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

Anurag Verma PhD[1,2†], Jessica Minnier PhD[3,4,5†], Emily S Wan MD, MPH[7,8†], Jennifer E Huffman PhD[6], Lina Gao PhD[9,10], Jacob Joseph MBBS, MD[11,12], Yuk-Lam Ho MPH[13], Wen-Chih Wu MD, MPH[14,15], Kelly Cho PhD, MPH[16,17], Bryan R Gorman PhD[18], Nallakkandi Rajeevan PhD[19,20], Saiju Pyarajan[21,22], Helene Garcon MD[6], James B Meigs MD, MPH[23], Yan V Sun PhD[24,25], Peter D Reaven MD[26,27], John E McGeary PhD[28,29], Ayako Suzuki MD, PhD, MSc[30,31], Joel Gelernter MD[32,33], Julie A Lynch PhD, RN[34,35], Jeffrey M Peterson MD[36,37], Seyedeh Maryam Zekavat BS[38,39], Pradeep Natarajan MD, MMSc[40,39,41], Sharvari Dalal MD[43,37], Darshana N Jhala MD, B Mus[44,45], Mehrdad Arjomandi MD[46], Elise Gatsby MPH[47], Kristine E Lynch MS PhD[47,48], Robert A Bonomo[49], Mat Freiberg[50], Gita A Pathak PhD[51,52], Jin J Zhou PhD[53,54], Curtis J Donskey MD[55], Ravi K Madduri MS[56], Quinn S Wells MD, PharmD[57], Rose DL Huang MS[6], Renato Polimanti PhD[51,52], *Million Veteran Program COVID-19 Science Initiative*, Kyong-Mi Chang MD[58], Katherine P Liao MD, MPH[59], Philip S Tsao PhD[60], Peter W.F. Wilson MD[25,61], Adriana Hung[62], Christopher J O'Donnell MD, MPH[63], John M Gaziano[64], Richard L Hauger MD[65,66], Sudha K. Iyengar PhD[67,68*] and Shiuh-Wen Luoh MD PhD[69*]

[1]Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, 19104, USA, [2]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA, [3]OHSU-PSU School of Public Health, Oregon Health & Science University, Portland, OR, 97239, USA, [4]Knight Cancer Institute, Biostatistics Shared Resource, Oregon Health & Science University, Portland, OR, 97239, USA, [5]VA Portland Health Care System, OR, 97239, USA, [6]MAVERIC, VA Boston Healthcare System, Boston, MA, 02130, USA, [7]Department of Medicine, Pulmonary, Critical Care, Sleep, and Allergy Section, VA Boston Healthcare System, Boston, MA, 02115, USA, [8]Channing Division of Network Medicine, Brigham & Women's Hospital, Boston, MA, 02115, USA, [9]Knight Cancer Institute, Biostatistics Shared Resource, Oregon Health & Science University, [10]VA Portland Health Care System, [11]Department of Medicine, VA Boston Healthcare System, Boston, MA 02115, USA, [12]Brigham & Women's Hospital, [13]MAVERIC, VA Boston Healthcare System, Boston, MA, 2130, USA, [14]Department of Medicine, Cardiology, Providence VA Healthcare System, Providence, RI, 02908, USA, [15]Alpert Medical School & School of

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

Public Health, Brown University, Providence, RI, 02903, USA, [16]MAVERIC, VA Boston Healthcare System, 150 S Huntington Ave, Boston, MA, 02130, USA, [17]Medicine, Aging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02130, USA, [18]VA Boston Healthcare System, 150 S Huntington Ave, Boston, MA, 02131, USA, [19]Yale Center for Medical Informatics, Yale School of Medicine, New Haven, CT, 06511', USA, [20]Clinical Epidemiology Research Center (CERC), VA Connecticut Healthcare System, West Haven, CT, 06516', USA, [21]VA Boston Healthcare System, Boston, MA, 2115, [22]Department of Medicine, Harvard Medical School, [23]Medicine, General Internal Medicine, Massachusetts General Hospital, 55 Fruit St, Boston, MA, 02115, USA, [24]Epidemiology, Emory University School of Public Health, 1518 Clifton Rd. NE, Atlanta, GA, 30322, USA, [25]Atlanta VA Health Care System, 1670 Clairmont Road, Decatur, GA, 30033, USA, [26]Deptartment of Medicine, Phoenix VA Healthcare System, Phoenix, AZ, 85012, USA, [27]Univ. of AZ, [28]Department of Psychiatry and Human Behavior, Providence VA Medical Center, Providence, RI, 2908, USA, [29]Brown University Medical School, [30]Department of Medicine, Gastroenterology, Durham VA Medical Center, 508 Fulton St, Durham, NC, 27705, USA, [31]Department of Medicine, Gastroenterology, Duke University, Durham, NC, 27710, USA, [32]Psychiatry, Human Genetics, Yale Univ. School of Medicine, 950 Campbell Avenue, West Haven, CT, 06516, USA, [33]VA CT Healthcare Center, [34]VA Informatics & Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT, USA, [35]Department of Medicine, University of Utah School of Medicine, Salt Lake City, UT, [36]Pathology and Laboratory Medicine, Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, 19104, USA, [37]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA, [38]Computational Biology & Bioinformatics, Yale School of Medicine, 333 Cedar St, New Haven, CT, 6510, USA, [39]Program in Medical and Population Genetics, Cardiovascular Disease Initiative, Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA, [40]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, 02114, USA, [41]Department of Medicine, Harvard Medical School, Boston, MA, 02115, USA, [43]Pathology and Laboratory Medicine, Corporal Michael J Crescenz VA Medical Center, Philadelphia, PA, 19104, USA, [44]Pathology and Laboratory Medicine, Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, 19104, [45]Perelman School of Medicine, University of Pennsylvania, 19104, [46]Medicine, Pulmonary and Critical Care, San Francisco VA Healthcare System; University of California San Francisco, San Francisco, CA, 94121, USA, [47]VA Informatics and Computing Infrastructure (VINCI), VA Salt Lake City Healthcare System, Salt Lake City, UT, [48]Internal Medicine, Epidemiology, University of Utah School of Medicine, Salt Lake City, UT, [49]Case

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

Western Reserve University, Cleveland, OH, [50]Vanderbilt University Medical Center, Nashville, TN, [51]Department of Psychiatry, Division of Human Genetics, Yale School of Medicine, New Haven, CT, 06511, USA, [52]VA Connecticut Healthcare System, West Haven, CT, 06516, USA, [53]Medicine, University of California, Los Angeles, Los Angeles, CA, 90024, USA, [54]Epidemiology and Biostatistics, University of Arizona, AZ, 85724, USA, [55]Infectious Disease Section, Louis Stokes Cleveland VA and Case Western Reserve University, Cleveland, OH, 44106, USA, [56]Data Science and Learning, Argonne National Laboratory, 9700 S Cass Ave, Lemont, IL, 60439, [57]Departments of Medicine, Biomedical Informatics, and Pharmacology, Vanderbilt University Medical Center, Nashville, TN, 37232, USA, [58]Corporal Michael J Crescenz VA Medical Center, [59]Medicine, Rheumatology, VA Boston Healthcare System, Boston, MA, 02130, [60]Precision Medicine, VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA, 94304, USA, [61]Emory University School of Medicine, Atlanta, GA, 30322, USA, [62](No affiliation data provided), [63]Medicine, Cardiology, VA Boston Healthcare System, 1400 VFW Parkway, Boston, MA, 2132, USA, [64]VA Boston Healthcare System, Boston, MA, [65]Center of Excellence for Stress & Mental Health, VA San Diego Healthcare System, San Diego, CA, 92161, USA, [66]Center for Behavioral Genetics of Aging, University of California San Diego, La Jolla, CA, 92093, USA, [67]Case Western Reserve University, Cleveland, OH, 44106, USA, [68]Louis Stokes Cleveland VA Medical Center, Cleveland, OH, 44106, USA, [69]VA Portland Health Care System, Portland, OR, 97239, USA

**Joint Authorship**

[†]These authors contributed equally to this work

*These authors jointly supervised this work

Supported by MVP035 award from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, MVP008 (1I01BX004188-01) (S-W L), Veteran Affairs of the United States Government BX 004831 (P.W./K.C.), National Institute of Health grant R01GM138597 (A.V.), National Heart, Lung, and Blood Institute grant R01HL142711 and

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

R01HL127564 (P.N.). This publication does not represent the views of the Department of Veteran Affairs of the United States Government.

Author contributions: A.V, J.E.H, S.I, S.L, L.G and J.M and E.S.W. analyzed data. J.E.H., A.V., L.G., E.S.W., S.I., S.L. supervised data collection. A.V., J.E.H, L.G., J.M., E.S.W., S.I. and S.L. wrote the manuscript. A.V., J.E.H, L.G. J.M., E.S.W., S.I., and S.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 should be addressed to Shiuh-Wen Luoh (Shiuh-Wen.Luoh@va.gov), Sudha K Iyengar (ski@case.edu)

Manuscript word count: **3500**

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 an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org.



AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

## At a Glance Commentary

### Scientific knowledge on the subject

Parenchymal fibrosis, a late feature of severe COVID-19 disease, 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 trans-ancestry meta-analysis conducted in the Million Veteran Program (MVP, ntotal=511,965; OR=0.89 [0.82-0.97]) and in joint meta-analysis with the Host Genetics Initiative (ntotal=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 post-infection pneumonia events.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

## Abstract

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

**Objectives**: To assess whether rs35705950-T confers differential risk for clinical outcomes associated with 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 trans-ancestry 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 (BMI, 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 ($N_{cases}$=4,325/, $N_{controls}$=507,640; OR=0.89 [0.82-0.97], p=6.86 x $10^{-03}$) in trans-ancestry meta-analysis within MVP 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 x $10^{-05}$). The rs35705950-T allele was not associated reduced COVID-19 positivity in trans- ancestry meta-analysis within MVP ($N_{cases}$=19,168/$N_{controls}$=492,854; OR=0.98 [0.95-1.01], p=0.06) but was nominally significant (p


AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

< 0.05) in the joint meta-analysis with HGI ($N_{cases}$=44,820/$N_{controls}$=1,775,827; OR=0.97 [0.95-1]; p=0.03). We did not observe associations with severe outcomes or mortality. Among MVP individuals of European ancestry, 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

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

## Introduction

A respiratory disease caused by a novel coronavirus, now known as SARS-CoV-2, was first reported towards the end of 2019. Despite massive public health measures and vaccination initiatives, the coronavirus disease-2019 (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 which 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 towards 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 to 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 is: 1) a functional variant located within an enhancer subject to epigenetic programming and 2) contributes to pathologic mis-expression in IPF(12). Notably, while rs3705950-T has been robustly associated with increased susceptibility, the same allele has also been associated with

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

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 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 multi- ancestry cohort of over 650,000 U.S. Veterans with detailed electronic health record (EHR) and genotyping data(15). Following 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 multi- ancestry genetic biobank sponsored by the United States 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 (HARE) derived from self-report and genetic ancestry data(17). Individuals from three major ancestry groups: European (EUR), African (AFR), Hispanic (HIS) were included.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

Demographic and pre-existing comorbidity data were collected from questionnaires and the VA EHR; "pre-COVID" data was 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 02/2020 - 04/2021 was assessed by either self-report (if testing was performed outside the VA) or by a positive polymerase chain reaction (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; and for a hospitalized patient, the admission date up to 15 days prior to the COVID-19 diagnosis date.

Our primary analyses used harmonized definitions with 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 due to symptoms of COVID-19; (3) COVID-19 Severe: individuals who 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 include individuals with unknown COVID-19 infection exposure. Additionally for the study of COVID-19 hospitalization, a fourth analysis comprised COVID-19 positive individuals that were

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

hospitalized were compared with COVID-19 positive individuals that 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) (**Supplemental Methods**)). 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 idiopathic pulmonary fibrosis (IPF)diagnosis(24). These variables were derived from the data within two years prior to the index date of diagnosis (**Table E1)**.

*Post-index pneumonia definition*

Post-index pneumonia was defined from ICD codes within 60 days after the index date. Association with COVID-19 pneumonia events (pneumonia60d) were 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 COVID-19 infection (pneumonia60d) are presented in **Table E3**.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**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 in supplemental methods**).

*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 since adjustment for multiple variables can lead to small sample 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 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

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

interaction included the independent variables of interest: COVID-19 infection status, rs35705950-T, and their multiplicative interaction term, adjusted for pre-index pneumonia (yes/no, within 2 years pre-index), 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 0,1,2) so that the interaction odds ratio (OR) is equal for heterozygous vs homozygous WT and homozygous mutation vs heterozygous. 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)*

Associations between rs35705950-T and pre-existing comorbid conditions and laboratory values were examined using clinical data prior to the COVID-19 pandemic (Sept 2019). Individuals with ≥2 Phecodes(25) were defined as cases. Phecodes with <200 cases within each ancestry group were excluded, resulting in 1618 (EUR, 1289 (AFR), 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 $1.2 \times 10^{-05}$ (0.05/3901) accounted for all the models tested across three ancestries for significance. Analyses were performed using PLINK2(26) (Additional details in **Supplemental Methods**).

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

## 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, HIS). The minor allele frequency for rs35705950-T was 9% among the entire MVP and 8% among COVID-19 positive individuals (**Table 1**); among these 14.4% (n=2,758) and 0.7% (n=138) were carriers of 1-copy and 2-copies of *MUC5B* rs35705950-T, respectively (**Table 1, Figure E1)**. Associations between rs35705950-T and clinical outcomes by ancestry group within MVP, trans-ancestry meta-analysis within MVP (MVP-ALL), and joint meta-analysis with HGI (META) 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 [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 trans- ancestry meta-analysis within 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 3 ancestry groups (EUR, AFR, 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

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

primary analyses (**Table E4**), with the association between rs35705950-T allele and reduced risk of hospitalization, compared with non-hospitalization, due to COVID-19 remaining robust (OR= 0.86 [0.77, 0.95], P=0.004, **Table E4**). The OR and 95% CI were very similar to the primary analyses (OR=0.88 [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 sub-population. 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=8541), the adjusted odds ratio for post-index 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 post-index pneumonia non-significantly 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 post-index 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 0, 1, and 2 copies of MUC5B rs35705950-T respectively among COVID-19 positive relative to negative subjects ($p < 0.0001$, **Table 3**). This differential effect of an additional *MUC5B* rs35705950-T allele on post-index pneumonia in COVID-19 positive vs. COVID-19 negative patients was statistically significant (multiplicative scale interaction OR = 0.77 [0.66, 0.89], p=0.0004) in EUR (**Table 3**). There was also evidence of a negative additive interaction, with the relative excess risk due to interaction (RERI) corresponding to an increase

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

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 was too low to permit further analysis.

*Exploring shared pathobiology with the MUC5B rs35705950-T through PheWAS and LabWAS*

To explore clinical conditions and biomarkers associated with the *MUC5B* rs35705950-T allele which may impact the susceptibility and severity of COVID-19, PheWAS and LabWAS using pre-COVID-era data (through Sept 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 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 x $10^{-289}$) and "post-inflammatory pulmonary fibrosis" (Phecode 502, OR = 2.85 [2.65 - 3.05], P = 8.90 x $10^{-186}$) were significantly associated ($P_{bonferroni} < 1.2 \times 10^{-5}$) with rs35705950-T in all three ancestry groups (EUR, AFR, 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

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

influenza and 10,579 cases of bacterial pneumonia in 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 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 following 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 confer the differential risks observed.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

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 towards 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, while 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 rs35 705950 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 which 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 review(27). It should be noted, however, that the underlying etiologies for ARDS in the population examined were heterogeneous, with over half of 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

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

abnormalities (ILA), defined as non-dependent parenchymal infiltrates which are frequently considered subclinical precursors to clinically-apparent interstitial lung disease (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 post-index 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 pre-pandemic 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 0, 1, and 2 copies of rs35705950-T, respectively), individuals with the "T" allele demonstrated a trend towards 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 post-index 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 non-respiratory disorders, such as shock and multi-system organ

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

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(40). 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)(41–43); 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 et al(44), reported minimal changes in 30-day mortality rates for heart failure and pneumonia hospitalizations following 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 $\geq$ 1 primary care visit within the preceding 18 months and comparable numbers of patients with $\geq$ 1 clinical encounter or admission in the preceding year. However, due to 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 & 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

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

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, due to the timing of the data freeze for our analysis, only short-term outcomes were captured. F uture 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(45), individuals harboring the *MUC5B* variant may exhibit behavior modifications (increased testing or more stringent self-isolation ) which could have introduced biases not 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 our results, which identify a potentially protective role of rs35705950-T in

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

COVID-19 outcomes, support that this functional polymorphism confers pleiotropic effects which may be modulated by epistasis and environmental interactions; longer term follow-up and future analyses into these complex relationships are warranted.

**Acknowledgments**

This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported in part by MVP035 and MVP008 (1I01BX004188-01). This publication does not represent the views of the Department of Veteran Affairs of the United States Government.

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 supplemental methods.

**Conflict of Interest**

CJO is an employee of Novartis Institute for Biomedical Research. PN reports grant support from Amgen, Apple, AstraZeneca, Boston Scientific, and Novartis, personal fees from Apple, AstraZeneca, Blackstone Life Sciences, Genentech, and Novartis, and spousal employment at Vertex, all unrelated to the present work.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

## References


1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020 May;20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1. Epub 2020 Feb 19. Erratum in: Lancet Infect Dis. 2020 Sep;20(9):e215. PMID: 32087114; PMCID: PMC7159018. "Accessed on Feb 16th, 2021"

2. Zhao Y-M, Shang Y-M, Song W-B, Li Q-Q, Xie H, Xu Q-F, Jia J-L, Li L-M, Mao H-L, Zhou X-M, Luo H, Gao Y-F, Xu A-G. Follow-up study of the pulmonary function and related physiological characteristics of COVID-19 survivors three months after recovery. *EClinicalMedicine* 2020;

3. Yan X, Huang H, Wang C, Jin Z, Zhang Z, He J, Yin S, Fan M, Huang J, Chen F, Zeng Y, Han X, Zhu Y. Follow-up study of pulmonary function among COVID-19 survivors 1 year after recovery. *J Infect* 2021;doi:10.1016/j.jinf.2021.05.034.

4. McGroder CF, Zhang D, Choudhury MA, Salvatore MM, D'Souza BM, Hoffman EA, Wei Y, Baldwin MR, Garcia CK. Pulmonary fibrosis 4 months after COVID-19 is associated with severity of illness and blood leucocyte telomere length. *Thorax* 2021;doi:10.1136/thoraxjnl-2021-217031.

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.


AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

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, Lindell KO, Branch RA, Nicolae D, Sciurba F, Garcia AN, Kaminski N. A Variant In The Promoter Of MUC5B Is Associated With Idiopathic Pulmonary Fibrosis And Not Chronic Obstructive Pulmonary Disease. *B102 INTERSTITIAL LUNG DISEASE: NOVEL MANAGEMENT AND OUTCOME STRATEGIES* 2011;doi:10.1164/ajrccm-conference.2011.183.1_meetingabstracts.a6395.

11. Moore C, Blumhagen RZ, Yang IV, Walts A, Powers J, Walker T, Bishop M, Russell P, Vestal B, Cardwell J, Markin CR, Mathai SK, Schwarz MI, Steele MP, Lee J, Brown KK, Loyd JE, Crapo JD, Silverman EK, Cho MH, James JA, Guthridge JM, Cogan JD, Kropski JA, Swigris JJ, Bair C, Kim DS, Ji W, Kim H, *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, Chu HW, Hou X, Poirion OB, Buchanan J, Preissl S, Ren B, Colgan SP, Dowell RD, Yang IV, Schwartz DA, Gerber AN. The MUC5B-associated variant rs35705950 resides within an enhancer subject to lineage- and disease-dependent epigenetic remodeling. *JCI Insight* 2021;6.:

13. Peljto AL, Zhang Y, Fingerlin TE, Ma S-F, Garcia JGN, Richards TJ, Silveira LJ, Lindell KO, Steele MP, Loyd JE, Gibson KF, Seibold MA, Brown KK, Talbert JL, Markin C, Kossen K, Seiwert SD, Murphy E, Noth I, Schwarz MI, Kaminski N, Schwartz DA. Association between the MUC5B promoter

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

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, Wain LV, Hingorani AD, Patel RS. 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, Whitbourne S, Deen J, Shannon C, Humphries D, Guarino P, Aslan M, Anderson D, LaFleur R, Hammond T, Schaa K, Moser J, Huang G, Muralidhar S, Przygodzki R, O'Leary TJ. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology* 2016;

16. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* 2021;doi:10.1038/s41586-021-03767-x.

17. Fang H, Hui Q, Lynch J, Honerlaw J, Assimes TL, Huang J, Vujkovic M, Damrauer SM, Pyarajan S, Gaziano JM, DuVall SL, O'Donnell CJ, Cho K, Chang K-M, Wilson PWF, Tsao PS, VA Million Veteran Program, Sun YV, Tang H. 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, Hilliard AT, Joseph J, O'Donnell CJ, Efird JT, Crawford DC, Wu W-C, Arjomandi M, Sun YV, Assimes TL, Huffman JE, VA Million Veteran Program COVID-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.:

19. Ioannou GN, Locke E, Green P, Berry K, O'Hare AM, Shah JA, Crothers K, Eastment MC, Dominitz JA,

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

Fan VS. 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, Kannengiesser C, Burgers JA, Grutters JC, van Moorsel CHM. The promoter risk allele for idiopathic pulmonary fibrosis predisposes to asbestosis. *Eur Respir J* 2020;55.:

23. Juge P-A, Lee JS, Ebstein E, Furukawa H, Dobrinskikh E, Gazal S, Kannengiesser C, Ottaviani S, Oka S, Tohma S, Tsuchiya N, Rojas-Serrano J, González-Pérez MI, Mejía M, Buendía-Roldán I, Falfán-Valencia R, Ambrocio-Ortiz E, Manali E, Papiris SA, Karageorgas T, Boumpas D, Antoniou K, van Moorsel CHM, van der Vis J, de Man YA, Grutters JC, Wang Y, Borie R, Wemeau-Stervinou L, *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, Fingerlin TE, Zhang W, Gudmundsson G, Groshong SD, Evans CM, Garantziotis S, Adler KB, Dickey BF, du Bois RM, Yang IV, Herron A, Kervitsky D, Talbert JL, Markin C, Park J, Crews AL, Slifer SH, Auerbach S, Roy MG, Lin J, Hennessy CE, Schwarz MI, Schwartz DA. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364:1503–1512.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

25. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. 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;

27. Rogers AJ, Solus JF, Hunninghake GM, Baron RM, Meyer NJ, Janz DR, Schwartz DA, May AK, Lawson WE, Blackwell TS, Ware LB. 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, COVID-19 Host Genetics Initiative, Baillie JK, Wain LV, Jenkins RG, Feenstra B. 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, Ruven HJT, Quanjel MJR, Grutters JC. The Promoter Polymorphism Associates With Severe COVID-19 in the European Population. *Front Med* 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, Schwarz MI, Schwartz DA, Reynolds SD. 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, Lara AR, Schwarz MI, Evans CM, Schwartz DA. MUC5BPromoter Variant rs35705950 Affects MUC5B Expression in the Distal Airways

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

in Idiopathic Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine* 2016;

33. Evans CM, Fingerlin TE, Schwarz MI, Lynch D, Kurche J, Warg L, Yang IV, Schwartz DA. 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, Stock CJ, Bonifazi M, Bonini M, Caramori G, Lindahl G, Blasi FB, Nicholson AG, Wells AU, Sestini P, Renzoni E. 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, Bowler RP, Regan EA, Curtis JL, Han MK, Boucher RC, O'Neal WK, Hatabu H, Lynch DA, Rosas IO, Hunninghake GM, San Jose Estepar R, Washko GR, 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, Alexander SN, Bellinghausen LK, Song AS, Petrova YM, Tuvim MJ, Adachi R, Romo I, Bordt AS, Bowden MG, Sisson JH, Woodruff PG, Thornton DJ, Rousseau K, De la Garza MM, Moghaddam SJ, Karmouty-Quintana H, Blackburn MR, Drouin SM, Davis CW, Terrell KA, Grubb BR, O'Neal WK, Flores SC, *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. Mukae H, Iiboshi H, Nakazato M, Hiratsuka T, Tokojima M, Abe K, Ashitani J, Kadota J, Matsukura S,

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

Kohno S. Raised plasma concentrations of alpha-defensins in patients with idiopathic pulmonary fibrosis. *Thorax* 2002;57:623–628.

39. Konishi K, Gibson KF, Lindell KO, Richards TJ, Zhang Y, Dhir R, Bisceglia M, Gilbert S, Yousem SA, Song JW, Kim DS, Kaminski N. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2009;180:167–175.

40. Magesh S, John D, Li WT, Li Y, Mattingly-App A, Jain S, Chang EY, Ongkeko WM. Disparities in COVID-19 Outcomes by Race, Ethnicity, and Socioeconomic Status: A Systematic-Review and Meta-analysis. *JAMA Netw Open* 2021;4:e2134147.

41. Silva GC, Jiang L, Gutman R, Wu W-C, Mor V, Fine MJ, Kressin NR, Trivedi AN. 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.

42. Kovesdy CP, Norris KC, Boulware LE, Lu JL, Ma JZ, Streja E, Molnar MZ, Kalantar-Zadeh K. Association of Race With Mortality and Cardiovascular Events in a Large Cohort of US Veterans. *Circulation* 2015;132:1538–1548.

43. Sheehan CM, Hayward MD. Black/white differences in mortality among veteran and non-veteran males. *Soc Sci Res* 2019;79:101–114.

44. Trivedi AN, Jiang L, Silva G, Wu W-C, Mor V, Fine MJ, Kressin NR, Gutman R. Evaluation of Changes in Veterans Affairs Medical Centers' Mortality Rates After Risk Adjustment for Socioeconomic Status. *JAMA Netw Open* 2020;3:e2024345.

45. Hunninghake GM, Hatabu H, Okajima Y, Gao W, Dupuis J, Latourelle JC, Nishino M, Araki T, Zazueta

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

OE, Kurugol S, Ross JC, San José Estépar R, Murphy E, Steele MP, Loyd JE, Schwarz MI, Fingerlin TE, Rosas IO, Washko GR, O'Connor GT, Schwartz DA. MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med* 2013;368:2192–2200.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

## Tables and Figures

**Table 1.** Demographics for COVID-19 tested positive and all MVP participants examined in this study.

| Characteristics | Million Veteran Program Number (%) | COVID-19 Positive Number (%) |
|---|---|---|
| Total Patients | 658,582[Ψ] | 19,168 |
| Male | 592,516 (90) | 17,151 (89) |
| **Genetic Ancestry** | | |
| 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** | | |
| 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) | 100,46 (1.5) | 335 (2) |
| **Outcomes** | | |
| Hospitalized | - | 4,234 (22) |
| Severe/ Deceased | - | 947 (5) |

[Ψ] *MVP participants who died before March 2020 were excluded from the analysis described in **Table 2**.*
** Due to small sample size, individuals from Asian and other ancestry group were not included*

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**Table 2**. Association between rs35705950-T in MUC5B with COVID-19 clinical outcomes.

| Analysis | Population | Group | Case \| Controls | Sample Size | EA: EAF | OR 95% CI | P |
|---|---|---|---|---|---|---|---|
| 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 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] | **$6.86 \times 10^{-03}$** |
| | HGI | META | 9,086 \| 1,001,201 | 1,010,287 | T: 0.106 | 0.91 [0.85-0.97] | **$4.12 \times 10^{-03}$** |
| | MVP + HGI | META | 13,320 \| 1,508,841 | 1,522,160 | T: 0.1 | 0.9 [0.86-0.95] | **$8.99 \times 10^{-05}$** |
| COVID-19 Hospitalized Vs COVID-19 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 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] | **$6.69 \times 10^{-05}$** |
| | MVP + HGI | META | 4,704 \| 1,210,959 | 1,215,663 | T: 0.1 | 0.93 [0.79-1.1] | 0.410 |

Odds ratio (OR) and 95% confidence interval (95% CI) reported for the minor (T) allele. Results are shown for VA Million Veteran Program (MVP) African Americans (AFR), European Americans (EUR), Hispanic/Latino Americans (HIS), and multi- ancestry meta-analysis (ALL), the COVID-19 Host Genetics Initiative (HGI) multi- ancestry round 5 meta-analysis excluding MVP and 23&Me, and the meta-analysis of MVP and HGI (META).

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**Table 3**. Associations between *MUC5B* rs35705950-T allele and pneumonia within 60 days of COVID-19 testing among MVP participants of European ancestry.

| **Odds Ratios (95% Confidence Intervals) for Pneumonia within 60 days of COVID-19 test** | | | |
|---|---|---|---|
| | | **OR 95% CI** | **P** |
| **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+ versus COVID-) 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)** | 0.77 (0.66, 0.89) p=0.0004 | | |
| **Additive Scale RERI (95% CI)** | -2.07 (-3.24, -0.9) | | |
| Odds ratios and 95% confidence intervals are estimated from Firth logistic regression adjusting for pre-index pneumonia, age, age$^2$, sex, BMI, CCI, smoking, IPF, asbestosis, RA-ILD, and principal components PC1-20, including a multiplicative interaction between additive MUC5B rs35705950-T allele and COVID-19 infection. An estimate of the relative excess risk due to interaction (RERI) and 95% confidence interval (Delta method) corresponding to an additive interaction effect is estimated from the logistic regression model. | | | |

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

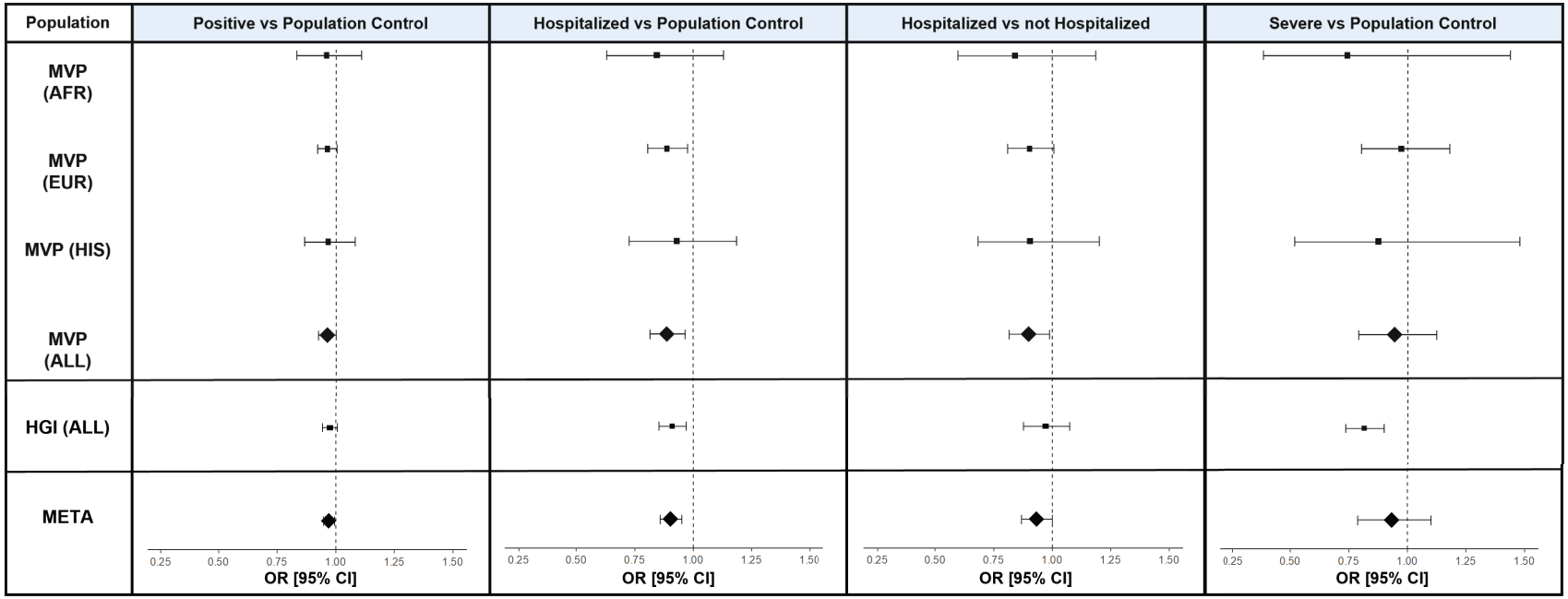
**Figure Legends**

**Figure 1. Forest plot association of rs35705950-T in MUC5B** with (i) COVID-19 Positive vs Population Controls, (ii) COVID-19 Positive, Hospitalized vs Population Controls, (iii) COVID-19 Positive, Hospitalized vs COVID-19 Positive, not Hospitalized, and (iv) Hospitalized for COVID-19 with high-flow oxygen or died of COVID-19 vs Population Controls. Odds ratio (OR) and 95% confidence interval (95% CI) is reported for the minor (T) allele, and results are shown for VA Million Veteran Program (MVP) African Americans (AFR), European Americans (EUR), Hispanic/Latino Americans (HIS), and multi- ancestry meta-analysis (ALL), the COVID-19 Host Genetics Initiative (HGI) multi- ancestry round 5 meta-analysis excluding MVP and 23&Me, and the meta-analysis of MVP and HGI (META).

**Figure 2**. **Phenome-Wide Association Study (PheWAS) of *MUC5B* rs35705950-T allele in the Million Veteran Program.** A PheWAS plot shows associations of rs35705950-T and phenotypes derived from the electronic health records data prior to COVID-19 in MVP participants from European ancestry. The phenotypes are shown on the x-axis and organized by disease categories. The p-value (-log10) 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 as increased risk and the downward triangle as reduced risk. The red line indicates the significance threshold based on the Bonferroni correction. The forest plot of Bonferroni significant associations are shown on the right top corner of the PheWAS plot.

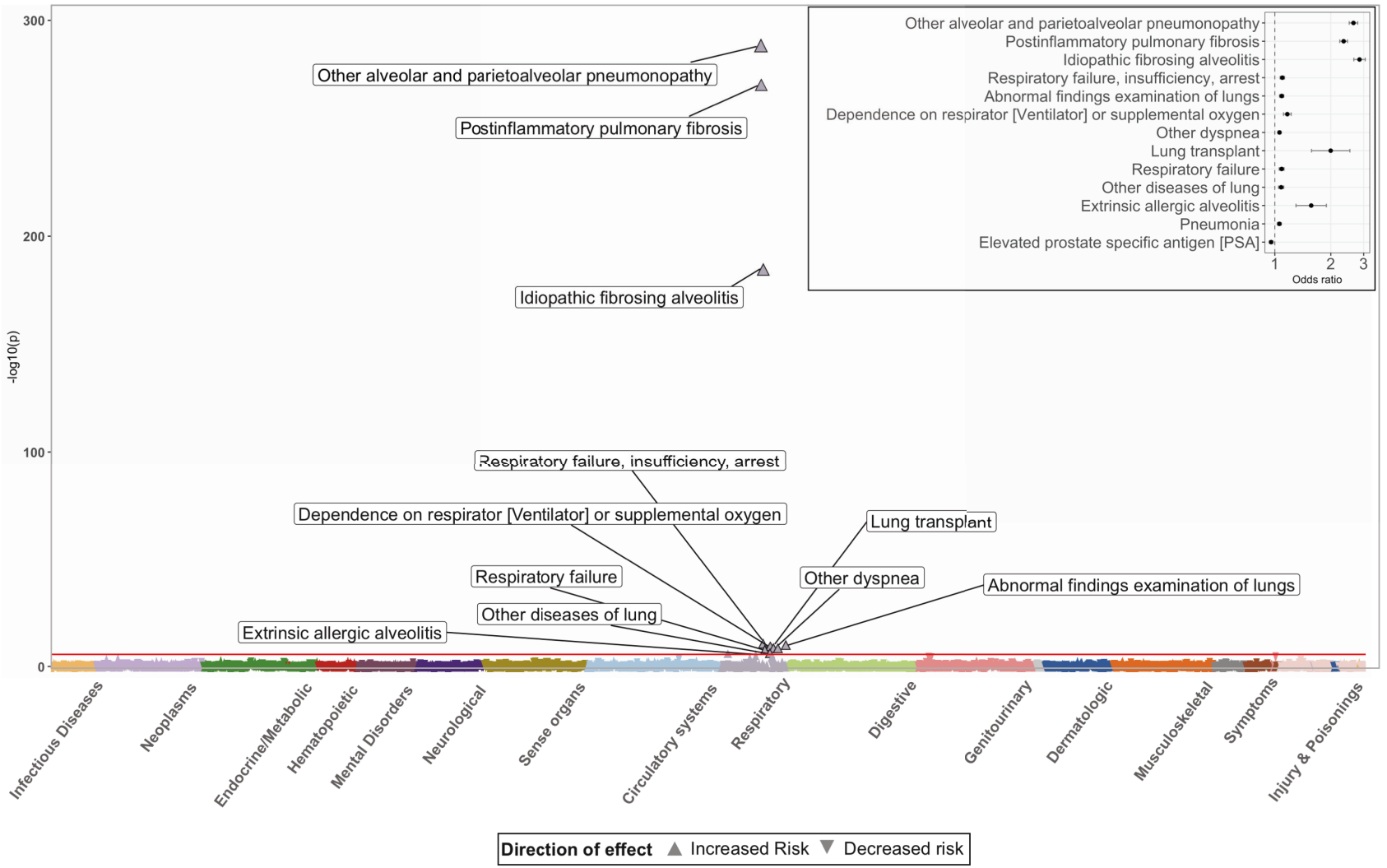
AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

Figure 1




AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

Figure 2




AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**Supplemental Materials**

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

**Supplemental Methods.**

- **Figure E1.** Overview of genetic association of *MUC5B* allele and COVID-19 phenotypes as well as phenome-wide association study.
- **Figure E2. Phenome-Wide Association Study (PheWAS) of *MUC5B* rs35705950-T allele in three Million Veteran Program ancestry groups.** A PheWAS plot shows associations of rs35705950-T and phenotypes derived from the electronic health records data prior to COVID-19 in MVP participants from A) European ancestry B) African ancestry and C) Hispanic ancestry.
- **Figure E3. Laboratory-Wide Association Study (PheWAS) of *MUC5B* rs35705950-T allele in the Million Veteran Program.** A LabWAS plot shows associations of rs35705950-T and median values of laboratory measures extracted from electronic health records data prior to COVID-19 in MVP participants. The bottom panel shows the -log10 (p-value) on the y-axis and laboratory test descriptions on the x-axis. Triangles points up have increasing effects and points down have decreasing effects. The colors represent the different ancestry groups. The top panel shows beta from the regression model for each laboratory measure. The significant results are highlighted in the color corresponding to ancestry groups and other results are plotted in grey.

- **Table E1**. The demographics and clinical conditions were documented within two years prior to the index dates for MVP subjects used for association studies with severe outcomes of COVID-19 infection and post-index pneumonia events (pneumonia60d) in this study. **See Excel Document**
- **Table E2.** Sample sizes and rates of pneumonia within 60 days post index date for MVP subjects

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

of European ancestry with no missing data for pre- or post- index pneumonia events after sensitivity analysis.

- **Table E3.** The ICD codes used for data pull of post-index pneumonia diagnosis (pneumonia60d) in this study. **See Excel Document**
- **Table E4.** Association of rs35705950-T allele with COVID-19 outcomes: 1) Testing positive, 2) hospitalization, 3) Severe COVID-19 + Death and 4) COVID-19 related Death in MVP participants after sensitivity analysis.
- **Table E5**. Full summary statistics of PheWAS associations with *MUC5B* rs35705950-T allele. **See Excel Document**
- **Table E6.** LabWAS reveals laboratory measurements associated with the presence of a *MUC5B* rs35705950-T allele. **See Excel Document**

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

## Supplemental Methods

### Data Sources

In this COVID-19 study, we used data from MVP, a large multi-ethnic genetic biobank at Veterans Affairs. MVP began enrolling veterans in 2011 and has now enrolled over 850,000 veterans and genotyped over 650,000. The Veterans' EHR data included diagnosis codes (ICD-9 and ICD-10), current procedural terminology codes (CPT), clinical laboratory measures, and demographic, lifestyle, and radiology reports (1). Pre-COVID EHR data were collected from participants from the time of enrollment to September 30, 2019.

The MVP cohort SNP data was generated using a custom Thermo Fisher Axiom genotyping platform named MVP 1.0. The MVP cohort quality control and genotyping imputation steps were previously reported(2). Ancestry of participants was defined using Harmonized ancestry, race, and ethnicity (HARE), a variable derived from self-reported survey data and genetically derived ancestry(3). Directly genotyped rs35705950-T information was extracted and utilized for association testing. Relatedness was estimated from Identity-by-descent probabilities and one individual from each related pair with kinship coefficient > 0.125 were dropped. We prioritize retained individuals for cases in post COVID-19 pneumonia, Muc5B rs35705950-T allele, covid death, severe covid outcome, covid hospitalization, and covid testing positive, in this order.

A VA-wide collective effort lead by the VA Informatics and Computing Infrastructure (VINCI) created a shared COVID-19 Shared Data Resource (SDR) during the April-August 2020 period, providing a new data domain related to COVID-19 for the VA. All metadata on documentation regarding conditions, laboratory measures, medications, and procedures pertaining to the COVID-19 pandemic and dissemination of information regarding SDR were provided by the VA Phenomics Library, Centralized

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

Interactive Phenomics Resource (CIPHER), to users across the VA healthcare systems. The VA COVID-19 SDR group including the VA National Surveillance Team provided researchers with curated data extracted from electronic health records (EHR) of Veterans, after stripping off identifiable elements, and ensuring data security by requiring restricted access to the data(4). This activity was undertaken centrally by an experienced VA research team in order to ensure consistent applications of dates across the EHR tables, as well as uniform definitions of events preceding and subsequent to COVID-19, in an effort to characterize the trajectory of COVID-19 and other diseases and conditions. VA-wide efforts describing these data are already in public domain(5). Structured data obtained via CPT and ICD9 or ICD10 codes, deposited in the VA Corporate Data Warehouse (CDW),was further enriched with rule-based unstructured events recorded in patient notes via natural language processing (NLP), as was done previously for other projects(6). The purpose of NLP-boosting was to fill gaps in knowledge about the severity of the disease, as well as extract specific dates when procedures (e.g. intubation and extubation) were performed. These curated data were then provisioned in the Million Veteran Program (MVP) study mart for COVID-19 in VINCI ensuring data security behind the VA firewall.

*COVID-19 outcome definitions*

The primary analysis *MUC5B* SNP rs35705950-T with COVID-19 infection and hospitalization included all the patients who were tested from Feb 2020 through April 2021 using RT-PCR based method(7, 8) or self-reported during primary care visit. These patients may have received their COVID-19 PCR testing inside as well as outside of the VA. However, for analysis of the outcome severity of COVID-19 infection such as ICU stay or death (more details below), we only included patients tested within VA. The *index date* was defined as a COVID-19 diagnosis date, i.e., specimen date, or a self-reported date of diagnosis; and for a hospitalized patient, the admission date up to 15 days prior to the COVID-19 case date. "*Pre-index*" and "*post-index*" are hereafter used to denote events and outcomes before and after this *index*

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

*date*. To define the COVID-19 related severity and death, we use a modified version of COVID-19 Severity Scale variable from WHO COVID-19 Disease Progression Scale(9). The current version of VA's adaptation of COVID-19 Severity Index classifies severity into mild, moderate, severe and death. The severity rating of COVID-19 infection for the MVP cohort was modified from the WHO criteria to classify into four groups: mild (testing positive for SARS-CoV-2 by PCR); moderate (SARS-CoV-2 positivity and hospitalization within 30 days of index dates); severe (SARS-CoV-2 positivity, hospitalized and having indication of treatment by noninvasive mechanical ventilation (NIV), high flow oxygen support, extracorporeal membrane oxygenation (ECMO), intubation, vasopressor support, and dialysis; and death within 30 days of index dates for COVID-19 infection (10)). All data and variables were provided by the MVP data core.

We applied two approaches to define COVID-19 outcomes. First, for the meta-analysis with HGI summary statistics, we synchronized the phenotype definitions as used by the consortium. The following binary COVID-19 outcomes were examined: (1) COVID Susceptibility: individuals who tested positive for COVID-19 (2) COVID Hospitalization: individuals who were hospitalized due to symptoms of COVID-19 (3) COVID-19 Critically ill: individuals who were hospitalized and required respiratory support.

Secondly, we conducted sensitivity analysis within VA cohort on further refined four binary outcomes using patients who received COVID-19 PCR testing at VA sites: 1) COVID-19 testing: individuals who tested positive vs individuals with negative results, 2) COVID-19 hospitalization: individuals who tested positive and required hospitalization due to COVID-19 vs individuals who tested positive but did not need hospitalization, 3) COVID-19 severity: individuals who tested positive and required respiratory support or died due to complication from COVID-19 vs individuals who were hospitalized due to COVID-19 but didn't require respiratory support 4) and COVID-19 mortality: death within 30 days of COVID-19 infection vs. all other COVID-19 test positives.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

*Phenotyping of pre- and post-index conditions for COVID-19*

To determine pre-index conditions, we used NLP-boosted unstructured notes, ICD and CPT codes, and medications taken 2 years prior. ICD and CPT codes and medications 60 days after the index date were used to derive post-index conditions. Pre- and post-index conditions were obtained from the Shared Data Resource's pre- and post-index tables (SDR). The study's post-index conditions' ICD codes are in the supplemental materials. Post-index pneumonia (Pneumonia60d) was studied 60 days after COVID-19 index dates (**Table E2**).

For the 60-day post-index analysis, the SDR post-index tables were used. We included only patients whose PCR testing was done at the VA. Additionally patients that were initially tested negative in the Patient table but then turned positive in the post-index table were excluded from the post index analysis. Participants who lost follow up in the post-index table were also excluded.

**Meta-analysis with HGI**

Results from the COVID-19 Host Genetics Initiative (HGI)(11) were utilized for replication and meta-analysis with our results. Since MVP contributed to HGI, we used the "leave one out" results that were available only to contributing cohorts. These are currently available only for Release 5 (available January 18, 2021), not the more recent Release 6 (June 15, 2021), and also excluded 23&Me, per their agreement with HGI. Results for rs35705950-T were extracted for both the European-only and trans-ethnic meta-analysis across all phenotype definitions.

Logistic regression was performed to determine the association of rs35705950-T with each of these outcomes within each of three HARE AFR, HIS, and EUR ancestry) within plink2a(12) adjusting for age, sex, sex$^2$, and the first 15 ethnicity-specific principal components. Inverse-variance weighted meta-analysis was performed in GWAMA(13) to combine MVP with HGI.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**Phenome-wide association study (PheWAS)**

For the PheWAS analysis on the *MUC5B* rs35705950-T allele, we used Phecodes (14) and laboratory measures from the clinical data available prior to the onset of COVID-19 infection (Sept 2019). Individuals with two or more Phecodes were defined as cases, while those without were defined as controls. In our previous study, we found that cases under 200 provide >80% statistical power and low type I error therefore we excluded phecodes with fewer than 200 cases within each ancestry group. It yielded 1618 (EUR), 1289 (AFR), 994 (HIS) Phecodes. We also performed a laboratory-wide association study using 69 clinical labs extracted from the EHR of MVP participants. We used the median of each person's entire lab history. We used logistic regression for Phecodes and linear regression for laboratory measurements in PLINK2(15). When the logistic regression model failed to converge for binary outcomes, firth regression was used. Regression models were adjusted for sex, age (at the time of data freeze), quadratic term of age, and the first 20 principal components. We applied Bonferroni correction to each ancestry specific analysis to adjust for multiple hypothesis testing, and following threshold were use select significant associations: EUR = $3.09 \times 10^{-05}$ (0.05/1618), AFR = $3.8 \times 10^{-05}$ (0.05/1289), EUR = $5.03 \times 10^{-05}$(0.05/994).

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**References**


1. Knight KE, Honerlaw J, Danciu I, Linares F, Ho Y-L, Gagnon DR, Rush E, Gaziano JM, Begoli E, Cho K, VA Million Veteran Program. Standardized Architecture for a Mega-Biobank Phenomic Library: The Million Veteran Program (MVP). *AMIA Jt Summits Transl Sci Proc* 2020;2020:326–334.

2. Hunter-Zinck H, Shi Y, Li M, Gorman BR, Ji S-G, Sun N, Webster T, Liem A, Hsieh P, Devineni P, Karnam P, Gong X, Radhakrishnan L, Schmidt J, Assimes TL, Huang J, Pan C, Humphries D, Brophy M, Moser J, Muralidhar S, Huang GD, Przygodzki R, Concato J, Gaziano JM, Gelernter J, O'Donnell CJ, Hauser ER, Zhao H, *et al.* Genotyping Array Design and Data Quality Control in the Million Veteran Program. *Am J Hum Genet* 2020;106:535–548.

3. Fang H, Hui Q, Lynch J, Honerlaw J, Assimes TL, Huang J, Vujkovic M, Damrauer SM, Pyarajan S, Gaziano JM, DuVall SL, O'Donnell CJ, Cho K, Chang K-M, Wilson PWF, Tsao PS, VA Million Veteran Program, Sun YV, Tang H. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am J Hum Genet* 2019;105:763–772.

4. COVID-19 Efforts. at <https://www.hsrd.research.va.gov/covid19.cfm>.

5. Song RJ, Ho Y-L, Schubert P, Park Y, Posner D, Lord EM, Costa L, Gerlovin H, Kurgansky KE, Anglin-Foote T, DuVall S, Huffman JE, Pyarajan S, Beckham JC, Chang K-M, Liao KP, Djousse L, Gagnon DR, Whitbourne SB, Ramoni R, Muralidhar S, Tsao PS, O'Donnell CJ, Gaziano JM, Casas JP, Cho K, VA Million Veteran Program COVID-19 Science Initiative. Phenome-wide association of 1809 phenotypes and COVID-19 disease progression in the Veterans Health Administration Million Veteran Program. *PLoS One* 2021;16:e0251651.

6. Patterson OV, Jones M, Yao Y, Viernes B, Alba PR, Iwashyna TJ, DuVall SL. Extraction of Vital Signs


AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

from Clinical Notes. *Stud Health Technol Inform* 2015;216:1035.

7. Ioannou GN, Locke E, Green P, Berry K, O'Hare AM, Shah JA, Crothers K, Eastment MC, Dominitz JA, Fan VS. Risk Factors for Hospitalization, Mechanical Ventilation, or Death Among 10 131 US Veterans With SARS-CoV-2 Infection. *JAMA Netw Open* 2020;3:e2022310.

8. 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.

9. 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.

10. 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.

11. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020 May;20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1. Epub 2020 Feb 19. Erratum in: Lancet Infect Dis. 2020 Sep;20(9):e215. PMID: 32087114; PMCID: PMC7159018. "Accessed on Feb 16th, 2021"

12. 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;

13. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* 2010;11:288.

14. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR,

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

Roden DM, Crawford DC. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 2010;26:1205–1210.

15. 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;

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**Supplemental Tables**

**Table E1**. The demographics and clinical conditions documented within two years prior to the index dates for MVP subjects used for association studies with severe outcomes of COVID-19 infection and post-index pneumonia events (pneumonia60d) in this study. **See Excel Document**

**Table E2.** Sample sizes and rates of pneumonia within 60 days post index date for MVP subjects of European ancestry with no missing data for pre- or post- index pneumonia events after sensitivity analysis.

| | **rs35705950-T _0 copy** | | **rs35705950-T _1 copy** | | **rs35705950-T _2 copies** | |
|---|---|---|---|---|---|---|
| *HARE* | Pneumonia_no | Pneumonia_yes | Pneumonia_no | Pneumonia_yes | Pneumonia_no | Pneumonia_yes |
| **COVID-19 Tested Positive** | | | | | | |
| *EUR* | 4994 | 1854 | 1228 | 372 | 76 | 17 |
| **COVID-19 Tested Negative** | | | | | | |
| *EUR* | 62076 | 2593 | 15267 | 689 | 941 | 47 |

**Table E3.** The ICD codes used for data pull of post-index pneumonia diagnosis (pneumonia60d) in this study. **See Excel Document**

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**Table E4.** Association of rs35705950-T allele with COVID-19 outcomes: 1) Testing positive, 2) hospitalization, 3) Severe COVID-19 + Death and 4) COVID-19 related Death in MVP participants after sensitivity analysis

| MVP Cohort Ancestry | Copy number_0 | | Copy number_1 | | Copy number_2 | | OR [95% CI] | P_value |
|---|---|---|---|---|---|---|---|---|
| COVID-19 Testing | COVID-19 Negative | COVID-19 Positive | COVID-19 Negative | COVID-19 Positive | COVID-19 Negative | COVID-19 Positive | | |
| African | 25992 | 3701 | 1126 | 151 | 18 | 2 | 0.98 [0.83, 1.16] | 0.83 |
| European | 64759 | 6860 | 15981 | 1604 | 989 | 93 | 0.95 [0.90, 1.00] | 0.05 |
| Hispanic | 9450 | 1654 | 1568 | 248 | 64 | 9 | 0.94 [0.82, 1.07] | 0.34 |
| Random Effect Meta-analysis | | | | | | | 0.95 [0.90, 0.99] | **0.03** |
| COVID-19 Hospitalization | COVID-19 Not-hospitalized | COVID-19 Hospitalized | COVID-19 Not-Hospitalized | COVID-19 Hospitalized | COVID-19 Not-Hospitalized | COVID-19 Hospitalized | | |
| African | 2322 | 1379 | 104 | 47 | 1 | 1 | 0.79 [0.55, 1.13) | 0.21 |
| European | 4607 | 2241 | 1143 | 457 | 65 | 28 | 0.86 [0.77, 0.97] | **0.015** |
| Hispanic | 1189 | 459 | 179 | 69 | 9 | 0 | 0.88 [0.65, 1.20] | 0.49 |
| Random Effect Meta-analysis | | | | | | | 0.86 [0.78, 0.95] | **0.004** |
| COVID-19 Severe + Death | COVID-19 Not-Severe | COVID-19 Severe + Death | COVID-19 Not-Severe | COVID-19 Severe + Death | COVID-19 Not-Severe | COVID-19 Severe + Death | | |
| African | 3314 | 387 | 138 | 13 | 2 | 0 | 0.82 [0.46, 1.45] | 0.46 |
| European | 6257 | 591 | 1463 | 137 | 87 | 6 | 1.01 [0.85, 1.21] | 0.71 |
| Hispanic | 1502 | 146 | 232 | 16 | 9 | 0 | 0.69 [0.41, 1.17] | 0.13 |
| Random Effect Meta-analysis | | | | | | | 0.92 [0.72, 1.16] | 0.48 |
| COVID-19 Related Death | COVID-19 Survivors | COVID-19 Deaths | COVID-19 Survivors | COVID-19 Deaths | COVID-19 Survivors | COVID-19 Deaths | | |
| African | 3547 | 154 | 143 | 8 | 2 | 0 | 1.41 [0.69, 2.89] | 0.42 |
| European | 6530 | 318 | 1535 | 65 | 90 | 3 | 0.91 [0.71, 1.17] | 0.59 |
| Hispanic | 1579 | 69 | 241 | 7 | 9 | 0 | 0.67 [0.32, 1.40] | 0.30 |
| Random Effect Meta-analysis | | | | | | | 0.92 [0.74, 1.16) | 0.50 |

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**Table E5**. Full summary statistics of PheWAS associations with *MUC5B* rs35705950-T allele. **See Excel Document**

**Table E6.** LabWAS reveals disease conditions associated with the presence of a *MUC5B* rs35705950-T allele. **See Excel Document**

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

**Acknowledgements**

**VA Million Veteran Program COVID-19 Science Initiative Membership & Acknowledgements**

**VA Million Veteran Program COVID-19 Science Initiative**

**MVP COVID-19 Science Program Steering Committee**

- Christopher J. O'Donnell, M.D., M.P.H. (Co-Chair)

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- J. Michael Gaziano, M.D., M.P.H. (Co-Chair)

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Philip S. Tsao, Ph.D. (Co-Chair)

  VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

- Sumitra Muralidhar, Ph.D.

  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

- Jean Beckham, Ph.D.

  Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705

- Kyong-Mi Chang, M.D.

  Philadelphia VA Medical Center, 3900 Woodland Avenue, Philadelphia, PA 19104

- Juan P. Casas, M.D., Ph.D.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Kelly Cho, M.P.H., Ph.D.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Saiju Pyarajan, Ph.D.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Jennifer Huffman, Ph.D.

VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Jennifer Moser, Ph.D.

  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

**MVP COVID-19 Science Program Steering Committee Support**

- Lauren Thomann, M.P.H. (P&P Committee Representative, Working Group Coordinator)

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Helene Garcon, M.D. (Program Coordinator, Working Group Coordinator)

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Nicole Kosik, M.P.H. (Working Group Coordinator)

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

**MVP COVID-19 Science Program Working Groups and Associated Chairs**

- COVID-19 Related PheWAS
  - Katherine Liao, M.D.

    VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
  - Scott Damrauer, M.D.

    Philadelphia VA Medical Center, 3900 Woodland Avenue, Philadelphia, PA 19104
- Disease Mechanisms
  - Richard Hauger, M.D.

    VA San Diego Healthcare System, 3350 La Jolla Village Drive, San Diego, CA 92161
  - Shiuh-Wen Luoh, M.D., Ph.D.

    Portland VA Medical Center, 3710 SW U.S. Veterans Hospital Road, Portland, OR 97239
  - Sudha Iyengar, Ph.D.

    VA Northeast Ohio Healthcare System, 10701 East Boulevard, Cleveland, OH 44106

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

- Druggable Genome
    - Juan P. Casas, M.D., Ph.D.

      VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Genomics for Risk Prediction, PRS, and MR
    - Themistocles Assimes, M.D., Ph.D.

      VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304
    - Panagiotis Roussos, M.D., Ph.D.

      James J. Peters VA Medical Center, 130 W Kingsbridge Rd, Bronx, NY 10468
    - Robert Striker, M.D., Ph.D.

      William S. Middleton Memorial Veterans Hospital, 2500 Overlook Terrace, Madison, WI 53705
- GWAS & Downstream Analysis
    - Jennifer Huffman, Ph.D.

      VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
    - Yan Sun, Ph.D.

      Atlanta VA Medical Center, 1670 Clairmont Road, Decatur, GA 30033
- Pharmacogenomics
    - Adriana Hung, M.D., M.P.H.

      VA Tennessee Valley Healthcare System, 1310 24th Avenue, South Nashville, TN 37212
    - Sony Tuteja, Pharm.D., M.S.

      Philadelphia VA Medical Center, 3900 Woodland Avenue, Philadelphia, PA 19104
- VA COVID-19 Shared Data Resource – Scott L. DuVall, Ph.D.; Kristine E. Lynch, Ph.D.; Elise Gatsby, M.P.H.

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

VA Informatics and Computing Infrastructure (VINCI), VA Salt Lake City Health Care System, 500 Foothill Drive, Salt Lake City, UT 84148

- MVP COVID-19 Data Core – Kelly Cho, M.P.H., Ph.D.; Lauren Costa, M.P.H.; Anne Yuk-Lam Ho, M.P.H.; Rebecca Song, M.P.H.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

**VA Million Veteran Program**

**MVP Executive Committee**

- Co-Chair: J. Michael Gaziano, M.D., M.P.H.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Co-Chair: Sumitra Muralidhar, Ph.D.

  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

- Rachel Ramoni, D.M.D., Sc.D., Chief VA Research and Development Officer

  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

- Jean Beckham, Ph.D.

  Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705

- Kyong-Mi Chang, M.D.

  Philadelphia VA Medical Center, 3900 Woodland Avenue, Philadelphia, PA 19104

- Christopher J. O'Donnell, M.D., M.P.H.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Philip S. Tsao, Ph.D.

  VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

- James Breeling, M.D., Ex-Officio

  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

- Grant Huang, Ph.D., Ex-Officio

  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

- Juan P. Casas, M.D., Ph.D., Ex-Officio

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

**MVP Program Office**

- Sumitra Muralidhar, Ph.D.

  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

- Jennifer Moser, Ph.D.

  US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

**MVP Recruitment/Enrollment**

- Recruitment/Enrollment Director/Deputy Director, Boston – Stacey B. Whitbourne, Ph.D.;

  Jessica V. Brewer, M.P.H.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- MVP Coordinating Centers

  - Clinical Epidemiology Research Center (CERC), West Haven – Mihaela Aslan, Ph.D.

    West Haven VA Medical Center, 950 Campbell Avenue, West Haven, CT 06516

  - Cooperative Studies Program Clinical Research Pharmacy Coordinating Center,

    Albuquerque – Todd Connor, Pharm.D.; Dean P. Argyres, B.S., M.S.

    New Mexico VA Health Care System, 1501 San Pedro Drive SE, Albuquerque, NM 87108

  - Genomics Coordinating Center, Palo Alto – Philip S. Tsao, Ph.D.

    VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

  - MVP Boston Coordinating Center, Boston - J. Michael Gaziano, M.D., M.P.H.

    VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

  - MVP Information Center, Canandaigua – Brady Stephens, M.S.

    Canandaigua VA Medical Center, 400 Fort Hill Avenue, Canandaigua, NY 14424

- VA Central Biorepository, Boston – Mary T. Brophy M.D., M.P.H.; Donald E. Humphries, Ph.D.; Luis E. Selva, Ph.D.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- MVP Informatics, Boston – Nhan Do, M.D.; Shahpoor (Alex) Shayan, M.S.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- MVP Data Operations/Analytics, Boston – Kelly Cho, M.P.H., Ph.D.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Director of Regulatory Affairs – Lori Churby, B.S.

  VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

**MVP Science**

- Science Operations – Christopher J. O'Donnell, M.D., M.P.H.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Genomics Core - Christopher J. O'Donnell, M.D., M.P.H.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

  Saiju Pyarajan Ph.D.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

  Philip S. Tsao, Ph.D.

  VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

- Data Core - Kelly Cho, M.P.H, Ph.D.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- VA Informatics and Computing Infrastructure (VINCI) – Scott L. DuVall, Ph.D.

  VA Salt Lake City Health Care System, 500 Foothill Drive, Salt Lake City, UT 84148

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

- Data and Computational Sciences – Saiju Pyarajan, Ph.D.

  VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Statistical Genetics – Elizabeth Hauser, Ph.D.

  Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705

  Yan Sun, Ph.D.

  Atlanta VA Medical Center, 1670 Clairmont Road, Decatur, GA 30033

  Hongyu Zhao, Ph.D.

  West Haven VA Medical Center, 950 Campbell Avenue, West Haven, CT 06516

**Current MVP Local Site Investigators**

- Atlanta VA Medical Center (Peter Wilson, M.D.)

  1670 Clairmont Road, Decatur, GA 30033

- Bay Pines VA Healthcare System (Rachel McArdle, Ph.D.)

  10,000 Bay Pines Blvd Bay Pines, FL 33744

- Birmingham VA Medical Center (Louis Dellitalia, M.D.)

  700 S. 19th Street, Birmingham AL 35233

- Central Western Massachusetts Healthcare System (Kristin Mattocks, Ph.D., M.P.H.)

  421 North Main Street, Leeds, MA 01053

- Cincinnati VA Medical Center (John Harley, M.D., Ph.D.)

  3200 Vine Street, Cincinnati, OH 45220

- Clement J. Zablocki VA Medical Center (Jeffrey Whittle, M.D., M.P.H.)

  5000 West National Avenue, Milwaukee, WI 53295

- VA Northeast Ohio Healthcare System (Frank Jacono, M.D.)

  10701 East Boulevard, Cleveland, OH 44106

- Durham VA Medical Center (Jean Beckham, Ph.D.)

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

508 Fulton Street, Durham, NC 27705

- Edith Nourse Rogers Memorial Veterans Hospital (John Wells., Ph.D.)

  200 Springs Road, Bedford, MA 01730

- Edward Hines, Jr. VA Medical Center (Salvador Gutierrez, M.D.)

  5000 South 5th Avenue, Hines, IL 60141

- Veterans Health Care System of the Ozarks (Gretchen Gibson, D.D.S., M.P.H.)

  1100 North College Avenue, Fayetteville, AR 72703

- Fargo VA Health Care System (Kimberly Hammer, Ph.D.)

  2101 N. Elm, Fargo, ND 58102

- VA Health Care Upstate New York (Laurence Kaminsky, Ph.D.)

  113 Holland Avenue, Albany, NY 12208

- New Mexico VA Health Care System (Gerardo Villareal, M.D.)

  1501 San Pedro Drive, S.E. Albuquerque, NM 87108

- VA Boston Healthcare System (Scott Kinlay, M.B.B.S., Ph.D.)

  150 S. Huntington Avenue, Boston, MA 02130

- VA Western New York Healthcare System (Junzhe Xu, M.D.)

  3495 Bailey Avenue, Buffalo, NY 14215-1199

- Ralph H. Johnson VA Medical Center (Mark Hamner, M.D.)

  109 Bee Street, Mental Health Research, Charleston, SC 29401

- Columbia VA Health Care System (Roy Mathew, M.D.)

  6439 Garners Ferry Road, Columbia, SC 29209

- VA North Texas Health Care System (Sujata Bhushan, M.D.)

  4500 S. Lancaster Road, Dallas, TX 75216

- Hampton VA Medical Center (Pran Iruvanti, D.O., Ph.D.)

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

100 Emancipation Drive, Hampton, VA 23667

- Richmond VA Medical Center (Michael Godschalk, M.D.)

  1201 Broad Rock Blvd., Richmond, VA 23249

- Iowa City VA Health Care System (Zuhair Ballas, M.D.)

  601 Highway 6 West, Iowa City, IA 52246-2208

- Eastern Oklahoma VA Health Care System (Douglas Ivins, M.D.)

  1011 Honor Heights Drive, Muskogee, OK 74401

- James A. Haley Veterans' Hospital (Stephen Mastorides, M.D.)

  13000 Bruce B. Downs Blvd, Tampa, FL 33612

- James H. Quillen VA Medical Center (Jonathan Moorman, M.D., Ph.D.)

  Corner of Lamont & Veterans Way, Mountain Home, TN 37684

- John D. Dingell VA Medical Center (Saib Gappy, M.D.)

  4646 John R Street, Detroit, MI 48201

- Louisville VA Medical Center (Jon Klein, M.D., Ph.D.)

  800 Zorn Avenue, Louisville, KY 40206

- Manchester VA Medical Center (Nora Ratcliffe, M.D.)

  718 Smyth Road, Manchester, NH 03104

- Miami VA Health Care System (Hermes Florez, M.D., Ph.D.)

  1201 NW 16th Street, 11 GRC, Miami FL 33125

- Michael E. DeBakey VA Medical Center (Olaoluwa Okusaga, M.D.)

  2002 Holcombe Blvd, Houston, TX 77030

- Minneapolis VA Health Care System (Maureen Murdoch, M.D., M.P.H.)

  One Veterans Drive, Minneapolis, MN 55417

- N. FL/S. GA Veterans Health System (Peruvemba Sriram, M.D.)

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

1601 SW Archer Road, Gainesville, FL 32608

- Northport VA Medical Center (Shing Shing Yeh, Ph.D., M.D.)

  79 Middleville Road, Northport, NY 11768

- Overton Brooks VA Medical Center (Neeraj Tandon, M.D.)

  510 East Stoner Ave, Shreveport, LA 71101

- Philadelphia VA Medical Center (Darshana Jhala, M.D.)

  3900 Woodland Avenue, Philadelphia, PA 19104

- Phoenix VA Health Care System (Samuel Aguayo, M.D.)

  650 E. Indian School Road, Phoenix, AZ 85012

- Portland VA Medical Center (David Cohen, M.D.)

  3710 SW U.S. Veterans Hospital Road, Portland, OR 97239

- Providence VA Medical Center (Satish Sharma, M.D.)

  830 Chalkstone Avenue, Providence, RI 02908

- Richard Roudebush VA Medical Center (Suthat Liangpunsakul, M.D., M.P.H.)

  1481 West 10th Street, Indianapolis, IN 46202

- Salem VA Medical Center (Kris Ann Oursler, M.D.)

  1970 Roanoke Blvd, Salem, VA 24153

- San Francisco VA Health Care System (Mary Whooley, M.D.)

  4150 Clement Street, San Francisco, CA 94121

- South Texas Veterans Health Care System (Sunil Ahuja, M.D.)

  7400 Merton Minter Boulevard, San Antonio, TX 78229

- Southeast Louisiana Veterans Health Care System (Joseph Constans, Ph.D.)

  2400 Canal Street, New Orleans, LA 70119

- Southern Arizona VA Health Care System (Paul Meyer, M.D., Ph.D.)

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

3601 S 6th Avenue, Tucson, AZ 85723

- Sioux Falls VA Health Care System (Jennifer Greco, M.D.)

  2501 W 22nd Street, Sioux Falls, SD 57105

- St. Louis VA Health Care System (Michael Rauchman, M.D.)

  915 North Grand Blvd, St. Louis, MO 63106

- Syracuse VA Medical Center (Richard Servatius, Ph.D.)

  800 Irving Avenue, Syracuse, NY 13210

- VA Eastern Kansas Health Care System (Melinda Gaddy, Ph.D.)

  4101 S 4th Street Trafficway, Leavenworth, KS 66048

- VA Greater Los Angeles Health Care System (Agnes Wallbom, M.D., M.S.)

  11301 Wilshire Blvd, Los Angeles, CA 90073

- VA Long Beach Healthcare System (Timothy Morgan, M.D.)

  5901 East 7th Street Long Beach, CA 90822

- VA Maine Healthcare System (Todd Stapley, D.O.)

  1 VA Center, Augusta, ME 04330

- VA New York Harbor Healthcare System (Scott Sherman, M.D., M.P.H.)

  423 East 23rd Street, New York, NY 10010

- VA Pacific Islands Health Care System (George Ross, M.D.)

  459 Patterson Rd, Honolulu, HI 96819

- VA Palo Alto Health Care System (Philip Tsao, Ph.D.)

  3801 Miranda Avenue, Palo Alto, CA 94304-1290

- VA Pittsburgh Health Care System (Patrick Strollo, Jr., M.D.)

  University Drive, Pittsburgh, PA 15240

- VA Puget Sound Health Care System (Edward Boyko, M.D.)

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

1660 S. Columbian Way, Seattle, WA 98108-1597

- VA Salt Lake City Health Care System (Laurence Meyer, M.D., Ph.D.)

  500 Foothill Drive, Salt Lake City, UT 84148

- VA San Diego Healthcare System (Samir Gupta, M.D., M.S.C.S.)

  3350 La Jolla Village Drive, San Diego, CA 92161

- VA Sierra Nevada Health Care System (Mostaqul Huq, Pharm.D., Ph.D.)

  975 Kirman Avenue, Reno, NV 89502

- VA Southern Nevada Healthcare System (Joseph Fayad, M.D.)

  6900 North Pecos Road, North Las Vegas, NV 89086

- VA Tennessee Valley Healthcare System (Adriana Hung, M.D., M.P.H.)

  1310 24th Avenue, South Nashville, TN 37212

- Washington DC VA Medical Center (Jack Lichy, M.D., Ph.D.)

  50 Irving St, Washington, D. C. 20422

- W.G. (Bill) Hefner VA Medical Center (Robin Hurley, M.D.)

  1601 Brenner Ave, Salisbury, NC 28144

- White River Junction VA Medical Center (Brooks Robey, M.D.)

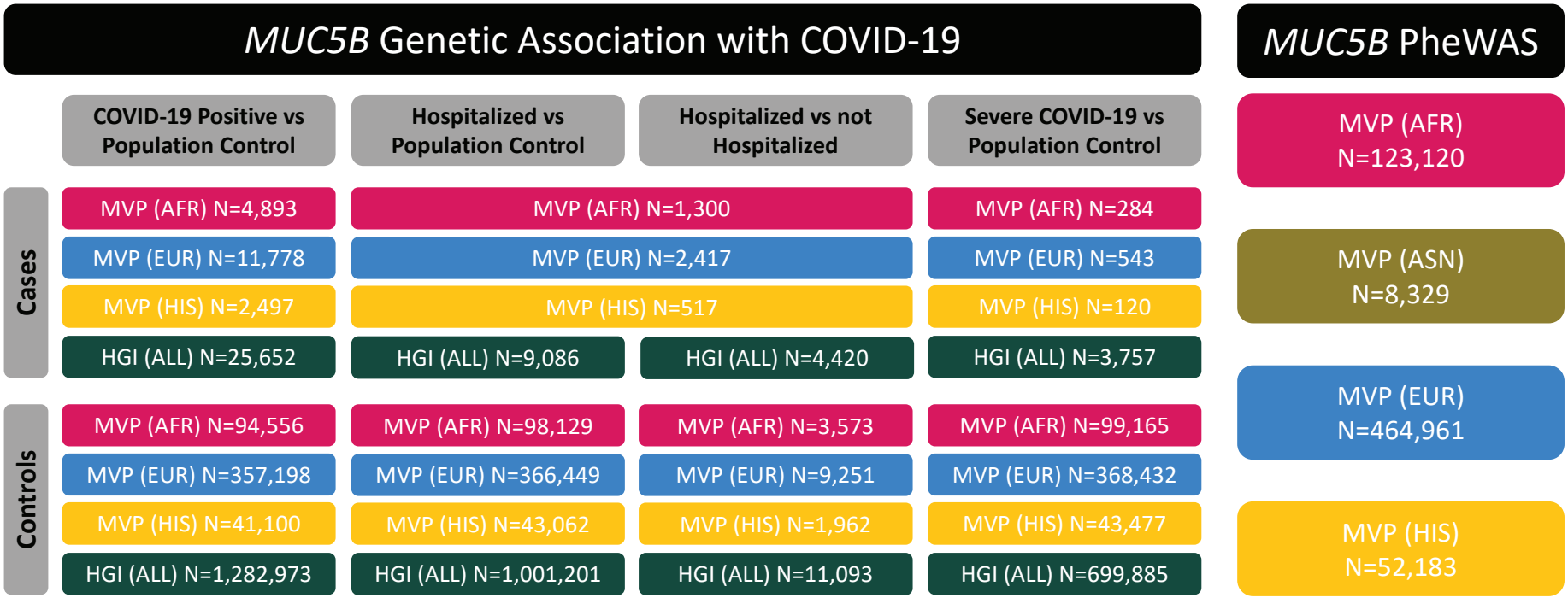
  163 Veterans Drive, White River Junction, VT 05009

- William S. Middleton Memorial Veterans Hospital (Robert Striker, M.D., Ph.D.)

  2500 Overlook Terrace, Madison, WI 53705

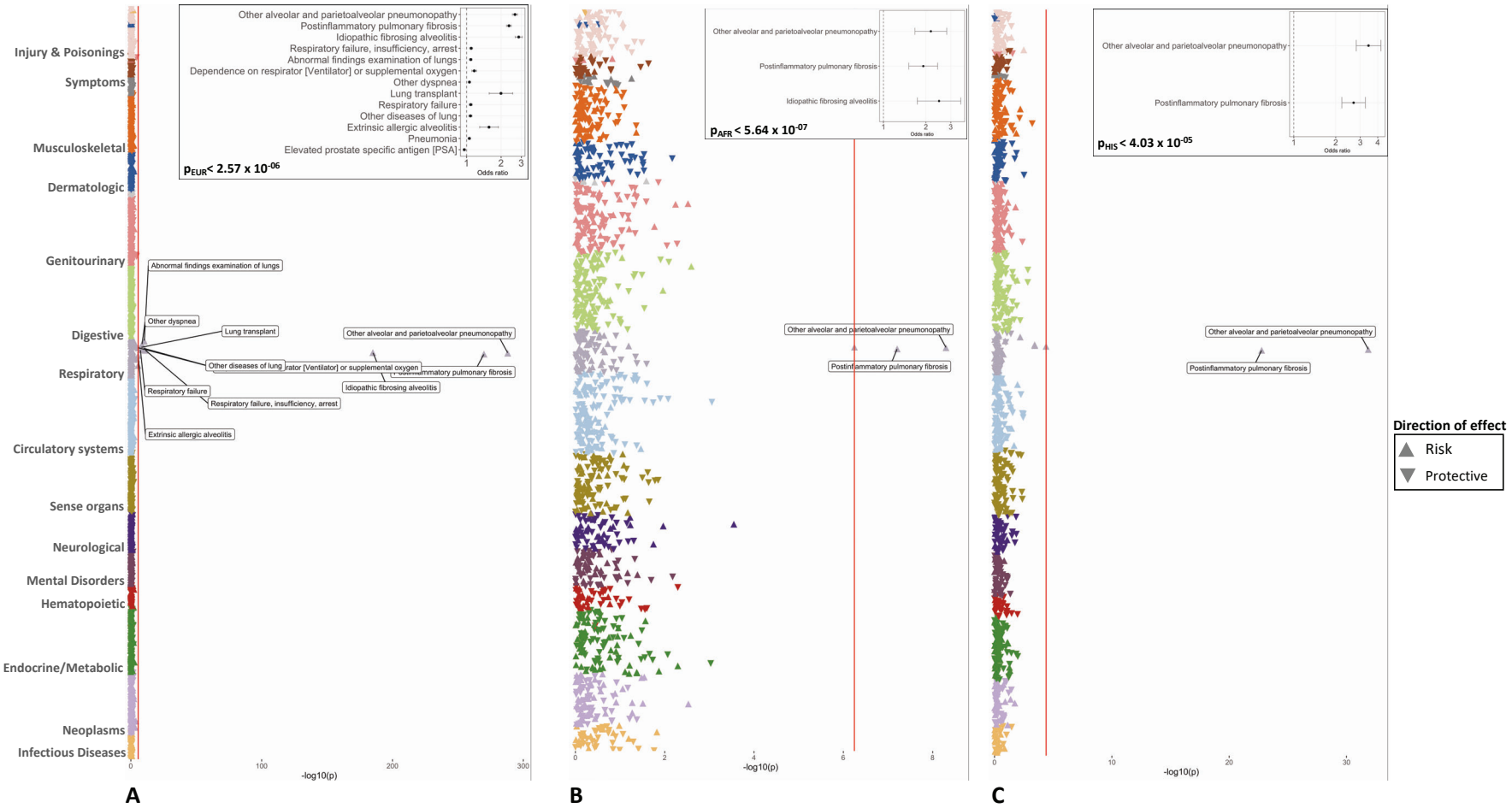
AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

# Figure E1



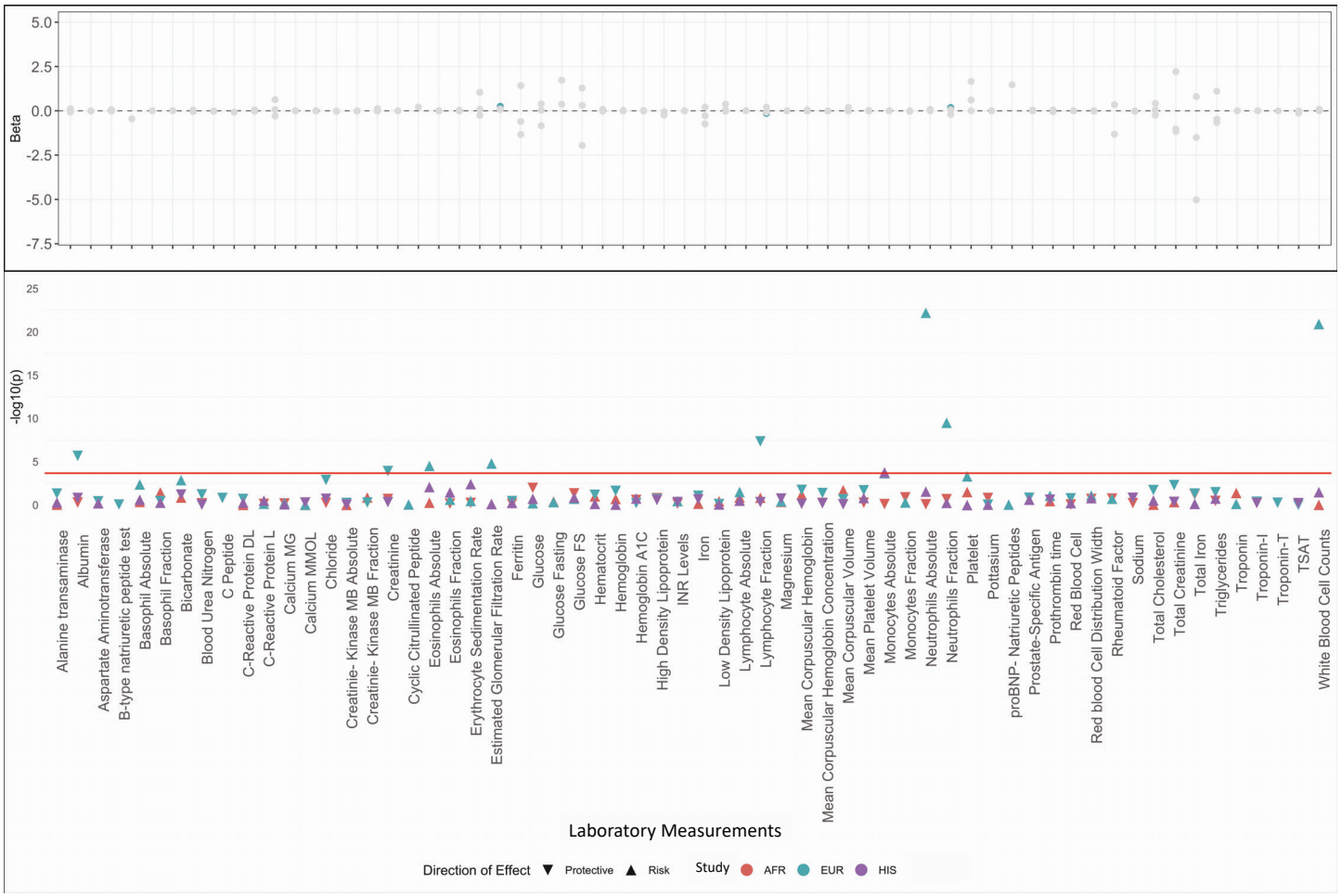

AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

# Figure E2




AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC

# Figure E3




AJRCCM Articles in Press. Published June 30, 2022 as 10.1164/rccm.202109-2166OC