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

Anurag Verma1,2, Jessica Minnier3,4,5*, Emily S. Wan6,7*, Jennifer E. Huffman8, Lina Gao4,5, Jacob Joseph9,10, Yuk-Lam Ho8, Wen-Chin Wu11,12, Kelly Cho8,13, Bryan R. Gorman8, Nallakkandi Rajeevan14,15, Saiju Pyarajan8,16, Helene Garcon8, James B. Meigs17, Yan V. Sun18,19, Peter D. Reaven20,21, John E. McGeary22,23, Ayako Suzuki24,25, Joel Gelernter26,27, Julie A. Lynch28,29, Jeffrey M. Petersen1,30, Seyedeh Maryam Zekavat31,32, Pradeep Natarajan16,32,33, Sharvari Dalal1,30, Darshana N. Jhala1,30, Mehrad Arjomandi34, Elise Gatsby28, Kristine E. Lynch28,35, Robert A. Bonomo36, Matthew Freiberg37, Gita A. Pathak26,27, Curtis J. Donskey40, Ravi K. Maddur41, Quinn S. Wells37,42,43, Rose D. L. Huang8, Renato Polimanti26,27, Kyong Mi Chang1, Katherine P. Liao44, Philip S. Tsao45, Peter W. F. Wilson19,46, Adriana M. Hung47, Christopher J. O’Donnell48, John M. Gaziano8, Richard L. Hauger49,50, Sudha K. Iyengar51,52, Shiu-Wen Luoh4,5‡, 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 (Ncases = 4,325; Ncontrols = 507,640; OR = 0.89 [0.82–0.97]; P = 6.86 × 10−3) and joint meta-analyses with the HGI (Ncases = 13,320; Ncontrols = 1,508,841; OR, 0.90 [0.86–0.95]; P = 8.99 × 10−5). The rs35705950-T allele was not associated with reduced COVID-19 positivity in transancestry meta-analysis within the MVP (Ncases = 19,168/Ncontrols = 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 (Ncases = 44,820; Ncontrols = 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)

This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0. For commercial usage and reprints, please contact Diane Gern (dgern@thoracic.org).

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.
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 42 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 infection and those underlying IPF. Notably, although rs35705950-T has been associated with decreased mortality in COVID-19 (11, 12), 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 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: I) 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).

Supported by an MVP035 award from the Million Veteran Program, Office of Research and Development, Veterans Health Administration; an MVP008 award (1101BX004188-01) (S.-W.L.), Veterans Affairs of the United States Government BX004831 (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 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 published 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.
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; and 3) 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 size.
sizes within covariate categories. Associations between COVID-19 outcomes and rs35705950-T were performed separately by ancestry, with adjustment for age, age², 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).

**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.

**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) | 227,755 (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) |

**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).

**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 first 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, age², 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
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                 | MVP                     | AFR   | 4,893| 94,556   | 99,449      | T: 0.021| 0.96 (0.83–1.11)| 0.572   |
| vs. population controls           |            | EUR   | 11,778| 557,198 | 535,420      | T: 0.109| 0.96 (0.92–1.01)| 0.076   |
|                                   |            | HIS   | 2,497| 41,100   | 43,597       | T: 0.074| 0.97 (0.86–1.08)| 0.555   |
|                                   |            | METE  | 19,168| 492,854 | 512,021      | T: 0.089| 0.96 (0.93–1.00)| 0.060   |
|                                   |            | HGI   | 25,652| 128,2973| 130,625      | T: 0.111| 0.98 (0.95–1.01)| 0.134   |
|                                   |            | MVP + HGI | 44,820| 1,775,827| 1,820,646    | T: 0.105| 0.97 (0.95–1.01)| 0.019   |
| COVID-19 positive and hospitalized| MVP         | AFR   | 1,300| 98,129   | 99,429       | T: 0.021| 0.84 (0.63–1.13)| 0.259   |
| vs. population controls           |            | 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   |
|                                   |            | METE  | 4,234| 1,507,640| 511,873      | T: 0.089| 0.89 (0.82–0.97)| 6.86 × 10⁻⁵ |
|                                   |            | HGI   | 9,086| 1,001,201| 1,010,287    | T: 0.106| 0.91 (0.85–0.97)| 4.12 × 10⁻⁵ |
|                                   |            | MVP + HGI | 13,320| 1,508,841| 1,522,160    | T: 0.1 | 0.9 (0.86–0.95)| 8.99 × 10⁻⁵ |
| COVID-19 positive and hospitalized| MVP         | AFR   | 1,300| 98,129   | 99,429       | T: 0.021| 0.84 (0.63–1.13)| 0.327   |
| vs. population controls           |            | 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   |
|                                   |            | METE  | 4,234| 1,507,640| 511,873      | T: 0.089| 0.89 (0.82–0.97)| 6.86 × 10⁻⁵ |
|                                   |            | HGI   | 9,086| 1,001,201| 1,010,287    | T: 0.106| 0.91 (0.85–0.97)| 4.12 × 10⁻⁵ |
|                                   |            | MVP + HGI | 13,320| 1,508,841| 1,522,160    | T: 0.1 | 0.9 (0.86–0.95)| 8.99 × 10⁻⁵ |

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 multiancestevity round 5 meta-analysis, excluding MVP and 23Me; 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 944 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

| Analysis                          | Group | Case | Controls | Sample Size | EA: EAF | OR (95% CI) | P Value |
|-----------------------------------|-------|------|----------|-------------|---------|------------|---------|
| COVID-19 negative                 | EVP   | 0.98 | 1.15     | 1.06 (0.98, 1.15) | 0.13    |            |         |
| by COVID-19 strata                | MVP   | 0.72 | 0.93     | 0.82 (0.72, 0.93) | 0.001   |            |         |
| Risk of pneumonia                 |       |      |          | 10.9 (10.2, 11.7) | <0.0001 |            |         |
| (COVID-19 positive vs. COVID-19 negative) by rs35705950-T copies |       |      |          | 8.42 (7.36, 9.61) | <0.0001 |            |         |
| Measures of Interaction between COVID-19 and MUC5B Allele |       |      |          | 6.47 (4.93, 8.48) | <0.0001 |            |         |
| Multiplicative scale, OR (95% CI), P | 0.77 (0.66, 0.89), P = 0.0004 |      |          | <0.0001 |            |         |

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, age², 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. 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](image-url)

**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 × 10−289) and “postinflammatory pulmonary fibrosis” (PheCode 502; OR = 2.85 [2.65–3.05], P = 8.90 × 10−186) were significantly associated (P Bonferroni < 1.2 × 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-19 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.

**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 2020;20:621.]

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 EA, Hoffmann 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. Gally F, Sasse SK, Kurche JS, Gruca MA, Cardwell JH, Okamoto T, et al. Revealing the pathogenic and aging-related remodeling of idiopathic pulmonary fibrosis. J Am J Respir Crit Care Med 2020;118:431–440.

8. Selman M, Pardo A. The leading role of epithelial cells in the pathogenesis of idiopathic pulmonary fibrosis. Cell Signal 2020;66:109482.

9. Chang CC, Chow CC, Teller LC, Vatikul SM, Lee JJ. Idiopathic pulmonary fibrosis honeycomb cysts contain a mucociliary pseudostratified epithelium. Chest 2021;160:477–485.

10. Moore C, Blumhagen RZ, Yang IV, Waits A, Powers J, Walker T, et al. Resequencing study confirms that host defense and cell senescence genes contribute to the risk of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2019;200:199–208.

11. Selman M, Pardo A. Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis: an integrative model. Am J Respir Crit Care Med 2014;189:1161–1172.

12. Selman M, Pardo A. The leading role of epithelial cells in the pathogenesis of idiopathic pulmonary fibrosis. Cell Signal 2020;66:109482.

13. 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;183:431–440.

14. Moores C, Blumhagen RZ, Yang IV, Waits A, Powers J, Walker T, et al. Resequencing study confirms that host defense and cell senescence genes contribute to the risk of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2019;200:199–208.

15. 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.

16. 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;309:2233–2239.

17. Legrande E, Allen RJ, Sheehan NA, Schmidt AF, Lee JC, Jenkins RG, 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.

18. 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.

19. McGroder CF, Zhang D, Choudhury MA, Salvatore MM, D Souza EA, Hoffmann 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.

20. King CS, Nathan SD. Idiopathic pulmonary fibrosis: effects and optimal management of comorbidities. Lancet Respir Med 2017;5:72–84.

21. George PM, Patterson CM, Reed AK, Thillai M. Lung transplantation for idiopathic pulmonary fibrosis. Lancet Respir Med 2019;7:271–282.

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. Jiang P-A, Lee JS, Ebstein E, Furukawa H, Dobrinshikkh E, Gazzal 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.
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.