Desmoplakin Variants Are Associated with Idiopathic Pulmonary Fibrosis

Susan K. Mathai¹, Brent S. Pedersen², Keith Smith¹, Pamela Russell³, Marvin I. Schwarz¹, Kevin K. Brown⁴, Mark P. Steele⁵, James E. Loyd⁵, James D. Crapo⁴, Edwin K. Silverman⁶, Deborah Nickerson⁷, Tasha E. Fingerlin⁸, Ivana V. Yang^{1,8}, and David A. Schwartz¹

¹Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, and ³Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Anschutz Medical Campus, Aurora, Colorado; ²Department of Genetics, University of Utah, Salt Lake City, Utah; ⁴Department of Medicine, National Jewish Health, Denver, Colorado; ⁵Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee; ⁶Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; ⁷Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington; and ⁸Center for Genes, Environment and Health, National Jewish Health, Denver, Colorado

Abstract

Rationale: Sequence variation, methylation differences, and transcriptional changes in desmoplakin (*DSP*) have been observed in patients with idiopathic pulmonary fibrosis (IPF).

Objectives: To identify novel variants in *DSP* associated with IPF and to characterize the relationship of these IPF sequence variants with *DSP* gene expression in human lung.

Methods: A chromosome 6 locus (7,370,061–7,606,946) was sequenced in 230 subjects with IPF and 228 control subjects. Validation genotyping of disease-associated variants was conducted in 936 subjects with IPF and 936 control subjects. *DSP* gene expression was measured in lung tissue from 334 subjects with IPF and 201 control subjects.

Measurements and Main Results: We identified 23 sequence variants in the chromosome 6 locus associated with IPF. Genotyping of selected variants in our validation cohort revealed that noncoding

intron 1 variant rs2744371 (odds ratio = 0.77, 95% confidence interval [CI] = 0.66–0.91, $P = 0.002$) is protective for IPF, and a previously described IPF-associated intron 5 variant (rs2076295) is associated with increased risk of IPF (odds ratio = 1.36, 95% CI = 1.19–1.56, $P < 0.001$) after controlling for sex and age. *DSP* expression is 2.3-fold increased (95% CI = 1.91–2.71) in IPF lung tissue ($P < 0.0001$). Only the minor allele at rs2076295 is associated with decreased *DSP* expression ($P = 0.001$). Staining of fibrotic and normal human lung tissue localized *DSP* to airway epithelia.

Conclusions: Sequence variants in *DSP* are associated with IPF, and rs2076295 genotype is associated with differential expression of *DSP* in the lung. *DSP* expression is increased in IPF lung and concentrated in the airway epithelia, suggesting a potential role for *DSP* in the pathogenesis of IPF.

Keywords: idiopathic pulmonary fibrosis; genetics; desmoplakin; idiopathic interstitial pneumonia; intron

(Received in original form September 21, 2015; accepted in final form December 7, 2015)

Supported by National Institutes of Health grants R01 HL097163, R21/R33 HL120770, UH2-HL123442, P01 HL092870, F32 HL123240, T32 HL007085, R01 HL089856, and R01 HL089897 and Department of Veterans Affairs grant 1101BX001534. Sequencing was provided by the Northwest Genomics Center at the University of Washington, Department of Genome Sciences, under U.S. Federal Government contract number HHSN268201100037C from the NHLBI for the Resequencing and Genotyping Service. Tissue samples were obtained from the Lung Tissue Research Consortium, a resource created by the NHLBI.

Author Contributions: S.K.M. performed genotyping and expression analysis described, performed statistical analysis, and wrote and edited the manuscript; K.S. assisted in genotyping and expression analysis. P.R. and T.E.F. aided in the presented statistical analyses. D.N. supervised sequencing performed at the University of Washington. B.S.P., T.E.F., and I.V.Y. analyzed sequencing data. I.V.Y., D.A.S., M.I.S., T.E.F., and S.K.M. contributed to study design and manuscript revision. D.A.S., K.K.B., M.P.S., J.E.L., E.K.S., and J.D.C. contributed to obtaining the samples analyzed in this study. All authors reviewed the final manuscript.

Correspondence and requests for reprints should be addressed to Susan K. Mathai, M.D., Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, 12631 East 17th Avenue, Aurora, Colorado 80045. E-mail: susan.mathai@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 Crit Care Med Vol 193, Iss 10, pp 1151–1160, May 15, 2016

Copyright © 2016 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201509-1863OC on December 15, 2015

Internet address: www.atsjournals.org

At a Glance Commentary

Scientific Knowledge on the

Subject: Idiopathic pulmonary fibrosis (IPF) is the most common of the idiopathic interstitial pneumonias. Variants in desmoplakin (*DSP*) are associated with fibrotic idiopathic interstitial pneumonia.

What This Study Adds to the

Field: This study describes the association of variants in *DSP* with idiopathic pulmonary fibrosis (IPF). We show that *DSP* expression is significantly elevated in the IPF lung and that a disease-associated *DSP* variant rs2076295 is associated with differential gene expression in human lung.

Idiopathic pulmonary fibrosis (IPF), the most common of the idiopathic interstitial pneumonias (IIPs), is a progressive disease characterized by the development of fibrosis within the lung parenchyma. IPF has a poor prognosis, often resulting in death within 2–3 years of diagnosis (1, 2), affecting 5 million people worldwide, and is increasing in prevalence (3). Although investigations have pointed to genetic and environmental factors that confer risk of disease, the etiology of IPF remains elusive.

Previously, we performed a genome-wide association study in subjects with fibrotic IIP, identified seven novel loci associated with disease (*FAM13A* [4q22], desmoplakin [*DSP*; 6p24], *OBFC1* [10q24], *ATP11A* [13q34], *DPP9* [19p13], and chromosomal regions 7q22 and 15q14–15), and confirmed disease associations at *MUC5B*, *TERT*, and *TERC* (4). This prior study also found that *DSP* gene expression was increased in lung tissue of individuals with fibrotic IIP and varied by copies of the minor allele for a variant (rs2076295) in intron 5, which was the most significant of the *DSP* variants associated with fibrotic IIP (4). *DSP* is a desmosomal protein critical to cell–cell adhesion in a variety of cell types (5–7) and important in wound healing and epithelial barrier function (8, 9). Desmosomes mechanically connect cells, stabilize tissue architecture, and are essential in cell migration, proliferation, and differentiation, linking the desmosome to tumorigenesis (10). The *DSP* variant

rs2076295 is within the binding site for transcription factor PU.1, which is implicated in alternative splicing of many genes, alternative macrophage activation, and airway remodeling in asthma (11, 12). However, there is no evidence for an effect of rs2076295 genotype on expression of the primary isoform compared with the alternative isoform of *DSP* in the lung (4).

We demonstrated that intron 1 of *DSP* is hypomethylated in lung tissue of patients with the most common fibrotic IIP, IPF (13). ENCODE data identify a DNase hypersensitivity site in intron 1, and there is a 1,954-bp CpG island spanning exon 1 and intron 1 of *DSP* (<https://genome.ucsc.edu>). This suggests that intron 1 of *DSP* is involved in gene regulation and could be involved in the differential expression of *DSP*.

Prior studies of *DSP* in human lung disease indicate that gene expression is decreased in the majority of lung cancer cell lines (14). These cell lines demonstrated hypermethylation in the promoter and intron 1 regions of the gene, indicating that epigenetic mechanisms may regulate *DSP* gene expression. Furthermore, *in vitro* experiments suggest that *DSP* functions as a tumor suppressor through inhibition of the Wnt/ β -catenin pathway in non-small cell lung cancer, and that *DSP* increased the sensitivity of cancer cells to anticancer drugs (14).

Based on these observations, we sought to identify biologically relevant sequence variants in *DSP* associated with IPF. We hypothesized that sequence variants in *DSP*, especially those localized to intron 1, would

be associated with IPF and with gene expression of *DSP* in lung tissue.

Methods

Sample Populations

Standard criteria established by the American Thoracic Society/European Respiratory Society were used to diagnose IPF (15). Cases with explanations for the development of fibrotic interstitial pneumonias (e.g., infections, systemic diseases, or relevant exposures) were excluded. Diagnoses were made based on multidisciplinary review integrating clinical, radiologic, and pathologic data at the sites where samples were obtained. Please see the online supplement and Table 1 for details regarding discovery and validation cohorts (16).

Lung tissue cohort for gene expression.

The Lung Tissue Research Consortium (LTRC) is a resource created by the NHLBI to provide human lung tissue and DNA to qualified investigators for use in research. Human lung tissue samples were obtained from the LTRC and National Jewish Health (NJH). A total of 201 control lung tissue samples (108 from the LTRC, 93 from NJH) and 334 IPF lung tissue samples from biopsy samples (188 from the LTRC, 146 from NJH) were used for gene expression. Control samples from the LTRC are pathologically normal areas of tissue obtained during tumor resection or biopsies for malignancy. Control samples from NJH are obtained from lungs of deceased donors deemed unsuitable for transplantation.

Table 1. Characteristics of Discovery and Validation Study Cohorts

Characteristics	Control	IPF	P Value
Discovery study cohort			
n	228	230	
Male sex, n (%)	151 (69)	159 (66)	0.51
Ever smoker, %	100	78	<0.0001
Age, yr, mean (95% CI)	64.5 (63.7–65.4)	65.4 (64.2–66.5)	0.11
Validation study cohort			
n	936	936	
Male sex, n (%)	653 (69.99)	647 (69.95)	0.98
Ever smoker, %	100	61.92	<0.0001
Age, yr, mean (95% CI)	62.73 (62.23–63.24)	64.97 (64.34–65.60)	<0.0001

Definition of abbreviations: CI = confidence interval; IPF = idiopathic pulmonary fibrosis. Clinical characteristics of subjects whose DNA was sequenced and genotyped. All control subjects were documented to be either current or former smokers (“ever smoker”), whereas 78% of sequenced subjects with IPF and 61.92% of genotyped subjects with IPF reported a smoking history. Sequenced subjects were of similar age and were predominantly male. Genotyped subjects were also predominantly male; genotyped subjects with IPF were slightly older than the control subjects. Statistical tests were performed using two-tailed *t* tests for means and proportions.

DNA was extracted from peripheral blood of study subjects.

Nucleic acids, tissues, and clinical data were obtained under protocols approved by NJH and the University of Colorado Anschutz Medical Campus (Aurora, CO; protocol NJH HS-1441a).

Identifying Variants in DSP Associated with IPF

We sequenced across the *DSP* locus (Chr6: 7,370,061–7,606,946) on 230 IPF samples and 228 control samples with a Nimblegen custom SeqCAP panel for the *DSP* locus using the Illumina HiSeq platform (Illumina, Inc., San Diego, CA) through the NHLBI Resequencing and Genotyping Service. The

DSP locus was defined by sequentially extending on either side from the most significant SNP from the genome-wide association study discovery (4) if there was another SNP with P value less than 1×10^{-4} within 50 kb of the current boundary. If either end of a resulting region was within 20 kb of a gene, the bounds were extended 20 kb beyond the end of that gene, resulting in a 237-kb region (Chr6: 7,370,061–7,606,946). Illumina sequence reads were aligned to the human genome (HG-19), and variants were called using the Genome Analysis Toolkit (GATK) (17). Variant detection and genotyping were performed using the UnifiedGenotyper (UG) from GATK (v1.6-11-g3b2fab9). Variant data were

flagged using the filtration walker (GATK) to mark sites of lower quality/false positives (e.g., low-quality scores [Q50], allelic imbalance [ABHet 0.75], long homopolymer runs [HRun > 3], low-quality by depth [QD < 5]).

Given the sample size, we were able only to test for association with IPF and individual common variants (>5% minor allele frequency). We initially identified all SNPs in the *DSP* locus associated with IPF using logistic regression and adjusting for age and sex. At this screening stage, we used an α of 0.005 as a practical compromise between power and type 1 error rate in anticipation of using a strict Bonferroni correction in the replication

Table 2. *DSP* Variants Associated with Idiopathic Pulmonary Fibrosis and Grouped by Linkage Disequilibrium

SNP	Minor Allele	Base Pair Change Implication	Case Minor Allele Frequency	Control Minor Allele Frequency	P Value	CADD Score	OR
Group 1							
rs72825047	G	9,449 bp upstream of <i>DSP</i> start site	0.048	0.1096	0.0007	8.77	0.4
rs72825054*	C	3,591 bp upstream of <i>DSP</i> start site	0.046	0.1118	0.0003	9.33	0.4
rs72825058	T	Intron 1	0.044	0.1075	0.0005	5.33	0.4
Group 2							
rs874646	G	Intron 1	0.14	0.23	0.002	3.71	0.6
rs2744362	C	Intron 1	0.14	0.23	0.002	4.77	0.6
rs2008748	T	Intron 1	0.14	0.23	0.001	2.79	0.6
rs2744365	G	Intron 1	0.14	0.23	0.002	4.88	0.6
rs2237102	C	Intron 1	0.14	0.23	0.002	4.28	0.6
rs2237103	G	Intron 1	0.14	0.23	0.002	6.03	0.6
rs2237104	T	Intron 1	0.14	0.23	0.002	4.63	0.6
rs2078438	A	Intron 1	0.14	0.23	0.002	2.81	0.6
rs2744368	T	Intron 1	0.14	0.23	0.002	0.32	0.6
rs2744369	T	Intron 1	0.14	0.23	0.001	2.62	0.6
rs2744370	C	Intron 1	0.14	0.23	0.002	8.33	0.6
rs2744371*	C	Intron 1	0.14	0.23	0.002	11.76	0.6
rs2744372	G	Intron 1	0.14	0.23	0.002	2.91	0.6
rs2744374	T	Intron 1	0.14	0.23	0.002	1.17	0.6
Group 3							
rs2744373*	A	Intron 1	0.10	0.16	0.004	3.39	0.5
rs2744375	T	Intron 1	0.10	0.16	0.004	0.18	0.5
Group 4							
rs17133512*	A [†]	Exon 1—likely benign	0.19	0.12	0.004	15.51	1.7
rs145125286	— [‡]	Intron 1	0.10	0.16	0.003	7.30	0.5
rs2008756	G	Intron 1	0.14	0.23	0.0009	0.96	0.6
rs2237105	A	Intron 1	0.14	0.23	0.002	0.71	0.6

Definition of abbreviations: CADD = combined annotation-dependent depletion; *DSP* = desmoplakin; OR = odds ratio; SNP = single-nucleotide polymorphism. Data from sequencing of 240 subjects with idiopathic pulmonary fibrosis and 240 control subjects around the chromosome (Chr) 6p24 locus

(Chr6: 7,370,061–7,606,946) revealed that variants in the promoter region and intron 1 of *DSP* were associated with disease status. Listed here are variants within 10 kb of the *DSP* start site with strong association ($P < 0.005$) with idiopathic pulmonary fibrosis diagnosis. The variants are grouped by their linkage disequilibrium (LD) with one another; variants with $R^2 > 0.8$ are grouped together (Groups 1–4). Bold variants are those that were further examined from each group based on CADD score and LD information. Group 4 variants (rs17133512, rs145125286, rs2008756, and rs2237105) did not have available LD data in the existing Centre d'Etude du Polymorphisme Humain databases (italicized), and therefore LD was examined within our data generated from sequencing. CADD scores estimate deleteriousness of single-nucleotide variants (20). Base pair change implications are with respect to *DSP* based on UCSC Genome Browser (<https://genome.ucsc.edu/index.html>) and dbSNP (<http://www.ncbi.nlm.nih.gov/snp>).

*Variants that were picked from each group to be genotyped using predesigned or custom quantitative polymerase chain reaction assays in 940 subjects with idiopathic pulmonary fibrosis and 940 control subjects.

[†]This variant is an insertion (A).

[‡]The minor allele is a deletion (C).

Table 3. Linkage Disequilibrium between Variants within Sequencing Data

SNP 1	SNP 1 Chr6 Position	SNP 2	SNP 2 Chr6 Position	R^2
rs145125286	7545870	rs2744373	7554527	0.99
rs145125286	7545870	rs2744371	7554174	0.65
rs145125286	7545870	rs72825054	7538279	0.49
rs2008756	7551792	rs2744371	7554174	0.94
rs2008756	7551792	rs2744373	7554527	0.62
rs2008756	7551792	rs72825054	7538279	0.31
rs2237105	7552555	rs2744371	7554174	0.99
rs2237105	7552555	rs2744373	7554527	0.66
rs2237105	7552555	rs72825054	7538279	0.31

Definition of abbreviations: Chr6 = chromosome 6; SNP = single-nucleotide polymorphism. Three SNPs that did not have reported R^2 values available were assessed for linkage disequilibrium with the other desmoplakin variants to be genotyped in this study based on sequencing data. Bold rows indicate analysis in which R^2 values were high (>0.8) with another variant.

phase. To focus analysis on *DSP*, sequence data greater than 10 kb from the *DSP* transcription start site were excluded.

To identify a parsimonious group of variants to take forward to the validation stage while still capturing the variation present, we selected one SNP among groups of significant SNPs in high linkage disequilibrium (LD) with each other ($R^2 > 0.8$ [18]) based on the Centre d'Etude du Polymorphisme Humain (Utah residents with ancestry from northern and western Europe) (CEU) 1,000 genomes population (19) using the SNP Annotation and Proxy Search (Broad Institute, <http://www.broadinstitute.org/mpg/snp/>). The variants in each LD bin were examined using combined annotation-dependent

depletion (CADD) scores to estimate relative pathogenicity based on an integration of diverse annotations of a genetic variant's potential biological consequences into a single value; higher CADD scores suggest higher pathogenicity (20). One variant from each LD bin was selected based on the highest CADD score, and genotyping was performed using pre-existing or custom TaqMan genotyping assays (Life Technologies, Carlsbad, CA) (Table 2). Of those variants without reported CEU LD data available, only one variant was genotyped (rs17133512) due to its high CADD score (15.51). The other three were not genotyped (Table 2) given the high estimated R^2 values with the other variants to be genotyped in sequencing data

(Table 3), as the high LD between the variants suggested that one would be a marker for the others.

Custom TaqMan assays were designed for rs72825054, rs17133512, and rs2744371 by Life Technologies. Predesigned Taqman genotyping assays (C_15921733_10 and C_16167921_10; Life Technologies) were used to genotype cases and controls at rs2744373 and rs2076295 using standard protocol. DNA samples from our validation cohort (936 subjects with IPF, 936 control subjects) were genotyped for all five variants (rs72825054, rs17133512, rs2744373, rs2744371, and rs2076295). Sample duplicates were included on each plate to ensure accuracy and reproducibility.

Association between IPF and the minor allele of the replication variants was assessed using logistic regression assuming a log-additive effect of the SNP, adjusting for sex and age. A P value less than 0.01 was considered statistically significant, corresponding to conservative Bonferroni adjustment for multiple testing (five variants).

To confirm that CEU population-derived LD estimates corresponded approximately to the LD that actually existed between these variants in our IPF and control validation cohorts, we estimated R^2 values for these variants within our study populations using the PLINK software (version 1.07; <http://pngu.mgh.harvard.edu/purcell/plink/>) (21).

DSP Gene Expression in Human Lung Tissue

Gene expression was performed on RNA extracted from human lung tissues using the mirVana kit (Ambion/Life Technologies, Carlsbad, CA). Complementary DNA (cDNA) was synthesized from 80 ng RNA using SuperScript III (Life Technologies) and diluted 1 to 10 with nuclease-free water. cDNA was used with real-time polymerase chain reaction (PCR) gene expression assay for *DSP* (Hs00189422_m1; Life Technologies) normalized to glyceraldehyde phosphate dehydrogenase (GAPDH; Hs03929097_g1; Life Technologies) using standard TaqMan protocol (TaqMan Fast Advanced Master Mix, 4444602; Applied Biosystems, Carlsbad, CA) and Vii7 real-time PCR system (Thermo Fisher Scientific, Waltham, MA). Relative gene expression was determined through the $\Delta\Delta$ cycle threshold (C_t) method (22).

Figure 1. Desmoplakin (*DSP*) variants and association with idiopathic pulmonary fibrosis (IPF). This figure illustrates the relative position in the *DSP* gene of the variants genotyped in this study. rs2076295 is the intron 5 variant previously described by Fingerlin and colleagues (4) to be associated with fibrosing idiopathic interstitial pneumonias, of which IPF is the most common. The variants in red were significantly associated with IPF when controlling for sex and age ($P < 0.01$). Odds ratios (ORs) are presented with 95% confidence intervals noted in parentheses.

Figure 2. Desmolapkin (*DSP*) variants and association with idiopathic pulmonary fibrosis (IPF) when controlling for rs2076295. This figure illustrates the relative position in the *DSP* gene of the variants genotyped in this study. rs2076295 is the intron 5 variant previously described by Fingerlin and colleagues (4) to be associated with fibrosing idiopathic interstitial pneumonias, of which IPF is the most common. The variant in red (rs2744371) is nominally associated with IPF when controlling for age, sex, and rs2076295 genotype ($P < 0.05$). Odds ratios (ORs) are presented with 95% confidence intervals noted in parentheses.

C_t values, the number of PCR cycles required for fluorescent signal to exceed background levels, are inversely proportional to the amount of target nucleic acid present in the sample. ΔC_t values were calculated by subtracting the C_t value for *GAPDH* from the C_t value for *DSP* for a sample. ΔC_t values were plotted and noted to be approximately normally distributed,

both in aggregate and among IPF and control diagnosis groups. We first tested for differential gene expression in the lung between 334 subjects with IPF and 201 control subjects using two-tailed Student's t test comparing ΔC_t values between groups. We similarly tested for differential expression using a one-way analysis of variance (ANOVA) for genotype at rs2076295. Differential gene expression

for genotype at rs2744371 was assessed with variant subjects grouped together given the small number of homozygous variant subjects in our sample ($n = 27$). A P value less than 0.025 was considered statistically significant when assessing individual SNP associations with expression, corresponding to Bonferroni correction for two tests. We used linear regression analyses to explore the relationship between *DSP* expression and genotype while controlling for age, diagnosis, smoking (ever/never and pack-years), and sex. ΔC_t for *DSP* was regressed on the continuous variables for age and rs2076295 genotype, as well as categorical variables for sex and diagnosis (IPF vs. control), and rs2744371 variant. A P value less than 0.025 was considered statistically significant when examining the effect of an individual variant on expression corresponding to Bonferroni correction for two variants tested. All statistical analyses were computed in GraphPad Prism (v. 5.01; GraphPad Software, Inc., La Jolla, CA) and Stata 14 (StataCorp LP, College Station, TX).

DSP Protein Expression in Human Lung Tissue
Deidentified, formalin-fixed, paraffin-embedded lung tissue from subjects with IPF and normal control subjects were obtained from the LTRC. Staining was performed on 5- μ m sections of lung tissues

Table 4. Linkage Disequilibrium between Variants within Genotyping Cohort, in Comparison to Centre d'Etude du Polymorphisme Humain Databases

SNP 1	SNP 1 IPF OR	SNP 1 MAF	SNP 2	CEU R^2	SNP 1 Control MAF	Control R^2	SNP 1 IPF MAF	IPF R^2
rs72825054	0.90	0.11	rs2744373	0.51	0.10	0.50	0.08	0.45
rs72825054	0.90	0.11	rs2744371	0.31	0.10	0.29	0.08	0.29
rs72825054	0.90	0.11	rs2076295	0.06	0.10	0.07	0.08	0.00
rs72825054	0.90	0.11	rs17133512	N/A	0.10	0.02	0.08	0.02
rs17133512	1.13	N/A	rs2744371	N/A	0.13	0.05	0.15	0.02
rs17133512	1.13	N/A	rs2744373	N/A	0.13	0.05	0.15	0.02
rs17133512	1.13	N/A	rs2076295	N/A	0.13	0.03	0.15	0.00
rs2744371	0.83	0.28	rs2076295	0.08	0.25	0.13	0.21	0.01
rs2744371	0.83	0.28	rs2744373	0.60	0.25	0.56	0.21	0.60
rs2744373	0.81	0.19	rs2076295	0.02	0.16	0.05	0.13	0.00

Definition of abbreviations: CEPH = Centre d'Etude du Polymorphisme Humain; CEU = Centre d'Etude du Polymorphisme Humain (Utah residents with ancestry from northern and western Europe); IPF = idiopathic pulmonary fibrosis; MAF = minor allele frequency; N/A = not available; OR = odds ratio; SNP = single-nucleotide polymorphism. To confirm the independence of the variants examined in this study, R^2 values were calculated between the genotyped SNPs. SNP 1 IPR OR reports the OR for IPF for the minor allele at that variant. SNP 1 MAF refers to reported minor allele frequency for the variant in the 1,000 Genomes European cohort (<http://www.broadinstitute.org/mpg/snp/snp/documentation.html>). CEU R^2 column reports previously reported values for these variants in a sample of Utah residents with Northern and Western European ancestry collected by the CEPH and reported as a HapMap population (CEU). Control R^2 and IPF R^2 columns refer to the linkage disequilibrium (LD) between these variants within the control and IPF sample populations genotyped in this study, illustrating that the LD observed between our variants in our study cohorts was similar to the LD observed between the same variants in the CEU population. There are no R^2 values greater than 0.6, suggesting relatively low LD between these variants, well below our cutoff for this study of 0.8.

Table 5. Clinical Characteristics of Lung Tissue Donors

Characteristics	Control (n = 201)	IPF (n = 334)	P Value
Age, yr, mean (95% CI)	56.2 (53.9–58.5)	63.1 (62.2–64.1)	<0.0001
Male sex, n (%)	102 (50.5)	227 (67.8)	0.0001
Never smoker, n (%)	63 (34.2)	108 (31.3)	0.50
Pack-years smoking, mean (95% CI)	17.3 (14.0–20.6)	33.91 (27.1–40.7)	0.02
% Predicted FVC, mean (95% CI)	—	63% (60.8–65.2%)	—
% Predicted DL _{CO} , mean (95% CI)	—	37% (33.6–40.5%)	—
DSP ΔC _t , mean (95% CI)	5.1 (4.9–5.3)	3.9 (3.8–4.1)	<0.0001

Definition of abbreviations: CI = confidence interval; C_t = cycle threshold; DL_{CO} = diffusing capacity of the lung for carbon monoxide; DSP = desmoplakin; IPF = idiopathic pulmonary fibrosis.

Clinical characteristics of the lung tissue subjects used for expression and genotyping analysis. Statistical tests performed using two-sample, two-tailed *t* tests for means and for proportions, where appropriate. Pulmonary function data were only available on a subset of IPF tissue donors (n = 234).

using standard protocols (please see the online supplement).

Results

DSP Variants and IPF

As shown in Table 2, 23 DSP variants meeting criteria for statistically significant association with IPF were found in the initial sequencing of the discovery cohort. Although the sequencing results revealed numerous DSP variants associated with disease, no variants in either *SNRNP48* or

RIOK1, the other genes in the sequenced locus (Chr6: 7,315,060–7,551,945), were significantly associated with IPF. Given the high recombination rates observed at sites flanking DSP, this finding is not surprising (4). Variants were grouped together, as described in the METHODS (Table 2). Variants rs17133512, rs145125286, rs2008756, and rs2237105 did not have data available in public databases regarding their LD with other variants on our list. The observed LD between these variants in our sequenced dataset is described in Table 3.

The results from the logistic regression analyses of the five genotyped SNPs (rs72825054, rs17133512, rs2744373, rs2744371, and rs2076295) in the validation cohort controlling for sex and age are depicted in Figure 1. Intron 1 variant rs2744371, as well as previously described IPF-associated intron 5 variant rs2076295, are significantly associated with IPF. The minor allele of rs2076295 confers increased risk of IPF (odds ratio [OR] = 1.36, 95% confidence interval [CI] = 1.19–1.56, *P* < 0.001); this effect is similar and is statistically significant (OR = 1.32, 95% CI = 1.14–1.52, *P* < 0.001) when controlling for both of the intron 1 SNP genotypes. The minor allele at rs2744371 (C) is protective (OR = 0.77, 95% CI = 0.66–0.91, *P* = 0.002, respectively; Figure 1) and remains significantly protective for IPF after controlling for the effect of rs2076295 (OR = 0.83, 95% CI = 0.71–0.98, *P* = 0.03; Figure 2). The effect of rs2076295 remains constant (OR = 1.32, 95% CI = 1.14–1.52, *P* < 0.001) when controlling for all intron 1 genotypes.

To assess whether variants in intron 1 were in strong LD with rs2076295 in subjects with IPF when compared with control subjects or when compared with reported CEU data, we examined the *R*² values for these variants within our genotyping data. We found that the variants were not in significant LD with one another (Table 4). All variants were in Hardy-Weinberg equilibrium in our validation cohorts.

DSP Expression in Lung

Disease status and age are significant drivers of DSP expression. Baseline characteristics of the lung tissue cohort subjects are listed in Table 5. Subjects with IPF were older than control subjects (mean age = 63.1 yr, 95% CI = 62.20–64.1 vs. 56.2 yr, 95% CI = 53.9–58.5), and had a higher average reported pack-year smoking history (33.9 pack-years, 95% CI = 27.1–40.7 vs. 17.3 pack-years, 95% CI = 14.0–20.6 for control subjects).

There was a 2.3-fold increase (95% CI = 1.91–2.71) in DSP gene expression in IPF lung compared with control lung tissue (*P* < 0.0001; Figure 3). DSP expression in the lung was not affected by age in all samples (*P* = 0.57) or in IPF samples (*P* = 0.87), but decreased with age in control lungs (*P* = 0.001) (Figures 4A and 4B).

Figure 3. Relative desmoplakin (DSP) expression in human lung tissue by quantitative polymerase chain reaction, grouped by diagnosis. Idiopathic pulmonary fibrosis (IPF) lung tissue had significantly higher DSP expression than control lung. Statistical tests (two-tailed Student's *t* test) were performed on Δ cycle threshold (C_t) values (DSP C_t values normalized to housekeeping gene *GAPDH*), and relative expression is visually depicted as 10 – ΔC_t. Means are indicated within each group with a horizontal line; error bars indicate SEM. UIP = usual interstitial pneumonia.

Figure 4. Desmoplakin (*DSP*) expression and age. In control lung samples, *DSP* expression falls with age ($P = 0.004$) (A), but in idiopathic pulmonary fibrosis lungs, *DSP* expression remains constant with increasing age (B). Data points are plotted as green circles; the line of best fit is also plotted with 95% confidence interval indicated by gray areas. Relative *DSP* expression is plotted on the y-axis ($10 - \Delta$ cycle threshold [C_t]), and age is plotted on the x-axis. Statistical analysis was performed using linear regression on ΔC_t values. CI = confidence interval.

DSP expression remained significantly higher in IPF lungs compared with control lungs after controlling for age, smoking, and sex ($P < 0.001$).

DSP expression and DSP intron 5 variant rs2076295. The number of copies of

the minor allele (G) at rs2076295 is associated with decreased whole-lung *DSP* expression in IPF lung tissue (ANOVA $P = 0.004$) and in control lung tissue (ANOVA $P < 0.0001$) (Figures 5A and 5B). In control lung tissue, there was a 2.63-fold decrease

(95% CI = 1.69–3.84-fold decrease, $P < 0.0001$) in *DSP* expression in G/G subjects compared with T/T. In IPF lung tissue, there was a 1.54-fold decrease (95% CI = 1.17–2.0-fold decrease, $P = 0.002$) in *DSP* expression in G/G subjects compared with T/T.

For all lung tissue specimens, age, sex, smoking history, and pack-years were not significantly associated with decreased *DSP* expression, but genotype at rs2076295 ($P = 0.002$) remained significantly associated with decreased gene expression after adjustment for these factors. In control lung tissue, rs2076295 genotype was significantly associated with decreased *DSP* expression ($P < 0.006$) controlling for age, pack-years of smoking, and sex. In IPF lung tissue, the rs2076295 variant remained significantly associated with decreased *DSP* expression ($P = 0.004$) when controlling for age, sex, and pack-years of smoking.

DSP expression and DSP intron 1 variant (rs2744371). We examined the relationship between *DSP* expression in the lung and rs2744371, the only intron 1 variant that was associated with IPF when controlling for genotype at rs2076295. Due to the small number of rs2744371 homozygous variant subjects ($n = 27$) (Table 2), data were analyzed for the rs2744371 variant as a dominant allele. *DSP* expression did not vary significantly between subjects with variant allele at rs2744371 in control subjects ($P = 0.21$) or subjects with IPF ($P = 0.27$). When controlling for age, sex, and pack-years of smoking, in control lung tissue, the rs2744371 variant was not associated with *DSP* gene expression ($P = 0.12$); similarly, there was no association between rs2744371 genotype and *DSP* expression in IPF lung tissue ($P = 0.42$).

Airway cells express DSP I/II in both normal and IPF lung. Staining for DSP in the normal and IPF lung tissue suggests that the protein is most strongly expressed in both proximal (Figure 6C) and distal (Figures 6A, 6C, and 6D) airways of normal lung. IPF lung tissue shows similar staining in both patches of normal lung parenchyma as well as in diseased areas (Figures 6B and 6E), with minimal staining in normal alveolar tissue (Figures 6A and 6D). Cystic structures within areas of fibrotic lung are lined with epithelial cells that stain positive for DSP (Figure 6F). There is no significant DSP staining in

Figure 5. Relative desmoplakin (*DSP*) expression by rs2076295 genotype. *DSP* expression as measured by quantitative polymerase chain reaction (normalized to *GAPDH*) varied by rs2076295 genotype. When lung samples were divided into idiopathic pulmonary fibrosis (A) and control (B) groups, there were significant differences between *DSP* gene expression by genotype at rs2076295. Means are indicated within each group with a horizontal line; error bars indicate SEM. Statistical tests were performed on raw Δ cycle threshold (C_t) values for each group; relative *DSP* expression is depicted visually as $10 - \Delta C_t$. ANOVA = analysis of variance.

the normal alveolar epithelia in control or IPF lung.

Discussion

Our results confirm that variants in *DSP* are associated with IPF, that *DSP* expression is higher in IPF lung when compared with healthy control subjects, and that intron 5 variant rs2076295 is associated with decreased gene expression, adding to evidence that this gene plays a

role in disease etiology. *DSP* is part of the desmosome, a structure that mechanically connects cells, stabilizes tissue, and is essential in cell migration, proliferation, and differentiation (5–7). Consequently, *DSP* is critical to cell–cell adhesion (8, 9), wound repair, and epithelial barrier function (10), and, given our findings, may prove biologically relevant to the development of IPF.

Analysis of lung tissue from a large number of cases and controls shows that *DSP* expression is higher in IPF

lung when compared with healthy controls. *DSP* lung expression falls with age in normal lung, but is not influenced by age in IPF lung. The increased *DSP* expression observed in IPF lung may be a consequence of robust and persistent epithelial response to injury in diseased pulmonary tissue. The correlation between decreased *DSP* expression and age also suggests that impaired pulmonary epithelial response to injury associated with age (23) may be due to decreased expression of cell–cell adhesion molecules, such as *DSP*.

We confirmed the association between intron 5 variant rs2076295 and IPF, and, in a large sample size, showed that, for each copy of the minor allele (G), *DSP* expression falls significantly in control and IPF lung tissue. The strength of this association between rs2076295 genotype and gene expression supports the hypothesis that this variant is important in gene regulation. rs2076295 is located within a DNaseI hypersensitivity cluster and within the binding site for PU.1, and a transcription factor associated with alternative splicing, alternative macrophage activation, and airway remodeling (11, 12). This study raises the possibility that PU.1 plays a role in *DSP* gene expression, or that another transcription factor binding in this region may be important in regulating *DSP*.

Although rs2076295 is strongly associated with IPF disease risk, the disease-associated minor allele is associated with decreased *DSP* expression, even in IPF lung. Individuals with the rs2076295 variant could comprise a distinct subpopulation within IPF patients—whereas IPF lung in general appears to be producing higher amounts of *DSP*, those with the rs2076295 variant are expressing less *DSP*. For those with the rs2076295 variant, decreased *DSP* expression in the setting of disease could be a marker of inadequate or disordered epithelial repair. However, our results suggest that, although the intron 1 *DSP* variant rs2744371 is associated with IPF, there does not appear to be an association between this variant and *DSP* expression in the lung tissue of patients with IPF.

Immunohistochemical staining for *DSP* in normal and IPF lung tissue illustrates strong staining of both proximal and distal airways. All observed airway epithelia stain strongly for *DSP*, as does the lining of cystic structures characteristic of

Figure 6. Immunohistochemistry of desmoplakin (DSP) in normal and fibrotic idiopathic pulmonary fibrosis (IPF) lungs. Immunohistochemical staining of formalin-fixed, paraffin-embedded human lung tissue with rabbit polyclonal antibody against DSP I/II (Santa Cruz Biotechnology, Inc., Dallas, TX). Normal lung stained for DSP shows strong staining along the airway epithelia (A), in the trachea (C), and in the more distal airways (D) without significant alveolar staining (A). IPF lung stained concurrently shows strong staining of airway epithelia in both normal-appearing and fibrotic regions (B and E). Cystic areas of the fibrotic lung are lined with epithelial cells that strongly stain positive for DSP I/II (F).

IPF. Other studies have also suggested that airway epithelia may be playing a role in pulmonary fibrosis. A common polymorphism in the promoter region of *MUC5B*, the gene encoding Mucin 5B (*MUC5B*), a mucin important in airway defense (24), is strongly associated with IPF and with increased gene expression (25).

MUC5B is expressed by airway cells in normal human lung, and also by epithelial cells lining the honeycomb cyst, a characteristic feature of IPF (26). Those with the *MUC5B* promoter polymorphism have increased accumulation of *MUC5B*-positive mucus in their distal airways (27), suggesting that alterations in mucin

amount or clearance may be leading distal airway injury, aberrant response to injury, and ultimately to fibrosis. Although many studies have focused on the role of type II alveolar cells in lung fibrosis, future investigation should focus on the airway epithelia in IPF, as both *MUC5B* and *DSP* are expressed by the airway epithelia in normal and fibrotic lung. These strong genetic associations, as well as the knowledge that bronchoalveolar stem cells are important in lung repair (28), emphasize the role that airway or airway-derived cells may be critical to the response to pulmonary injury.

Our study has several limitations. A total of 228 control subjects and 230 subjects with IPF were sequenced to determine our initial list of *DSP* variants. Sequencing a larger number of cases and controls across this locus would yield greater power to detect important *DSP* variants in intron 1 and other gene regions. This would be particularly important if IPF is a genetically heterogeneous disease. Furthermore, gene expression studies were performed on RNA extracted from whole lung—although *DSP* expression, as well as our group's other findings regarding mucins in IPF, suggest a role for the airway epithelium in disease pathogenesis, definitive conclusions cannot be drawn in terms of what cell type(s) are driving findings of differential *DSP* expression between groups.

Our findings reveal intron 1 variants in *DSP* associated with IPF and confirm the robust association of intron 5 variant rs2076295 with disease. The association of rs2076295 with differential *DSP* expression in lung tissue suggests that it affects gene regulation and that differential *DSP* expression plays a role in disease etiology in a subset of patients with pulmonary fibrosis, and implicates the airway epithelium in the pathogenesis of IPF. ■

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

References

- King TE Jr, Albera C, Bradford WZ, Costabel U, du Bois RM, Leff JA, Nathan SD, Sahn SA, Valeyre D, Noble PW. All-cause mortality rate in patients with idiopathic pulmonary fibrosis: implications for the design and execution of clinical trials. *Am J Respir Crit Care Med* 2014;189:825–831.
- King TE Jr, Tooze JA, Schwarz MI, Brown KR, Cherniack RM. Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. *Am J Respir Crit Care Med* 2001;164:1171–1181.
- Olson AL, Swigris JJ, Lezotte DC, Norris JM, Wilson CG, Brown KK. Mortality from pulmonary fibrosis increased in the United States from 1992 to 2003. *Am J Respir Crit Care Med* 2007;176:277–284.

4. 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.
5. Asimaki A, Saffitz JE. Remodeling of cell–cell junctions in arrhythmogenic cardiomyopathy. *Cell Commun Adhes* 2014;21:13–23.
6. Rampazzo A, Nava A, Malacrida S, Beffagna G, Bauce B, Rossi V, Zimbello R, Simionati B, Basso C, Thiene G, *et al.* Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy. *Am J Hum Genet* 2002;71:1200–1206.
7. Uzunçu A, Norgett EE, Dindar A, Uyguner O, Nisli K, Kayserili H, Sahin SE, Dupont E, Severs NJ, Leigh IM, *et al.* Loss of desmoplakin isoform I causes early onset cardiomyopathy and heart failure in a Naxos-like syndrome. *J Med Genet* 2006;43:e5.
8. Ou J, Lowes C, Collinson JM. Cytoskeletal and cell adhesion defects in wounded and Pax6^{+/-} corneal epithelia. *Invest Ophthalmol Vis Sci* 2010;51:1415–1423.
9. McAleer MA, Pohler E, Smith FJ, Wilson NJ, Cole C, MacGowan S, Koetsier JL, Godsel LM, Harmon RM, Gruber R, *et al.* Severe dermatitis, multiple allergies, and metabolic wasting syndrome caused by a novel mutation in the N-terminal plakin domain of desmoplakin. *J Allergy Clin Immunol* 2015;1:1–9.
10. Huber O, Petersen I. 150th anniversary series: desmosomes and the hallmarks of cancer. *Cell Commun Adhes* 2015;22:15–28.
11. Guillouf C, Gallais I, Moreau-Gachelin F. Spi-1/PU.1 oncoprotein affects splicing decisions in a promoter binding–dependent manner. *J Biol Chem* 2006;281:19145–19155.
12. Qian F, Deng J, Lee YG, Zhu J, Karpurapu M, Chung S, Zheng JN, Xiao L, Park GY, Christman JW. The transcription factor PU.1 promotes alternative macrophage polarization and asthmatic airway inflammation. *J Mol Cell Biol* 2015;7:557–567.
13. 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.
14. Yang L, Chen Y, Cui T, Knösel T, Zhang Q, Albring KF, Huber O, Petersen I. Desmoplakin acts as a tumor suppressor by inhibition of the Wnt/β-catenin signaling pathway in human lung cancer. *Carcinogenesis* 2012;33:1863–1870.
15. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, *et al.*; ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183:788–824.
16. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, Curran-Everett D, Silverman EK, Crapo JD. Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010;7:32–43.
17. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
18. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–120.
19. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA, De La Vega FM, Donnelly P, *et al.*; 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–1073.
20. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–315.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575.
22. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 2008;3:1101–1108.
23. Yin L, Zheng D, Limmon GV, Leung NHN, Xu S, Rajapakse JC, Yu H, Chow VTK, Chen J. Aging exacerbates damage and delays repair of alveolar epithelia following influenza viral pneumonia. *Respir Res* 2014;15:116.
24. 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.
25. 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.
26. Seibold MA, Smith RW, Urbanek C, Groshong SD, Cosgrove GP, Brown KK, Schwarz MI, Schwartz DA, Reynolds SD. The idiopathic pulmonary fibrosis honeycomb cyst contains a mucociliary pseudostratified epithelium. *PLoS One* 2013;8:e58658.
27. Nakano Y, Evans CM, Yang IV, Schwarz MI, Schwartz DA. MUC5B promoter variant Rs35705950 effects MUC5B expression and this effect is localized to the distal airways in IPF [abstract]. *Am J Respir Crit Care Med* 2015;191:A3446.
28. Kim CFB, Jackson EL, Woolfenden AE, Lawrence S, Babar I, Vogel S, Crowley D, Bronson RT, Jacks T. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* 2005;121:823–835.