Evolution of the Gain-of-Function *MUC5B* Promoter Variant

Evolution represents genomic adaptions to dynamic environmental challenges, and these genomic changes often come with both beneficial and detrimental consequences. Genomic changes, or alleles, that confer a survival advantage to offspring are more likely to be retained and increase in frequency. For instance, advantages resulting in positive selection may relate to increased early-life survival and/or reproductive success. Alleles that confer a disadvantage in these regards may be removed from the population via purifying selection. However, some alleles that confer increased risk for a disease that is unrelated to survival of offspring may be under positive selection due to their early-life survival and/or reproductive advantages. In those situations, a consequence of positive selection is that a detrimental genomic change that confers risk of disease is not removed via purifying selection and may even increase to a common frequency; we hypothesize that the *MUC5B* promoter variant is a case in point of this phenomenon. The work of Verma and colleagues (pp. 1220–1229) in this issue of the *Journal* (1) helps us understand the struggle for survival between different functional alleles of the *MUC5B* promoter and the potential beneficial and detrimental effects of rs35705950.

Over the past decade, we have found that: *1*) a common gain-of-function *MUC5B* promoter variant rs35705950 is the dominant idiopathic pulmonary fibrosis (IPF) risk factor (2), is present in >50% with IPF, and is also a primary risk factor for other types of usual interstitial pneumonia (UIP), including chronic hypersensitivity pneumonitis (3), asbestosis (4), and rheumatoid arthritis associated interstitial lung disease (5); *2*) the *MUC5B* promoter variant resides within an enhancer that is subject to epigenetic remodeling and contributes to pathologic misexpression of *MUC5B* (6); *3*) among patients with IPF or those with the *MUC5B* promoter variant, *MUC5B* is *misexpressed* specifically in bronchiolar epithelia and alveolar epithelial type 2 cells (1, 7,); and *4*) *MUC5B* appears to be involved in the pathogenesis of IPF (8, 9). Interestingly, the *MUC5B* promoter variant is relatively frequent among individuals of European ancestry (MAF∼9%) but rare in Asian (MAF∼1%) and nearly absent in African ancestry (10).

These findings, along with the reported relationship between SARS-CoV-2 lung infection and lung fibrosis prompted Verma and colleagues to investigate the relationship between the *MUC5B* promoter variant and SARS-CoV-2 infection (1). These investigators found that the gain-of-function *MUC5B* promoter variant was associated with 10% fewer coronavirus disease (COVID-19)-related hospitalizations and approximately 20% less COVID-19–related pneumonia in their study population. These differences were primarily observed in European study subjects. Interestingly, there was no clear relationship between the *MUC5B* promoter variant and COVID-19–associated infection, infection severity, or mortality. While the authors present nominally relevant data that suggest that their findings are specific to COVID-19, the specificity of host defense benefits of the *MUC5B* promoter variant remains uncertain. Additional research is needed to definitively address the relationship between the *MUC5B* promoter variant and post–COVID-19–associated lung fibrosis. Regardless, the potentially protective role of the *MUC5B* promoter variant and COVID-19 outcomes is provocative and supported by independent observations (11).

As discussed by the investigators (1), the *MUC5B* promoter variant may reduce the impact of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections by improving host defense through increased *MUC5B* expression. Enhanced expression of *Muc5b* in mice (12) and the gain-of-function *MUC5B* promoter variant in patients with IPF (13) have been reported to be associated with improved lung hygiene. Too little *Muc5b* in mice results in reduced mucociliary clearance, reduced IL-23 production, impaired macrophage phagocytosis, and chronic respiratory infections (12), and congenital absence of *MUC5B* in humans is associated with impaired mucociliary clearance, apoptotic macrophages, and repeated respiratory infections with *Staphylococcus aureus* (14). In mice, these host defense defects were ameliorated by increased expression of *Muc5b* (12). These findings along with the observations by Verma and colleagues (1) highlight the importance of understanding how *MUC5B*, and mucins in general, alter lung host defense.

Given likely beneficial host defense effects of the *MUC5B* promoter variant (1, 11–14) and high frequency of this allele in European populations, it is logical to speculate that this allele has undergone positive selection. To address this, we evaluated potential for positive selection on Chr11p15, and found the following evidence to suggest that the *MUC5B* region has undergone positive selection among Europeans. First, the rs35705950 allele is consistently out of Hardy-Weinberg equilibrium among IPF cases but not among controls (2), and the average shared haplotype length among chromosomes carrying the *MUC5B* risk variant is ∼10× longer than among chromosomes carrying the ancestral allele (Figure 1). Second, examination of the cross-population extended haplotype homozygosity statistic (surrogate indicator of recent positive selection) across the *MUC5B* region shows a peak in the promoter region of the *MUC5B* gene. Third, we observed differences in fine-scale European ancestry associated with the *MUC5B* risk variant. Specifically, the North–South gradient was significantly associated with number of copies of the *MUC5B* risk allele in IPF cases (*P*-value = 0.004; Figure 2), indicating that the risk variant is more prevalent in southern European regions. In aggregate, these findings suggest that the *MUC5B* promoter variant first appeared in humans in a region that produced migrants to Southern Europe and Asia and that the variant confers an early life survival advantage potentially through improved lung host defense. These observations may provide the evolutionary basis for the protective role of the *MUC5B* promoter variant and COVID-19 outcomes (1, 11).

The *MUC5B* promoter variant appears to have both beneficial (1, 11, 14) and detrimental (2–5) effects. Importantly, the beneficial

Supported by NIH (NHLBI): R01-HL149836 (D.A.S.), R01-HL158668 (D.A.S.), P01-HL0928701 (D.A.S. and T.E.F.), UG3-HL151865 (D.A.S.), and X01-HL134585 (D.A.S. and T.E.F.); Veterans Administration: IO1BX005295 (D.A.S.); and Department of Defense: W81XWH-17-1-0597 (D.A.S.).

Originally Published in Press as DOI: 10.1164/rccm.202207-1300ED on July 13, 2022

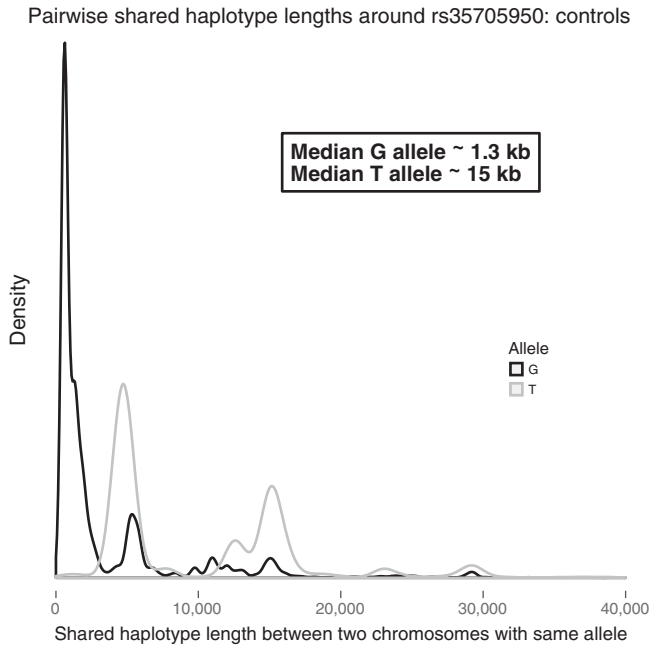

Figure 1. We calculated and compared the shared haplotype length among chromosomes carrying the T allele compared with chromosomes carrying the G allele at rs35705950. We imputed genotypes and phased haplotypes over a 5 Mb region around rs35705950. For previously reported IPF cases (N = 1,320) and controls (N = 1,730) (15) separately, and for each allele, we subsampled pairs of chromosomes and calculated the number of contiguous base pairs shared between the pair of chromosomes proceeding outward from rs35705950. To put the difference in shared haplotype length between the two alleles in the context of other alleles of a similar frequency, we similarly compared shared haplotype lengths among controls between alleles at 3,728 other SNPs across the genome that had the same observed minor allele frequency of 0.105 and were at least 1 Mb away from the boundaries of our 5 Mb imputed blocks around rs35705950. The difference in haplotype length between the T and G alleles among controls was in the 80th percentile of differences between alleles among other loci with the same observed minor allele frequency across the genome when we used physical distance to measure haplotype lengths and in the 91st percentile when using genetic distance.

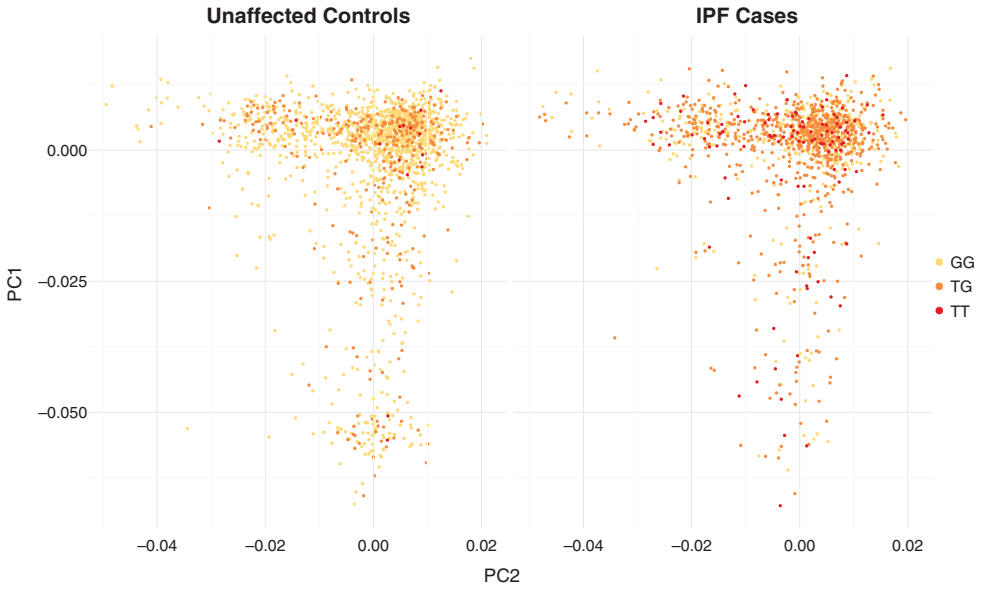

Figure 2. Using genome-wide SNP data, we computed principal components that reflect axes of European ancestry among previously reported unaffected controls (15) using EIGENSTRAT (16). We used the country-of-origin information for each unaffected control to interpret each principal component in terms of the direction of ancestry represented by that principal component. The first principal component (PC1) represented a North–South gradient. The rs35705950 risk variant was not included when computing principal components to derive estimates of European ancestry independent of the genotype at the risk variant at the *MUC5B* locus. This derived principal component was projected onto the previously reported IPF cases and unaffected control subjects (15) to generate estimates of North-South European Ancestry that are shown here and were used to test for association with the T allele at rs35705950. IPF = idiopathic pulmonary fibrosis.

effects are likely to occur across the age spectrum and are inclusive of the reproductive years, and the detrimental effects occur in the elderly. Consequently, one can understand how a gene variant that is the dominant cause of UIP (2–5) may have undergone positive selection. Moreover, there are numerous examples of highly morbid genetic modifications, such as mutations in the hemoglobin-β gene or cystic fibrosis transmembrane receptor, that have both beneficial and detrimental effects on the host.

The lung offers an unusual opportunity to understand the relationship between environmental challenges, genomic responses, and human disease. This relationship is particularly well illustrated in considering the role of host defense, the gain-of-function *MUC5B* promoter variant, and lung fibrosis. ■

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

David A. Schwartz, M.D.
Department of Medicine
and
Department of Immunology
University of Colorado Anschutz Medical Campus
Aurora, Colorado

Rachel Z. Blumhagen, Ph.D.
Tasha E. Fingerlin, Ph.D.
Center for Genes, Environment and Health
National Jewish Health
Denver, Colorado

ORCID ID: 0000-0001-6743-8443 (D.A.S.).

References

1. Verma A, Minnier J, Wan ES, Hufamn JE, Gao L, Joseph J, *et al.*; Million Veteran Program COVID-19 Science Initiative. A *MUC5B* gene polymorphism, rs35705950-T confers protective effects against COVID-19 hospitalization but not severe disease or mortality. *Am J Respir Crit Care Med* 2022;206:1220–1229.
2. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, *et al.* A common *MUC5B* promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364:1503–1512.
3. Furusawa H, Peljto AL, Walts AD, Cardwell J, Molyneaux PL, Lee JS, *et al.* Common idiopathic pulmonary fibrosis risk variants are associated with hypersensitivity pneumonitis. *Thorax* 2022;77:508–510.
4. Platenburg MGJP, Wiertz IA, van der Vis JJ, Crestani B, Borie R, Dieude P, *et al.* The *MUC5B* promoter risk allele for idiopathic pulmonary fibrosis predisposes to asbestosis. *Eur Respir J* 2020;55:55.
5. Juge PA, Lee JS, Ebstein E, Furukawa H, Dobrinskikh E, Gazal S, *et al. MUC5B* promoter variant and rheumatoid arthritis with interstitial lung disease. *N Engl J Med* 2018;379:2209–2219.
6. Gally F, Sasse SK, Kurche J, Cardwell JH, Gruca MA, Okamoto T, *et al.* Gain-of-function *MUC5B* promoter variant, strongly linked to idiopathic pulmonary fibrosis, defines an epigenetically poised enhancer. *JCI Insight* 2021;6:e144294.
7. Nakano Y, Yang IV, Walts AD, Watson AM, Helling BA, Fletcher AA, *et al. MUC5B* promoter variant rs35705950 affects *MUC5B* expression in the distal airways in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016;193:464–466.
8. Hancock LA, Hennessy CE, Solomon GM, Dobrinskikh E, Estrella A, Hara N, *et al. Muc5b* overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice. *Nat Commun* 2018;9:5363.
9. Stancil IT, Michalski JE, Davis-Hall D, Chu HW, Park JA, Magin CM, *et al.* Pulmonary fibrosis distal airway epithelia are dynamically and structurally dysfunctional. *Nat Commun* 2021;12:4566.
10. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, *et al.*; Genomes Project C. A global reference for human genetic variation. *Nature* 2015;526:68–74.
11. van Moorsel CHM, van der Vis JJ, Duckworth A, Scotton CJ, Benschop C, Ellinghaus D, *et al.* The *MUC5B* promoter polymorphism associates with severe COVID-19 in the European population. *Front Med (Lausanne)* 2021;8:668024.
12. Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, *et al. Muc5b* is required for airway defence. *Nature* 2014; 505:412–416.
13. Molyneaux PL, Cox MJ, Willis-Owen SA, Mallia P, Russell KE, Russell AM, *et al.* The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014;190: 906–913.
14. Costain G, Liu Z, Mennella V, Radicioni G, Goczi AN, Albulescu A, *et al.* Hereditary mucin deficiency caused by biallelic loss of function of *MUC5B*. *Am J Respir Crit Care Med* 2022;205:761–768.
15. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, *et al.* Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013;45:613–620.
16. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38: 904–909.