Proteomic Profiling in Biracial Cohorts Implicates DC-SIGN as a Mediator of Genetic Risk in COVID-19 — Source link ⧉

Daniel H. Katz, Usman A. Tahir, Debby Ngo, Mark D. Benson ...+22 more authors

Institutions: Harvard University, Broad Institute, University of Mississippi, National Institutes of Health ...+1 more institutions

Related papers:

- SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor

- DC-SIGN and DC-SIGNR Interact with the Glycoprotein of Marburg Virus and the S Protein of Severe Acute Respiratory Syndrome Coronavirus

- CD209L/L-SIGN and CD209/DC-SIGN act as receptors for SARS-CoV-2 and are differentially expressed in lung and kidney epithelial and endothelial cells

- Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.

- Genomewide Association Study of Severe Covid-19 with Respiratory Failure.

Share this paper: 𝐟 🐦 in ✉

View more about this paper here: https://typeset.io/papers/proteomic-profiling-in-biracial-cohorts-implicates-dc-sign-1ycnxk0god

Proteomic Profiling in Biracial Cohorts Implicates DC-SIGN as a Mediator of Genetic Risk in COVID-19

Daniel H. Katz[1], Usman A. Tahir[1], Debby Ngo[1], Mark Benson[1], Alexander G. Bick[2], Akhil Pampana[2], Yan Gao[3], Michelle J. Keyes[1,4], Adolfo Correa[3], Sumita Sinha[1], Dongxiao Shen[1], Qiong Yang[5], Jeremy M. Robbins[1], Zsu-Zsu Chen[1], Daniel E. Cruz[1], Bennet Peterson[1], Pradeep Natarajan[2,6,7], Ramachandran S. Vasan[4,5], J. Gustav Smith[2,8,9,10,11], Thomas J. Wang[12], James G. Wilson[1], Robert E. Gerszten[1,2]

[1]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA; [2]Broad Institute of Harvard and MIT, Cambridge, MA [3]Univ of Mississippi Medical Center, Jackson, MS; [4]Framingham Heart Study, Framingham, MA; [5]Department of Biostatistics, Boston University School of Public Health, Boston, MA; [6]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; [7]Department of Medicine Harvard Medical School, Boston, MA; [8]Department of Cardiology, Clinical Sciences, Lund University, Lund, Sweden; [9]Department of Heart Failure and Valvular Disease, Skåne University Hospital, Lund, Sweden; [10]The Wallenberg Laboratory/Department of Molecular and Clinical Medicine, Institute of Medicine, Gothenburg University and the Department of Cardiology, Sahlgrenska University Hospital, Gothenburg, Sweden; [11]Wallenberg Center for Molecular Medicine and Lund University Diabetes Center, Lund University, Lund, Sweden; [12]Department of Medicine, UT Southwestern Medical Center, Dallas, TX;

Correspondence:

Robert E. Gerszten, MD,
Division of Cardiovascular Medicine
Beth Israel Deaconess Medical Center,
185 Pilgrim Road, Baker 408
Boston, MA 02215
Email: rgerszte@bidmc.harvard.edu

Brief Summary

COVID-19 is one of the most consequential pandemics in the last century, yet the biological mechanisms that confer disease risk are incompletely understood. Further, heterogeneity in disease outcomes is influenced by race, though the relative contributions of structural/social and genetic factors remain unclear.[1,2] Very recent unpublished work has identified two genetic risk loci that confer greater risk for respiratory failure in COVID-19: the *ABO* locus and the 3p21.31 locus.[3] To understand how these loci might confer risk and whether this differs by race, we utilized proteomic profiling and genetic information from three cohorts including black and white participants to identify proteins influenced by these loci. We observed that variants in the *ABO* locus are associated with levels of CD209/DC-SIGN, a known binding protein for SARS-CoV and other viruses,[4] as well as multiple inflammatory and thrombotic proteins, while the 3p21.31 locus is associated with levels of CXCL16, a known inflammatory chemokine.[5] Thus, integration of genetic information and proteomic profiling in biracial cohorts highlights putative mechanisms for genetic risk in COVID-19 disease.

Introduction

SARS-CoV-2 infection displays a wide array of clinical manifestations and degrees of severity. While there is evidence that comorbidities, particularly cardiovascular and metabolic disease, are risk factors for disease severity and outcomes,[6] the underlying biologic mechanisms that cause some to develop life threatening disease while others remain asymptomatic are not well understood.

The association recently observed between the *ABO* locus on chromosome 9 and susceptibility to respiratory failure in COVID-19 is consistent with earlier work showing an association between blood type and COVID-19 disease, though the mechanism(s) by which this locus might confer susceptibility to respiratory failure is unknown.[7,8] The association with the 3p21.31 region observed in the same recent study was particularly novel, but also of unclear significance.[3]

Emerging proteomic technologies enable large-scale protein profiling in population-based studies. Leveraging available genetic data, investigators have identified the genetic architecture of the circulating proteome.[9–11] Conversely, combing this information can identify proteomic signatures associated with specific loci or disease variants. Here we used measurements of 1,305 circulating proteins on the SOMAScan[TM] platform and genetic data from 4,859 participants in three large population-based studies: the Jackson Heart Study (JHS), a cohort of black participants, as well as meta-analyzed data from two white cohorts, the Framingham Heart Study (FHS) and the Malmö Diet and Cancer Study (MDCS). We tested for associations between genetic variants at the *ABO* and 3p21.31 loci and protein levels in the three cohorts to identify possible mediators of disease.

Results

Cohort Characteristics

Participants in the JHS with proteomics had an average age of 56 years and were 61% female. They had multiple comorbidities including hypertension (56% on treatment), diabetes mellitus (24%), and obesity (mean BMI 32). Baseline characteristics in FHS/MDCS have been reportedly previously.[9] In brief, participants in FHS/MDCS were of similar age to JHS but with fewer females (49-53%), and lower prevalence of treated hypertension, obesity, and diabetes mellitus.

pQTLs in the ABO locus

Table 1 shows the 56 proteins that associate with variants within 1MB of the transcription start site (TSS) of the *ABO* gene in either JHS or FHS/MDCS or both at a p-value < 5×10^{-8}. Such variants are termed protein quantitative trait loci (pQTLs). Twenty-three proteins had significant genetic associations in both black and white subjects, while 15 were specific to JHS and 18 were specific to the FHS/MDCS. As might be expected given the *ABO* region's known association with thrombosis,[12] proteins associated with variants in this locus across all cohorts included ADAMTS13, von Willebrand Factor (vWF), Tie-1, Angiopoetin-1 receptor, VEGFR-2 and VEGFR-3. Inflammatory mediators, including P-selectin and E-

selectin, Immunoglobulin superfamily containing leucine-rich repeat protein 2, and FAM3D (which has known cytokine activity) were also observed. Strikingly, CD209 antigen/DC-SIGN, which is the known binding site for multiple viruses including SARS-CoV, and a theorized binding site for SARS-CoV-2,[13] showed a strong association in both white and black cohorts. The specific variant most strongly associated with levels differed between JHS and FHS/MDCS. To demonstrate the specificity of the aptamer for DC-SIGN protein, we separately identified 47 variants within the gene encoding DC-SIGN protein (on chromosome 19) that associated with DC-SIGN protein levels in JHS at genome wide significance (p < 5×10^{-8}). Supplementary Table 1 summarizes data that similarly support aptamer specificity for all proteins described herein, using the presence of such variants at or near the gene coding for the target protein that also associate with measured protein levels (termed cis-pQTLs), mass spectrometry data, or immunoassay data.

Some protein associations were observed only in one racial group or the other. Among JHS participants, pQTLs in the *ABO* locus were associated with multiple inflammatory proteins including Cytotoxic T-lymphocyte protein 4, Interleukin-13, Interleukin-6 receptor subunit beta, Immunoglobulin alpha Fc receptor, T-cell surface glycoprotein CD4, and Programmed cell death 1 ligand 2. No pQTLs were identified for these proteins in FHS/MDCS. On the other hand, in FHS/MDCS but not JHS, variants in the *ABO* locus were associated with multiple adhesion molecules including Intercellular adhesion molecule 2, Neural cell adhesion molecule L1, Intercellular adhesion molecule 5, and Junctional adhesion molecule B.

Proteins associated with rs657152

Given the above findings, we examined the specific variant identified by Ellinghaus et al., rs657152, which was the sentinel SNP at the *ABO* locus, having the strongest association with respiratory failure in COVID-19. Table 2 shows that this variant is a pQTL for a subset of the proteins associated with the larger locus described above. Notably, CD209 antigen/DC-SIGN, Basal Cell Adhesion Molecule,

FAM3D, and vWF were all associated with this variant in both cohorts. For each of these four proteins, the effect allele that conferred higher risk for respiratory failure, A, was associated with higher measured levels of protein.

pQTLs in the chr3:45800446-46135604 locus

Given the multiple genes spanned by the other susceptibility locus, on chromosome 3, we similarly looked for pQTLs in this locus. Table 3 shows the proteins associated with variants in this region. Two proteins were found to have pQTLs in this locus across all cohorts: C-X-C motif chemokine 16 (CXCL16) and Teratocarcinoma growth factor 1. The *TDGF1* gene is near this locus making this a cis-pQTL. Further, while not a true cis relationship, CXCL16 is the ligand for CXCR6, whose gene is within this locus. No proteins were significantly associated with the specific variant identified by Ellinghaus et al., rs11385942.

Finally, we examined whether circulating levels of CXCL16 or DC-SIGN are associated with known risk factors for COVID-19 in JHS using unadjusted associations. While CXCL16 showed modest associations with age, sex, BMI, smoking, and renal function, DC-SIGN only showed a weak association with coronary disease (Table 4).

Discussion

The COVID-19 pandemic is a continually evolving public health crisis, and the biological mechanisms that confer the heterogeneous outcomes of infection remain unclear. Given the recently identified COVID-19 risk loci, our data identify several possible pathways by which these loci might confer risk in COVID-19.

The *ABO* locus is highly pleiotropic in our pQTL data, being associated with the levels of 56 proteins across the black and white cohorts. This likely reflects, in part, its role as a glycosyltransferase, altering the overall structure of multiple glycoproteins. An association between ABO blood group and disease is

emerging in COVID-19. At least one published study observed a higher proportion of type A blood among individuals hospitalized with COVID-19,[14] and unpublished data from China and the United States suggests that blood group A is associated with increased risk of acquiring COVID-19. In China, an increased proportion of Type A blood was observed among those with COVID-19 as compared with local controls.[8] Zietz and Tatonetti found similar results among patients from New York Presbyterian Hospital, and meta-analyzed with the data from China to confirm the association.[7] Most recently, the *ABO* locus was shown to be a risk locus for COVID-19 severity; in a meta-analysis of 1,610 cases of COVID-19 and respiratory failure and 2,205 COVID-19 cases without respiratory failure across a population of patients from Spain and Italy, Ellinghaus et al. observed multiple variants in the *ABO* gene locus that conferred increased risk.[3] They, and others, have speculated about the possible ways in which ABO might play a role in COVID-19. The thrombotic risk associated with this locus is one plausible element, and indeed we observe the well-known association of this locus with vWF levels.[12,15] Others suggest that blood type associates with ACE2 levels.[16] Still others hypothesized that circulating anti-A antibodies in individuals with Type O or B blood might also confer some level of immunity to COVID-19.[14,17] Our data, however, strongly suggest that *ABO* influences CD209 antigen/DC-SIGN, which is a known binding site for SARS-CoV.[4] DC-SIGN is expressed by dendritic cells, and it has been shown that the SARS-CoV spike protein utilizes this protein for cell entry, as do HIV and dengue virus.[4] Interestingly, DC-SIGN is thought to increase with age, and pre-print data suggest that it is also increased in smokers, associating it with two risk factors for COVID-19,[18] though we did not observe these associations in our JHS data. Additionally, we did not observe an association with ACE2 levels measured on the SOMAScan[TM] platform. The association between *ABO* and DC-SIGN is consistent with data from other GWAS studies of the human proteome,[10,11] and we now show the association in a black population. The frequency of the risk allele is slightly higher in JHS, one potential (though likely modest) reason for racial differences in COVID-19. Taken together, these data suggest that the *ABO* locus may confer its risk in part by modulating DC-SIGN, a putative binding site for SARS-CoV-2 cell entry.

Our data further suggest that the *ABO* locus may influence disease through pleiotropic effects, which may differ between black and white individuals. We show that *ABO* variants are associated with proteins involved in endothelial function and thrombosis, important complicating factors in COVID-19.[19] However, there were key differences between the cohorts. In JHS, levels of multiple inflammatory proteins were associated with *ABO* variants, whereas in FHS/MDCS, *ABO* variation was associated with proteins involved in cellular adhesion. These differences may mediate increased risk, perhaps for inflammatory complications of COVID-19, such as cytokine storm.[20]

We also observed interesting associations at the other risk locus identified by Ellinghaus et al., at which the sentinel effect allele conferred higher risk in their analysis, compared to *ABO*.[3] The validity of the 3p21.31 locus is noted by them to be supported by the Covid-19 Host Genetics Consortium data, which also showed increased COVID-19 risk, albeit at a lower level of significance.[21] In our analysis, two key proteins emerged, CXCL16 and Teratocarcinoma growth factor 1. TDGF-1 is a signaling protein, and has an unclear relationship to infection or propagation of respiratory disease. CXCL16, on the other hand, is the chemokine ligand for the CXCR6 receptor, whose coding gene is at the chromosome 3 locus of interest. This makes the *CXCR6* gene a strong candidate for further investigation among the large number of genes at this locus. Indeed CXCL16/CXCR6 has been implicated in LPS-induced acute lung injury and alveolar inflammation previously.[5,22]

Limitations

Our study has several important limitations. We have not studied the relationship between these proteins and genes in COVID-19 cases, thus confirming their role in pathogenesis requires further study. While the proteomic profiling discussed here is extensive, it does not cover the entire proteome, and important protein associations may be missed as a result. Further, while changes in aptamer binding are typically reflective of protein levels, it may be the case that ABO-mediated glycosylation alters aptamer binding, thus changing the measurement of the protein without a true change in protein levels. Nonetheless, our

data demonstrate that ABO affects the identified proteins, motivating further investigation. Additionally, in the case of DC-SIGN (and CXCL16), we and others have found that variants at the cognate gene are associated with levels of the protein as measured by the aptamer (Supplementary Table 1). Finally, we acknowledge we do not observe an association between the lead risk SNP at 3p21.31 identified by Ellinghaus et al. and CXCL16. Formal colocalization in larger cohorts will be useful in clarifying the gene-protein relationships.

Conclusions

We show here extensive proteomic profiling of genetic variation proposed to confer risk in COVID-19. We have shown in black and white cohorts that the *ABO* locus is associated with DC-SIGN, a putative binding site for SARS-CoV-2, suggesting ABO-mediated alteration of DC-SIGN may play a key role in disease pathogenesis. We further identify the CXCL16/CXCR6 pair as another potential disease mediator. Further study, specifically in COVID-19 infected patients, is needed to confirm these findings and determine whether levels or other modifications of these proteins could alter disease processes.

Methods

Study Approval

The human study protocols were approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center, Boston University Medical Center, Lund University, and University of Mississippi Medical Center, and all participants provided written informed consent.

Proteomic samples

The Jackson Heart Study is a community-based longitudinal cohort study begun in 2000 of 5306 self-identified African Americans from the Jackson, Mississippi metropolitan statistical area, the design of which is previously described.[23] Baseline characteristics were assessed at Visit 1 between 2000 and 2004.

Included in the present study are 1813 individuals with proteomic profiling and whole genome sequencing. Resting blood pressure was measured while sitting by recording two measurements with a Hawksley random zero sphygmomanometer using one of four cuff sizes selected by measured arm circumference. Glomerular filtration rate was estimated using the CKD-EPI equation.[24] Prevalent coronary heart disease (CHD) at Visit 1 was determined as a composite of patient reported angina, patient reported myocardial infarction, and evidence of previous myocardial infarction on ECG.

JHS plasma samples were collected at Visit 1 in EDTA tubes maintained in -70°C freezers.[25] Proteomic measurements were performed using SOMAscan™, a single-stranded DNA aptamer-based proteomics platform, which contained 1,305 aptamers.[26] Samples were run in three separate batches for cost efficiency. Batch 1 was run as a nested case-control study of incident coronary disease, excluding those with prevalent coronary heart disease at Visit 1. Batches 2 and 3 were a randomly selected sample of the remaining JHS participants. Each batch was divided into several plates containing a subset of the samples.

The FHS Offspring study has been previously described.[27,28] Included in the present study are 1625 individuals with proteomic profiling and genotyping performed at Visit 5. Proteomic profiling in FHS was also performed on the SOMAscan™ platform. In FHS, plasma samples were collected in citrate-treated tubes, which were then centrifuged within 15 minutes at 2000 g for 10 minutes and the supernatant plasma was aliquoted and stored at −80°C without freeze thaw cycles until assayed. This was completed in 2 different batches. In batch one, 1129 proteins (1.1k) were profiled in 695 individuals. As a result of platform enhancements that occurred in the interval between the first and second set of samples being run, batch 2 included an expanded panel of 1305 proteins (1.3k), which was assayed in 930 participants.

The MDCS is a Swedish population-based, prospective, observational cohort recruited between 1991 and 1996.[29] Included in the present study are 1421 individuals with proteomics and genotyping. Proteomic profiling in MDCS was also performed on the SOMAscan™ 1.3k platform as above, except samples were collected in EDTA-treated tubes. All assays in all cohorts were performed using SOMAscan™ reagents according to the manufacturer's detailed protocol.[30]

Genotyping and Imputation

Whole genome sequencing (WGS) in JHS has been described previously.[31] JHS participants underwent 30× WGS through the Trans-Omics for Precision Medicine (TOPMed) project at the Northwest Genome Center at University of Washington; genotype calling was performed by the Informatics Resource Center at the University of Michigan.

Genome-wide genotyping methods for the FHS have been described previously.[32] Briefly, genotyping was conducted using the Affymetrix 500K mapping array and the Affymetrix 50K gene-focused MIP supplemental array. Genotypes were called using Chiamo (http://www.stats.ox.ac.uk/~marchini/software/gwas/chiamo.html). We used the 1000 Genomes Phase I version 3 (August 2012) reference panel to perform imputation using a hidden Markov model implemented in MaCH (version 1.0.16)[33] for all SNPs passing the following criteria: call rate $\geq$ 97%, pHWE $\geq 1 \times 10^{-6}$, Mishap P $\geq 1 \times 10^{-9}$, Mendel errors $\leq$ 100, and MAF $\geq$ 1%.

In the MDCS, genotyping was conducted using the Illumina Omni Express Exome BeadChip kit. Genotypes were called using Illumina GenomeStudio and imputation performed to the same 1000 Genomes version as for FHS using IMPUTE (v2) for SNPs passing the following criteria: call rate $\geq$ 95%, pHWE $\geq 1 \times 10^{-6}$, minor allele frequency $\geq$ 0.01.

Statistical analysis

In JHS, because proteomic data varied by batch, measurements were first standardized to a set of control samples that were part of each plate. Because of the non-normal distribution of the resulting protein levels, age, sex, and batch adjusted residuals were generated and inverse normalized. The association between these values and genetic variants was tested using linear mixed effects models adjusted for age, sex and the genetic relationship matrix to adjust for relatedness using the fastGWA model implemented in the GCTA software package.[34] Variants with a minor allele count less than 5 were excluded.

In FHS and MDCS, inverse normalized transformed values of protein levels were also used. The association of genetic variants and protein levels were tested using linear mixed effects models to accommodate pedigree structure under an additive genetic model, adjusted for age and sex. Genome-wide association analyses were performed using the R GWAF package.[35]

Association analyses were limited to within 1MB of the transcription start site (TSS) of the ABO locus on chromosome 9 as well as the locus on chromosome 3 identified by Ellinghaus et al.: chr3:45800446-46135604 in Build hg38. Since FHS loci are identified by Build hg37 location, the region was shifted using the sentinel SNP rs11385942 as the reference point to chr3:45841939-46177097 in Build hg37. As we had hypothesis-driven genomic loci of interest, statistical significance for protein quantitative trait loci (pQTLs) was kept at a level of 5×10^{-8} despite multiple comparisons.

To determine associations between clinical variables and DC-SIGN or CXCL16, we used Pearson correlations for continuous variables and logistic regression for dichotomous variables. All protein levels were log-transformed and scaled by batch to normalize the data and reduce batch effects.

Disclaimer

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services

References

1. Gross, C. P. *et al.* Racial and Ethnic Disparities in Population Level Covid-19 Mortality. *medRxiv* 2020.05.07.20094250 (2020) doi:10.1101/2020.05.07.20094250.

2. Mahajan, U. V. & Larkins-Pettigrew, M. Racial demographics and COVID-19 confirmed cases and deaths: a correlational analysis of 2886 US counties. *J. Public Health Oxf. Engl.* (2020) doi:10.1093/pubmed/fdaa070.

3. Ellinghaus, D. *et al.* The ABO blood group locus and a chromosome 3 gene cluster associate with SARS-CoV-2 respiratory failure in an Italian-Spanish genome-wide association analysis. *medRxiv* 2020.05.31.20114991 (2020) doi:10.1101/2020.05.31.20114991.

4. Yang, Z.-Y. *et al.* pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN. *J. Virol.* **78**, 5642–5650 (2004).

5. Tu, G.-W. *et al.* CXCL16/CXCR6 is involved in LPS-induced acute lung injury via P38 signalling. *J. Cell. Mol. Med.* **23**, 5380–5389 (2019).

6. Driggin, E. *et al.* Cardiovascular Considerations for Patients, Health Care Workers, and Health Systems During the COVID-19 Pandemic. *J. Am. Coll. Cardiol.* **75**, 2352–2371 (2020).

7. Zietz, M. & Tatonetti, N. P. Testing the association between blood type and COVID-19 infection, intubation, and death. *medRxiv* 2020.04.08.20058073 (2020) doi:10.1101/2020.04.08.20058073.

8. Zhao, J. *et al.* Relationship between the ABO Blood Group and the COVID-19 Susceptibility. *medRxiv* 2020.03.11.20031096 (2020) doi:10.1101/2020.03.11.20031096.

9. Benson, M. D. *et al.* The Genetic Architecture of the Cardiovascular Risk Proteome. *Circulation* **137**, 1158–1172 (2018).

10. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).

11. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).

12. Vasan Senthil K. *et al.* ABO Blood Group and Risk of Thromboembolic and Arterial Disease. *Circulation* **133**, 1449–1457 (2016).

13. Brufsky, A. & Lotze, M. T. DC/L-SIGNs of Hope in the COVID-19 Pandemic. *J. Med. Virol.* (2020) doi:10.1002/jmv.25980.

14. Li, J. *et al.* Association between ABO blood groups and risk of SARS-CoV-2 pneumonia. *Br. J. Haematol.* (2020) doi:10.1111/bjh.16797.

15. Franchini, M., Capra, F., Targher, G., Montagnana, M. & Lippi, G. Relationship between ABO blood group and von Willebrand factor levels: from biology to clinical implications. *Thromb. J.* **5**, 14 (2007).

16. Dai, X. ABO blood group predisposes to COVID-19 severity and cardiovascular diseases. *Eur. J. Prev. Cardiol.* 2047487320922370 (2020) doi:10.1177/2047487320922370.

17. Gérard, C., Maggipinto, G. & Minon, J.-M. COVID-19 and ABO blood group: another viewpoint. *Br. J. Haematol.* (2020) doi:10.1111/bjh.16884.

18. Cai, G., Cui, X., Zhu, X. & Zhou, J. A Hint on the COVID-19 Risk: Population Disparities in Gene Expression of Three Receptors of SARS-CoV. (2020) doi:10.20944/preprints202002.0408.v1.

19. Levi, M., Thachil, J., Iba, T. & Levy, J. H. Coagulation abnormalities and thrombosis in patients with COVID-19. *Lancet Haematol.* **7**, e438–e440 (2020).

20. Jose, R. J. & Manuel, A. COVID-19 cytokine storm: the interplay between inflammation and coagulation. *Lancet Respir. Med.* **0**, (2020).

21. 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. *Eur. J. Hum. Genet. EJHG* **28**, 715–718 (2020).

22. Agostini, C. *et al.* Role for CXCR6 and its ligand CXCL16 in the pathogenesis of T-cell alveolitis in sarcoidosis. *Am. J. Respir. Crit. Care Med.* **172**, 1290–1298 (2005).

23. Taylor, H. A. *et al.* Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn. Dis.* **15**, S6-4–17 (2005).

24. Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).

25. Carpenter, M. A. *et al.* Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. *Am. J. Med. Sci.* **328**, 131–144 (2004).

26. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PloS One* **5**, e15004 (2010).

27. Nayor Matthew *et al.* Aptamer-Based Proteomic Platform Identifies Novel Protein Predictors of Incident Heart Failure and Echocardiographic Traits. *Circ. Heart Fail.* **13**, e006749 (2020).

28. Kannel, W. B., Feinleib, M., McNamara, P. M., Garrison, R. J. & Castelli, W. P. An Investigation of Coronary Heart Disease in Families: The Framingham Offspring Study. *Am. J. Epidemiol.* **110**, 281–290 (1979).

29. Berglund, G., Elmstähl, S., Janzon, L. & Larsson, S. A. The Malmo Diet and Cancer Study. Design and feasibility. *J. Intern. Med.* **233**, 45–51 (1993).

30. Candia, J. *et al.* Assessment of Variability in the SOMAscan Assay. *Sci. Rep.* **7**, 14248 (2017).

31. Raffield Laura M. *et al.* D-Dimer in African Americans. *Arterioscler. Thromb. Vasc. Biol.* **37**, 2220–2227 (2017).

32. Wilk, J. B. *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.* **5**, e1000429 (2009).

33. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).

34. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51**, 1749–1755 (2019).

35. Chen, M.-H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580–581 (2010).

Table 1. Proteins with pQTLs in the *ABO* Locus

| SomaId | Target Full Name | JHS | | | | | | FHS/MDCS | | | | | |
		rsID	Effect AF	Distance to ABO TSS	BETA	SE	P	rsID	Effect AF	Distance to ABO TSS	BETA	SE	P
SL002081	Cadherin-5	rs8176722	0.133	17847	1.22	0.05	2.60E-135	rs8176741	0.92	19144	1.11	0.05	1.10E-129
SL006610	A disintegrin and metalloproteinase with thrombospondin motifs 13	rs36090624	0.063	165217	-1.69	0.07	1.06E-134	rs28647808	0.91	154925	1.02	0.04	3.28E-128
SL003199	Tyrosine-protein kinase receptor Tie-1	rs8176722	0.133	17847	1.13	0.05	3.24E-118	rs10793962	0.92	21490	-1	0.05	3.60E-97
SL005157	CD209 antigen/DC-SIGN	rs8176719	0.291	17693	0.80	0.04	3.38E-106	rs8176663	0.67	6178	-0.84	0.02	0
SL003200	Angiopoietin-1 receptor	rs8176722	0.133	17847	0.93	0.05	3.44E-81	rs8176741	0.92	19144	0.81	0.05	5.60E-66
SL001945	E-selectin	rs992108547	0.902	1231	1.06	0.06	6.01E-78	rs532436	0.82	775	-1.04	0.03	0
SL012769	Protein FAM3D	rs879055593	0.844	4032	-0.77	0.05	2.01E-61	rs550057	0.75	4008	0.82	0.03	2.80E-177
SL018587	Immunoglobulin superfamily containing leucine-rich repeat protein 2	rs8176722	0.133	17847	0.70	0.05	1.77E-46	rs2519093	0.82	8735	-0.56	0.04	1.10E-54
SL001902	Basal Cell Adhesion Molecule	rs8176722	0.133	17847	0.66	0.05	2.30E-44	rs75179845	0.93	17651	-0.44	0.05	2.60E-18
SL005223	Interleukin-27 receptor subunit alpha	rs8176722	0.133	17847	0.66	0.05	1.48E-41	rs7470777	0.91	35429	-0.47	0.05	1.20E-18
SL014248	Semaphorin-6B	rs8176722	0.133	17847	0.66	0.05	6.40E-41	rs4424335	0.90	37046	-0.51	0.06	1.30E-19
SL004125	Insulin receptor	rs977371848	0.903	8758	0.73	0.06	5.31E-38	rs2519093	0.82	8735	-0.7	0.03	2.50E-111
SL000017	von Willebrand factor	rs8176719	0.291	17693	0.39	0.03	4.10E-29	rs687289	0.68	13499	0.41	0.03	7.00E-57
SL000560	P-selectin	rs635634	0.901	4213	0.55	0.06	1.94E-22	rs2519093	0.82	8735	-0.53	0.03	7.90E-60
SL004482	Endoglin	rs8176722	0.133	17847	0.45	0.05	3.41E-20	rs2519093	0.82	8735	-0.43	0.03	7.80E-39
SL005703	Neurogenic locus notch homolog protein 1	rs8176722	0.133	17847	0.42	0.05	4.13E-18	rs8176644	0.92	1455	0.39	0.05	5.70E-15
SL004714	Leukemia inhibitory factor receptor	rs8176722	0.133	17847	0.37	0.05	2.93E-14	rs2519093	0.82	8735	-0.35	0.04	5.30E-21
SL003322	Vascular endothelial growth factor receptor 3	rs977371848	0.903	8758	0.43	0.06	5.77E-14	rs532436	0.82	775	-0.58	0.03	1.20E-73
SL000134	Hepatocyte growth factor receptor	rs992108547	0.902	1231	0.4	0.06	1.58E-12	rs2519093	0.82	8735	-0.37	0.05	1.40E-14
SL014268	OX-2 membrane glycoprotein	rs977371848	0.903	8758	0.38	0.05	5.02E-12	rs507666	0.82	1206	-0.36	0.03	3.10E-27
SL008709	Desmocollin-3	rs142596951	0.001	50564	2.05	0.3	1.12E-11						
SL003201	Vascular endothelial growth factor receptor 2	rs1381383189	0.069	11852	-0.44	0.07	3.35E-11	rs532436	0.82	775	-0.52	0.03	5.50E-58
SL004625	A disintegrin and metalloproteinase with thrombospondin motifs 4	rs1019902671	0.001	214677	1.4	0.22	1.16E-10						
SL018625	Toll-like receptor 4:Lymphocyte antigen 96 complex	rs635634	0.901	4213	0.36	0.06	1.82E-10	rs635634	0.82	4395	-0.31	0.03	1.30E-19
SL001753	Sialoadhesin	rs957365367	0.002	979	1.12	0.18	1.24E-09						
SL008590	Olfactomedin-4	rs146749579	0.001	964494	1.66	0.28	2.78E-09						
SL005160	Endothelial cell-selective adhesion molecule	rs8176722	0.133	17847	0.28	0.05	6.35E-09						
SL007284	Cysteine-rich secretory protein 3	rs548435692	0.002	779015	1.51	0.26	6.69E-09						
SL002922	Intercellular adhesion molecule 1	rs1381383189	0.069	11852	-0.39	0.07	9.51E-09	rs635634	0.82	4395	-0.19	0.03	1.80E-08

		SNP	Freq	N	Beta	SE	P-value	SNP	Freq	N	Beta	SE	P-value
SL007361	Galectin-10	rs541531301	0.001	787328	1.92	0.34	1.21E-08						
SL000384	Cytotoxic T-lymphocyte protein 4	rs571481216	0.004	339063	1.43	0.25	1.26E-08						
SL001718	Interleukin-13	rs28551042	0.156	135050	-0.24	0.04	1.43E-08						
SL004690	WNT1-inducible-signaling pathway protein 3	rs28568310	0.510	854709	-0.18	0.03	1.60E-08						
SL003872	Interleukin-6 receptor subunit beta	rs8176722	0.133	17847	0.28	0.05	1.84E-08						
SL011498	Pituitary adenylate cyclase-activating polypeptide 38	rs925632408	0.001	280707	1.72	0.31	2.03E-08						
SL010373	Immunoglobulin alpha Fc receptor	rs551079828	0.001	257980	2.06	0.37	2.05E-08						
SL002524	T-cell surface glycoprotein CD4	rs899928301	0.001	298436	1.27	0.23	3.88E-08						
SL004862	Programmed cell death 1 ligand 2	rs8176722	0.133	17847	0.27	0.05	4.26E-08						
SL000668	Platelet glycoprotein 4							rs2519093	0.82	8735	-0.52	0.03	1.90E-59
SL007502	Carbohydrate sulfotransferase 15							rs550057	0.75	4008	-0.38	0.03	3.40E-39
SL003177	Intercellular adhesion molecule 2							rs507666	0.82	1206	-0.46	0.04	4.50E-35
SL004516	Mannose-binding protein C							rs597988	0.71	6321	0.3	0.03	1.30E-24
SL008773	CD109 antigen							rs2519093	0.82	8735	-0.31	0.03	2.70E-20
SL003304	Insulin-like growth factor 1 receptor							rs550057	0.75	4008	-0.27	0.03	4.00E-20
SL007328	Protein jagged-1							rs2519093	0.82	8735	-0.29	0.03	2.40E-18
SL007306	Protein FAM3B							rs550057	0.75	4008	0.28	0.03	2.20E-17
SL004154	Neural cell adhesion molecule L1							rs635634	0.82	4395	-0.28	0.03	2.40E-16
SL005169	Intercellular adhesion molecule 5							rs2519093	0.82	8735	-0.26	0.03	2.00E-14
SL005214	Semaphorin-6A							rs550057	0.75	4008	-0.19	0.03	1.80E-10
SL008504	N-acetylglucosamine-6-sulfatase							rs507666	0.82	1206	-0.21	0.03	2.50E-10
SL003303	C-C motif chemokine 28							rs2519093	0.82	8735	-0.21	0.03	3.00E-10
SL004337	Fibroblast growth factor 19							rs4357365	0.87	36771	0.28	0.05	4.00E-09
SL005193	Junctional adhesion molecule B							rs8176672	0.92	8420	0.29	0.05	6.50E-09
SL006523	Lactadherin							rs201379627		34280	0.34	0.06	1.60E-08
SL003060	Fibroblast growth factor receptor 1							rs2519093	0.82	8735	-0.21	0.04	1.90E-08
SL015046	Amphoterin-induced protein 2							rs550057	0.75	4008	-0.18	0.03	3.50E-08

Proteins associated with any variant within 1MB of transcription start site of *ABO* gene in either Jackson Heart Study (JHS) or Framingham Heart Study (FHS)/Malmö Diet and Cancer Study (MDCS). The SNP with lowest p-value for association with that protein in that cohort is displayed for simplicity. P values of 0 are reported where the value is below the limit of the software.

Table 2. Proteins Associated with rs657152

SomaId	Target Full Name	JHS				FHS/MDCS			
		A Allele Frequency	BETA	SE	P	A Allele Frequency	BETA	SE	P
SL005157	CD209 antigen/DC-SIGN	0.42	0.50	0.03	1.77E-51	0.35	0.79	0.02	1.70E-271
SL003199	Tyrosine-protein kinase receptor Tie-1	0.42	0.45	0.03	1.19E-41				
SL002081	Cadherin-5	0.42	0.36	0.03	7.38E-27				
SL001945	E-selectin	0.42	-0.33	0.03	3.57E-22				
SL001902	Basal Cell Adhesion Molecule	0.42	0.29	0.03	4.73E-19	0.35	0.17	0.03	2.20E-10
SL003200	Angiopoietin-1 receptor	0.42	0.28	0.03	2.56E-17				
SL012769	Protein FAM3D	0.42	0.25	0.03	6.84E-14	0.35	0.61	0.03	1.60E-108
SL014248	Semaphorin-6B	0.42	0.23	0.03	6.62E-12				
SL000017	von Willebrand factor	0.42	0.22	0.03	8.20E-12	0.35	0.38	0.03	5.40E-49
SL005223	Interleukin-27 receptor subunit alpha	0.42	0.18	0.03	4.29E-08				
SL004125	Insulin receptor					0.35	-0.38	0.03	1.10E-48
SL000560	P-selectin					0.35	-0.33	0.03	1.10E-35
SL000668	Platelet glycoprotein 4					0.35	-0.31	0.03	3.20E-34
SL003322	Vascular endothelial growth factor receptor 3					0.35	-0.32	0.03	5.30E-33
SL003201	Vascular endothelial growth factor receptor 2					0.35	-0.27	0.03	2.30E-25
SL004516	Mannose-binding protein C					0.35	0.26	0.03	4.00E-23
SL003177	Intercellular adhesion molecule 2					0.35	-0.28	0.03	3.40E-20
SL007306	Protein FAM3B					0.35	0.23	0.03	1.90E-14
SL014268	OX-2 membrane glycoprotein					0.35	-0.19	0.03	2.90E-13
SL018587	Immunoglobulin superfamily containing leucine-rich repeat protein 2					0.35	-0.19	0.03	2.70E-10
SL004482	Endoglin					0.35	-0.16	0.03	1.50E-09
SL003303	C-C motif chemokine 28					0.35	-0.15	0.03	7.90E-09
SL008504	N-acetylglucosamine-6-sulfatase					0.35	-0.15	0.03	1.50E-08
SL003304	Insulin-like growth factor 1 receptor					0.35	-0.15	0.03	2.60E-08
SL004154	Neural cell adhesion molecule L1					0.35	-0.15	0.03	2.70E-08
SL000134	Hepatocyte growth factor receptor					0.35	-0.2	0.04	3.10E-08
SL005169	Intercellular adhesion molecule 5					0.35	-0.15	0.03	3.30E-08

Proteins with association with rs657152 in either Jackson Heart Study (JHS) or Framingham Heart Study (FHS)/Malmö Diet and Cancer Study (MDCS) meta-analysis. All Beta estimates are for the presence of the A allele.

Table 3. Proteins with pQTLs in the *chr3:45800446-46135604* Locus of hg38

		JHS					FHS/MDCS				
SomaId	Target Full Name	rsID	Effect AF	BETA	SE	P	rsID	Effect AF	BETA	SE	P
SL004016	C-X-C motif chemokine 16	rs2234355	0.44	0.5	0.03	8.78E-50	rs4396933	0.69	0.18	0.03	6.50E-09
SL005155	Teratocarcinoma-derived growth factor 1	rs1546079	0.11	0.47	0.05	2.20E-18	rs13434029	0.67	-0.28	0.03	5.49E-19
SL000076	Cyclin-dependent kinase inhibitor 1B	rs9870445	0.86	0.31	0.04	5.49E-12					
SL004739	Inter-alpha-trypsin inhibitor heavy chain H4	rs1860264	0.83	-0.25	0.04	1.90E-08					
SL007361	Galectin-10	rs373364694	0.001	1.89	0.34	2.01E-08					
SL004356	C-C motif chemokine 4-like						rs1500004	0.88	0.55	0.05	5.40E-28

Proteins associated with any variant within *chr3:45800446-46135604* of hg38 in either Jackson Heart Study (JHS) or Framingham Heart Study (FHS)/Malmö Diet and Cancer Study (MDCS) meta-analysis. The SNP with lowest p-value for association with that protein in that cohort is displayed for simplicity. As GWAS in FHS/MDCS was performed using hg37, the locus has been translated to *chr3:45841939-46177097*.

Table 4. Associations between clinical risk factors and protein levels of DC-SIGN and CXCL16 in the Jackson Heart Study

Variable	*DC-SIGN*		*CXCL16*	
	Coefficient	P Value	Coefficient	P Value
Age (y)	-0.03	0.2	0.08	0.0002*
Gender (male)	-0.06	0.14	-0.29	2.00E-11*
Body Mass Index (m^2/kg)	-0.03	0.17	-0.19	<2E-16*
Systolic Blood pressure (mmHg)	-0.003	0.88	0.26	0.24
HbA1c(%)	-0.01	0.57	0.02	0.25
Estimated glomerular filtration rate (mL/min/1.73m2)	0.02	0.27	-0.16	2.30E-14*
Total Cholesterol (mg/dL)	0.03	0.12	-0.02	0.45
Coronary Heart Disease History	0.09	0.02*	0.09	0.70
Current Smoker	0.01	0.89	0.14	0.02*

Unadjusted associations shown, using Pearson correlations for continuous variables and logistic regression for dichotomous variables. All protein levels were log-transformed and scaled by batch to normalize the data and reduce batch effects. *Highlights p-values <0.05

Supplementary Table 1. Data Sources Corroborating Aptamer Specificity

SomaId	Target Full Name	Emillson	Butterworth	JHS	FHS	Suhre	Emillson MS MRM	Emillson MS DDA	OLINK
SL002081	Cadherin-5				X			X	X
SL006610	A disintegrin and metalloproteinase with thrombospondin motifs 13	X	X	X	X	X		X	
SL003199	Tyrosine-protein kinase receptor Tie-1								X
SL005157	CD209 antigen	X	X	X	X	X			
SL003200	Angiopoietin-1 receptor	X	X		X	X			
SL001945	E-selectin	X			X		X		X
SL012769	Protein FAM3D	X	X	X	X				
SL018587	Immunoglobulin superfamily containing leucine-rich repeat protein 2	X	X	X	X				
SL001902	Basal Cell Adhesion Molecule	X	X		X				
SL005223	Interleukin-27 receptor subunit alpha	X	X	X	X				
SL014248	Semaphorin-6B	X			X				
SL004125	Insulin receptor								
SL000017	von Willebrand factor	X			X			X	X
SL000560	P-selectin	X	X	X	X	X			X
SL004482	Endoglin	X			X	X			
SL005703	Neurogenic locus notch homolog protein 1				X	X			
SL004714	Leukemia inhibitory factor receptor	X			X				
SL003322	Vascular endothelial growth factor receptor 3	X	X	X	X				
SL000134	Hepatocyte growth factor receptor	X			X	X	X		
SL014268	OX-2 membrane glycoprotein	X			X	X			
SL008709	Desmocollin-3			X					
SL003201	Vascular endothelial growth factor receptor 2	X	X	X	X	X			
SL004625	A disintegrin and metalloproteinase with thrombospondin motifs 4				X				
SL018625	Toll-like receptor 4:Lymphocyte antigen 96 complex				X	X	X		
SL001753	Sialoadhesin				X	X	X		
SL008590	Olfactomedin-4								
SL005160	Endothelial cell-selective adhesion molecule	X	X		X	X		X	X
SL007284	Cysteine-rich secretory protein 3								
SL002922	Intercellular adhesion molecule 1	X	X	X	X	X			
SL007361	Galectin-10								
SL000384	Cytotoxic T-lymphocyte protein 4								

SL ID	Protein							
SL001718	Interleukin-13							
SL004690	WNT1-inducible-signaling pathway protein 3			X				
SL003872	Interleukin-6 receptor subunit beta	X	X	X	X	X	X	
SL011498	Pituitary adenylate cyclase-activating polypeptide 38							
SL010373	Immunoglobulin alpha Fc receptor							
SL002524	T-cell surface glycoprotein CD4	X		X	X			
SL004862	Programmed cell death 1 ligand 2	X	X	X	X	X		
SL000668	Platelet glycoprotein 4	X		X	X		X	
SL007502	Carbohydrate sulfotransferase 15	X		X				
SL003177	Intercellular adhesion molecule 2			X			X	
SL004516	Mannose-binding protein C	X	X	X	X	X	X	X
SL008773	CD109 antigen	X	X	X	X	X	X	X
SL003304	Insulin-like growth factor 1 receptor			X				
SL007328	Protein jagged-1			X				
SL007306	Protein FAM3B	X	X	X	X		X	X
SL004154	Neural cell adhesion molecule L1							
SL005169	Intercellular adhesion molecule 5	X	X	X	X	X		
SL005214	Semaphorin-6A		X	X				
SL008504	N-acetylglucosamine-6-sulfatase			X	X			
SL003303	C-C motif chemokine 28	X						
SL004337	Fibroblast growth factor 19						X	X
SL005193	Junctional adhesion molecule B			X				X
SL006523	Lactadherin	X	X	X	X	X	X	
SL003060	Fibroblast growth factor receptor 1			X			X	
SL015046	Amphoterin-induced protein 2	X		X	X		X	X
SL004016	C-X-C motif chemokine 16	X	X	X				X
SL005155	Teratocarcinoma-derived growth factor 1	X	X	X	X	X		
SL000076	Cyclin-dependent kinase inhibitor 1B			X				
SL004739	Inter-alpha-trypsin inhibitor heavy chain H4	X		X	X		X	
SL004356	C-C motif chemokine 4-like	X		X	X	X		

Listed above are the proteins significantly associated with either locus. Aptamer specificity for target proteins were confirmed/inferred via orthogonal methods (i.e. genetics, mass spectrometry, immunoassay). Genetic studies with cis-pQTLs in literature and unpublished data cited. The Olink Spearman Rho data was generated from 400 samples run on both SOMAscan Version 1.3 and Olink immunoassay (12 panels, unpublished data). Mass spectrometry data from previously published literature. * Genetic studies may not be able to determine aptamer protein isoform specificity.

Emilsson et. al., Science 2018 Aug 24;361(6404):769-773. doi: 10.1126/science.aaq1327

Sun et. al., Nature. 2018 Jun;558(7708):73-79. doi: 10.1038/s41586-018-0175-2

JHS whole genome association analysis (unpublished data)

FHS-MDC exome array and GWAS meta-analysis (unpublished data)

Suhre et. al., Nat Commun. 2017 Feb 27;8:14357. doi: 10.1038/ncomms14357.

Ngo et al. Circulation. 2016 Jul 26;134(4):270-85. doi: 10.1161/CIRCULATIONAHA.116.021803.