ORIGINAL ARTICLE

A Common *MUC5B* Promoter Polymorphism and Pulmonary Fibrosis

Max A. Seibold, Ph.D., Anastasia L. Wise, Ph.D., Marcy C. Speer, Ph.D.,*
Mark P. Steele, M.D., Kevin K. Brown, M.D., James E. Loyd, M.D.,
Tasha E. Fingerlin, Ph.D., Weiming Zhang, Ph.D.,
Gunnar Gudmundsson, M.D., Ph.D., Steve D. Groshong, M.D., Ph.D.,
Christopher M. Evans, Ph.D., Stavros Garantziotis, M.D.,
Kenneth B. Adler, Ph.D., Burton F. Dickey, M.D., Roland M. du Bois, M.D.,
Ivana V. Yang, Ph.D., Aretha Herron, B.A., Dolly Kervitsky, B.A., Janet L. Talbert, M.S.,
Cheryl Markin, B.A., Joungjoa Park, B.A., Anne L. Crews, B.A., Susan H. Slifer, Ph.D.,
Scott Auerbach, Ph.D., Michelle G. Roy, B.A., Jia Lin, B.A., Corinne E. Hennessy, M.S.,
Marvin I. Schwarz, M.D., and David A. Schwartz, M.D.

ABSTRACT

BACKGROUND

The mutations that have been implicated in pulmonary fibrosis account for only a small proportion of the population risk.

METHODS

Using a genomewide linkage scan, we detected linkage between idiopathic interstitial pneumonia and a 3.4-Mb region of chromosome 11p15 in 82 families. We then evaluated genetic variation in this region in gel-forming mucin genes expressed in the lung among 83 subjects with familial interstitial pneumonia, 492 subjects with idiopathic pulmonary fibrosis, and 322 controls. *MUC5B* expression was assessed in lung tissue.

RESULTS

Linkage and fine mapping were used to identify a region of interest on the p-terminus of chromosome 11 that included gel-forming mucin genes. The minor-allele of the single-nucleotide polymorphism (SNP) rs35705950, located 3 kb upstream of the *MUC5B* transcription start site, was present at a frequency of 34% among subjects with familial interstitial pneumonia, 38% among subjects with idiopathic pulmonary fibrosis, and 9% among controls (allelic association with familial interstitial pneumonia, $P=1.2\times10^{-15}$; allelic association with idiopathic pulmonary fibrosis, $P=2.5\times10^{-37}$). The odds ratios for disease among subjects who were heterozygous and those who were homozygous for the minor allele of this SNP were 6.8 (95% confidence interval [CI], 3.9 to 12.0) and 20.8 (95% CI, 3.8 to 113.7), respectively, for familial interstitial pneumonia and 9.0 (95% CI, 6.2 to 13.1) and 21.8 (95% CI, 5.1 to 93.5), respectively, for idiopathic pulmonary fibrosis. *MUC5B* expression in the lung was 14.1 times as high in subjects who had idiopathic pulmonary fibrosis as in those who did not ($P<0.001$). The variant allele of rs35705950 was associated with up-regulation in *MUC5B* expression in the lung in unaffected subjects (expression was 37.4 times as high as in unaffected subjects homozygous for the wild-type allele, $P<0.001$). MUC5B protein was expressed in lesions of idiopathic pulmonary fibrosis.

CONCLUSIONS

A common polymorphism in the promoter of *MUC5B* is associated with familial interstitial pneumonia and idiopathic pulmonary fibrosis. Our findings suggest that dysregulated *MUC5B* expression in the lung may be involved in the pathogenesis of pulmonary fibrosis. (Funded by the National Heart, Lung, and Blood Institute and others.)

From National Jewish Health (M.A.S., A.L.W., K.K.B., S.D.G., R.M.B., I.V.Y., D.K., J.L.T., J.L., C.E.H., M.I.S., D.A.S.); the University of Colorado–Denver, School of Medicine (A.L.W., K.K.B., S.D.G., M.I.S., D.A.S.); and the University of Colorado–Denver, School of Public Health (T.E.F., W.Z.) — all in Denver; Duke University Medical Center, Durham (M.C.S., M.P.S., A.H.); the National Institute of Environmental Health Sciences, Research Triangle Park (S.G., S.A.); and North Carolina State University, Raleigh (K.B.A., J.P., A.L.C.) — all in North Carolina; Vanderbilt University School of Medicine, Nashville (J.E.L., C.M.); Landspitali University Hospital and the Faculty of Medicine, University of Iceland — both in Reykjavik, Iceland (G.G.); the University of Texas M.D. Anderson Cancer Center, Houston (C.M.E., B.F.D., M.G.R.); and the University of Miami, Miami (S.H.S.). Address reprint requests to Dr. Schwartz at the Department of Medicine, University of Colorado, 12631 E. 17th Ave., Aurora, CO 80045, or at david.schwartz@udenver.edu.

*Deceased.

N Engl J Med 2011;364:1503-12.

The evidence that there is a genetic basis for idiopathic interstitial pneumonia is substantial, with familial aggregation confirmed through studies in twins, siblings raised apart, and multigenerational families.[1] Interstitial lung disease has been associated with several pleiotropic genetic disorders.[2] The development of pulmonary fibrosis has been associated with private mutations in surfactant protein C,[3] surfactant protein A2,[4] and genes that maintain telomere length.[5] However, collectively these mutations account for a small proportion of the population risk of idiopathic interstitial pneumonia.

In this article, we used a genomewide linkage scan to identify a risk locus for idiopathic interstitial pneumonia on the p-terminus of chromosome 11. Through fine mapping of this risk locus, we identified a common variant in the putative promoter of the gene encoding mucin 5B (*MUC5B*) that is associated with the development of both familial interstitial pneumonia and sporadic idiopathic pulmonary fibrosis.

Methods

Study Populations

National Jewish Health, Vanderbilt University, Duke University, and InterMune[6] identified and phenotyped subjects with familial interstitial pneumonia or idiopathic pulmonary fibrosis. The diagnosis of idiopathic interstitial pneumonia was established in accordance with conventional criteria.[7,8] Eligible subjects were at least 38 years of age and reported having symptoms of idiopathic interstitial pneumonia for at least 3 months. A high-resolution computed tomographic scan was required to show definite or probable idiopathic interstitial pneumonia in accordance with predefined criteria,[7,8] and a specimen from a surgical lung biopsy was obtained in 46% of affected subjects. Familial interstitial pneumonia was defined by the presence of two or more cases of definite or probable idiopathic interstitial pneumonia within three generations of a family, with at least one case of idiopathic interstitial pneumonia established as a definite or probable case of idiopathic pulmonary fibrosis. Subjects with clinically significant exposure to known fibrogenic agents or another cause of interstitial lung disease were excluded. Control subjects for the genetic analysis were recruited by Duke University and National Jewish Health (see Fig. S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). All protocols were approved by the institutional review boards at each institution, and all subjects provided written informed consent.

Linkage Analysis

A genomewide linkage screen was completed in 82 multiplex families (Fig. S2 in the Supplementary Appendix) with the use of a deCODE linkage panel consisting of a total of 884 markers (Table S1 in the Supplementary Appendix), with an average intermarker distance of 4.2 cM. Multipoint, nonparametric linkage analysis was performed with the use of Merlin.[9] Kong and Cox LOD scores[10] were calculated with the S_{pairs} statistic[11] with the use of an exponential model; the one-LOD-score-down method was used to support intervals.

Fine Mapping of Chromosome 11

To interrogate the linked region on the p-terminus of chromosome 11 (8.4 Mb bounded by rs702966 and rs1136966), we performed fine mapping by genotyping 306 tagging single-nucleotide polymorphisms (SNPs)[12] in 145 unrelated subjects with familial interstitial pneumonia, 152 subjects with idiopathic pulmonary fibrosis, and 233 controls (all subjects were white). Tests of association comparing subjects who had familial interstitial pneumonia and those who had idiopathic pulmonary fibrosis with controls were calculated with the use of an additive model for the minor allele.

Resequencing of *MUC2* and *MUC5AC*

Primer pairs, used to generate overlapping amplicons for resequencing the proximal promoter and most exons of the genes encoding mucin 2 (*MUC2*) and mucin 5AC (*MUC5AC*), were designed on sequences masked for repetitive elements, SNPs, and homology with other regions of the genome.

Genetic Screen of Gel-Forming Mucins in the Lung

To confirm and extend the linkage, fine mapping, and resequencing, we conducted a case–control association study in an independent population of 83 subjects with familial interstitial pneumonia, 492 subjects with sporadic idiopathic pulmonary fibrosis, and 322 controls (Table S2 in the Supplementary Appendix) with the use of tagging and other SNPs localized across the gel-forming mucin genes on chromosome 11 expressed in the lung (Table S3 in the Supplementary Appendix). A total of 175 SNPs were successfully genotyped with Sequenom iPLEX assays, and Haploview[13] was

used to test SNPs for allelic association with familial interstitial pneumonia and idiopathic pulmonary fibrosis. For associations that remained significant after Bonferroni correction, odds ratios were estimated with the use of an additive model for the rare allele after adjustment for age and sex with the use of logistic regression. Goodness-of-fit chi-square tests were used to evaluate the evidence for genotypic departures from Hardy–Weinberg equilibrium.[14] For the SNP most highly associated with disease, we used the linkage and association modeling in pedigrees[15] to determine whether, in the original linkage families, the SNP was linked to the disease locus, was in linkage disequilibrium with the disease locus, and could account for the linkage signal.

STATISTICAL ANALYSIS

A description of the statistical analysis, along with a comprehensive report on all methods, is available in the Supplementary Appendix.

RESULTS

IDENTIFICATION OF MUCIN REGION ON CHROMOSOME 11

On the basis of data gathered on the 82 families with familial interstitial pneumonia, the strongest evidence for linkage occurred on chromosome 11, where the maximum multipoint LOD score was 3.3 (P=0.00004; marker D11S1318) (Fig. S3 in the Supplementary Appendix). The 1-LOD support interval for this linked region was bounded by markers D11S4046 and D11S1760, spanning 3.4 Mb. Since D11S4046 was the most telomeric marker typed, the region of interest was inclusive of the p-terminus of chromosome 11 (Fig. 1A). Within the larger, 8.4-Mb region, 306 tagging SNPs were selected for fine mapping in a case–control association analysis (145 subjects with familial interstitial pneumonia, 152 subjects with idiopathic pulmonary fibrosis, and 233 controls) (Fig. 1B).

Allelic association testing revealed seven SNPs within *MUC2* that were significantly associated with either familial interstitial pneumonia or idiopathic pulmonary fibrosis (Fig. 1B). *MUC2* is contained in a genomic region harboring four gel-forming mucin genes (telomere to centromere: the gene encoding mucin 6 [*MUC6*], *MUC2*, *MUC5AC*, and *MUC5B*). Although recombination hotspots located between *MUC6* and *MUC2* and within the proximal portion of *MUC5B* have been reported, markers within *MUC2* and *MUC5AC* exhibit strong linkage disequilibrium.[16] Thus, *MUC2* and *MUC5AC* were selected for resequencing with the use of the oligonucleotide primers listed in Tables S5 and S6, respectively, in the Supplementary Appendix. Resequencing analysis identified 330 genetic variants in *MUC2* and 195 genetic variants in *MUC5AC* (Table S7 and Table S8, respectively, in the Supplementary Appendix). Allelic association testing between these genetic variants and disease status yielded seven independent SNPs in both *MUC2* and *MUC5AC* that were significantly associated with either familial interstitial pneumonia or idiopathic pulmonary fibrosis (Fig. 1C, and Tables S9 and S10 in the Supplementary Appendix).

A *MUC5B* PROMOTER SNP AND PULMONARY FIBROSIS

To further evaluate the region of interest, we designed a genetic screen (Tables S3 and S4 in the Supplementary Appendix) for common genetic variation across the three gel-forming mucin genes expressed in the lung (*MUC2*, *MUC5AC*, and *MUC5B*) in an independent population of subjects with idiopathic interstitial pneumonia (83 with familial interstitial pneumonia and 492 with idiopathic pulmonary fibrosis) and 322 controls (Fig. S1 and Table S2 in the Supplementary Appendix). On the basis of allelic testing, we observed significant associations of 19 independent SNPs with familial interstitial pneumonia, idiopathic pulmonary fibrosis, or both, after Bonferroni correction for multiple comparisons (Fig. 1D and Table 1). Of these 19 SNPs, 6 occurred in *MUC2*, 1 in the *MUC2–MUC5AC* intergenic region, 4 in *MUC5AC*, 3 in the *MUC5AC–MUC5B* intergenic region, and 5 in the putative *MUC5B* promoter, within 4 kb of the *MUC5B* transcription start site[17,18] (Table 1).

A SNP in the putative promoter of *MUC5B*, 3 kb upstream of the transcription start site (rs35705950), exhibited the strongest association with both familial interstitial pneumonia and idiopathic pulmonary fibrosis. The minor allele of this SNP was present at a frequency of 34% among subjects with familial interstitial pneumonia, 38% among those with idiopathic pulmonary fibrosis, and 9% among controls (allelic association with familial interstitial pneumonia, $P=1.2\times10^{-15}$; allelic association with idiopathic pulmonary fibrosis, $P=2.5\times10^{-37}$) (Fig. 1D). The genotypic frequencies for rs35705950 were consistent with Hardy–Weinberg equilibrium among controls but not among subjects with idiopathic pulmonary fibrosis ($P=6.0\times10^{-11}$) and were nearly consistent with

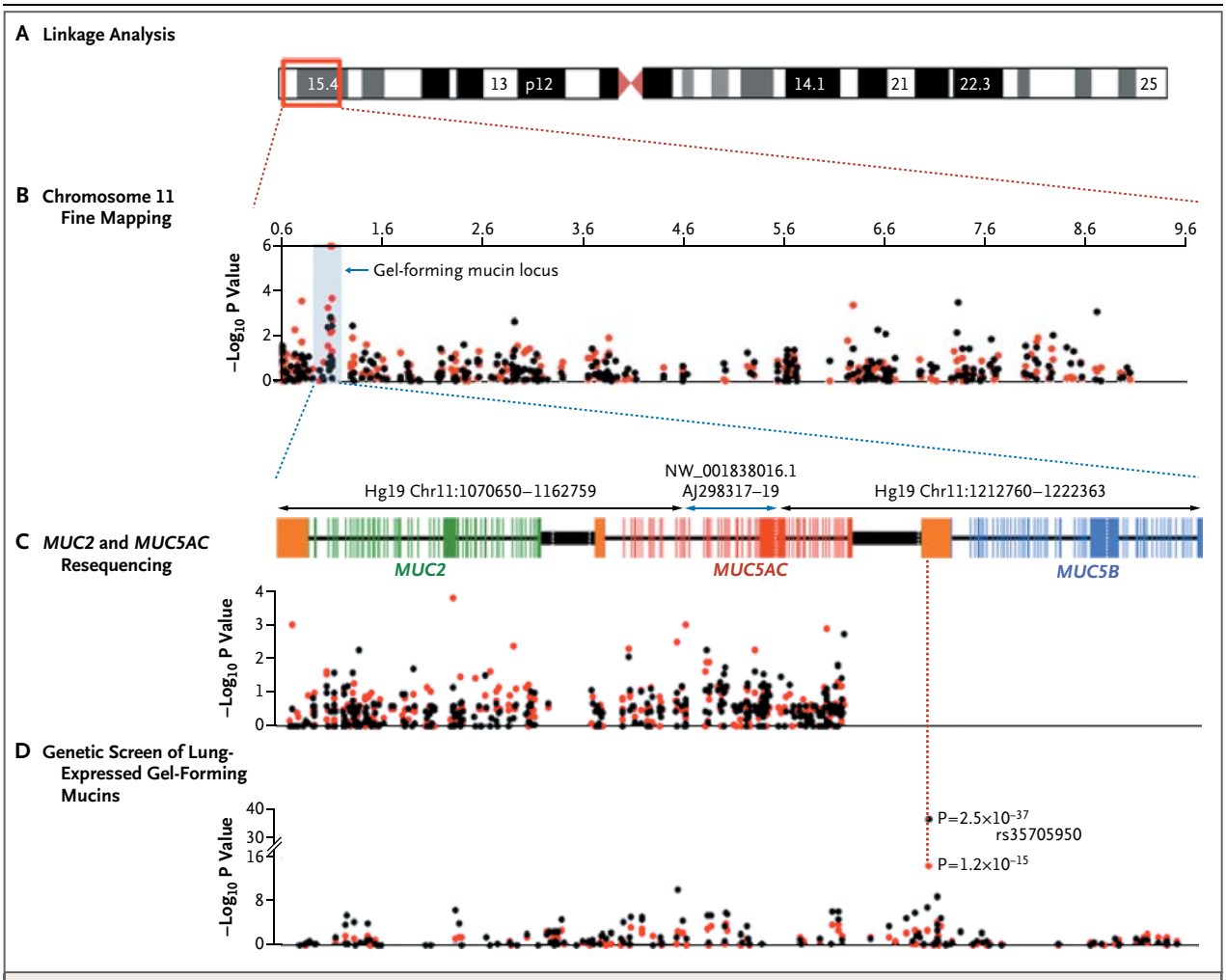

Figure 1. Summary of Genetic Screen in Familial Interstitial Pneumonia and Idiopathic Pulmonary Fibrosis.

In a linkage analysis involving 82 multiplex families with familial interstitial pneumonia, 884 markers were tested across the genome. Panel A shows the location of the linkage peak on chromosome 11 (maximum LOD score, 3.3) and the adjacent area of fine mapping (red box). Panel B shows the results of fine mapping of the chromosome 11 linkage peak. A total of 306 tagging SNPs were typed across the linkage peak in 145 subjects with familial interstitial pneumonia, 152 subjects with idiopathic pulmonary fibrosis, and 233 healthy controls. Chromosomal location (in megabases) is shown on the x axis. The resequencing diagram in Panel C shows the entire genomic region from *MUC2* to *MUC5B*, drawn to scale except where the dotted line appears. The green, red, and blue blocks represent gene exons, and the orange blocks putative promoter regions. The black blocks represent intergenic regions, and the black horizontal line represents introns. SNPs identified by means of resequencing and tested for association in 54 controls who were spouses of subjects, 96 subjects with sporadic idiopathic pulmonary fibrosis, and 69 subjects with familial interstitial pneumonia are displayed in Panel B as dots. SNPs genotyped in an independent group of 83 subjects with familial interstitial pneumonia, 492 subject with sporadic idiopathic pulmonary fibrosis, and 322 controls are shown in Panel D. In Panels B, C, and D, the results of allelic association testing for each SNP are denoted by red dots for familial interstitial pneumonia and by black dots for idiopathic pulmonary fibrosis. (See the Supplementary Appendix for a description of gene models of gel-forming mucins.)

Hardy–Weinberg equilibrium among subjects with familial interstitial pneumonia (P=0.11).

A comparison of the genotype frequencies observed in case subjects and controls with the frequencies that would be expected if rs35705950 were a true risk locus revealed that the observed genotypic frequencies are consistent with an additive genotypic effect on the risk of disease conferred by rs35705950 (P=0.88 for familial interstitial pneumonia and P=0.77 for idiopathic pulmonary fibrosis [P values indicate findings consistent with an additive model]). In addition, the estimates of frequency and penetrance for the disease allele suggest that the disease models for

Table 1. Additive Genotypic Associations of *MUC2*, *MUC5AC*, and *MUC5B* Single-Nucleotide Polymorphisms (SNPs) in Subjects with Familial Interstitial Pneumonia or Idiopathic Pulmonary Fibrosis and Controls.*

SNP	Nucleotide and Amino Acid Change	Mucin Region	Nucleotide Position†	Minor-Allele Frequency			Genotypic Association Test			
				FIP (N=83)	IPF (N=492)	Controls (N=322)	Odds Ratio for FIP (95% CI)	P Value	Odds Ratio for IPF (95% CI)	P Value
				percent						
rs10902081	C→T	*MUC2* Int7	1079809	37.2	38.6	47.9	0.6 (0.4–0.9)	0.011	0.7 (0.5–0.8)	4.3×10^{-4}
rs7127117‡	T→C	*MUC2* Int7	1079879	49.3	60.0	47.4	1.0 (0.7–1.5)	0.826	1.6 (1.3–2.0)	6.9×10^{-5}
rs41453346	C→T Tyr426Tyr	*MUC2* Ex10	1080894	5.0	6.5	2.2	1.9 (0.8–4.3)	0.124	2.8 (1.6–5.2)	0.001
rs41480348	G→A Thr618Thr	*MUC2* Ex15	1082605	8.4	6.5	12.1	0.7 (0.4–1.2)	0.188	0.5 (0.4–0.8)	0.001
rs7934606‡	C→T	*MUC2* Int31	1093945	49.4	54.0	40.5	1.4 (1.0–2.0)	0.055	1.7 (1.4–2.2)	3.8×10^{-6}
rs10902089‡	A→G	*MUC2* Int31	1094357	57.9	58.8	48.5	1.5 (1.0–2.1)	0.031	1.5 (1.2–1.9)	2.9×10^{-4}
rs9667239	C/T	*MUC2-5AC* intergenic region	1143101	22.5	21.0	12.5	2.2 (1.4–3.6)	0.001	1.9 (1.4–2.7)	5.6×10^{-4}
rs55846509	G→A Arg47Gln	*MUC5AC* Ex2	1154294	3.1	5.5	1.6	1.7 (0.6–5.1)	0.316	3.6 (1.7–7.3)	0.001
rs28403537	C→T Ala497Val	*MUC5AC* Ex12	1161315	8.9	13.0	3.4	2.7 (1.3–5.3)	0.006	4.6 (2.8–7.6)	3.2×10^{-9}
MUC5AC-025447‡	C→T	*MUC5AC* Int26	826476§	20.1	21.0	13.8	1.6 (1.0–2.5)	0.053	1.6 (1.2–2.2)	0.003
rs35288961	G→T	*MUC5AC* Int46	1220462	28.8	26.6	15.9	2.2 (1.4–3.5)	3.2×10^{-4}	2.0 (1.5–2.6)	3.7×10^{-6}
rs35671223	C→T	*MUC5AC–5B* intergenic region	1227069	42.6	42.4	33.4	1.4 (1.0–2.0)	0.05	1.5 (1.2–1.9)	0.001
rs28654232	C→T	*MUC5AC–5B* intergenic region	1229227	21.6	22.8	32.9	0.6 (0.4–0.9)	0.009	0.6 (0.5–0.8)	1.1×10^{-4}
rs34595903‡	C→T	*MUC5AC–5B* intergenic region	1230393	21.5	23.3	34.8	0.5 (0.3–0.7)	0.001	0.5 (0.4–0.7)	2.4×10^{-6}
rs2672794	C→T	*MUC5B* promoter	1241005	27.2	27.5	40.4	0.5 (0.3–0.8)	0.001	0.5 (0.4–0.7)	1.9×10^{-7}
rs35705950	G→T	*MUC5B* promoter	1241221	33.8	37.5	9.1	6.2 (3.7–10.4)	3.7×10^{-12}	8.3 (5.8–11.9)	4.6×10^{-31}
rs35619543‡	G→T	*MUC5B* promoter	1242250	40.3	39.0	23.8	2.4 (1.6–3.6)	3.3×10^{-5}	2.1 (1.6–2.8)	1.5×10^{-8}
rs12804004	G→T	*MUC5B* promoter	1242299	39.2	39.4	48.9	0.6 (0.4–0.9)	0.019	0.6 (0.5–0.8)	1.2×10^{-4}
rs868903‡	T→C	*MUC5B* promoter	1242690	65.4	61.0	49.5	1.8 (1.3–2.6)	0.001	1.6 (1.3–2.1)	2.8×10^{-5}

* FIP denotes familial interstitial pneumonia, and IFP idiopathic pulmonary fibrosis.
† Nucleotide positions are based on data from human genome browser hg19, except where denoted otherwise.
‡ For these SNPs, DNA was available for 304 controls.
§ The position provided for this nucleotide is based on genome build NW_001838016.1.

familial interstitial pneumonia and idiopathic pulmonary fibrosis are similar. The odds ratio for disease among subjects who were heterozygous and those who were homozygous for the rarer allele of this SNP were 6.8 (95% confidence interval [CI], 3.9 to 12.0) and 20.8 (95% CI, 3.8 to 113.7), respectively, for familial interstitial pneumonia, and 9.0 (95% CI, 6.2 to 13.1) and 21.8 (95% CI, 5.1 to 93.5), respectively, for idiopathic pulmonary fibrosis (Table 1). To be certain that this SNP was not tagging another SNP in the *MUC5B* promoter region, we resequenced the 4-kb region upstream of the *MUC5B* transcription start site in 48 subjects with idiopathic pulmonary fibrosis and 48 controls (Table S11 in the Supplementary Appendix). We observed 34 genetic variants, but none had a pairwise r^2 statistic for linkage disequilibrium with rs35705950 that was above 0.2 (Table S12 in the Supplementary Appendix). Finally, among the families in the original linkage analysis, rs35705950 was found to be both linked to the disease locus (P=0.04) and in linkage disequilibrium with it ($P=1.5\times10^{-9}$). Although there is some evidence for other linked variants in the region (P=0.054), these results verify the relevance of this SNP to disease in these families.

RELATIONSHIP OF RS35705950 WITH OTHER MUCIN SNPS

Next, we explored the relationship between the rs35705950 SNP and the 18 other SNPs significantly associated with idiopathic interstitial pneu-

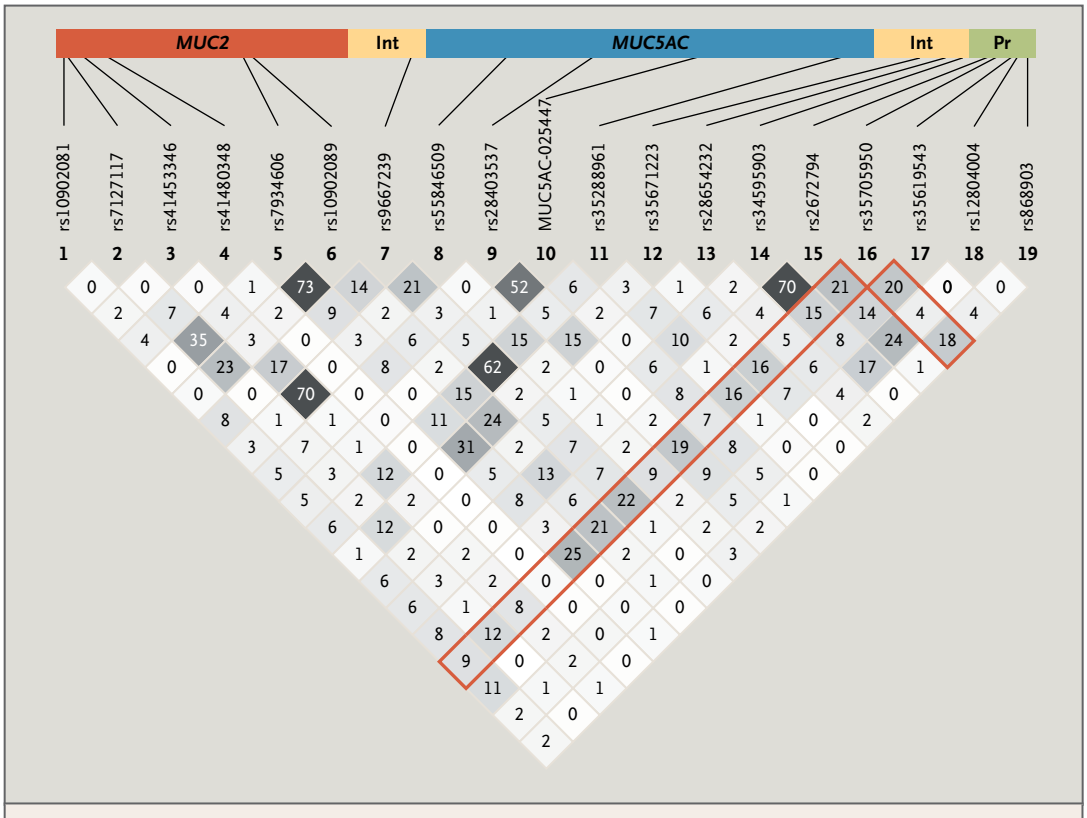

Figure 2. Pairwise Linkage Disequilibrium Plot for Single-Nucleotide Polymorphisms (SNPs) Significantly Associated with Idiopathic Pulmonary Fibrosis or Familial Interstitial Pneumonia in a Genetic Screen of Lung-Expressed Gel-Forming Mucins.

The linkage-disequilibrium values displayed were calculated with the use of the r^2 statistic for the mucin genetic screen performed in 492 subjects with idiopathic pulmonary fibrosis. The bar above the plot indicates the approximate location of these SNPs within the gel-forming mucin region. The highly significant *MUC5B* promoter SNP (rs35705950) and its pairwise linkage-disequilibrium values are outlined in red. Linkage-disequilibrium patterns were qualitatively similar among controls, although in most instances the linkage disequilibrium was also weaker among controls (Fig. S5 in the Supplementary Appendix). Areas with darker shading represent stronger linkage disequilibrium. Int denotes intergenic region, and Pr *MUC5B* promoter.

monia. Testing pairwise linkage disequilibrium among these 18 SNPs with the use of the r^2 statistic, we found that 10 exhibited a low but substantial level of linkage disequilibrium with rs35705950 (r^2=0.15 to 0.27) among subjects with idiopathic pulmonary fibrosis, suggesting that the association of these SNPs with idiopathic interstitial pneumonia is due to linkage disequilibrium with rs35705950 (Fig. 2). Using genotypic logistic-regression models to adjust for the effects of rs35705950, we observed that the coefficients and corresponding P values were substantially reduced for all 18 SNPs that had previously been associated with familial interstitial pneumonia or idiopathic pulmonary fibrosis (Table S13 in the Supplementary Appendix). In fact, after adjustment for rs35705950, only one SNP retained nominal significance for idiopathic pulmonary fibrosis (rs41480348, P=0.04). The significance of the association of the rs35705950 SNP with disease was largely unaffected by adjustment for any of the 18 SNPs tested ($P<1.7\times10^{-9}$ for familial interstitial pneumonia and $P<1.1\times10^{-24}$ for idiopathic pulmonary fibrosis) (Table S13 in the Supplementary Appendix). These results show that the rs35705950 SNP has a strong independent effect on both familial interstitial pneumonia and idiopathic pulmonary fibrosis, with a minimal effect of other mucin variants on the risk of disease.

EFFECT OF RS35705950 ON *MUC5B* EXPRESSION

The wild-type G allele of the rs35705950 SNP is conserved across primate species, is directly 5′ (or adjacent) to a highly conserved region across vertebrate species, and is in the middle of a sequence that has been predicted to be involved in gene regulation (Fig. S4 in the Supplementary Appendix).[19,20] Bioinformatic analysis of the effect of the rs35705950 SNP predicted disruption of an E2F binding site and the creation of two new binding sites (HOX9 and PAX2). On the basis of these analyses, we explored the effect of rs35705950 on *MUC5B* expression in lung tissue from 33 subjects with idiopathic pulmonary fibrosis and 47 unaffected subjects. *MUC5B* expression was 14.1 times as high among the subjects with idiopathic pulmonary fibrosis as it was among the unaffected subjects (P<0.001) (Fig. 3A). Expression of the gene among unaffected subjects carrying at least one copy of the variant allele was 37.4 times as high as it was among unaffected subjects who were homozygous for the wild-type allele (P<0.001) (Fig. 3B). In contrast, no significant difference in *MUC5B* expression was observed between subjects with idiopathic pulmonary fibrosis who had at least one variant allele of rs35705950 and subjects with idiopathic pulmonary fibrosis who were homozygous for the wild-type allele (Fig. 3C). Smoking, a potential confounder of *MUC5B* expression, appeared to have little effect on the association between the rs35705950 variant allele and *MUC5B* expression among subjects with idiopathic pulmonary fibrosis or among those who were unaffected (Fig. 3B and 3C).

SITES OF MUC5B EXPRESSION IN THE LUNG

Immunohistochemical staining of lung tissue showed that MUC5B was present in the cytoplasm of the secretory columnar cells of the bronchi and in larger proximal bronchioles (>200 μm) in subjects with idiopathic pulmonary fibrosis and in controls (Fig. 4A). In subjects with idiopathic pulmonary fibrosis, regions of dense accumulation of MUC5B were observed in areas of microscopical honeycombing and involved patchy staining of the metaplastic epithelia lining the honeycomb cysts (Fig. 4B); there was also staining of the mucous plugs within the cysts (Fig. 4C). No obvious differences in MUC5B staining characteristics were observed in subjects with idiopathic pulmonary fibrosis who had the *MUC5B* promoter polymorphism.

DISCUSSION

We identified a common variant in the putative promoter of *MUC5B* that is associated with the development of familial interstitial pneumonia and idiopathic pulmonary fibrosis. Linkage, fine mapping, selective resequencing of *MUC2* and *MUC5AC,* and a genetic analysis of the gel-forming mucin region at the p-terminus of chromosome 11 resulted in the identification of a SNP (rs35705950) in the putative *MUC5B* promoter that is strongly associated with both *MUC5B* expression in the lung among unaffected subjects and the development of familial interstitial pneumonia and idiopathic pulmonary fibrosis. Moreover, subjects with idiopathic pulmonary fibrosis had significantly higher levels of expression of *MUC5B* in the lungs than did controls, and the MUC5B protein was expressed in lesions of idiopathic pulmonary fibrosis. On the basis of our findings, the population attributable risk of fa-

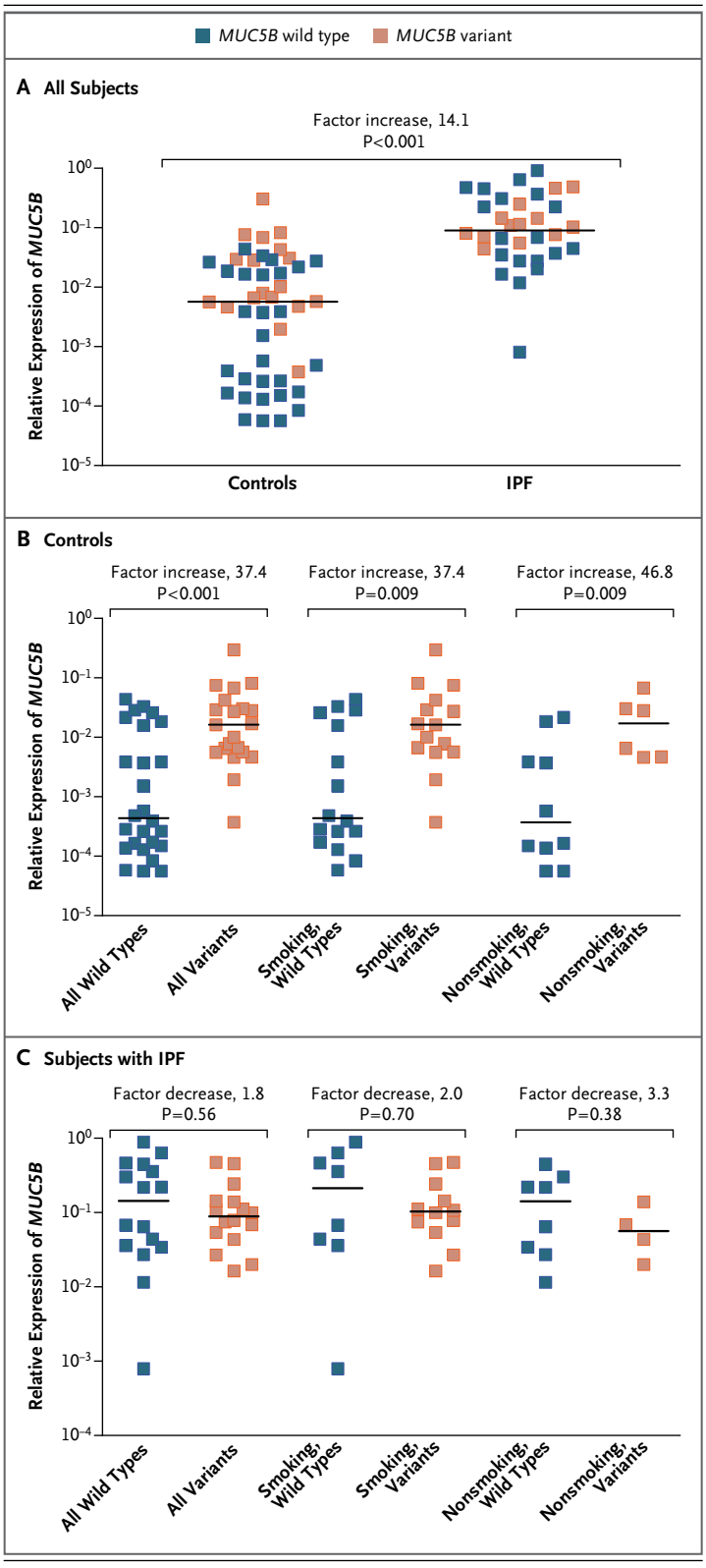

Figure 3. *MUC5B* Expression in 33 Subjects with Idiopathic Pulmonary Fibrosis (IPF) and 47 Healthy Controls, Stratified According to *MUC5B* Promoter Single-Nucleotide Polymorphism (SNP) (rs35705950) Genotype and Smoking Status.

Panel A shows the distribution of *MUC5B* expression among subjects with wild-type or heterozygous rs35705950 genotypes of *MUC5B* according to the presence or absence of IPF. Panel B shows *MUC5B* expression among all controls, controls who smoked, and controls who did not smoke, according to rs35705950 genotype. Panel C compares *MUC5B* expression in all subjects with IPF, those who smoked, and those who did not smoke, according to rs35705950 genotype. In all the panels, the horizontal lines indicate group medians, and the expression of *MUC5B* is shown in relation to the expression of the glyceraldehyde 3-phosphate dehydrogenase gene (*GAPDH*).

milial interstitial pneumonia or idiopathic pulmonary fibrosis for this promoter polymorphism is likely to be substantial. In aggregate, our findings suggest that dysregulated *MUC5B* expression in the lung may be involved in the pathogenesis of pulmonary fibrosis.

The prevailing opinion is that idiopathic pulmonary fibrosis develops as a result of excessive, sequential lung injury or aberrant wound healing.[21] Although the mechanisms that account for excessive lung injury or aberrant repair in persons with the SNP in the putative *MUC5B* promoter remain unknown, our findings point to at least three possibilities. First, on the basis of the relationship between the SNP and excess production of MUC5B, we hypothesize that too much MUC5B impairs mucosal host defense, results in excessive lung injury from inhaled substances, and over time leads to the development of idiopathic interstitial pneumonia. In addition to the *MUC5B* promoter SNP, common exposures and basic biologic processes can influence either the expression of the gene or the clearance of the protein. For instance, *MUC5B* expression can be enhanced in the lung by cigarette smoke,[17,22] acrolein,[23] oxidative stress,[24] interleukin-6,[25] interleukin-8,[26] interleukin-13,[27] interleukin-17,[25] 17β-estradiol,[28] extracellular nucleotides,[22] or epigenetic changes that alter DNA methylation or chromatin structure.[29] In addition, clearance of lung mucus is dependent on effective ciliary motion, adequate hydration of the periciliary liquid layer, and an intact cough.[26] However, regardless of whether excess MUC5B in the air space is

caused by overexpression or impaired clearance, our findings raise the possibility that excess concentrations of MUC5B compromise mucosal host defense, reducing lung clearance of inhaled particles, dissolved chemicals, and microorganisms. Given the importance of environmental exposures (e.g., exposure to asbestos and silica) in the development of other forms of interstitial lung disease, it is logical to speculate that common inhaled particles, such as those associated with cigarette smoke or air pollution, might cause exaggerated interstitial injury in persons who have defects in mucosal host defense.

Second, excess MUC5B in the respiratory bronchioles may interfere with alveolar repair. It has been established that alveolar injury results in the collapse of bronchoalveolar units and that this focal lung injury is repaired through re-epithelialization of the alveolus by type II alveolar epithelial cells.[30,31] Thus, an alternative hypothesis is that MUC5B impedes alveolar repair, either by interfering with the interaction between the type II alveolar epithelial cells and the underlying matrix or by interfering with the surface-tension properties of surfactant. Either the failure to re-epithelialize the basal lamina of the alveolus or the presence of suboptimal surfactant activity could enhance ongoing alveolar collapse and fibrosis of adjacent bronchoalveolar units and eventually lead to the development of idiopathic interstitial pneumonia.

Third, the lesions of idiopathic pulmonary fibrosis are spatially heterogeneous,[21] suggesting that the disorder is multifocal, originating in individual bronchoalveolar units. Since the rs35705950 SNP occurs in the putative promoter region of *MUC5B* and is predicted to disrupt transcription-factor binding sites, one must consider ectopic production of MUC5B in cells or locations that cause injury to the bronchoalveolar unit. Although all three mechanisms are plausible and may act alone or together to contribute to the development of idiopathic interstitial pneumonia in persons with the *MUC5B* promoter SNP, further research is needed to define the mechanisms involved in *MUC5B*-associated interstitial lung disease and to address the possibility that unscreened genetic variants (especially in the inaccessible repetitive mucin regions) in linkage disequilibrium with the *MUC5B* promoter SNP affect the function of other lung mucins.

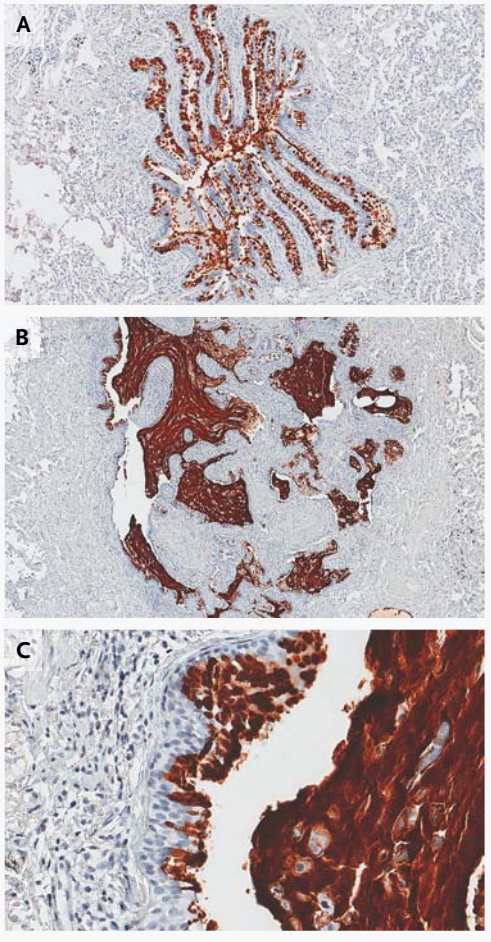

Figure 4. Immunohistochemical Staining of MUC5B in Lung Tissue from Subjects with Idiopathic Pulmonary Fibrosis and Controls.

Immunohistochemical staining showed MUC5B distribution in the cytoplasm of the secretory columnar cells of the bronchi and larger proximal bronchioles in a specimen of lung tissue from a control subject (Panel A). In subjects with idiopathic pulmonary fibrosis, regions of dense accumulation of MUC5B were observed in areas of microscopical honeycombing and involved patchy staining of the metaplastic epithelia lining the honeycomb cysts (Panel B). Accumulation was also observed in the mucous plugs within the cysts (Panel C).

Our results could potentially alter the clinical approach to idiopathic interstitial pneumonia. The implication of secreted airway mucins in the pathogenesis of pulmonary fibrosis suggests that the air space plays a role in the development of idiopathic interstitial pneumonia. Although identification of rs35705950 in the putative *MUC5B*

promoter can be used to target persons at risk for the development of idiopathic interstitial pneumonia (especially those who are members of families subject to idiopathic interstitial pneumonia), our observation that the biologic features of mucins may be important etiologic factors in this disease could reorient the focus of pathogenic and therapeutic studies in interstitial lung disease to lung mucins, the air space, and the bronchoalveolar unit. In addition, the genetic causes of idiopathic interstitial pneumonia — *MUC5B*, surfactant protein C, surfactant protein A2, and two telomerase genes — could provide insight into the particular clinical manifestations of this complex disease process and consequently lead to earlier detection, more predictable prognosis, and personalized therapeutic strategies.

Supported by grants from the National Heart, Lung, and Blood Institute (U01-HL067467, R01-HL095393, R01-HL097163, P01-HL092870, RC2-HL101715, and R37-HL36982), the National Institutes of Health Intramural Research Program of the National Institute of Environmental Health Sciences (Z01 ES101947), the National Cancer Institute (P30-CA016672), the American Lung Association (RG-22720-N), and the Cystic Fibrosis Foundation (06IO) and by the Chapman Foundation and InterMune.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Dr. Kenneth Beckman at the University of Minnesota for his support with regard to genotyping and all the study subjects who provided personal information that allowed us to further understand the causes of pulmonary fibrosis.

REFERENCES

1. Steele MP, Speer MC, Loyd JE, et al. Clinical and pathologic features of familial interstitial pneumonia. Am J Respir Crit Care Med 2005;172:1146-52.
2. Raghu G, Hert R. Interstitial lung diseases: genetic predisposition and inherited interstitial lung diseases. Sem Respir Med 1993;14:323-32.
3. Nogee LM, Dunbar AE III, Wert SE, Askin F, Hamvas A, Whitsett JA. A mutation in the surfactant protein C gene associated with familial interstitial lung disease. N Engl J Med 2001;344:573-9.
4. Wang Y, Kuan PJ, Xing C, et al. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. Am J Hum Genet 2009;84:52-9.
5. Armanios MY, Chen JJ, Cogan JD, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med 2007;356:1317-26.
6. King TE Jr, Albera C, Bradford WZ, et al. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. Lancet 2009;374:222-8.
7. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment: international consensus statement. Am J Respir Crit Care Med 2000;161:646-64.
8. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias: this joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. Am J Respir Crit Care Med 2002;165:277-304. [Erratum, Am J Respir Crit Care Med 2002;166:426.]
9. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin — rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30:97-101.
10. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 1997;61:1179-88.
11. Whittemore AS, Halpern J. A class of tests for linkage using affected pedigree members. Biometrics 1994;50:118-27.
12. Xu H, Gregory SG, Hauser ER, et al. SNPselector: a web tool for selecting SNPs for genetic association studies. Bioinformatics 2005;21:4181-6.
13. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.
14. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:967-86.
15. Li M, Boehnke M, Abecasis GR. Joint modeling of linkage and association: identifying SNPs responsible for a linkage signal. Am J Hum Genet 2005;76:934-49.
16. Rousseau K, Byrne C, Griesinger G, et al. Allelic association and recombination hotspots in the mucin gene (MUC) complex on chromosome 11p15.5. Ann Hum Genet 2007;71:561-9.
17. Chen Y, Zhao YH, Di YP, Wu R. Characterization of human mucin 5B gene expression in airway epithelium and the genomic clone of the amino-terminal and 5′-flanking region. Am J Respir Cell Mol Biol 2001;25:542-53.
18. Wu DY, Wu R, Chen Y, Tarasova N, Chang MM. PMA stimulates MUC5B gene expression through an Sp1-based mechanism in airway epithelial cells. Am J Respir Cell Mol Biol 2007;37:589-97.
19. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005;15:1034-50.
20. King DC, Taylor J, Elnitski L, Chiaromonte F, Miller W, Hardison RC. Evaluation of regulatory potential and conservation scores for detecting cis-regulatory modules in aligned mammalian genome sequences. Genome Res 2005;15:1051-60.
21. Gross TJ, Hunninghake GW. Idiopathic pulmonary fibrosis. N Engl J Med 2001;345:517-25.
22. Chen Y, Zhao YH, Wu R. Differential regulation of airway mucin gene expression and mucin secretion by extracellular nucleotide triphosphates. Am J Respir Cell Mol Biol 2001;25:409-17.
23. Borchers MT, Carty MP, Leikauf GD. Regulation of human airway mucins by acrolein and inflammatory mediators. Am J Physiol 1999;276:L549-L555.
24. Casalino-Matsuda SM, Monzon ME, Day AJ, Forteza RM. Hyaluronan fragments/CD44 mediate oxidative stress-induced MUC5B up-regulation in airway epithelium. Am J Respir Cell Mol Biol 2009;40:277-85.
25. Chen Y, Thai P, Zhao YH, Ho YS, DeSouza MM, Wu R. Stimulation of airway mucin gene expression by interleukin (IL)-17 through IL-6 paracrine/autocrine loop. J Biol Chem 2003;278:17036-43.
26. Rose MC, Voynow JA. Respiratory tract mucin genes and mucin glycoproteins in health and disease. Physiol Rev 2006;86:245-78.
27. Zhen G, Park SW, Nguyenvu LT, et al. IL-13 and epidermal growth factor receptor have critical but distinct roles in epithelial cell mucin production. Am J Respir Cell Mol Biol 2007;36:244-53.
28. Choi HJ, Chung YS, Kim HJ, et al. Signal pathway of 17beta-estradiol-induced MUC5B expression in human airway epithelial cells. Am J Respir Cell Mol Biol 2009;40:168-78.
29. Vincent A, Perrais M, Desseyn JL, Aubert JP, Pigny P, Van Seuningen I. Epigenetic regulation (DNA methylation, histone modifications) of the 11p15 mucin genes (MUC2, MUC5AC, MUC5B, MUC6) in epithelial cancer cells. Oncogene 2007;26:6566-76.
30. Crouch E. Pathobiology of pulmonary fibrosis. Am J Physiol 1990;259:L159-L184.
31. Kuhn C III, Boldt J, King TE Jr, Crouch E, Vartio T, McDonald J. An immunohistochemical study of architectural remodeling and connective tissue synthesis in pulmonary fibrosis. Am Rev Respir Dis 1989;140:1693-703.