HHS Public Access

Author manuscript

Semin Arthritis Rheum. Author manuscript; available in PMC 2024 December 01.

Published in final edited form as:

Semin Arthritis Rheum. 2023 December ; 63: 152254. doi:10.1016/j.semarthrit.2023.152254.

Gene-respiratory disease interactions for rheumatoid arthritis risk

Vanessa L. Kronzer, MD MSCI[1], **Keigo Hayashi, MD PhD MPH**[2], **Cynthia S. Crowson, PhD**[1,3], **John M. Davis III, MD MS**[1], **Gregory C. McDermott, MD**[2], **Jing Cui, PhD**[4], **Elena Losina, PhD**[5], **Pierre-Antoine Juge, MD PhD**[6], **James R. Cerhan, MD PhD**[3], **Jeffrey A. Sparks, MD MMSc**[2]

[1]Division of Rheumatology, Mayo Clinic, Rochester, Minnesota, USA.

[2]Division of Rheumatology, Inflammation, and Immunity; Brigham and Women's Hospital; Harvard Medical School; Boston, USA

[3]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, Minnesota, USA.

[4]Department of Medicine; Brigham and Women's Hospital; Harvard Medical School; Boston, USA

[5]Department of Orthopedic Surgery; Brigham and Women's Hospital; Boston, USA

[6]Dept of Rheumatology, DMU Locomotion, INSERM UMR1152, Hôpital Bichat-Claude Bernard, APHP, Université de Paris, Paris, France.

Abstract

Objective: We aimed to identify gene by respiratory tract disease interactions that increase RA risk.

Methods: In this case-control study using the Mass General Brigham Biobank, we matched incident RA cases, confirmed by ACR/EULAR criteria, to four controls on age, sex, and electronic health record history. Genetic exposures included a validated overall genetic risk score (GRS) for RA, a Human Leukocyte Antigen (*HLA*) GRS for RA, and the *MUC5B* promoter variant, an established risk factor for RA-associated interstitial lung disease (ILD). Preceding respiratory tract

Corresponding Author: Vanessa L. Kronzer, kronzer.vanessa@mayo.edu, 200 First Street SW, Rochester, MN 55905, P: 651-308-1523 F: 507-266-1799.

Author contributions:

Vanessa Kronzer: conceptualization, funding acquisition, investigation, methodology, project administration, writing – original draft.
Keigo Hayashi: data curation, formal analysis, methodology, software, validation, writing – review & editing.
Cynthia Crowson: methodology, writing – review & editing.
John Davis: methodology, writing – review & editing.
Gregory McDermott: methodology, writing – review & editing.
Jing Cui: data curation, methodology, writing – review & editing.
Elena Losina: methodology, writing – review & editing.
Pierre-Antoine Juge: methodology, writing – review & editing.
James Cerhan: methodology, writing – review & editing.
Jeffrey Sparks: conceptualization, data curation, funding acquisition, investigation, methodology, project administration, resources, supervision, writing – review & editing.

Declarations of interest: None related to this work.

diseases came from diagnosis codes (positive predictive value 86%). We estimated attributable proportions (AP) and multiplicative odds ratios (OR) with 95% confidence intervals (CI) for RA for each genetic and respiratory exposure using conditional logistic regression models, adjusting for potential confounders.

Results: We identified 653 incident RA cases and 2,607 matched controls (mean 54 years, 76% female). The highest tertile of the overall GRS and the *HLA* GRS were both associated with increased RA risk (OR 2.28, 95% CI 1.89,2.74; OR 2.02, 95% CI 1.67–2.45). ILD and the *HLA* GRS exhibited a synergistic relationship for RA risk (OR for both exposures 4.30, 95% CI 1.28,14.38; AP 0.51, 95% CI-0.16,1.18). Asthma and the *MUC5B* promoter variant also exhibited a synergistic interaction for seropositive RA (OR for both exposures 2.58, 95% CI 1.10,6.07; AP 0.62, 95% CI 0.24,1.00).

Conclusion: ILD-*HLA* GRS and asthma-*MUC5B* promoter variant showed synergistic interactions for RA risk. Such interactions may prove useful for RA prevention and screening.

Keywords

Rheumatoid arthritis; genetics; respiratory tract diseases; epidemiology; MUC5B; interaction

INTRODUCTION

Increasing evidence suggests RA may originate in mucosal sites such as the lungs.[1–4] Using the Epidemiological Investigation of Rheumatoid Arthritis (EIRA) case-control study, we showed that acute and chronic upper and lower respiratory disease groups were associated with over two-fold increased risk of RA, particularly in nonsmokers.[5] Recent work from our own group also showed that asthma, pharyngitis, and sinusitis may also be novel risk factors for RA.[3,6,7]

Numerous studies have shown an interaction between other respiratory exposures and genes for RA risk. Uncovering such effect modifiers can facilitate breakthroughs in understanding disease. For example, cigarette smoking and the human leukocyte antigen (*HLA) DRB1* shared epitope interact in a synergistic fashion to increase risk of RA over 40-fold for anti-citrullinated peptide antibody (ACPA) positive RA.[8–10] Recent studies have shown textile dust and occupational inhalants may also interact strongly with the shared epitope to increase ACPA positive RA risk (odds ratios [OR] 39 and 18 respectively).[11,12] Despite these strong gene-respiratory interactions, the interplay between RA risk genes and respiratory tract diseases for RA risk has not yet been studied.

Although the strongest genetic risk factor for RA is the *HLA* shared epitope, multiple additional RA risk genes outside of the *HLA* region have been identified that may also interact with respiratory tract diseases. Over 200 non-*HLA* single nucleotide polymorphisms (SNPs) have been associated with RA risk to date[13,14] and combined into a genetic risk score (GRS) for RA.[15] Since many of these SNPs relate to immune function, they could also influence susceptibility to respiratory tract diseases. In addition, a gain-of-function promoter variant in *MUC5B* was found to increase the risk of RA-associated interstitial lung disease (ILD) by three-fold.[16] A recent study by our group showed that this association was stronger

for late-onset RA,[17] possibly identifying a novel RA phenotype. These findings raise the question of whether RA risk alleles or *MUC5B* modify the relationship between other lung diseases and RA in certain RA subgroups.

To address these two gaps, we leveraged the Mass General Brigham (MGB) Biobank. We aimed to (1) identify how the RA GRS modifies the relationship between respiratory tract diseases and RA and (2) determine the interaction between *MUC5B* and prior respiratory tract diseases for RA risk. We hypothesized that preceding respiratory tract diseases including asthma, sinusitis, and pharyngitis interact synergistically with genetic risk factors for RA risk. Furthermore, we hypothesized that *MUC5B* increases the impact of both ILD and other respiratory disease such as asthma on risk of RA.

METHODS

Study Participants and Design

This case-control study took place within the MGB Biobank. This ongoing study includes blood samples, survey, and clinical data from the electronic health record (EHR) at MGB for over 130,000 participants.[18] It recruited these participants from 2010 to present from clinical visits at Brigham and Women's Hospital or Massachusetts General Hospital and affiliated sites. Participants provided blood samples at the time of biobank enrollment, yielding current genome-wide association study (GWAS) data for approximately 57,000 individuals.

For this MGB biobank sub-study, we only analyzed participants with GWAS data available. We matched each incident RA case to four controls based on age (±5 years) at index date of RA onset, sex, and duration of prior EHR history at index date (±3 years). Index date for this study was the time of RA clinical diagnosis (or matched date for controls), as indicated by the first contact with a clinician related to RA symptoms or diagnosis on manual medical record review. This study received approval from the MGB institutional review board (protocol #2019P000264), obtained written informed consent, followed the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines, and complied with the Declaration of Helsinki.

Incident RA Cases

We identified potential RA cases using a validated Electronic Medical Records and Genomics (eMERGE) Network algorithm combined with natural language processing.[19] This algorithm has 97% specificity and 95% positive predictive value for RA by 2010 American College of Rheumatology/European Alliance of Associations for Rheumatology (ACR/EULAR) criteria.[20] We then verified all RA cases met 2010 ACR/EULAR criteria using medical record review.[21] Inclusion criteria for this study included age 18 and older at index date of RA onset, available GWAS data, and at least five years of preceding EHR data in order to allow time for respiratory exposure accrual. We defined "seropositive" RA as presence of positive rheumatoid factor (RF) and/or ACPA, which both came from testing in clinical care. In our post-hoc analysis, we used a subset of RA cases with RA-ILD as

defined by clinically-obtained high-resolution CT showing interstitial lung abnormalities as reviewed by two radiologists and one pulmonologist.[22]

Controls

We required controls also have GWAS data and at least five years EHR history at MGB. In addition, we required they have available smoking data from Biobank enrollment questionnaire, no RA ever by the eMERGE algorithm,[19] and no billing code for any other systemic rheumatic disease including autoinflammatory syndromes, ankylosing spondylitis, psoriatic arthritis, systemic connective tissue diseases, enteropathic arthropathies, juvenile idiopathic arthritis. We reviewed 100 records meeting our control definition and none had RA (negative predictive value 100%).

Genetic Factors

The MGB Biobank performed genotyping using the Illumina Multi-Ethnic Genotyping Array (MEGA, Illumina Inc., San Diego, CA) and Global Screening Array (GSA) chips. They performed imputation using the Michigan imputation server with Haplotype Reference Consortium (HRC) as the reference panel. We identified *HLA* amino acid information using the SNP2HLA computational strategy.[23]

The three genetic exposures of interest for this study included (1) overall RA weighted GRS (continuous), which added the score derived from recently-published non-*HLA* SNPs[15] to the score from five RA risk amino acid positions in the *HLA* region,[24] (2) HLA weighted GRS (continuous), composed of the score five RA risk amino acids alone,[24] and (3) the *MUC5B* promoter variant rs35705950 (present [TT or GT] vs. absent [GG]).[16] The GRS were weighted by the natural logarithm of the effect size estimate in prior large genetic studies.[15,25] In this study, the overall GRS included only SNPs with information score ≥0.5, or 96 out of the 122 SNPs in the most recently published RA GWAS (see Supplementary Table S1).[15]

Respiratory Tract Disease Exposures

Respiratory tract disease exposures for this study included history of any respiratory tract disease, along with history of specific respiratory tract diseases including asthma, ILD, pharyngitis (both acute and chronic), pneumonia and acute lower respiratory tract diseases, and sinusitis (both acute and chronic). We previously ascertained these respiratory exposures using diagnosis codes,[7] which had mean positive predictive value (PPV) of 86% for true respiratory disease as diagnosed by a physician. To be “exposed,” we required participants to have at least one inpatient or emergency room code or at least two outpatient diagnosis codes (at least 30 days apart for chronic respiratory diseases) before index date of RA onset. Participants could be “exposed” to more than one respiratory tract disease exposure. The reference or “unexposed” group included any participants not meeting the above criteria for that particular exposure.

Covariates

Covariates for this study included age at index date (continuous), sex (male, female), duration of EHR history at MGB prior to index date (continuous), biobank enrollment

year (continuous), genetic ancestry (10 principal components obtained from GWAS data),[26] self-reported education (four-year college or master's/doctoral/professional degree, less than 4-year college degree), body mass index (BMI) at index date (<20, 20-<25, 25-<30, 30+), smoking status as of biobank enrollment year (never, past, current), and smoking pack-years as of biobank enrollment year (continuous). Age, sex, EHR history, and BMI were obtained from the EHR. Enrollment year, education, smoking, and missing BMIs were obtained from MGB Biobank enrollment data. We obtained any remaining missing BMI and smoking data from manual medical record review.

Statistical Analysis

We compared normally distributed continuous variables using t-tests, non-normally distributed continuous variables using the Wilcoxon rank sum tests, and categorical variables using chi-squared tests. We studied the association between each genetic (n=3) and respiratory exposure (n=10) and RA using multivariable conditional logistic regression models to obtain OR with 95% confidence intervals (CI). Models adjusted for age, sex, EHR history, enrollment year, race and ethnicity by principal components, education, BMI, smoking status, and pack-years. Using the same models, we studied interactions on both the additive and multiplicative scales, as both can be biologically plausible depending on the circumstances.[27] For additive interactions, we divided genetic risk into higher and lower risk groups using the cutoff of 67 percentile (highest tertile) in controls. We prespecified this threshold in our study protocol to achieve roughly even "GRS-low" and "GRS-high" groups for cases for power purposes. We then reported the attributable proportion due to interaction (AP) and relative excess risk due to interaction (RERI) when compared to the reference group without that respiratory tract disease and lower genetic risk.[28] For multiplicative interactions, we used the continuous GRS for greater power and reported the interaction OR (per unit increase in GRS score). We then repeated analyses with RA serostatus as the outcome in secondary analyses.

The only missing covariate used in multivariable models was education history. We imputed missing education history using multiple imputation with all case/control and covariate data.[29] Participants missing *MUC5B* data were excluded from those particular analyses. Pre-planned sensitivity analyses stratified results by smoking status (never vs. ever), replicated the known gene-smoking interaction for RA risk as a positive control, and reported characteristics of participants missing education history or *MUC5B* compared to those with complete data. We conducted one post-hoc analysis to explore a significant interaction identified from the main analyses. That is, among RA-ILD cases, we calculated the OR for pre-RA asthma based on presence of *MUC5B* promoter variant. Throughout this study, we considered two-sided $p<0.05$ as statistically significant. All analyses were pre-specified in a study protocol and performed using SAS version 9.4 (SAS Institute Inc., Cary, NC).

RESULTS

Characteristics

We identified 2,283 confirmed RA cases in the MGB Biobank, of whom 1,625 had GWAS data available and were age 18 or older at RA onset. Among these, 653 had at least five

years EHR history prior to the index date of RA diagnosis and were thus included in this study. Out of the 641 RA cases with serostatus data available, 373 (58%) were seropositive. We matched the 653 RA cases to 2,607 controls (mean age 54, 76% female, median 12 years EHR history). RA cases were less likely to have White race or higher education and more likely to have obesity or history of smoking compared to controls (Table 1).

Individual Genetic and Respiratory Exposures

The highest tertile of RA GRS scores (both overall and *HLA*) were associated with over two-fold increased risk RA compared ot the other tertiles, especially seropositive RA (Table 2). In addition, respiratory tract diseases including pharyngitis, pneumonia, and sinusitis were also associated with increased risk of RA (Table 2).

Interactions between GRS and Respiratory Tract Diseases

The only respiratory tract disease that showed evidence of a synergistic relationship with the GRS for RA risk was ILD (Table 3). This synergistic interaction was more prominent for the *HLA* GRS, where the OR for RA for ILD alone was 1.10 (95% CI 0.32,3.77), for the high *HLA* GRS alone was 2.00 (95% CI 1.66,2.43), but in the presence of both exposures was 4.30 (95% CI 1.28,14.38; AP 0.51, 95% CI-0.16,1.18).

In contrast, most respiratory tract diseases and each GRS interacted in a negative or “antagonistic” fashion on the multiplicative scale (Table 3). That is, the risk of RA in the presence of both these genetic and respiratory tract disease exposures was less than expected (i.e. the sum) for the association between each exposure and RA alone. This antagonistic interaction was statistically significant for any respiratory tract disease as well as acute sinusitis (Table 3). For example, the odds of RA in the presence of any respiratory tract disease alone was 1.18 (95% CI 0.91,1.53) and the high overall GRS alone was 2.70 (95% CI 2.08,3.51). However, in the presence of both exposures, the OR for RA was only 2.25 (95% CI 1.72,2.95). Although no interactions were statistically significant on the additive scale, a similar (antagonistic) direction of effect was observed for both any respiratory disease and acute sinusitis, and even more so for chronic rhinitis and pharyngitis and its interaction with the overall GRS (Table 3).

Interactions for RA Serostatus

The magnitude of the antagonistic interaction between the overall GRS with any respiratory tract disease was larger and statistically significant in seropositive RA (multiplicative OR 0.74, 95% CI 0.59,0.94) compared to seronegative RA (OR 0.84, 95% CI 0.64,1.10). Although the antagonistic interaction between the GRS and chronic sinusitis was not statistically significant for all RA, it was significant for seropositive RA, with additive RERI −2.58 (95% CI −5.15,−0.21) and multiplicative OR 0.64 (95% CI 0.40,1.02). In contrast, asthma, pharyngitis (especially chronic), and acute sinusitis demonstrated stronger antagonistic interactions for seronegative RA, though none were statistically significant (Table 4).

Interaction between *MUC5B* and Respiratory Tract Diseases

When subdivided by RA serostatus, asthma and the *MUC5B* promoter variant exhibited a synergistic interaction for seropositive RA (Table 5). That is, the OR for RA in the presence of both asthma and *MUC5B* was 2.58 (95% CI 1.10,6.07; AP 0.62, 95% CI 0.24,1.00; multiplicative OR 2.64, 95% CI 0.98,7.09). This interaction was not present for seronegative RA (see Supplementary Table S2). We found no evidence of an interaction between ILD and *MUC5B* for RA risk, though confidence intervals were wide (Table 5).

To examine the relationship between asthma and *MUC5B* in the context of RA-ILD, we performed a post-hoc analysis among 32 known RA-ILD cases in this cohort. RA-ILD cases with the *MUC5B* promoter variant had higher history of pre-RA asthma (25%) than those without the *MUC5B* promoter variant (9.5%; OR 3.2, 95% CI 0.5,22).

Sensitivity Analyses

The main gene-respiratory interaction results were largely similar by point estimates regardless of smoking status (see Supplementary Table S3). However, among ever smokers, the GRS exhibited a synergistic additive interaction with acute pharyngitis (AP 0.59, 95% CI 0.26,0.92) and an antagonistic multiplicative interaction with acute sinusitis (OR 0.44, 95% CI 0.25,0.79) for RA risk. As a positive control, we replicated the known gene-smoking interaction for RA risk using the overall GRS and smoking status (never vs. ever). Indeed, we observed a positive association with AP 0.26 (95% CI 0.06,0.47) and RERI 0.86 (95% CI 0.10,1.62). Finally, only 333 (10.5%) participants were missing education history and 249 (7.9%) missing *MUC5B* status. Participants with complete data were more likely to have White, non-Hispanic race and ethnicity and less smoking history compared to those with missing data (Supplementary Table S4).

DISCUSSION

In this study, we found a synergistic relationship between *HLA* risk alleles and ILD for RA risk. We also found that the *MUC5B* promoter variant may interact synergistically with history of asthma for an elevated risk of seropositive RA. Finally, we found that the RA GRS generally exhibits a mild negative or “antagonistic” interaction with most respiratory tract diseases for RA risk. These findings identify potential novel pathways for RA pathogenesis with potential clinical implications for RA prevention and screening.

The first key finding from this study was that we observed a synergistic relationship between the *HLA* GRS and ILD for RA risk. Although RA is known to increase risk of ILD,[30,31] this study and one prior study suggest that ILD may also increase risk of RA.[5] These findings raise new questions about the direction of causality between ILD and RA. In addition, although a previous study showed that smoking and *HLA* alleles interacted for risk of RA-ILD,[32] the interaction between *HLA* and ILD for RA risk has not previously been reported. This interaction was not statistically significant, but it had high point estimates. In the presence of both *HLA* and ILD, the odds of RA were increased over four-fold with AP over 0.5. This AP is similar in magnitude to the interaction between *HLA* shared epitope alleles and smoking for RA risk.[33,34] It is possible that participants with ILD already had

RA (i.e., reverse causation). However, if participants already had RA, an interaction should have been present for both RA genetic risk scores rather than just the *HLA* GRS. The high risk we observed suggests that screening for *HLA* risk alleles in newly diagnosed ILD may be helpful in flagging patients for rheumatology evaluation. Due to small sample sizes, however, the association between ILD and RA and its interaction with *HLA* for RA risk should be replicated in larger cohorts.

The second main finding from this study was that the *MUC5B* promoter variant and asthma interacted synergistically for seropositive RA risk. Individuals with asthma have previously been shown to have altered *MUC5B* gene expression[35] and increased risk for RA.[5,36] Based on these relationships, asthma could simply mediate the relationship between the *MUC5B* promoter variant and RA risk. However, the *MUC5B* promoter variant is not associated with asthma in GWAS.[37] Furthermore, we observed little association for either *MUC5B* or asthma separately for RA risk, but over a two-fold association in the presence of both. Another possibility to consider is that the asthma exposure in this study could simply represent misclassified ILD. However, we did not observe a synergistic relationship between ILD and *MUC5B* for RA risk. Instead, *MUC5B* and asthma could indeed interact to induce a certain phenotype of RA. Supporting this conclusion, we previously showed *MUC5B* was associated with RA-ILD particularly in the late-onset RA phenotype.[17] Furthermore, our data suggested that a *MUC5B*-asthma pathway may not only increase risk of RA but possibly also RA-ILD. If so, preventing asthma flares might help reduce RA and/or RA-ILD onset in individuals with the *MUC5B* promoter variant. Therefore, this potential mechanistic pathway for RA onset merits replication. Future work should also study each RA risk locus separately with each respiratory disease to determine whether other gene-respiratory interaction pathways explain other subsets of RA.

Other than the interactions of ILD with *HLA* and asthma with *MUC5B*, we observed that RA risk genes interact with respiratory tract diseases in a generally negative or "antagonistic" fashion for RA risk. That is, the odds of RA in the presence of both exposures were lower than expected given the positive associations for each exposure alone. This antagonistic interaction was most prominent for any respiratory tract disease and sinusitis. The observation that this finding was stronger in seropositive RA for any respiratory disease and chronic sinusitis aligns with prior literature showing an interaction between RA risk genes and smoking or textile dust but only in ACPA-positive RA.[8,11] It is possible that this interaction was spurious especially since the additive interactions, which are thought to be more biologically relevant,[38] were not statistically significant. However, the measures of additive interaction trended in the same direction. Biologically speaking, RA risk alleles, which presumably bolster the immune system, might better help fight infections such as sinusitis and thereby reduce the inflammation that triggers RA. Alternatively, this negative or antagonistic interaction may have resulted from the "risk factor paradox," where negative associations between risk factors arise from conditioning on the presence of RA in the cases but not the controls.[39]

Another explanation for the antagonistic interactions could be the fact that longer-term respiratory tract exposures are believed to be more associated with RA risk than recent ones.

For example, our past study found respiratory tract diseases 5–10 years or >10 years prior to RA were most associated with RA risk.[6]

A different study found air pollutants to be associated with RA only 10 or more years before RA onset,[40] which might explain the lack of statistically significant interaction between air pollution and genes for RA risk.[41] Similarly, distant respiratory tract disease exposures could have been missed by our 5 year EHR requirement. If they interact with RA risk alleles synergistically for RA risk, but more recent respiratory tract diseases do not, we expect the GRS exposed/respiratory "unexposed" group to have the higher odds of RA than the GRS exposed/respiratory "exposed" group. This is exactly what we observed. Thus, longer-term studies will be important to resolve this question.

Strengths of this study included its manual verification of RA status and date of diagnosis, data on RA serostatus, adjustment for important confounders such as smoking status and pack-years, and the relatively large sample size of incident RA cases and controls with GWAS data. There are also several important limitations to consider. First, the single geographical location of this study and the predominately White, non-Hispanic ancestry of participants limits generalizability to other locations and populations, especially for genetic data.[42] Second, the small sample size of RA cases with certain respiratory tract diseases such as ILD led to wide confidence intervals. Larger studies are needed to replicate these results and improve precision of these estimates. Third, selection bias is possible for MGB Biobank participants in general and for this study's controls given the requirement for smoking data. This requirement likely explains why controls were significantly higher educated than RA cases. Fourth, misclassification of exposures is possible. Although our previous study showed that diagnosis codes had high PPV for true respiratory tract diseases,[7] participants classified as unexposed might have had exposures before entry into the MGB HealthCare system, in outside healthcare systems, or simply from not coming to clinical attention at all. Any of these possibilities would bias results towards the null. In addition, the *HLA* GRS we used did not incorporate protective alleles such as DERAA-encoding *HLA* alleles.[43] Fifth, smoking measured at time of MGB Biobank enrollment may not reflect smoking history prior to RA. This discrepancy may explain why we observed no difference in the association between respiratory diseases and RA by smoking status, as we had seen previously.[5] Sixth, we performed many statistical tests, though all were pre-specified in our study protocol. Finally, unmeasured confounding is possible such as occupational inhalants or environmental pollutants, which are associated with both respiratory tract diseases and RA.[12,40]

Conclusion

In summary, ILD in the presence of *HLA* risk alleles and asthma in the presence of the *MUC5B* promoter variant both synergistically increased RA risk. In contrast, RA risk genes and most respiratory tract diseases interact in an antagonistic fashion for RA risk. Future studies should replicate these findings and seek to uncover additional interaction pathways, as they may have important implications for RA prevention and screening.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We would like to thank the participants and staff of the Mass General Brigham Biobank that enabled this research study.

Funding:

This study was supported by the Rheumatology Research Foundation Scientist Development Award (VLK), the National Institute of Arthritis and Musculoskeletal and Skin Diseases awards VERITY Pilot & Feasibility Award from P30-AR072577 (VLK), along with R01 AR46849 (CSC), R01 AR080659 (JAS), R01 AR077607 (JAS), P30 AR070253 (JAS), and T32 AR007530 (GCM). It was also supported by the R. Bruce and Joan M. Mickey Research Scholar Fund, and the Llura Gund Award for Rheumatoid Arthritis Research and Care (JAS). The funders had no role in the study design; collection, analysis, and interpretation of data; writing of the report; or the decision to submit this manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University, its affiliated academic health care centers, or the National Institutes of Health.

REFERENCES

1. Khuder SA, Peshimam AZ, Agraharam S. Environmental risk factors for rheumatoid arthritis. Rev Environ Health. Oct-Dec 2002;17(4):307–15. doi:10.1515/reveh.2002.17.4.307 [PubMed: 12611472]
2. Di Giuseppe D, Discacciati A, Orsini N, Wolk A. Cigarette smoking and risk of rheumatoid arthritis: a dose-response meta-analysis. Arthritis Res Ther. Mar 5 2014;16(2):R61. doi:10.1186/ar4498 [PubMed: 24594022]
3. Kronzer VL, Crowson CS, Sparks JA, Vassallo R, Davis JM 3rd. Investigating Asthma, Allergic Disease, Passive Smoke Exposure, and Risk of Rheumatoid Arthritis. Arthritis Rheumatol. Aug 2019;71(8):1217–1224. doi:10.1002/art.40858 [PubMed: 30747496]
4. Ford JA, Liu X, Chu SH, et al. Asthma, Chronic Obstructive Pulmonary Disease, and Subsequent Risk for Incident Rheumatoid Arthritis Among Women: A Prospective Cohort Study. Arthritis Rheumatol. May 2020;72(5):704–713. doi:10.1002/art.41194 [PubMed: 32129572]
5. Kronzer VL, Westerlind H, Alfredsson L, et al. Respiratory Diseases as Risk Factors for Seropositive and Seronegative Rheumatoid Arthritis and in Relation to Smoking. Arthritis Rheumatol. Jan 2021;73(1):61–68. doi:10.1002/art.41491 [PubMed: 32799411]
6. Kronzer VL, Huang W, Crowson CS, et al. Timing of sinusitis and other respiratory tract diseases and risk of rheumatoid arthritis. Semin Arthritis Rheum. Feb 2022;52:151937. doi:10.1016/j.semarthrit.2021.11.008 [PubMed: 35042150]
7. Kronzer VL, Huang W, Zaccardelli A, et al. Association of Sinusitis and Upper Respiratory Tract Diseases With Incident Rheumatoid Arthritis: A Case-control Study. J Rheumatol. Oct 15 2021;doi:10.3899/jrheum.210580
8. Klareskog L, Stolt P, Lundberg K, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum. Jan 2006;54(1):38–46. doi:10.1002/art.21575 [PubMed: 16385494]
9. Kim K, Jiang X, Cui J, et al. Interactions between amino acid-defined major histocompatibility complex class II variants and smoking in seropositive rheumatoid arthritis. Arthritis Rheumatol. Oct 2015;67(10):2611–23. doi:10.1002/art.39228 [PubMed: 26098791]
10. Too CL, Yahya A, Murad S, et al. Smoking interacts with HLA-DRB1 shared epitope in the development of anti-citrullinated protein antibody-positive rheumatoid arthritis: results from the Malaysian Epidemiological Investigation of Rheumatoid Arthritis (MyEIRA). Arthritis Res Ther. Apr 26 2012;14(2):R89. doi:10.1186/ar3813 [PubMed: 22537824]

11. Too CL, Muhamad NA, Ilar A, et al. Occupational exposure to textile dust increases the risk of rheumatoid arthritis: results from a Malaysian population-based case-control study. Ann Rheum Dis. Jun 2016;75(6):997–1002. doi:10.1136/annrheumdis-2015-208278 [PubMed: 26681695]
12. Tang B, Liu Q, Ilar A, et al. Occupational inhalable agents constitute major risk factors for rheumatoid arthritis, particularly in the context of genetic predisposition and smoking. Ann Rheum Dis. Mar 2023;82(3):316–323. doi:10.1136/ard-2022-223134 [PubMed: 36600175]
13. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. Feb 20 2014;506(7488):376–81. doi:10.1038/nature12873 [PubMed: 24390342]
14. Rostami S, Hoff M, Brown MA, Hveem K, Videm V. Comparison of methods to construct a genetic risk score for prediction of rheumatoid arthritis in the population-based Nord-Trondelag Health Study, Norway. Rheumatology (Oxford). Jan 13 2020;doi:10.1093/rheumatology/kez638
15. Ishigaki K, Sakaue S, Terao C, et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. Nat Genet. Nov 2022;54(11):1640–1651. doi:10.1038/s41588-022-01213-w [PubMed: 36333501]
16. Juge PA, Lee JS, Ebstein E, et al. MUC5B Promoter Variant and Rheumatoid Arthritis with Interstitial Lung Disease. N Engl J Med. Dec 6 2018;379(23):2209–2219. doi:10.1056/NEJMoa1801562 [PubMed: 30345907]
17. McDermott G, Gill R, Gagne S, et al. Associations of the MUC5B Promoter Variant with Timing of Interstitial Lung Disease and Rheumatoid Arthritis Onset. Rheumatology (Oxford). Mar 15 2022;doi:10.1093/rheumatology/keac152
18. Karlson EW, Boutin NT, Hoffnagle AG, Allen NL. Building the Partners HealthCare Biobank at Partners Personalized Medicine: Informed Consent, Return of Research Results, Recruitment Lessons and Operational Considerations. J Pers Med. Jan 14 2016;6(1)doi:10.3390/jpm6010002
19. Carroll RJ, Thompson WK, Eyler AE, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. J Am Med Inform Assoc. Jun 2012;19(e1):e162–9. doi:10.1136/amiajnl-2011-000583 [PubMed: 22374935]
20. Liao KP, Cai T, Gainer V, et al. Electronic medical records for discovery research in rheumatoid arthritis. Arthritis Care Res (Hoboken). Aug 2010;62(8):1120–7. doi:10.1002/acr.20184 [PubMed: 20235204]
21. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum. Sep 2010;62(9):2569–81. doi:10.1002/art.27584 [PubMed: 20872595]
22. Sparks JA, He X, Huang J, et al. Rheumatoid arthritis disease activity predicting incident clinicallyapparent RA-associated interstitial lung disease: A prospective cohort study. Arthritis Rheumatol. Apr 5 2019;doi:10.1002/art.40904
23. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One. 2013;8(6):e64683. doi:10.1371/journal.pone.0064683 [PubMed: 23762245]
24. Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. Jan 29 2012;44(3):2916. doi:10.1038/ng.1076
25. Sparks JA, Chen CY, Jiang X, et al. Improved performance of epidemiologic and genetic risk models for rheumatoid arthritis serologic phenotypes using family history. Ann Rheum Dis. Aug 2015;74(8):1522–9. doi:10.1136/annrheumdis-2013-205009 [PubMed: 24685909]
26. Byun J, Han Y, Gorlov IP, Busam JA, Seldin MF, Amos CI. Ancestry inference using principal component analysis and spatial analysis: a distance-based analysis to account for population substructure. BMC Genomics. Oct 16 2017;18(1):789. doi:10.1186/s12864-017-4166-8 [PubMed: 29037167]
27. Weinberg CR. Interaction and exposure modification: are we asking the right questions? Am J Epidemiol. Apr 1 2012;175(7):602–5. doi:10.1093/aje/kwr495 [PubMed: 22306562]
28. Andersson T, Alfredsson L, Kallberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. Eur J Epidemiol. 2005;20(7):575–9. doi:10.1007/s10654-005-7835-x [PubMed: 16119429]

29. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. Jun 2007;16(3):219–42. doi:10.1177/0962280206074463 [PubMed: 17621469]
30. Samhouri BF, Vassallo R, Achenbach SJ, et al. Incidence, Risk Factors, and Mortality of Clinical and Subclinical Rheumatoid Arthritis-Associated Interstitial Lung Disease: A Population-Based Cohort. Arthritis Care Res (Hoboken). Jan 7 2022;doi:10.1002/acr.24856
31. Hyldgaard C, Hilberg O, Pedersen AB, et al. A population-based cohort study of rheumatoid arthritis-associated interstitial lung disease: comorbidity and mortality. Ann Rheum Dis. Oct 2017;76(10):1700–1706. doi:10.1136/annrheumdis-2017-211138 [PubMed: 28611082]
32. Restrepo JF, del Rincon I, Battafarano DF, Haas RW, Doria M, Escalante A. Clinical and laboratory factors associated with interstitial lung disease in rheumatoid arthritis. Clin Rheumatol. Sep 2015;34(9):1529–36. doi:10.1007/s10067-015-3025-8 [PubMed: 26255186]
33. Lundstrom E, Kallberg H, Alfredsson L, Klareskog L, Padyukov L. Gene-environment interaction between the DRB1 shared epitope and smoking in the risk of anti-citrullinated protein antibody-positive rheumatoid arthritis: all alleles are important. Arthritis Rheum. Jun 2009;60(6):1597–603. doi:10.1002/art.24572 [PubMed: 19479873]
34. Kokkonen H, Brink M, Hansson M, et al. Associations of antibodies against citrullinated peptides with human leukocyte antigen-shared epitope and smoking prior to the development of rheumatoid arthritis. Arthritis Res Ther. May 20 2015;17:125. doi:10.1186/s13075-015-0638-x [PubMed: 25990747]
35. Woodruff PG, Modrek B, Choy DF, et al. T-helper type 2-driven inflammation defines major subphenotypes of asthma. Am J Respir Crit Care Med. Sep 1 2009;180(5):388–95. doi:10.1164/rccm.200903-0392OC [PubMed: 19483109]
36. Charoenngam N, Ponvilawan B, Rittiphairoj T, et al. Patients with asthma have a higher risk of rheumatoid arthritis: A systematic review and meta-analysis. Semin Arthritis Rheum. Oct 2020;50(5):968–976. doi:10.1016/j.semarthrit.2020.07.015 [PubMed: 32906033]
37. Valette K, Li Z, Bon-Baret V, et al. Prioritization of candidate causal genes for asthma in susceptibility loci derived from UK Biobank. Commun Biol. Jun 8 2021;4(1):700. doi:10.1038/s42003-02102227-6 [PubMed: 34103634]
38. Klareskog L, Ronnelid J, Saevarsdottir S, Padyukov L, Alfredsson L. The importance of differences; On environment and its interactions with genes, and immunity in the causation of rheumatoid arthritis. J Intern Med. Mar 16 2020;doi:10.1111/joim.13058
39. Choi HK, Nguyen US, Niu J, Danaei G, Zhang Y. Selection bias in rheumatic disease research. Nat Rev Rheumatol. Jul 2014;10(7):403–12. doi:10.1038/nrrheum.2014.36 [PubMed: 24686510]
40. Hart JE, Källberg H, Laden F, et al. Ambient air pollution exposures and risk of rheumatoid arthritis: results from the Swedish EIRA case-control study. Ann Rheum Dis. Jun 2013;72(6):888–94. doi:10.1136/annrheumdis-2012-201587 [PubMed: 22833374]
41. Zhang J, Fang XY, Wu J, et al. Association of Combined Exposure to Ambient Air Pollutants, Genetic Risk, and Incident Rheumatoid Arthritis: A Prospective Cohort Study in the UK Biobank. Environ Health Perspect. Mar 2023;131(3):37008. doi:10.1289/ehp10710 [PubMed: 36913237]
42. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun. Jul 25 2019;10(1):3328. doi:10.1038/s41467-019-11112-043. [PubMed: 31346163]
43. van der Helm-van Mil AH, Huizinga TW, Schreuder GM, Breedveld FC, de Vries RR, Toes RE. An independent role of protective HLA class II alleles in rheumatoid arthritis severity and susceptibility. Arthritis Rheum. Sep 2005;52(9):2637–44. doi:10.1002/art.21272 [PubMed: 16142711]

Table 1.

Characteristics of the 653 incident Mass General Brigham Biobank RA cases and 2,607 matched controls with GWAS data and at least 5 years of EHR history

Characteristic	RA cases (n=653)	Controls (n=2607)	p-value
Age in years, mean (SD)	54 (13)	54 (12)	*
Female sex, n (%)	494 (76)	1971 (76)	*
EHR history in years, median (IQR)	12 (9,17)	12 (8,16)	*
Enrollment year, median (IQR)	2015 (2014–2017)	2016 (2014–2017)	<0.001
Race and ethnicity, n (%)			<0.0001
Asian	18 (3)	50 (2)	
Black	49 (8)	91 (4)	
Hispanic	9 (1)	20 (1)	
Other	20 (3)	43 (2)	
White, non-Hispanic	529 (85)	2344 (92)	
Education college or higher, n (%) **	377 (58)	1882 (72)	<0.001
BMI, kg/m^2, mean (SD)	29 (7)	28 (6)	<0.001
Smoking status, n (%) **			<0.001
Never	320 (49)	1526 (59)	
Past	274 (42)	949 (36)	
Current	59 (9)	132 (5)	
Smoking pack-years, mean (SD) **	10 (20)	6.6 (13)	<0.001

BMI = body mass index, CI = confidence interval, EHR = electronic health record, GWAS = genome-wide association study, IQR = interquartile range, kg = kilograms, m = meters, RA = rheumatoid arthritis, SD = standard deviation

*Matching factor

**As of Mass General Brigham Biobank enrollment

Table 2.

Association between individual genetic factors and respiratory tract diseases for incident RA risk, overall and by serostatus

	Number (%) or Mean (SD)		Adjusted OR* (95% CI) for RA		
Exposure	**RA cases (n=653)**	**Controls (n=2607)**	**All RA (n=653)**	**Seropositive RA (n=373)**	**Seronegative RA (n=268)**
Genetic factors					
Overall genetic risk score					
Highest tertile, n (%)	344 (53)	868 (33)	2.28 (1.89,2.74)	3.57 (2.72,4.68)	1.33 (1.00,1.78)
Continuous score	10.03 (1.3)	9.47 (1.0)	1.60 (1.47,1.75)	1.99 (1.76,2.26)	1.18 (1.03,1.34)
HLA genetic risk score					
Highest tertile, n (%)	318 (49)	866 (33)	2.02 (1.67,2.45)	2.76 (2.12,3.60)	1.23 (0.91,1.66)
Continuous score	0.77 (1.0)	0.40 (0.8)	1.64 (1.48,1.82)	2.00 (1.73,2.31)	1.18 (1.00,1.40)
MUC5B promoter variant present, n (%)	99/607 (16)	416/2404 (17)	1.06 (0.82,1.36)	1.05 (0.73,1.50)	1.05 (0.72,1.53)
*Respiratory tract diseases before index date**					
Any respiratory tract disease, n (%)	331 (51)	1198 (46)	1.02 (0.85,1.24)	1.06 (0.82,1.38)	0.97 (0.72,1.30)
Asthma	92 (14)	302 (12)	1.00 (0.76,1.31)	1.22 (0.84,1.77)	0.86 (0.57,1.31)
Interstitial lung disease	11 (1.7)	20 (0.8)	1.57 (0.71,3.49)	2.48 (0.88,7.00)	0.69 (0.14,3.40)
Pharyngitis	80 (12)	220 (8.4)	1.34 (1.00,1.79)	1.19 (0.78,1.79)	1.66 (1.08,2.56)
Acute pharyngitis or nasopharyngitis	63 (9.6)	160 (6.1)	1.45 (1.05,2.02)	1.17 (0.73,1.89)	2.08 (1.28,3.39)
Chronic rhinitis and pharyngitis	23 (3.5)	76 (2.9)	1.06 (0.64,1.74)	1.41 (0.72,2.74)	0.74 (0.33,1.67)
Pneumonia and acute lower diseases	76 (12)	196 (7.5)	1.35 (1.00,1.81)	1.35 (0.90,2.01)	1.29 (0.81,2.06)
Sinusitis	70 (11)	198 (7.6)	1.38 (1.02,1.88)	1.34 (0.86,2.06)	1.32 (0.83,2.12)
Acute sinusitis	38 (5.8)	96 (3.7)	1.56 (1.03,2.37)	1.37 (0.77,2.43)	1.76 (0.93,3.35)
Chronic sinusitis	44 (6.7)	133 (5.1)	1.26 (0.86,1.83)	1.24 (0.72,2.12)	1.12 (0.64,1.97)

CI = confidence interval, *HLA* = human leukocyte antigen, OR = odds ratio, RA = rheumatoid arthritis, SD = standard deviation

*Reference group was individuals with no respiratory tract disease codes prior to index date of RA diagnosis (or assigned date for matched controls). Adjusting for age, sex, EHR history, enrollment year, race and ethnicity by principal components, education, body mass index, smoking status, and pack-years.

Table 3.

Interactions between the RA GRS and respiratory tract diseases for incident RA risk in the Mass General Brigham Biobank *

	Adjusted* OR (95% CI)			Gene-Respiratory Interactions		
Respiratory tract disease	**GRS-low/ Resp+**	**GRS-high/ Resp−**	**GRS-high/ Resp+**	**AP (95% CI)**	**RERI (95% CI)**	**Multiplicative OR**
Overall GRS						
Any respiratory tract disease	1.18 (0.91,1.53)	2.70 (2.08,3.51)	2.25 (1.72,2.95)	−0.28 (−0.64,0.08)	−0.63 (−1.39,0.13)	**0.80 (0.68,0.94)**
Asthma	1.12 (0.78,1.61)	2.35 (1.93,2.88)	2.05 (1.38,3.06)	−0.20 (−0.71,0.30)	−0.42 (−1.34,0.50)	0.88 (0.68,1.10)
Interstitial lung disease	1.65 (0.52,5.26)	2.28 (1.89,2.75)	3.03 (0.97,9.45)	0.03 (−1.00,1.00)	0.10 (−3.84,4.04)	1.37 (0.58,3.22)
Pharyngitis	1.52 (1.03,2.23)	2.34 (1.92,2.86)	2.83 (1.82,4.38)	−0.01 (−0.48,0.46)	−0.03 (−1.34,1.28)	0.86 (0.67,1.11)
Acute pharyngitis	1.48 (0.96,2.30)	2.28 (1.88,2.77)	3.47 (2.12,5.69)	0.20 (−0.22,0.62)	0.71 (−1.04,2.46)	0.87 (0.66,1.16)
Chronic rhinitis/ pharyngitis	1.48 (0.80,2.75)	2.34 (1.94,2.83)	1.52 (0.67,3.48)	−0.85 (−2.47,0.76)	−1.30 (−2.86,0.26)	0.76 (0.49,1.19)
Pneumonia and acute lower diseases	1.29 (0.84,1.96)	2.26 (1.86,2.75)	3.08 (1.99,4.77)	0.17 (−0.22,0.57)	0.53 (−0.88,1.94)	0.91 (0.70,1.19)
Sinusitis	1.61 (1.05,2.49)	2.36 (1.94,2.88)	2.45 (1.58,3.80)	−0.22 (−0.80,0.37)	−0.53 (−1.79,0.74)	**0.73 (0.55,0.96)**
Acute sinusitis	1.94 (1.12,3.37)	2.34 (1.93,2.83)	2.71 (1.46,5.03)	−0.21 (−1.03,0.61)	−0.56 (−2.50,1.37)	**0.68 (0.47,0.99)**
Chronic sinusitis	1.47 (0.86,2.52)	2.34 (1.93,2.84)	2.07 (1.22,3.50)	−0.36 (−1.00,0.44)	−0.74 (−2.10,0.61)	0.73 (0.52,1.02)
HLA GRS						
Any respiratory tract disease	1.14 (0.88,1.47)	2.28 (1.76,2.97)	2.03 (1.54,2.67)	−0.19 (−0.54,0.15)	−0.39 (−1.06,0.28)	**0.78 (0.64,0.95)**
Asthma	0.89 (0.62,1.29)	1.95 (1.60,2.39)	2.26 (1.51,3.39)	0.18 (−0.17,0.54)	0.42 (−0.53,1.37)	0.99 (0.74,1.32)
Interstitial lung disease	1.10 (0.32,3.77)	2.00 (1.66,2.43)	4.30 (1.28,14.38)	0.51 (−0.16,1.00)	2.19 (−3.16,7.54)	1.24 (0.51,2.98)
Pharyngitis	1.36 (0.93,2.00)	2.03 (1.66,2.48)	2.75 (1.75,4.32)	0.13 (−0.29,0.55)	0.36 (−0.93,1.66)	0.88 (0.65,1.19)
Acute pharyngitis	1.47 (0.96,2.25)	2.03 (1.66,2.47)	3.13 (1.85,5.28)	0.20 (−0.25,0.65)	0.63 (−1.06,2.32)	0.86 (0.61,1.22)
Chronic rhinitis/ pharyngitis	1.11 (0.58,2.11)	2.03 (1.67,2.46)	2.03 (0.93,4.44)	−0.05 (−0.94,0.83)	−0.11 (−1.84,1.62)	0.93 (0.54,1.62)
Pneumonia and acute lower diseases	1.54 (1.04,2.27)	2.09 (1.71,2.55)	2.38 (1.49,3.81)	−0.10 (−0.67,0.46)	−0.24 (−1.50,1.01)	0.89 (0.64,1.23)
Sinusitis	1.47 (0.97,2.22)	2.05 (1.68,2.51)	2.44 (1.53,3.90)	−0.03 (−0.57,0.50)	−0.08 (−1.36,1.19)	0.78 (0.56,1.09)
Acute sinusitis	1.89 (1.12,3.20)	2.07 (1.71,2.52)	2.47 (1.26,4.83)	−0.20 (−1.00,0.68)	−0.50 (−2.39,1.39)	**0.63 (0.40,0.99)**
Chronic sinusitis	1.35 (0.80,2.26)	2.05 (1.69,2.50)	2.05 (1.17,3.56)	−0.17 (−0.91,0.56)	−0.36 (−1.70,0.98)	0.78 (0.52,1.19)

AP = attributable proportion, CI = confidence interval, COPD = chronic obstructive pulmonary disease, GRS = genetic risk score, *HLA* = human leukocyte antigen, OR = odds ratio, RA = rheumatoid arthritis, RERI = relative excess risk due to interaction, Resp = respiratory tract exposure

*Reference group was individuals without respiratory tract disease exposure and lowest two GRS tertiles. Adjusting for age, sex, EHR history, enrollment year, race and ethnicity by principal components, education, body mass index, smoking status, and pack-years. Bold values are statistically significant ($p<0.05$).

Table 4.

Interaction between the overall GRS and respiratory tract diseases for incident RA risk, stratified by serostatus

	Adjusted* OR (95% CI)			Overall GRS-Respiratory Interactions		
Respiratory tract disease	**GRS-low/ Resp+**	**GRS-high/ Resp−**	**GRS-high/ Resp+**	**AP (95% CI)**	**RERI (95% CI)**	**Multiplicative OR**
Seropositive RA outcomes (n=373)						
Any respiratory tract disease	1.40 (0.94,2.07)	4.68 (3.21,6.84)	3.76 (2.54,5.57)	−0.35 (−0.83,0.13)	−1.32 (−2.98,0.35)	**0.74 (0.59,0.94)**
Asthma	1.26 (0.71,2.22)	3.61 (2.70,4.81)	4.15 (2.44,7.07)	0.07 (−0.42,0.56)	0.29 (−1.88,2.45)	0.91 (0.65,1.26)
Interstitial lung disease	2.60 (0.43,15.59)	3.57 (2.71,4.69)	5.54 (1.46,21.12)	0.07 (−1.00,1.00)	0.38 (−8.33,9.09)	2.33 (0.58,9.47)
Pharyngitis	1.33 (0.72,2.45)	3.61 (2.71,4.82)	4.58 (2.51,8.36)	0.14 (−0.38,0.66)	0.64 (−2.05,3.33)	0.85 (0.59,1.22)
Acute pharyngitis	1.13 (0.54,2.35)	3.51 (2.65,4.66)	5.08 (2.59,9.95)	0.28 (−0.20,0.77)	1.44 (−1.87,4.76)	0.90 (0.60,1.35)
Chronic rhinitis/ pharyngitis	2.00 (0.81,4.93)	3.67 (2.78,4.84)	4.33 (1.50,12.51)	−0.08 (−1.00,1.00)	−0.33 (−5.18,4.51)	0.86 (0.46,1.64)
Pneumonia and acute lowerdiseases	1.39 (0.77,2.53)	3.59 (2.70,4.77)	4.86 (2.64,8.94)	0.18 (−0.33,0.69)	0.88 (−2.07,3.82)	0.95 (0.65,1.39)
Sinusitis	1.66 (0.85,3.23)	3.76 (2.83,5.01)	3.45 (1.91,6.24)	−0.28 (−1.07,0.51)	−0.97 (−3.25,1.32)	0.77 (0.52,1.15)
Acute sinusitis	1.68 (0.72,3.94)	3.66 (2.76,4.83)	3.93 (1.82,8.51)	−0.10 (−1.00,0.79)	−0.41 (−3.65,2.84)	0.73 (0.44,1.22)
Chronic sinusitis	2.02 (0.90,4.55)	3.85 (2.90,5.10)	2.19 (1.04,4.61)	−1.23 (−3.03,0.57)	**−2.58 (−5.15,−0.21)**	0.64 (0.40,1.02)
Seronegative RA (n=268)						
Any respiratory tract disease	1.02 (0.71,1.48)	1.44 (0.95,2.19)	1.27 (0.83,1.94)	−0.15 (0.75,0.45)	−0.19 (−0.94,0.55)	0.84 (0.64,1.10)
Asthma	1.10 (0.67,1.81)	1.47 (1.07,2.00)	0.80 (0.39,1.66)	−0.96 (−1.00,0.62)	−0.77 (−1.66,0.12)	0.80 (0.55,1.16)
Interstitial lung disease	0.95 (0.18,4.92)	1.34 (1.00,1.79)	N/A	N/A	−1.29 (−2.92,0.35)	0.33 (0.06,1.99)
Pharyngitis	1.98 (1.17,3.34)	1.42 (1.04,1.92)	1.69 (0.82,3.49)	−0.41 (−1.00,0.73)	−0.70 (−2.26,0.86)	0.84 (0.56,1.26)
Acute pharyngitis	2.33 (1.29,4.21)	1.38 (1.02,1.87)	2.35 (1.05,5.24)	−0.15 (−1.00,0.90)	−0.36 (−2.61,1.89)	0.82 (0.53,1.28)
Chronic rhinitis/ pharyngitis	1.10 (0.45,2.68)	1.39 (1.04,1.86)	0.30 (0.04,2.33)	N/A	−1.19 (−2.41,0.03)	0.52 (0.21,1.32)
Pneumonia and acute lowerdiseases	1.29 (0.69,2.40)	1.33 (0.81,1.80)	1.64 (0.81,3.32)	0.01 (−0.83,0.86)	0.02 (−1.38,1.41)	0.83 (0.54,1.26)
Sinusitis	1.44 (0.79,2.64)	1.36 (1.01,1.85)	1.53 (0.74,3.16)	−0.19 (−1.00,0.83)	−0.28 (−1.69,1.12)	0.73 (0.47,1.14)
Acute sinusitis	2.23 (1.03,4.80)	1.39 (1.03,1.86)	1.55 (0.50,4.75)	−0.69 (−1.00,1.00)	−1.06 (−3.47,1.34)	0.67 (0.36,1.25)
Chronic sinusitis	0.95 (0.44,2.08)	1.30 (0.97,1.76)	1.67 (0.75,3.73)	0.25 (−0.51,1.00)	0.41 (−1.10,1.93)	0.90 (0.53,1.50)

AP = attributable proportion, CI = confidence interval, COPD = chronic obstructive pulmonary disease, GRS = genetic risk score, *HLA* = human leukocyte antigen, N/A = not applicable due to small sample size, OR = odds ratio, RA = rheumatoid arthritis, RERI = relative excess risk due to interaction, Resp = respiratory tract exposure

*Reference group was individuals without respiratory tract disease exposure and lowest two GRS tertiles. Adjusting for age, sex, EHR history, enrollment year, race and ethnicity by principal components, education, body mass index, smoking status, and pack-years. Bold values are statistically significant.

Table 5.

Interaction between *MUC5B* promoter variant and respiratory tract diseases for incident seropositive RA risk

				MUC5B-Respiratory Interactions		
Respiratory tract disease	***MUC5B-/ Resp+***	***MUC5B+/ Resp−***	***MUC5B+/ Resp+***	**AP (95% CI)**	**RERI (95% CI)**	**Multiplicative OR**
Any respiratory tract disease	1.07 (0.80,1.44)	1.19 (0.73,1.93)	1.04 (0.59,1.82)	−0.21 (−1.04,0.62)	−0.22 (−1.02,0.57)	0.81 (0.40,1.67)
Asthma	1.03 (0.68,1.58)	0.95 (0.63,1.43)	2.58 (1.10,6.07)	**0.62 (0.24,1.00)**	1.60 (−0.61,3.81)	2.64 (0.98,7.09)
Interstitial lung disease	2.71 (0.53,13.8)	1.07 (0.73,1.57)	1.97 (0.34,11.4)	−0.41 (−1.00,1.00)	−0.80 (−6.40,4.79)	0.68 (0.06,7.57)
Pharyngitis	1.43 (0.89,2.30)	1.11 (0.75,1.65)	1.22 (0.42,3.53)	−0.26 (−1.00,1.19)	−0.32 (−1.81,1.16)	0.77 (0.23,2.52)
Acute pharyngitis	1.41 (0.83,2.40)	1.12 (0.76,1.64)	1.10 (0.26,4.65)	−0.40 (−1.00,1.00)	−0.44 (−2.20,1.33)	0.69 (0.15,3.28)
Chronic rhinitis/ pharyngitis	1.45 (0.65,3.27)	1.06 (0.72,1.56)	1.88 (0.50,7.12)	0.20 (−1.00,1.00)	0.37 (−2.39,3.13)	1.22 (0.25,5.92)
Pneumonia and acute lower diseases	1.44 (0.92,2.24)	1.12 (0.75,1.66)	1.18 (0.42,3.35)	−0.31 (−1.00,1.00)	−0.37 (−1.79,1.05)	0.74 (0.23,2.37)
Sinusitis	1.31 (0.77,2.22)	1.04 (0.70,1.55)	1.77 (0.71,4.42)	0.24 (−0.54,1.00)	0.43 (−1.29,2.14)	1.31 (0.45,3.78)
Acute sinusitis	1.17 (0.58,2.36)	1.04 (0.70,1.54)	2.05 (0.62,6.83)	0.41 (−0.41,1.00)	0.84 (−1.74,3.42)	1.69 (0.41,6.94)
Chronic sinusitis	1.16 (0.59,2.28)	1.04 (0.70,1.54)	2.00 (0.69,5.80)	0.40 (−0.34,1.00)	0.80 (−1.42,3.03)	1.66 (0.47,5.87)

CI = confidence interval, N/A = not applicable due to sample size, OR = odds ratio, RA = rheumatoid arthritis

*Reference group was individuals without respiratory tract disease exposure and lowest two GRS tertiles. Adjusting for age, sex, EHR history, enrollment year, race and ethnicity by principal components, education, body mass index, smoking status, and pack-years. Bold values are statistically significant.