

# Integrated Genomics Reveals Convergent Transcriptomic Networks Underlying Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis

Rebecca L. Kusko<sup>1\*</sup>, John F. Brothers II<sup>1\*</sup>, John Tedrow<sup>2</sup>, Kusum Pandit<sup>2</sup>, Luai Huleihel<sup>2</sup>, Catalina Perdomo<sup>1</sup>, Gang Liu<sup>1</sup>, Brenda Juan-Guardela<sup>3</sup>, Daniel Kass<sup>2</sup>, Sherry Zhang<sup>1</sup>, Marc Lenburg<sup>1</sup>, Fernando Martinez<sup>4</sup>, John Quackenbush<sup>5</sup>, Frank Scirba<sup>2</sup>, Andrew Limper<sup>6</sup>, Mark Geraci<sup>7</sup>, Ivana Yang<sup>7</sup>, David A. Schwartz<sup>7</sup>, Jennifer Beane<sup>1</sup>, Avrum Spira<sup>1‡</sup>, and Naftali Kaminski<sup>3‡</sup>

<sup>1</sup>Computational Biomedicine, Boston University School of Medicine, Boston, Massachusetts; <sup>2</sup>Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania; <sup>3</sup>Pulmonary, Critical Care, and Sleep Medicine, Yale School of Medicine, New Haven, Connecticut; <sup>4</sup>Pulmonary & Critical Care Medicine, University of Michigan, Ann Arbor, Michigan; <sup>5</sup>Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; <sup>6</sup>Mayo Clinic, Rochester, Minnesota; and <sup>7</sup>Pulmonary Sciences and Critical Care Medicine, UC Denver, Denver, Colorado

ORCID IDs: 0000-0003-1467-0489 (A.S.); 0000-0001-5917-4601 (N.K.).

## Abstract

**Rationale:** Despite shared environmental exposures, idiopathic pulmonary fibrosis (IPF) and chronic obstructive pulmonary disease are usually studied in isolation, and the presence of shared molecular mechanisms is unknown.

**Objectives:** We applied an integrative genomic approach to identify convergent transcriptomic pathways in emphysema and IPF.

**Methods:** We defined the transcriptional repertoire of chronic obstructive pulmonary disease, IPF, or normal histology lungs using RNA-seq (n = 87).

**Measurements and Main Results:** Genes increased in both emphysema and IPF relative to control were enriched for the p53/hypoxia pathway, a finding confirmed in an independent cohort using both gene expression arrays and the nCounter Analysis System (n = 193). Immunohistochemistry confirmed overexpression of *HIF1A*, *MDM2*, and *NFKB1B* members of this pathway in tissues from

patients with emphysema or IPF. Using reads aligned across splice junctions, we determined that alternative splicing of p53/hypoxia pathway-associated molecules *NUMB* and *PDGFA* occurred more frequently in IPF or emphysema compared with control and validated these findings by quantitative polymerase chain reaction and the nCounter Analysis System on an independent sample set (n = 193). Finally, by integrating parallel microRNA and mRNA-Seq data on the same samples, we identified *MIR96* as a key novel regulatory hub in the p53/hypoxia gene-expression network and confirmed that modulation of *MIR96* *in vitro* recapitulates the disease-associated gene-expression network.

**Conclusions:** Our results suggest convergent transcriptional regulatory hubs in diseases as varied phenotypically as chronic obstructive pulmonary disease and IPF and suggest that these hubs may represent shared key responses of the lung to environmental stresses.

**Keywords:** network; COPD; ILD; IPF; transcriptome

Chronic lung diseases affect a significant portion of the population and account for more than 100,000 deaths a year (1). Although most of these deaths can

be attributed to chronic obstructive pulmonary disease (COPD), the major smoking-induced lung disease, more than 15,000 deaths a year can be attributed to

idiopathic pulmonary fibrosis (IPF), a relentless, nearly always fatal fibrotic lung disease also associated with smoking (2). COPD, defined by the Global Initiative for

(Received in original form October 18, 2015; accepted in final form March 27, 2016)

\*Co-first authors.

‡Co-senior authors.

Correspondence and requests for reprints should be addressed to either Avrum Spira, M.D., 72 East Concord Street, E631, Boston, MA 02118. E-mail: [aspira@bu.edu](mailto:aspira@bu.edu) or to Naftali Kaminski, M.D., 300 Cedar Street, Ste S441D, New Haven, CT 06519. E-mail: [naftali.kaminski@yale.edu](mailto:naftali.kaminski@yale.edu)

This article has an online supplement, which is accessible from this issue's table of contents at [www.atsjournals.org](http://www.atsjournals.org)

Am J Respir Crit Care Med Vol 194, Iss 8, pp 948–960, Oct 15, 2016

Copyright © 2016 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201510-2026OC on April 22, 2016

Internet address: [www.atsjournals.org](http://www.atsjournals.org)

## At a Glance Commentary

### Scientific Knowledge on the

**Subject:** Although idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease share risk factors, such as cigarette smoking, overlapping disease pathogenesis mechanisms have yet to be described.

### What This Study Adds to the

**Field:** We show that chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis share transcriptional mRNA–microRNA hypoxia/p53 regulation and alternative splicing of p53/hypoxia-associated genes *NUMB* and *PDGFA*.

Chronic Obstructive Lung Disease as a disease state characterized by exposure to a noxious agent (which is almost entirely cigarette smoke in the United States) resulting in airflow limitation that is not fully reversible (3), is thought to result from recruitment of inflammatory cells in response cigarette smoke. A subset of patients sustain destruction of lung elastin and other extracellular matrix proteins, death of alveolar cells by apoptosis, or repair failure that leads to airspace enlargement characteristic of emphysema. Often contrasted with emphysema, IPF is characterized by the findings of usual interstitial pneumonia, including the interposition of patchy foci of active fibroblast proliferation associated with minimal inflammation, extracellular matrix deposition, and abnormal alveolar remodeling (2). In recent years there has been significant progress in understanding of the molecular mechanisms that underlie the lung response to

injury and lead to fibrosis or emphysema in animal models of disease, but there is still little understanding of the molecular mechanisms that lead to the abnormal lung phenotype in both diseases in humans.

The emergence of high throughput transcript technologies allows investigators to glean mechanisms from diseased human tissues. In chronic lung disease, there have been only a few high-throughput gene expression microarray studies of COPD (4–7) or IPF (8–11) and studies of miRNA expression in COPD (12, 13) or IPF (14, 15). Despite common risk factors, evidence that parallel pathways may be activated and suggestions that direct comparisons of the diseases may be mechanistically informative (16), there has been no effort to determine whether convergent molecular pathways could be identified in IPF and COPD. Thus, in this study we examined IPF, COPD, and normal lungs in parallel, aiming to provide an unbiased description of the transcriptional repertoire of the lung and its response to chronic injury and to identify convergent transcriptional regulatory networks in COPD and IPF using integrative genomic approaches.

Accordingly, we performed mRNA-Seq and mRNA and miRNA profiling using microarrays on 89 lung tissue samples representing IPF, COPD, and control from subjects obtained through the NHLBI Lung Tissue Research Consortium (LTRC) as part of the Lung Genomic Research Consortium (LGRC). With the objective of identifying convergent disease-associated alterations in gene expression and splicing, we characterized the lung transcriptome and identified molecular alterations shared by both chronic lung diseases. The p53/hypoxia pathway was up-regulated in both COPD and

IPF compared with normal histology controls, a finding we validated in an independent cohort of samples and localized to the epithelial layer by immunohistochemistry. In addition, we identified both miRNAs and alternative splicing events related to the p53/hypoxia pathway. Our work provides the first RNA-seq study of the adult human lung transcriptional repertoire in health or response to chronic injury as represented by COPD and IPF together. It also represents the first time that common genome scale transcript alteration alterations have been identified between IPF and COPD. Our approach (Figure 1) and findings provide insights into convergent mechanisms that underlie both diseases and highlight a relatively unrecognized central role of the p53/hypoxia pathway in both diseases.

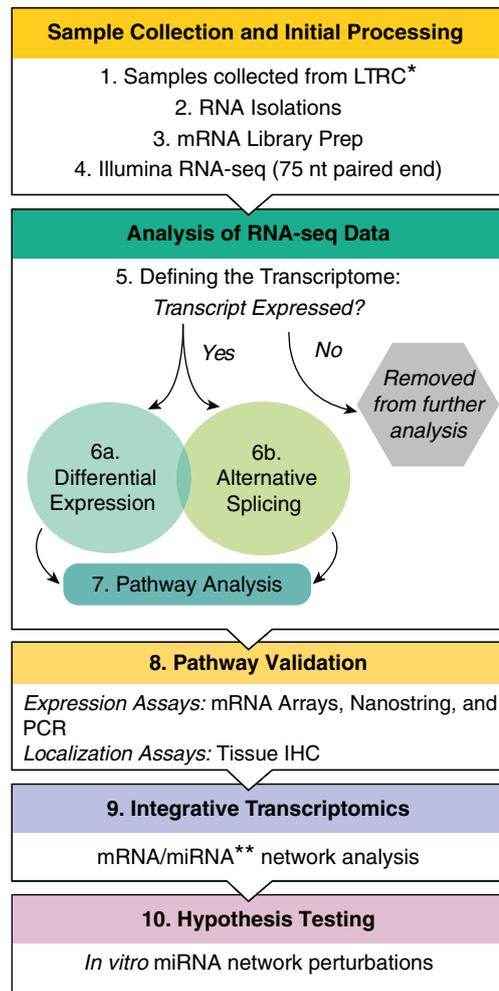
## Methods

### Patient Population

Lung tissue samples were obtained from the NHLBI funded LTRC as previously described (17). We evaluated the initial 89 samples for potential field of cancerization effects, because some samples were collected from areas adjacent to cancer. This evaluation included mining for abnormal gene fusions and cytogenetic abnormalities in the lung and blood allele balance. Two control samples contained between 12% and 25% abnormal cells by allele balance via the Illumina (San Diego, CA) Infinium assay and were thus removed from further analysis. Remaining were  $n = 19$  subjects with COPD with predominant emphysema phenotype,  $n = 17$  subjects with COPD without predominant emphysema phenotype (suspected airways

Supported by National Institutes of Health/NHLBI grants R01-HL095393 and U01HL108642 and Lung Genomics Research Consortium grant RC2-HL101715.

Author Contributions: R.L.K. participated in design of work, analyzed and interpreted data, drafted the work, and revised the work; J.F.B. participated in design of work, analyzed data, interpreted work, drafted the work, and revised the work; J.T. acquired and annotated samples, interpreted results, and revised the work; K.P. led microRNA experiments, performed key experiments, analyzed the work, and revised the work; L.H. performed key validation experiments, analyzed the work, and revised the work; C.P. led experimental validation design and implementation, analyzed the work, and revised the work; G.L. acquired samples, led sequencing effort, and revised the work; B.J.-G. generated microarray data, interpreted results, and revised the work; D.K. analyzed the work and revised the work; S.Z. conceived of the work, acquired samples, and revised the work; M.L. conceived of and designed the work, interpreted data, and revised the work; F.M. participated in study design, acquired samples, and interpreted and revised the work; J.Q. participated in study design, interpreted data, and revised the work; F.S. participated in study design, interpreted data, and revised the work; A.L. participated in study design, interpreted data, and revised the work; M.G. contributed to data interpretation, study design, and manuscript generation; I.Y. contributed to data interpretation, study design, and manuscript generation; D.A.S. was involved in all stages of work from initial conceptualization, study design, data generation and interpretation, experimental validation, and manuscript generation; J.B. led the sequencing effort, conceived of and designed the work, interpreted data, and revised the work; A.S. led all stages of the work from initial conceptualization, study design, data generation and interpretation, and manuscript generation; and N.K. led all stages of work from initial conceptualization, study design, data generation and interpretation, experimental validation, and manuscript generation.



\* LTRC: Lung Tissue Research Consortium

\*\* microRNA expression data generated using Agilent microRNA microarrays

**Figure 1.** Overview of analysis. IHC = immunohistochemistry; LTRC = Lung Tissue Research Consortium; miRNA = microRNA; PCR = polymerase chain reaction.

disease),  $n = 19$  subjects with IPF,  $n = 13$  subjects with COPD with intermediate emphysema phenotype, and  $n = 20$  histologically normal tissue samples.

Seventy-five of the remaining 87 samples were selected by pathologists as displaying the most distinct phenotypes (Table 1). The COPD categories were defined based on the percent emphysema: samples that had less than 10% emphysema and greater than 30% emphysema were used to define the COPD airway and emphysema phenotypes, respectively. The interstitial lung disease (ILD) samples were subset down to those with “IPF” phenotype as per American Thoracic Society criteria (18).

### RNA Extraction

We extracted total RNA from all lung samples using the QIAcube system (QIAGEN Inc., Valencia, CA) with the miRNeasy kit. RNA quality was determined using a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) with a RNA integrity number greater than 7.0 as the criterion for acceptable quality.

### mRNA-Seq

Library preparation and mRNA sequencing was performed on each of the 89 LGRC samples. The library preparation was done using Illumina’s mRNA Sequencing Sample Preparation kit starting with 1  $\mu\text{g}$

of total RNA (see the METHODS section in the online supplement for additional information).

### mRNA and miRNA Microarray Processing

We hybridized RNA from all LGRC samples (Table 2) to Agilent V2 Human Whole Genome microarrays and then quantile normalized the data using GeneSpring (see the METHODS section in the online supplement for additional information).

### Initial Processing of mRNA-Seq Data

To perform initial data quality control (QC), we used standard Illumina metrics, our own custom perl scripts, and FastQC (Babraham Bioinformatics, Babraham, UK). We aligned samples that passed the QC filter to hg19 using Tophat and completed an additional QC step involving review of standard alignment metrics. We generated gene level expression estimates using Cufflinks (19).

### Gene Filtering

First we  $\log_2$  transformed FPKM gene expression data from Cufflinks. Using their “on” or “off” status and coefficient of variance, we filtered genes. To determine gene status we used a modified version of the mixture model in the SCAN.UPC Bioconductor package (see the METHODS section in the online supplement) (20). For a gene to be included in differential expression analysis, it had to be classified as “on” in at least 25% of samples out of the two phenotypic groups being compared. After this, the bottom 20% of genes were filtered out based on their coefficient of variation.

### Differential Expression

We identified differentially expressed genes with the limma package. For emphysema, we included only samples with greater than 30% emphysema. Out of the ILD population, we included only samples with pure IPF. We included only genes annotated as “known” in Ensembl. Overall we chose samples such that age, pack-years, smoking status, and sex were not confounding between the groups being compared. Functional enrichment was tested with Gene Set Enrichment Analysis (GSEA).

### Comparing mRNA Arrays with mRNA-Seq

Using a  $t$  test in limma (same as for RNAseq analysis) we identified disease-associated

**Table 1.** Demographics of the 75 Lung Tissue Samples Used for Analysis of Differential Expression and Splicing (Figures 3 and 4)

	IPF	Emphysema (>30% Emphysema)	Airway COPD (<10% Emphysema)	Control
Samples, n	19	19	17	20
Age	64.0 ± 9.2	56.3 ± 8.7	68.4 ± 10.4	62.3 ± 10.0
Sex	15 M, 4 F	10 M, 9 F	12 M, 5 F	11 M, 9 F
Race	19 white	18 white, 1 African American	17 white	18 white, 2 African American
Smoking status	17 former, 2 never	18 former, 1 never	15 former (2)	1 current, 15 former, 2 never (2)
Pack-years	31 ± 23 (2)	48 ± 27 (1)	51 ± 23 (2)	27 ± 23 (4)
% Emphysema	1.7 ± 2.6 (7)	47.5 ± 9.1	2.3 ± 2.0	0.8 ± 1.3 (9)

*Definition of abbreviations:* COPD = chronic obstructive pulmonary disease; IPF = idiopathic pulmonary fibrosis. These samples are a subset of Table 2. Values in parentheses are missing demographics.

changes in gene expression. Genes in Ensembl without Agilent probes mapping to them were excluded from analyses. Platforms were compared by evaluating the overlap between genes identified as differentially expressed. Correlation between *t*-statistics on the two platforms was found using a Pearson correlation.

### Immunohistochemistry

We acquired formalin-fixed paraffin-embedded tissues from the LTRC. As previously described (21), we performed immunohistochemistry using mouse monoclonal antibodies directed against *MDM2* (Millipore, Temecula, CA), *HIF1A* (Stressgen, Victoria, BC, Canada), and *NFKB1B* (ABD Serotec, Raleigh, NC), and a rabbit polyclonal antibody directed against *PDGFA* (Santa Cruz Biotechnology, Santa Cruz, CA). We took all brightfield images with an Olympus (Billerica, MA) DP25 camera on an Olympus CH2 microscope (21).

**Table 2.** Demographics of the 87 Lung Tissue Samples Used to Define Transcriptomic Landscape (Figure 2)

	ILD/IPF	COPD	Control
Samples, n	23	43	21
Age	63.5 ± 8.7 (1)	62.0 ± 11.2 (7)	62.0 ± 9.8
Sex	17 M, 5 F (1)	22 M, 14 F (7)	12 M, 9 F
Race	22 white (1)	35 white, 1 African American (7)	19 white, 2 African American
Smoking status	19 former, 3 never (1)	33 former, 1 never (9)	1 current, 15 former, 3 never (2)
Pack-years	28.7 ± 22.5 (3)	49.2 ± 25.3 (10)*	27.3 ± 22.6 (5)
% Emphysema	1.7 ± 2.6 (11)	24.9 ± 22.2*	1.1 ± 1.5 (4)

*Definition of abbreviations:* COPD = chronic obstructive pulmonary disease; ILD = interstitial lung disease; IPF = idiopathic pulmonary fibrosis.

Values in parentheses are missing demographics.

\*Significant with  $P < 0.05$  compared with control.

### Integrating miRNA and mRNA Data

We integrated miRNA array and mRNA-Seq data with MirConnX (22). This tool combines a prior, static network created from miRNA binding predictions and literature validation with user submitted data to create a transcriptomic gene regulatory network. For each condition-control comparison, we filtered to only differentially expressed mRNAs. We inspected resulting regulatory networks for potential regulatory hubs (miRNAs with a high number of connected mRNAs).

### Identification of Alternative Splicing Events

First, we ran Tophat (19), allowing only known junctions. Using the junction.bed output file, we collected the number of reads that span splice junctions. We ran a linear model with an interaction term, using the RPM of reads spanning each splice junction (*see* online supplement for model). We examined switching behavior and the structure of the significant splice

junctions to identify alternative splicing between the disease and control samples.

### Data Access

Genomic data (RNA-seq and Agilent mRNA and miRNA expression arrays) and associated clinical data are available for download on the LGRC website (<https://www.lung-genomics.org/research/>); miRNA regulatory networks are available at [http://mirconnx.csb.pitt.edu/job\\_results?job\\_id=example10103](http://mirconnx.csb.pitt.edu/job_results?job_id=example10103), [http://mirconnx.csb.pitt.edu/job\\_results?job\\_id=example10102](http://mirconnx.csb.pitt.edu/job_results?job_id=example10102), and [http://mirconnx.csb.pitt.edu/job\\_results?job\\_id=example10101](http://mirconnx.csb.pitt.edu/job_results?job_id=example10101).

## Results

### Study Population from the LTRC

We acquired all lung samples from the LTRC, which were used by the LGRC, both supported by the NHLBI. Patient clinical information, pulmonary functions, demographics, imaging results, pathology, and clinical diagnoses were available. Tables 1–3 provide the demographics and clinical characteristics of all research subjects included in the different phases of the study (Figure 1). We used 87 samples to analyze the global lung transcriptional repertoire (Table 2). To perform analysis of convergent pathways, we used 75 of the 87 samples that exhibited relatively distinct phenotypes of emphysema, COPD without emphysema, IPF, or normal histology controls (Table 1). The patient samples were matched for age, smoking history, and sex. Institutional review boards approved all studies at participating institutions and per LTRC protocol all patients signed informed consent. All of the transcriptomic results, and associated clinical data, are available for download on the LGRC

website (<https://www.lung-genomics.org/research/>).

### RNA-Sequencing Data Alignment to the Human Genome

Each sample yielded a range of 18–46 million (average, 32 million) 75-nt paired-end reads, and a high percentage (average, 28 million) of these reads aligned to the genome using conservative alignment parameters. Specifically,  $85.9 \pm 6.9\%$  of reads aligned to the genome, and  $81.4 \pm 3.1\%$  aligned uniquely. Of the aligned reads,  $90.3 \pm 4.8\%$  were aligned as paired ends (of which  $88.7 \pm 3.8\%$  were properly paired), and  $9.04 \pm 4.8\%$  were aligned as singletons. The alignment statistics across the samples indicate that the RNA-seq data obtained from the LTRC tissue samples were of high quality (see Figures E1 and E2 in the online supplement).

### Identifying the Core Lung Transcriptome in Health and Disease

To identify genes reliably expressed across all lung tissues, we developed a method to distinguish between genes with clear expression signal (called “signal” or “reliably expressed”) and genes with zero or so few aligned reads that they could not be distinguished from either statistical or biologic noise (possibly occurring because of leaky transcription) (23) or the combination of transcription rate and mRNA degradation in the cells (24). We performed this analysis on each of the 87 samples (Table 2) and each sample showed a similar number of genes in the three categories (signal, noise, zero) (Figure 2A). Next, we examined each gene to see if it was “always reliably expressed,” “never reliably expressed,” or “variably expressed” across all 87 samples to characterize the landscape of expression

across the lung transcriptome (Figure 2B). Of the 24,297 experimentally known Ensembl59 genes, 7,767 were constitutively expressed (“reliably expressed” across the 87 samples), 8,756 were never reliably expressed, and 7,774 were variably expressed.

To better understand the transcriptional repertoire of the lung, we looked at which genes were always reliably expressed in each disease subtype (control,  $n = 20$ ; COPD with emphysema,  $n = 19$ ; IPF,  $n = 19$ ; airway COPD,  $n = 17$ ) (Table 1) and the overlap across all 87 samples (Table 2). Most of the genes (77.9%) that were always reliably expressed in each of those subsets overlapped (Figure 2C) and those can be considered the core lung transcriptome. We performed gene ontology analysis on the lists of genes identified as always expressed across all samples and within each phenotype, which showed that most (76%) of the KEGG pathways (25) that were significantly enriched in at least one group ( $q < 0.05$ ) were shared across all five groups (Figure 2D). The data suggest that there is a core lung transcriptome that is reliably expressed in the lung regardless of disease subtypes.

### Differential Expression of Genes in IPF and Emphysema Reveals Overexpression of the p53/Hypoxia Pathway

Starting with 10,512 and 10,267 filtered genes for IPF and emphysema, respectively (intersecting genes, 9,472; union of genes, 11,307), we identified 2,490 genes significantly differentially expressed between IPF and control subjects, and 337 genes between emphysema and control subjects ( $P < 0.005$ ; 55 and 53 genes expected by chance for IPF or emphysema, respectively). We validated the RNAseq results using gene

expression microarrays run on the same 87 samples (Table 2). The  $t$ -statistics were significantly correlated between RNAseq and gene expression microarrays (emphysema vs. control,  $R = 0.75$ ,  $P < 0.001$ ; IPF vs. control,  $r = 0.83$ ,  $P < 0.001$ )

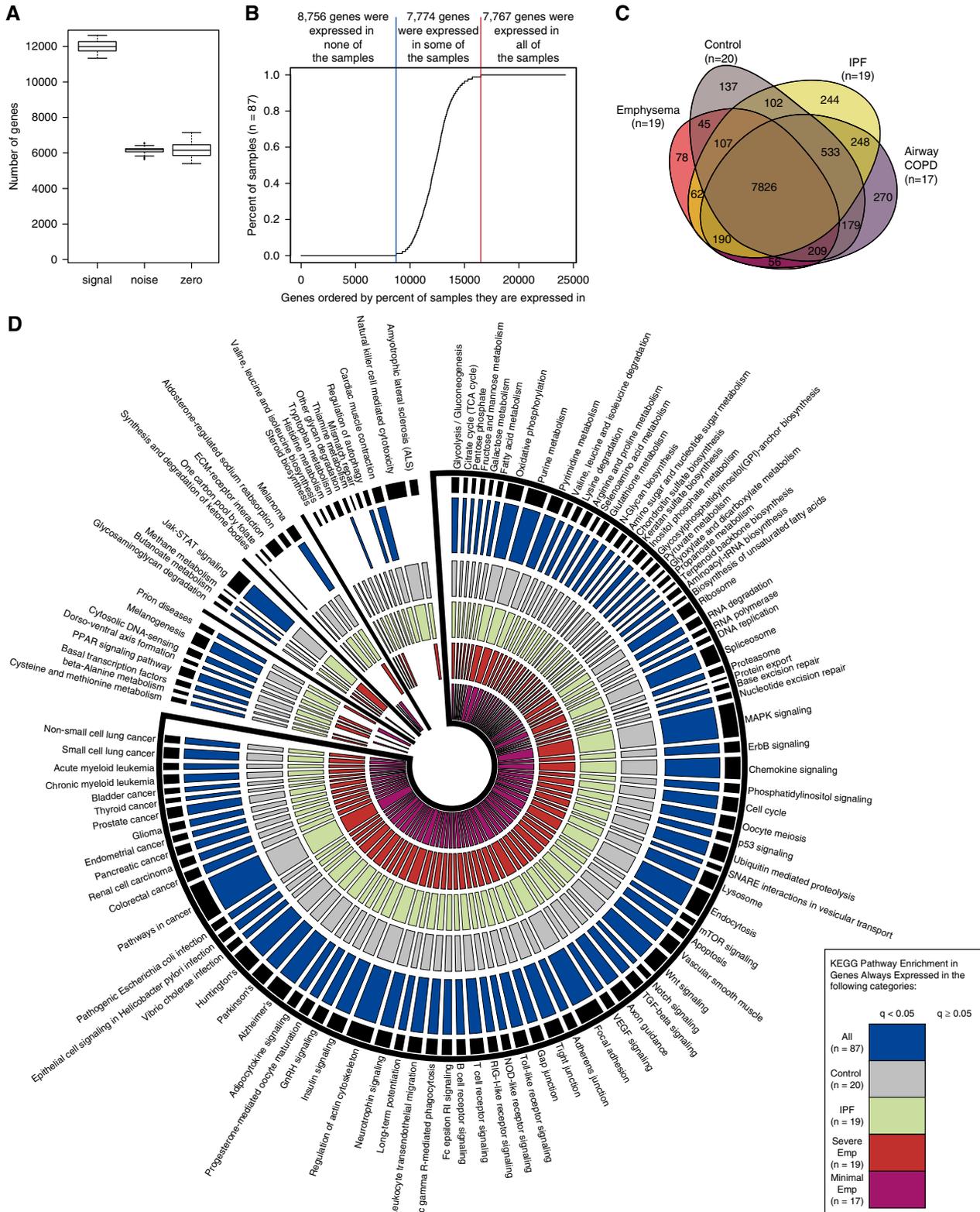
Importantly, the genes that distinguished IPF or emphysema from normal histology controls were changed in concordant directions, even if not always at the same magnitude (Figure 3A). The overlap of the differentially expressed genes common to emphysema and IPF was significantly higher than what would be expected by chance (214 genes, hypergeometric  $P < 3.8e^{-62}$ ). The 214 genes are available in Table E1. GSEA of IPF and emphysema differential expression revealed shared pathways including the KEGG p53 pathway (emphysema vs. control,  $P = 0.003$ ; IPF vs. control,  $P = 0.003$ ), Biocarta p53/hypoxia pathway (emphysema vs. control,  $P = 0.01$ ; IPF vs. control,  $P = 0.026$ ), and other biologic processes outlined in Table E2.

Analysis of a nonoverlapping cohort of lung tissue samples from the LTRC (Table 3) by gene expression microarrays confirmed that up-regulation of the p53/hypoxia pathway characterized the genes that distinguished emphysema or IPF from normal histology controls. GSEA (26) corroborated significant enrichment of the KEGG p53 and Biocarta p53/hypoxia leading edge from GSEA of the initial analysis set among genes up-regulated in emphysema or IPF tissues compared with normal histology controls ( $P < 0.001$ ) (see Figure E3). Because cell type differences were a concern, we used immunohistochemistry to confirm the location of select differentially expressed genes in the p53 pathway in control, emphysema, and IPF samples ( $n = 5$  for each).

**Table 3.** Demographics of Independent Cohort Used for Validation

	ILD/IPF	COPD	Control
Samples, n	77	34	82
Age	$64.4 \pm 8.7$	$60.6 \pm 9.5$	$63.8 \pm 11.9$
Sex	54 M, 23 F	15 M, 19 F	35 M, 47 F
Race	69 white, 2 African American, 2 Asian, 1 other (3)	33 white, 1 African American	76 white, 1 Hispanic, 1 African American, 3 Asian, 1 other
Smoking status	2 current, 42 former, 29 never (4)	2 current, 32 former	1 current, 43 former, 29 never (9)
Pack-years	$24 \pm 18$ (33)	$51 \pm 27$	$37 \pm 32$ (38)
% Emphysema	$0.9 \pm 1.6$ (63)	$36.6 \pm 9.9$ (21)	$0.6 \pm 0.9$ (71)

*Definition of abbreviations:* COPD = chronic obstructive pulmonary disease; ILD = interstitial lung disease; IPF = idiopathic pulmonary fibrosis. Values in parentheses are missing demographics.



**Figure 2.** Overview of the transcriptomic landscape of Lung Genomic Research Consortium lung tissue samples. (A) Box plot of the number of genes characterized as real signal, missing, or noise in all sequenced samples (Table 2). (B) Distribution of genes present in all 87 samples (core lung transcriptome), those never expressed, and those that are variably expressed. (C) Venn diagram of genes always present within the unique lung condition signatures (Table 1). (D) Circos plot highlighting the overlap of functional enrichment of genes always expressed in the core lung transcriptome and across the unique lung conditions. COPD = chronic obstructive pulmonary disease; EMP = emphysema; IPF = idiopathic pulmonary fibrosis; KEGG = Kyoto Encyclopedia for Genes and Genomes.

The changes in expression of *HIF1A*, *MDM2*, and *NFKB1B* were all confirmed by immunohistochemistry, with *MDM2* localizing to the airway epithelium (Figure 3C).

### Overlapping Differential Splicing in Emphysema and IPF Identifies p53-associated *NUMB* and *PDGFA*

We have leveraged the RNA-seq data (Table 1) to identify disease-associated alternative splicing events using an interaction term linear model (see METHODS and online supplement). Our analysis best selected genes with cassette exon skipping or inclusion events (27). Differential splicing analysis revealed one gene in emphysema versus control samples and 27 genes in IPF versus control samples that had significant changes associated with both splice junctions and disease ( $q < 0.25$ ) compared with controls (see Table E3).

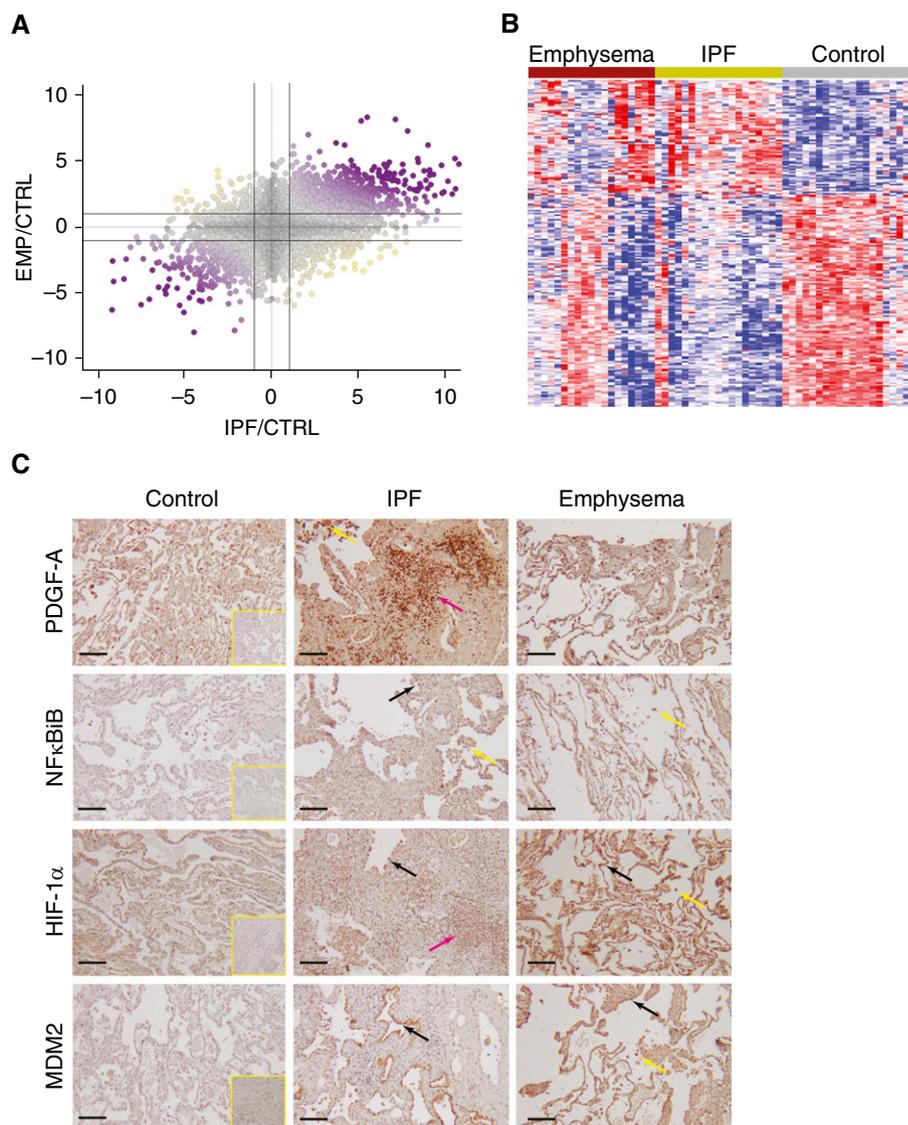
Given that the heterogeneity of disease tissue samples and the large number of splice junctions could leave us underpowered to detect significant alternative splicing, we decided to relax the significance value ( $P < 0.01$ ) to identify a set of alternative splicing events shared in both diseases, because these were less likely to be affected by a change in cell type content, given the histologic/pathologic differences between IPF and COPD. To identify disease-associated splicing events, we focused on genes with splice junctions that showed a switching in expression associated with disease (i.e., at least one splice junction was down-regulated with disease and one splice junction was up-regulated with disease). Two genes, *PDGFA* and *NUMB* (Figures 4A and 4D), showed significant concordant changes in isoform proportions in both emphysema and IPF samples compared with control. *PDGFA* and *NUMB* are likely associated with the p53/hypoxia pathway (see DISCUSSION and supplemental RESULTS).

Analysis of *PDGFA* showed exclusion of an exon in both IPF and emphysema samples compared with normal histology controls. Specifically, whereas all samples express both isoforms of *PDGFA*, expression of one splice junction supporting *PDGFA* isoform 001 increased and expression two splice junctions supporting *PDGFA* isoform 002 were decreased in the disease samples (Figure 4B). Oppositely *NUMB* demonstrated preferential inclusion of an exon among both IPF and emphysema

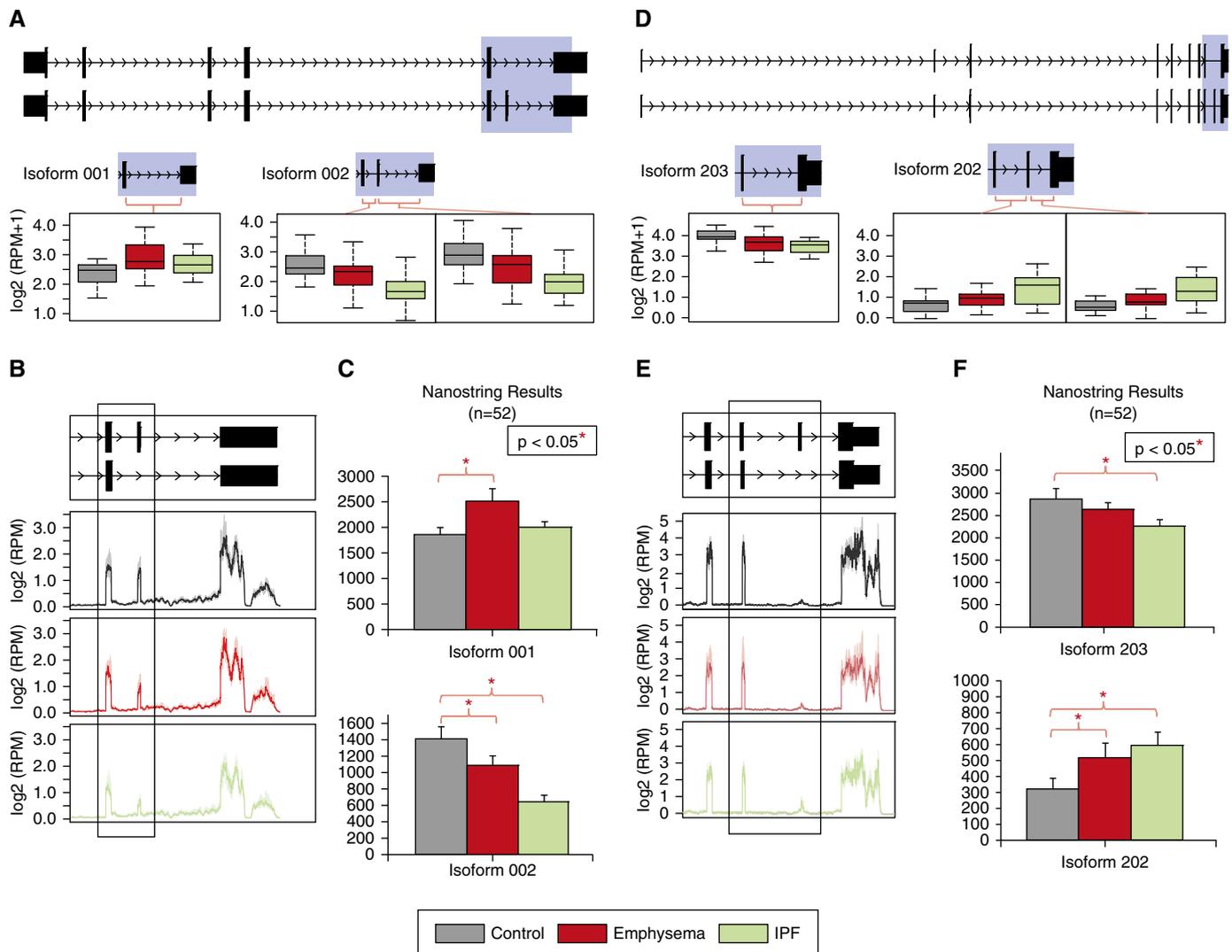
samples compared with control. There was a difference in ratios of isoform 203 and isoform 202 (contains extra exon 9) observed by changes in expression of one splice junction versus two splice junctions, respectively (Figure 4D). Although the *NUMB* coverage plots was more subtle,

there is a significant change ( $P < 0.01$ ) in the proportions of these two isoforms reflected in the ratio of reads mapping to the genomic regions spanning exon 8 and exon 9 in Figure 4E.

We validated these splicing events using the nCounter gene expression analysis



**Figure 3.** Convergent gene expression patterns in idiopathic pulmonary fibrosis (IPF) and emphysema. (A) Scatter plot of fold ratios of emphysema to control ( $y$ -axis) versus fold ratio of IPF to control ( $x$ -axis). Fold ratios for all genes are depicted in log base 2. Spots are colored in increasing shades of purple based on concurrent directions and yellow if going to the opposite direction. (B) Heatmap of significant shared changes in gene expression in emphysema and IPF. Red indicates relative higher expression, blue means relative lower expression. (C) Immunohistochemistry for  $\text{NF}\kappa\text{B}1\text{B}$ ,  $\text{HIF-1}\alpha$ , and  $\text{MDM2}$  ( $n = 5$  per group). For all three targets, greater staining intensity was observed in IPF and emphysema samples compared with controls. For  $\text{NF}\kappa\text{B}1\text{B}$  and  $\text{HIF-1}\alpha$ , there is no clear localization to specific lung cell lineages in contrast to  $\text{MDM2}$ , which seems to be localized to epithelial cells and macrophages. Inset images show staining with nonimmune control IgG (magnification  $\times 100$ ; scale bar = 100  $\mu\text{m}$ ). Black arrows indicate epithelial cells, yellow arrows indicate macrophages, and pink arrows indicate cells in lymphoid aggregates. CTRL = control; EMP = emphysema;  $\text{HIF-1}\alpha$  = hypoxia-inducible factor 1- $\alpha$ ; IPF = idiopathic pulmonary fibrosis;  $\text{MDM2}$  = E3 ubiquitin-protein ligase;  $\text{PDGF-A}$  = platelet-derived growth factor  $\alpha$ .



**Figure 4.** PDGFA and NUMB are differentially spliced in chronic lung disease. The structure of the two isoforms for both *PDGFA* and *NUMB* are shown at the top of A and D, respectively. (A) Box plots of *PDGFA* isoform 001 and 001 splice junctions. (B) Coverage plots showing the interquartile range (lighter shading) of normalized reads and the mean (darker line) for *PDGFA*. (C) Barplot of nanostring validation showing the mean and the SE of the normalized number of times a transcript was counted by the nanostring method. (D) Box plots of the splice junctions in *NUMB*. (E) Coverage plots of *NUMB*. (F) Barplot of *NUMB* alternative splicing nanostring validation. IPF = idiopathic pulmonary fibrosis; RPM = reads per million.

system (28, 29) across the same cohort of samples with RNA-Seq data (subset of 52 samples with leftover mRNA) and an independent replication cohort of samples from the LTRC ( $n = 193$ ) (Table 3). The results on the same subset of samples ( $n = 52$ ) corroborated that *PDGFA* isoform 001 was up-regulated in emphysema versus control samples and unchanged in IPF versus control samples. In validation, *PDGFA* isoform 002 was down-regulated in both IPF and emphysema samples versus control. This verifies that the isoform proportions of *PDGFA* are significantly different between disease and control samples in the original cohort (Figure 4C).

The nCounter results across independent cohort of 193 lung tissue samples confirmed that *PDGFA* is differentially spliced in a more general set of COPD samples compared with control ( $P < 0.05$ ). However, when moving to a more general set of ILD samples (not limited to IPF samples), *PDGFA* seems to be significantly differentially expressed rather than differentially spliced, suggesting that the differential splicing is likely to be more specific to IPF samples than to the ILD samples (see Figure E4).

For *NUMB*, the nCounter expression analysis on the same subset ( $n = 52$ ) showed that *NUMB* isoform 202 was significantly

up-regulated in both IPF and emphysema samples versus control ( $P < 0.05$ ). It also confirmed that *NUMB* isoform 203 was significantly down-regulated in IPF versus control samples ( $P < 0.05$ ) and not significantly changed in COPD with emphysema versus control samples (Figure 4F). This confirmed our RNA-seq results, which showed that isoform proportions of *NUMB* were significantly different between chronic lung diseases and control samples (Figure 4E). The nCounter validation on the independent cohort also confirmed that the *NUMB* isoform 202 was significantly up-regulated across a more general set of ILD versus control samples,

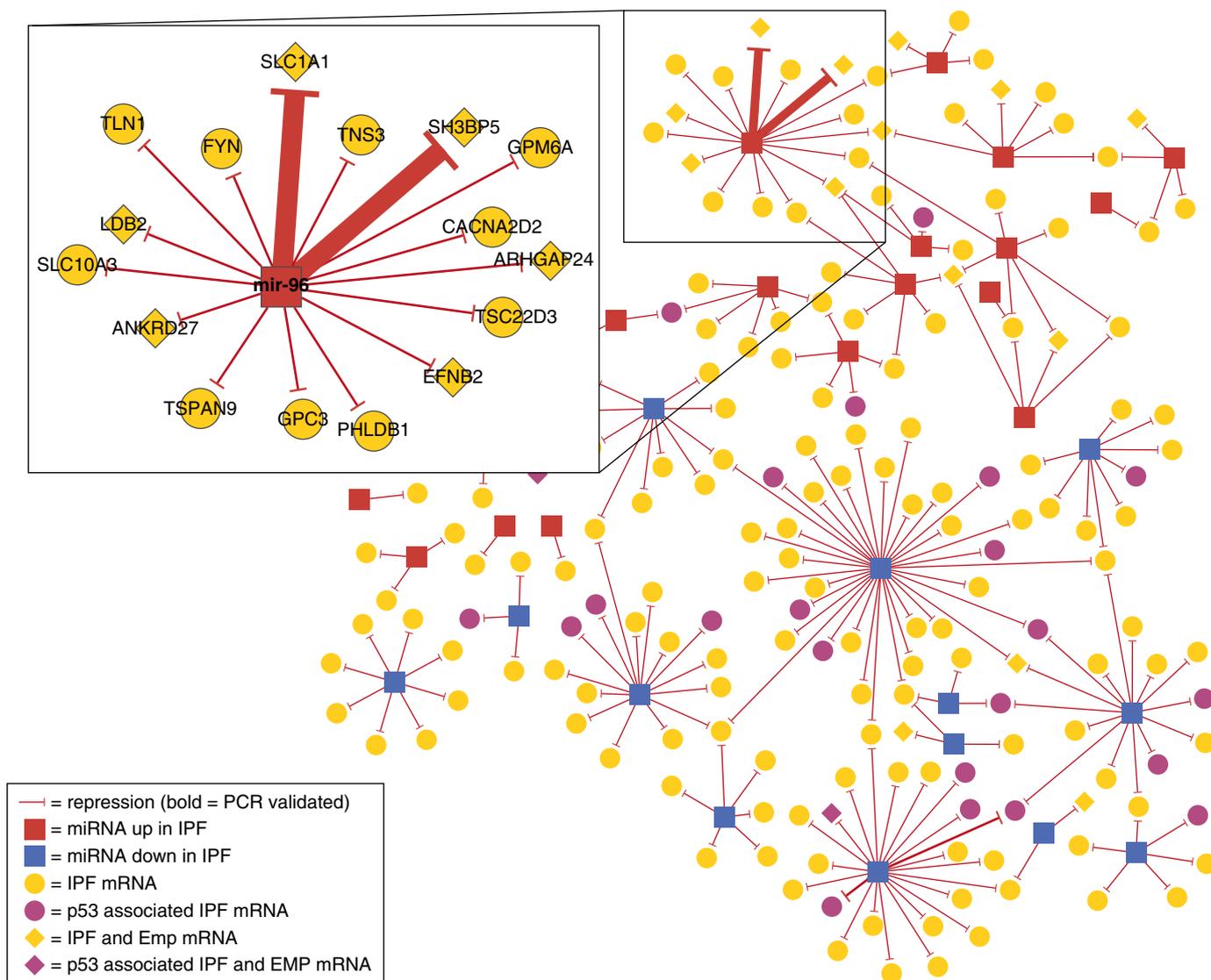
whereas the *NUMB* isoform 203 was significantly down-regulated in ILD samples (see Figure E4).

### Identification and Validation of MicroRNAs That Regulate Transcriptional Networks Shared in COPD and IPF

Integrating mRNA-Seq and miRNA microarray expression data on the same samples (Table 1) uncovered additional insights into the transcriptomic regulation of the p53/hypoxia pathway in emphysema and IPF. Using miRconnX (22), we created a data-driven and prior knowledge based gene/miRNA regulatory network. Initially, we constructed a regulatory network using

genes differentially expressed ( $P < 0.05$ ) in the same direction in both emphysema and IPF to explore shared regulatory mechanisms between the two diseases ([http://mirconnx.csb.pitt.edu/job\\_results?job\\_id=example10103](http://mirconnx.csb.pitt.edu/job_results?job_id=example10103)). The network contains 15 miRNA including *MIR96* and 31 genes. We created two additional networks by submitting emphysema versus control genes ([http://mirconnx.csb.pitt.edu/job\\_results?job\\_id=example10102](http://mirconnx.csb.pitt.edu/job_results?job_id=example10102)) and IPF versus control genes (Figure 5, or [http://mirconnx.csb.pitt.edu/job\\_results?job\\_id=example10101](http://mirconnx.csb.pitt.edu/job_results?job_id=example10101)) with the same  $P$  value cutoff. Both of these networks featured *MIR96* as the most connected miRNA.

Because glutamate transporter *SCL1A1* and BTK inhibitor *SH3BP5* were down-regulated in both diseases (see Table E1) and predicted to be repressed by *MIR96* in the shared regulatory network generated using miRConnX we tested the effect of overexpression of *MIR96* on their gene expression levels *in vitro*. Overexpression of *MIR96* in primary lung fibroblasts and epithelial cells significantly repressed the expression of both genes (see Figure E5). To test whether up-regulation of *MIR96* induced global changes in gene expression similar to those observed in the lung we ran gene expression arrays on the same samples. GSEA analysis revealed that overexpression of *MIR96* recapitulated



**Figure 5.** Shared emphysema (EMP) and idiopathic pulmonary fibrosis (IPF) microRNA (miRNA) regulatory network. Regulatory miRNA–mRNA network showing regulation in both diseases. *Red lines* indicate direction of repression. *Bold red lines* indicate interactions that were selected and validated by polymerase chain reaction (PCR).

*in vitro* some of the global gene expression changes observed in IPF lungs relative to controls (see Figure E6;  $P = 0.008$ ).

## Discussion

In this study we provide the most in depth profile of the core lung transcriptome in health and disease to date using next generation sequencing. By comparing the two radically different but similarly devastating lung diseases, emphysema and IPF, with healthy, normal histology lungs we discover that that activation of p53/hypoxia pathway is common to both diseases. Despite their extremely divergent phenotypes, IPF and emphysema share the same environmental risk factors and especially exposure to cigarette smoke. This unexpected and novel finding may shed significant light on the potential core pathways that initiate lung chronic lung remodeling in response to environmental injury. Taking advantage of the full depth of our transcriptomic analyses and large well characterized cohort we identified and validated alternative splicing of p53/hypoxia pathway associated molecules *NUMB* and a role for *MIR96* as a regulator of the p53/hypoxia lung response to environmental injury gene-expression network.

Given the depth of RNA sequencing coverage across a relatively large and diverse set of lung tissues at different stages of health and disease, our study provides a view of the core transcriptome of the lung and the wide range of its transcriptional potential. Using a mixed modeling approach, we classified genes as either having no detected expression by RNA-seq, reliably expressed, and detected but likely to be transcriptional noise rather than active expression. Generally, within each sample, approximately half of the known transcriptome is reliably expressed as shown by Figure 2A. We then classified the transcriptome by counting the number of samples a gene was reliably expressed in: approximately one-third of genes were reliably expressed in all 87 samples, one-third were never reliably expressed in any of 87 samples, and one-third were variably expressed in a subset of the 87 samples as seen in Figure 2B. This relatively large transcriptional plasticity likely reflects both the diversity of cell types in the lung, and the ability of individual cells types to

dramatically change their transcriptional profiles, both important characteristics in an organ continuously and unpredictably exposed to a wide array of inhaled environmental exposures.

Aside from its novelty the importance of our description of the lung transcriptome lies in its comprehensive nature. In recent years there have been multiple efforts to target drugs to distinct organ systems. Many of these efforts failed because most of the information about organ specificity of gene expression is historical, based on a small number of samples and usually containing only normal tissues or one disease state. In contrast our dataset is wide, contains a diverse compendium of lung tissues in ranges from normal histology to fulminant disease, and thus should serve as a unique resource to those seeking to understand lung transcriptional networks, or more importantly, develop lung-specific interventions.

Our initial analysis aimed at identifying the common molecular networks in emphysema and IPF was conducted at the gene-expression level. One of the striking findings from that analysis was the relatively large number of genes that were differentially expressed between IPF and normal lung, as compared with the number of genes that are differentially expressed between emphysema and normal lung. These findings could potentially stem from the distinct cellular changes that characterize fibrotic foci that were profiled as compared with a more heterogeneous cell type composition in emphysema. However, despite the contrasting differences in cellular content and lung structure between IPF and emphysema, we identified a shared molecular network related to the up-regulation of the p53/hypoxia pathway in both conditions. When the p53 pathway is triggered by hypoxia instead of DNA damage, apoptosis is not triggered. Our results confirm previous studies that found up-regulation of specific members of the p53/hypoxia pathway when studying the diseases separately. Up-regulation of *HIF1A*, *TP53*, *MDM2*, *CDKN1A*, and *BAX* were described in IPF (30, 31) and up-regulation of *TP53* and *BAX* were described in emphysema (32). However, these studies did not compare emphysema and IPF together with control and thus did not highlight the fact that this pathway may serve as a core lung response to environmental injury. Similarly, our

analysis uncovered additional transcriptional regulatory dimensions of this pathway were dysregulated in both diseases including alternative splicing and microRNA regulation.

Considering the commonality of the findings between emphysema and IPF, it is tempting to hypothesize that these findings represent indeed a core pulmonary injury response, but further studies are needed to evaluate whether this dysregulation is in fact causal. The mechanisms driving P53/hypoxia pathways in both diseases are unclear. It is possible that this is simply a response to the impaired gas exchange typical for both diseases. However, it could also represent a response to injury as has been preciously suggested. Regardless, chronic activation of these pathways has important downstream cellular and inflammatory effects that may sustain the phenotypic changes observed in the diseases. Detailed mechanistic studies are required to answer this question, and to identify what are the points of divergence for both diseases.

Another discovery is the identification of IPF- and emphysema-related splicing events. Analysis of RNAseq reads that overlap transcript splice junctions allowed us to identify disease-associated changes in alternative splicing. As with differential gene expression, there were more differentially spliced genes in IPF versus control lung as compared with emphysema versus control lung, which may again likely reflect diversity of cell types. *PDGFA* and p53-associated *NUMB* were differentially spliced in both emphysema and IPF tissues compared with normal histology controls, with the change in isoform expression in the same direction. *NUMB* has four primary isoforms that occur from the alternative splicing of two regions: a phosphotyrosine-binding domain and a proline-rich region (PRR), which is a SH3-binding domain. In IPF and emphysema, we saw evidence for a change in the alternative splicing of the PRR based on the splice junctions that distinguish between the isoforms that contain a 144-nt (48 amino acid) insert in the PRR and the isoforms that do not include that insert in the PRR. In both diseases, the *NUMB* isoform with the longer PRR was increased and the *NUMB* isoform with the shorter PRR was decreased. This result is of particular interest because of its relation to the p53 pathway. *NUMB* is known to

bind to both *TP53* and *MDM2* separately and may potentially form a triplex with *TP53* and *MDM2*, stabilizing and preventing *TP53* from being degraded (33–35). Further experiments are required to determine whether the observed splicing event affects the binding of *NUMB* to *TP53* and *MDM2*.

*PDGFA* has two primary transcript products (Ensembl isoforms *PDGFA-001* and *PDGFA-002*) that are 196 and 211 amino acids long, respectively (36, 37). The longer (*PDGFA-002*) version that contains exon 6 is decreased in IPF and emphysema, whereas the shorter (*PDGFA-001*) does not change. The function of this extra exon is well described. It is transcribed into a retention domain composed of a 15 amino acid long carboxy-terminus that is ensnared by binding to heparin sulfate in the extracellular matrix on secretion, preventing this isoform from diffusing away from the cell. The shorter isoform, however, is able to diffuse away and trigger other cells (38–43). Although there is limited understanding how the individual isoforms may play a role in the p53/hypoxia pathway, a number of groups have shown that *PDGFA* expression associates with hypoxic lung injury in rodent models of lung disease (44, 45). We have also shown that there is a significant enrichment of genes associated with the p53/hypoxia pathway in a list of genes whose expression correlates significantly with the ratio of measurements of the two different isoforms of *PDGFA* and is further evidence that this overlapping alternatively spliced gene in IPF and COPD might also be related to the p53/hypoxia pathway (see RESULTS in the online supplement). Although this suggests a relationship that may affect the p53/hypoxia pathway, such as *NUMB*, the specifics of how the two isoforms may affect the p53/hypoxia pathway are unknown and a potential area of future study.

Exploring the shared molecular network between IPF and emphysema involved an integrative analysis of the mRNA-seq and microRNA array data generated on the same lung tissue samples. This analysis also revealed evidence

suggesting that both diseases share common transcriptional regulatory motifs, with several of the same microRNAs being included in the regulatory networks of both. Specifically of interest is *MIR96*, which goes up in both diseases and could regulate a number of genes differentially expressed in both IPF and emphysema including *SCL1A1*, *SH3BP5*, *LDB2*, and *ARGHAP24*. *SCL1A1* is a glutamate transporter that is down-regulated under hypoxic conditions (46, 47). *SH3BP5* inhibits *BTB* (48), which is a binding partner of hypoxia-induced mitogenic factor (*HIMF*) (49). *LDB2* binds to LIM domain-binding proteins (50), which inhibit *HIF1A* (51). Importantly, our results demonstrate that overexpression of *MIR96* in both lung epithelial cells and fibroblasts *in vitro* recapitulates components of the shared emphysema-IPF gene-expression network, providing evidence that *MIR96* may regulate a part of the shared disease gene-expression network.

Although our unique study design and comprehensive transcriptional profiling provided an unprecedented resolution of the lung transcriptome in health and disease, there are several critical limitations to our work. We profiled whole lung tissue and thus some of the differential gene expression observed between conditions is driven by differing proportions of lung cell types. To address this concern, we pursued immunohistochemistry to validate the cell type responsible for expression of a select number of genes. Interestingly, we observed convergence of molecular pathways between IPF and emphysema despite the clear difference in cell types between these conditions, an observation that may suggest that indeed these are core pathways that underlie the lung response to environmental injury.

Given the cross-sectional nature of our study, it is difficult to distinguish gene expression changes that are causal versus consequence of the disease process. The integrative mRNA-miRNA network and the subsequent functional validation studies *in vitro* provide some evidence for a causal relationship between regulatory miRNA

and the disease-associated gene expression network. However, further functional studies including animal knockout models are needed to conclusively test for causality. Some of the control lungs were collected from smokers with adjacent lung cancer, raising the potential for some of the differential expression to be driven by the local “field cancerization” effect. To limit this confounding effect, we used matched single-nucleotide polymorphism array from the lung and blood to filter out samples with abnormal gene fusions and cytogenetic abnormalities. Although the size of our initial cohort was limited, we confirmed our sequencing results on multiple levels, including validating and replicating the results using an additional cohort and additional techniques. Finally, our study did not explore the potential for RNA-seq to uncover unannotated alternative splicing events, novel transcripts associated with disease, or comparative network analysis, which were beyond the scope of the current paper.

In summary, our paper demonstrates the potential for next-generation sequencing to provide resolution of the transcription potential of the lung in health and disease. Importantly, by profiling multiple distinct lung diseases in parallel within the same study, we uncovered molecular networks that are shared among smokers with IPF and emphysema. Expanding this approach to other disease of the lung and other organ systems may enable us to begin to redefine the clinical and pathologic boundaries that have traditionally divided disease entities to the molecular pathways that are shared or distinct. Our study also demonstrates the value of integrating diverse molecular data from the same specimen to gain insight into the regular networks that associate with disease. These networks may not only provide insight into disease pathogenesis (additional function studies are needed), but if proved to be causal may suggest novel diagnostic biomarkers and therapeutic targets for chronic lung disease. ■

**Author disclosures** are available with the text of this article at [www.atsjournals.org](http://www.atsjournals.org).

## References

1. Mannino DM, Homa DM, Akinbami LJ, Ford ES, Redd SC. Chronic obstructive pulmonary disease surveillance—United States, 1971–2000. *Respir Care* 2002;47:1184–1199.
2. Selman M, King TE, Pardo A; American Thoracic Society; European Respiratory Society; American College of Chest Physicians. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. *Ann Intern Med* 2001;134:136–151.

3. Pauwels RA, Buist AS, Calverley PMA, Jenkins CR, Hurd SS; GOLD Scientific Committee. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med* 2001;163:1256–1276.
4. Wang I-M, Stepaniants S, Boie Y, Mortimer JR, Kennedy B, Elliott M, Hayashi S, Loy L, Coulter S, Cervino S, *et al.* Gene expression profiling in patients with chronic obstructive pulmonary disease and lung cancer. *Am J Respir Crit Care Med* 2008;177:402–411.
5. Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, Celli B, Brody JS. Gene expression profiling of human lung tissue from smokers with severe emphysema. *Am J Respir Cell Mol Biol* 2004;31:601–610.
6. Ning W, Li C-J, Kaminski N, Feghali-Bostwick CA, Alber SM, Di YP, Otterbein SL, Song R, Hayashi S, Zhou Z, *et al.* Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. *Proc Natl Acad Sci USA* 2004;101:14895–14900.
7. Golpon HA, Coldren CD, Zamora MR, Cosgrove GP, Moore MD, Tudor RM, Geraci MW, Voelkel NF. Emphysema lung tissue gene expression profiling. *Am J Respir Cell Mol Biol* 2004;31:595–600.
8. Kass DJ, Kaminski N. Evolving genomic approaches to idiopathic pulmonary fibrosis: moving beyond genes. *Clin Transl Sci* 2011;4:372–379.
9. Pardo A, Gibson K, Cisneros J, Richards TJ, Yang Y, Becerril C, Yousem S, Herrera I, Ruiz V, Selman M, *et al.* Up-regulation and profibrotic role of osteopontin in human idiopathic pulmonary fibrosis. *PLoS Med* 2005;2:e251.
10. Selman M, Pardo A, Barrera L, Estrada A, Watson SR, Wilson K, Aziz N, Kaminski N, Zlotnik A. Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. *Am J Respir Crit Care Med* 2006;173:188–198.
11. Zuo F, Kaminski N, Eugui E, Allard J, Yakhini Z, Ben-Dor A, Lollini L, Morris D, Kim Y, DeLustro B, *et al.* Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci USA* 2002;99:6292–6297.
12. Ezzie ME, Crawford M, Cho J-H, Orellana R, Zhang S, Gelinas R, Batte K, Yu L, Nuovo G, Galas D, *et al.* Gene expression networks in COPD: microRNA and mRNA regulation. *Thorax* 2012;67:122–131.
13. Van Pottelberge GR, Mestdagh P, Bracke KR, Thas O, van Durme YMTA, Joos GF, Vandesompele J, Brusselle GG. MicroRNA expression in induced sputum of smokers and patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2011;183:898–906.
14. Milosevic J, Pandit K, Magister M, Rabinovich E, Ellwanger DC, Yu G, Vuga LJ, Weksler B, Benos PV, Gibson KF, *et al.* Profibrotic role of miR-154 in pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2012;47:879–887.
15. Pandit KV, Corcoran D, Yousef H, Yarlagaadda M, Tzouveleakis A, Gibson KF, Konishi K, Yousem SA, Singh M, Handley D, *et al.* Inhibition and role of let-7d in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2010;182:220–229.
16. Chilosi M, Poletti V, Rossi A. The pathogenesis of COPD and IPF: distinct horns of the same devil? *Respir Res* 2012;13:3.
17. 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.
18. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier J-F, 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.
19. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562–578.
20. Piccolo SR, Withers MR, Francis OE, Bild AH, Johnson WE. Multiplatform single-sample estimates of transcriptional activation. *Proc Natl Acad Sci USA* 2013;110:17778–17783.
21. Kass DJ, Yu G, Loh KS, Savir A, Borczuk A, Kahloon R, Juan-Guardela B, Deiliis G, Tedrow J, Choi J, *et al.* Cytokine-like factor 1 gene expression is enriched in idiopathic pulmonary fibrosis and drives the accumulation of CD4+ T cells in murine lungs: evidence for an antifibrotic role in bleomycin injury. *Am J Pathol* 2012;180:1963–1978.
22. Huang GT, Athanassiou C, Benos PV. mirConnX: condition-specific mRNA-microRNA network integrator. *Nucleic Acids Res* 2011;39:W416–23.
23. Hebenstreit D, Fang M, Gu M, Charoensawan V, van Oudenaarden A, Teichmann SA. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol Syst Biol* 2011;7:497.
24. Hayles B, Yellaboina S, Wang D. Comparing transcription rate and mRNA abundance as parameters for biochemical pathway and network analysis. *PLoS One* 2010;5:e9908.
25. Kanehisa M. The KEGG database. *Novartis Found Symp* 2002;247:91–101; discussion 101–103, 119–128, 244–252.
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;102:15545–15550.
27. Breitbart RE, Andreadis A, Nadal-Ginard B. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu Rev Biochem* 1987;56:467–495.
28. Kulkarni MM. Digital multiplexed gene expression analysis using the NanoString nCounter System. *Curr Protoc Mol Biol* 2011;94:25B.10.1–25B.10.17.
29. Malkov VA, Serikawa KA, Balantac N, Watters J, Geiss G, Mashadi-Hosseini A, Fare T. Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System. *BMC Res Notes* 2009;2:80.
30. Thannickal VJ, Horowitz JC. Evolving concepts of apoptosis in idiopathic pulmonary fibrosis. *Proc Am Thorac Soc* 2006;3:350–356.
31. Nakashima N, Kuwano K, Maeyama T, Hagimoto N, Yoshimi M, Hamada N, Yamada M, Nakanishi Y. The p53-Mdm2 association in epithelial cells in idiopathic pulmonary fibrosis and non-specific interstitial pneumonia. *J Clin Pathol* 2005;58:583–589.
32. Morissette MC, Vachon-Beaudoin G, Parent J, Chakir J, Milot J. Increased p53 level, Bax/Bcl-x(L) ratio, and TRAIL receptor expression in human emphysema. *Am J Respir Crit Care Med* 2008;178:240–247.
33. Colaluca IN, Tosoni D, Nuciforo P, Senic-Matuglia F, Galimberti V, Viale G, Pece S, Di Fiore PP. Numb controls p53 tumour suppressor activity. *Nature* 2008;451:76–80.
34. Carter S, Vousden KH. A role for Numb in p53 stabilization. *Genome Biol* 2008;9:221.
35. Yogosawa S, Miyauchi Y, Honda R, Tanaka H, Yasuda H. Mammalian Numb is a target protein of Mdm2, ubiquitin ligase. *Biochem Biophys Res Commun* 2003;302:869–872.
36. Bonthron DT, Morton CC, Orkin SH, Collins T. Platelet-derived growth factor A chain: gene structure, chromosomal location, and basis for alternative mRNA splicing. *Proc Natl Acad Sci USA* 1988;85:1492–1496.
37. Rorsman F, Bywater M, Knott TJ, Scott J, Betsholtz C. Structural characterization of the human platelet-derived growth factor A-chain cDNA and gene: alternative exon usage predicts two different precursor proteins. *Mol Cell Biol* 1988;8:571–577.
38. Raines EW, Ross R. Compartmentalization of PDGF on extracellular binding sites dependent on exon-6-encoded sequences. *J Cell Biol* 1992;116:533–543.
39. Pollock RA, Richardson WD. The alternative-splice isoforms of the PDGF A-chain differ in their ability to associate with the extracellular matrix and to bind heparin in vitro. *Growth Factors* 1992;7:267–277.
40. Kelly JL, Sánchez A, Brown GS, Chesterman CN, Sleight MJ. Accumulation of PDGF B and cell-binding forms of PDGF A in the extracellular matrix. *J Cell Biol* 1993;121:1153–1163.

41. Somasundaram R, Schuppan D. Type I, II, III, IV, V, and VI collagens serve as extracellular ligands for the isoforms of platelet-derived growth factor (AA, BB, and AB). *J Biol Chem* 1996;271:26884–26891.
42. Andersson M, Ostman A, Westermark B, Heldin CH. Characterization of the retention motif in the C-terminal part of the long splice form of platelet-derived growth factor A-chain. *J Biol Chem* 1994;269:926–930.
43. Andersson M, Ostman A, Bäckström G, Hellman U, George-Nascimento C, Westermark B, Heldin CH. Assignment of interchain disulfide bonds in platelet-derived growth factor (PDGF) and evidence for agonist activity of monomeric PDGF. *J Biol Chem* 1992;267:11260–11266.
44. Mizuno S, Bogaard HJ, Kraskauskas D, Alhussaini A, Gomez-Arroyo J, Voelkel NF, Ishizaki T. p53 Gene deficiency promotes hypoxia-induced pulmonary hypertension and vascular remodeling in mice. *Am J Physiol Lung Cell Mol Physiol* 2011;300:L753–L761.
45. Katayose D, Ohe M, Yamauchi K, Ogata M, Shirato K, Fujita H, Shibahara S, Takishima T. Increased expression of PDGF A- and B-chain genes in rat lungs with hypoxic pulmonary hypertension. *Am J Physiol* 1993;264:L100–L106.
46. Bianchi MG, Bardelli D, Chiu M, Bussolati O. Changes in the expression of the glutamate transporter EAAT3/EAAC1 in health and disease. *Cell Mol Life Sci* 2014;71:2001–2015.
47. Fujita H, Sato K, Wen T-C, Peng Y, Sakanaka M. Differential expressions of glycine transporter 1 and three glutamate transporter mRNA in the hippocampus of gerbils with transient forebrain ischemia. *J Cereb Blood Flow Metab* 1999;19:604–615.
48. Ortutay C, Nore BF, Vihinen M, Smith CIE. Phylogeny of Tec family kinases: identification of a premetazoan origin of Btk, Bmx, Itk, Tec, Txk, and the Btk regulator SH3BP5. In: Hall JC, editor. *Advances in genetics*. Amsterdam, the Netherlands: Academic Press; 2008. pp. 51–80.
49. Su Q, Zhou Y, Johns RA. Bruton's tyrosine kinase (BTK) is a binding partner for hypoxia induced mitogenic factor (HIMF/FIZZ1) and mediates myeloid cell chemotaxis. *FASEB J* 2007;21:1376–1382.
50. Matthews JM, Visvader JE. LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO Rep* 2003;4:1132–1137.
51. Lin J, Qin X, Zhu Z, Mu J, Zhu L, Wu K, Jiao H, Xu X, Ye Q. FHL family members suppress vascular endothelial growth factor expression through blockade of dimerization of HIF1 $\alpha$  and HIF1 $\beta$ . *IUBMB Life* 2012;64:921–930.