Genome-Wide Association Study: Functional Variant rs2076295 Regulates Desmoplakin Expression in Airway Epithelial Cells

Yuan Hao^{1*}, Samuel Bates^{1*}, Hongmei Mou², Jeong H. Yun¹, Betty Pham¹, Jiangyuan Liu¹, Weiliang Qiu¹, Feng Guo¹, Jarrett D. Morrow¹, Craig P. Hersh¹, Christopher J. Benway¹, Lu Gong¹, Yihan Zhang², Ivan O. Rosas³, Michael H. Cho¹, Jin-Ah Park⁴, Peter J. Castaldi¹, Fei Du¹, and Xiaobo Zhou¹

¹Channing Division of Network Medicine and ³Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; ²Mucosal Immunology and Biology Research Center, Massachusetts General Hospital, Boston, Massachusetts; and ⁴Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

ORCID IDs: 0000-0002-6624-7757 (H.M.); 0000-0002-4361-8295 (J.H.Y.); 0000-0002-1342-4334 (C.P.H.); 0000-0002-7127-2869 (X.Z.).

Abstract

Rationale: Genetic association studies have identified rs2076295 in association with idiopathic pulmonary fibrosis (IPF). We hypothesized that rs2076295 is the functional variant regulating *DSP* (*desmoplakin*) expression in human bronchial epithelial cells, and *DSP* regulates extracellular matrix-related gene expression and cell migration, which is relevant to IPF development.

Objectives: To determine whether rs2076295 regulates *DSP* expression and the function of *DSP* in airway epithelial cells.

Methods: Using CRISPR (clustered regularly interspaced short palindromic repeat)/Cas9 editing (including regional deletion, indel, CRISPR interference, and single-base editing), we modified rs2076295 and measured *DSP* expression in edited 16HBE14o- and primary airway epithelial cells. Cellular integrity, migration, and genome-wide gene expression changes were examined in 16HBE14o- single colonies with *DSP* knockout. The expression of *DSP* and its

relevant matrix genes was measured by quantitative PCR and also analyzed in single-cell RNA-sequencing data from control and IPF lungs.

Measurements and Main Results: *DSP* is expressed predominantly in bronchial and alveolar epithelial cells, with reduced expression in alveolar epithelial cells in IPF lungs. The deletion of the DNA region-spanning rs2076295 led to reduced expression of *DSP*, and the edited rs2076295GG 16HBE14o- line has lower expression of *DSP* than the rs2076295TT lines. Knockout of *DSP* in 16HBE14o- cells decreased transepithelial resistance but increased cell migration, with increased expression of extracellular matrix-related genes, including *MMP7* and *MMP9*. Silencing of *MMP7* and *MMP9* abolished increased migration in *DSP*-knockout cells.

Conclusions: rs2076295 regulates *DSP* expression in human airway epithelial cells. The loss of *DSP* enhances extracellular matrix-related gene expression and promotes cell migration, which may contribute to the pathogenesis of IPF.

Keywords: *DSP*; CRISPR/Cas9 genome-editing; bronchial epithelial cells; cell migration; rs2076295

(Received in original form October 11, 2019; accepted in final form June 18, 2020)

*These authors contributed equally to this work.

Supported by U.S. NIH grants R01HL127200, 1P01HL132825, R01HL148667 (X.Z.), and R01HL137927 and R01HL147148 (M.H.C. and X.Z.); Cystic Fibrosis Foundation grant MOU19G0 (H.M.); and a Child Health Research Award (H.M.) from the Charles H. Hood Foundation.

Author Contributions: Y.H., F.D., and X.Z. designed the study. Y.H. designed, conducted, and analyzed the biological experiment. S.B. and F.D. designed and conducted CRISPR/Cas9 genome-editing and RNA-sequencing experiments. Y.H. and S.B. performed immunofluorescence and cell migration experiment. J.H.Y. performed single-cell RNA-sequencing data analysis. B.P. performed quantitative PCR and luciferase reporter assay. H.M., Y.Z., and J.-A.P. performed air-liquid interface culture and identification. J.L., W.Q., C.J.B., and M.H.C. analyzed RNA-sequencing data. J.D.M. and C.P.H. analyzed gene expression data in the human brushing sample. L.G. performed online data mining on open chromatin status nearby rs2076295. F.G. isolated human alveolar type II epithelial cells. I.O.R. provided normal and idiopathic pulmonary fibrosis human lung tissue samples. Y.H., H.M., J.-A.P., P.J.C., F.D., and X.Z. participated in the manuscript writing.

Correspondence and requests for reprints should be addressed to Xiaobo Zhou, Ph.D., Channing Division of Network Medicine, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA 02115. E-mail: xiaobo.zhou@channing.harvard.edu.

This article has a related editorial.

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

Am J Respir Crit Care Med Vol 202, Iss 9, pp 1225–1236, Nov 1, 2020

Copyright © 2020 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201910-1958OC on June 18, 2020

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Noncoding variant rs2076295 is the most significant genetic variant at chromosome 6q24 associated with the susceptibility of idiopathic pulmonary fibrosis (IPF) in genome-wide association studies (GWAS). This variant is also associated with chronic obstructive pulmonary disease in the opposite direction. However, the causal gene and causal variant for the association in this locus are not entirely known yet.

What This Study Adds to the Field:

Using a series of CRISPR (clustered regularly interspaced short palindromic repeat)-based genome-editing approaches, we demonstrated that rs2076295 regulates *DSP* (desmoplakin) expression in airway epithelial cells, which supported *DSP* as the causal gene in this GWAS locus implicated in IPF pathogenesis. We also showed that reduced levels of *DSP*, associated with the IPF risk allele in this GWAS locus, leads to increased expression of extracellular matrix genes and promotes the migration of airway epithelial cells. Our findings represent the critical and necessary first step on the post-GWAS functional studies, preceding future in-depth investigations on how the loss of *DSP* may promote pulmonary fibrosis.

Idiopathic pulmonary fibrosis (IPF) is a chronic lung disease characterized by progressive lung scarring and fibrosis in which alveolar architecture is destroyed and replaced by aberrant extracellular matrix (ECM) and activated myofibroblasts, eventually leading to respiratory failure and death (1). IPF affects more than 3 million people worldwide, with 2–3 years of median survival after diagnosis (2, 3). The etiology of IPF is complex, attributable to microenvironmental factors in subjects with genetic predispositions (4).

Genome-wide association studies (GWAS) have identified 14 disease susceptibility loci associated with IPF in an unbiased way (5), including a region at chromosome 6p24, with rs2076295 as the

most significant variant (6–8). rs2076295 is a noncoding variant located in the intron 5 of *DSP* (*desmoplakin*) gene, encoding DSP that forms cell–cell adhesion complexes, enabling tissues to resist mechanical forces. Subsequent analysis of expression quantitative trait loci (eQTLs) identified rs2076295 as an eQTL SNP for *DSP* gene expression in human lungs (7, 9), suggesting that *DSP* is likely the causal gene in this locus.

Despite multiple reports on the association of rs2076295 with IPF susceptibility, these following questions remain unaddressed: 1) Does rs2076295 directly regulate *DSP* expression? 2) What is the directional regulation of rs2076295 on the expression of *DSP* in the lung-relevant cell type? 3) What is the molecular function of *DSP* that may determine IPF susceptibility? The identification of functional variants in each GWAS locus is the important and necessary first step to understanding the biological function of GWAS loci (10). In this study, we identified rs2076295 as a functional variant at the 6q24 *DSP* GWAS locus by showing the allele-specific expression of *DSP* in isogenic human bronchial epithelial cell line (16HBE14o- cells) edited by CRISPR (clustered regularly interspaced short palindromic repeat)/Cas9-based techniques targeting rs2076295. To determine the function of *DSP* in lung epithelial cells, we used the CRISPR/Cas9 method to generate knockout clones of *DSP* in 16HBE14o- cells that showed increased migration and reduced transepithelial electrical resistance (TEER). By RNA-sequencing (RNA-seq) in isogenic clones, we found that *DSP* knockout led to increased expression of *MMP7* (matrix metalloproteinase 7) and *MMP9*, which in turn enhanced cell migration and might contribute to the pathogenesis of IPF. Expression levels of *DSP* and other ECM genes were measured by quantitative PCR and further explored using publicly available single-cell RNA-seq data in human lungs (normal and with IPF).

Methods

Cell Line, Primary Cells, and Human Lung Tissues

Human primary cells were isolated from healthy donor lungs or purchased from Lonza (normal human bronchial epithelial [NHBE] cells; CC-2540). The human lung tissue samples were approved by the

institutional review board of Partners HealthCare. Detailed information on human IPF and control lung samples are shown in Table E1 in the online supplement.

CRISPR/Cas9 Guide RNA Delivery

Guide RNAs (gRNAs) used in CRISPR/Cas9-based methods targeting rs2076295 are summarized in Table E2. The gRNA was delivered into 16HBE14o- cells by transfection with Lipofectamine 3000 and into NHBE cells by the ribonucleoprotein method, using Amaxa 4D-Nucleofector electroporation.

Single-Cell RNA-Seq Dataset Analysis

The publicly available human single-cell RNA-seq dataset (GSE135893) (11) was reanalyzed for cell type-specific gene expression using Seurat version 3 (12).

Migration Assay

Cells were seeded in 96-well Oris Cell Migration Assay (Platypus Technologies) using Calcein-AM (Invitrogen) for live-cell fluorescence staining based on the manufacturer's instructions.

Statistical Analysis

Statistical analyses were performed using GraphPad Prism 7 and R 3.6.2. All data were tested for normal distribution with the Shapiro-Wilk normality test before we used an unpaired Student's *t* test or the Mann-Whitney test for comparisons. Details on statistical analyses are described in the figure legends and the online supplement.

Detailed methods are described in online supplement.

Results

rs2076295 Regulates the Expression of *DSP*

Regulatory regions near functional SNPs are typically characterized as open chromatin regions harboring DNase hypersensitivity peaks or assay for transposase-accessible chromatin using sequencing (ATAC-Seq) peak. Indeed, based on data from the Roadmap Epigenomics Project and ENCODE Project (13–16), strong DNase hypersensitivity signals were found near rs2076295 in bronchial epithelial cells, in contrast to lung endothelial cells and lung fibroblasts (Figure E1A). Consistently, in our assay for transposase-accessible chromatin sequencing analysis, we detected

Figure 1. Genetic variant rs2076295 regulates *DSP* (desmoplakin) expression. (A) Expression of *DSP* in human bronchial epithelial cells collected from brushing with various genotypes of rs2076295. $n = 7$ – 11 subjects/genotype group, as indicated by numbers inside each column. (B) Expression of *DSP* in edited 16HBE14o- cells targeting rs2076295 for regional deletion by CRISPR/Cas9. Means $\pm$ SEM are from six biological replicates. (C) Upper: Sequencing results in empty vector-transfected wild-type (WT) and three clonal cell lines with CRISPR (clustered regularly interspaced short palindromic repeat) (CRISPR)/Cas9-edited indel mutations targeting rs2076295. The inserted DNA bases are displayed in red, with the number of deleted (–) or inserted (+) bases for each clone shown on the right. Lower: Measurements on the expression of *DSP* in three indel single clones of 16HBE14o- cells and

open chromatin peaks near exon 5 of *DSP* around rs2076295 (Figure E1B) in both human primary airway epithelial cells and alveolar type II epithelial cells isolated from healthy donors. These results suggested human bronchial epithelial cells as the relevant cell type for studying the function of rs2076295 in the *DSP* GWAS region.

In human primary bronchial epithelial cells (NHBE) obtained by bronchoscopy brushing (17), the samples with GG homozygous genotype at rs2076295 exhibited a lower expression of *DSP* compared with the TT homozygous samples (Figure 1A), prioritizing *DSP* as a potential candidate gene associated with rs2076295.

Given the pervasive nature of eQTLs, many SNPs that are associated with both a disease phenotype and gene expression are not necessarily functional variants for GWAS. Nevertheless, the pattern of GWAS association in the 6q24 region (with statistical colocalization) (9) implicates SNP rs2076295 as the likely functional variant in this region. To experimentally confirm this, we performed four sets of CRISPR/Cas9-based gene editing experiments in human bronchial epithelial cells (16HBE14o- cells). First, we designed two gRNAs and generated a 55 bp deletion of nucleotide region-spanning rs2076295. This resulted in >90% deletion efficiency, as shown by Droplet Digital PCR (Figure E1C), accompanied by a 30% reduction in *DSP* expression (Figure 1B) in a mixed population of 16HBE14o- cells. Second, we designed a single gRNA targeting rs2076295 for the generation of small indels, and three single colonies were confirmed with variably sized indels near or spanning rs2076295. These resulting clonal cells exhibited significantly reduced expression of *DSP* (20–45%) (Figure 1C). Third, we applied CRISPR interference (CRISPRi) method (18), in which dCas-9 is coupled to the transcriptional repressor Krüppel-associated box to determine the regulatory effects of rs2076295 on the expression of *DSP* without chromatin cutting. Deactivating rs2076295 using the CRISPRi approach significantly

decreased the expression of *DSP* to 50% in mixed cell populations (Figure 1D). All these results support the notion that the DNA region near rs2076295 is likely an enhancer that promotes the expression of *DSP* in 16HBE14o- cells. Lastly, we designed gRNA for CRISPR/Cas9-mediated homology-directed repair targeting rs2076295 in heterozygous 16HBE14o- cells (Figure 1E). This resulted in single colonies homozygous for the TT or GG genotype at rs2076295, which was confirmed by Sanger sequencing. By quantitative PCR analysis, cells with the GG genotype showed lower expression of *DSP* compared with those with the TT genotype, the same trend as seen in human primary bronchial epithelial cells (Figure 1A).

In addition to applying the CRISPR/Cas9 method to determine endogenous regulation of *DSP* expression by rs2076295, we also examined the allele-specific regulation of the *DSP* promoter by rs2076295 by luciferase reporter assays in 16HBE14o- cells. Indeed, the DNA constructs containing the T allele of rs2076295 led to higher *DSP* promoter activity compared with matched G allele constructs (Figure 1F).

Consistently, in primary NHBE cells, significantly decreased expression of *DSP* (~20%) was detected (Figure 1G) after CRISPR/Cas9-mediated gRNA delivery targeting rs2076295, with more than 99% deletion efficiency shown by Droplet Digital PCR (Figure E1D).

Therefore, our results, obtained from a series of CRISPR/Cas9 genome-editing approaches and reporter assays, strongly support the idea that rs2076295 regulates *DSP* expression, with IPF risk allele GG decreasing expression of *DSP* in human bronchial epithelial cells.

DSP Is Localized in Both Human Airway and Alveolar Epithelial Cells

The expression of *DSP* was detected primarily in human lung epithelial cells (Figure E2A). By immunohistochemistry staining, *DSP* is detected in both airway

and alveolar epithelial cells in normal human lungs (Figure 2A). Specifically, *DSP* showed a punctate staining pattern in the cell boundary between alveolar type I (AT1; marked with T1 α) and type II cells (AT2; marked with ABCA3) in alveoli as well as the apical and lateral sides of airway epithelial cells, including ciliated cells and club cells (Figure 2B). The expression of *DSP* in airway basal cells (marked with Krt5) was also detected (Figure 2B). However, airway smooth muscle cells and lung fibroblasts showed no *DSP* staining, as expected (Figure E2B). In addition, in NHBE cells (Figure 2C) as well as in mouse tracheal epithelial cells differentiated at the air-liquid interface (Figures E2C and E2D), *DSP* demonstrated strong expression at the apical surface of the airway epithelium, suggesting that the desmosomes formed by *DSP* may be important for maintaining airway epithelial integrity.

Knockout of *DSP* Reduces TEER, Increases Epithelial-Mesenchymal Transition, and Promotes Cell Migration in Epithelial Cells

Using gRNA targeting exon 2 of *DSP*, we generated three *DSP*-knockout clonal lines in 16HBE14o- cells (Figures 3A and 3B). In *DSP*-knockout clonal lines, the loss of punctate staining of *DSP* between cell boundaries is accompanied by the perinuclear accumulation of keratin 5 (in red), in contrast to a normal filament network pattern of keratin 5 attached to desmosomes in wild-type cells (Figure 3C) (19). As a key component of the desmosome complex, *DSP* enables epithelial cells to resist mechanical stress, which is required to maintain the integrity of the epithelial barrier (20, 21). To directly examine the role of *DSP* in maintaining cellular integrity and permeability of bronchial epithelial cells cultured at monolayers, we assessed TEER in *DSP*-knockout clones. Indeed, decreased TEER was detected in the *DSP*-knockout clones compared with wild-type cells after monolayer culture in transwells (Figures 3D and 3E).

Figure 1. (Continued). one WT control are shown. Means $\pm$ SEM are from at least four biological replicates. (D) Upper: Expression of *DSP* in 16HBE14o- cells edited by CRISPR interference (CRISPRi) targeting rs2076295. Means $\pm$ SEM are two biological replicates. Location of three single guide RNAs designed for CRISPRi targeting rs2076295 are indicated below. (E) Expression of *DSP* in 16HBE14o- lines (GT) edited by CRISPR/Cas9 method into homozygous (TT or GG) at rs2076295. Means $\pm$ SEM are from four to nine biological replicates of each isogenic line. (F) Relative luciferase activity with either G or T allele of rs2076295 as measured in reporter assay with endogenous *DSP* promoter cloned in pGL3 basic vector and transfected into 16HBE14o- cells. Means $\pm$ SEM are from two biological replicates representative of four biological repeats. (G) Expression of *DSP* in primary normal human bronchial epithelial cells with or without rs2076295 regional deletion by CRISPR/Cas9. Means $\pm$ SEM are from three biological replicates. WT control cells are cells transfected with empty vector. * P < 0.05, unpaired t test (for A–E and G) and Mann-Whitney test (for F). NHBE = normal human bronchial epithelial; sgRNA = single guide RNA.

Figure 2. Localization of DSP (desmoplakin) in human lungs. (A) Representative images and enlarged images (in the black squares) of immunohistochemistry staining for DSP in the bronchi and alveoli in human lung tissue. Scale bars, 10 μ m. (B) Representative images and enlarged images (in the white squares) of immunofluorescence staining of DSP (green) and various cell-type markers (red), including T1 α , ABCA3, acetylated tubulin, CC10, and keratin 5, with DAPI (blue) in human lung tissue. Scale bars, 20 μ m. The white dashed line indicates the basal membrane. (C) Images of immunofluorescence staining of DSP in differentiated human airway epithelial cells at the air-liquid interface after 7 days of differentiation. The three dashed lines (a–c) indicate the levels where images were taken (the top, middle, and bottom areas of the luminal epithelial cells at cross-section). Scale bars, 50 μ m. ABCA3 = ATP-binding cassette subfamily A member 3; ACTTUB = acetylated tubulin; ALI = air-liquid interface; IHC = immunohistochemistry; Krt5 = keratin 5; T1 α = lung type I cell membrane-associated glycoprotein.

Figure 3. Generation of *DSP* (desmoplakin)-knockout (KO) lines in 16HBE14o- human bronchial epithelial cell line. (A) Expression of *DSP* in three single colonies of 16HBE14o- cells with *DSP*-KO (KO1, KO2, and KO3) using CRISPR (clustered regularly interspaced short palindromic repeat)/Cas9 targeting the coding region of *DSP*. Means $\pm$ SEM are from four biological replicates. (B) The levels of *DSP* as measured in wild-type (WT) and *DSP*-KO 16HBE14o- cells by Western blotting. Vinculin was applied as a loading control. (C) The representative immunofluorescence staining of *DSP* (green) and keratin 5 (red) in *DSP*-KO cells and control cells. Nuclei are counterstained with DAPI (blue). Scale bars, 10 μ m. (D) The transepithelial electrical resistance (TEER) measurement in WT and three *DSP*-KO 16HBE14o- cells cultured on transwells for 18 continuous days. Means $\pm$ SEM are shown from triplicate wells within each line. One representative result is shown from two biological replicates. A linear mixed effects model with Dunnett's *post hoc* test used to compare difference among four groups demonstrated significant difference among groups over time. Significantly reduced TEER was found in at least two KO lines compared with WT cells, starting from day 7 to Day 18. (E) TEER measurements on Day 18 in D were shown in WT and *DSP*-KO cells. * $P < 0.05$, unpaired *t* test. WT control cells are cells transfected with empty vector. Krt5 = keratin 5.

Epithelial cells in IPF lungs usually have epithelial-mesenchymal transition (EMT) pathological changes (22). Indeed, significant downregulation of epithelial markers (E-cadherin, Syndecan-1, Claudin-1, and COL4A1) and upregulation of mesenchymal markers (Snail, vimentin, S100A4, N-cadherin, and Slug) were

detected in *DSP*-knockout clones compared with in wild-type (empty vector-transfected) 16HBE14o- cells (Figures 4A and 4B). Similarly, decreased E-cadherin and increased vimentin were also detected by immunofluorescence staining in *DSP*-knockout cells (Figure 4C), indicating that the loss of *DSP* may promote EMT in

human bronchial epithelial cells. Consistent with EMT gene expression changes, 16HBE14o- cells with reduced expression of *DSP* by either knockout or rs2076295 indel mutation exhibited increased cell migration (Figures 4D and 4E), a key feature of fibrotic cells in IPF lungs (23). We also confirmed increased cell migration upon *DSP* knockout in primary airway epithelial cells isolated from *DSP* flox/flox mice treated with AAV-Cre *ex vivo* (Figures 4F and 4G), in which the knockout efficiency of *DSP* was more than 80% (Figure 4H). In contrast, cell death and cell proliferation are comparable between control and *DSP*-knockout 16HBE14o- clonal lines (Figures E3A and E3B).

Differentially Expressed Genes in *DSP*-Knockout 16HBE14o- Cells Include Differentially Expressed Genes Found in Human IPF Lungs

To further dissect the signaling pathways regulated by *DSP*, RNA-seq was performed in *DSP*-knockout and wild-type 16HBE14o- cells. A total of 209 genes exhibited significantly decreased expression, and 50 genes exhibited significantly increased expression in *DSP*-knockout 16HBE14o- cells compared with wild-type cells (Table E3).

Given that the IPF risk allele G at rs2076295 (24) is associated with reduced expression of *DSP* (Figures 1A and 1E), we hypothesized that the knockout of *DSP* in 16HBE14o- cells might lead to gene expression changes that resemble transcriptomic signatures of IPF lungs. Therefore, we compared previously published RNA-seq data in human IPF lungs (25) with RNA-seq data in *DSP*-knockout clones (Figure E4). Overlapped genes include four upregulated ECM-related genes (*COL6A3*, *MMP7*, *MFAP2*, and *DNAH3*) and six downregulated genes (*SEMA3E*, *KLF15*, *PTX3*, *MGAM*, *TNNC1*, and *BDNF*), supporting the idea that the downregulation of *DSP* may trigger similar pathways to those altered in human IPF lungs.

DSP Modulates Bronchial Epithelial Cell Migration via Regulating ECM-related Genes

Pathway analysis on differentially expressed genes in *DSP*-knockout cells demonstrated that genes with decreased expression in knockout clones were involved in cell-cell adhesion and cell junction, whereas genes with increased expression were involved in

Figure 4. Influences of *DSP* (desmoplakin)-knockout (KO) on epithelial–mesenchymal transition, and bronchial epithelial cell migration. (A) Expression of epithelial markers (*CDH1*, *COL4A1*, and *SDC1*) and mesenchymal markers (*SNAI1*, *VIM*, and *S100A4*) in control and *DSP*-KO 16HBE14o- cells measured by quantitative PCR. Means $\pm$ SEM are from four biological replicates. GAPDH was applied as a reference gene. *False discovery rate–adjusted *P* value < 0.05 , unpaired *t* tests with false discovery rate correction for multiple testing. (B) The Western blotting result of mesenchymal and epithelial

ECM remodeling. We then validated the expression changes of upregulated ECM genes in cells with either *DSP* knockout or indel mutation at rs2076295 (Figure 5A), including *MFAP2*, *ADAM19*, *SRGN*, *MMP7*, *MMP9*, *COL6A3*, and *COL13A1* by quantitative PCR (Figures 5B and 5C). Among these seven genes, five of them, including *MFAP2*, *ADAM19*, *MMP7*, *MMP9*, and *COL6A3*, showed consistently increased expression when either *DSP* was knocked out or rs2076295 was modified with an indel. Furthermore, increased expression of *MMP7*, *MMP9*, and *ADAM19* was confirmed in NHBE cells cultured at the air–liquid interface (Figure 5D) after *DSP* knockout by CRISPR/Cas9 as well as primary EpCAM+ lung epithelial cells from *DSP* flox/flox mice with *ex vivo* treatment of Adeno-Cre in three-dimensional organoid models (Figure 5E). These results suggest that *DSP* represses the expression of ECM genes in both murine and human primary lung epithelial cells.

To determine whether these *DSP*-regulated ECM genes contribute to *DSP*-mediated cell migration, we knocked down five ECM-related genes individually in wild-type and *DSP*-knockout 16HBE14o- cells and measured cell migration. The silencing of *MMP7* or *MMP9* reversed the increased migration in *DSP*-knockout cells (Figures 5F and 5G), indicating that *MMP7* and *MMP9* may mediate *DSP*-regulated cell migration.

Cell Type-Specific Expression

Changes of *DSP* in Human IPF Lungs

To determine the expression changes of *DSP* and its regulated genes in human IPF lungs, we performed a series of analyses in human IPF lung samples. First, the expression of *DSP* showed an approximately threefold increase at the mRNA level and a twofold increase at the protein level in total IPF lungs compared with that found in normal control lungs (Figures 6A–6C). Moreover, reduced concentrations of the epithelial cell marker

E-cadherin and increased concentrations of the mesenchymal cell marker vimentin were detected in IPF lungs, which is consistent with increased EMT in IPF lungs (Figures 6D and 6E), as reported previously (22, 26). Second, decreased expression of E-cadherin and increased staining of vimentin were evident in airway regions from IPF lungs (Figure 6F). However, we were unable to detect EMT changes in distal lungs because of the complete loss of normal alveolar structure in the end stage of IPF. Lastly, publicly available single-cell RNA-seq data in normal and IPF human lungs(11) were analyzed for cell type-specific expressions of *DSP* and its regulated genes. Despite the increased expression in whole lungs, *DSP* showed significantly reduced expression in alveolar epithelial cells, including AT1, AT2, and ciliated cells (Figures 6G and 6H). Among the nine EMT marker genes we examined, seven genes showed detectable expression (Figure 6I) in at least one lung epithelial cell type. As an important cell type for IPF, AT2 cells showed significant downregulation of epithelial genes (*CDH1*) and upregulation of mesenchymal genes (*VIM* and *FN1*) as well as ECM-related genes (*MMP7* and *SRGN*). In summary, human IPF lungs demonstrated reduced expression of epithelial cells genes including *DSP* and increased expression of mesenchymal and ECM genes in AT2 cells, as we observed in *DSP*-knockout 16HBE14o- cells.

Discussion

Multiple publications have consistently reported the association of rs2076295, an intronic variant within the *DSP* gene, with IPF susceptibility by GWAS (14, 27), yet no functional studies have been conducted on this variant (7, 28). Using comprehensive CRISPR/Cas9-based methods in both cell line and primary NHBE cells, we demonstrate for the first time 1) the direct causal link between rs2076295 and *DSP*

expression, 2) the directional regulation of rs2076295 on the expression of *DSP* in human airway epithelial cells, 3) the location of *DSP* at the apical surface of airway epithelial cells, and 4) that the loss of *DSP* in lung epithelial cells leads to reduced epithelial features (reduced barrier function) with transitional changes toward fibroblasts (increased expression of EMT and ECM genes with increased cell migration). Our findings represent the important and necessary first step for functional studies on the *DSP* GWAS locus in IPF.

The IPF risk allele rs2076295G is genetically associated with decreased expression of *DSP* (8, 24); however, increased expression of *DSP* at both RNA and protein levels in human IPF lungs has been reported (8), as we observed as well (Figures 6A–6C). One possible reason that explains such seemingly conflicting changes is the stiffness-induced expression of *DSP* in IPF lungs, which have a significantly increased number of myofibroblasts and stiffness of lung matrix (29). Therefore, increased expression of *DSP* in total lungs is a consequence instead of a driving force of IPF development. Second, elevated expression levels of *DSP* in total IPF lungs may represent a mixture of altered gene expression in multiple lung cell types, as supported by single-cell RNA-seq analysis (Figures 6G and 6H). We found that in healthy lungs, ciliated cells have the most *DSP* expression compared with other lung epithelial cells, whereas in IPF lungs, aberrant basal cells (30) have significantly increased expression of *DSP* (Figure 6H), suggesting that basal cells may be the most sensitive to matrix stiffness-induced expression of *DSP*. Meanwhile, alveolar epithelial cells, including AT1 and AT2 cells, have reduced expression of *DSP*, supporting the theory that reduced *DSP* expression increased the risk of IPF. Therefore, although altered expression levels of a given GWAS gene in human diseased tissues can

Figure 4. (Continued). markers in control and *DSP*-KO 16HBE14o- cells. α -Tubulin was applied as a loading control. (C) The representative images and enlarged images of immunofluorescence staining of E-cadherin (green) and vimentin (red) with DAPI (blue) staining in wild-type (WT) and *DSP*-KO 16HBE14o- cells. Scale bars, 50 μ m. (D) The representative images of cell migration assays in WT, *DSP*-KO, and rs2076295 indel mutant 16HBE14o- cells at 0 and 24 hours after migration initiation. The leading edges of the migrated cells were outlined by black lines. (E) Quantification on cell migration rate by fluorescence measurement with Calcein-AM dye. Means $\pm$ SEM are from three biological replicates. (F) The representative images of cell migration assays in WT and two *ex vivo* induced *DSP*-KO mouse tracheal basal cells at 0 and 48 hours after migration initiation. The leading edges of the migrated cells are outlined by black lines. * P < 0.05, unpaired t test. (G) Quantification on cell migration rate by fluorescence measurement with Calcein-AM dye. Means $\pm$ SEM are from triplicate wells in the tracheal basal cells isolated from WT (*DSP*^{fl/fl}) and two *ex vivo* induced *DSP*-KO mice (*DSP*^{fl/fl}+Cre). * P < 0.05, unpaired t test. (H) Expression of *DSP* in mouse tracheal basal cells used in G. Means $\pm$ SEM are from duplicate wells. WT control cells in A–E are cells transfected with empty vector. E-cad = E-cadherin; Indel = insertions or deletions of three or multiples of three base pairs; VIM = vimentin.

Figure 5. *DSP* (desmoplakin) regulates extracellular matrix (ECM)-related gene expression and cell migration in bronchial epithelial cells. (A) The heatmap of RNA-sequencing results showing differentially expressed ECM-related genes in *DSP*-knockout (KO) 16HBE140- cells compared with wild-type (WT) control cells. (B) Relative expression of *DSP*-regulated ECM-related genes identified from RNA-sequencing results in WT and three clones of *DSP*-KO 16HBE140- cells (KO1, KO2, and KO3). Means $\pm$ SEM are from two biological replicates for each individual clone. (C) Relative expression of *DSP*-regulated ECM-related genes in rs2076295 indel mutation cells (Indel) versus WT control. Means $\pm$ SEM are from two biological replicates. (D) Relative expression of *DSP*- and ECM-related genes in WT and *DSP*-KO primary human bronchial epithelial cells cultured at the air-liquid interface for 7 days. Means $\pm$ SEM are from duplicate wells. (E) Relative expression of ECM-related genes in EpCAM+ lung epithelial cells from a *DSP* flox/flox mouse treated with Adeno-Cre followed by subsequent culture at three-dimensional organoid models. Means $\pm$ SEM are from two independent replicates. *False discovery rate-adjusted P value < 0.05 , unpaired t tests with false discovery rate correction for multiple testing for B–E. (F) The representative images of WT and *DSP*-KO 16HBE140- cells with or without knockdown of MMP7 and MMP9 in migration assay at 24 hours after migration initiation. The leading edges of the migrated cells were outlined in black. (G) The cell migration rate in WT and *DSP*-KO 16HBE140- cells after knockdown of individual ECM-related genes by siRNAs. Means $\pm$ SEM are from three biological replicates. * or $^{\#}P < 0.05$, two-way ANOVA with Dunnett's *post hoc* test comparing cells transfected with siRNA targeting various genes to cells transfected with negative control within *DSP*-KO or WT groups, respectively. $^{\#}P < 0.05$, unpaired t test comparing between WT and *DSP*-KO cells. WT control cells are cells transfected with empty vector. ALI = air-liquid interface; EpCAM = epithelial cell adhesion molecule; NHBE = normal human bronchial epithelial cells.

Figure 6. Detection of *DSP* (desmoplakin) and its regulated genes in human idiopathic pulmonary fibrosis (IPF) lungs. (A) Expression of *DSP* in human IPF ($n = 16$) and control lungs ($n = 16$). * $P < 0.05$, unpaired t test. (B) Protein levels of *DSP*, E-cadherin, and vimentin measured in healthy control and IPF lungs by Western blotting. β -actin was applied as a loading control. (C–E) The quantification of protein levels for *DSP*, E-cadherin, and vimentin in the lung of healthy control ($n = 15$) and IPF ($n = 14$) lungs. Means $\pm$ SEM are shown. * $P < 0.05$, unpaired t test (for C) or Mann-Whitney test (for D and E). (F) The representative images of immunofluorescence staining of E-cadherin (green), vimentin (red), and DAPI (blue) in human control and IPF lung tissue. Scale bars, 20 μ m. (G) Heatmap showing average expression of *DSP* in control and IPF lungs using data from single-cell RNA-sequencing in human IPF lungs. (H) Heatmap showing the average expression of *DSP* within various lung epithelial subtypes. (I) Heatmap showing average log₂ fold difference in gene expression between IPF and control lungs by cell type categories. Genes with an adjusted P value lower than 0.10 were plotted. AT1 = type I alveolar epithelial cells; AT2 = alveolar type II epithelial cells; avg_diff = average log₂ fold difference; E-Cad = E-cadherin; ECM = extracellular matrix; VIM = vimentin.

support its relevance to human disease, the directional impacts of the GWAS gene on disease susceptibility and progression are more reliably inferred by the combined analysis of genetic association studies, eQTL analysis in relevant tissues, and CRISPR/Cas9-based genome-editing approaches in relevant cell types.

In addition to IPF, rs2076295 is also associated with chronic obstructive pulmonary disease (COPD) in GWAS but with an opposite direction of the effects for the risk allele (9), similar to other COPD/IPF-overlapped GWAS loci such as *FAM13A* and *MAPT* loci, where the fibrosis risk alleles are protective for COPD (24). It would be intriguing to determine how these overlapped GWAS loci may divert the disease susceptibility toward either IPF or COPD in future animal models. Despite lacking *in vivo* evidence to conclude the roles of *DSP* in the development of IPF and COPD, our studies have provided a series of solid evidence to improve our understanding of the function of this GWAS locus. First, the functional variant rs2076295, as the top GWAS SNP for both COPD and IPF, regulates the expression of *DSP* (supported by comprehensive CRISPR/Cas9 genome-editing approaches). Second, *DSP* is localized in various lung epithelial cell types that are relevant for both COPD and IPF. Third, in a loss-of-function model, *DSP* regulates ECM gene expression and inhibits cell migration, an important cellular function for both COPD and IPF pathogenesis.

Bronchial epithelial cells were used as the main cellular model for rs2076295 functional studies because of feasibilities and efficiency related with transfection experiments and CRISPR/Cas9-mediated genome-editing methods. Furthermore, open chromatin near rs2076295 in both bronchial epithelial cells and AT2 cells

(Figures E1A and E1B) also supports airway epithelial cells as a relevant cellular model to study the effects of rs2076295. Admittedly, critical roles of alveolar epithelial cells for the pathogenesis of progressive fibrosis are nonnegligible, future open chromatin profiling studies on normal and IPF AT2 cells may illuminate more comprehensive and mechanistic insights into putative functional variants in all IPF GWAS loci.

DSP knockout led to an increased expression of genes related to the ECM (Figure 5A). The ECM not only provides a scaffold for epithelial cells, it also provides essential cues for tissue morphogenesis during development and for tissue repair during homeostasis and disease. The ECM composition is markedly changed in the pathogenesis of both COPD and IPF (31, 32). IPF is characterized by an accumulation of excess fibrous material in the lung, resulting from the excessive production and deposition of ECM (32), whereas emphysema, a major pathological manifestation of COPD, results from parenchymal destruction due to significant ECM loss and defective alveolar repair. We speculate that increased ECM gene expression under *DSP* deficiency may prevent emphysema development but promote fibrotic changes because we see that *MFAP2*, *MMP7* and *COL6A3* (Figures 5B and 5C) were all upregulated in human IPF lungs (Figure E4).

Furthermore, *DSP*-deficient 16HBE14o- cells have increased cell migration, as reported previously, and depletion of *DSP* in mice is associated with increased tumor invasion in pancreatic cancer (33). More importantly, we found that the silencing of *MMP7* or *MMP9* reversed *DSP* knockout-induced cell migration (Figures 5F and 5G). *MMP7* and *MMP9* belong to matrix metalloproteinases, a pivotal family of zinc enzymes capable of

degrading ECM components, including basement membrane collagen, interstitial collagen, fibronectin, and various proteoglycans, during tissue remodeling and repair processes (34). Several matrix metalloproteinases (including *MMP1*, *MMP7*, *MMP8*, and *MMP9*) showed increased levels in human IPF lungs, possibly contributing to ECM remodeling and basement membrane disruption during IPF development (35–37). For example, *MMP7* induced epithelial-to-mesenchymal transition, whereas *MMP9* may promote abnormal epithelial cell migration and other aberrant repair processes (37), thus promoting the pathogenesis of fibrotic lungs. Admittedly, future mechanistic studies on the possible regulation of *DSP* on Wnt/ β -catenin (38), TGF β (39), and Notch pathways (40, 41) modulating cell migration will illuminate additional insights into *DSP*.

In conclusion, our study provides multiple evidences to elucidate the role of IPF-susceptibility variant rs2076295 in regulating *DSP* expression in human bronchial epithelial cells, with the risk allele G leading to lower *DSP* expression. *DSP* is normally expressed predominantly in lung epithelial cells, and reduced *DSP* expression promotes cell migration possibly by upregulating ECM-related genes, which may contribute to IPF development. ■

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

Acknowledgment: The authors thank Dr. Scott H. Randell from the University of North Carolina at Chapel Hill for generously providing healthy human lung samples for primary lung epithelial cell isolation used in quantitative PCR and assay for transposase-accessible chromatin sequencing analysis. They also thank Dr. Xingbin Ai from Massachusetts General Hospital, Harvard Medical School, for her valuable comments on this study.

References

- Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. *Lancet* 2017;389:1941–1952.
- Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, et al. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* 2017;3:17074.
- Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;183:431–440.
- Xaubet A, Ancochea J, Molina-Molina M. Idiopathic pulmonary fibrosis [in English, Spanish]. *Med Clin (Barc)* 2017;148:170–175.
- Allen RJ, Guillen-Guio B, Oldham JM, Ma SF, Dressen A, Paynton ML, et al. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2020;201:564–574.
- Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013;1:309–317.
- 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.
- Mathai SK, Pedersen BS, Smith K, Russell P, Schwarz MI, Brown KK, et al. Desmoplakin variants are associated with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016;193:1151–1160.
- Hobbs BD, de Jong K, Lamontagne M, Bossé Y, Shrine N, Artigas MS, et al.; COPD Gene Investigators; ECLIPSE Investigators; LifeLines Investigators; SPIROMICS Research Group; International COPD Genetics Network Investigators; UK BiLEVE Investigators;

- International COPD Genetics Consortium. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017;49:426–432.
10. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 2017;101:5–22.
 11. Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, Calvi CL, et al. Single-cell RNA-sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* 2020;6:eaba1972.
 12. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM III, et al. Comprehensive integration of single-cell data. *Cell* 2019;177:1888–1902, e21.
 13. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al.; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–330.
 14. Ernst J, Kellis M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat Biotechnol* 2015;33:364–376.
 15. Kazachenka A, Bertozzi TM, Sjöberg-Herrera MK, Walker N, Gardner J, Gunning R, et al. Identification, characterization, and heritability of murine metastable epialleles: implications for non-genetic inheritance. *Cell* 2018;175:1259–1271, e13. [published erratum appears in *Cell* 2018;175:1717.]
 16. Inoue F, Kircher M, Martin B, Cooper GM, Witten DM, McManus MT, et al. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res* 2017;27:38–52.
 17. Morrow JD, Chase RP, Parker MM, Glass K, Seo M, Divo M, et al. RNA-sequencing across three matched tissues reveals shared and tissue-specific gene expression and pathway signatures of COPD. *Respir Res* 2019;20:65.
 18. Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* 2015;12:1143–1149.
 19. Hopkins G, Kimura TE, Garrod DR. Desmoplakin is essential for epidermal sheet formation. *J Invest Dermatol* 2007;127:E12.
 20. Kottke MD, Delva E, Kowalczyk AP. The desmosome: cell science lessons from human diseases. *J Cell Sci* 2006;119:797–806.
 21. Albrecht LV, Zhang L, Shabanowitz J, Purejav E, Towbin JA, Hunt DF, et al. GSK3- and PRMT-1-dependent modifications of desmoplakin control desmoplakin-cytoskeleton dynamics. *J Cell Biol* 2015;208:597–612.
 22. Selman P, Pardo A, Kaminski N. Idiopathic pulmonary fibrosis: aberrant recapitulation of developmental programs? *PLoS Med* 2008;5:e62.
 23. Suganuma H, Sato A, Tamura R, Chida K. Enhanced migration of fibroblasts derived from lungs with fibrotic lesions. *Thorax* 1995;50:984–989.
 24. van Moersel CHM. Trade-offs in aging lung diseases: a review on shared but opposite genetic risk variants in idiopathic pulmonary fibrosis, lung cancer and chronic obstructive pulmonary disease. *Curr Opin Pulm Med* 2018;24:309–317.
 25. Gangwar I, Kumar Sharma N, Panzade G, Awasthi S, Agrawal A, Shankar R. Detecting the molecular system signatures of idiopathic pulmonary fibrosis through integrated genomic analysis. *Sci Rep* 2017;7:1554.
 26. Kim KK, Kugler MC, Wolters PJ, Robillard L, Galvez MG, Brumwell AN, et al. Alveolar epithelial cell mesenchymal transition develops *in vivo* during pulmonary fibrosis and is regulated by the extracellular matrix. *Proc Natl Acad Sci USA* 2006;103:13180–13185.
 27. Najor NA. Desmosomes in human disease. *Annu Rev Pathol* 2018;13:51–70.
 28. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM, et al. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir Med* 2017;5:869–880.
 29. Qu J, Zhu L, Zhou Z, Chen P, Liu S, Locy ML, et al. Reversing mechanoinductive DSP expression by CRISPR/dCas9-mediated epigenome editing. *Am J Respir Crit Care Med* 2018;198:599–609.
 30. Xu Y, Mizuno T, Sridharan A, Du Y, Guo M, Tang J, et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016;1:e90558.
 31. Kulkarni T, O'Reilly P, Antony VB, Gaggari A, Thannickal VJ. Matrix remodeling in pulmonary fibrosis and emphysema. *Am J Respir Cell Mol Biol* 2016;54:751–760.
 32. Burgess JK, Mauad T, Tjin G, Karlsson JC, Westergren-Thorsson G. The extracellular matrix: the under-recognized element in lung disease? *J Pathol* 2016;240:397–409.
 33. Broussard JA, Getsios S, Green KJ. Desmosome regulation and signaling in disease. *Cell Tissue Res* 2015;360:501–512.
 34. Ohbayashi H. Matrix metalloproteinases in lung diseases. *Curr Protein Pept Sci* 2002;3:409–421.
 35. Dancer RC, Wood AM, Thickett DR. Metalloproteinases in idiopathic pulmonary fibrosis. *Eur Respir J* 2011;38:1461–1467.
 36. Pardo A, Selman M. Role of matrix metalloproteinases in idiopathic pulmonary fibrosis. *Fibrogenesis Tissue Repair* 2012;5:S9.
 37. Craig VJ, Zhang L, Hagood JS, Owen CA. Matrix metalloproteinases as therapeutic targets for idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2015;53:585–600.
 38. Yang L, Chen Y, Cui T, Knösel T, Zhang Q, Albring KF, et al. Desmoplakin acts as a tumor suppressor by inhibition of the Wnt/ β -catenin signaling pathway in human lung cancer. *Carcinogenesis* 2012;33:1863–1870.
 39. Yoshida M, Romberger DJ, Illig MG, Takizawa H, Sacco O, Spurzem JR, et al. Transforming growth factor-beta stimulates the expression of desmosomal proteins in bronchial epithelial cells. *Am J Respir Cell Mol Biol* 1992;6:439–445.
 40. Namba T, Tanaka KI, Ito Y, Hoshino T, Matoyama M, Yamakawa N, et al. Induction of EMT-like phenotypes by an active metabolite of leflunomide and its contribution to pulmonary fibrosis. *Cell Death Differ* 2010;17:1882–1895.
 41. Shao S, Zhao X, Zhang X, Luo M, Zuo X, Huang S, et al. Notch1 signaling regulates the epithelial-mesenchymal transition and invasion of breast cancer in a Slug-dependent manner. *Mol Cancer* 2015;14:28.