Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans

Fengrong Zuo*†, Naftali Kaminski*‡§, Elsie Eugui†, John Allard†, Zohar Yakhini¶, Amir Ben-Dor¶, Lance Lollini†, David Morris‖, Yong Kim†, Barbara DeLustro†, Dean Sheppard‖, Annie Pardo, Moises Selman††, and Renu A. Heller†**

†Roche Bioscience, Palo Alto, CA 94304; ‡Functional Genomics, and Institute of Respiratory Medicine, Sheba Medical Center, Tel Hashomer, 52621 Israel; ¶Agilent Technologies, Haifa, 32000 Israel; ‖Lung Biology Center, Cardiovascular Research Institute, and Department of Medicine, University of California, San Francisco, CA 94143; and **Facultad de Ciencias, Universidad Nacional Autonoma de Mexico, and ††Instituto Nacional de Enfermedades Respiratorias, Mexico City, 14080 Mexico

Communicated by Matthew P. Scott, Stanford University School of Medicine, Stanford, CA, March 7, 2002 (received for review October 11, 2001)

Pulmonary fibrosis is a progressive and largely untreatable group of disorders that affects up to 100,000 people on any given day in the United States. To elucidate the molecular mechanisms that lead to end-stage human pulmonary fibrosis we analyzed samples from patients with histologically proven pulmonary fibrosis (usual interstitial pneumonia) by using oligonucleotide microarrays. Gene expression patterns clearly distinguished normal from fibrotic lungs. Many of the genes that were significantly increased in fibrotic lungs encoded proteins associated with extracellular matrix formation and degradation and proteins expressed in smooth muscle. Using a combined set of scoring systems we determined that matrilysin (matrix metalloproteinase 7), a metalloprotease not previously associated with pulmonary fibrosis, was the most informative increased gene in our data set. Immunohistochemisry demonstrated increased expression of matrilysin protein in fibrotic lungs. Furthermore, matrilysin knockout mice were dramatically protected from pulmonary fibrosis in response to intratracheal bleomycin. Our results identify matrilysin as a mediator of pulmonary fibrosis and a potential therapeutic target. They also illustrate the power of global gene expression analysis of human tissue samples to identify molecular pathways involved in clinical disease.

usual interstitial pneumonia | microarray analysis | informative genes | bleomycin | matrix metalloproteases

The molecular mechanisms that lead to end-stage pulmonary fibrosis are poorly understood. Although a multitude of conditions, including granulomatous disorders, exposure to environmental dusts and toxins, autoimmune diseases, and drugs, have been associated with pulmonary fibrosis, in many cases the causative agent remains unknown. The most lethal and refractory form of pulmonary fibrosis is idiopathic pulmonary fibrosis (IPF). IPF is defined histologically by the findings of usual interstitial pneumonia (UIP), characterized by the interposition of patchy foci of active fibroblast proliferation associated with minimal inflammation, areas of advanced scarring and fibrosis (honeycombing), and normal-appearing lung (1, 2). The etiology of IPF/UIP has remained elusive, and the mean survival from the of diagnosis is only 3 years regardless of treatment (3).

Much of our understanding of the mechanisms of pulmonary fibrosis comes from studies of bleomycin-induced lung fibrosis in mice. The common paradigm postulates an initial alveolar injury, which triggers an inflammatory response and, subsequently, fibrosis (4). The mainstay of the current therapeutic approach to UIP is suppression of inflammation. However, this approach is ineffective in most patients (3). Recently, evidence has emerged that UIP may be the result of disordered fibroblast regulation. Under this paradigm, the pathological process of fibrosis is the result of persistent remodeling of the lung interstitium by fibroblasts and smooth muscle cells rather than the result of persistent inflammation (5).

The advent of microarray technology allows the simultaneous monitoring of the transcriptional behavior of thousands of genes. This technology has been repeatedly shown to be useful in the analysis of the response of a variety of cellular systems to stimuli, in the classification of human cancer, and in the analysis of animal models of human disease (6–8). To characterize the transcriptional profile of UIP, we analyzed gene expression patterns in lung tissue from patients with UIP by using oligonucleotide microarrays and compared them with gene expression patterns in normal lung tissue. Our results demonstrate a clear distinction between gene expression patterns in the lungs of patients with UIP and control lungs. In addition to the expected increase in extracellular matrix proteins, we found a coordinated induction of genes that encode metalloproteases. One of them, matrilysin [matrix metalloproteinase (MMP) 7], was the gene that was most distinctive between fibrotic and normal lungs. To analyze the role of matrilysin in pulmonary fibrosis we injected bleomycin into the tracheas of matrilysin knockout mice and found that matrilysin knockout mice are protected from bleomycin-induced pulmonary fibrosis. Our data suggest that matrilysin is a key regulator of pulmonary fibrosis in mice and humans. Blocking matrilysin activity may thus serve as a novel therapeutic intervention in this devastating disease.

Methods

Lung Specimens. For microarray analysis lung specimens from patients with pulmonary fibrosis were either obtained from lungs removed for lung transplantation (patients 4 and 5) or by diagnostic thoracoscopy (patients 6–8). Only one of the patients had been treated with corticosteroids at the time of biopsy (patient 8, prednisone 15 mg/day). Control specimens were obtained from the normal parts of lungs removed for lung cancer (patients 1–3). An additional control was a purchased pool of total RNA from five lungs obtained from CLONTECH (sample 0). Patients with pulmonary fibrosis (two smokers, three nonsmokers) showed a restrictive physiology with decreased diffusion capacity, ground glass attenuation and honeycombing on high-resolution CT scan, and a typical pattern of UIP on lung histology. Two suffered from an autoimmune disease (patient 7: rheumatoid arthritis; patient 8: Sjogren's syndrome). The other three patients were diagnosed as having IPF/UIP. The experimental protocol was approved by the Institutional Internal

Abbreviations: IPF, idiopathic pulmonary fibrosis; UIP, usual interstitial pneumonia; MMP, matrix metalloproteinase; TNoM, threshold number of misclassifications.

*F.Z. and N.K. contributed equally to this work.

§To whom reprint requests should be addressed: Functional Genomics, Institute of Respiratory Medicine, Sheba Medical Center, Tel Hashomer, 52621, Israel. E-mail: kamins@sheba.health.gov.il.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

 www.pnas.org/cgi/doi/10.1073/pnas.092134099

Review Board. All surgical specimens for gene expression analysis were snap-frozen in liquid nitrogen.

For immunohistochemistry, lung samples were obtained from seven other patients with IPF/UIP (five male and two female; five nonsmokers, two ex-smokers), aged 59 ± 6 years by open lung biopsy. None of the patients had been treated with corticosteroids or immunosuppressive drugs. The diagnosis of IPF was based on clinical, radiological, and functional findings and was corroborated by the characteristic morphology of UIP. Control lung tissue samples were obtained from autopsies of patients who died from nonlung-related causes (two men and one woman, mean age 48 ± 11 years).

Preparation of Labeled cRNA. Probe preparation was performed as recommended by the manufacturer of the microarrays (9). Briefly, total RNA was isolated by homogenization of lung specimens in ice-cold Trizol. Twice-purified poly(A) mRNA was isolated (Oligotex, Qiagen, Chatsworth, CA) and used as template for double-stranded cDNA synthesis with an oligo$(dT)_{24}$ primer containing a T7 RNA polymerase promoter site added to the 3′ end (Genset, La Jolla, CA). The cDNA was extracted with phenol/chloroform, ethanol-precipitated, and used as a template for *in vitro* transcription (Ambion T7 Megascript system) with biotin-labeled nucleotides (Enzo Diagnostics). Labeled cRNA was fragmented and a hybridization mix was generated as recommended (9).

Hybridization of Microarrays. Aliquots of each sample (10 μg cRNA in 200 μl hybridization mix) were hybridized to a Genechip Hugene FL array. After hybridization, each array was washed, stained with streptavidin phycoerythrin (Molecular Probes), washed again, hybridized with biotin labeled anti-streptavidin phycoerythrin antibodies, restained with streptavidin phycoerythrin (Molecular Probes), and scanned (Hewlett–Packard, GeneArray scanner G2500A).

Analysis of GENECHIP Data. Scanned output files were visually inspected for hybridization artifacts. Arrays lacking significant artifacts were analyzed by using GENECHIP 3.3 software (Affymetrix, Santa Clara, CA). Arrays were scaled to an average intensity of 150 per gene and analyzed independently. The expression value for each gene was determined by calculating the average of differences (perfect match intensity minus mismatch intensity) of the probe pairs in use for this gene. The expression analysis files created by GENECHIP 3.3 software were then transferred to a database (Microsoft Access) linked to internet genome databases (e.g., National Heart, Lung, and Blood Institute, Swiss Prot, and GeneCards) to update gene definitions. A value of 20 was assigned to all measurements lower than 20. For cluster analysis we used CLUSTER and TREEVIEW programs described by Michael Eisen *et al*. (10). We did not include in the analysis genes that did not have at least one average difference intensity value ≥100 or one present call by Affymetrix criteria. Fold ratios were calculated for each sample against the median of the controls. Average fold ratios are the averages of the fold ratios of the UIP samples.

Separation Studies. To determine the most informative genes we used scoring methods previously described by Ben-Dor *et al.* (11) and applied to the analysis of breast cancer and melanoma by using cDNA arrays (12). For detailed descriptions see ref. 13. In brief, a gene is designated as informative based on the degree to which its tissue expression level is predictive of an independent classification of the tissue sample as ''diseased'' or ''not diseased'' (11). The scores used in this study were:

Threshold number of misclassifications (TNoM). TNoM is count of the number of classification errors committed when using the best simple threshold to distinguish between two classes (diseased or not diseased) based on the expression levels of a specific gene.

Info. Info is an estimate of the uncertainty remaining about the sample classification (diseased or not diseased) after the incorporation of predictions based on expression of an individual gene is given (a lower Info score indicates a higher predictive value for a given gene).

Gaussian. Gaussian is the overlap between distributions of expression levels for genes in two classes. The score is based on normality assumptions.

Immunohistochemistry. Immunohistochemistry was performed as described (14). Briefly, after deparaffinization and rehydration tissue sections were blocked with 3% H_2O_2 in methanol, followed by antigen retrieval in citrate buffer (10 mM pH 6.0). Lung sections were incubated with normal serum for 30 min, followed by overnight incubation at 4°C with primary monoclonal anti-MMP-1 (20 mg/ml), anti-MMP-9 (5 mg/ml) (Fuji), or anti-MMP-7 (20 mg/ml) (Chemicon). A secondary biotinylated anti-Ig was applied followed by horseradish peroxidase-conjugated streptavidin (BioGenex Laboratories, San Ramon, CA) according to the manufacturer. 3-Amino-9-ethyl-carbazole (BioGenex Laboratories) in acetate buffer containing 0.05% H2O2 was used as substrate. The sections were counterstained with hematoxylin. The primary antibody was replaced by non-immune serum for negative control slides.

Bleomycin Treatment. Age- and sex-matched, 8- to 16-week-old 129/Sv *MMP*-$7^{+/+}$ (The Jackson Laboratory) and *MMP*-$7^{-/-}$ mice (a gift from L. M. Matrisian, Vanderbilt University School of Medicine, Nashville), and 14- to 23-week-old C57BL/6 *MMP*-$7^{+/+}$ (Charles River Breeding Laboratories) and *MMP*-$7^{-/-}$ (L. M. Matrisian) mice were maintained in a specific pathogen-free environment. Mice were anesthetized by methoxyflurane, and a 24-gauge needle was inserted into the trachea by means of the oral cavity. Fifty microliters of bleomycin (0.05–0.08 units in 0.9% saline, Sigma) or saline was slowly injected. Mice were killed 14 or 21 days after bleomycin or saline injection, and lungs were collected for either hydroxyproline determination or histology.

Hydroxyproline Assay. To estimate the total amount of collagen in the lungs, hydroxyproline was measured as described (15). In brief, both lungs were removed and submitted to acid hydrolysis in 5 ml of 6 M HCl for 21 h at 110°C. Precipitates were removed by centrifugation. Supernatants were dried overnight and dissolved in 100 μl of distilled water at room temperature. Each sample was tested in duplicate. Fifty microliters of chloramine T (Sigma) in acetate-citrate buffer (pH 6.0), 50 μl of perchloric acid, and 50 μl of Erlich's reagent was added sequentially to each sample as described (15). Absorbance was measured at 570 nm, and the amount of hydroxyproline was calculated.

Histology and Morphometry. Murine lungs were fixed in 10% formalin by perfusion followed by submersion and processed by routine histologic methods through paraffin into hematoxylin/eosin-stained sections for evaluation by light microscopy. A single longitudinal section of each lung lobe was examined from each mouse. Scores were assigned for pulmonary fibrosis, increased numbers of pulmonary alveolar macrophages, the presence of leukocytes other than macrophages within alveoli, and interstitial leukocytic infiltrates. Each parameter was marked as being essentially normal or given a severity score from 1 (minimal severity) to 5 (most severe). A score of 1 indicated a threshold level. A score of 2 indicated slight changes that did not involve a significant amount of tissue. A score of 3 indicated the effect was easily discernable in multiple sites and/or was present in larger foci in from one to a few sites. A score of 4 indicated

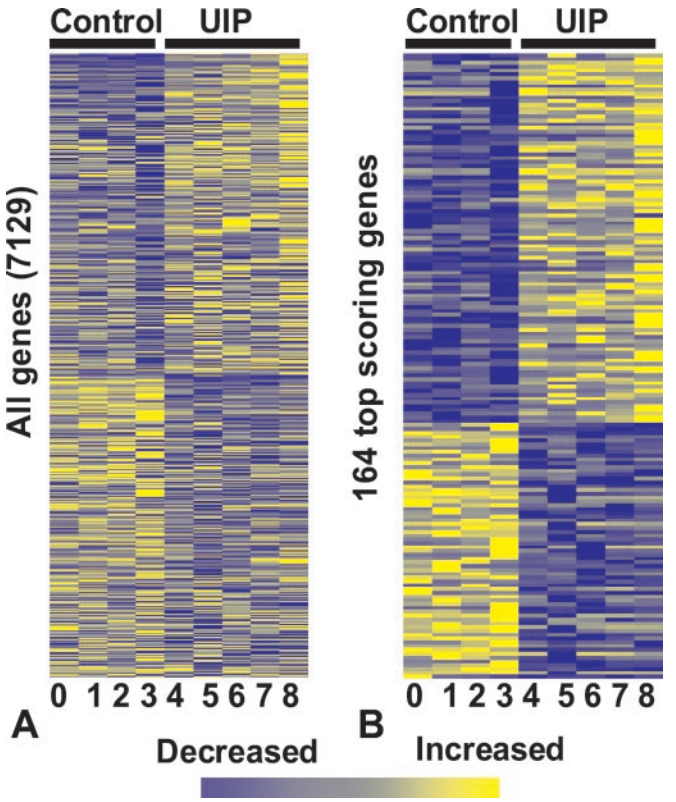

Fig. 1. Gene expression infogram for all 7,129 genes (*A*) and for the 164 most informative genes (TNoM = 0) (*B*). To eliminate outlier effect, genes were normalized to a range of (0,1), meaning that the maximum value for every gene was set to be 1, the minimum value to be 0, and the rest of the values were linearly fitted to this range. Yellow is maximal expression and blue is minimal.

a relatively severe change that was widespread within multiple lobes or that completely filled a single lobe. A score of 5 indicated that a change was severe and involved all lobes.

Results

Gene Expression Patterns Are Distinct in Fibrotic and Nonfibrotic Lungs. Global gene expression patterns clearly distinguished between fibrotic and normal histology lungs (Fig. 1*A*). Furthermore, when only the genes that received the most significant score (TNoM = 0) are presented, the differences in gene expression patterns are even clearer (Fig. 1*B*). This distinction was also evident when we applied hierarchical clustering to the samples (data not shown).

A known issue in analysis of microarray data are the small numbers of samples compared with the large number of parameters measured—a situation that can lead to spurious statistical associations. To address this issue we subjected TNoM to rigorous statistical benchmarking. We calculated the number (of 7,129) of genes expected for every score (TNoM 0, 1, 2, etc.), when a classification of the tissues is uniformly drawn at random. In this case we expected 111 genes to have the best score (TNoM = 0). We actually observed 164 such highly informative genes. Such an overabundance of highly informative genes has a P value $< 10^{-6}$ when genes are independently drawn and their TNoM is calculated with respect to a random partition. In performing the calculations we treated the pool of five normal lungs as a single sample, thus the figures presented here may even underestimate the information content of this study. The complete data set is provided on our web site (http://FGUSheba.cs.huji.ac.il/).

Smooth Muscle Markers Are Increased in Fibrotic Lungs. One of the most intriguing phenomena that occurs during the development of UIP is the formation of small aggregates of actively proliferating myofibroblasts and fibroblasts called myofibroblast/fibroblast foci (5). In addition, abnormal collections of smooth muscle cells have been identified in UIP lungs (16). Interestingly, we observed a marked increase in the expression of genes that encode for muscle proteins. These included markers of smooth muscle differentiation such as vascular α smooth muscle actin, γ smooth muscle actin, calponin, and integrin $\alpha 7\beta 1$. Genes that encode for proteins associated with cell contraction and actin filament organization, such as myosin, SM22, tropomyosin, and S100 A9 protein are also increased. The complete list is provided in Table 1.

Genes that Encode Immunoglobulins, Complement, and Some Chemokines Were Up-Regulated in Fibrotic Lungs. Surprisingly, gene expression patterns in patients with pulmonary fibrosis did not suggest increased acute inflammatory activity. We did not observe any activation of an IL-1 or tumor necrosis factor α-related pathway, nor did we observe any increase in apoptosis-related genes. However, we did observe an increase in genes that encode for immunoglobulins, potentially reflecting chronic B cell infiltration, and an increase in some complement factors. The genes that encode the chemokines MCP 4, GRO 1, and eotaxin were increased (Table 1).

Genes that Encode Proteins that Are Involved in Extracellular Matrix Formation, Degradation, and Signaling Are Increased in Fibrotic Lungs. As expected, expression of the genes that encode for collagens I and III were increased in fibrotic lungs, as was expression of the genes that encode for collagen VI, tenascin C, osteopontin, and fibronectin (Table 1). Several other genes that encode for proteins involved in extracellular matrix formation such as BIGH3, filamin, and fibrillin were increased, as were genes associated with extracellular matrix-related signaling such as cell adhesion kinase β and the collagen receptor tyrosine kinase DDR (Table 1). Another interesting subset was of genes that encoded proteases. The expression levels of MMP-1, MMP-2, MMP-7, MMP-9 (Table 1) and cathepsin E all were relatively increased in fibrotic lungs. MMP-7 (matrilysin) was the gene that scored as the most informative (TNoM = 0, Info = 0, Gaussian = 0.00625, complete data set, http://FGUSheba.cs.huji.ac.il/) among the genes that were up-regulated (Fig. 2*A*).

Increased Expression of MMP-7 Is Confirmed by Immunohistochemistry. To verify the gene expression data at the level of protein expression and to identify the cellular source of MMP-7, we performed immunohistochemiistry on lung tissue from patients with UIP. Immunoreactive MMP-7 was abundantly localized in the epithelium including in alveolar and bronchiolar epithelial cells (Fig. 2 *B* and *C*, and Fig. 4 *A–C*, which is published as supporting information on the PNAS web site, www.pnas.org). MMP-7 was also detected in the extracellular space as shown by staining in thickened alveolar septa and the extracellular matrix (Fig. 2*B*). MMP-7 was minimally detected in normal lung parenchyma (Fig. 2*D*). We also stained for MMP-1 and MMP-9. MMP-1 was primarily present on hyperplastic type 2 pneumocytes and bronchiolar cells lining honeycomb cysts, as well as in reactive alveolar epithelial cells (Fig. 5 *A* and *B*, which is published as supporting information on the PNAS web site). Alveolar macrophages also intensely stained for MMP-1 in tissue sections from patients with fibrotic lung disease (Figs. 1*A* and 5*A*). MMP-9 was detected in neutrophils, both inside vessels and in the interstitium (Fig. 6*A*, which is published as supporting information on the PNAS web site). MMP-9 was also localized in the extracellular matrix and in some areas of dense scars (Fig. 6*B*). Control lungs showed scattered staining for MMP-1 (Fig.

Table 1. Functional groups of selected genes substantially up-regulated in fibrotic lungs (fold induction)

Muscle markers
- [D17408] calponin (9.9)
- [S81419] dystrophin (2.9)
- [D00654] enteric smooth muscle γ-actin (4.2)
- [M12125] fibroblast tropomyosin muscle-type (26)
- [X74295] Integrin $\alpha7\beta1$ (1.7)
- [U48959] myosin light chain kinase MLCK (6.7)
- [U14391] myosin-IC (4.2)
- [AF001548] myosin heavy chain (2.9)
- [M63603] PLN (2.5)
- [L10678] profilin II (3.2)
- [M95787] SM22 (2.3)
- [M26311] S100 calcium-binding protein A9 (2.7)
- [Z24727] tropomyosin isoform (2.1)
- [X13839] vascular smooth muscle α-actin (2.6)

Extracellular matrix, growth factors, and proteases
- [M22489] BMP2a (3.9)
- [M77349] BIGH3 (2.2)
- [X65784] CAR (1.7)
- [J05036] cathepsin E (9.8)
- [U43522] cell adhesion kinase β (4.5)
- [M55998] collagen I-α1 (2.4)
- [Z74616] collagen I-α2 (5.2)
- [X06700] collagen III-α1 (4.7)
- [X52022] collagen VI-α3 (2.5)
- [U48705] DDR (11.3)
- [L13923] fibrillin (2.0)
- [D83920] ficolin-1 (4.5)
- [X53416] filamin (4.3)
- [S37730] IGFBP2 (5.8)
- [L27560] IGFBP5 (2.4)
- [M62403] IGFBP4 (2.0)
- [M21389] keratin 5 (20.3)
- [X74929] keratin 8 (2.9)
- [X07696] keratin 15 (14.1)

Cytokines, chemokines, and antioxidants
- [U25182] antioxidant enzyme AOE37 (25.6)
- [U72511] BAP (2.3)
- [L33930] CD24 (3.4)
- [D49372] eotaxin (4.0)
- [X54489] GRO1 (2.9)
- [X65727] GSTα (7.0)
- [L76191] IL-1 r-associated kinase (6.1)
- [U00672] IL-10 r (2.8)
- [S74221] IK (4.5)
- [U46767] MCP-4 (8.6)
- [M37766] MEM-102 glycoprotein (6.8)
- [L19686] MIF (8.9)
- [M94250] retinoic acid-inducible factor MK (16.5)
- [X65965] SOD-2 (4.6)
- [D50663] TCTEL1 (3.6)
- [X16662] vascular anticoagulant-β (2.7)

Complement, immunoglobulins, and amyloid
- [U50939] amyloid protein-binding protein 1 (3.2)
- [M84526] adipsin/complement factor D (2.2)
- [L15702] complement factor B (2.7)
- [K02765] complement C3, α and β (2.0)
- [M65292] factor H homologue (2.7)
- [U28488] G protein-coupled r AZ3B (2.5)
- [M63438] Ig rearranged γ chain, V-J-C region (4.8)
- [S71043] Ig α2 (4.3)
- [M12759] Ig J chain (3.6)
- [L02326] λ-17 (4.7)
- [M34516] o light chain protein (2.4)
- [U22178] microseminoprotein b (8.4)
- [X57809] rearranged Ig λ light chain (4.2)
- [X51441] SAA protein (10.9)
- [M63379] TRPM-2 (1.8)
- [V00563] V00563 (3.9)

Numbers in parentheses are average fold changes.

5*C*) and minimal MMP-9 staining in neutrophils (Fig. 6*D*). Control samples incubated with nonimmune sera were negative as shown in Fig. 4*D*.

Matrylisin Knockout Mice Do Not Develop Lung Fibrosis After Treatment with Bleomycin. To determine whether MMP-7 was actually involved in the fibrotic process, we administered bleomycin to lungs of *MMP-7*$^{-/-}$ mice, and to age- and sex-matched wild-type controls from two different backgrounds: 129Sv and C57BL/6. Two to 3 weeks after intratracheal bleomycin, hydroxyproline levels increased by 96–119% in wild-type 129Sv mice compared with 55–59% in matrilysin knockout mice, in three separate experiments using 5–9 mice/group (Fig. 3*A*). The differences were statistically significant with *P* values from 0.0002 to 0.02 in all three experiments. This finding did not depend on the genetic background of the mice, because comparable results were obtained in three separate experiments with mice of the C57BL/6 background (Fig. 3*A*). These results were supported by histologic evaluation that revealed milder changes in the lungs of matrilysin knockout mice (Fig. 3*D*). The mean severity for all histopathologic parameters assessed was significantly less in the bleomycin-challenged *MMP-7*$^{-/-}$ mice than in the corresponding wild-type controls (Fig. 3*B*).

Discussion

The results of this study support the view of UIP as a disease of persistent matrix deposition and destruction that is associated with relatively modest, chronic inflammation. Interestingly, despite the prevailing view of "honeycomb" lung as an irreversibly scarred and quiescent tissue, our gene expression data show a tissue that is quite actively remodeling with a high level of gene expression of both matrix degrading proteins and proteins involved in matrix deposition.

One of the most impressive features of our results is the clear distinction in gene expression patterns, and therefore the fundamental tissue biology, of UIP lungs compared with nonfibrotic lungs. A usual concern with microarray analysis of diseases, in which tissue availability is a challenge, is the difficulty in obtaining data that are both biologically meaningful and statistically significant. The use of traditional statistics is complicated in microarray experiments because of the troubling asymmetry between the numbers of samples and the number of parameters measured, which may result in spurious statistical associations. One of the advantages of the analytic methods that we used here is their amenability to rigorous statistical benchmarking (11). We can calculate the number of informative genes per score expected in a random classification and then compare this estimated number of high scoring (or informative) genes to the actual number of informative genes (per score) measured in our data set (13). In our study, despite the small sample size and the genetic heterogeneity of our patients we observed overabundance of informative genes, suggesting that the differences in gene expression that we observed between the two groups were biologically meaningful. Indeed, gene expression patterns were highly robust and also evident when we applied other statistical tools (*t* test, Wilcoxon), clustering methods, and self-organizing

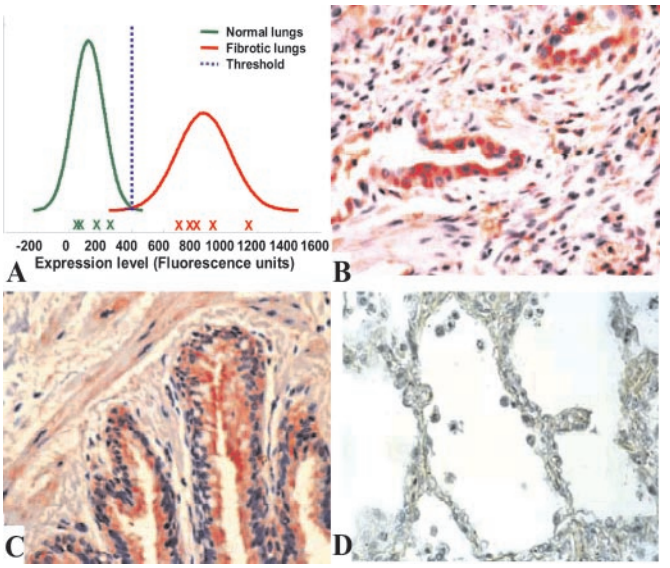

Fig. 2. MMP-7 gene and protein expression levels are significantly increased in fibrotic lungs. The Gaussian distribution for MMP-7 reveals minimal overlap (score of 0.00625) between normal lungs (green line, *A*) and fibrotic lungs (red line, *A*). All actual expression levels of MMP-7 in fibrotic lungs (red X marks, *A*) were higher than in normal lungs (green X marks, *A*) and no samples were misclassified (TNoM and Info = 0). The *y* axis in the Gaussian figure (*A*) is the estimated density of the expression level. This density is computed based on the normality assumption of expression level (for each class). MMP-7 staining verified the gene expression predictions and demonstrated both cell-associated and matrix MMP-7 in distal airspaces (*B*) and proximal airway (*C*) from fibrotic lungs. No MMP-7 immunoreactivity was seen in normal lungs (*D*). (Magnification: *B–D*, ×40).

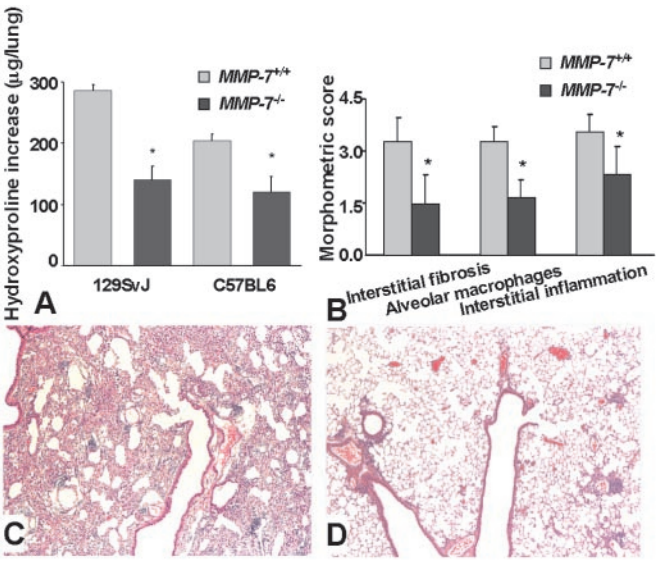

Fig. 3. *MMP7*$^{-/-}$ mice are protected against bleomycin-induced pulmonary fibrosis. Increase in lung hydroxyproline content in lungs of wild-type (*MMP-7*$^{+/+}$) mice 14 days after bleomycin injection is significantly higher than in *MMP-7*$^{-/-}$ mice on both C57BL6 and 129SvJ background. Increase values were calculated by subtracting mean hydroxyproline content (μg hydroxyproline/lung) of lungs of saline-injected animals from the hydroxyproline content of lungs 14 days after bleomycin (*A*). Comparison of semiquantitative morphometric analysis of lung histology 14 days after bleomycin (*B*) and representative histology of *MMP-7*$^{+/+}$ (*C*) and *MMP-7*$^{-/-}$ mice (*D*). Data (means ± SE) are representative of at least two comparable experiments with six mice per group; *, $P < 0.01$ relative to bleomycin-treated *MMP*$^{+/+}$ mice. (Magnification: *B* and *C*, ×10).

maps (data not shown). The results of our analysis are an example of overcoming some of the limitations of a small sample size by using computational tools specifically designed for microarray experiments.

Close inspection of the informative genes in our data set sheds light on several important aspects of fibrotic lung disease. One impressive feature is the coordinated up-regulation of MMPs, predominantly MMP-1 and MMP-7, in the lungs of patients with UIP. It has been suggested that the tissue phenotype of fibrosis (primarily the exaggerated extracellular matrix deposition) results from dysregulation of the synthesis and degradation of extracellular matrix proteins—a process that could involve several members of the MMP family. Thus, MMP-1 (collagenase-1 that degrades fibrillar collagens) as well as MMP-2 and MMP-9 (gelatinases A and B, which have a broad range of substrates including type IV collagens from basement membranes) have previously been reported to be up-regulated in human pulmonary fibrosis and animal models of pulmonary fibrosis (17–19). In our data set, MMP-1, MMP-2, and MMP-9 were significantly higher in UIP lungs. These observations were verified at the protein expression level for MMP-1 and MMP-9 (Figs. 5 and 6). We did not observe any significant increase in gene expression levels of tissue inhibitors of metalloproteinases in UIP lungs. Surprisingly we detected a decrease in TIMP3 gene expression (complete data set table, http://FGUSheba.cs.huji.ac.il/), an observation that has been seen in other microarray experiments (A.P., unpublished observation).

MMP-7, a metalloproteinase not previously associated with pulmonary fibrosis, was the most informative gene among the increased genes in UIP lungs. This observation was confirmed by immunohistochemical staining on lung tissue from a separate group of patients with UIP. The intense staining confirmed increased MMP-7 protein expression by epithelial cells as well as an increase in MMP-7 bound to the extracellular matrix. The protection of *MMP-7*$^{-/-}$ knockouts from bleomycin-induced fibrosis shows that this metalloproteinase plays an important role in the development of pulmonary fibrosis. MMP-7 is expressed in epithelial tissues that are in contact with the environment. Although it is not expressed by mice living in a germ-free environment (20), its expression is rapidly induced by infection and air pollution (21, 22). MMP-7 is required for activation of defensins in the small bowel and *MMP-7*$^{-/-}$ mice cannot kill enteric pathogens (23). Recently, MMP-7 has been shown to play a role in two pathways that are considered important in the development of bleomycin-induced pulmonary fibrosis, the tumor necrosis factor α, and Fas pathways (24, 25). Like MMP-1, MMP-7 was expressed primarily by alveolar epithelial cells, and both may have a role in cell migration as has been proposed for keratinocytes in skin wound healing (26, 27). Similarly, MMP-1 and MMP-7 may participate in alveolar and bronchiolar cell migration over different matrices during UIP lung remodeling (28). In addition, MMP-7 may enhance a procoagulant microenvironment in the alveolar spaces because it is able to rapidly cleave tissue factor pathway inhibitor, an inhibitor of tissue factor-induced coagulation (29). Increased local procoagulant activity has been found in fibrotic lung disorders, including idiopathic pulmonary fibrosis, and alveolar epithelial cells seem to contribute to the increased procoagulant and antifibrinolytic activities in this disorder (5, 30). Based on our data and its known functions, matrilysin is an excellent candidate to be a key player in sustained tissue injury and fibrosis. On one side it is induced by epithelial injury, a common pathological process leading to pulmonary fibrosis, and on the other hand it may be an activator of cell death, chronic inflammation, and procoagulant activity. The focus on MMP-7 in our work is a good example of large-scale gene expression experiments as hypothesis-generating tools.

Based on previous knowledge we had little reason to suspect that MMP-7 played a role in pulmonary fibrosis. However, no matter what type of analysis was used, MMP-7 was the most significant gene that was induced in UIP samples, thus almost demanding its analysis.

Among the genes that were significantly up-regulated, we did not find many inflammatory-related genes. We observed evidence of a chronic inflammatory infiltration probably associated with B cell infiltration (immunoglobulins, etc.), but we did not observe any evidence for T cell activation or any significant activation of apoptotic pathways. Although in animal models of pulmonary fibrosis the fibrotic response may depend on intact tumor necrosis factor α or Fas signaling pathways (31–35), activation of these pathways was not apparent in this study.

One limitation of using whole lung homogenates in our study is our relative inability to assess the exact cells that are overexpressing genes that were higher in UIP lungs. For some typical cellular markers it is very hard to even guess whether the changes result from changes in cellular content or real gene expression patterns. For instance we observed changes in three cytokeratins (5, 8, 15). This finding may represent a change in epithelial gene expression, a change in epithelial cellular subset (epithelial migration), or a phenotypic change in another cell type. So far such analysis is cumbersome, but the creation of databases that will contain gene expression signatures of specific cell types and analytic approaches to dissect them out computationally could ameliorate this hardship. Despite this limitation, one of the impressive clusters of genes that that were significantly increased in UIP lungs was that of muscle-related genes. An important feature in the histological description of UIP is the presence of fibroblastic foci (5). Fibroblast foci possibly represent microscopic areas in which fibroblasts migrate, proliferate, and maintain the abnormal architecture of the lung. A major subpopulation of the fibroblasts in these foci expresses markers of smooth muscle differentiation, such as α-smooth muscle actin (16, 17). In addition, clusters of smooth-muscle cells usually are seen in UIP lungs, probably contributing to the decreased lung compliance (16). We observed an increase in muscle-related genes that ranged from 2.5- to 25-fold (Table 1). These coordinated increases in smooth muscle markers, and in cellular contractile machinery, reflect our ability to detect a significant phenotypic change in even in a relatively small cell population in the lung.

In this study we used oligonucelotide microarrays to analyze gene expression patterns in UIP. We distinguished specific subsets of genes that differentiated UIP from nonfibrotic lungs. These genes included genes that encoded proteins associated with B cells and with chronic inflammation, muscle-related genes, and genes that encoded proteins involved in extracellular matrix production, degradation, and signaling. Using advanced computational scoring methods we determined that MMP-7, a multifunctional MMP not previously associated with pulmonary fibrosis, was the most informative increased gene in our data set. The hypothesis that MMP-7 is indeed involved in the pathogenesis of pulmonary fibrosis was supported by our observation that MMP-7 knockout mice were protected from bleomycin-induced fibrosis. Our results support the view of UIP as a chronic destructive process, in which an increase in myofibroblasts and cells with smooth muscle-like phenotype, active abnormal reepithelialization, and aberrant repair are interwoven.

We thank Gady Cojocaro and Dr. Hal van Wart for their support and Dr. Tom Raffin and Ms. Susan Jacob for providing fresh human lung tissue samples from Stanford Medical Center, Stanford University, Stanford, CA. Dr. Isasshar Ben-Dov provided helpful comments on the manuscript. N.K.'s work was supported in part by a generous grant from the Tel-Aviv chapter of the Israeli Lung Association.

1. Costabel, U. & King, T. E. (2001) *Eur. Respir. J.* **17,** 163–167.
2. Ryu, J. H., Colby, T. V. & Hartman, T. E. (1998) *Mayo Clin. Proc.* **73,** 1085–1101.
3. King, T. E., Jr. (1998) *Ann. Intern. Med.* **129,** 806–812.
4. Coker, R. K. & Laurent, G. J. (1998) *Eur. Respir. J.* **11,** 1218–1221.
5. Selman, M., King, T. E. & Pardo, A. (2001) *Ann. Intern. Med.* **134,** 136–151.
6. Cojocaru, G. S., Rechavi, G. & Kaminski, N. (2001) *Isr. Med. Assoc. J.* **3,** 292–296.
7. van Berkum, N. L. & Holstege, F. C. (2001) *Curr. Opin. Biotechnol.* **12,** 48–52.
8. Kaminski, N., Allard, J. D., Pittet, J. F., Zuo, F., Griffiths, M. J., Morris, D., Huang, X., Sheppard, D. & Heller, R. A. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 1778–1783.
9. Affymetrix (1998) *Eukaryotic Expression Analysis Target Preparation* (Affymetrix, Santa Clara, CA).
10. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14863–14868.
11. Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M. & Yakhini, Z. (2000) *J. Comput. Biol.* **7,** 559–583.
12. Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O. P., *et al.* (2001) *N. Engl. J. Med.* **344,** 539–548.
13. Ben-Dor, A., Friedman, N. & Yakhini, Z. (2000) *Agilent Laboratories Technical Report AGL-2000–13*, http://www.labs.agilent.com/resources/techreports.html.
14. Pardo, A., Barrios, R., Gaxiola, M., Segura-Valdez, L., Carrillo, G., Estrada, A., Mejia, M. & Selman, M. (2000) *Am. J. Respir. Crit. Care Med.* **161,** 1698–1704.
15. Woessner, J. F. (1961) *Arch. Biochem. Biophys.* **93,** 440–447.
16. Ohta, K., Mortenson, R. L., Clark, R. A., Hirose, N. & King, T. E., Jr. (1995) *Am. J. Respir. Crit. Care Med.* **152,** 1659–1665.
17. Selman, M., Ruiz, V., Cabrera, S., Segura, L., Ramirez, R., Barrios, R. & Pardo, A. (2000) *Am. J. Physiol.* **279,** L562–L574.
18. Lemjabbar, H., Gosset, P., Lechapt-Zalcman, E., Franco-Montoya, M. L., Wallaert, B., Harf, A. & Lafuma, C. (1999) *Am. J. Respir. Cell Mol. Biol.* **20,** 903–913.
19. Swiderski, R. E., Dencoff, J. E., Floerchinger, C. S., Shapiro, S. D. & Hunninghake, G. W. (1998) *Am. J. Pathol.* **152,** 821–828.
20. Parks, W. C., Lopez-Boado, Y. S. & Wilson, C. L. (2001) *Chest* **120,** S36–S41.
21. Su, W. Y., Jaskot, R. H., Richards, J., Abramson, S. R., Woessner, J. F., Jr., Yu, W. H. & Dreher, K. L. (2000) *Am. J. Physiol.* **279,** L152–L160.
22. Lopez-Boado, Y. S., Wilson, C. L., Hooper, L. V., Gordon, J. I., Hultgren, S. J. & Parks, W. C. (2000) *J. Cell Biol.* **148,** 1305–1315.
23. Wilson, C. L., Ouellette, A. J., Satchell, D. P., Ayabe, T., Lopez-Boado, Y. S., Stratman, J. L., Hultgren, S. J., Matrisian, L. M. & Parks, W. C. (1999) *Science* **286,** 113–117.
24. Mitsiades, N., Yu, W. H., Poulaki, V., Tsokos, M. & Stamenkovic, I. (2001) *Cancer Res.* **61,** 577–581.
25. Haro, H., Crawford, H. C., Fingleton, B., Shinomiya, K., Spengler, D. M. & Matrisian, L. M. (2000) *J. Clin. Invest.* **105,** 143–150.
26. Dunsmore, S. E., Saarialho-Kere, U. K., Roby, J. D., Wilson, C. L., Matrisian, L. M., Welgus, H. G. & Parks, W. C. (1998) *J. Clin. Invest.* **102,** 1321–1331.
27. Planus, E., Galiacy, S., Matthay, M., Laurent, V., Gavrilovic, J., Murphy, G., Clerici, C., Isabey, D., Lafuma, C. & d'Ortho, M. P. (1999) *J. Cell Sci.* **112,** 243–252.
28. Parks, W. C., Sudbeck, B. D., Doyle, G. R. & Saariahlo-Kere, U. K. (1998) in *Matrix Metalloproteinases*, eds. Parks, W. C. & Mecham, R. P. (Academic, San Diego), pp. 263–297.
29. Belaaouaj, A. A., Li, A., Wun, T. C., Welgus, H. G. & Shapiro, S. D. (2000) *J. Biol. Chem.* **275,** 27123–27128.
30. Loskutoff, D. J. & Quigley, J. P. (2000) *J. Clin. Invest.* **106,** 1441–1443.
31. Ortiz, L. A., Lasky, J., Hamilton, R. F., Jr., Holian, A., Hoyle, G. W., Banks, W., Peschon, J. J., Brody, A. R., Lungarella, G. & Friedman, M. (1998) *Exp. Lung Res.* **24,** 721–743.
32. Zhang, K., Gharaee-Kermani, M., McGarry, B., Remick, D. & Phan, S. (1997) *J. Immunol.* **158,** 954–959.
33. Aoshiba, K., Yasui, S., Tamaoki, J. & Nagai, A. (2000) *Am. J. Respir. Crit. Care Med.* **162,** 695–700.
34. Kuwano, K., Hagimoto, N., Tanaka, T., Kawasaki, M., Kunitake, R., Miyazaki, H., Kaneko, Y., Matsuba, T., Maeyama, T. & Hara, N. (2000) *J. Pathol.* **190,** 221–229.
35. Kuwano, K., Hagimoto, N., Kawasaki, M., Yatomi, T., Nakamura, N., Nagata, S., Suda, T., Kunitake, R., Maeyama, T., Miyazaki, H. & Hara, N. (1999) *J. Clin. Invest.* **104,** 13–19.