Title

A MUC5B Gene Polymorphism, rs35705950-T, Confers Protective Effects Against COVID-19 Hospitalization but Not Severe Disease or Mortality

Journal

American Journal of Respiratory and Critical Care Medicine, 206(10)

ISSN

1073-449X

Authors

Verma, Anurag
Minnier, Jessica
Wan, Emily S
et al.

Publication Date

2022-11-15

DOI

10.1164/rccm.202109-2166oc

Peer reviewed

ORIGINAL ARTICLE

A *MUC5B* Gene Polymorphism, rs35705950-T, Confers Protective Effects Against COVID-19 Hospitalization but Not Severe Disease or Mortality

Anurag Verma[1,2*], Jessica Minnier[3,4,5*], Emily S. Wan[6,7*], Jennifer E. Huffman[8], Lina Gao[4,5], Jacob Joseph[9,10], Yuk-Lam Ho[8], Wen-Chih Wu[11,12], Kelly Cho[8,13], Bryan R. Gorman[8], Nallakkandi Rajeevan[14,15], Saiju Pyarajan[8,16], Helene Garcon[8], James B. Meigs[17], Yan V. Sun[18,19], Peter D. Reaven[20,21], John E. McGeary[22,23], Ayako Suzuki[24,25], Joel Gelernter[26,27], Julie A. Lynch[28,29], Jeffrey M. Petersen[1,30], Seyedeh Maryam Zekavat[31,32], Pradeep Natarajan[16,32,33], Sharvari Dalal[1,30], Darshana N. Jhala[1,30], Mehrdad Arjomandi[34], Elise Gatsby[28], Kristine E. Lynch[28,35], Robert A. Bonomo[36], Matthew Freiberg[37], Gita A. Pathak[26,27], Jin J. Zhou[38,39], Curtis J. Donskey[40], Ravi K. Madduri[41], Quinn S. Wells[37,42,43], Rose D. L. Huang[8], Renato Polimanti[26,27], Kyong-Mi Chang[1], Katherine P. Liao[44], Philip S. Tsao[45], Peter W. F. Wilson[19,46], Adriana M. Hung[47], Christopher J. O'Donnell[48], John M. Gaziano[8], Richard L. Hauger[49,50], Sudha K. Iyengar[51,52‡], Shiuh-Wen Luoh[4,5‡], and the Million Veteran Program COVID-19 Science Initiative

Abstract

Rationale: A common *MUC5B* gene polymorphism, rs35705950-T, is associated with idiopathic pulmonary fibrosis (IPF), but its role in severe acute respiratory syndrome coronavirus 2 infection and disease severity is unclear.

Objectives: To assess whether rs35705950-T confers differential risk for clinical outcomes associated with coronavirus disease (COVID-19) infection among participants in the Million Veteran Program (MVP).

Methods: The *MUC5B* rs35705950-T allele was directly genotyped among MVP participants; clinical events and comorbidities were extracted from the electronic health records. Associations between the incidence or severity of COVID-19 and rs35705950-T were analyzed within each ancestry group in the MVP followed by transancestry meta-analysis. Replication and joint meta-analysis were conducted using summary statistics from the COVID-19 Host Genetics Initiative (HGI). Sensitivity analyses with adjustment for additional covariates (body mass index, Charlson comorbidity index, smoking, asbestosis, rheumatoid arthritis with interstitial lung disease, and IPF) and associations with post–COVID-19 pneumonia were performed in MVP subjects.

Measurements and Main Results: The rs35705950-T allele was associated with fewer COVID-19 hospitalizations in transancestry meta-analyses within the MVP ($N_{cases} = 4{,}325$; $N_{controls} = 507{,}640$; OR = 0.89 [0.82–0.97]; $P = 6.86 \times 10^{-3}$) and joint meta-analyses with the HGI ($N_{cases} = 13{,}320$; $N_{controls} = 1{,}508{,}841$; OR, 0.90 [0.86–0.95]; $P = 8.99 \times 10^{-5}$). The rs35705950-T allele was not associated with reduced COVID-19 positivity in transancestry meta-analysis within the MVP ($N_{cases} = 19{,}168$/$N_{controls} = 492{,}854$; OR, 0.98 [0.95–1.01]; $P = 0.06$) but was nominally significant ($P < 0.05$) in the joint meta-analysis with the HGI ($N_{cases} = 44{,}820$; $N_{controls} = 1{,}775{,}827$; OR, 0.97 [0.95–1.00]; $P = 0.03$). Associations were not observed with severe outcomes or mortality. Among individuals of European ancestry in the MVP, rs35705950-T was associated with fewer post–COVID-19 pneumonia events (OR, 0.82 [0.72–0.93]; $P = 0.001$).

Conclusions: The *MUC5B* variant rs35705950-T may confer protection in COVID-19 hospitalizations.

Keywords: coronavirus disease 2019; severe acute respiratory syndrome coronavirus 2; idiopathic pulmonary fibrosis; electronic health records; genetic association

(Received in original form September 21, 2021; accepted in final form June 30, 2022)

A complete list of Million Veteran Program COVID-19 Science Initiative members may be found in the online supplement.

*These authors contributed equally to this work.

‡These authors jointly supervised this work.

Am J Respir Crit Care Med Vol 206, Iss 10, pp 1220–1229, Nov 15, 2022

Originally Published in Press as DOI: 10.1164/rccm.202109-2166OC on June 30, 2022.
Internet address: www.atsjournals.org

[1]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, Pennsylvania; [2]Department of Medicine, Perelman School of Medicine, and [30]Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; [3]OHSU-PSU School of Public Health and [4]Knight Cancer Institute, Oregon Health & Science University, Portland, Oregon; [5]VA Portland Health Care System, Portland, Oregon; [6]Department of Medicine, Pulmonary, Critical Care, Sleep, and Allergy Section, [8]MAVERIC, [9]Department of Medicine, [44]Medicine, Rheumatology, and [48]Medicine, Cardiology, VA Boston Healthcare System, Boston, Massachusetts; [7]Channing Division of Network Medicine and [10]Medicine, Cardiovascular, Brigham & Women's Hospital, Boston, Massachusetts; [13]Medicine, Aging, Brigham & Women's Hospital and [16]Department of Medicine, Harvard Medical School, Boston, Massachusetts; [11]Department of Medicine, Cardiology, Providence VA Healthcare System, Providence, Rhode Island; [12]Alpert Medical School & School of Public Health, Brown University, Providence, Rhode Island; [14]Yale Center for Medical Informatics, [26]Division of Human Genetics, Department of Psychiatry, and [31]Computational Biology & Bioinformatics, Yale University School of Medicine, New Haven, Connecticut; [15]Clinical Epidemiology Research Center (CERC), [27]VA Connecticut Healthcare System, West Haven, Connecticut; [17]Medicine, General Internal Medicine and [33]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts; [18]Epidemiology, School of Public Health and [46]Emory University, Atlanta, Georgia; [19]Atlanta VA Healthcare System, Decatur, Georgia; [20]Department of Medicine, Phoenix VA Healthcare System, Phoenix, Arizona; [21]College of Medicine, University of Arizona, Phoenix, Arizona; [22]Department of Psychiatry and Human Behavior, Providence VA Medical Center, Providence, Rhode Island; [23]Department of Psychiatry and Human Behavior, Brown University Medical School, Providence, Rhode Island; [24]Department of Medicine, Gastroenterology, Durham VA Medical Center, Durham, North Carolina; [25]Department of Medicine, Gastroenterology, Duke University, Durham, North Carolina; [28]VA Informatics & Computing Infrastructure (VINCI), VA Salt Lake City Healthcare System, Salt Lake City, Utah; [29]Department of Medicine and [35]Internal Medicine, Epidemiology, University of Utah School of Medicine, Salt Lake City, Utah; [32]Program in Medical and Population Genetics, Cardiovascular Disease Initiative, Broad Institute of Harvard and MIT, Cambridge, Massachusetts; [34]Medicine, Pulmonary and Critical Care, San Francisco VA Healthcare System, University of California, San Francisco, San Francisco, California; [36]Department of Medicine and [51]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio; [37]Department of Medicine, [42]Department of Biomedical Informatics, and [43]Department of Pharmacology, Vanderbilt University Medical Center, Nashville, Tennessee; [38]Department of Medicine, University of California, Los Angeles, Los Angeles, California; [39]Epidemiology and Biostatistics, University of Arizona, Tucson, Arizona; [40]Infectious Disease Section and [52]Louis Stokes Cleveland VA Medical Center, Cleveland, Ohio; [41]Data Science and Learning, Argonne National Laboratory, Lemont, Illinois; [45]Precision Medicine, VA Palo Alto Health Care System, Palo Alto, California; [47]Department of Veteran's Affairs, Tennessee Valley Healthcare System, Vanderbilt University Medical Center, Division of Nephrology & Hypertension, Nashville, Tennessee; [49]Center of Excellence for Stress & Mental Health, VA San Diego Healthcare System, San Diego, California; and [50]Center for Behavioral Genetics of Aging, University of California, San Diego, La Jolla, California

ORCID ID: 0000-0002-0116-9217 (M.A.).

A respiratory disease caused by a novel coronavirus, now known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first reported near the end of 2019. Despite massive public health measures and vaccination initiatives, the coronavirus disease (COVID-19) pandemic remains a major global health threat. By February 2022, the COVID-19 pandemic had caused more than 424 million confirmed infections, resulting in more than 5.8 million deaths worldwide (1).

Parenchymal fibrosis is a late complication of severe respiratory infections due to COVID-19 (2–4). Among chronic lung diseases, idiopathic pulmonary fibrosis (IPF) (5), a disorder characterized by progressive pulmonary scarring that is associated with a median survival of 2–3 years in the absence of lung transplantation (6), shares several risk factors with those for severe COVID-19 disease, including advanced age (7), cardiovascular disease, diabetes, and a history of smoking (5). Thus, common pathological processes may be shared between the fibrotic response toward COVID-19 infection and those underlying IPF.

IPF likely develops from a multifaceted interaction between genetic and environmental factors, age-related mechanisms, and epigenetic profibrotic reprogramming (8, 9). One of the most robust genetic risk factors identified for IPF susceptibility is rs35705950-T, a common G → T transversion located approximately 3 kb upstream of the mucin 5B, oligomeric mucus/gel-forming (*MUC5B*) gene (10, 11). Laboratory evidence supports that rs35705950-T: *1*) is a functional variant located within an enhancer subject to epigenetic programming and *2*) contributes to pathologic misexpression in IPF (12). Notably, although rs35705950-T has been robustly associated with increased susceptibility, the same allele has also been associated with decreased mortality in IPF (13), though whether this paradoxical effect is attributable to pleiotropy or index event bias remains controversial (14).

Given the high minor allele frequency (MAF) of rs35705950-T (~11% among

Supported by an MVP035 award from the Million Veteran Program, Office of Research and Development, Veterans Health Administration; an MVP008 award (1I01BX004188-01) (S.-W.L.), Veterans Affairs of the United States Government BX 004831 (P.W.F.W. and K.C.); National Institutes of Health grant R01GM138597 (A.V.); and National Heart, Lung, and Blood Institute Grants R01HL142711 and R01HL127564 (P.N.). This publication does not represent the views of the Department of Veterans Affairs of the U.S. Government.

Author Contributions: A.V., J.E.H., S.K.I., S.-W.L., L.G., J.M., and E.S.W. analyzed data. J.E.H., A.V., L.G., E.S.W., S.K.I., and S.-W.L. supervised data collection. A.V., J.E.H., L.G., J.M., E.S.W., S.K.I., and S.-W.L. wrote the manuscript. A.V., J.E.H., L.G., J.M., E.S.W., S.K.I., and S.-W.L. conceived the study design, supervised data collection and analysis, and wrote and edited the manuscript. All authors revised and approved the final version to be published.

Correspondence and requests for reprints should be addressed to Shiuh-Wen Luoh, M.D., Ph.D., or Sudha K. Iyengar, Ph.D., 3710 SW US Veterans Hospital Road, Portland, OR 97239-2964. Email: Shiuh-Wen.Luoh@va.gov or ski@case.edu.

Some of the results of these studies have been previously reported in the form of a preprint (medRxiv [29 September 2021]; https://doi.org/10.1101/2021.09.28.21263911).

This article has a related editorial.

This article has a data supplement, which is accessible from this issue's table of contents online at www.atsjournals.org.

At a Glance Commentary

Scientific Knowledge on the Subject: Parenchymal fibrosis, a late feature of severe coronavirus disease (COVID-19), shares characteristics with idiopathic pulmonary fibrosis (IPF). The impact of rs35705950-T, a functional polymorphism upstream of the *MUC5B* gene and an established risk factor for IPF, on COVID-19 outcomes in an ancestrally diverse population is unclear.

What This Study Adds to the Field: rs35705950-T was associated with fewer COVID-19 hospitalizations in a transancestry meta-analysis conducted in the Million Veteran Program (MVP; total $n = 511{,}965$; OR, 0.89 [0.82–0.97]) and in joint meta-analysis with the Host Genetics Initiative (total $n = 506{,}174$; OR, 0.90 [0.86–0.95]). In subgroup analyses of MVP participants of European ancestry, rs35705950-T was associated with fewer post–COVID-19 pneumonia events, with evidence supportive of a protective dose–response relationship for each copy of the rs35705950-T allele. These results support a potentially protective effect of rs35705950-T against COVID-19 hospitalizations and postinfection pneumonia events.

individuals of European ancestry) and possible shared pathophysiological pathways between IPF and severe COVID-19 disease, we examined the association between rs35705950-T and the clinical outcomes of COVID-19 infection in the Million Veteran Program (MVP), a multiancestry cohort of over 650,000 U.S. veterans with detailed electronic health record (EHR) and genotyping data (15). After our primary analysis in the MVP, we validated our results with a comparable analysis conducted in the Host Genetics Initiative (HGI), a global collaboration of over 160 genetic studies assembled to facilitate rapid discovery and dissemination of COVID-19–related science (16).

Methods

Data Sources

Data from the MVP, a multiancestry genetic biobank sponsored by the U.S. Department of Veterans Affairs (VA), were analyzed (15). All protocols were approved by the VA Central Institutional Review Board, and all participants provided written informed consent. Genotyping was performed using a custom Thermo Fisher Axiom genotyping platform (MVP 1.0), which included direct genotyping of rs35705950-T. Ancestry was defined using harmonized ancestry, race, and ethnicity derived from self-report and genetic ancestry data (17). Individuals from three major ancestry groups—European (EUR), African (AFR), and Hispanic (HIS)—were included. Demographic and preexisting comorbidity data were collected from questionnaires and the VA EHR; "Pre-COVID" data were from the time of enrollment into the MVP to September 30, 2019. The cohort demographics and a description of the clinical conditions for all genotyped MVP participants and COVID-19–positive MVP participants (18) that were evaluated in this study are shown in Table 1.

COVID-19 outcome definitions. COVID-19 infection status from February 2020 to April 2021 was assessed by either self-report (if testing was performed outside the VA) or by a positive PCR-based test (19, 20) performed within the VA health system. The index date was defined as a COVID-19 diagnosis date (i.e., specimen date); for a hospitalized patient, it was defined as the admission date up to 15 days before the COVID-19 diagnosis date.

Our primary analyses used definitions that were harmonized with those used by the COVID-19 HGI to maximize sample sizes, examine consistent endpoints, and facilitate meta-analyses. In accordance with the HGI phenotype definitions, the following three analyses were performed: *1*) COVID-19 susceptibility: individuals who tested positive for COVID-19 *2*); COVID-19 hospitalization: individuals who were hospitalized for symptoms of COVID-19 *3*); and COVID-19 severe: individuals who either were hospitalized and required respiratory support beyond nasal cannula oxygen or died. For each phenotype, population controls were selected; i.e., all the MVP participants who were not defined as a case were assigned as control. Controls also included individuals for whom status of exposure to COVID-19 was unknown. Additionally, in the study of hospitalization for COVID-19, a fourth analysis was conducted in which individuals who tested positive and were hospitalized for COVID-19 were compared with COVID-19–positive individuals who were not hospitalized.

Separate sensitivity analyses were conducted within the MVP cohort to evaluate the effect of possible confounding variables that were not evaluated in the HGI. To ensure complete capture of relevant clinical information, we included only individuals who were tested for COVID-19 at one of the VA sites; individuals who did not have COVID testing within the VA were excluded. The four phenotypes evaluated included COVID-19 testing positive rate, COVID-19 hospitalization, severe COVID-19 infection or death, and COVID-19 death; modified from the WHO Working Group criteria (21) in the online supplement). All models were adjusted for additional possible confounders, including body mass index (BMI), Charlson comorbidity index (CCI), smoking history, asbestosis (22), rheumatoid arthritis with interstitial lung disease (RA-ILD) (23) and IPF diagnosis (24). These variables were derived from the data within 2 years before the index date of diagnosis (Table E1).

Postindex pneumonia definition. Postindex pneumonia was defined from codes from the International Classification of Diseases (10th revision; ICD) within 60 days after the index date. Association with COVID-19 pneumonia events (within 60 days of infection; pneumonia60d) was performed among patients who received COVID-19 PCR testing at VA sites (Table E2). The ICD codes used to define pneumonia events within 60 days of COVID-19 infection (pneumonia60d) are presented in Table E3.

Statistical Analysis

Meta-analysis with HGI. The analysis recommendations from the COVID-19 HGI were utilized to test for associations between the rs35705950-T allele and COVID-19 outcomes. First, we conducted analyses within MVP by each ancestry group using plink2a. Inverse variance weighted meta-analyses were then performed with summary statistics from Release 5 (01/18/2021) of the HGI using GWAMA (18) (Additional details are given in the online supplement).

Sensitivity analysis. Firth logistic regression, as implemented in the R (v3.6.1) package "brglm2" (version 0.7.1) was used in the sensitivity analyses, as it provides a bias-reduced estimate in the setting of small sample sizes and is most powerful for analyses of genetic mutations. This was relevant because adjustment for multiple variables can lead to small sample

Table 1. Demographics for COVID-19–Positive and All MVP Participants Examined in This Study

Characteristics	MVP	COVID-19 Positive
Total patients, *n*	658,582*	19,168
Male, *n* (%)	592,516 (90)	17,151 (89)
Genetic ancestry, *n* (%)		
European	464,961 (70)	11,778 (61)
African	123,120 (19)	4,893 (26)
Hispanic	52,183 (8)	2,497 (13)
Asian†	8,329 (1)	N/A
Other†	9,989 (2)	N/A
MUC5B rs35705950-T		
GG	547,846 (83.2)	16,272 (84.9)
GT	104,834 (15.9)	2,758 (14)
TT	5,902 (0.9)	138 (0.7)
Comorbidities, *n* (%)		
Obesity (Phecode = 278)	283,197 (43)	10,844 (56)
Hypertension (Phecode = 401.1)	451,998 (69)	14,036 (73)
Type 2 diabetes (Phecode = 250.2)	227,575 (34)	8,190 (43)
Coronary artery disease (Phecode = 411.4)	152,136 (23)	4,664 (24)
Chronic kidney disease (Phecode = 585.2)	10,046 (1.5)	335 (2)
Outcomes, *n* (%)		
Hospitalized	—	4,234 (22)
Severe/deceased	—	947 (5)

Definition of abbreviations: COVID-19 = coronavirus disease; MVP = Million Veteran Program; N/A = not applicable.
*MVP participants who died before March 2020 were excluded from the analysis described in Table 2.
†Because of the small sample sizes, individuals from Asian and Other ancestry groups were not included.

sizes within covariate categories. Associations between COVID-19 outcomes and rs35705950-T were performed separately by ancestry, with adjustment for age, age^2, sex, BMI, CCI, smoking history, asbestosis, RA-ILD, and IPF (Table E1), and the first 20 principal components. The meta-analyses were performed using random-effects models in "metafor" (version 2.4–0).

Post-index pneumonia. Interactions between COVID-19 infection status and rs35705950-T on the outcome of COVID-19 pneumonia within 60 days (pneumonia60d) were assessed using a multiplicative interaction term followed by stratified analyses by COVID-19 infection status. The firth logistic model for interaction included the independent variables of interest: COVID-19 infection status, rs35705950-T, and their multiplicative interaction term, adjusted for preindex pneumonia (yes/no, within 2 years preindex), age, age^2, sex, BMI, CCI, smoking history, asbestosis, RA-ILD and IPF diagnosis, and first 20 ancestry-specific principal components (PC1–20). The SNP rs35705950-T was modeled as a continuous variable (additive genetic model with values 0, 1, and 2) so that the interaction odds ratio (OR) is equal for heterozygous versus homozygous wild-type and homozygous mutation versus heterozygous mutation. The additive interaction of COVID-19 and rs35705950-T (for every increase in one allele; i.e., 0 to 1, or 1 to 2 copies) was also assessed with the estimate of excess risk due to interaction (RERI) and the 95% confidence interval estimated with the delta method.

Phenome-wide and laboratory-wide association studies (PheWAS and LabWAS, respectively). Associations between rs35705950-T and preexisting comorbid conditions and laboratory values were

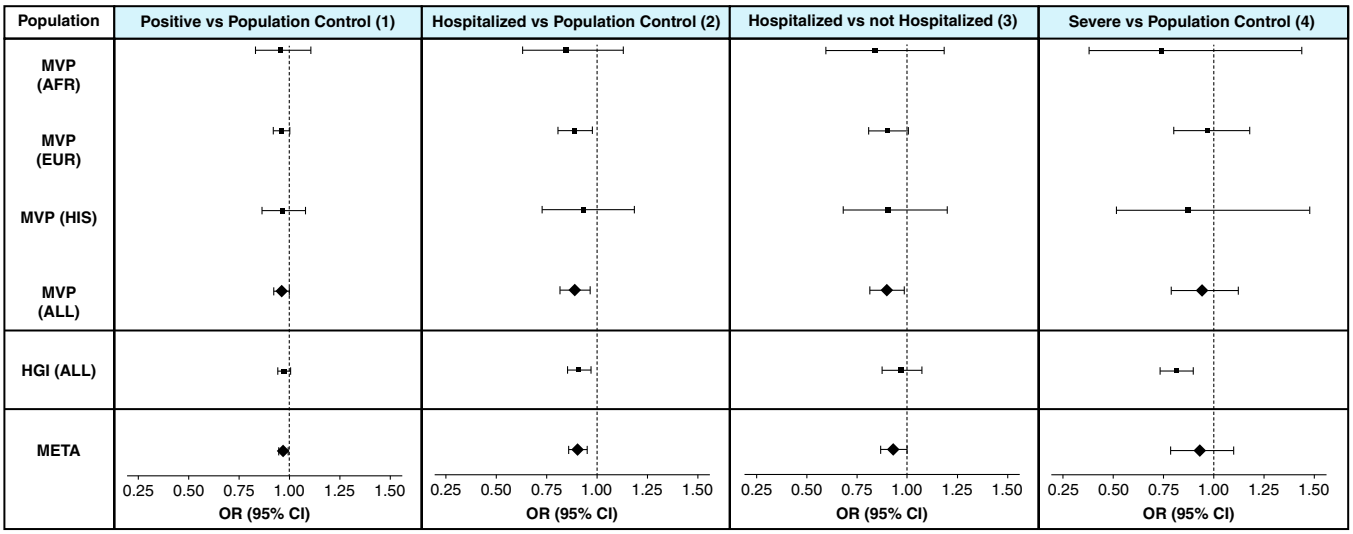

Figure 1. Forest plot association of rs35705950-T in *MUC5B* with *1*) COVID-19–positive participants versus population controls; *2*) COVID-19–positive, hospitalized participants versus population controls, *3*) COVID-19–positive, hospitalized participants versus COVID-19–positive participants who were not hospitalized, and *4*) participants who were hospitalized for COVID-19 with high-flow oxygen or died of COVID-19 versus population controls. Odds ratio (OR) and 95% confidence interval (95% CI) are reported for the minor (T) allele, and results are shown for VA Million Veteran Program (MVP) African American (AFR), European American (EUR), and Hispanic/Latino American (HIS) participants and for multiancestry meta-analysis (ALL); the COVID-19 Host Genetics Initiative (HGI) multiancestry round 5 meta-analysis, excluding MVP and 23andMe; and the meta-analysis of MVP and HGI (META).

Table 2. Association between rs35705950-T in *MUC5B* with COVID-19 Clinical Outcomes

Analysis	Analysis and Population	Group	Case | Controls	Sample Size	EA: EAF	OR (95% CI)	*P* Value
COVID-19 positive vs. population controls	MVP	AFR	4,893 | 94,556	99,449	T: 0.021	0.96 (0.83–1.11)	0.572
		EUR	11,778 | 357,198	368,975	T: 0.109	0.96 (0.92–1)	0.076
		HIS	2,497 | 41,100	43,597	T: 0.074	0.97 (0.86–1.08)	0.555
		META	19,168 | 492,854	512,021	T: 0.089	0.96 (0.93–1)	0.060
	HGI	META	25,652 | 128,2973	1,308,625	T: 0.111	0.98 (0.95–1.01)	0.134
	MVP + HGI	META	44,820 | 1,775,827	1,820,646	T: 0.105	0.97 (0.95–1)	**0.019**
COVID-19 positive and hospitalized vs. population controls	MVP	AFR	1,300 | 98,129	99,429	T: 0.021	0.84 (0.63–1.13)	0.259
		EUR	2,417 | 366,449	368,865	T: 0.109	0.89 (0.81–0.98)	0.015
		HIS	517 | 43,062	43,579	T: 0.074	0.93 (0.73–1.19)	0.562
		META	4,234 | 507,640	511,873	T: 0.089	0.89 (0.82–0.97)	$\mathbf{6.86 \times 10^{-3}}$
	HGI	META	9,086 | 1,001,201	1,010,287	T: 0.106	0.91 (0.85–0.97)	$\mathbf{4.12 \times 10^{-3}}$
	MVP + HGI	META	13,320 | 1,508,841	1,522,160	T: 0.1	0.9 (0.86–0.95)	$\mathbf{8.99 \times 10^{-5}}$
COVID-19 positive and hospitalized vs. COVID-19 positive but not hospitalized	MVP	AFR	1,300 | 3,573	4,873	T: 0.02	0.84 (0.6–1.19)	0.327
		EUR	2,417 | 9,251	11,668	T: 0.106	0.9 (0.81–1.01)	0.063
		HIS	517 | 1,962	2,479	T: 0.071	0.91 (0.68–1.2)	0.493
		META	4,234 | 14,786	19,020	T: 0.079	0.9 (0.81–0.99)	**0.030**
	HGI	META	4,420 | 11,093	15,513	T: 0.162	0.97 (0.88–1.08)	0.575
	MVP + HGI	META	8,654 | 25,879	34,533	T: 0.116	0.93 (0.87–1)	**0.049**
COVID-19 positive, severe vs. population controls	MVP	AFR	284 | 99,165	99,449	T: 0.021	0.74 (0.38–1.44)	0.375
		EUR	543 | 368,432	368,975	T: 0.109	0.97 (0.8–1.18)	0.774
		HIS	120 | 43,477	43,597	T: 0.074	0.88 (0.52–1.48)	0.620
		META	947 | 511,074	512,021	T: 0.089	0.94 (0.79–1.12)	0.509
	HGI	META	3,757 | 699,885	703,642	T: 0.108	0.81 (0.73–0.9)	$\mathbf{6.69 \times 10^{-5}}$
	MVP + HGI	META	4,704 | 1,210,959	1,215,663	T: 0.1	0.93 (0.79–1.1)	0.410

Definition of abbreviations: AFR = African American; COVID-19 = coronavirus disease; EA = effect allele; EAF = effect allele frequency; EUR = European American; HGI = COVID-19 Host Genetics Initiative; HIS = Hispanic/Latino American; META = meta-analysis of MVP and HGI; MVP = Million Veteran Program; OR = odds ratio; 95% CI = 95% confidence interval.
ORs and 95% CIs are reported for the minor (T) allele. Results are shown for VA MVP AFR, EUR, and HIS participants; for the multiancestry meta-analysis (ALL); the HGI multiancestry round 5 meta-analysis, excluding MVP and 23&Me; and for META.
Bold represent significant results i.e., *P* value <0.01.

examined using clinical data before the COVID-19 pandemic (September 2019). Individuals with ⩾2 Phecodes (25) were defined as cases. Phecodes with <200 cases within each ancestry group were excluded, resulting in 1,618 EUR, 1,289 AFR, and 994 HIS Phecodes. LabWAS was conducted for 69 clinical tests; for individuals with repeated measures, the median of the individuals' EHR record was used. Logistic/firth regression and linear regression were used for Phecodes and laboratory measurements, respectively. A Bonferroni-adjusted *P* value threshold of 1.2×10^{-5} (0.05/3901) accounted for all the models tested across

Table 3. Associations between the *MUC5B* rs35705950-T Allele and Pneumonia within 60 Days of COVID-19 Testing among MVP Participants of European Ancestry

		OR (95% CI)	*P* Value
OR (95% CI) for Pneumonia within 60 Days of COVID-19 Test			
Effect of rs35705950-T allele by COVID-19 strata	COVID-19 negative	1.06 (0.98, 1.15)	0.13
	COVID-19 positive	0.82 (0.72, 0.93)	0.001
Risk of pneumonia (COVID-19 positive vs. COVID-19 negative) by rs35705950-T copies	Copy number = 0	10.9 (10.2, 11.7)	< 0.0001
	Copy number = 1	8.42 (7.36, 9.61)	< 0.0001
	Copy number = 2	6.47 (4.93, 8.48)	< 0.0001
Measures of Interaction between COVID-19 and *MUC5B* Allele			
Multiplicative scale, OR (95% CI), *P*	0.77 (0.66, 0.89), *P* = 0.0004		
Additive scale RERI, OR (95% CI)	−2.07 (−3.24, −0.9)		

Definition of abbreviations: COVID-19 = coronavirus disease; MVP = Million Veteran Program; RERI = relative excess risk due to interaction; 95% CI = 95% confidence interval.
ORs and 95% CIs are estimated from Firth logistic regression adjusting for preindex pneumonia, age, age^2, sex, body mass index, Charlson Comorbidity Index, smoking, idiopathic pulmonary fibrosis, asbestosis, rheumatoid arthritis with interstitial lung disease, and principal components 1–20, including a multiplicative interaction between additive *MUC5B* rs35705950-T allele and COVID-19 infection. An estimate of the RERI and 95% CI (delta method) corresponding to an additive interaction effect is estimated from the logistic regression model.

three ancestries for significance. Analyses were performed using PLINK2 (26) (additional details are given in the online supplement).

Results

Association between rs35705950-T with COVID-19 Positivity and Outcome Severity

Our study included 19,168 COVID-19–positive patients from three major ancestry groups (EUR, AFR, and HIS). The minor allele frequency for rs35705950-T was 9% among the entire MVP participants and 8% among COVID-19–positive individuals (Table 1); among these, 14.4% ($n = 2{,}758$) and 0.7% ($n = 138$) were carriers of one copy and two copies of *MUC5B* rs35705950-T, respectively (Table 1; Figure E1). Associations between rs35705950-T and clinical outcomes by ancestry group within the MVP, the transancestry meta-analysis within the MVP, and the joint meta-analysis with HGI are shown in Figure 1 and Table 2. In the joint meta-analysis, the most significant association was between rs35705950-T and fewer hospitalization events compared with population controls (OR, 0.90 [95% CI, 0.86–0.93], $P = 8.99 \times 10^{-5}$; Figure 1 and Table 2). The *MUC5B* rs35705950-T allele was not associated with reduced COVID-19 positivity in transancestry meta-analysis within the MVP (OR, 0.98 [0.95–1.01], $P = 0.06$) but was nominally significant ($P < 0.05$) in the joint meta-analysis with HGI (OR, 0.97 [0.95–1]; $P = 0.02$). The *MUC5B* rs35705950-T allele was not associated with severe COVID-19 disease (critically ill).

To further evaluate the robustness of our findings from the primary analyses, we performed sensitivity analyses among MVP participants from three ancestry groups (EUR, AFR, and HIS) who had COVID-19 testing performed within the VA ($n = 136{,}164$). Associations between rs35705950-T and COVID-19 positivity and outcomes with adjustment for additional covariates, including BMI, smoking status, CCI, asbestosis, RA-ILD, and IPF diagnosis, which were not available in HGI, were performed in MVP participants. We observed similar effect sizes as the primary analyses (Table E4), with the association between rs35705950-T allele and reduced risk of hospitalization, compared with nonhospitalization, due to COVID-19 remaining robust (OR, 0.86 [0.77, 0.95], $P = 0.004$; Table E4). The OR and 95% confidence interval (CI) were very similar to the primary analyses (OR, 0.88 [95% CI, 0.81, 0.96]). Notably, rs35705950-T was associated with a reduced risk of testing positive for COVID-19 (OR, 0.95 [0.90, 0.99], $P = 0.03$) in this MVP subpopulation.

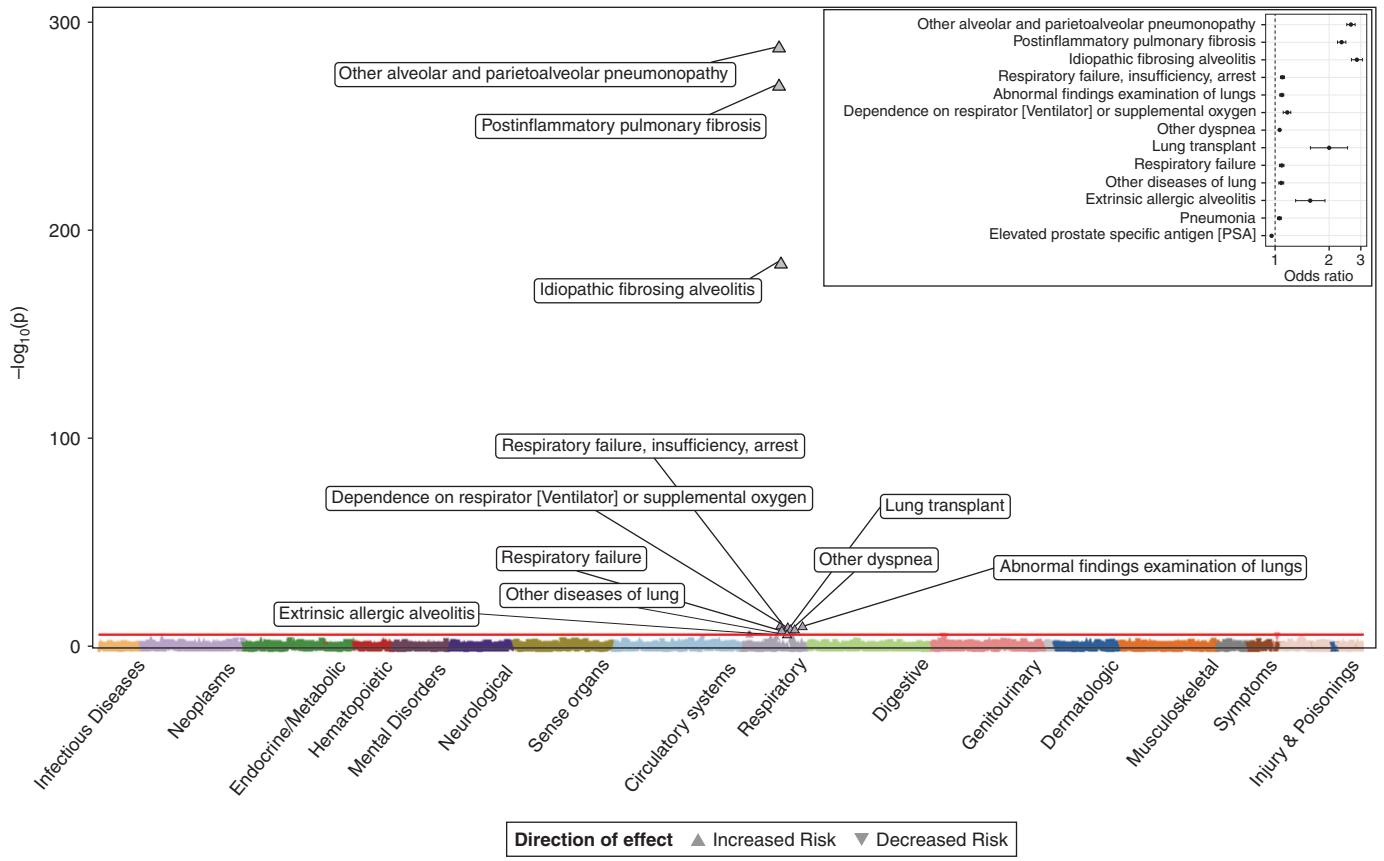

Figure 2. Phenome-wide association study (PheWAS) of *MUC5B* rs35705950-T allele in the MVP. A PheWAS plot shows associations of rs35705950-T and phenotypes derived from electronic health records data before COVID-19 in MVP participants of European ancestry. The phenotypes are shown on the x axis and organized by disease categories. The *P* value ($-\log_{10}$) of each association is shown on the y axis. The direction of the triangle represents the direction of effect of the associations, with the upward triangle indicating increased risk and the downward triangle indicating reduced risk. The red line indicates the significance threshold based on the Bonferroni correction. The forest plot of Bonferroni significant associations is shown at the right top corner of the PheWAS plot.

The *MUC5B* rs35705950-T allele was not associated with severe outcomes plus mortality (OR, 0.89 [0.67–1.20], $P = 0.45$) or mortality alone (OR, 0.93 [0.74–1.17], $P = 0.56$; Table E4).

Association between rs35705950-T and Pneumonia Events within 60 Days of COVID-19 Infection

Among MVP participants of European ancestry who tested positive for COVID-19 ($n = 8{,}541$), the adjusted OR for postindex pneumonia was 18% less with each additional *MUC5B* rs35705950-T allele (OR, 0.82 [0.72, 0.93], $P = 0.001$). Among COVID-19–negative MVP participants, the adjusted odds for postindex pneumonia nonsignificantly increased with each additional *MUC5B* rs35705950-T allele (OR, 1.06 [0.98, 1.15], $P = 0.13$). Increasing copy numbers of the *MUC5B* rs35705950-T allele were associated with progressive reduction of the postindex pneumonia risk, with adjusted ORs of 10.9 [10.2, 11.7], 8.41 [7.36, 9.61], and 6.47 [4.93, 8.48] in patients with zero, one, and two copies of *MUC5B* rs35705950-T respectively among COVID-19–positive relative to COVID-19–negative subjects ($P < 0.0001$; Table 3). This differential effect of an additional *MUC5B* rs35705950-T allele on postindex pneumonia in COVID-19–positive versus COVID-19–negative patients was statistically significant (multiplicative scale interaction OR, 0.77 [0.66, 0.89], $P = 0.0004$) in EUR individuals (Table 3). There was also evidence of a negative additive interaction, with the RERI corresponding to an increase of one copy of rs35705950-T allele estimated as −2.07 (95% CI = −3.24, −0.9). This suggests that the relative risk for pneumonia in COVID-19–positive patients is −2.07 less for each increase in one copy of *MUC5B* rs35705950-T allele than if there were no interaction between COVID-19 and *MUC5B* rs35705950-T. The number of post–COVID-19 pneumonia events in HIS or AFR individuals was too low to permit further analysis.

Exploring Shared Pathobiology with the *MUC5B* rs35705950-T through PheWAS and LabWAS Analyses

To explore clinical conditions and biomarkers associated with the *MUC5B* rs35705950-T allele that may impact the susceptibility and severity of COVID-19, PheWAS and LabWAS analyses using pre–COVID-era data (through September 2019) were performed. The sample sizes for MVP participants included in PheWAS and COVID-19 association studies are shown in Table 1 (Figure E1). The results of the PheWAS analyses are shown in Figure 2, with full summary statistics presented in Table E5.

In PheWAS analyses between rs35705950-T and PheCodes with >200 cases, significant associations were identified exclusively with pulmonary processes. Increased risk for "other alveolar and parietoalveolar pneumonopathy" (Phecode 504; OR = 2.64 [2.50 – 2.78], $P = 7.07 \times 10^{-289}$) and "postinflammatory pulmonary fibrosis" (Phecode 502; OR = 2.85 [2.65–3.05], $P = 8.90 \times 10^{-186}$) were significantly associated ($P_{\text{Bonferroni}} < 1.2 \times 10^{-5}$) with rs35705950-T in all three ancestry groups (EUR, AFR, and HIS). Among the EUR and AFR groups, rs35705950-T was also associated with increased risk for idiopathic fibrosing alveolitis (Phecode 504.1). Notably, there were no significant associations identified between rs35705950-T and influenza infection (Phecode 481) or bacterial pneumonia (Phecode 480.1) in any of the ancestry groups. The power to detect a difference with these conditions was >95%, as there were 4,728 cases of influenza and 10,579 cases of bacterial pneumonia in the EUR cohort. There were also no significant associations between rs35705950-T and non–pulmonary conditions potentially relevant to COVID-19 outcomes, including pulmonary embolism/deep vein, systemic inflammatory syndrome/sepsis, or acute renal failure (Figure E2 and Table E5).

The LabWAS of *MUC5B* rs35705950-T with median values of pre-COVID clinical laboratory tests is shown in Table E6 and Figure E3. Among EUR participants, 10 laboratory tests were significant after Bonferroni adjustment, with the majority belonging to white blood cell counts or fractions, with an increase in monocyte count also significant in the HIS subgroup. There were no significant associations among individuals in the AFR subgroup.

Discussion

Our study supports that the "T" allele of rs35705950 in *MUC5B*, which has been associated with an *increased* risk for the development of IPF, confers a *decreased* risk for COVID-related hospitalization among MVP participants; among participants of European descent, a decreased incidence of pneumonia after COVID-19 infection was also observed. The protective effect of the rs35705950-T may appear to contradict previous work demonstrating an increased risk for acute respiratory distress syndrome (ARDS) in the pre-COVID era (27) as well as the increased risk of severe COVID-19 disease observed for other well-established causal variants of IPF located in the *TERC*, *DEPTOR*, and *FAM13A* (28) genes. However, our findings are consistent with previous studies conducted in European cohorts (29) and in the HGI (28), which support that pathophysiological changes due to rs35705950-T may result in distinct interactions with SARS-CoV-2, which confers the differential risks observed.

The rs35705950 polymorphism is located within an enhancer region of the *MUC5B* gene, the protein product of which is a major gel-forming mucin in the lung that plays a key role in mucociliary clearance and host defense (30–33). Consistent with this, mouse knockout models for *MUC5B* demonstrate increased susceptibility toward bacterial infections of the respiratory tract and persistent inflammation. In contrast to the loss-of-function of knockout models, the "T" allele of rs35705950 results in a gain-of-function and is associated with enhanced expression of the *MUC5B* transcript in lung tissue from human subjects (24) without clinical IPF. Notably, although excess *MUC5B* protein is observed in the epithelial cells of respiratory bronchioles and honeycomb cysts (31, 32, 34) of individuals with IPF, differential expression in lung tissue by rs35705950 genotype has not been consistently observed in patients with advanced disease and may be attributable to the universally elevated *MUC5B* protein levels in IPF patients, regardless of their rs35705950 allelic configuration (24, 32, 34).

The functional impact of increased *MUC5B* expression on clinical outcomes in populations both with and without IPF remains incompletely understood. In a study that antedated the COVID-19 pandemic, a modestly increased risk for the development of ARDS, a major feature of severe SARS-CoV-2 infection, was observed among homozygotes for rs35705950-T who were >50 years old, none of whom had clinical or radiographic evidence of IPF on *post hoc* chart reviews (27). It should be noted, however, that the underlying etiologies for ARDS in the population examined were

heterogeneous, with over half of the cases attributed to either trauma or non–pneumonia-related sepsis. In contrast to the increased risk for pre–COVID-era ARDS, the rs35705950-T allele was associated with decreased risk for acute respiratory exacerbation events among non-Hispanic White ever-smokers with interstitial lung abnormalities (ILAs), defined as nondependent parenchymal infiltrates that are frequently considered subclinical precursors to clinically apparent ILD, in the COPDGene study (35). Whether increased mucin production confers protection against viral infections (36), which are believed to contribute to a substantial proportion of acute respiratory exacerbation events in chronic lung disease (37), should be explored in future studies.

Our analyses suggest that the decreasing risk of postindex pneumonia associated with rs35705950-T may be specific to COVID-19. First, to our knowledge, there have been no previously published reports of differential susceptibility or clinical outcomes in non–COVID-19 respiratory viral or bacterial infections by rs35705950-T genotype. Within our own MVP data, there were no associations between prepandemic influenza infection or bacterial pneumonia and rs35705950-T allele. Second, although rs35705950-T carrier status or number of allele copies did not impact testing rates for COVID-19 (which were 22.2%, 22.3%, and 22.4% for individuals of European descent with zero, one, and two copies of rs35705950-T, respectively), individuals with the "T" allele demonstrated a trend toward decreasing rates of testing positive in MVP, supporting that rs35705950-T may modulate an individual's susceptibility to infection by SARS-CoV-2. Third, in addition to attenuating the risk for postindex pneumonia exclusively among COVID-19–positive individuals, differential risk by rs35705950-T copy number was observed, supporting a dose–response relationship for this outcome.

We did not observe associations between rs35705950 and severe COVID-19 illness or mortality. This may be due to the low incidence and/or multifactorial causes of these severe outcomes, which may include nonrespiratory disorders such as shock and multisystem organ failure. This is supported by the PheWAS analysis, which found nearly exclusive associations between rs35705950-T with pulmonary processes, with no evidence of association (with >80% power to detect differential effects) with other processes potentially relevant to severe COVID-19 disease, including thromboembolic disease, septic shock, and acute renal failure.

The COVID-19 pandemic amplified healthcare disparities due to socioeconomic factors and likely contributed to differences in outcomes (38). However, heterogeneity in the magnitude of disparities between healthcare systems exists, with numerous studies demonstrating *reduced* levels of disparities and relatively equitable access to care among VA healthcare users (relative to non–VA healthcare users) (39–41); thus, differences in outcomes attributable to socioeconomic factors within the VA may be less than those in the general population. Consistent with this, a study by Trivedi and colleagues (42) reported minimal changes in 30-day mortality rates for heart failure and pneumonia hospitalizations after adjustment for a comprehensive panel of socioeconomic factors, including poverty, housing, education, and rurality. Within our cohort, access to care was uniformly high, with >98% of individuals having ⩾1 primary care visit within the preceding 18 months and comparable numbers of patients with ⩾1 clinical encounter or admission in the preceding year. However, because of the complexities of socioeconomic factors and their potential impact on and interactions with rs35705950-T, we acknowledge that future studies in this area are needed.

Strengths and Limitations

The strengths of our study include a large, ethnically and geographically diverse cohort with harmonized prospective outcomes data, directly genotyped rs35705950 data, and the use of robust statistical approaches. In addition to interrogating the relationships between rs35705950-T and clinical outcomes, the availability of rich clinical phenotyping data permitted exploration of the underlying pathobiological mechanisms through PheWAS and LabWAS and support that the differential risks observed by rs35705950-T are likely mediated by pulmonary-specific processes. Despite these strengths, we acknowledge the following limitations. First, although the MVP is one of the largest and most diverse genomic medicine databases established to date, participants are predominantly male and of European ancestry; this, and the lower minor allele frequency of rs35705950-T among non-European populations, may have impacted our power to detect associations between rs35705950-T genotype and clinical outcomes among females and racial/ethnic minorities. Second, because of the timing of the data freeze for our analysis, only short-term outcomes were captured. Future studies with extended follow-up to explore the potential relationship between rs35705950-T and post-COVID syndromes, including lung fibrosis, are needed. Third, although our analyses were adjusted for critical variables, residual confounding and bias may still exist. Given the association between rs35705950-T and an increased incidence of subclinical ILA and respiratory symptoms (43), individuals harboring the *MUC5B* variant may exhibit behavior modifications (increased testing or more stringent self-isolation) that could have introduced biases that cannot be captured or adjusted for in our analyses (28). Fourth, viral subtypes, vaccination status, and treatment approaches evolved with time, which may have impacted severity outcomes. Fifth, we acknowledge that the effect estimates associated with rs35705950-T in our study are modest and may limit its application in clinical or predictive algorithms. However, despite these limitations, we assert that our results, which identify a potentially protective role of rs35705950-T in COVID-19 outcomes, support that this functional polymorphism confers pleiotropic effects that may be modulated by epistasis and environmental interactions; longer term follow-up and future analyses into these complex relationships are warranted. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: We are grateful to our veterans for their contribution to MVP. Full acknowledgments for the VA Million Veteran Program COVID-19 Science Initiative can be found in the online supplement.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20:533–534. [Published erratum appears in *Lancet Infect Dis* 20:e215].
2. Zhao Y-M, Shang Y-M, Song W-B, Li Q-Q, Xie H, Xu Q-F, *et al.* Follow-up study of the pulmonary function and related physiological characteristics of COVID-19 survivors three months after recovery. *EClinicalMedicine* 2020;25:100463.
3. Yan X, Huang H, Wang C, Jin Z, Zhang Z, He J, *et al.* Follow-up study of pulmonary function among COVID-19 survivors 1 year after recovery. *J Infect* 2021;83:381–412.
4. McGroder CF, Zhang D, Choudhury MA, Salvatore MM, D'Souza BM, Hoffman EA, *et al.* Pulmonary fibrosis 4 months after COVID-19 is associated with severity of illness and blood leucocyte telomere length. *Thorax* 2021;76:1242–1245.
5. King CS, Nathan SD. Idiopathic pulmonary fibrosis: effects and optimal management of comorbidities. *Lancet Respir Med* 2017;5:72–84.
6. George PM, Patterson CM, Reed AK, Thillai M. Lung transplantation for idiopathic pulmonary fibrosis. *Lancet Respir Med* 2019;7:271–282.
7. Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;183: 431–440.
8. Selman M, Pardo A. Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis. an integral model. *Am J Respir Crit Care Med* 2014;189:1161–1172.
9. Selman M, Pardo A. The leading role of epithelial cells in the pathogenesis of idiopathic pulmonary fibrosis. *Cell Signal* 2020;66:109482.
10. Zhang Y, Noth I, Gibson KF, Ma S-F, Richards TJ, Bon JM, *et al.* A variant in the promoter of *MUC5B* is associated with idiopathic pulmonary fibrosis and not chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2011;183:A6395.
11. Moore C, Blumhagen RZ, Yang IV, Walts A, Powers J, Walker T, *et al.* Resequencing study confirms that host defense and cell senescence gene variants contribute to the risk of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2019;200:199–208.
12. Gally F, Sasse SK, Kurche JS, Gruca MA, Cardwell JH, Okamoto T, *et al.* The *MUC5B*-associated variant rs35705950 resides within an enhancer subject to lineage- and disease-dependent epigenetic remodeling. *JCI Insight* 2021;6:e144294.
13. Peljto AL, Zhang Y, Fingerlin TE, Ma S-F, Garcia JGN, Richards TJ, *et al.* Association between the *MUC5B* promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA* 2013;309:2232–2239.
14. Dudbridge F, Allen RJ, Sheehan NA, Schmidt AF, Lee JC, Jenkins RG, *et al.* Adjustment for index event bias in genome-wide association studies of subsequent events. *Nat Commun* 2019;10:1561.
15. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016;70:214–223.
16. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* 2021;600:472–477.
17. Fang H, Hui Q, Lynch J, Honerlaw J, Assimes TL, Huang J, *et al.*; VA Million Veteran Program. Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *Am J Hum Genet* 2019;105:763–772.
18. Peloso GM, Tcheandjieu C, McGeary JE, Posner DC, Ho Y-L, Zhou JJ, *et al.*; VA Million Veteran ProgramCOVID-19 Science Initiative. Genetic loci associated with COVID-19 positivity and hospitalization in White, Black, and Hispanic veterans of the VA Million Veteran Program. *Front Genet* 2022;12:777076.
19. Ioannou GN, Locke E, Green P, Berry K, O'Hare AM, Shah JA, *et al.* Risk factors for hospitalization, mechanical ventilation, or death among 10 131 US veterans with SARS-CoV-2 infection. *JAMA Netw Open* 2020; 3:e2022310.
20. Obeidat M, Frank AR, Icardi MS, Klutts JS. VA-wide, multicenter verification study of the Cepheid Xpert SARS-CoV-2 assay. *Acad Pathol* 2021;8:23742895211011911.
21. WHO Working Group on the Clinical Characterisation and Management of COVID-19 Infection. A minimal common outcome measure set for COVID-19 clinical research. *Lancet Infect Dis* 2020;20:e192–e197.
22. Platenburg MGJP, Wiertz IA, van der Vis JJ, Crestani B, Borie R, Dieude P, *et al.* The *MUC5B* promoter risk allele for idiopathic pulmonary fibrosis predisposes to asbestosis. *Eur Respir J* 2020;55: 1902361.
23. Juge P-A, Lee JS, Ebstein E, Furukawa H, Dobrinskikh E, Gazal S, *et al. MUC5B* promoter variant and rheumatoid arthritis with interstitial lung disease. *N Engl J Med* 2018;379:2209–2219.
24. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, *et al.* A common *MUC5B* promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364:1503–1512.
25. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 2010; 26:1205–1210.
26. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7.
27. Rogers AJ, Solus JF, Hunninghake GM, Baron RM, Meyer NJ, Janz DR, *et al. MUC5B* promoter polymorphism and development of acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2018;198: 1342–1345.
28. Fadista J, Kraven LM, Karjalainen J, Andrews SJ, Geller F, Baillie JK, *et al.*; COVID-19 Host Genetics Initiative. Shared genetic etiology between idiopathic pulmonary fibrosis and COVID-19 severity. *EBioMedicine* 2021;65:103277.
29. van Moorsel CHM, van der Vis JJ, Duckworth A, Scotton CJ, Benschop C, Ellinghaus D, *et al.* The *MUC5B* promoter polymorphism associates with severe COVID-19 in the European population. *Front Med (Lausanne)* 2021;8:668024.
30. Rose MC, Voynow JA. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol Rev* 2006;86: 245–278.
31. Seibold MA, Smith RW, Urbanek C, Groshong SD, Cosgrove GP, Brown KK, *et al.* The idiopathic pulmonary fibrosis honeycomb cyst contains a mucocilary pseudostratified epithelium. *PLoS One* 2013;8: e58658.
32. Nakano Y, Yang IV, Walts AD, Watson AM, Helling BA, Fletcher AA, *et al. MUC5B* promoter variant rs35705950 affects *MUC5B* expression in the distal airways in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016;193:464–466.
33. Evans CM, Fingerlin TE, Schwarz MI, Lynch D, Kurche J, Warg L, *et al.* Idiopathic pulmonary fibrosis: a genetic disease that involves mucociliary dysfunction of the peripheral airways. *Physiol Rev* 2016;96: 1567–1591.
34. Conti C, Montero-Fernandez A, Borg E, Osadolor T, Viola P, De Lauretis A, *et al.* Mucins *MUC5B* and MUC5AC in distal airways and honeycomb spaces: comparison among idiopathic pulmonary fibrosis/usual interstitial pneumonia, fibrotic nonspecific interstitial pneumonitis, and control lungs. *Am J Respir Crit Care Med* 2016; 193:462–464.
35. Ash SY, Harmouche R, Putman RK, Ross JC, Martinez FJ, Choi AM, *et al.*; COPDGene Investigators. Association between acute respiratory disease events and the promoter polymorphism in smokers. *Thorax* 2018;73:1071–1074.
36. Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, *et al. MUC5B* is required for airway defence. *Nature* 2014;505:412–416.

37. Sapey E, Bafadhel M, Bolton CE, Wilkinson T, Hurst JR, Quint JK. Building toolkits for COPD exacerbations: lessons from the past and present. *Thorax* 2019;74:898–905.
38. Magesh S, John D, Li WT, Li Y, Mattingly-App A, Jain S, *et al.* Disparities in COVID-19 outcomes by race, ethnicity, and socioeconomic status: a systematic-review and meta-analysis. *JAMA Netw Open* 2021;4: e2134147.
39. Silva GC, Jiang L, Gutman R, Wu W-C, Mor V, Fine MJ, *et al.* Racial/ethnic differences in 30-day mortality for heart failure and pneumonia in the Veterans Health Administration using claims-based, clinical, and social risk-adjustment variables. *Med Care* 2021;59: 1082–1089.
40. Kovesdy CP, Norris KC, Boulware LE, Lu JL, Ma JZ, Streja E, *et al.* Association of race with mortality and cardiovascular events in a large cohort of US veterans. *Circulation* 2015;132:1538–1548.
41. Sheehan CM, Hayward MD. Black/white differences in mortality among veteran and non-veteran males. *Soc Sci Res* 2019;79:101–114.
42. Trivedi AN, Jiang L, Silva G, Wu W-C, Mor V, Fine MJ, *et al.* Evaluation of changes in Veterans Affairs medical centers' mortality rates after risk adjustment for socioeconomic status. *JAMA Netw Open* 2020;3: e2024345.
43. Hunninghake GM, Hatabu H, Okajima Y, Gao W, Dupuis J, Latourelle JC, *et al. MUC5B* promoter polymorphism and interstitial lung abnormalities. *N Engl J Med* 2013;368:2192–2200.