Regulation of *MUC5B* Expression in Idiopathic Pulmonary Fibrosis

Britney A. Helling¹, Anthony N. Gerber^{1,2}, Vineela Kadiyala², Sarah K. Sasse², Brent S. Pedersen¹, Lenore Sparks¹, Yasushi Nakano¹, Tsukasa Okamoto¹, Christopher M. Evans¹, Ivana V. Yang^{1,3}, and David A. Schwartz^{1,2}

¹Department of Medicine, School of Medicine, University of Colorado–Denver, Denver, Colorado; ²Department of Medicine, National Jewish Health, Denver, Colorado; and ³University of Colorado–Denver, School of Public Health, Denver, Colorado

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

Abstract

The gain-of-function mucin 5B (*MUC5B*) promoter variant, rs35705950, confers the largest risk, genetic or otherwise, for the development of idiopathic pulmonary fibrosis; however, the mechanisms underlying the regulation of *MUC5B* expression have yet to be elucidated. Here, we identify a critical regulatory domain that contains the *MUC5B* promoter variant and has a highly conserved forkhead box protein A2 (FOXA2) binding motif. This region is differentially methylated in association with idiopathic pulmonary fibrosis, *MUC5B* expression, and rs35705950. In addition, we show that this locus binds FOXA2 dynamically, and that binding of FOXA2 is necessary for enhanced expression of *MUC5B*. In aggregate, our findings identify novel targets to regulate the expression of *MUC5B*.

Keywords: mucin 5B; idiopathic pulmonary fibrosis; gene regulation; DNA methylation; mucins

Clinical Relevance

Mucin dysregulation is a common phenotype in a number of pulmonary diseases, namely, idiopathic pulmonary fibrosis (IPF). Currently, there are limited treatment options for IPF; however, our research highlights critical regulatory mechanisms for mucin 5B, which may be targeted as a therapeutic for IPF and perhaps other pulmonary diseases.

Idiopathic pulmonary fibrosis (IPF) is a rapidly progressive and often fatal fibrosing interstitial lung disease. Currently, there is no cure for IPF, and treatment options are only modestly effective. Proteomics (1), epigenomics (2–4), and genetics (5, 6) have begun to define the etiology and heterogeneity of this complex disease, and may provide insight into more targeted therapeutics.

Recent genome-wide association studies identified a paradoxical link between a gene encoding a mucin produced in the conducting airways and IPF, the etiology of which is typically thought to be restricted to peripheral alveolar lung origins. The gain-of-function promoter variant, rs35705950, is

3 kb upstream of the mucin 5B (*MUC5B*) transcriptional start site, and is the strongest risk factor (genetic or otherwise) for the development of either sporadic or familial IPF (5). The association between rs35705950 and IPF has since been replicated in 10 independent cohorts (6–12), including a Mexican cohort (13) and two Asian cohorts (14, 15). However, the penetrance of the *MUC5B* promoter variant is low, with up to 20% of non-Hispanic White subjects having at least one copy of the variant and IPF occurring infrequently. This suggests that factors, in addition to the *MUC5B* promoter variant, are involved in the development of IPF.

Previous work by our laboratory has determined that the *MUC5B* promoter variant is associated with a significant increase in *MUC5B* expression in nonfibrotic lung (5). Furthermore, the rs35705950 minor allele is sufficient to enhance promoter activity *in vitro* (16). Despite the significance of these observations, molecular mechanisms remain unclear. The *MUC5B* promoter variant, rs35705950, is a G-to-T transversion that occurs in an area of the *MUC5B* 5' flanking region with elements that are suggestive of a regulatory region (see Figure E1 in the online supplement). Public databases and published work

(Received in original form February 6, 2017; accepted in final form March 6, 2017)

Author Contributions: B.A.H., I.V.Y., and D.A.S. conceived the study; B.A.H., A.N.G., V.K., S.K.S., I.V.Y., and D.A.S. designed the experiments; B.A.H. performed the experiments; B.A.H., A.N.G., S.K.S., and D.A.S. wrote and edited the manuscript; B.A.H., A.N.G., S.K.S., B.S.P., and C.M.E. analyzed the data; L.S. and T.O. performed replication studies; Y.N. created the reported construct; D.A.S. evaluated the results of the experiments.

Correspondence and requests for reprints should be addressed to David A. Schwartz, M.D., Department of Medicine, University of Colorado, School of Medicine, 12631 E. 17th Avenue, Room 8607, MS B178, Aurora, CO 80045. E-mail: david.schwartz@ucdenver.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Cell Mol Biol Vol 57, Iss 1, pp 91–99, Jul 2017

Copyright © 2017 by the American Thoracic Society

Originally Published in Press as DOI: 10.1165/rcmb.2017-0046OC on March 8, 2017

Internet address: www.atsjournals.org

identify evidence for enhancer activity, transcription factor binding, and epigenetic regulation. For example, the area surrounding the *MUC5B* promoter variant has additional characteristics of an enhancer, including H3K27 acetylation histone marks, a DNase I hypersensitivity site marking open chromatin, and occupancy of a histone modifier (P300) (17–20). In addition, motif mapping software suggests that the *MUC5B* promoter variant, rs35705950, also disrupts an E2F binding site and creates both HOX9 (homeobox A9 [HOXA9]-Meis homeobox 1 [Meis1] heterodimer) and paired box 2 (PAX2) sites, and 19 transcription factors have Encyclopedia of DNA Elements (ENCODE)-reported chromatin immunoprecipitation sequencing (ChIP-seq) peaks across rs35705950 (Figure E1). Furthermore, in a gastric cancer cell line, global DNA demethylation through 5-aza-2'-deoxycytidine treatment enhances expression of *MUC5B* (21), suggesting a potential role for DNA methylation in *MUC5B* regulation. Indeed rs35705950 is in a CpG island (22), and rs35705950 itself disrupts a CpG motif.

In the present study, our goal is to further elucidate the mechanisms involved in regulation of *MUC5B* expression in IPF. Based on this evidence of *cis*-acting genetic and epigenetic regulator elements in the *MUC5B* promoter, which are strongly associated with rs3570590, we chose to focus on DNA methylation and transcription factor binding in the *MUC5B* regulatory region. This focus is directly related to critical disease associations in humans. There are established risk factors for IPF that are known to alter the methylome, such as cigarette smoking (23–25), asbestos exposure (25, 26), male sex (27), and aging (25, 27). Notably, cigarette smoke is associated with dynamic changes in mucin expression, including induction of *MUC5B* (28). Moreover, DNA methylation patterns in lung tissue are known to be modified in IPF (4). Given these observations, we hypothesize that DNA methylation and transcription factor binding contribute to *MUC5B* regulation, and could consequently contribute to the association between the single nucleotide polymorphism (SNP) and disease risk.

Materials and Methods

Samples

Genotyping, expression, and methylation.

Human lung tissue-derived RNA and DNA were obtained from the NHLBI–Lung

Tissue Research Consortium and National Jewish Health–Interstitial Lung Disease Program, including IPF (n = 203) and nonfibrotic control (n = 139) samples. Control tissue was obtained from transplant specimens that failed regional lung selection (National Jewish Health) or biopsies of normal tissue surrounding peripheral lung lesions (Lung Tissue Research Consortium). The diagnosis of

IPF was determined by a pathology core consisting of two pulmonary pathologists, a radiology core consisting of three pulmonary radiologists, and a clinical core consisting of five pulmonary physicians. All diagnoses were made in accordance with established criteria (29, 30). Sample pool for methylation analysis is summarized in Table E1.

Primary cell culture. Normal human tracheobronchial epithelial cells were

Figure 1. The mucin 5B (*MUC5B*) promoter variant, rs35705950, and idiopathic pulmonary fibrosis (IPF) contribute to the expression of *MUC5B* in the lung. (A) Sample data: samples are divided into control and IPF, and then further divided by the rs35705950 genotype. Sample number is indicated by raw counts and percent of population, sex is represented as percent male, and average age is the mean age ($\pm$ SD). Smoking data are represented as Ever (those who no longer smoke, but have a history of smoking, and current smokers) and Never (those who report no history of smoking). (B) Lung *MUC5B* expression levels are plotted with each point representing an IPF (red) or a control (black) sample. Samples are further separated by genotype at the rs35705950 locus, where G is the wild-type allele and T is the common variant. Relative expression of *MUC5B* on the y axis is calculated by subtracting the change in cycle threshold between *MUC5B* and *GAPDH* by 20 (the highest reading). Data are presented as mean ($\pm$ SEM). This process is used to invert the values for graphical purposes. Statistical analyses (two-tailed *t* tests) were conducted on noninverted values. ***P* < 0.0001.

obtained from the tracheas of deidentified organ donors whose lungs did not meet the criteria for transplantation (31, 32). The Institutional Review Board at National Jewish Health approved the collection and the use of these cells.

Genotyping and Expression

Whole lung (DNeasy; Qiagen, Valencia, CA) and peripheral blood DNA (PAXgene blood kits; Qiagen) were used to genotype each individual for rs35705950 using a prevalidated assay (5). Total RNA isolated from lung tissue from the cases and control subjects was reverse transcribed using SuperScript III reverse transcription kits (Invitrogen, Grand Island, NY) according to the manufacturer's protocol. *MUC5B* expression levels were determined using a prevalidated *MUC5B* (Hs00861588_m1) Taqman gene expression assay (Applied Biosystems, Foster City, CA) and normalized to the prevalidated *GAPDH* (Hs99999905_g1) assay (5). Relative gene expression data were determined and analyzed using the $\Delta\Delta$ cycle threshold method (33).

DNA Methylation

CpG-level resolution of DNA methylation was determined using a custom EpiTyper panel by Agena (San Diego, CA; formally Sequenom). Detailed methods are available in supplemental MATERIALS AND METHODS.

ChIP

ChIP was conducted in two cancer cell lines (A549 and LC-2/ad) and two primary human tracheal cells grown at air-liquid interface (ALI). Details for ChIP and quantitative PCR (qPCR) are available in the supplemental MATERIALS AND METHODS.

Reporter Assay

The *MUC5B* promoter reporter construct (16) was mutated to disrupt the conserved forkhead box protein A2 (FOXA2) binding motif using the Agilent Quikchange II Site-Directed Mutagenesis kit (Santa Clara, CA). Cancer cells (A549 and H358) were transfected with the reporter plasmid, SV40-renilla, and mCherry at a ratio of 10:1:1 using Lipofectamine 2000 (Thermo Fisher Scientific, Waltham, MA). After a 24-hour incubation, luciferase activity was assayed using the Dual Luciferase Reporter assay (Promega, Chicago, IL), and luciferin levels were normalized to renilla levels. Each assay was replicated with similar results.

Short Interfering RNA Knockdown

Dharmafect smartpools were used to knock down the tested transcription factors. Suggested transfection reagent, concentrations, and volumes were used for A549s (96-well format, 0.2 μ l Dharmafect 1, and 0.5 μ l of 5 μ M short interfering RNA (siRNA) per well).

Results

rs35705950 and IPF Contribute to the Expression of *MUC5B* in the Lung

To explore the relationships between rs35705950, IPF, and *MUC5B* expression, we performed genotyping and expression analysis in a study population of 203 IPF and 139 control samples. Congruent with prior publications (5–12, 14, 15, 34), the T allele at rs35705950 was present in 51% of patients with IPF and only 23% of control subjects (odds ratio = 3.4; 95% confidence interval [CI] = 2.1–5.5; $P = 1.8 \times 10^{-8}$; Figure 1A). As previously shown, rs35705950 was associated with increased *MUC5B* gene expression in control subjects (fold change = 34.1; 95% CI = 7.1–162.7; $P = 9.8 \times 10^{-6}$). In addition, we show a novel correlation between rs35705950 and *MUC5B* gene expression in subjects with IPF (fold change = 5.7; 95% CI = 2.3–12.6; $P = 6.5 \times 10^{-5}$; Figure 1B). *MUC5B* transcripts were also elevated in subjects with IPF compared with control subjects even after controlling for the variant (fold change = 13.0; 95% CI = 6.1–27.9 fold; $P = 1.94 \times 10^{-10}$). Taken together, these findings provide further support for the role of the *MUC5B* promoter variant, rs35705950, and *MUC5B* expression in lung tissue, and also indicate that factors other than rs35705950 are involved in the transcriptional regulation of *MUC5B* in patients with IPF.

Methylation of CpG Motifs Adjacent to the *MUC5B* Promoter Variant Alter the Expression of *MUC5B* in the IPF Lung

Methylation of the region 4 kb upstream of the *MUC5B* transcription start site (TSS) was assayed using a mass spectrometry-based method, which allowed for resolution of methylation at individual CpG motifs. Although *MUC5B* promoter methylation has been previously assayed, the depth and range of detection were limited (21) and methylation of the region 4 kb upstream of the *MUC5B* TSS has not

Figure 2. Methylation analysis of the *MUC5B* promoter (A). Disease-associated differentially methylated region (DMR). Linear regression modeling of IPF with DNA methylation and covariates (age, sex, smoking status, and genotype at rs35705950) with multiple testing and regional interaction correction. (B) *MUC5B* expression-associated DMR: linear regression modeling of *MUC5B* expression and DNA methylation with multiple testing and regional interaction correction. (C) rs35705950-associated DMR: linear regression modeling of the common *MUC5B* promoter variant with DNA methylation with multiple testing and regional interaction correction. The horizontal line indicates a stringent multiple testing correction threshold for significance.

been previously evaluated in lung tissue from subjects with IPF. In this analysis, we focused on the association between DNA methylation and rs35705950, because the risk allele disrupts a methylation site. We also focused on the association between DNA methylation and *MUC5B* expression and disease, given that *MUC5B* expression is dysregulated in IPF.

IPF was found to be associated with two differentially methylated regions (DMRs) in the region 4 kb upstream of the *MUC5B* TSS (Figure 2A). The 5'-most disease-associated DMR consists of 25 CpG motifs that are hypermethylated in the IPF lung, and includes the *MUC5B* promoter variant, rs35705950 (Chr11: 1241086–1241658; $P = 8.8 \times 10^{-59}$). The smaller disease-associated DMR consists of four CpG motifs that are hypomethylated in IPF compared with control subjects (Chr11:1244003–1244077; $P = 2.5 \times 10^{-7}$) and is immediately upstream of the *MUC5B* TSS. Both disease-associated DMRs are within areas of open chromatin, and have P300 and H3K27 acetylation ChIP-seq marks (Figure E1), which suggests potential enhancer activity.

Irrespective of disease state, *MUC5B* expression was found to be associated with a DMR that contains 11 CpG motifs that are hypermethylated (Chr11: 1241139–1241412; $P = 4.19 \times 10^{-6}$) in association with increased expression of *MUC5B*, and contains the *MUC5B* promoter variant, rs35705950 (Figure 2B). Thus, this noncanonical association between expression and methylation is likely explained either by inhibition of repressor binding or binding of a transcription factor with higher affinity to methylated DNA (35, 36).

Given that the minor allele at rs35705950 disrupts a CpG motif, we investigated the association between the variant and DNA methylation. The *MUC5B* promoter variant, rs35705950, is associated with a single DMR containing 34 CpG motifs and includes the promoter variant (Chr11: 1240989–1241950; $P = 1.08 \times 10^{-39}$; Figure 2C). When this association is analyzed in nonfibrotic control subjects alone, four DMRs are associated with the polymorphism (Chr11:1240712–1241950, $P = 2.13 \times 10^{-213}$; Chr11:1242061–1242473, $P = 2.69 \times 10^{-14}$; Chr11:1242702–1243421, $P = 3.67 \times 10^{-77}$; Chr11:1243606–1243691, $P = 6.14 \times 10^{-4}$). In subjects with IPF, the associated region is more specific and identifies a single DMR (Chr11:1241559–1241849, $P = 2.82 \times 10^{-14}$; Figure E2). This finding suggests that the *MUC5B* promoter variant itself may differentially affect methylation patterns in the *MUC5B* promoter in subjects with IPF.

When the DMRs are aligned, an overlapping DMR is evident that includes 11

CpG motifs that are associated with IPF, *MUC5B* expression, and the *MUC5B* promoter variant. This region is hypermethylated in association with IPF, increased expression of *MUC5B* in lung tissue, and the rs35705950 risk allele (T). These findings define a regulatory region and support a role for DNA hypermethylation in the regulation of *MUC5B* expression in the lung and as a risk factor for the development of IPF.

Identification of a Critical Regulatory Domain for *MUC5B*

Having identified a region of differential methylation associated with IPF, *MUC5B* expression, and the *MUC5B* promoter variant, we next applied conservation analysis as a tool to identify potential regulatory sites within this region. In the 1,000 Genomes Project, only three variants were identified within the overlapping DMR, and these variants were found in less

Figure 3. Forkhead box protein A2 (FOXA2) binding site is critical for promoter activity. (A) The FOXA2 binding motif in the *MUC5B* promoter construct was disrupted using site-directed mutagenesis. A “CT” insert, “AAA” → “C” in/del and a “C” → “T” point mutation were made. (B) The *MUC5B* wild-type (WT) promoter construct clones (blue; pGL4.10 with *MUC5B* promoter insert) show that the 4.2-kb region directly upstream of the *MUC5B* transcriptional start site has increased promoter activity relative to the empty vector (gray; pGL4.10 with no insertions). After site-directed mutagenesis of the conserved FOXA2 motif in the WT reporter, the FOXA2-mutant reporter clones (yellow) show a significant decrease in promoter activity (H358 data shown; $P = 1.7 \times 10^{-8}$). Two-tailed *t* tests were used for all analyses. Each bar represents an individual clone, with the error bars indicating SEM for technical replicates. EV, empty vector; In/del, dual insertion and deletion.

than 1% of the population (Chr11:1241281, 0.137%; Chr11:1241329, 0.137%; Chr11:1241361, 0.458%). Cross-species analysis indicated that the wild-type G allele at rs35705950 is the major allele in primates, including gorilla and gibbon (Figure E3). Collectively, these data suggest that this region of the *MUC5B* promoter is highly conserved in mammals, and that the bases encoded at the site where the rs35705950 variant occurs in humans represent the 5' border of a region of an unusually strong area of conservation, suggesting that the region contains a conserved element that serves critical gene regulatory functions.

To this end, 32 bp downstream of rs35705950 is a highly conserved region (Chr11:1241262–1241269) containing a FOXA2 binding site (Chr11:1241254–1241270) that is also conserved across mammals ranging from primates to rodents, and includes both terrestrial and aquatic groups (Figure E3). In addition, there are no reported mutations in the FOXA2 binding site (Chr11:1241254–1241270) in the 1,000 Genomes Project or our sample population, further suggesting the evolutionary importance of this region. Based on this conservation across and within species, we postulated that FOXA2 binding may play a role in *MUC5B* gene regulation.

To determine whether the FOXA2 binding site contributes to transcriptional regulation of *MUC5B*, we disrupted the conserved FOXA2 motif within a *MUC5B* promoter reporter that we have described previously (16) (Figure E4). Congruent with our prior findings, the parent construct with an intact FOXA2 motif generates significant luciferase activity (16). However, when the conserved FOXA2 site is disrupted, luciferase activity is markedly decreased (H358 cells, $P = 1.7 \times 10^{-8}$; A549 cells, $P = 5.9 \times 10^{-9}$; Figure 3B), suggesting a critical role for the FOXA2 site in regulating *MUC5B* expression and further defining the critical regulatory domain of *MUC5B*. These data were further replicated in an additional cell line (Figure E5).

FOXA2 Binds Dynamically to the Critical Regulatory Domain of *MUC5B*

FOXA2 binding at the predicted motif (Figure 4A) was assayed using ChIP-qPCR in the pulmonary adenocarcinoma cell line, LC-2/ad, which expresses high levels of *MUC5B* (Table E3). In LC-2/ad cells,

Figure 4. Chromatin immunoprecipitation (ChIP)-quantitative PCR (qPCR) occupancy of FOXA2 and RNA polymerase II (A). The two regions targeted by qPCR were at the predicted enhancer site (red) and the *MUC5B* transcription start site (TSS; blue). The arrows at each site represent the locations of the qPCR primers. In addition, positive control primers were designed at the ENCODE FOXA2 peaks with the highest occupancy (in green in B) and at the transcriptional start site of a transcriptionally active gene for Pol2 (green in C). (B) FOXA2 shows high occupancy at the predicted enhancer with a 201-fold enrichment ($P = 1.6 \times 10^{-9}$) over the negative controls (gray). (C) RNA polymerase II (Pol2) occupancy is enriched for at the TSS (15-fold, $P = 0.015$), indicating that *MUC5B* is being actively transcribed. In addition, Pol2 is enriched for at the enhancer, increasing the likelihood that this is an active enhancer motif (14-fold, $P = 0.0009$). Two-tailed *t* tests were used for all analyses with results shown as mean ($\pm$ SEM).

FOXA2 binds in the enhancer region with a 201-fold enrichment relative to negative control regions ($P = 1.6 \times 10^{-9}$; Figure 4B). Enrichment of RNA polymerase II binding

was assayed at the TSS to verify the transcriptional activity of *MUC5B* and at the enhancer as a predictive measure of enhancer activity (37). RNA polymerase II

showed a 15-fold enrichment ($P = 0.02$) at the TSS and a 14-fold enrichment of occupancy ($P = 0.0009$) at the enhancer (Figure 4C). In aggregate, these findings indicate that *MUC5B* is actively transcribed in LC-2/ad cells, and also suggest that the *MUC5B* promoter variant, rs35705950, is located within an active enhancer. Similar results were observed in an additional lung cancer cell line, A549 (Figure E6), which expresses lower levels of *MUC5B* transcript.

Primary human tracheal cells do not produce *MUC5B* in submerged conditions, but, when cultured at ALI, they differentiate into mucosal and ciliated cells that produce high concentrations of *MUC5B* (38). This provides a dynamic model for studying transcriptional control of *MUC5B* expression. Primary human tracheal cells were grown to confluence and either harvested in a submerged state or moved to ALI and then harvested for ChIP-qPCR. ALI culture is associated with increased transcription complex assembly in comparison to submerged cells (Figure E7). We also measured occupancy of FOXA2 at the 4-kB enhancer in both the submerged and ALI culture conditions, and found that ALI cultures show robust occupancy of FOXA2 at the enhancer site (Figure 5A; 62-fold, $P = 5.1 \times 10^{-12}$). Interestingly, the submerged cultures also have binding of FOXA2 at the enhancer site (Figure 5B; 12-fold, $P = 1.3 \times 10^{-9}$); however, the ALI cultures have a 5.1-fold higher occupancy ($P = 1.9 \times 10^{-8}$) compared with the submerged cultures (Figure 5A), indicating that binding of FOXA2 is dynamic in association with *MUC5B* expression. Furthermore, binding of FOXA2 in the submerged cultures may suggest that other components of the transcriptional complex are necessary for the initiation of transcription, thus making FOXA2 a necessary, but not sufficient, initiator of the transcriptional complex.

FOXA2 Affects *MUC5B* Expression

There are 24 transcription factors that have either conserved binding motifs or have validated binding through ChIP-SEquation (20) within the critical regulatory domain of *MUC5B* (Figure E1). To determine whether FOXA2 plays a role in the regulation of *MUC5B* and to identify other components of the transcriptional complex, a comprehensive siRNA knockdown approach of the 24 transcription factors was completed. Of the tested transcription factors, four had a significant effect on

MUC5B expression. FOXA2 (40.1-fold decrease; 95% CI = 6.4–53.2; $P = 3.2 \times 10^{-4}$) had the strongest effect on *MUC5B* expression (Figure 6). Knockdown (KD) of signal transducer and activator of transcription 3 (STAT3) (24.2-fold decrease; 95% CI = 4.0–32.1; $P = 9.4 \times 10^{-4}$), HOXA9 (26.5-fold decrease; 95% CI = 4.2–35.3; $P = 8.7 \times 10^{-4}$), and zinc finger and BTB domain containing 7A (ZBTB7A) (17.7-fold decrease; 95% CI = 2.9–23.4; $P =$

3.3×10^{-7}) also decreased *MUC5B* expression. Interestingly, the T allele at rs35705950 creates a *de novo* putative HOXA9-MEIS1 binding site, and STAT3 has been shown to regulate FOXA2 (39). The FOXA2 siRNA KD was replicated in three additional cell lines, and, in each case, a significant decrease in *MUC5B* expression was observed (H358, $P = 0.039$, 1.3-fold; LC-2/ad, $P = 0.0084$, 1.2-fold; H2122, $P = 0.041$, 1.3-fold).

Figure 5. Primary tracheal cell FOXA2 ChIP-qPCR binding of FOXA2 to the enhancer was assayed in undifferentiated submerged cultures (B) and differentiated air-liquid interface (ALI) cultures (A). There is increased occupancy of FOXA2 in both the submerged (62-fold, $P = 5.1 \times 10^{-12}$) and ALI cultures (12-fold, $P = 1.3 \times 10^{-9}$). The relative occupancy of FOXA2 at the enhancer was increased by 5.1-fold ($P = 1.9 \times 10^{-8}$) in the *MUC5B*-expressing ALI cultures compared with the nonexpressing submerged cultures. The red bars represent the predicted enhancer site (see Figure 4A), the green bars represent positive control primers, and the gray bars represent negative control primers. Two-tailed *t* tests were used for all analyses with results shown as mean ($\pm$ SEM).

Figure 6. Effects of short interfering RNA (siRNA) knockdown of transcription factors on *MUC5B* expression. A total of 24 transcription factors predicted to have binding near rs35705950 were knocked down using siRNA, and the effect on *MUC5B* expression was measured using qPCR. Fold change in *MUC5B* expression was calculated using the $\Delta\Delta CT$ method, and significance was calculated using a standard *t* test with a corrected *P* value of significance of 0.002. Four transcription factors significantly down-regulated *MUC5B* expression: FOXA2, HOXA9, STAT3, and ZBTB7A ($P < 0.002$). CEBPB, CCAAT/enhancer binding protein β ; E2F1, E2F transcription factor 1; ELF-1, E74 like ETS transcription factor 1; EP300, E1A binding protein p300; ESR1, estrogen receptor α ; HDAC2, histone deacetylase 2; HNF4, hepatocyte nuclear factor 4; HOXA9, homeobox A9; JUND, JunD proto-oncogene, AP-1 transcription factor subunit; NFIC, nuclear factor I C; PAX, paired box family of transcription factors; REST, RE1 silencing transcription factor; RXRA, retinoid X receptor α ; SP1, Sp1 transcription factor; STAT3, signal transducer and activator of transcription 3; TCF12, transcription factor 12; TEAD4, TEA domain transcription factor 4; ZBTB, transcription repressor. Results are shown as the mean of the replicates ($\pm$ SEM).

Discussion

We have identified a highly conserved *MUC5B* enhancer element that contains the well known *MUC5B* genetic variant, rs35705950. Here, we show that this enhancer is composed of an overlapping DMR that is associated with expression, IPF, and the presence of the *MUC5B* promoter variant. This enhancer region in the *MUC5B* promoter dynamically binds a pioneering transcription factor, FOXA2, as well as RNA polymerase II, together strongly suggestive of an important regulatory function in controlling *MUC5B* expression. Identification of FOXA2 as a necessary component of the *MUC5B* transcriptional complex and isolation of the relevant binding regions provide us with the framework both to modify the expression of *MUC5B* and identify additional factors that contribute to regulation of *MUC5B*.

DNA methylation remains of particular interest in IPF, given that many of the known risk factors of IPF—male sex, increased age, cigarette smoke exposure—are also known to cause changes to the methylome. Here, we indicate that

increased methylation is associated with IPF, the risk allele T at rs35705950, and increased expression of *MUC5B*.

Interestingly, although this positive correlation between expression and methylation was once considered paradoxical, more recent studies have indicated that the canonical negative correlation is mostly observed in promoters, whereas enhancers show increased enrichment for positive correlations. This is commonly explained by either increased methylation inhibiting the binding of repressive complexes or binding of a subset of transcription factors that bind with higher affinity to methylated DNA, such as CCCTC-binding factor (CTCF) (40).

Although the previously identified *MUC5B* promoter variant, rs35705950, is 32 bp upstream from the FOXA2 binding motif, it is likely that the variant and the binding site function together within a composite regulatory element. There are two probable hypotheses for the interaction between these two sites. First, the *MUC5B* promoter variant is associated with regional changes in methylation surrounding the FOXA2 binding motif, and this could directly affect FOXA2 occupancy. In this case, we would hypothesize that

increased methylation of the surrounding area leads to increased binding of FOXA2. Alternatively, the *MUC5B* promoter variant may alter the binding affinity/specificity for other transcription factors that interact with or are recruited by FOXA2 to regulate *MUC5B* expression. In that regard, our data indicate that knockdown of HOXA9, which is predicted to preferentially bind the promoter variant, leads to decreased expression of *MUC5B*. Intriguingly, the HOX and FOX family of proteins are already known to interact in gene regulation (41, 42). Moreover, activation of STAT3 is associated with down-regulation of FOXA2 (39), and, here, we show that knockdown of STAT3 reduces the expression of *MUC5B*, suggesting that STAT3 may be involved upstream in *MUC5B* regulation. In addition, FOXA2 is well known to play a key role in early lung development, lung homeostasis (43), surfactant expression, and goblet cell metaplasia (44), in association with transcription factors, such as GATA binding protein 6 (GATA6) and transcription termination factor 1 (TTF-1). Whether the regulation of *MUC5B* during disease pathogenesis recapitulates interaction between these transcription factor families during development remains to be determined.

Although age, sex, smoke exposure, and mucin dysregulation are commonly associated with a number of pulmonary diseases (chronic obstructive pulmonary disease, asthma, cystic fibrosis, etc.), the location of mucus production may have a profound effect on the consequent pathophysiological phenotype (45). For instance, in contrast to IPF, where the excess mucus production is localized to the bronchiole (16), cystic fibrosis (46, 47) and chronic bronchitis (48) are associated with enhanced mucus secretion in more proximal airways and submucosal glands, and consequently these diseases are airway centric. However, too little Muc5b also appears to be detrimental, as we have reported that Muc5b-deficient mice have chronic mucociliary clearance impairment, persistent lung injury and inflammation, and die from bacterial pneumonia (49). In aggregate, our findings suggest that too little or too much Muc5b leads to different types of lung injury, and that the regulatory components and post-transcriptional processing of Muc5b are finely tuned in the normal lung to

maintain host defense in response to a dynamic and potentially detrimental environment. In addition, we find that the clinical manifestations may, in part, be affected by the location of the defect in mucociliary function.

Defining the *MUC5B* enhancer and identifying FOXA2 as a critical regulator establishes a rationale for pharmacologic targeting of FOXA2 to reduce the expression of *MUC5B* in IPF. Our work also provides a framework to define

additional factors, likely including HOXA9, STAT3, and ZBTB7A that increase *MUC5B* expression, which is now clearly implicated in IPF pathogenesis, both in association with and independent of the *MUC5B* promoter variant. Our findings add to the growing body of evidence linking noncoding variants to regulation of gene expression in lung disease, and highlight the importance of integrating genetic and epigenetic variation to uncover novel pathways that are disrupted in lung disease.

Understanding the biological mechanisms underlying the association between genome-wide association study variants and disease is an important next step in the elucidation of critical pathways, identification of targetable components within those pathways, and, ultimately, creation of treatments and cures for a wide range of complex diseases. ■

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

References

- Korfei M, Schmitt S, Ruppert C, Henneke I, Markart P, Loeh B, Mahavadi P, Wygrecka M, Klepetko W, Fink L, *et al.* Comparative proteomic analysis of lung tissue from patients with idiopathic pulmonary fibrosis (IPF) and lung transplant donor lungs. *J Proteome Res* 2011;10:2185–2205.
- Korfei M, Skwarna S, Henneke I, MacKenzie B, Klymenko O, Saito S, Ruppert C, von der Beck D, Mahavadi P, Klepetko W, *et al.* Aberrant expression and activity of histone deacetylases in sporadic idiopathic pulmonary fibrosis. *Thorax* 2015;70:1022–1032.
- Sanders YY, Ambalavanan N, Halloran B, Zhang X, Liu H, Crossman DK, Bray M, Zhang K, Thannickal VJ, Hagood JS. Altered DNA methylation profile in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2012;186:525–535.
- Yang IV, Pedersen BS, Rabinovich E, Hennessy CE, Davidson EJ, Murphy E, Guardela BJ, Tedrow JR, Zhang Y, Singh MK, *et al.* Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014;190:1263–1272.
- Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, Fingerlin TE, Zhang W, Gudmundsson G, Groshong SD, *et al.* A common *MUC5B* promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364:1503–1512.
- Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, Loyd JE, Cosgrove GP, Lynch D, Groshong S, *et al.* Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013;45:613–620.
- Zhang Y, Noth I, Garcia JG, Kaminski N. A variant in the promoter of *MUC5B* and idiopathic pulmonary fibrosis. *N Engl J Med* 2011;364:1576–1577.
- Stock CJ, Sato H, Fonseca C, Banya WA, Molyneaux PL, Adamali H, Russell AM, Denton CP, Abraham DJ, Hansell DM, *et al.* Mucin 5B promoter polymorphism is associated with idiopathic pulmonary fibrosis but not with development of lung fibrosis in systemic sclerosis or sarcoidosis. *Thorax* 2013;68:436–441.
- Noth I, Zhang Y, Ma S-F, Flores C, Barber M, Huang Y, Broderick SM, Wade MS, Hysi P, Scuirba J, *et al.* Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013;1:309–317.
- Borie R, Crestani B, Dieude P, Nunes H, Allanore Y, Kannengiesser C, Airo P, Matucci-Cerinic M, Wallaert B, Israel-Biet D, *et al.* The *MUC5B* variant is associated with idiopathic pulmonary fibrosis but not with systemic sclerosis interstitial lung disease in the European Caucasian population. *PLoS One* 2013;8:e70621.
- Coghlan MA, Shifren A, Huang HJ, Russell TD, Mitra RD, Zhang Q, Wegner DJ, Cole FS, Hamvas A. Sequencing of idiopathic pulmonary fibrosis-related genes reveals independent single gene associations. *BMJ Open Respir Res* 2014;1:e000057.
- Lee MG, Lee YH. A meta-analysis examining the association between the *MUC5B* rs35705950 T/G polymorphism and susceptibility to idiopathic pulmonary fibrosis. *Inflamm Res* 2015; 64:463–470.
- Peljto AL, Zhang Y, Fingerlin TE, Ma SF, Garcia JG, Richards TJ, Silveira LJ, Lindell KO, Steele MP, Loyd JE, *et al.* Association between the *MUC5B* promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA* 2013;309: 2232–2239.
- Wang C, Zhuang Y, Guo W, Cao L, Zhang H, Xu L, Fan Y, Zhang D, Wang Y. Mucin 5B promoter polymorphism is associated with susceptibility to interstitial lung diseases in Chinese males. *PLoS One* 2014;9:e104919.
- Horimasu Y, Ohshimo S, Bonella F, Tanaka S, Ishikawa N, Hattori N, Kohno N, Guzman J, Costabel U. *MUC5B* promoter polymorphism in Japanese patients with idiopathic pulmonary fibrosis. *Respirology* 2015;20:439–444.
- Nakano Y, Yang IV, Walts AD, Watson AM, Helling BA, Fletcher AA, Lara AR, Schwarz MI, Evans CM, Schwartz DA. *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.
- Pott S, Lieb JD. What are super-enhancers? *Nat Genet* 2015;47:8–12.
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, *et al.* The accessible chromatin landscape of the human genome. *Nature* 2012;489:75–82.
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 2009;459:108–112.
- Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, *et al.* ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* 2013;41:D56–D63.
- Vincent A, Perrais M, Desseyn JL, Aubert JP, Pigny P, Van Seuning I. Epigenetic regulation (DNA methylation, histone modifications) of the 11p15 mucin genes (*MUC2*, *MUC5AC*, *MUC5B*, *MUC6*) in epithelial cancer cells. *Oncogene* 2007;26:6566–6576.
- Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci USA* 2006;103: 1412–1417.
- Hubbard R, Johnston I, Britton J. Survival in patients with cryptogenic fibrosing alveolitis: a population-based cohort study. *Chest* 1998; 113:396–400.
- Lyn-Cook L, Word B, George N, Lyn-Cook B, Hammons G. Effect of cigarette smoke condensate on gene promoter methylation in human lung cells. *Tob Induc Dis* 2014;12:15.
- Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 2009; 5:e1000602.
- Tsou JA, Galler JS, Wali A, Ye W, Siegmund KD, Groshen S, Laird PW, Turla S, Koss MN, Pass HI, *et al.* DNA methylation profile of 28 potential marker loci in malignant mesothelioma. *Lung Cancer* 2007; 58:220–230.

27. Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, Kahn RS, Ophoff RA. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One* 2009;4:e6767.
28. Preciado D, Lin J, Wuertz B, Rose M. Cigarette smoke activates NF κ B and induces *Muc5b* expression in mouse middle ear cells. *Laryngoscope* 2008;118:464–471.
29. Travis WD, Hunninghake G, King TE Jr, Lynch DA, Colby TV, Galvin JR, Brown KK, Chung MP, Cordier JF, du Bois RM, *et al.* Idiopathic nonspecific interstitial pneumonia: report of an American Thoracic Society project. *Am J Respir Crit Care Med* 2008;177:1338–1347.
30. Travis WD, Costabel U, Hansell DM, King TE Jr, Lynch DA, Nicholson AG, Ryerson CJ, Ryu JH, Selman M, Wells AU, *et al.*; ATS/ERS Committee on Idiopathic Interstitial Pneumonias. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2013;188:733–748.
31. Wu Q, Jiang D, Smith S, Thaikoottathil J, Martin RJ, Bowler RP, Chuw HW IL-13 dampens human airway epithelial innate immunity through induction of IL-1 receptor-associated kinase M. *J Allergy Clin Immunol* 2012;129:825–833.e2.
32. Berman R, Huang C, Jiang D, Finigan JH, Wu Q, Chu HW. MUC18 differentially regulates pro-inflammatory and anti-viral responses in human airway epithelial cells. *J Clin Cell Immunol* 2014;5:257.
33. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* 2001;25:402–408.
34. Peljto AL, Selman M, Kim DS, Murphy E, Tucker L, Pardo A, Lee JS, Ji W, Schwarz MI, Yang IV, *et al.* The MUC5B promoter polymorphism is associated with idiopathic pulmonary fibrosis in a Mexican cohort but is rare among Asian ancestries. *Chest* 2015;147:460–464.
35. Meehan RR, Lewis JD, Bird AP. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res* 1992;20:5085–5092.
36. Sarraf SA, Stancheva I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell* 2004;15:595–605.
37. Savic D, Roberts BS, Carleton JB, Partridge EC, White MA, Cohen BA, Cooper GM, Gertz J, Myers RM. Promoter-distal RNA polymerase II binding discriminates active from inactive CCAAT/ enhancer-binding protein β binding sites. *Genome Res* 2015;25:1791–1800.
38. Gray TE, Guzman K, Davis CW, Abdullah LH, Nettesheim P. Mucociliary differentiation of serially passaged normal human tracheobronchial epithelial cells. *Am J Respir Cell Mol Biol* 1996;14:104–112.
39. Hao Y, Kuang Z, Jing J, Miao J, Mei LY, Lee RJ, Kim S, Choe S, Krause DC, Lau GW. Mycoplasma pneumoniae modulates STAT3-STAT6/EGFR-FOXA2 signaling to induce overexpression of airway mucins. *Infect Immun* 2014;82:5246–5255.
40. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* 2012;22:1680–1688.
41. Pfaff SL. Developmental neuroscience: Hox and Fox. *Nature* 2008;455:295–297.
42. Brayer KJ, Lynch VJ, Wagner GP. Evolution of a derived protein–protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. *Proc Natl Acad Sci USA* 2011;108:E414–E420.
43. Wan H, Kaestner KH, Ang SL, Ikegami M, Finkelman FD, Stahlman MT, Fulkerson PC, Rothenberg ME, Whitsett JA. Foxa2 regulates alveolarization and goblet cell hyperplasia. *Development* 2004;131:953–964.
44. Sun L, Tang XJ, Luo FM. Forkhead box protein A2 and T helper type 2-mediated pulmonary inflammation. *World J Methodol* 2015;5:223–229.
45. Evans CM, Fingerlin TE, Schwarz MI, Lynch D, Kurche J, Warg L, Yang IV, Schwartz DA. Idiopathic pulmonary fibrosis: a genetic disease that involves mucociliary dysfunction of the peripheral airways. *Physiol Rev* 2016;96:1567–1591.
46. Boucher RC. Airway surface dehydration in cystic fibrosis: pathogenesis and therapy. *Annu Rev Med* 2007;58:157–170.
47. Button B, Cai LH, Ehre C, Kesimer M, Hill DB, Sheehan JK, Boucher RC, Rubinstein M. A periciliary brush promotes the lung health by separating the mucus layer from airway epithelia. *Science* 2012;337:937–941.
48. Hogg JC, Chu F, Utokeaparch S, Woods R, Elliott WM, Buzatu L, Cherniack RM, Rogers RM, Sciurba FC, Coxson HO, *et al.* The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med* 2004;350:2645–2653.
49. Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, Alexander SN, Bellinghausen LK, Song AS, Petrova YM, *et al.* Muc5b is required for airway defence. *Nature* 2014;505:412–416.