver	Lung	Muscle	Ovary	Spleen	Testis	F value from ANOVA test	P-value	F value from Levene's test	P-value
ACE2	Mean	0.582	-	8.758	12.289	1.442	0.826	-	3.519	-	8.094	15.665	0.000		
ACE2	Standard deviation	0.290	-	2.855	1.803	1.008	0.359	-	3.561	-	8.052			6.319	0.000
BSG	Mean	196.266	114.523	408.696	248.148	53.430	148.512	29.377	115.300	107.690	250.406	8.682	0.000		
BSG	Standard deviation	137.358	23.340	74.744	39.235	16.600	23.837	13.663	35.254	10.264	121.820			5.671	0.000
TMPRSS2	Mean	-	-	-	16.134	13.741	34.173	-	-	-	41.957	30.671	0.000		
TMPRSS2	Standard deviation	-	-	-	3.393	9.485	9.105	-	-	-	24.115			2.072	0.107

- indicates the expression is less than 0.1

Table 2. Nonsynonymous and potential methylation loci with allele or genotype frequency of 2-3% in two populations

Gene	Chr	Pos	SNP	Ref	Alt	Freq	Population
ABITRAM	9	111696795	rs76817627	C	T	CT 0.0224	1000 Genome
ACOXL	2	111562875	rs77331476	C	T	CT 0.0208	1000 Genome
ADM5	19	50193437	rs45613034	C	A	CA 0.0200	1000 Genome
ALDH4A1	1	19211987	rs113846237	C	T	CT 0.0240	1000 Genome
ANKS1B	12	100169425	rs116525095	C	T	CT 0.0220	1000 Genome
AP1G2	14	24029037	rs74849041	C	T	CT 0.0216	1000 Genome
ARHGAP40	20	37272380	rs61737953	C	T	CT 0.0272	1000 Genome
ARSD	X	2825596	rs2229557	C	T	CT 0.0212	1000 Genome
AXDND1	1	179460806	rs17369441	C	T	T 0.0216	1000 Genome
BTNL2	6	32370879	rs28362681	C	T	TT 0.0204	1000 Genome
BUB1B	15	40457337	rs56079734	C	T	T 0.0232	1000 Genome
C9orf135	9	72459437	rs55859531	C	T	T 0.0242	1000 Genome
CAGE1	6	7374210	rs142828071	C	T	CT 0.0216	1000 Genome
CCDC114	19	48821738	rs58966182	C	G	CG 0.0200	1000 Genome
CEP164	11	117279712	rs61995733	C	T	T 0.0290	1000 Genome
CORIN	4	47625753	rs75141391	C	T	CT 0.0216	1000 Genome
CRYBG3	3	97594786	rs79581944	C	T	T 0.0260	1000 Genome
CRYGC	2	208994274	rs61751949	C	T	CT 0.0236	1000 Genome
CXXC4	4	105412536	rs552060137	C	G	CG 0.0212	1000 Genome
DNAH17	17	76571121	rs61744339	C	T	CT 0.0264	1000 Genome
EPCAM	2	47604176	rs74531854	C	T	CT 0.0268	1000 Genome
ERAP2	5	96228072	rs75263594	C	T	CT 0.0224	1000 Genome
ERP27	12	15073952	rs35042193	C	T	CT 0.0252	1000 Genome
FGL1	8	17739639	rs78484373	C	A	A 0.0212	1000 Genome
FOLR3	11	71850130	rs61734430	C	T	T 0.0218	1000 Genome
GPR35	2	241569692	rs3749171	C	T	TT 0.0248	1000 Genome
GPR63	6	97247547	rs35358396	C	T	CT 0.0276	1000 Genome
HIPK4	19	40886742	rs56365273	C	T	T 0.0238	1000 Genome
HLA-DMA	6	32917411	rs41555121	C	T	T 0.0214	1000 Genome
HLA-DPA1	6	33037060	rs115722167	C	T	T 0.0240	1000 Genome
IFNL2	19	39759335	rs554971	C	T	T 0.0264	1000 Genome
ITGAL	16	30492823	rs1064524	C	T	CT 0.0232	1000 Genome
IYD	6	150716612	rs73617694	C	T	CT 0.0268	1000 Genome
KIAA1210	X	118284520	rs142034189	C	A	A 0.0252	1000 Genome
KRT14	17	39739524	rs59780231	C	T	T 0.0278	1000 Genome
KRT7	12	52636874	rs7963792	C	T	CT 0.0224	1000 Genome
KRTAP10-12	21	46117285	rs201215667	C	T	CT 0.0200	1000 Genome
KRTAP4-11	17	39274491	rs113376601	C	T	TT 0.0200	1000 Genome
LRG1	19	4540003	rs115514161	C	T	CT 0.0220	1000 Genome
MAEA	4	1303499	rs116098348	C	T	CT 0.0292	1000 Genome

MAP7D1	1	36636835	rs2296266	C	T	CT 0.0272	1000 Genome
MCCD1	6	31496949	rs78957773	C	T	CT 0.0300	1000 Genome
MICA	6	31382831	rs41546114	C	T	CT 0.0264	1000 Genome
MIIP	1	12089946	rs117146839	C	G	CG 0.0244	1000 Genome
MIP	12	56848079	rs74641138	C	T	T 0.0220	1000 Genome
MROH7	1	55119654	rs12074499	C	T	T 0.0282	1000 Genome
MSH5	6	31725978	rs28399976	C	G	CG 0.0296	1000 Genome
MUC5B	11	1270361	rs201532622	C	T	CT 0.0288	1000 Genome
NACAD	7	45125663	rs73692203	C	T	CT 0.0232	1000 Genome
ODF3L2	19	464080	rs76592524	C	T	T 0.0282	1000 Genome
OR10AD1	12	48596708	rs114759358	C	T	T 0.0212	1000 Genome
OR1D5	17	2966828	rs531952330	C	T	T 0.0282	1000 Genome
OR1Q1	9	125377176	rs187659423	C	T	CT 0.0212	1000 Genome
OR2W3	1	248059252	rs61750779	C	T	TT 0.0276	1000 Genome
OSCP1	1	36887813	rs17442970	C	A	CA 0.0216	1000 Genome
PDLIM3	4	186435443	rs11944325	C	T	T 0.0286	1000 Genome
PEX2	8	77895682	rs112108739	C	T	T 0.0282	1000 Genome
PLCH1	3	155200562	rs74737650	C	T	T 0.0254	1000 Genome
PLEKHJ1	19	2230593	rs112397317	C	T	CT 0.0272	1000 Genome
POLQ	3	121202345	rs3218637	C	T	CT 0.0244	1000 Genome
POLRMT	19	621561	rs10421235	C	A	AA 0.0264	1000 Genome
PPFIBP2	11	7627239	rs76465494	C	T	T 0.0218	1000 Genome
PRAMEF2	1	12919079	rs9659529	C	G	GG 0.0200	1000 Genome
PRR14L	22	32109927	rs115688748	C	T	CT 0.0260	1000 Genome
PSD	10	104174738	rs150603372	C	T	CT 0.0248	1000 Genome
RAET1E	6	150211071	rs61747261	C	T	CT 0.0260	1000 Genome
REEP4	8	21996546	rs117397164	C	T	CT 0.0236	1000 Genome
REG1B	2	79313990	rs62640882	C	T	CT 0.0204	1000 Genome
RGS3	9	116356389	rs57845277	C	T	CT 0.0208	1000 Genome
RIMBP2	12	130935796	rs142989103	C	G	CG 0.0276	1000 Genome
RPL3L	16	1997004	rs113956264	C	T	T 0.0212	1000 Genome
SAPCD1	6	31731881	rs6905572	C	T	TT 0.0228	1000 Genome
SEMA4C	2	97527490	rs79331914	C	T	T 0.0256	1000 Genome
SH3PXD2B	5	171766566	rs62621449	C	T	CT 0.0248	1000 Genome
SIM2	21	38081480	rs116988298	C	T	CT 0.0280	1000 Genome
SLC29A4	7	5338736	rs150889269	C	T	CT 0.0216	1000 Genome
SNAPC2	19	7987428	rs116635738	C	T	CT 0.0216	1000 Genome
TACC2	10	123843946	rs60531929	C	T	T 0.0300	1000 Genome
TAP1	6	32821365	rs57640466	C	G	CG 0.0260	1000 Genome
TAP2	6	32800412	rs1800454	C	T	TT 0.0228	1000 Genome
TAP2	6	32800427	rs111303994	C	T	CT 0.0288	1000 Genome
TAP2	6	32802938	rs140654840	C	T	CT 0.0216	1000 Genome
TAPBP	6	33281576	rs117394742	C	T	CT 0.0204	1000 Genome
TBXT	6	166572045	rs3127328	C	T	TT 0.0280	1000 Genome

TECPR2	14	102900956	rs118141823	C	T	CT 0.0212	1000 Genome
TEP1	14	20850434	rs2229100	C	T	CT 0.0268	1000 Genome
TFRC	3	195791240	rs41295879	C	T	CT 0.0208	1000 Genome
TMEM240	1	1470807	rs146206869	C	T	CT 0.0264	1000 Genome
TPSG1	16	1272275	rs143120059	C	T	T 0.0250	1000 Genome
TRMU	22	46746188	rs35338668	C	T	CT 0.0276	1000 Genome
TRPM2	21	45820196	rs35288229	C	T	CT 0.0244	1000 Genome
TSPAN8	12	71533592	rs76495455	C	T	CT 0.0296	1000 Genome
TTLL10	1	1115461	rs116383664	C	T	CT 0.0220	1000 Genome
TUBB1	20	57599303	rs35565630	C	T	CT 0.0220	1000 Genome
UCP3	11	73718063	rs8179180	C	T	CT 0.0288	1000 Genome
UHRF1BP1	6	34826612	rs61732794	C	T	CT 0.0276	1000 Genome
USP36	17	76798460	rs77261246	C	T	CT 0.0300	1000 Genome
VSIG4	X	65241977	rs146489734	C	A	CA 0.0228	1000 Genome
WFS1	4	6303534	rs71530907	C	T	CT 0.0200	1000 Genome
ZDHHC11B	5	751299	rs112221486	C	T	T 0.0280	1000 Genome
ZFHX2	14	24002201	rs61744308	C	T	T 0.0280	1000 Genome
ZNF230	19	44514436	rs75657250	C	T	CT 0.0232	1000 Genome
ZNF747	16	30544067	rs115499189	C	T	CT 0.0232	1000 Genome
COL11A2	6	33132693	rs1799912	C	T	CT 0.0222	CONVERGE
GSDME	7	24756951	rs12540919	C	T	TT 0.0269	CONVERGE
H1FNT	12	48723324	rs2732441	C	G	CG 0.0208	CONVERGE
HLA-DMA	6	32917411	rs41555121	C	T	T 0.0206	CONVERGE
HLA-DPA1	6	33037060	rs115722167	C	T	CT 0.0291	CONVERGE
HSPA1L	6	31779728	rs141059651	C	G	CG 0.0216	CONVERGE
KIAA1210	X	118284520	rs142034189	C	A	CA 0.0273	CONVERGE
LMO7	13	76430665	rs75375399	C	T	CT 0.0206	CONVERGE
MEX3D	19	1556410	rs184017550	C	G	G 0.0225	CONVERGE
MICA	6	31382831	rs116789740	C	T	CT 0.0284	CONVERGE
MIP	12	56848079	rs74641138	C	T	CT 0.0239	CONVERGE
MRGPRX1	11	18956196	rs78179510	C	T	TT 0.0213	CONVERGE
OR2H1	6	29430334	rs140963208	C	T	CT 0.0290	CONVERGE
OXNAD1	3	16327909	rs6777976	C	T	TT 0.0285	CONVERGE
PASK	2	242077496	rs1470414	C	T	TT 0.0251	CONVERGE
POLRMT	19	621561	rs10421235	C	A	A 0.0223	CONVERGE
PRX	19	40902634	19:40902634	C	T	CT 0.0272	CONVERGE
SCGB1C1;SCGB1C2	11	193112	rs7951297	C	T	TT 0.0273	CONVERGE
SPTB	14	65260321	rs3742602	C	T	T 0.0213	CONVERGE
SRBD1	2	45640374	rs3755073	C	A	AA 0.0285	CONVERGE
TAP2	6	32800412	rs104893997	C	T	TT 0.0270	CONVERGE
TAPBP	6	33281576	rs117394742	C	T	CT 0.0202	CONVERGE
TECPR2	14	102918789	rs75004775	C	A	A 0.0230	CONVERGE
THNSL2	2	88478458	rs34250376	C	T	CT 0.0256	CONVERGE
TNXB	6	32064161	rs118086587	C	T	CT 0.0227	CONVERGE

TPSD1	16	1306346	rs3865205	C	T	TT 0.0290	CONVERGE
TRIM26	6	30166475	rs142601029	C	T	CT 0.0281	CONVERGE
TRIM31	6	30080405	rs143020056	C	T	CT 0.0292	CONVERGE
TTN	2	179403750	rs4894028	C	T	CT 0.0300	CONVERGE
WDR25	14	100847617	rs34331240	C	T	CT 0.0277	CONVERGE
WNT10A	2	219754840	rs116998555	C	T	CT 0.0263	CONVERGE
ZNF600	19	53269486	rs143951651	C	T	CT 0.0297	CONVERGE
ZSCAN10	16	3142911	rs11644978	C	G	GG 0.0211	CONVERGE

Chromosome position is based on hg37 for population CONVERGE

Chromosome position is based on hg38 for population 1000 genome

Table 3. Proteins showing distinct expression pattern between young and elderly in the plasma

Protein	Corr_elderly	P_elderly	Corr_young	P_young
ABL2	-0.195	0.095	0.068	0.770
ACY1	-0.204	0.082	0.167	0.470
ADAMTS15	-0.263	0.024	0.002	0.994
ADSL	-0.311	0.007	0.168	0.467
AGT	-0.364	0.001	0.025	0.915
ANGPT1	-0.205	0.080	0.050	0.828
ANK2	-0.259	0.026	0.306	0.177
APOA1	-0.406	0.000	0.277	0.224
APP	-0.218	0.062	0.007	0.978
BCL2A1	-0.195	0.096	0.439	0.047
BIRC5	-0.196	0.095	0.031	0.895
BTC	-0.321	0.005	0.021	0.929
C1QBP	-0.315	0.006	0.155	0.503
C1R	-0.274	0.018	0.189	0.411
C2	-0.324	0.005	0.217	0.344
C4A_C4B	-0.285	0.014	0.193	0.403
C5_1	-0.248	0.033	0.495	0.023
CAMK1D	-0.246	0.035	0.074	0.749
CASP10	-0.363	0.001	0.358	0.111
CBX5	-0.305	0.008	0.327	0.148
CCL24	-0.211	0.071	0.252	0.270
CD207	-0.193	0.100	0.314	0.165
CD27	-0.222	0.057	0.421	0.058
CD300C	-0.251	0.031	0.129	0.577
CD63	-0.291	0.012	0.405	0.069
CD70	-0.278	0.017	0.070	0.762
CDC2_CCNB1	-0.232	0.047	0.205	0.373
CDH15	-0.216	0.065	0.407	0.067
CDH6	-0.413	0.000	0.156	0.501
CDON	-0.314	0.006	0.046	0.844
CEBPB	-0.220	0.060	0.139	0.548
CFB	-0.306	0.008	0.281	0.217
CFHR5	-0.199	0.090	0.154	0.505
CHKB	-0.351	0.002	0.104	0.653
CNDP2	-0.213	0.068	0.015	0.947
CRISP3	-0.254	0.029	0.525	0.014
CTSC	-0.222	0.058	0.433	0.050
CTSD	-0.210	0.073	0.267	0.241
CTSS	-0.332	0.004	0.108	0.641
CTSV	-0.215	0.066	0.182	0.429

CXCL5	-0.302	0.009	0.246	0.282
CYCS	-0.339	0.003	0.141	0.543
DKKL1	-0.362	0.002	0.346	0.124
DPP7	-0.297	0.010	0.128	0.580
DSC3	-0.257	0.027	0.134	0.562
DYRK3	-0.256	0.028	0.012	0.960
EGFR	-0.300	0.009	0.222	0.333
ENPP7	-0.264	0.023	0.242	0.292
EPHB4	-0.302	0.009	0.239	0.297
ERBB3	-0.342	0.003	0.038	0.869
ERBB4	-0.350	0.002	0.059	0.801
ETHE1	-0.197	0.093	0.072	0.756
F3	-0.197	0.092	0.277	0.224
FAM107A	-0.196	0.095	0.249	0.276
FAM3B	-0.211	0.071	0.016	0.946
FCAR	-0.223	0.056	0.009	0.969
FCN2	-0.388	0.001	0.371	0.098
FCN3	-0.215	0.066	0.040	0.864
FGF1	-0.279	0.016	0.049	0.833
FGF10	-0.324	0.005	0.300	0.186
FGF17	-0.263	0.024	0.053	0.818
FGF9	-0.244	0.036	0.001	0.996
GDF11_MSTN	-0.201	0.086	0.015	0.949
GDNF	-0.233	0.045	0.031	0.895
GHR	-0.255	0.029	0.271	0.235
GPC6	-0.196	0.094	0.179	0.437
GPT	-0.213	0.069	0.351	0.119
GRP	-0.275	0.018	0.103	0.656
GZMB	-0.292	0.012	0.096	0.678
HBEGF	-0.266	0.022	0.188	0.414
HCK	-0.264	0.023	0.002	0.993
HFE2	-0.347	0.002	0.220	0.337
HIF1A	-0.285	0.014	0.095	0.683
HMGN1	-0.241	0.039	0.172	0.455
HNRNPA2B1	-0.214	0.067	0.259	0.258
HPGD	-0.267	0.022	0.325	0.150
HSD17B1	-0.215	0.066	0.241	0.292
ICAM3	-0.233	0.046	0.051	0.825
IL10RA	-0.264	0.023	0.216	0.348
IL13	-0.244	0.036	0.074	0.751
IL17A	-0.375	0.001	0.113	0.627
IL1R2	-0.347	0.002	0.089	0.701
IL37	-0.279	0.016	0.345	0.126
KDR	-0.302	0.009	0.069	0.765

KIR3DS1	-0.276	0.017	0.010	0.965
KLK14	-0.203	0.083	0.113	0.627
KLK5	-0.287	0.013	0.250	0.275
KYNU	-0.285	0.014	0.197	0.391
LDLR	-0.211	0.071	0.096	0.678
LGALS7	-0.258	0.027	0.248	0.277
LGMN	-0.412	0.000	0.261	0.253
LIN7B	-0.246	0.034	0.401	0.072
LMAN2	-0.320	0.006	0.141	0.541
LRRK2	-0.258	0.027	0.218	0.341
LTBR	-0.245	0.036	0.190	0.409
LY86	-0.287	0.013	0.048	0.836
MIF	-0.386	0.001	0.230	0.315
MSR1	-0.228	0.051	0.038	0.871
MST1R	-0.201	0.086	0.158	0.495
MSTN	-0.216	0.065	0.100	0.667
NOG	-0.277	0.017	0.137	0.555
NRG4	-0.216	0.065	0.065	0.779
NXPH1	-0.329	0.004	0.236	0.302
OAS1	-0.392	0.001	0.029	0.902
OPCML	-0.375	0.001	0.338	0.133
OSM	-0.193	0.099	0.210	0.361
PDE2A	-0.199	0.090	0.242	0.290
PGF	-0.264	0.023	0.053	0.818
PGLYRP1	-0.405	0.000	0.183	0.426
PIAS4	-0.261	0.025	0.069	0.765
PIGR	-0.226	0.053	0.111	0.631
PLCG1	-0.303	0.009	0.307	0.175
PLG	-0.232	0.046	0.101	0.664
PLG	-0.232	0.046	0.101	0.664
PLG_2	-0.365	0.001	0.056	0.808
PPP3R1	-0.263	0.024	0.321	0.155
PROC	-0.284	0.014	0.127	0.583
PROK1	-0.282	0.015	0.283	0.215
PRSS27	-0.268	0.021	0.242	0.292
PRTN3	-0.227	0.051	0.080	0.729
RASA1	-0.246	0.035	0.059	0.799
RBP4	-0.323	0.005	0.084	0.717
SEMA3A	-0.234	0.044	0.095	0.683
SEMA5A	-0.197	0.093	0.499	0.021
SERPINA4	-0.209	0.075	0.145	0.532
SERPINA7	-0.224	0.055	0.192	0.404
SERPIND1	-0.295	0.011	0.255	0.264
SERPINE1	-0.249	0.033	0.158	0.493

SIGLEC1	-0.226	0.053	0.190	0.408
SMPDL3A	-0.411	0.000	0.101	0.664
SPOCK1	-0.207	0.076	0.402	0.071
SPOCK2	-0.338	0.003	0.199	0.388
STK17B	-0.203	0.083	0.050	0.829
TG	-0.224	0.055	0.032	0.891
TNFRSF11B	-0.198	0.091	0.210	0.361
TP53	-0.342	0.003	0.256	0.262
TYK2	-0.202	0.084	0.117	0.615
UBB	-0.215	0.065	0.109	0.639
VEGFA_1	-0.240	0.039	0.110	0.635
WNK3	-0.212	0.070	0.045	0.847
ZNRF3	-0.198	0.090	0.066	0.777
A2M	0.219	0.061	-0.320	0.157
C7	0.197	0.093	-0.066	0.775
CADM1	0.210	0.072	-0.076	0.743
CCDC80	0.320	0.005	-0.205	0.373
CCL23	0.239	0.040	-0.117	0.613
CCL23	0.239	0.040	-0.117	0.613
CCL23_1	0.315	0.006	-0.042	0.858
CCL3	0.304	0.008	-0.119	0.607
CD109	0.226	0.053	-0.247	0.281
CHI3L1	0.196	0.094	-0.037	0.873
CHIT1	0.195	0.096	-0.160	0.490
CSNK2A2_CSNK2B	0.233	0.046	-0.007	0.975
CST3	0.245	0.035	-0.200	0.385
CST5	0.365	0.001	-0.047	0.840
CTSB	0.267	0.021	-0.076	0.743
CXCL13	0.232	0.047	-0.282	0.215
CXCL8	0.275	0.018	-0.258	0.260
CXCL9	0.286	0.014	-0.019	0.934
DKK3	0.287	0.013	-0.003	0.991
DPT	0.328	0.004	-0.113	0.625
EDA2R	0.208	0.075	-0.070	0.763
EFNA5	0.226	0.052	-0.188	0.416
EFNB2	0.247	0.034	-0.026	0.912
EPHA2	0.318	0.006	-0.018	0.937
FCGR2A	0.226	0.053	-0.141	0.541
FGA_FGB_FGG	0.316	0.006	-0.163	0.480
FLRT2	0.228	0.051	-0.037	0.873
H2AFZ	0.233	0.046	-0.236	0.304
IGFBP6	0.212	0.070	-0.024	0.919
IL15RA	0.222	0.057	-0.092	0.690
LAYN	0.268	0.021	-0.079	0.732

LGALS3	0.196	0.095	-0.336	0.137
LRP8	0.193	0.100	-0.084	0.716
MICA	0.194	0.098	-0.245	0.284
MMP12	0.378	0.001	-0.024	0.917
MMP3	0.217	0.063	-0.096	0.680
MMP7	0.348	0.002	-0.202	0.379
NBL1	0.326	0.005	-0.179	0.438
NRCAM	0.237	0.042	-0.130	0.574
NRP1	0.222	0.057	-0.562	0.008
NTN1	0.206	0.079	-0.086	0.711
OMD	0.201	0.086	-0.639	0.002
PIANP	0.274	0.018	-0.073	0.753
PLAU	0.197	0.092	-0.014	0.951
PLAUR	0.203	0.082	-0.005	0.982
PTN	0.400	0.000	-0.205	0.372
RSPO3	0.281	0.015	-0.091	0.696
SHH	0.258	0.026	-0.089	0.703
SMOC1	0.322	0.005	-0.107	0.646
SPON1	0.228	0.051	-0.142	0.540
SPP1	0.224	0.055	-0.464	0.034
TFF1	0.244	0.036	-0.239	0.297
TFF2	0.271	0.020	-0.327	0.148
THBS4	0.298	0.010	-0.175	0.447
TNFRSF1A	0.203	0.082	-0.076	0.745
UNC5D	0.225	0.054	-0.393	0.078
WFIKKN2	0.301	0.009	-0.486	0.025

Table 4. Candidate nonsynonymous mutations within candidate genes in two populations with allele or genotype frequency of 2-3%

Gene	Chr	Pos	SNP	Ref	Alt	Freq	SIFT	FATHMM	Population	Group
ABCC6	16	16182435	rs59593133	T	C	C 0.0230	T	D	1000 Genome	A
ABCC6	16	16165739	rs41278174	G	A	GA 0.0232	D	D	1000 Genome	A
ABCC6	16	16169805	rs61340537	G	T	GT 0.0200	T	D	1000 Genome	A
ABCC6	16	16182488	rs58073789	C	T	CT 0.0268	T	D	1000 Genome	A
ABL2	1	179142975	rs202125612	G	T	GT 0.0204	D	T	1000 Genome	A
ADAMTS15	11	130462562	rs116897071	T	C	C 0.0210	D	T	1000 Genome	A
AFP	4	73445051	rs41265657	C	G	CG 0.0260	D	D	1000 Genome	A
ARHGEF25	12	57610287	rs113892147	G	A	GA 0.0204	D	T	1000 Genome	A
ARMC5	16	31458422	rs115527944	G	A	GA 0.0280	D	T	1000 Genome	A
ARSK	5	95555355	rs61733082	G	C	GC 0.0236	T	D	1000 Genome	A
ATP2A3	17	3941051	rs9895012	G	A	GA 0.0244	T	D	1000 Genome	A
AXIN2	17	65537421	rs9913621	C	T	CT 0.0268	D	T	1000 Genome	A
AXIN2	17	65537801	rs115931022	T	C	TC 0.0204	T	D	1000 Genome	A
BRDT	1	92002069	rs78267346	A	G	AG 0.0216	D	T	1000 Genome	A
C2	6	31936027	rs9332739	G	C	C 0.0298	T	D	1000 Genome	A
C7	5	40936439	rs2271708	T	C	C 0.0262	D	T	1000 Genome	A
CA1	8	85333547	rs7821248	G	A	A 0.0238	D	T	1000 Genome	A
CCL25	19	8056208	rs143413416	C	T	CT 0.0296	D	T	1000 Genome	A
CD109	6	73811140	rs61754903	C	T	T 0.0264	D	T	1000 Genome	A
CD27	12	6445462	rs25680	G	A	AA 0.0300	T	D	1000 Genome	A
CD300C	17	74544797	rs11870245	G	A	AA 0.0272	D	T	1000 Genome	A
CDNF	10	14820083	rs61738953	C	G	CG 0.0264	D	.	1000 Genome	A
CDON	11	126015388	rs35665264	G	C	GC 0.0276	D	T	1000 Genome	A
CFB	6	31946247	rs4151667	T	A	A 0.0300	D	T	1000 Genome	A
CFB	6	31946403	rs641153	G	A	AA 0.0212	T	D	1000 Genome	A
CFB	6	31951241	rs4151660	T	G	TG 0.0208	T	D	1000 Genome	A
CHIT1	1	203217835	rs9943208	C	T	T 0.0254	D	D	1000 Genome	A
CHIT1	1	203219815	rs61745299	G	C	C 0.0216	D	T	1000 Genome	A
CLC	19	39734391	rs146776010	G	C	GC 0.0264	D	T	1000 Genome	A
COL6A5	3	130471687	rs79145375	T	C	C 0.0264	D	T	1000 Genome	A
COL6A5	3	130376336	rs113396273	T	C	TC 0.0208	D	T	1000 Genome	A
COL6A5	3	130403649	rs114408026	A	G	AG 0.0236	D	D	1000 Genome	A
COL6A5	3	130405625	rs77730506	G	A	GA 0.0212	D	D	1000 Genome	A
COL6A5	3	130469371	rs115907998	A	T	AT 0.0200	D	D	1000 Genome	A
COLGALT1	19	17579483	rs544877640	T	G	G 0.0232	D	T	1000 Genome	A
CRELD1	3	9938029	rs2302787	C	G	CG 0.0244	D	T	1000 Genome	A
CSF2RB	22	36937615	rs1801122	C	A	CA 0.0208	D	D	1000 Genome	A
CSH1	17	63896230	rs200765645	G	A	A 0.0206	T	D	1000 Genome	A
CST8	20	23491822	rs35190670	C	T	T 0.0210	D	T	1000 Genome	A
CUBN	10	17129177	rs12259370	C	T	T 0.0216	D	T	1000 Genome	A

CUBN	10	16890491	rs1801238	G	T	GT 0.0264	D	T	1000 Genome	A
DDC	7	50528222	rs6262	G	A	GA 0.0248	D	T	1000 Genome	A
DKK3	11	11964597	rs114873269	T	C	C 0.0218	D	T	1000 Genome	A
DKKL1	19	49374796	rs111233735	G	A	GA 0.0292	D	T	1000 Genome	A
DLL1	6	170283949	rs16901311	C	T	CT 0.0256	D	D	1000 Genome	A
DNASE2B	1	84402058	rs72943980	T	C	TC 0.0244	D	T	1000 Genome	A
DOK2	8	21911921	rs56094005	A	G	AG 0.0228	D	T	1000 Genome	A
DPT	1	168729060	rs998688	C	T	CT 0.0248	D	T	1000 Genome	A
DPYSL5	2	26944749	rs41288789	G	A	GA 0.0296	T	D	1000 Genome	A
EDA2R	X	66602765	rs1385698	T	C	TT 0.0220	T	D	1000 Genome	A
EMILIN3	20	41366471	rs192531993	T	G	TG 0.0256	D	T	1000 Genome	A
ETHE1	19	43527117	rs116440799	C	A	A 0.0234	T	D	1000 Genome	A
F3	1	94530506	rs3789683	C	T	CT 0.0280	D	T	1000 Genome	A
FAM171B	2	186751157	rs79964543	C	T	CT 0.0204	D	T	1000 Genome	A
FCN2	9	134887245	rs7851696	G	T	TT 0.0272	D	T	1000 Genome	A
FEN1	11	61795891	rs552145732	C	A	CA 0.0220	D	T	1000 Genome	A
FGG	4	154610181	rs2066870	A	G	G 0.0232	T	D	1000 Genome	A
FRZB	2	182834857	rs7775	G	C	CC 0.0284	D	T	1000 Genome	A
GAPDHS	19	35538338	rs61745734	G	T	GT 0.0220	D	T	1000 Genome	A
GNGT2	17	49207390	rs9895097	C	A	CA 0.0276	D	T	1000 Genome	A
GZMA	5	55108225	rs2270627	G	T	GT 0.0204	D	T	1000 Genome	A
GZMA	5	55108350	rs61730831	A	G	AG 0.0284	D	D	1000 Genome	A
HGD	3	120674935	rs138846036	C	A	CA 0.0260	T	D	1000 Genome	A
HNF4A	20	44413724	rs1800961	C	T	T 0.0240	T	D	1000 Genome	A
HS3ST3B1	17	14345560	rs9906590	G	A	GA 0.0284	T	D	1000 Genome	A
HSD17B14	19	48815123	rs35299026	G	A	GA 0.0284	D	T	1000 Genome	A
HTN1	4	70055510	rs61744465	T	C	C 0.0200	D	T	1000 Genome	A
IFNA10	9	21206744	rs112963053	G	T	GT 0.0284	D	T	1000 Genome	A
IL12B	5	159316780	rs3213119	C	A	CA 0.0232	D	T	1000 Genome	A
IL16	15	81290495	rs200730945	G	A	GA 0.0272	D	T	1000 Genome	A
IL16	15	81299475	rs34460207	A	G	AG 0.0236	D	T	1000 Genome	A
IL27RA	19	14050824	rs76543168	C	T	T 0.0292	D	T	1000 Genome	A
IL7R	5	35876347	rs2229232	C	T	CT 0.0252	D	T	1000 Genome	A
INSL5	1	66801073	rs549148	T	A	TT 0.0240	T	D	1000 Genome	A
INSL6	9	5185460	rs62620190	G	A	A 0.0228	T	D	1000 Genome	A
ITGAL	16	30481502	rs1064524	C	T	CT 0.0232	D	D	1000 Genome	A
KCNA10	1	110518027	rs11806812	T	C	C 0.0300	T	D	1000 Genome	A
KIAA1324L	7	86909913	rs34321015	A	G	AG 0.0296	D	T	1000 Genome	A
KLRC3	12	10415651	rs149679250	T	A	TA 0.0264	D	T	1000 Genome	A
KLRC3	12	10415654	rs115905385	A	T	AT 0.0264	D	T	1000 Genome	A
KMT2C	7	152252650	rs77652527	T	C	TC 0.0232	T	D	1000 Genome	A
KMT2C	7	152358656	rs111826855	T	C	TC 0.0200	D	D	1000 Genome	A
KRT17	17	41622360	rs140804147	C	T	T 0.0240	D	D	1000 Genome	A
KRT5	12	52519884	rs11170164	C	T	T 0.0266	D	D	1000 Genome	A

KRT5	12	52515088	rs11549949	C	T	TT 0.0204	T	D	1000 Genome	A
KRT5	12	52515133	rs11549950	T	C	CC 0.0296	T	D	1000 Genome	A
KRT7	12	52248734	rs74761279	G	A	A 0.0278	T	D	1000 Genome	A
KRT7	12	52248741	rs111432010	G	A	A 0.0232	T	D	1000 Genome	A
KRT7	12	52243090	rs7963792	C	T	CT 0.0224	D	D	1000 Genome	A
LAYN	11	111544009	rs11827718	G	A	GA 0.0200	D	T	1000 Genome	A
LCN8	9	136756542	rs78263063	G	A	GA 0.0276	D	T	1000 Genome	A
LILRB4	19	54667873	rs112611821	G	C	C 0.0244	D	T	1000 Genome	A
LRP1	12	57184890	rs34577247	G	A	GA 0.0236	T	D	1000 Genome	A
LRTM1	3	54918617	rs61735203	C	T	T 0.0210	D	T	1000 Genome	A
LTBR	12	6390130	rs35681405	G	A	A 0.0256	T	D	1000 Genome	A
LTBR	12	6385344	rs41393248	C	T	CT 0.0292	D	T	1000 Genome	A
MEPE	4	87846108	rs74593863	T	G	TG 0.0228	D	T	1000 Genome	A
MGA	15	41729314	rs17677811	T	C	TC 0.0252	D	T	1000 Genome	A
MGAT1	5	180792838	rs61743256	C	T	CT 0.0300	T	D	1000 Genome	A
MICA	6	31411155	rs41556715	G	A	GA 0.0268	D	T	1000 Genome	A
MICA	6	31411357	rs1131897	C	G	GG 0.0276	D	T	1000 Genome	A
MICA	6	31412153	rs41554616	C	G	CG 0.0284	D	T	1000 Genome	A
MMP17	12	131851047	rs11835665	G	A	AA 0.0264	D	T	1000 Genome	A
MMP7	11	102527598	rs17884789	C	T	T 0.0218	D	T	1000 Genome	A
MMRN2	10	86943015	rs200124789	C	T	CT 0.0272	D	T	1000 Genome	A
MPP7	10	28120369	rs145570841	C	T	CT 0.0260	D	T	1000 Genome	A
MRAP	21	32311866	rs114530014	C	T	CT 0.0272	D	D	1000 Genome	A
MRVI1	11	10626448	rs35468145	G	A	A 0.0250	D	T	1000 Genome	A
MSR1	8	16120603	rs138987923	G	A	GA 0.0220	T	D	1000 Genome	A
MTPAP	10	30349161	rs114438816	C	A	A 0.0288	D	T	1000 Genome	A
MYO6	6	75894825	rs9447572	A	G	G 0.0212	T	D	1000 Genome	A
MYO6	6	75855180	rs61732664	T	C	TC 0.0260	T	D	1000 Genome	A
MYOM2	8	2069280	rs34823600	C	T	CT 0.0264	D	T	1000 Genome	A
NBR1	17	43191443	rs115552058	G	A	GA 0.0300	D	T	1000 Genome	A
NCR2	6	41350686	rs2273961	T	A	AA 0.0288	D	T	1000 Genome	A
NEFH	22	29489027	rs59371099	G	A	A 0.0294	D	D	1000 Genome	A
NEFH	22	29489484	rs5763269	C	T	TT 0.0268	D	D	1000 Genome	A
NEO1	15	73283089	rs118087147	C	T	CT 0.0228	D	T	1000 Genome	A
NEURL4	17	7327526	rs11655578	G	T	TT 0.0280	D	T	1000 Genome	A
NID1	1	236042041	rs34406281	C	T	T 0.0216	T	D	1000 Genome	A
NID1	1	236042140	rs16833183	C	T	T 0.0232	T	D	1000 Genome	A
NID1	1	236013545	rs35714220	C	T	CT 0.0252	T	D	1000 Genome	A
NPC2	14	74480634	rs61738594	G	A	GA 0.0220	D	D	1000 Genome	A
NQO1	16	69714966	rs1131341	G	A	A 0.0214	D	T	1000 Genome	A
NQO1	16	69715025	rs11555215	C	G	CG 0.0256	D	T	1000 Genome	A
NTRK1	1	156879126	rs6336	C	T	T 0.0244	D	D	1000 Genome	A
NTRK1	1	156879154	rs6339	G	T	T 0.0230	D	D	1000 Genome	A
PAM	5	103003107	rs35658696	A	G	AG 0.0268	D	D	1000 Genome	A

PCDHA7	5	140834646	rs17844305	C	G	G 0.0300	D	T	1000 Genome	A
PDILT	16	20362484	rs11648131	C	T	T 0.0260	D	T	1000 Genome	A
PLG	6	160706526	rs4252070	G	A	GA 0.0240	T	D	1000 Genome	A
PLOD3	7	101210630	rs75592752	G	C	C 0.0236	D	D	1000 Genome	A
PNKP	19	49861464	rs3739206	A	C	AC 0.0232	D	T	1000 Genome	A
PNKP	19	49864229	rs3739186	A	T	AT 0.0220	D	T	1000 Genome	A
POLI	18	54293839	rs3218786	T	C	TC 0.0208	D	T	1000 Genome	A
PRIM1	12	56752285	rs2277339	T	G	GG 0.0268	D	T	1000 Genome	A
PRNP	20	4699875	rs1800014	G	A	GA 0.0300	D	D	1000 Genome	A
PROK1	1	110454030	rs62623571	C	T	T 0.0236	D	D	1000 Genome	A
QRFP	9	130893636	rs12340616	A	T	AT 0.0220	D	T	1000 Genome	A
RNASE4	14	20699417	rs3748338	A	T	TT 0.0216	D	T	1000 Genome	A
ROBO1	3	78635874	rs35456279	C	T	T 0.0292	D	T	1000 Genome	A
SECTM1	17	82322180	rs145723287	G	A	GA 0.0268	D	.	1000 Genome	A
SEMG1	20	45207596	rs61729393	T	A	TA 0.0236	D	T	1000 Genome	A
SERPIND1	22	20779331	rs5905	G	A	GA 0.0272	T	D	1000 Genome	A
SIGLEC1	20	3701479	rs34924243	C	T	T 0.0274	D	T	1000 Genome	A
SORBS3	8	22564314	rs75023966	T	C	TC 0.0244	D	T	1000 Genome	A
SORBS3	8	22564469	rs3758036	C	T	CT 0.0240	D	T	1000 Genome	A
SPP1	4	87982660	rs575565830	A	C	AC 0.0224	D	T	1000 Genome	A
SPP1	4	87982664	rs78051010	A	C	AC 0.0224	D	T	1000 Genome	A
STAB1	3	52517957	rs77803411	C	T	T 0.0214	T	D	1000 Genome	A
STX2	12	130808702	rs76596600	T	C	TC 0.0200	D	T	1000 Genome	A
SULT2A1	19	47883735	rs11569681	C	G	G 0.0240	T	D	1000 Genome	A
TBX22	X	80025703	rs34244923	G	A	A 0.0209	D	D	1000 Genome	A
TENM4	11	78658435	rs76975254	C	T	CT 0.0220	T	D	1000 Genome	A
TG	8	132941495	rs61744749	A	C	C 0.0280	D	T	1000 Genome	A
THBS4	5	80065442	rs1866389	G	C	CC 0.0232	D	D	1000 Genome	A
THBS4	5	80078171	rs2229398	G	A	GA 0.0236	T	D	1000 Genome	A
TMEM119	12	108592241	rs73404789	G	A	A 0.0262	D	T	1000 Genome	A
TNFAIP3	6	137874854	rs146534657	A	G	AG 0.0252	D	T	1000 Genome	A
TNFRSF10D	8	23137998	rs34622674	T	G	G 0.0238	T	D	1000 Genome	A
TNFRSF10D	8	23137974	rs112449156	G	T	GT 0.0244	T	D	1000 Genome	A
TNFRSF10D	8	23138194	rs61752042	C	T	CT 0.0228	T	D	1000 Genome	A
TNFRSF10D	8	23144577	rs55636833	C	T	CT 0.0288	T	D	1000 Genome	A
TNFRSF10D	8	23144611	rs78332874	G	A	GA 0.0244	T	D	1000 Genome	A
TNFRSF10D	8	23145716	rs150861587	A	T	AT 0.0240	T	D	1000 Genome	A
TNFRSF10D	8	23145719	rs139303833	C	A	CA 0.0240	T	D	1000 Genome	A
TNFRSF10D	8	23145779	rs35213435	G	T	GT 0.0248	T	D	1000 Genome	A
TNFRSF10D	8	23145881	rs111569580	G	A	GA 0.0276	T	D	1000 Genome	A
TNFRSF14	1	2559867	rs2234163	G	A	A 0.0212	T	D	1000 Genome	A
TNFRSF1A	12	6333835	rs4149637	G	A	A 0.0282	D	D	1000 Genome	A
TNFRSF8	1	12138305	rs144498730	C	A	CA 0.0248	D	T	1000 Genome	A
TNS2	12	53059649	rs11558984	G	A	A 0.0274	T	D	1000 Genome	A

TNS4	17	40487344	rs33923045	G	T	T 0.0220	D	T	1000 Genome	A
TRAT1	3	108853726	rs57744779	A	G	G 0.0210	D	T	1000 Genome	A
TYK2	19	10359299	rs12720356	A	C	C 0.0282	D	T	1000 Genome	A
TYK2	19	10352442	rs34536443	G	C	GC 0.0204	D	T	1000 Genome	A
VWC2L	2	214575691	rs16852050	A	T	T 0.0200	D	.	1000 Genome	A
XRCC1	19	43553422	rs1799782	G	A	AA 0.0216	D	T	1000 Genome	A
ZNF843	16	31436564	rs116147135	G	T	GT 0.0220	D	T	1000 Genome	A
ABCC6	16	16278863	rs8058694	G	T	TT 0.0272	T	D	CONVERGE	A
ABCC6	16	16281007	rs12931472	A	G	GG 0.0255	T	D	CONVERGE	A
ADAMTS15	11	130319265	rs185269810	G	C	GC 0.0269	D	T	CONVERGE	A
ARHGEF1	19	42402593	rs2303797	C	T	T 0.0216	D	T	CONVERGE	A
AXIN2	17	63533904	rs201460658	G	A	GA 0.0227	T	D	CONVERGE	A
C2	6	31903804	rs9332739	G	C	C 0.0232	T	D	CONVERGE	A
C7	5	40936541	rs2271708	T	C	C 0.0251	D	T	CONVERGE	A
CASS4	20	55020999	rs77627768	C	T	T 0.0204	D	T	CONVERGE	A
CDH15	16	89261471	rs149307887	G	A	A 0.0260	D	T	CONVERGE	A
CDHR1	10	85961608	rs151204356	C	T	CT 0.0268	D	T	CONVERGE	A
CEBPB	20	48807605	20:48807605	T	C	C 0.0250	D	T	CONVERGE	A
CFB	6	31914024	rs4151667	T	A	A 0.0240	D	T	CONVERGE	A
COL6A5	3	130116696	rs12488457	A	C	CC 0.0214	D	T	CONVERGE	A
COL6A5	3	130125116	rs1497312	G	C	CC 0.0240	T	D	CONVERGE	A
COL6A5	3	130134492	rs16827497	T	C	CC 0.0248	T	D	CONVERGE	A
CTSB	8	11708479	rs74996838	T	C	TC 0.0214	T	D	CONVERGE	A
CUBN	10	16960683	rs140806389	T	A	A 0.0271	D	T	CONVERGE	A
CUBN	10	16961995	rs2271460	A	C	C 0.0281	D	T	CONVERGE	A
CUBN	10	17153023	rs78201384	C	T	CT 0.0262	T	D	CONVERGE	A
CUL4B	X	119693637	rs113213082	C	T	CT 0.0271	D	T	CONVERGE	A
DKKL1	19	49867899	rs2303757	T	G	TG 0.0262	D	T	CONVERGE	A
EDAR	2	109513572	rs146567337	T	G	G 0.0203	D	D	CONVERGE	A
EHD4	15	42235316	rs11549015	C	T	CT 0.0289	T	D	CONVERGE	A
EPHA2	1	16461581	rs55747232	G	A	A 0.0207	D	T	CONVERGE	A
EPHB4	7	100417364	rs55720981	G	A	GA 0.0205	T	D	CONVERGE	A
ERBB3	12	56495023	rs2271188	G	A	GA 0.0251	D	T	CONVERGE	A
FCN2	9	137777132	rs12684476	G	A	A 0.0205	D	T	CONVERGE	A
FLNA	X	153593613	rs143873938	C	T	T 0.0257	D	D	CONVERGE	A
GHR	5	42719092	rs6183	C	A	CA 0.0230	D	T	CONVERGE	A
GZMA	5	54404053	rs2270627	G	T	GT 0.0299	D	T	CONVERGE	A
HHLA2	3	108081217	rs3792332	C	A	AA 0.0212	D	T	CONVERGE	A
IL7R	5	35874575	rs6897932	C	T	TT 0.0243	T	D	CONVERGE	A
KCNA10	1	111061066	rs150926393	C	T	CT 0.0270	D	T	CONVERGE	A
KMT2C	7	151859683	rs74483926	G	A	AA 0.0259	T	D	CONVERGE	A
LILRB4	19	55179120	rs147216101	C	T	CT 0.0209	D	T	CONVERGE	A
LRIT2	10	85984444	rs11200927	T	G	GG 0.0222	D	T	CONVERGE	A
LRP1	12	57602881	rs79435985	A	T	AT 0.0231	D	D	CONVERGE	A

LRP8	1	53712727	rs5174	C	T	CT 0.0284	D	D	CONVERGE	A
MICA	6	31379134	rs115257533	C	G	G 0.0292	D	T	CONVERGE	A
MICA	6	31378864	rs148069425	A	G	AG 0.0258	D	T	CONVERGE	A
MNX1	7	156801700	rs78596384	G	T	TT 0.0247	T	D	CONVERGE	A
MNX1	7	156802998	7:156802998	T	G	TG 0.0227	D	D	CONVERGE	A
MST1R	3	49924940	rs1062633	T	C	CC 0.0284	T	D	CONVERGE	A
MYOM2	8	2017399	rs34823600	C	T	T 0.0248	D	T	CONVERGE	A
MYOM2	8	2024280	rs117189614	G	A	GA 0.0256	D	T	CONVERGE	A
NCL	2	232326517	rs117088491	T	G	TG 0.0271	D	T	CONVERGE	A
NGRN	15	90814664	rs11073922	C	T	CT 0.0215	D	T	CONVERGE	A
NID1	1	236228288	rs2071529	C	A	A 0.0213	D	T	CONVERGE	A
NID1	1	236176845	rs35714220	C	T	CT 0.0233	T	D	CONVERGE	A
PCDHA7	5	140214231	rs17844305	C	G	CG 0.0245	D	T	CONVERGE	A
PNKP	19	50364721	rs3739206	A	C	C 0.0245	D	T	CONVERGE	A
PRNP	20	4680251	rs1799990	A	G	G 0.0232	D	D	CONVERGE	A
PSG4	19	43699278	rs2355442	A	G	AG 0.0257	D	T	CONVERGE	A
RBM19	12	114282496	rs2075387	C	T	TT 0.0242	D	T	CONVERGE	A
RBM19	12	114377835	rs2290788	A	G	GG 0.0292	D	T	CONVERGE	A
ROBO3	11	124744033	rs74787566	G	A	GA 0.0282	T	D	CONVERGE	A
SORBS3	8	22421982	rs3758036	C	T	T 0.0240	D	T	CONVERGE	A
TENM4	11	78565314	rs17137261	C	G	G 0.0220	D	T	CONVERGE	A
TNFAIP3	6	138195991	rs146534657	A	G	AG 0.0289	D	T	CONVERGE	A
TNS2	12	53442956	rs12369033	G	C	CC 0.0200	D	D	CONVERGE	A
TNS4	17	38643384	rs140876567	C	A	CA 0.0200	D	T	CONVERGE	A
YES1	18	745840	rs34580680	T	C	TC 0.0202	T	D	CONVERGE	A
ZP4	1	238050801	rs35921460	G	A	GA 0.0211	T	D	CONVERGE	A
ABCB1	7	87600124	rs9282564	T	C	C 0.0260	T	D	1000 Genome	B
ABCB1	7	87550493	rs2229109	C	T	CT 0.0236	T	D	1000 Genome	B
ADAL	15	43348983	rs77284705	A	G	AG 0.0200	T	D	1000 Genome	B
AMPD1	1	114679616	rs34526199	T	A	TA 0.0204	D	D	1000 Genome	B
AMPD3	11	10482189	rs11042836	C	T	T 0.0230	D	T	1000 Genome	B
AMPD3	11	10495666	rs36003153	T	C	C 0.0238	T	D	1000 Genome	B
ANO1	11	70161664	rs2186797	T	C	C 0.0202	D	T	1000 Genome	B
BEST1	11	61962955	rs115979721	G	T	T 0.0214	D	D	1000 Genome	B
BEST1	11	61956981	rs74653691	C	A	CA 0.0232	D	D	1000 Genome	B
BEST1	11	61959966	rs1801390	C	T	CT 0.0280	D	D	1000 Genome	B
BEST1	11	61963081	rs17185413	T	C	CC 0.0220	D	D	1000 Genome	B
BEST2	19	12752615	rs79300835	G	A	A 0.0292	T	D	1000 Genome	B
BEST2	19	12757736	rs61737519	C	T	CT 0.0272	T	D	1000 Genome	B
BEST3	12	69643740	rs61747221	G	A	AA 0.0252	D	D	1000 Genome	B
BEST3	12	69676990	rs114061225	C	T	CT 0.0232	T	D	1000 Genome	B
BEST4	1	44784335	rs59895800	C	G	GG 0.0280	T	D	1000 Genome	B
CDHR2	5	176578480	rs61743422	G	A	GA 0.0212	D	T	1000 Genome	B
CFTR	7	117587820	rs75789129	A	G	AG 0.0216	T	D	1000 Genome	B

CLCN3	4	169660419	rs4692739	G	A	G 0.0252	T	D	1000 Genome	B
CR1	1	207609424	rs41274768	G	A	GA 0.0240	D	T	1000 Genome	B
DDX58	9	32481341	rs61752945	C	T	CT 0.0220	D	T	1000 Genome	B
DPP4	2	162018839	rs1129599	C	G	CG 0.0200	T	D	1000 Genome	B
DPP9	19	4682832	rs141294259	T	C	TC 0.0248	D	T	1000 Genome	B
GHRHR	7	30974060	rs28371560	G	A	GA 0.0244	D	T	1000 Genome	B
IFNK	9	27524626	rs34933275	T	A	A 0.0228	D	T	1000 Genome	B
IL12B	5	159316780	rs3213119	C	A	CA 0.0232	D	T	1000 Genome	B
IL33	9	6256144	rs16924241	C	G	CG 0.0296	D	T	1000 Genome	B
IL7	8	78798167	rs201412253	C	T	CT 0.0212	D	T	1000 Genome	B
ITGAX	16	31382471	rs4889652	C	G	G 0.0246	D	T	1000 Genome	B
KCNJ1	11	128840045	rs41302407	T	C	TC 0.0288	T	D	1000 Genome	B
MTUS1	8	17755585	rs61733703	G	T	GT 0.0220	D	T	1000 Genome	B
NEDD4	15	55915910	rs73416291	G	A	A 0.0272	D	T	1000 Genome	B
NEDD4	15	55916265	rs113176671	A	C	AC 0.0200	D	T	1000 Genome	B
NLRP3	1	247425556	rs35829419	C	A	A 0.0224	T	D	1000 Genome	B
NPHS2	1	179557040	rs61747727	G	A	A 0.0206	D	D	1000 Genome	B
NPHS2	1	179557079	rs61747728	C	T	CT 0.0268	T	D	1000 Genome	B
SCNN1A	12	6349184	rs5742912	A	G	AG 0.0260	D	T	1000 Genome	B
SCNN1D	1	1291471	rs61730205	C	T	T 0.0238	D	T	1000 Genome	B
SCNN1D	1	1287578	rs111819661	C	T	CT 0.0232	D	T	1000 Genome	B
TLR1	4	38797778	rs76796448	G	T	T 0.0230	D	T	1000 Genome	B
TLR1	4	38797918	rs3923647	T	A	A 0.0292	D	T	1000 Genome	B
TLR1	4	38798593	rs5743611	C	G	G 0.0226	D	T	1000 Genome	B
TLR1	4	38796634	rs5743621	G	A	GA 0.0236	D	T	1000 Genome	B
TLR10	4	38774018	rs11466658	G	A	A 0.0278	D	T	1000 Genome	B
TLR5	1	223112604	rs5744167	T	G	G 0.0204	D	T	1000 Genome	B
VCAM1	1	100738209	rs3783615	A	T	T 0.0226	D	T	1000 Genome	B
ANXA13	8	124710729	rs2294013	C	T	TT 0.0289	D	T	CONVERGE	B
BEST1	11	61725901	rs2524295	G	A	GA 0.0257	D	D	CONVERGE	B
CFTR	7	117227874	rs75789129	A	G	AG 0.0278	T	D	CONVERGE	B
MTUS1	8	17612526	rs138713013	C	T	CT 0.0211	D	T	CONVERGE	B
MX2	21	42749081	rs56680307	T	C	C 0.0203	T	D	CONVERGE	B
SCNN1D	1	1222958	rs111819661	C	T	CT 0.0254	D	T	CONVERGE	B
TICAM1	19	4818295	rs79591246	G	A	A 0.0212	D	T	CONVERGE	B

Group A: candidate proteins for COVID-19 inactivation or aging marker proteins in plasma

Group B: candidate genes for COVID-19 infection, morbidity and clearance

Chromosome position is based on hg37 for population CONVERGE

Chromosome position is based on hg38 for population 1000 genome

Table 5. Candidate nonsynonymous eQTL mutations of candidate genes in two populations with allele or genotype frequency of 2-3%

Target_gene	Chr	Pos	SNP	Ref	Alt	Freq	eQTL_gene	SIFT	FATHMM	Population	Group
C2	6	31270250	rs41540117	C	A	AA 0.0276	HLA-C	D	T	1000 Genome	A
CCL16 SOCS7	17	40771762	rs9898164	A	G	GG 0.0272	KRT26	D	D	1000 Genome	A
CCL23	17	41515472	rs2305556	T	C	CC 0.0252	KRT15	T	D	1000 Genome	A
CD58	11	5454276	rs11037502	G	A	AA 0.0276	OR51I2	D	T	1000 Genome	A
DHFR	5	80654905	rs2001675	G	C	C 0.0208	MSH3	T	D	1000 Genome	A
FAF2	2	166456941	rs11888208	T	C	CC 0.0212	SCN7A	T	D	1000 Genome	A
FGF9	6	26092913	rs1800562	G	A	GA 0.0236	HFE	D	D	1000 Genome	A
GAL	11	68433827	rs3736228	C	T	TT 0.0224	LRP5	T	D	1000 Genome	A
HSD17B14	19	45629998	rs7252175	G	A	AA 0.0204	EML2	D	T	1000 Genome	A
IL17RC	3	9835235	rs1057278	G	A	AA 0.0268	TTLL3	D	T	1000 Genome	A
LTBR	2	45413195	rs3755072	T	C	CC 0.0272	SRBD1	D	T	1000 Genome	A
MMP3 MMP10	15	41879285	rs1197682	G	A	AA 0.0204	SPTBN5	D	T	1000 Genome	A
NCR3	6	32832635	rs1800454	C	T	TT 0.0228	TAP2	T	D	1000 Genome	A
NMB	15	84839914	rs3803403	C	G	GG 0.0300	ALPK3	D	T	1000 Genome	A
NPDC1	9	136474501	rs3812594	G	A	AA 0.0272	SEC16A	D	T	1000 Genome	A
NQO1	16	69714966	rs1131341	G	A	A 0.0214	NQO1	D	T	1000 Genome	A
PPP3R1	2	68137346	rs13009282	T	C	CC 0.0284	WDR92	T	D	1000 Genome	A
PRIM1	12	56752285	rs2277339	T	G	GG 0.0268	PRIM1	D	T	1000 Genome	A
PRKAG1	12	48118548	rs11168414	C	T	TT 0.0232	PFKM	D	D	1000 Genome	A
SERPIND1	22	25835345	rs17704912	G	C	C 0.0226	MYO18B	D	T	1000 Genome	A
STAB1	3	52395649	rs419752	C	T	T 0.0300	DNAH1	D	T	1000 Genome	A
THBS4	5	80065442	rs1866389	G	C	CC 0.0232	THBS4	D	D	1000 Genome	A
TMEM119	12	108623488	rs7300972	T	C	CC 0.0280	SELPLG	D	T	1000 Genome	A
BUB1	2	111598958	rs1554005	C	T	TT 0.0244	ACOXL	D	D	CONVERGE	A
CD38	4	15688604	rs2302469	G	C	C 0.0297	FAM200B	D	T	CONVERGE	A
CDH15	16	89613123	rs2292954	A	G	GG 0.0253	SPG7	T	D	CONVERGE	A
CHKB	22	51012775	rs2269383	C	T	TT 0.0237	CPT1B	D	D	CONVERGE	A
ERO1L	14	50092471	rs9989177	T	C	CC 0.0226	DNAAF2	D	T	CONVERGE	A
FAF2	2	167313451	rs11888208	T	C	C 0.0217	SCN7A	T	D	CONVERGE	A
GLIPR1	12	75900382	rs11540407	C	T	TT 0.0236	KRR1	D	T	CONVERGE	A
HLA-DQA2	6	32486442	rs147669022	T	A	A 0.0208	HLA-DRB5	D	T	CONVERGE	A
IGFBP7 KRT7	12	52284668	rs12368048	C	A	AA 0.0272	ANKRD33	D	T	CONVERGE	A
INSR	19	8161450	rs3829817	C	T	TT 0.0243	FBN3	T	D	CONVERGE	A
MST1R	3	49924940	rs1062633	T	C	CC 0.0284	MST1R	T	D	CONVERGE	A
NELL2	12	45417666	rs2731038	T	C	TT 0.0298	DBX2	T	D	CONVERGE	A
NMB	15	85383145	rs3803403	C	G	GG 0.0212	ALPK3	D	T	CONVERGE	A
NMB	15	85383640	rs3803405	G	A	AA 0.0211	ALPK3	D	T	CONVERGE	A
NMB	15	85405995	rs187316	T	C	CC 0.0231	ALPK3	D	T	CONVERGE	A
OAS1	12	113386779	rs2285933	C	G	GG 0.0286	OAS3	D	T	CONVERGE	A

OAS1	12	114282496	rs2075387	C	T	TT 0.0242	RBM19	D	T	CONVERGE	A
RCVRN	17	8700982	rs2242373	C	T	TT 0.0281	MFSD6L	D	D	CONVERGE	A
SCIN	10	129906364	rs4750685	G	A	AA 0.0247	MKI67	D	T	CONVERGE	A
TNS4	17	39116728	rs17843021	G	A	AA 0.0204	KRT39	D	D	CONVERGE	A
WISP3	1	98165091	rs2297595	T	C	TC 0.0263	DPYD	D	T	CONVERGE	A
AGTR2	11	6877264	rs2595453	T	C	C 0.0214	OR10A4	D	T	1000 Genome	B
CD40	20	46043236	rs16985442	C	G	G 0.0220	SLC12A5	D	D	1000 Genome	B
DPP9	2	166456941	rs11888208	T	C	CC 0.0212	SCN7A	T	D	1000 Genome	B
GHSR	3	173117292	rs16846616	C	T	TT 0.0300	SPATA16	D	T	1000 Genome	B
MAS1	6	160249250	rs316019	A	C	AA 0.0220	SLC22A2	D	T	1000 Genome	B
NMB	15	84839914	rs3803403	C	G	GG 0.0300	ALPK3	D	T	1000 Genome	B
CECR1	22	17669306	rs2231495	T	C	CC 0.0226	ADA2	T	D	CONVERGE	B
DPP9	2	167313451	rs11888208	T	C	C 0.0217	SCN7A	T	D	CONVERGE	B
ERLIN1	12	45417666	rs2731038	T	C	TT 0.0298	DBX2	T	D	CONVERGE	B
NMB	15	85383145	rs3803403	C	G	GG 0.0212	ALPK3	D	T	CONVERGE	B
NMB	15	85383640	rs3803405	G	A	AA 0.0211	ALPK3	D	T	CONVERGE	B
NMB	15	85405995	rs187316	T	C	CC 0.0231	ALPK3	D	T	CONVERGE	B

Group A: candidate proteins for COVID-19 inactivation or aging marker proteins in plasma

Group B: candidate genes for COVID-19 infection, morbidity and clearance

Chromosome position is based on hg37 for population CONVERGE

Chromosome position is based on hg38 for population 1000 genome