Gene Expression Profiling of Human Lung Tissue from Smokers with Severe Emphysema

Avrum Spira, Jennifer Beane, Victor Pinto-Plata, Aran Kadar, Gang Liu, Vishal Shah, Bartolome Celli, and Jerome S. Brody

The Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston; Bioinformatics Program, College of Engineering, Boston University, Boston; and COPD Center at St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, Massachusetts

The mechanism by which inhaled smoke causes the anatomic lesions and physiologic impairment of chronic obstructive pulmonary disease remains unknown. We used high-density microarrays to measure gene expression in severely emphysematous lung tissue removed from smokers at lung volume reduction surgery (LVRS) and normal or mildly emphysematous lung tissue from smokers undergoing resection of pulmonary nodules. Class prediction algorithms identified 102 genes that accurately distinguished severe emphysema from non-/mildly emphysematous lung tissue. We also defined a number of genes whose expression levels correlated strongly with lung diffusion capacity for carbon monoxide and/or forced expiratory volume at 1 s. Genes related to oxidative stress, extracellular matrix synthesis, and inflammation were increased in severe emphysema, whereas expression of endothelium-related genes was decreased. To identify candidate genes that might be causally involved in the pathogenesis of emphysema, we linked gene expression profiles to chromosomal regions previously associated with chronic obstructive pulmonary disease in genome-wide linkage analyses. Unsupervised hierarchical clustering of the LVRS samples revealed distinct molecular subclasses of severe emphysema, with body mass index as the only clinical variable that differed between the groups. Class prediction models established a set of genes that predicted functional outcome at 6 mo after LVRS. Our findings suggest that the gene expression profiles from human emphysematous lung tissue may provide insight into pathogenesis, uncover novel molecular subclasses of disease, predict response to LVRS, and identify targets for therapeutic intervention.

Chronic obstructive pulmonary disease (COPD) is the fourth leading cause of death in this country and is projected to be the number three cause of death globally by 2020 (1, 2). Despite the well-documented role that cigarette smoking plays in the genesis of COPD, it is unclear what steps are involved in its pathogenesis and why only 10–20% of smokers actually develop the disease (3). Most if not all patients with COPD develop the characteristic features of lung emphysema with its pattern of alveolar destruction and abnormal repair, as well as abnormal airway and alveolar inflammatory responses to noxious particles and gases that persist even years after smoking cessation (4, 5). Current pathogenetic theories for the development of emphysema include chronic airway inflammation, an imbalance between protease and antiprotease activity, dysregulation of oxidative stress, and amplifying mechanisms that may perpetuate the chronic inflammatory processes, leading to the progressive destruction and aberrant repair of the lung connective tissue matrix (6). Recent studies have suggested that increased apoptosis of the alveolar wall accounts in part for the loss of lung tissue that characterizes emphysema (7, 8). Transgenic and null-mutant mouse studies have identified a number of genes and pathways that, when altered, result in the morphologic changes of emphysema (9, 10).

The emphysematous phenotype of COPD is associated with increased resting lung volumes and low lung diffusion capacity for carbon monoxide ($D_{L_{CO}}$), tools that have allowed for the proper selection of patients likely to benefit from lung volume reduction surgery (LVRS) (11). In addition, although emphysema is a disease primarily of the lungs, it is associated with important systemic consequences, which include malnutrition with a low body mass index (BMI) (12) and impaired peripheral muscle function (13). These clinically relevant expressions of the disease have been associated with detectable systemic changes, including evidence of oxidative stress, activation of circulating inflammatory cells, and increased levels of the proinflammatory cytokines nitrogen oxide, interleukin 8, and tumor necrosis factor-α (14). These peripheral manifestations are so important that an integrative clinical score that includes BMI, degree of obstruction, perception of dyspnea, and the 6 min walk distance (also known as the BODE index) proved to be an excellent predictor of mortality in a large cohort of patients followed for 5 yr (15).

Despite the many studies aimed at defining the pathogenesis of emphysema, there have been few reports of altered gene expression profiles in small numbers of human emphysematous lung tissue (16, 17), and, to our knowledge, no studies that utilize high-density microarray technology to define gene expression profiles of emphysematous human lungs across a large number of smokers with severe emphysema. Microarray technology produces a global picture of gene expression in lung tissue that may provide insights into the pathogenetic mechanisms involved in COPD. Using high-density DNA microarrays, we compared severely emphysematous tissue removed at LVRS to that of normal or mildly emphysematous lung tissue resected from smokers with nodules suspicious for lung cancer. Class prediction algorithms identified genes that distinguish severe emphysema from mild or no emphysema. By mapping these genes onto their chromosomal loci, we identified those that fall within chromosomal regions previously associated with COPD in genome-wide linkage analyses. The combination of differential expression and chromosomal linkage identifies candidate genes that may be causally involved in COPD. We also defined genes that strongly correlate with pulmonary phenotypic expressions of the disease, such as diffusion capacity for carbon monoxide ($D_{L_{CO}}$) and forced expiratory volume at 1 s (FEV_1), in an attempt to link gene expression

(*Received in original form August 25, 2004 and in revised form September 13, 2004*)

A.S., J.B., and V.P.-P. contributed equally to this work and should be considered co-first authors.

Address correspondence to: Avrum Spira, M.D., The Pulmonary Center, Boston Medical Center, 715 Albany Street, R304, Boston, MA 02118. E-mail: aspira@lung.bumc.bu.edu

Abbreviations: body mass index, BMI; chronic obstructive pulmonary disease, COPD; diffusion capacity for carbon monoxide, $D_{L_{CO}}$; extracellular matrix, ECM; forced expiratory volume at 1 s, FEV_1; insulin-like growth factor binding protein, IGFBP; logarithm of the odds ratio, LOD; lung volume reduction surgery, LVRS; matrix γ-carboxyglutamic acid protein, MGP; Overholt–Blue Cross Emphysema Surgery Trial, OBEST; real-time polymerase chain reaction, RT-PCR; short-tandem repeat, STR.

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

Am. J. Respir. Cell Mol. Biol. Vol. 31, pp. 601–610, 2004
Originally Published in Press as DOI: 10.1165/rcmb.2004-0273OC on September 16, 2004
Internet address: www.atsjournals.org

to molecular processes that predominate in lung parenchyma or in small conducting airways. Through unsupervised analysis, we uncovered a novel molecular subclass of severe emphysema in patients with clinically important systemic manifestations of the disease, such as lower BMI. Finally, our data raise the possibility that gene expression profiles in severely emphysematous lung tissue may predict BODE outcome after LVRS.

Methods

Patient Recruitment

Human lung tissue was obtained from two groups of patients. The patients with severe emphysema ($n = 20$) ($FEV_1 < 50\%$ predicted) underwent lung volume reduction surgery and were enrolled in the Overholt–Blue Cross Emphysema Surgery Trial (OBEST) between December 1998 and December 2002 in eastern Massachusetts. The control subjects were enrolled during the same time period from a population of current or former smokers with no or mild airflow limitation undergoing thoracotomy for localized pulmonary lesions suspicious for malignancy ($n = 14$) at Caritas St. Elizabeth's Medical Center, Boston. Patients with severe emphysema were included if they had a clinical and radiologic diagnosis of emphysema, were less than 75 yr of age, had dyspnea (Medical Research Council scale ≥ 2), and were on maximal medical therapy. Exclusion criteria included patients with α_1-antitrypsin deficiency, tobacco use within 16 wk preceding surgery, significant comorbidities, $FEV_1 > 50\%$, homogeneous emphysema, or an inability to walk more than 150 m after pulmonary rehabilitation (a complete list of inclusion and exclusion criteria is available in the online supplement). For the control group, all patients were former smokers and all underwent high-resolution computed tomography scans of the chest to exclude bullous emphysema before surgery. In addition, patients were excluded from the control group if their FEV_1 was less than 45% predicted or their D_{LCO} was less than 50% predicted.

The forced vital capacity, FEV_1, lung volumes, and single breath D_{LCO} were determined following ATS guidelines. The 6-min walk distance, BMI, and modified Medical Research Council dyspnea scale used to calculate the BODE index have been previously described (15). The patients were evaluated before and 6 mo after surgery. All patients provided written informed consent and the study was approved by the human studies committees of the participating centers.

Specimen Collection

At the time of surgical resection, lung tissue specimens were immediately frozen in dry ice and stored at −80°C. Each specimen was accompanied by an adjacent sample of lung tissue for histologic confirmation. For control subjects, histologically normal lung tissue adjacent to the resected nodule was collected. RNA was extracted from the lung tissue using TRIzol (Invitrogen, Carlsbad, CA) per the manufacturer's protocol. Integrity of the RNA was confirmed with RNA denaturing gel. Each tissue sample yielded 10–20 μg of RNA.

Microarray Data Acquisition and Preprocessing

A sample of 8–10 μg of total RNA from lung tissue was processed, labeled, and hybridized to the Affymetrix HG-U133A Genechip (containing ~ 22,500 human transcripts) (Affymetrix, SantaClara, CA) as described previously (18) (see online supplement for detailed protocol). A single weighted mean expression level for each gene along with a detection p-value (which indicates whether the transcript was reliably detected) was derived using Microarray Suite 5.0 software (Affymetrix). We scaled the data from each array (target intensity of 100) to normalize the results for interarray comparisons. The list of genes on this array is available at http://www.affymetrix.com/support/technical/byproduct.affx?product=hgu133.

To filter out arrays of poor quality, several quality control measures on each array were assessed, including review of the scanned image for significant artifacts, background and noise measurements that differ significantly from other chips, and presence on the array of bacterial genes spiked into the hybridization mix. Furthermore, arrays failing two out of the three following quality control measures were excluded from the study: the 3′ to 5′ ratio of the intensities for glyceraldehyde-3-phosphate dehydrogenase (ratio < 4 considered suitable), percent of genes detected (> 20% of genes detected was acceptable), and percent of gene outliers (< 8% gene outliers was satisfactory). The percent of gene outliers is based on a computational algorithm developed to filter outlier arrays by quantifying the percent of genes on each array that are more than 2 SDs from the mean for the gene across all arrays. A total of 4 of 34 samples (2 patients undergoing LVRS and 2 control patients) were excluded based on the quality control filters described above. To remove genes that were not reliably detected, we filtered out genes whose detection P-value is not < 0.05 in at least 20% of all samples (18), leaving 9,336 genes for the statistical and computational analysis described below.

Class Prediction: Severe Emphysema Versus Normal Lung Tissue

Several class prediction algorithms were used to identify a group of genes whose expression in the lung distinguished severe emphysema ($n = 18$) from no or mild emphysema ($n = 12$). The rationale for this approach stems from the fact that the effectiveness of any single class prediction algorithm will depend on the nature of the data set, and it is therefore prudent to evaluate microarray data using a diverse set of computational algorithms and assign potential relevance to those genes that are reproducibly selected by many protocols (19). The BRB ArrayTools package (available at http://linus.nci.nih.gov/BRB-ArrayTools.html) includes several class prediction methods: the compound covariate predictor, diagonal linear discriminant analysis, one and three nearest neighbors predictor, nearest centroid predictor, and support vector machines. A t test on log-normalized data with a user-specified cutoff for significance level identified differentially expressed genes between the two groups for use in these predictors. Leave-one-out cross-validation was performed and the overall cross-validation misclassification rate reported. Different significance-level cutoffs were tried, and the cutoff giving the lowest cross-validation misclassification rate for a given method was chosen and the corresponding gene list was saved. In addition to BRB ArrayTools, we used three other class prediction algorithms: prediction analysis of microarrays (20) (obtained at http://www-stat.stanford.edu/~tibs/PAM/), the weighted voting (21) class prediction algorithm (implemented using GeneCluster software obtained at http://www-genome.wi.mit.edu/cancer/software/genecluster2/gc2.html), and the genetic algorithm and k-nearest neighbor method described by Li and colleagues (22) (obtained at http://dir.niehs.nih.gov/microarray/datamining/). The results of running the genetic algorithm and k-neareast neighbor algorithm represent the top-20 most frequently chosen genes in 10,000 near-optimal solutions. All of the class prediction methods utilize leave one out cross-validation except for prediction analysis of microarrays which uses 10-fold cross-validation. Genes reported by at least four of the methods above (102 genes) were chosen for further analysis. *See* the online supplement for details of each class-prediction algorithm applied.

Two-dimensional hierarchical clustering was performed across all 30 patients (severe emphysema and no or mild emphysema) and across the genes, distinguishing severe from mild or no emphysema. Prior to clustering, each gene was normalized across all samples to have a mean of zero and a standard deviation of one. Clustering was performed using an uncentered Pearson correlation similarity metric and average linkage clustering with CLUSTER and TREEVIEW software programs (http://rana.lbl.gov/EisenSoftware.htm). Functional classification of these genes was obtained from the Genecards database (http://bioinformatics.weizmann.ac.il/cards/), and statistical over/under-representation of functional categories was determined using EASE software (http://david.niaid.nih.gov/david/ease.htm).

Pearson Correlation

To determine the relationship between gene expression and D_{LCO}, FEV_1, or BMI, Pearson correlations were computed using R software version 1.6.2 (available at http://www.r-project.org/). The correlations were computed for each of the 9,336 genes.

Mapping Gene Expression to Chromosomal Regions

Using genome-wide linkage analyses, Silverman and colleagues have identified chromosomal short-tandem repeat (STR) polymorphic markers with logarithm of the odds ratio (LOD) scores > 1 that are linked to various phenotypes of early-onset COPD (23–25) (summarized in the online supplement, Table E7). To investigate if any of the class-prediction genes that distinguish between severe and mild/no emphy-

sema were located in susceptibility loci, 23 chromosomal regions containing consecutive STR markers with LOD scores > 1 were visually selected from the graphic linkage analyses presented in the various studies. Genetic linkage units (cM) estimated from Silverman and colleagues were converted to their corresponding physical positions (base pairs) using the Marshfield Genetic map (http://hgdownload.cse.ucsc.edu/goldenPath/10april2003/database/stsMap.txt.gz). The start and stop physical positions of each chromosomal region were estimated by finding their respective closest markers on the Marshfield map (the markers have data specifying both their physical and genetic linkage positions). Chromosomal regions containing probe-sets on the class prediction list were flagged (Table E8). These regions were further investigated to explore whether or not other differentially expressed probe-sets between severe and mild/no emphysema, not included on the class prediction list, were present. Probe-sets from the U133A array contained in the filtered gene list of 9,336 probe-sets were mapped to their physical chromosomal locations using probe-set annotation from Affymetrix (http://www.affymetrix.com/support/technical/byproduct.affx?product=hgu133).

Additional differentially expressed probe-sets between severe and mild/no emphysema were identified by conducting a two-sample unequal variance *t* test on log-transformed microarray data for each of the 9,336 probe-sets. The q-value proposed by Tibshirani and colleagues (26) was used as a multiple comparison correction. Each gene was reported with its *P*-value and q-value. The q-value of a particular gene represents the proportion of false positives present in the group of genes with equal or smaller *P*-values than the gene. The q-values were calculated using the program Q-Value, downloaded from http://faculty.washington.edu/~jstorey/qvalue/.

Class Discovery

Several filters were applied to the 9,336 genes to determine subsets of genes that varied among the 18 patients undergoing LVRS. For each gene, the maximum and minimum signals were identified across all samples to calculate the difference in signal intensity as well as the ratio of the two extremes. Gene lists were generated using the following cutoffs for difference and ratio: a max/min difference of at least 300 and ratio of at least 5 (414 genes), 100 and 10 (583 genes), 500 and 3 (597 genes), 300 and 3 (1,028 genes), and 100 and 3 (3,386 genes). For each of these gene lists, two-dimensional hierarchal clustering of the genes and the 18 LVRS samples was performed using *z*-score–normalized data, a Pearson correlation (uncentered) similarity metric, and average linkage clustering with CLUSTER and TREEVIEW software programs.

Class Prediction: Outcome after LVRS

To distinguish between patients undergoing LVRS with a favorable outcome after surgery versus patients whose condition worsened or did not change (based on BODE score before and 6 mo after surgery), the same class prediction algorithms described above were applied. Multidimensional scaling (MDS) using Partek software 5.1 was constructed of all samples according to the expression of these class prediction genes. The MDS plot was constructed from the raw expression data for the genes across all the samples using orthogonal initialization and Euclidean distance as the similarity metric. Principal component analysis using the same data was also performed.

Real-time Quantitative RT-PCR

Quantitative real-time polymerase chain reaction (RT-PCR) was used to confirm the differential expression of a select number of genes within biologically relevant functional categories. Primer sequences were designed with Primer Express software (Applied Biosystems, Foster City, CA) based on alignments of candidate gene sequences. RNA samples (500 ng of residual sample from array experiment) were treated with DNAfree (Ambion, Austin, TX), as per the manufacturer protocol, to remove contaminating genomic DNA. Total RNA was reverse-transcribed using Superscript II (Gibco, Carlsbad, CA). A total of 5 μl of the reverse transcription reaction was added to 45 μl of SYBR Green PCR master mix (Applied Biosystems). Forty cycles of amplification, data acquisition, and data analysis were performed in an ABI Prism 7700 Sequence Detector (Applied Biosystems). All RT-PCR experiments were performed in triplicate on each sample.

TABLE 1. Demographic features for 30 subjects whose microarrays passed the quality control filters

Category	Severe Emphysema	Mild/No Emphysema	*P* value	Test‡
No. of patients, *n*	18	12		
Age, yr	64 (3)	69 (7)	0.02	*t* Test
Sex, female/male	10/8	4/8	0.49	χ^2
BMI	23.2 (3.4)	28.3 (7.2)	0.04	*t* Test
Pack-years	61 (31)	43 (30)	0.12	*t* Test
FEV_1 before BD (% pred)	21 (6)	65 (14)	1.26E-07	*t* Test
FEV_1 after BD (% pred)	23 (6)	67 (12)*	7.59E-07	*t* Test
D_{LCO} (% pred)	29 (7)	82 (25)†	9.20E-05	*t* Test

Definition of abbreviations: BD, bronchodilator; BMI, body mass index; D_{LCO}, lung diffusion capacity for carbon monoxide; FEV_1, forced expiratory volume at 1 s.
Mean (± SD) is shown for continuous variables.
* Three missing datapoints.
† Two missing datapoints.
‡ *t* Test = two sample *t* test.

Supplemental Information

Additional information from this study, including the raw image data from all microarray samples (.DAT files), expression levels for all genes in all samples (stored in a relational MYSQL database), user-defined statistical and graphical analysis of data, and clinical data on all subjects is available at http://pulm.bumc.bu.edu/copdb/. Data from our microarray experiments have also been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (series reference number GSE1650).

Results

Study Population

Microarrays from 30 subjects passed the quality control filters described above and have been included in this study. The clinical characteristics of the 18 patients undergoing LVRS and the 12 smokers with mild or no emphysema are summarized in Table 1. As expected, the patients undergoing LVRS expressed more severe airflow limitation, were slightly younger, and had lower values for D_{LCO} and BMI compared with those who had curative lung resection. Patients undergoing LVRS had a more intense exposure to cigarette smoke, although the difference between groups was not statistically significant.

Genes that Distinguish Severe Emphysema from Mild/No Emphysema

There were 102 genes in common between 4 or more of the class prediction algorithms described above (*see* Figure 1). The *P* value for each of these genes (as determined by a two-sample unequal variance *t* test performed on log-transformed expression data) along with an associated Q value (which represents the proportion of false positives) is reported in Table E2 in the online supplement . A total of 76 genes (75%) were upregulated in the lungs of smokers with severe emphysema as compared with those with mild or no emphysema. On leave-one-out cross-validation, 86–93% of samples were correctly classified by the various class prediction algorithms applied (*see* Table E1 in the online supplement for details). A large number of genes upregulated in the lungs of smokers with emphysema were extracellular matrix (ECM)–related genes, whereas many immune and cell signaling–related genes were downregulated in the lungs of smokers with severe emphysema. Statistically overrepresented gene ontology categories of molecular function include ECM constituents and stress-related genes having oxidoreductase, isomerase, and complement activity (*see* Table E3 in the online supplement).

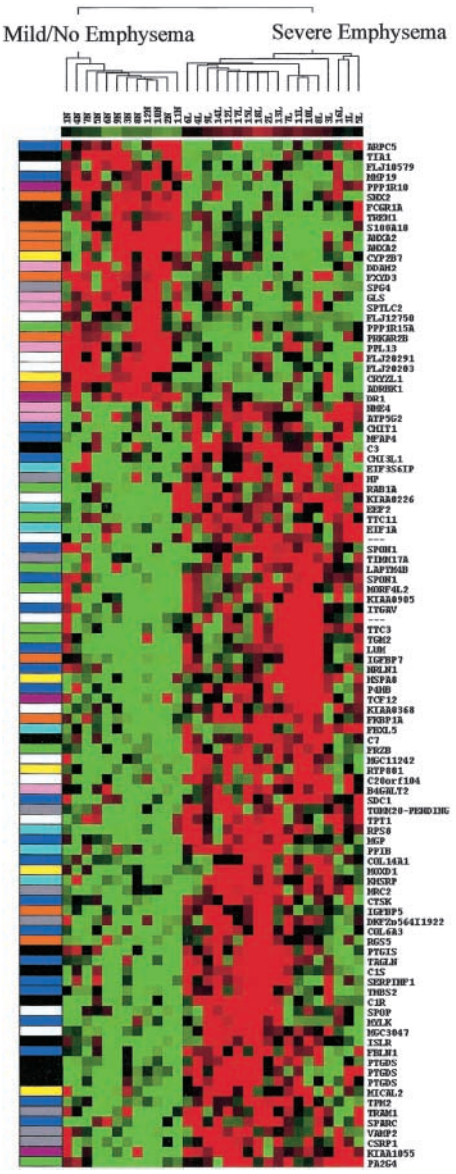

Figure 1. Class prediction genes that distinguish severe from mild/no emphysema. Two-dimensional hierarchical clustering of all 30 samples according to the expression of the 102 genes chosen by 4 or more class prediction algorithms to distinguish severe from mild/no emphysema. HUGO gene ID is listed for all 102 genes. Functional classification of genes (*color-coded*) is shown. Samples cluster with their appropriate clinical classes. *Red* represents a high level of gene expression, *green* represents a low level of gene expression, and *black* represents the mean level of expression. N = no/mild emphysema; L = severe emphysema. *Dark blue*, ECM/structural protein; *orange*, cell signaling; *black*, immune/inflammation; *yellow*, redox/stress; *pink*, metabolism; *red*, transcription factors; *light blue*, ribosomal/protein synthesis; *green*, cell cycle/apoptosis; *gray*, other; *white*, unknown.

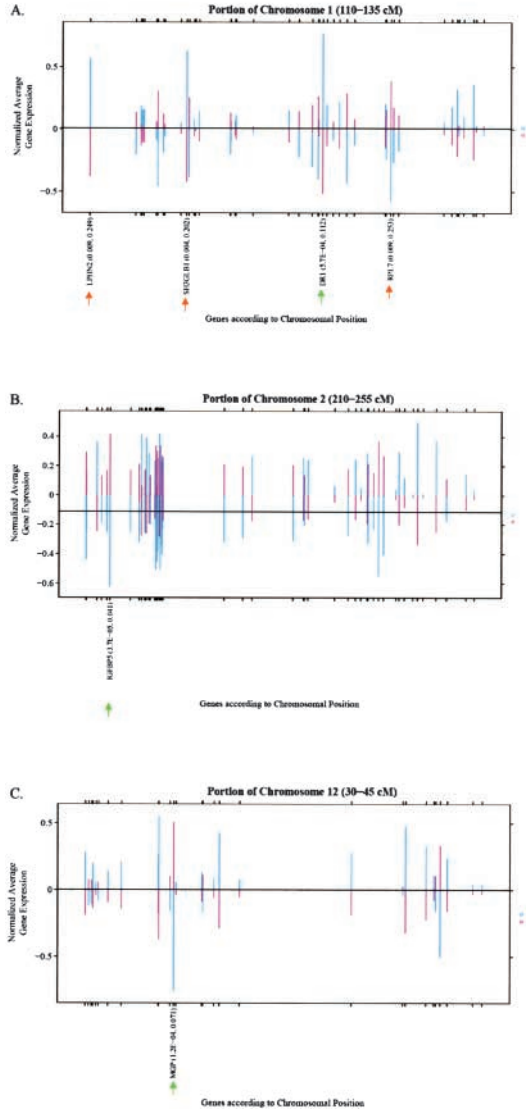

Figure 2. Selected chromosomal regions linked (LOD scores > 1) with early-onset COPD that contain at least 1 gene from the class prediction list of 102 genes that distinguish between severe and mild/no emphysema (*see* Figure 1). The average gene expression values for the severe (*red*) and mild/no emphysema (*blue*) groups of patients are plotted for each gene as *bars* across each chromosomal region (SD is not shown). The microarray expression data obtained from the 18 patients with severe emphysema and 12 patients with mild/no emphysema was normalized to a mean of 0 and an SD of 1 for each gene before computing the averages. Differential gene expression between severe and mild/no emphysema was assessed by a two-sample unequal-variance *t* test performed on log-transformed expression data. The *P* values and q values are reported for each gene with a *P* value < 0.01. The genes highlighted with a *green arrow* are on the list of 102 probe-sets that distinguish between severe and mild/no emphysema. The genes highlighted with an *orange arrow* are not on the list, but have a *P* value < 0.01.

Genes that Tightly Correlate with D_{LCO} and FEV_1

There were 92 genes whose level of expression strongly correlated ($P < 0.001$) with D_{LCO}; 41% were negatively correlated (i.e., were expressed at higher levels in patients with severe emphysema) and 59% were positively correlated (i.e., were expressed at lower levels in patients with severe emphysema). There were 73 genes that correlated significantly ($P < 0.001$) with FEV_1. Thirty-three of the above genes correlated with both D_{LCO} and FEV_1. Table 2 shows selected genes illustrating potential pathogenetic features of emphysema. Genes associated with redox stress and hypoxia were present in both lists, as were genes associated with apoptosis. Inflammation-related and immunomodulatory genes were also evident, although genes associated with lymphocyte maturation and with prostaglandin synthesis negatively correlated only with FEV_1. In addition, vascular genes positively correlated with these clinical measures of emphysema and were expressed at lower levels in severe emphysema. The complete list of genes correlated with D_{LCO}, FEV_1, and BMI can be found in the online supplement (Tables E4, E5, and E6).

TABLE 2. Correlation of gene expression with physiology. Selected genes whose expression tightly correlates with percent-predicted postbronchodilator forced expiratory volume at 1 s or forced expiratory volume at 1 s and diffusing capacity

Correlation, FEV_1			Correlation, FEV_1 and D_{LCO}		
Function	Gene	Accession	Function	Gene	Accession
Redox/stress			Redox/stress		
	MICAL2	BE965029		P4HB	J02783
	HIG2	NM_013332		SERP1	BG107676
	RTP801	NM_019058		CRYZL1	NM_005111
	MOXD1	AY007239		*SCL31A2*	NM_001860
Immune/inflammation			Immune/inflammation		
(in both)	PTGS1	S36219		SERP1	BG107676
	PTGDS	BC005939		C1R	AL573058
	FKBP1A	NM_000801	ECM		
	HCCR1	NM_015416		NRLN1	AL049176
	NOTCH2	AA291203		P4HB	J02783
	ID4	U16153		ELF3	AF017307
Secretion				ACTA2	NM_001613
	Tram1	BC000687		PFN2	NM_002628
	GORASP2	NM_015530		LUM	NM_002345
	SSR2	NM_003145	Apoptosis/cell cycle		
	SPUVE	NM_007173		LDOC1	NM_012317
Injury/repair				PA2G4	BF669264
	FOXF1	NM_001451		LUM	NM_002345
	klotho	NM_004795		MORF4L2	NM_012286
				PUM1	D87078
				ELF3	AF017307

Definition of abbreviations: D_{LCO}, lung diffusion capacity for carbon monoxide; ECM, extracellular matrix; FEV_1, forced expiratory volume at 1 s.

Mapping Gene Expression to Chromosomal Regions

Chromosomal regions containing STR markers with LOD scores > 1 (23–25) were chosen to investigate if genes that distinguish between severe and mild/no emphysema were located within these regions. Susceptibility loci were selected from the various linkage studies that included markers with LOD scores > 1, a statistically significant ($P < 0.05$) threshold in the above datasets. A total of 23 chromosomal regions were investigated, and 17 of the 23 different regions were found to contain 1 or more of the 102 class prediction probe-sets that distinguish severe from mild/no emphysema (Table E9). Several of these regions also contained additional genes, not present on the class prediction list, that are differentially expressed ($P < 0.01$) between severe versus mild/no emphysema (*see* Table E10). Regions from chromosomes 1, 2, and 12 are graphically displayed in Figure 2.

Class Discovery

Clustering dendograms were constructed for each of the 5 filtered gene lists derived from variation filters applied to the 9,336 genes expressed in all resected lungs (*see* Figure E1 in the online supplement). BMI was the only clinical variable in our study that was associated with the pattern of clustering. Each figure had one cluster of samples, ranging from 5–7 patients that had significantly lower BMI values ($P \leq 0.01$) when compared with the remaining samples outside the cluster. Four patients were present in the low BMI cluster in all of the five figures constructed. To identify genes differentially expressed between the patients in the lower BMI cluster, an unpaired unequal variance *t* test was performed between the 6 patients identified by clustering on the filtered group of 597 genes and the remaining 12 patients undergoing LVRS across all 9,336 genes. A total of 96 genes with a *P* value < 0.001 were identified (*see* Figure E2). Using EASE, KEGG, and PubMatrix, a number of functional categories were identified as overrepresented among this list of genes, including transcription factors, inflammatory mediators, and a number of genes within the mitogen-activated protein kinase pathway (Table 3).

Class Prediction: Outcome after LVRS

A total of 9 out of the 14 patients had BODE scores that improved at least one unit after surgery, whereas the remaining 5 patients had BODE scores that remained the same or worsened after surgery. Using the class prediction algorithms described above, 17 genes were found in common to 4 or more of the methods (Figure 3A). On leave-out-one cross-validation, 66–92% of samples were correctly classified (*see* Table E11). On MDS analysis, samples separated into their 2 classes according to the expression of these 17 genes (Figure 3B).

RT-PCR

Using RT-PCR, we validated the expression of 10 genes that were differentially expressed between severe emphysema and no/mild emphysema (*see* Figure 4).

Discussion

The pathogenetic mechanism of COPD in the lung involves an amplification of the normal inflammatory response to cigarette smoke, but the molecular reasons for this amplification have yet to be elucidated. Whereas the molecular mechanisms underlying COPD remain unclear, a number of pathways have been implicated, including an imbalance between protease and antiprotease activity, release of multiple inflammatory mediators with resulting high levels of oxidative stress, and apoptosis of alveolar cell walls (6). In addition, chronic inflammation and connective tissue deposition have been shown to narrow small conducting airways (5). In this high-throughput gene expression study of emphysematous human lung tissue, we have identified a large number of genes differentially expressed in the lungs of smokers with severe emphysema as compared with lungs of smokers with mild/no emphysema, along with a number of genes the expression level of which tightly correlates with the physiologic severity and phenotypic expression of emphysema. These gene expression profiles have the potential of providing insight into the molecular pathways involved in the pathogenesis of the disease. Further, gene expression modifications associated with low BMI and BODE response after surgery provide support for the importance of the search for the mechanisms responsible for the

TABLE 3. Select genes differentially expressed among patients with chronic obstructive pulmonary disease with a low body-mass index

Inflammation		Transcriptional Regulation		MAP Kinase		Response to Stress	
BC004490	FOS	M58297	ZNF42	NM_004417	DUSP1	NM_004417	DUSP1
BG491844	JUN	AL049942	ZNF337	BC003143	DUSP6	NM_005627	SGK
AI984479	MAX	M91083	c11orf13	K03193	EGFR	M91083	C11orf13
BE311760	HMGB1	BC004490	FOS	BC004490	FOS	BC004490	FOS
NM_014267	SCAP	BG491844	JUN	BG491844	JUN	BC005979	UBE2B
		AI984479	MAX	AI984479	MAX	BF246436	SUI1
		NM_005342	HMGB3	BC005365	MAP2K7	BC005365	MAP2K7
		NM_004235	KLF4			BC006325	GTSE
		NM_022454	SOX17			NM_002913	RFC1
		NM_014267	SCAP			NM_012072	C1QR1
		NM_002913	RFC1			NM_014267	SCAP
		NM_000435	NOTCH3				
		BE311760	HMGB1				
		K03193	EGR1				

Definition of abbreviation: MAP, mitogen-activated protein.
Functional categories generated with EASE, KEGG, and PubMatrix.

systemic manifestations of COPD and the possible identification of clinical tests that will predict the outcome of LVRS.

Class Prediction of COPD

It has long been proposed that various proteases break down connective tissue components in lung parenchyma to produce emphysema (6). The balance between protease and antiprotease activity in the lung parenchyma is hypothesized to be disrupted in patients with emphysema. Increased proteolysis in the lung leads to aberrant remodeling and/or degradation of the ECM. Cigarette smoking may induce inflammation and increase the release of proteases from both neutrophils and macrophages, and, in smokers who develop COPD, the production of antiproteases may be inadequate to neutralize these effects. In our study, surprisingly few proteases were upregulated in patients with severe emphysema. Those proteases included cathepsin K, a cysteine proteinase that degrades elastin, collagen, and gelatin, as well as the metalloproteinase, MMP2, which specifically cleaves type IV collagen (27), the major structural component of basement membranes. Interestingly, two antiproteases are upregulated in severe emphysema: SERPINF1, a neurotrophic protein belonging to the serine protease inhibitor (serpin) family (28), and TIMP1, a collegenase inhibitor (29).

One of the striking findings in our study is the large number of ECM-related genes upregulated in severe emphysema (Figure 1 and Table 2). In experimental emphysema produced by pancreatic or neutrophil elastase, connective tissue components are degraded, leading to enlargement of distal airspaces (30). Both elastin and collagen are rapidly resynthesized in these animal models, and mRNA levels for both are increased, but the connective tissue remodeling process is ineffective and lung mechanical properties remain abnormal (31). The increased levels of ECM mRNAs in severe emphysema (Figure 1) and those associated with low D_{LCO} (Table 2) suggest that connective tissue remodeling continues even in severe "end-stage" emphysema in humans, but that ECM-related proteins fail to effectively restore the mechanical properties of the emphysematous lung.

In COPD, chronic inflammation leads to a fixed narrowing of small airways and alveolar wall destruction. This inflammation is characterized by increased numbers of alveolar macrophages, neutrophils, and T lymphocytes, along with the release of multiple inflammatory mediators that result in a high level of oxidative stress (6). Multiple oxidant-related genes in our study were expressed at lower levels in severe emphysema and correlate positively with D_{LCO}. Specifically, the gene CRYZL1 bears structural resemblance to quinone reductase (32), a gene with polymorphisms that have been associated with increased susceptibility to oxidant damage (33). Diminished transcription of this product, along with decreased levels of the copper transporter SLC31A2, may promote inflammation or further impair the normal response to stress in emphysematous lung (34). Alteration of inflammatory pathways may further disrupt the normal response to stress, as suggested by the tight correlation between D_{LCO} and CD97, a gene involved in cell adhesion and signaling after leukocyte activation (35).

It has recently been reported that angiogenesis and apoptosis of the alveolar cell wall may play a role in emphysema (8). Blockade of vascular endothelial growth factor receptor 2 in rats induces apoptosis of the alveolar cell wall and results in an emphysema-like pathology (7). The role of vascular endothelial growth factor in human emphysema remains unclear. In our study, a number of angiogenesis-related genes were positively correlated with D_{LCO} (Table 2), including endothelial cell growth factor 1, which stimulates endothelial cell growth and chemotaxis *in vitro* and angiogenesis *in vivo* (36), and endomucin 2, which has a role in tumor angiogenesis (37). The decreased expression level of vascular-related genes in patients with low D_{LCO} (i.e., severe emphysema) likely does not play a causative role, but rather is secondary to the decreased vascular surface area in these patients. However, one of the antiproteases that was elevated in severe emphysema, SERPINF1, is also a potent inhibitor of angiogenesis through FasL-mediated apoptosis of endothelial cells (38). Although its role in precipitating apoptosis of endothelial cells has been limited to the retina, elevated levels of this gene in the lung of patients with severe emphysema suggests a possible pathologic role in mediating apoptosis of the vasculature and alveolar cell wall.

Correlation with FEV_1 and D_{LCO}

The low FEV_1 characteristic of COPD is a function of both intrinsic obstruction to airflow in smaller conducting airways and loss of lung elasticity reducing the recoil forces that tether intrapulmonary airways during forced expiration. Lung diffusing capacity is a function of the total gas exchange surface of the lung, which is decreased in proportion to the degree of emphysema. Contrasting genes, the expression level of which correlates with FEV_1 but not D_{LCO}, might represent biologic events intrinsic to small conducting airways, whereas genes that correlate with D_{LCO} would be more likely to reflect biologic processes involved in the genesis of emphysema in gas exchanging portions of the lung. These findings may represent the genotypic expressions of the different phenotypes frequently detected in clinical practice (39).

Expression of several genes associated with hypoxia or redox stress increase as FEV_1 worsens. Low FEV_1 is also strongly correlated with increased expression of golgi and endoplasmic reticulum translocation genes, along with a secretory serine–protease, consistent with increased mucus secretion recently documented in a histologic study of pulmonary emphysema (5). A number of genes associated with inflammation and with lymphocyte accumulation also tightly correlate with low FEV_1. Two prostaglandin synthases, the expression of which leads to lymphocytic inflammation, increase with decreasing FEV_1, as does NOTCH2, which potentiates lymphopoesis and differentiation of CD8 cells (40), and which inhibits expression of Id4, another gene that decreases as FEV_1 falls and which normally acts to block T-cell differentiation (41). Two genes, the absence of which has been associated with ineffective lung repair after inflammatory cell injury and with increased lung cell apoptosis, FoxF1 (42) and Klotho, also decrease with falling FEV_1. Klotho expression is suppressed in hypoxia and in inflammatory states, so its decreased expression may amplify the detrimental effects of airway inflammation.

Linking Gene Expression to Chromosomes and Candidate Genes

As illustrated by the above discussion, one of the challenges in analyzing microarray gene expression data is separation of genes causally involved in a disease from "bystander" genes, the expression levels of which have been secondarily altered by primary changes elsewhere (19). Several recent studies in yeast and mouse models have used microarray-measured gene expression levels as quantitative traits and genome-wide genotype data to identify genetic loci that can explain variations in expression levels (43–45). Schadt and colleagues (45), for example, used a mouse model of obesity and identified chromosomal loci linked to gene expression patterns associated with different subtypes of the complex disease. Human orthologs of the mouse chromosomal loci were identified as possible candidates for further studies of obesity in humans. Integrating various sources of clinical, genetic, and genomic data has proven to be a powerful strategy in elucidating complex diseases and, in this study, we have attempted to combine our gene expression data with genetic linkage studies to identify candidate genes that may be causally involved in the pathogenesis of COPD. Several chromosomal regions linked to various phenotypes of early onset COPD have been identified by Silverman and colleagues (23–25). We mapped genes differentially expressed between severe and mild/no emphysema onto these chromosomal regions to find gene candidates causally implicated in COPD. Future sequence analysis of single-nucleotide polymorphisms (SNPs) located within or near the genes we identified is needed to determine if an SNP or multiple SNPs are associated with COPD. Additional studies of these genes will also be needed to ascertain whether or not the sequence

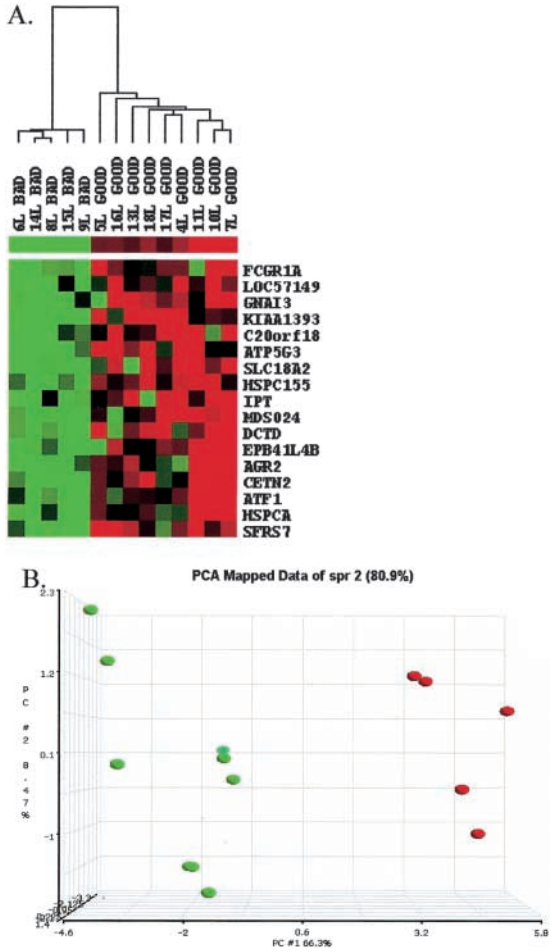

Figure 3. Gene expression profile that may predict outcome after LVRS. (*A*) Two-dimensional hierarchical clustering plot of LVRS samples according to the expression of the 17 genes chosen by 4 or more class-prediction algorithms to predict functional outcome at 6 mo after LVRS. Samples cluster with their appropriate class. Good = good outcome where there was improvement in BODE score at 6 mo after LVRS; bad = bad outcome where there was worsening or no change in BODE score at 6 mo after LVRS. *Red* represents a high level of gene expression, *green* represents a low level of gene expression, and *black* represents the mean level of expression. (*B*) Multidimensional scaling plot of good (*green boxes*) and bad (*red boxes*) outcome patients undergoing LVRS in 17 dimensional space (represented in 30 space in this figure) according to the expression of the 17 genes that predict outcome in these patients. The two outcome groups separate into their two classes according to the expression of these genes.

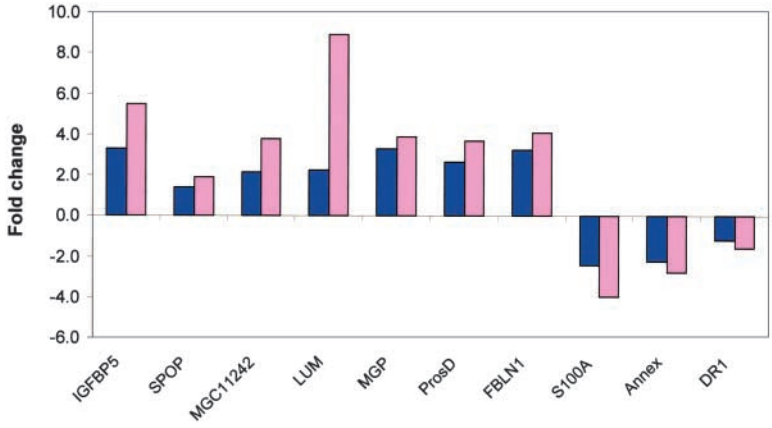

Figure 4. Quantitative RT-PCR (*blue bars*) and microarray (*pink bars*) data for select genes found to be differentially expressed between severe emphysema and normal lung tissue on microarray analysis. For all 10 genes, gene expression was measured, via quantitative RT-PCR, on 2 normal lung tissues and 2 lungs with severe emphysema. For each of those four samples, the level of expression of the gene was calculated relative to one of the non/mild emphysematous lungs. Fold change refers to the average level of expression of the gene across both severe emphysema samples divided by the average level of expression across both normal lung samples.

variations alter the expression of the genes and in turn modulate the disease phenotype. Although a number of differentially expressed genes mapped into the chromosomal regions (*see* Table E10), we have focused, in the discussion below, on two illustrative examples where suggestive or significant linkage to various phenotypes of COPD were found.

Whereas upregulation of ECM-related genes in severe emphysema likely represents a compensatory increase in the expression of those connective tissue components, matrix γ-carboxyglutamic acid protein (MGP), an ECM-related gene, mapped onto a region of chromosome 12 linked to FEV_1, suggesting a causal role in the pathogenesis of this disease (*see* Figure 2). MGP has recently been reported to play a role in lung growth and development, likely via temporally and spatially specific interactions with other branching morphogenesis–related proteins (46). Whether aberrant expression of MGP results in an ineffective connective tissue remodeling process is unclear, but further study of this gene is warranted. Insulin-like growth factor binding protein (IGFBP) 5 is another gene upregulated in severe emphysema, and this gene mapped on to a region of chromosome 2 strongly linked to FEV_1/forced vital capacity. IGFBPs bind IGF with high affinity, preventing IGF binding to its receptors and signaling. IGFBP5-overexpressing transgenic mice have recently been shown to have decreased growth and muscle development (47). IGFBPs have also been reported to act through IGF-independent mechanisms (48) and to affect apoptosis and cell survival (49). Further functional studies of MGP and IGFBP5 and their role in the lung are needed, as is genotyping of SNPs (in and around these genes) in cohorts with and without COPD, to investigate the genes' association with the disease.

Molecular Subclass of COPD

In addition to providing insight into disease pathogenesis, high-throughput gene expression analysis of emphysematous lung tissue has uncovered potential novel molecular subclasses of the disease. Several different approaches to unsupervised clustering of all LVRS samples revealed 4–6 patients who grouped together across each analysis. Compared with the remainder of the patients undergoing LVRS, this subgroup demonstrated a significantly decreased BMI. COPD is increasingly recognized as a disease of inflammation that reaches beyond the lungs (14, 50). Although the precise mechanism of weight loss in COPD has not been described, it appears to occur largely through muscle wasting, and resembles cachexia as opposed to weight loss from caloric restriction (14). Among this class of patients in our study, the significant number of differentially expressed genes involved in gene transcription, response to stress, and inflammation (Table 3) may reflect an active, persistent inflammatory process initiated in the lung that results in systemic effects such as decreased BMI. Identification of a subgroup of patients with COPD with more pronounced inflammation, using biomarkers in serum, could guide the use of steroids and other anti-inflammatory medications to prevent the damaging local and systemic responses.

Gene Expression May Predict LVRS Outcome

For patients with severe COPD with inhomogeneous emphysema, medical therapy has proven largely ineffective in improving dyspnea and functional status, and does not alter pulmonary function. LVRS has been proposed as a palliative treatment for certain subgroups of COPD patients with emphysema, but initial enthusiasm over its application had been confounded by uncertainty about the potential cost and morbidities associated with LVRS, as well as its beneficial effects (51). Recent reports suggest that only a subset of patients with emphysema benefit from LVRS (11), and refinement in patient selection remains a current goal in the surgical approach to COPD (51). Although our sample size is small and cross-validation results are inconclusive, our study raises the possibility that gene expression profiles in severely emphysematous lung tissue may predict functional outcome after LVRS. Given the short duration of follow-up in our study, we elected to use a surrogate marker of survival in emphysema, the BODE index, which predicts the risk of respiratory and all-cause mortality among patients with COPD (15). If our gene expression signature can be further developed, validated, and refined on a larger number of subjects undergoing LVRS, we may be able to develop assays to measure the protein products of these genes in the serum. These serum protein products could then be used as a noninvasive preoperative test to select patients who are likely to benefit from the procedure.

Limitations

There are a number of important limitations to our study. Given the limited number of lung tissue specimens available from LVRS, our sample size for the microarray analyses performed in the study was relatively small. For the class prediction analyses, we attempted to address the small sample sizes by the process of cross-validation, by selecting genes chosen via multiple class prediction methodologies, and by validating changes for a subset of genes using quantitative RT-PCR. For the class discovery, we performed hierarchical clustering of LVRS samples on a number of different gene lists (determined by variability filters) to test for a consistent subset of samples that clustered separately. Given the multiple comparison problem inherent in our correlation analyses, we selected a P value threshold of 0.001 for significance, and have validated a number of the more highly correlated genes via RT-PCR.

Additional studies on larger numbers of severely emphysematous lung tissue are needed to validate our findings and to address variability in the cellular composition of the tissues given the heterogeneous nature of the disease. A broader spectrum of emphysema subjects (i.e., those with mild and moderate disease) is needed to identify those genes differentially expressed between mild/moderate emphysema and no emphysema, which may be implicated in the early pathogenesis of the disease. Due to the limited sample size of our study, smokers without emphysema were grouped together with smokers having mild emphysema as a control group for our patients undergoing LVRS. Comparison of smokers without emphysema ($n = 5$) versus those with mild emphysema ($n = 7$) yielded fewer differentially expressed genes than would be expected by chance (data not shown), although our study was not powered for this subgroup analysis. However, hierarchical clustering of all 12 control subjects according to the expression of the 102 genes that distinguish subjects of LVRS versus control subjects did not result in separation of those smokers with mild emphysema from those with normal spirometry (Figure 1). An additional limitation to our control group was the fact that a significant number of these patients ($n = 9$) had primary lung cancer adjacent to the "normal" lung tissue studied, raising the concern that gene expression in the "normal" lung may have been influenced by the adjacent tumor. Whereas our study was inadequately powered to address the effects of cancer on the surrounding normal lung tissue, a nonparametric t test comparing the nine subjects with cancer versus the three subjects with benign nodules yielded fewer genes at any given P value threshold than would be expected by chance (data not shown). In addition, unsupervised analysis of all 12 control samples according to the expression of all genes that passed a number of variability filters did not reveal clustering of the cancer and benign nodule subjects into separate groups (*see* Figure E3). Finally, although combining genetic and genomic datasets can be a powerful methodology, it is unclear whether or

not the linkage analysis data in the early-onset COPD studies can be generalized to the COPD population at-large.

In summary, we have identified genes the expression levels of which can distinguish severely emphysematous lung from normal lung tissue, as well as genes that tightly correlate with overall lung function. By linking these gene expression profiles to chromosomal regions previously associated, via genome-wide linkage analyses, with severe early-onset COPD, we have identified candidate genes that may be causally involved in the pathogenesis of COPD. We also have documented a possible association between gene expression and extrapulmonary manifestations of the disease, such as low BMI. To our knowledge, this is the first study to combine high-throughput gene expression studies with genome-wide linkage analysis in human lung disease. A search for functional polymorphisms in these candidate genes that may predispose a smoker to COPD may be warranted.

Conflict of Interest Statement: A.S. has no declared conflict of interest; J.B. has no declared conflict of interest; V.P.-P. has no declared conflict of interest; A.K. has no declared conflict of interest; G.L. has no declared conflict of interest; V.S. has no declared conflict of interest; B.C. has no declared conflict of interest; and J.S.B. has no declared conflict of interest.

Acknowledgments: The authors thank Dr. Robert Berger for support in obtaining clinical data on OBEST subjects and Dr. Mike Pedrick for histologic review of the LVRS samples. Supported in part by a Doris Duke Charitable Foundation Clinical Scientist Development Award (A.S.), National Institutes of Health grant HL71771 (J.S.B.), and ES10377 (J.S.B.). This project was also partially supported by a grant from GlaxoSmithKline (B.C.).

References

1. Ezzati, M., and A. D. Lopez. 2003. Estimates of global mortality attributable to smoking in 2000. *Lancet* 362:847–852.
2. Anderson, R., and B. Smith. 2003. Deaths: leading causes for 2001. *Natl. Vital Stat. Rep.* 2003 Nov 7;52:1–85.
3. Mannino, D. M. 2003. Chronic obstructive pulmonary disease: definition and epidemiology. *Respir. Care* 48:1185–1191.
4. Pauwels, R. A., A. S. Buist, P. M. Calverley, C. R. Jenkins, and S. S. Hurd. 2001. 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.* 163:1256–1276.
5. Hogg, J. C., F. Chu, S. Utokaparch, R. Woods, W. M. Elliott, L. Buzatu, R. M. Cherniack, R. M. Rogers, F. C. Sciurba, H. O. Coxson, and P. D. Pare. 2004. The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N. Engl. J. Med.* 350:2645–2653.
6. Barnes, P. J., S. D. Shapiro, and R. A. Pauwels. 2003. Chronic obstructive pulmonary disease: molecular and cellular mechanisms. *Eur. Respir. J.* 22: 672–688.
7. Tuder, R. M., L. Zhen, C. Y. Cho, L. Taraseviciene-Stewart, Y. Kasahara, D. Salvemini, N. F. Voelkel, and S. C. Flores. 2003. Oxidative stress and apoptosis interact and cause emphysema due to vascular endothelial growth factor receptor blockade. *Am. J. Respir. Cell Mol. Biol.* 29:88–97.
8. Aoshiba, K., N. Yokohori, and A. Nagai. 2003. Alveolar wall apoptosis causes lung destruction and emphysematous changes. *Am. J. Respir. Cell Mol. Biol.* 28:555–562.
9. Fehrenbach, H. 2002. Animal models of chronic obstructive pulmonary disease: some critical remarks. *Pathobiology* 70:277–283.
10. Mahadeva, R., and S. D. Shapiro. 2002. Chronic obstructive pulmonary disease * 3: experimental animal models of pulmonary emphysema. *Thorax* 57:908–914.
11. Fishman, A., F. Martinez, K. Naunheim, S. Piantadosi, R. Wise, A. Ries, G. Weinmann, and D. E. Wood. 2003. A randomized trial comparing lung–volume–reduction surgery with medical therapy for severe emphysema. *N. Engl. J. Med.* 348:2059–2073.
12. Schols, A. M., J. Slangen, L. Volovics, and E. F. Wouters. 1998. Weight loss is a reversible factor in the prognosis of chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 157:1791–1797.
13. Maltais, F., A. A. Simard, C. Simard, J. Jobin, P. Desgagnes, and P. LeBlanc. 1996. Oxidative capacity of the skeletal muscle and lactic acid kinetics during exercise in normal subjects and in patients with COPD. *Am. J. Respir. Crit. Care Med.* 153:288–293.
14. Agusti, A. G., A. Noguera, J. Sauleda, E. Sala, J. Pons, and X. Busquets. 2003. Systemic effects of chronic obstructive pulmonary disease. *Eur. Respir. J.* 21:347–360.
15. Celli, B. R., C. G. Cote, J. M. Marin, C. Casanova, D. O. Montes, R. A. Mendez, V. Pinto Plata, and H. J. Cabral. 2004. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N. Engl. J. Med.* 350:1005–1012.
16. Golpon, H. A., M. W. Geraci, M. D. Moore, H. L. Miller, G. J. Miller, R. M. Tuder, and N. F. Voelkel. 2001. HOX genes in human lung: altered expression in primary pulmonary hypertension and emphysema. *Am. J. Pathol.* 158:955–966.
17. Zhang, W., S. D. Yan, A. Zhu, Y. S. Zou, M. Williams, G. C. Godman, B. M. Thomashow, M. E. Ginsburg, D. M. Stern, and S. F. Yan. 2000. Expression of Egr-1 in late stage emphysema. *Am. J. Pathol.* 157:1311–1320.
18. Spira, A., J. Beane, V. Shah, G. Liu, F. Schembri, X. Yang, J. Palma, and J. S. Brody. 2004. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc. Natl. Acad. Sci. USA* 101:10143–10148.
19. Miklos, G. L., and R. Maleszka. 2004. Microarray reality checks in the context of a complex disease. *Nat. Biotechnol.* 22:615–621.
20. Tibshirani, R., T. Hastie, B. Narasimhan, and G. Chu. 2002. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. USA* 99:6567–6572.
21. Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531–537.
22. Li, L., C. R. Weinberg, T. A. Darden, and L. G. Pedersen. 2001. Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics* 17:1131–1142.
23. Palmer, L. J., J. C. Celedon, H. A. Chapman, F. E. Speizer, S. T. Weiss, and E. K. Silverman. 2003. Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. *Hum. Mol. Genet.* 12:1199–1210.
24. Silverman, E. K., J. D. Mosley, L. J. Palmer, M. Barth, J. M. Senter, A. Brown, J. M. Drazen, D. J. Kwiatkowski, H. A. Chapman, E. J. Campbell, M. A. Province, D. C. Rao, J. J. Reilly, L. C. Ginns, F. E. Speizer, and S. T. Weiss. 2002. Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. *Hum. Mol. Genet.* 11:623–632.
25. Silverman, E. K., L. J. Palmer, J. D. Mosley, M. Barth, J. M. Senter, A. Brown, J. M. Drazen, D. J. Kwiatkowski, H. A. Chapman, E. J. Campbell, M. A. Province, D. C. Rao, J. J. Reilly, L. C. Ginns, F. E. Speizer, and S. T. Weiss. 2002. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. *Am. J. Hum. Genet.* 70:1229–1239.
26. Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 100:9440–9445.
27. Collier, I. E., S. M. Wilhelm, A. Z. Eisen, B. L. Marmer, G. A. Grant, J. L. Seltzer, A. Kronberger, C. S. He, E. A. Bauer, and G. I. Goldberg. 1988. H-ras oncogene–transformed human bronchial epithelial cells (TBE-1) secrete a single metalloprotease capable of degrading basement membrane collagen. *J. Biol. Chem.* 263:6579–6587.
28. Steele, F. R., G. J. Chader, L. V. Johnson, and J. Tombran-Tink. 1993. Pigment epithelium–derived factor: neurotrophic activity and identification as a member of the serine protease inhibitor gene family. *Proc. Natl. Acad. Sci. USA* 90:1526–1530.
29. Docherty, A. J., A. Lyons, B. J. Smith, E. M. Wright, P. E. Stephens, T. J. Harris, G. Murphy, and J. J. Reynolds. 1985. Sequence of human tissue inhibitor of metalloproteinases and its identity to erythroid-potentiating activity. *Nature* 318:66–69.
30. Snider, G. L., E. C. Lucey, and P. J. Stone. 1986. Animal models of emphysema. *Am. Rev. Respir. Dis.* 133:149–169.
31. Lucey, E. C., R. H. Goldstein, P. J. Stone, and G. L. Snider. 1998. Remodeling of alveolar walls after elastase treatment of hamsters. Results of elastin and collagen mRNA *in situ* hybridization. *Am. J. Respir. Crit. Care Med.* 158:555–564.
32. Kim, M. Y., H. K. Lee, J. S. Park, S. H. Park, H. B. Kwon, and J. Soh. 1999. Identification of a zeta-crystallin (quinone reductase)–like 1 gene (CRYZL1) mapped to human chromosome 21q22.1. *Genomics* 57:156–159.
33. Bergamaschi, E., G. De Palma, P. Mozzoni, S. Vanni, M. V. Vettori, F. Broeckaert, A. Bernard, and A. Mutti. 2001. Polymorphism of quinone-metabolizing enzymes and susceptibility to ozone-induced acute effects. *Am. J. Respir. Crit. Care Med.* 163:1426–1431.
34. Lominadze, D., J. T. Saari, S. S. Percival, and D. A. Schuschke. 2004. Proinflammatory effects of copper deficiency on neutrophils and lung endothelial cells. *Immunol. Cell Biol.* 82:231–238.
35. Hamann, J., W. Eichler, D. Hamann, H. M. Kerstens, P. J. Poddighe, J. M. Hoovers, E. Hartmann, M. Strauss, and R. A. van Lier. 1995. Expression cloning and chromosomal mapping of the leukocyte activation antigen CD97, a new seven-span transmembrane molecule of the secretion receptor superfamily with an unusual extracellular domain. *J. Immunol.* 155:1942–1950.
36. Ishikawa, F., K. Miyazono, U. Hellman, H. Drexler, C. Wernstedt, K. Hagiwara, K. Usuki, F. Takaku, W. Risau, and C. H. Heldin. 1989. Identification of angiogenic activity and the cloning and expression of platelet-derived endothelial cell growth factor. *Nature* 338:557–562.
37. Liu, C., Z. M. Shao, L. Zhang, P. Beatty, M. Sartippour, T. Lane, E. Livingston, and M. Nguyen. 2001. Human endomucin is an endothelial marker. *Biochem. Biophys. Res. Commun.* 288:129–136.
38. Volpert, O. V., T. Zaichuk, W. Zhou, F. Reiher, T. A. Ferguson, P. M. Stuart, M. Amin, and N. P. Bouck. 2002. Inducer-stimulated Fas targets activated

endothelium for destruction by anti-angiogenic thrombospondin-1 and pigment epithelium–derived factor. *Nat. Med.* 8:349–357.
39. Celli, B. R., and W. MacNee. 2004. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. *Eur. Respir. J.* 23:932–946.
40. Witt, C. M., V. Hurez, C. S. Swindle, Y. Hamada, and C. A. Klug. 2003. Activated Notch2 potentiates CD8 lineage maturation and promotes the selective development of B1 B cells. *Mol. Cell. Biol.* 23:8637–8650.
41. Rivera, R., and C. Murre. 2001. The regulation and function of the Id proteins in lymphocyte development. *Oncogene* 20:8308–8316.
42. Kalinichenko, V. V., Y. Zhou, B. Shin, D. B. Stolz, S. C. Watkins, J. A. Whitsett, and R. H. Costa. 2002. Wild-type levels of the mouse Forkhead Box f1 gene are essential for lung repair. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 282:L1253–L1265.
43. Jansen, R. C., and J. P. Nap. 2001. Genetical genomics: the added value from segregation. *Trends Genet.* 17:388–391.
44. Brem, R. B., G. Yvert, R. Clinton, and L. Kruglyak. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* 296:752–755.
45. Schadt, E. E., S. A. Monks, T. A. Drake, A. J. Lusis, N. Che, V. Colinayo, T. G. Ruff, S. B. Milligan, J. R. Lamb, G. Cavet, P. S. Linsley, M. Mao, R. B. Stoughton, and S. H. Friend. 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422:297–302.
46. Gilbert, K. A., and S. R. Rannels. 2004. Matrix GLA protein modulates branching morphogenesis in fetal rat lung. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 286:L1179–L1187.
47. Salih, D. A., G. Tripathi, C. Holding, T. A. Szestak, M. I. Gonzalez, E. J. Carter, L. J. Cobb, J. E. Eisemann, and J. M. Pell. 2004. Insulin-like growth factor–binding protein 5 (Igfbp5) compromises survival, growth, muscle development, and fertility in mice. *Proc. Natl. Acad. Sci. USA* 101:4314–4319.
48. Schneider, M. R., E. Wolf, A. Hoeflich, and H. Lahm. 2002. IGF-binding protein-5: flexible player in the IGF system and effector on its own. *J. Endocrinol.* 172:423–440.
49. Butt, A. J., K. A. Dickson, F. McDougall, and R. C. Baxter. 2003. Insulin-like growth factor-binding protein-5 inhibits the growth of human breast cancer cells *in vitro* and *in vivo*. *J. Biol. Chem.* 278:29676–29685.
50. Gan, W. Q., S. F. Man, A. Senthilselvan, and D. D. Sin. 2004. Association between chronic obstructive pulmonary disease and systemic inflammation: a systematic review and a meta-analysis. *Thorax* 59:574–580.
51. Trow, T. K. 2004. Lung-volume reduction surgery for severe emphysema: appraisal of its current status. *Curr. Opin. Pulm. Med.* 10:128–132.