

Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease

Wen Ning^{*†}, Chao-Jun Li^{*†‡}, Naftali Kaminski^{*}, Carol A. Feghali-Bostwick^{*}, Sean M. Alber[§], Yuanpu P. Di[¶], Sherrie L. Otterbein^{*}, Ruiping Song^{*}, Shizu Hayashi^{||}, Zhihong Zhou^{*}, David J. Pinsky^{**}, Simon C. Watkins[§], Joseph M. Pilewski^{*}, Frank C. Sciurba^{*}, David G. Peters^{*}, James C. Hogg^{||}, and Augustine M. K. Choi^{*††}

^{*}Division of Pulmonary Allergy and Critical Care Medicine, Department of Medicine, and [§]Department of Cell Biology and Physiology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213; [¶]Department of Environmental and Occupational Health, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA 15213; [‡]Jiangsu Key Laboratory for Molecular Medical Biotechnology, Nanjing Normal University, Nanjing 210097, China; ^{||}Department of Pathology and Laboratory Medicine, St. Paul's Hospital-Providence Health Care, University of British Columbia, Vancouver, BC, Canada V6Z 1Y6; and ^{**}Department of Medicine, College of Physicians and Surgeons of Columbia University, New York, NY 10032

Edited by Bert Vogelstein, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, and approved August 30, 2004 (received for review February 19, 2004)

To better understand the molecular basis of chronic obstructive pulmonary disease (COPD), we used serial analysis of gene expression (SAGE) and microarray analysis to compare the gene expression patterns of lung tissues from COPD and control smokers. A total of 59,343 tags corresponding to 26,502 transcripts were sequenced in SAGE analyses. A total of 327 genes were differentially expressed (1.5-fold up- or down-regulated). Microarray analysis using the same RNA source detected 261 transcripts that were differentially expressed to a significant degree between GOLD-2 and GOLD-0 smokers. We confirmed the altered expression of a select number of genes by using real-time quantitative RT-PCR. These genes encode for transcription factors (EGR1 and FOS), growth factors or related proteins (CTGF, CYR61, CX3CL1, TGF β 1, and PDGFRA), and extracellular matrix protein (COL1A1). Immunofluorescence studies on the same lung specimens localized the expression of Egr-1, CTGF, and Cyr61 to alveolar epithelial cells, airway epithelial cells, and stromal and inflammatory cells of GOLD-2 smokers. Cigarette smoke extract induced Egr-1 protein expression and increased Egr-1 DNA-binding activity in human lung fibroblast cells. Cytomix (tumor necrosis factor α , IL-1 β , and IFN- γ) treatment showed that the activity of matrix metalloproteinase-2 (MMP-2) was increased in lung fibroblasts from EGR1 control (+/+) mice but not detected in that of EGR1 null (-/-) mice, whereas MMP-9 was regulated by EGR1 in a reverse manner. Our study represents the first comprehensive analysis of gene expression on GOLD-2 versus GOLD-0 smokers and reveals previously unreported candidate genes that may serve as potential molecular targets in COPD.

Chronic obstructive pulmonary disease (COPD) is a slowly progressive and irreversible disorder characterized by the functional abnormality of airway obstruction, which is a significant cause of morbidity, mortality, and health care costs. COPD is a collective term describing two separate chronic lung diseases: emphysema and chronic bronchitis, which are caused largely by a common agent, cigarettes (1, 2). Cigarette smoke has been generally accepted as the most important of many risk factors for the development of COPD, which accounts for \approx 80–90% of COPD cases in the United States (3). However, only 15–20% of heavy smokers develop clinically significant airflow obstruction, which suggests a genetic susceptibility to the development of the disease (4). The genes that determine this genetic susceptibility to cigarette smoking and disease progression to COPD are poorly understood.

To better understand the candidate genes involved in the development of COPD in smokers, we performed serial analysis of gene expression (SAGE) and microarray analysis as a complementary approach to analyze the global gene expression profiles of lung tissue from smokers who are at risk (GOLD-0) and who have developed moderate (GOLD-2) COPD (5). SAGE (6), based on

the 10-bp tag for sufficient individual gene identification and the concatenation of SAGE tags for high-efficient gene identification, has the additional advantage of allowing unbiased and comprehensive analysis of a large number of differentially expressed genes without prior knowledge of the genes when applied to any particular cell system in different conditions (7, 8). This tag-based approach has the potential to identify new genes expressed at lower levels (9).

In this study, we compared the tags present in the GOLD-2 smoker samples with GOLD-0 smoker controls and generated a comprehensive profile of gene expression patterns in these lung tissues. We selected genes that were consistently differentially expressed in GOLD-2 smokers versus GOLD-0 smoker controls by both SAGE and microarray analysis, with confirmation of expression in the lung tissue of GOLD-2 smokers and fibroblasts from emphysema patients. We demonstrate that gene expression profiling analysis represents a powerful approach to provide insights to novel pathways involved in the pathogenesis of COPD.

Materials and Methods

Human Lung Tissue Acquisition. Lung tissue samples from two groups of smokers were obtained from surgical specimens. The 14 smokers with obstruction ranged from 48 to 76 years of age and had pulmonary function test results consistent with moderate COPD (GOLD-2 classification). Twelve control smokers (45–76 years of age) exhibited nonobstruction (GOLD-0 classification) [forced expiratory volume in 1 sec (FEV₁), >90% predicted] (Table 1). Three severe COPD (GOLD-4 classification; lung transplant explants) and three normal lung control samples were also obtained from the University of Pittsburgh Tissue Bank.

SAGE Libraries and SAGE Analysis. Within the two groups, that is, GOLD-2 (G₂1 to G₂6) and GOLD-0 (G₀1 to G₀5), equal quantities of total RNA from each individual were pooled. The same RNA source was used for SAGE, microarray, and real-time quantitative RT-PCR analysis. We used 20 μ g of total RNA to construct each SAGE library, as recommended in the MICROSAGE DETAILED

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: COPD, chronic obstructive pulmonary disease; SAGE, serial analysis of gene expression; QRT-PCR, real-time quantitative RT-PCR; CSE, cigarette smoke extract; MMP, matrix metalloproteinase; FEV₁, forced expiratory volume in 1 sec; TNF- α , tumor necrosis factor α ; GO, gene ontology.

[†]W.N. and C.-J.L. contributed equally to this work.

^{††}To whom correspondence should be addressed at: University of Pittsburgh, Division of Pulmonary Allergy and Critical Care Medicine, Montefiore University Hospital 628 NW, 3459 Fifth Avenue, Pittsburgh, PA 15213. E-mail: choiam@upmc.edu.

© 2004 by The National Academy of Sciences of the USA

Table 1. Clinical information of human lung specimens

	Nos.	Age	Pkyr	FEV ₁	FVC	FEV ₁ /FVC
GOLD-2	G ₂ 1	71	Smoker*	46	69	51
	G ₂ 2	71	Smoker*	69	74	73
	G ₂ 3	59	45	65	84	62
	G ₂ 4	62	52	64	96	52
	G ₂ 5	48	9	68	101	54
	G ₂ 6	76	Smoker*	53	58	71
	G ₂ 7	59	33	62	82	59
	G ₂ 8	64	36	73	88	65
	G ₂ 9	63	68	71	89	63
	G ₂ 10	62	52	64	96	52
	G ₂ 11	76	9	60	77	60
	G ₂ 12	55	30	66	84	54
	G ₂ 13	69	51	69	99	54
	G ₂ 14	50	56	63	79	62
GOLD-0	G ₀ 1	69	49	91	87	82
	G ₀ 2	76	40	95	94	77
	G ₀ 3	50	20	120	112	87
	G ₀ 4	45	30	112	109	83
	G ₀ 5	58	17	114	112	84
	G ₀ 6	66	17	87	83	81
	G ₀ 7	72	37	105	106	75
	G ₀ 8	56	50	89	93	75
	G ₀ 9	62	42	93	93	78
	G ₀ 10	57	44	97	109	70
	G ₀ 11	74	3	123	122	77
	G ₀ 12	73	27	96	95	77

Pkyr, packs per year; FVC, forced vital capacity.

*Packs per year unknown.

PROTOCOL (version 1.0e) with some minor modifications as described (10). In brief, an anchoring enzyme (*Nla*III) and a tagging enzyme (*Bsm*F) were used to release SAGE tags. The self-ligated concatemers were cloned into vector pZero (Invitrogen) for high-throughput sequencing. The transcript identity of each SAGE tag was obtained by matching the unitag list against the human tag-to-gene “reliable” map (<ftp://ncbi.nlm.nih.gov/pub/sage/map>). Each specific transcript abundance was then determined by its unique tag count (10).

Microarray and Data Analysis. The microarray experiments using pooled samples were performed on UniGEM-V high-density cDNA microarrays (Incyte Pharmaceuticals, Fremont, CA). We individually hybridized 5 μ g of total RNA from each group to the arrays according to protocols (11). Data analysis was performed with GEMTOOLS. Incyte’s data analysis tools recommend genes whose absolute fold changes (balanced Cy3/Cy5 or Cy5/Cy3 signal intensity ratio) ≥ 1.6 , a real expression difference between the GOLD-2 and GOLD-0 smokers at the 95% confidence level (11). For studies involving microarray analysis of independent individual samples (G₂7–G₂12 and G₀6–G₀11), we used Codelink cDNA microarrays according to the manufacturer’s protocol (Amersham Pharmacia Biosciences). Arrays were scanned by using GENEPIX 4000B (Axon Instruments, Union City, CA), and data were generated by using CODELINK EXPRESSION ANALYSIS (version 4.0). All arrays were median-normalized. Gene fold ratios were obtained by dividing the individual gene values by the geometric mean of all control values for this gene. Statistical analysis was performed by using the SCOREGENE software suite (www.cs.huji.ac.il/labs/compbio/scoregenes) (12). Functional analysis of microarray data was performed by using GENEXPRESS (<http://genexpress.stanford.edu>). The significant enrichment of gene ontology (GO) annotations within genes that significantly changed in COPD lungs [*t* test and threshold number of misclassification (TNOM) *P* values of <0.05] was determined by using hypergeometric models. All *P* values were corrected to a false discovery rate of 5% (13).

Real-Time Quantitative RT-PCR (QRT-PCR). QRT-PCR was carried out as described (10, 14). PCR primers and probes were designed with PRIMER EXPRESS 1.5A software (Applied Biosystems), and their sequences are given in Table 5, which is published as supporting information on the PNAS web site.

Primary Lung Fibroblast Isolation and Culture. Lung fibroblasts, hereafter referred to as emphysematous lung fibroblasts, were cultured from the explanted lung tissue of patients with severe emphysema (GOLD-4 FEV₁ $< 30\%$ predicted) who underwent lung transplantation. Human normal lung fibroblasts were from lungs of non-COPD donors that were not used for transplant surgery at the University of Pittsburgh Medical Center. Mice lung fibroblasts were cultured from lung tissue of EGR1 null (–/–) or control littermates. Fibroblast cell culture was carried out as described (15). Human fetal lung fibroblasts (MRC-5) were obtained from the American Type Culture Collection.

Immunofluorescence. Immunofluorescence was performed by using standard methods (16). In brief, lung frozen-tissue sections (5 μ m) from same-patient samples (*n* = 4 for each group) used for SAGE analysis were incubated with respective use of rabbit anti-Egr-1, Cyr61, or goat anti-CTGF antibody (Santa Cruz Biotechnology), then anti-rabbit Cy3 (Jackson ImmunoResearch) or anti-goat Alexa 488 (Molecular Probes). The first antibodies were replaced by rabbit or goat IgG and then exposed to secondary antibodies as blank control. Images were taken with an LSM 510 Meta laser-scanning confocal microscope (Zeiss). Quantitative measurement was made with METAMORPH software. Ten random fields at $\times 200$ magnification were selected from each sample (*n* = 4 for each group). The images were taken under the same exposure setting.

Preparation and Treatment of Cigarette Smoke Extract (CSE). Non-filtered research reference cigarettes 2R1 were purchased from the University of Kentucky (Lexington). CSE was prepared freshly at a concentration of one cigarette per 5 ml in serum-free DMEM with a modification from ref. 17. This medium was defined as 100% CSE and can be used after adjusting to pH 7.4 and filtering through a 0.22- μ m filter.

Egr-1 Protein Expression and DNA-Binding Activity Assay. Total protein of fibroblast was prepared and Western blotting was performed by using rabbit anti-Egr-1 antibody as described (18). Nuclear protein was prepared and Egr1 DNA-binding activity was measured by a BD Mercury TransFactor kit according to the manufacturer’s instructions (BD Biosciences, Palo Alto, CA).

Gelatin Zymography. Gelatin zymography for assaying matrix metalloproteinase (MMP) was carried out as described (19) with minor modifications. In brief, mouse lung fibroblasts were treated with cytomix [tumor necrosis factor α (TNF- α), 10 ng/ml; IL-1 β , 5 ng/ml; IFN- γ , 10 ng/ml] for 1–4 days. Conditioned media were concentrated by using the Centricon-10 system (Amicon-Millipore). We electrophoresed 10 μ g of total protein on 10% precast polyacrylamide gel containing 2 mg/ml gelatin (Invitrogen). The gel was incubated in 1 \times Zymogram developing buffer for 4 h at 37°C for maximum protease activity.

Statistical Analysis. Data are expressed as the mean \pm SEM. Differences in measured variables between experimental and control group were assessed by using Student’s *t* tests. Results were considered statistically significant at *P* < 0.05 .

Results

SAGE Expression Profiles of GOLD-2 and GOLD-0 Smokers. To characterize genes that are differentially expressed in the lung tissues of GOLD-2 and GOLD-0 smokers, two SAGE libraries were constructed. See Tables 6 and 7, which are published as supporting

Table 2. Differentially expressed genes between GOLD-2 and GOLD-0 smokers

Gene	Fold change*		
	SAGE	Microarray	QRT-PCR
EGR1	1.8	2	3.0 ± 0.84
FOS	2.7	1.8	3.6 ± 0.73
CTGF	1.6	1.5	1.7 ± 0.38
CYR61	2.8	1.7	1.9 ± 0.60
CX3CL1	8	1.6	2.2 ± 0.44
TGFB1	1.9	–	1.2 ± 0.07
PDGFRA	–5	–	–1.5 ± 0.10
COL1A1	–5.5	–2.6	–4.3 ± 0.58

*Positive fold change is the ratio of transcript level of GOLD-2 to GOLD-0; a negative fold change is the ratio of GOLD-0 to GOLD-2.

information on the PNAS web site, for the entire SAGE database. A total of 59,343 tags were sequenced, nearly equal for GOLD-0 (29,273) and GOLD-2 (30,070), which represented 20,191 unitags (unique sequence tags). We classified unitags into four abundance classes: “high (69 unitags)” (total tag counts, >80); “moderate (240 unitags)” (total tag counts, ≤80 and >20); “low (1,392 unitags)” (total tag counts, ≤20 and ≥5); and “rare (18,478)” (total tag counts, <5). The 8% of unitags that were recorded five to several hundred times comprised 58% of the mRNA population based on tag-count analysis. The remaining 18,478 unitags (92%) were detected less than five times, but in aggregate, this rare-abundance class represented 42% of the mRNA population. This distribution fits with the overall pattern of gene expression in mammalian cells, in which only a small percentage of mRNA species reach high-copy number, and most mRNA display faint levels (20).

By matching unitag sequences to NCBI human tag-to-gene “reliable” SAGEmap, we identified a total of 26,502 transcripts. 9,731 tags (37%) and 7,613 tags (32%) matched to characterized GenBank entries and expressed-sequence tag/other uncharacterized Unigene entries, respectively. 2,741 redundant tags (10%) were derived from mitochondrial genes/ribosome RNA, matched multiple transcripts, or presented within so-called Alu repeats (21). A total of 6,417 tags (27%) had no known match to publicly available databases. These tags may represent entirely novel, previously unidentified transcripts. We focused our analysis on the tags that matched known genes to identify genes that are differentially expressed between GOLD-2 and GOLD-0 smokers.

Differential Gene Expression Between GOLD-2 and GOLD-0 Smokers by SAGE. Comparison of gene expression patterns between GOLD-2 and GOLD-0 smokers revealed that most transcripts were ex-

pressed at similar levels. However, a total of 327 genes were determined to be differentially expressed by using the tag ratio (greater than or equal to ±1.5) and tag counts as the criteria to compare the two libraries. Of these genes, 97 were underexpressed or overexpressed by ≥5-fold in GOLD-2 smokers compared with GOLD-0 smokers, 85 moderately abundant genes (tag counts ≥20) were underexpressed or overexpressed 1.5- to 5-fold, and 145 low-abundant transcripts (tag counts >5 and <20) were underexpressed or overexpressed 2- to 5-fold in GOLD-2 than that in GOLD-0 smokers. Given the challenges associated with classifying many genes by function in the context of a single experiment, we used online tools that are designed to assist investigators in this task. These include ONTOEX and FATIGO (22, 23), which are united by their use of the GO Database provided by the GO consortium (www.geneontology.org). A broad functional classification by biological process of the genes, in which expression was altered in GOLD-2 compared with GOLD-0, is given in Table 8 (which is published as supporting information on the PNAS web site). Specific terms for these processes were extracted from multiple levels within the GO hierarchy so that only major gene categories are listed. This broad classification includes genes encoding molecules for signal transduction, receptor function, growth factor, nuclear chromatin and DNA binding, adhesion and cytoskeleton, metabolism, matrix, cell cycle, and oxidative stress.

As expected, transcripts encoding proteins associated with inflammation were overexpressed in GOLD-2 smokers. These included TGF-β1, CX3CL1, CTGF, CYR61, TNFSF10, and receptor such as IL1R. SAGE also indicated that mRNAs that encode several proteins involved in angiogenesis were among differentially expressed transcripts in GOLD-2 versus GOLD-0 smokers. Several transcripts that inhibit cell proliferation were highly expressed in the lung of smokers with COPD, such as CDKN1A and CDC2L1. We also observed significant elevation of expression of a number of apoptosis-related genes in patients with COPD relative to controls, such as TEGT, TXNL, GRIM19, NCKAP1, and BCAP31, which have not previously been reported to be associated with smoking or COPD. In contrast, we also detected more transcripts derived from the ECM genes COL1A1, COL3A1, COL4A1, COL6A1, and COL18A1 in the lungs of GOLD-0 smokers. Transcripts encoding transcription factors, including FOS, EGR1, KLF2, HEYL, HAX1, and ILF3, were highly expressed in samples from GOLD-2 smokers.

Differential Gene Expression Between GOLD-2 and GOLD-0 Smokers Revealed by Microarray Analysis. In addition to SAGE, we used microarray analysis as a complementary approach to survey the same pooled RNA samples we used for SAGE for relative gene expression patterns. Microarray analysis detected the expression of

Table 3. Functional analysis of genes that significantly distinguish COPD lungs

Attribute	Cluster ID	Cluster name	P value	%*
Up in COPD	241	Growth factor activity	6.41 × 10 ⁻³	26.7
Up in COPD	300	Signal transductor activity	3.65 × 10 ⁻³	9.8
Up in COPD	402	Development	9.70 × 10 ⁻³	10.5
Up in COPD	1099	Cell-surface receptor-linked signal transduction	4.55 × 10 ⁻³	15.7
Up in COPD	1116	Receptor binding	4.22 × 10 ⁻³	17.5
Up in COPD	1203	Hormone activity	4.29 × 10 ⁻³	42.9
Up in COPD	1207	Secretion	8.22 × 10 ⁻³	15.6
Down in COPD	100	Mitochondrion	5.49 × 10 ⁻³	30.1
Down in COPD	202	NADH dehydrogenase (ubiquinone) activity	3.88 × 10 ⁻³	50
Down in COPD	231	Sodium ion transporter activity	9.97 × 10 ⁻³	37.9
Down in COPD	912	26S proteasome	3.95 × 10 ⁻³	60
Down in COPD	1117	NADH dehydrogenase activity	3.88 × 10 ⁻³	50
Change in COPD	1238	Lipoprotein biosynthesis	3.48 × 10 ⁻³	83.3

Codelink microarray analysis was performed on independent individual lung tissue samples from GOLD-2 and GOLD-0 smokers.

*Percentage of the genes in the annotation that are changed in COPD.

Table 4. Confirmation of expression levels of EGR1, CYR61, and CTGF by QRT-PCR in individual lung tissues

Gene	Fold change*	
	GOLD-2 (<i>n</i> = 6)	GOLD-4 (<i>n</i> = 3)
EGR1	4.8 ± 1.10	2.6 ± 0.95
CTGF	1.4 ± 0.21	2.4 ± 0.44
CYR61	1.8 ± 0.46	1.9 ± 0.76

*The ratio of mean relative expression level of COPD to control smokers.

5,201 distinct transcripts among 10,000 total transcripts represented on the chip, of which 261 were differentially expressed to a significant degree between GOLD-2 and GOLD-0 smokers. Remarkably, the same groups of cytokine and growth factor genes, ECM genes, and transcription factors that were differentially expressed by SAGE were also detected by microarray analysis. These data support the highly reproducible nature of SAGE for most differentially expressed genes. Six representative genes were selected as differentially expressed with both techniques. Their fold changes are given in Table 2.

The rationale for using the pooled samples above was to directly compare the gene expression profiling between SAGE and microarray analysis. We also performed additional microarray analysis by using independent lung tissue samples. Table 3 shows the microarray data obtained from lung tissues of GOLD-2 smokers (*n* = 6) and GOLD-0 smoker controls (*n* = 6). We also observed similar changes in functional, genes such as those involving growth factors, signal transduction, receptor function, DNA binding, adhesion and cytoskeleton, and metabolism (Table 3).

Confirmation of Candidate Genes by QRT-PCR. Independent assays were performed to measure the expression levels of selected genes by using QRT-PCR on the same RNA samples used for the SAGE and microarray analysis. Table 2 documents the confirmation by QRT-PCR of six genes that were differentially expressed in GOLD-2 and GOLD-0 by both SAGE and microarray analysis results. These genes are EGR1, FOS, CTGF, and CYR61 CX3CL1, which were consistently elevated in GOLD-2, and COL1A1, which was consistently decreased in GOLD-2 relative to GOLD-0. QRT-PCR also verified TGF- β 1 and PDGFRA expression, which were detected by SAGE alone (Table 2). Of these eight genes, EGR-1 (24), FOS (25), and TGF- β 1 (26) are well known COPD-responsive

genes. To our knowledge, the other genes are not known to be associated with COPD or smoking.

We have also confirmed the expression of several of these genes including Egr-1, CTGF, and Cyr61 in independent lung tissues from six GOLD-2 and six GOLD-0 smokers (Table 4). Similar to the QRT-PCR results from pooled RNAs, genes identified by SAGE as differentially expressed in GOLD-2 smokers relative to GOLD-0 smokers were confirmed to be reproducibly expressed in these individuals (Table 4). We also examined the expression of these three genes in independent lung tissue samples from severe COPD patients (GOLD-4 classification). We observed similar levels of expression in the GOLD-4 independent lung tissue samples as in the GOLD-2 samples (Table 4).

Localization of Egr-1, Cyr61, and CTGF Expression in GOLD-2 Lungs. We sought to localize the expression of three selected genes, namely Egr-1, Cyr61, and CTGF in the lung tissues of GOLD-2 and GOLD-0 smokers by using immunohistochemical methods (Fig. 1). Egr-1, Cyr61, and CTGF were detected in the alveolar epithelial cells, small airway epithelial cells, stromal cells, and inflammatory cells in lung tissues from both sources. CTGF was also found in small vessel endothelial cells. Although Egr-1, CTGF, and Cyr61 were observed both in GOLD-2 and GOLD-0 lung tissues, we observed a higher level of expression in the GOLD-2 samples than in the GOLD-0 samples as assessed by percentage of positive staining cells and staining intensity. A higher percentage of positive staining cells was observed in GOLD-2 samples (Egr-1, $9.8 \pm 6.7\%$ of GOLD-2 versus $2.1 \pm 1.6\%$ of GOLD-0, $P < 0.01$; CTGF, $15.5 \pm 15.1\%$ of GOLD-2 versus $4.2 \pm 1.5\%$ of GOLD-0, $P < 0.05$; Cyr61, $5.4 \pm 7.8\%$ of GOLD-2 versus $1.1 \pm 0.8\%$ of GOLD-0, respectively). The relative cell-staining intensity measurement was also increased in the GOLD-2 samples (Egr-1, 95.6 ± 9.6 of GOLD-2 versus 85.8 ± 6.5 of GOLD-0, $P < 0.05$; CTGF, 117.6 ± 2.2 of GOLD-2 versus 105.2 ± 3.5 of GOLD-0, $P < 0.01$; Cyr61, 103.1 ± 6.1 of GOLD-2 versus 100.1 ± 6.2 of GOLD-0, respectively).

Confirmation by QRT-PCR of Candidate Genes in Emphysema Lung Fibroblasts. COPD is the result of complex interactions among epithelial cells, inflammatory cells, and fibroblasts (27). In this experiment, we determined the expression levels of the six selected genes in primary lung fibroblasts from emphysema patients and compared this expression to those from normal donors by using QRT-PCR. The mean relative expression levels of these genes are

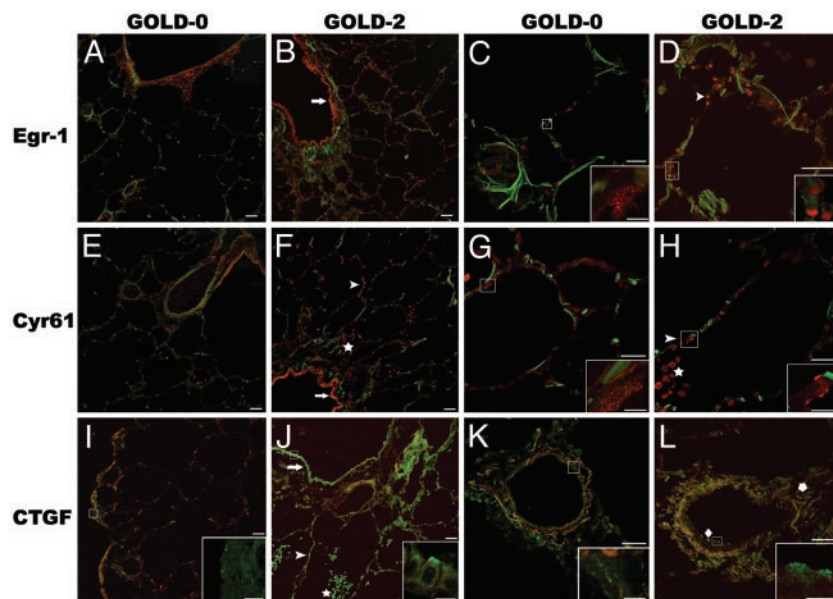


Fig. 1. Localization of Egr-1, Cyr61, and CTGF in the lung tissues of GOLD-2 and GOLD-0 smokers. Egr-1, Cyr61 and CTGF expression, identified by SAGE and microarray analysis and confirmed by QRT-PCR, were examined for protein expression by immunofluorescence staining. (A–D) Egr-1 expression (red stain) was located in small airways (arrow) and alveolar epithelia (arrowhead) of GOLD-0 and GOLD-2 smokers. (E–H) Cyr61 expression (red stain) was located in small airways (arrow), alveolar epithelia (arrowhead), and inflammatory cells (star) of GOLD-0 and GOLD-2 smokers. (I–L) CTGF expression (green stain) was located in small airways (arrow), alveolar epithelia (arrowhead), small vessel endothelial cells (diamond), stromal cells (pentagon), and inflammatory cells (star) of GOLD-0 and GOLD-2 smokers. (Insets) Egr-1, Cyr61, or CTGF staining at higher magnification. Green (Egr-1 and Cyr61) and yellow (CTGF) stains are the autofluorescence. (Scale bars inside the box indicate 20 μ m; scale bars outside the box indicate 100 μ m.)

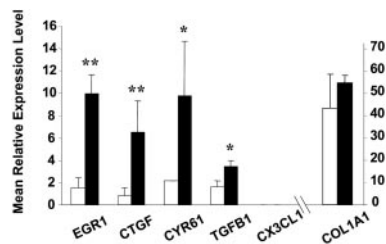


Fig. 2. The mean relative expression level of candidate genes in human lung fibroblasts as determined by QRT-PCR. QRT-PCR was performed on total RNA isolated from human lung fibroblasts. Open bars, normal lung ($n = 2$); filled bars, emphysema lung ($n = 6$). Significant difference between emphysema and normal lung fibroblasts is shown (*, $P < 0.05$; **, $P < 0.01$).

shown in Fig. 2. In agreement with the results obtained by using the whole lung tissues, the expression of EGR1, CTGF, CYR61, and TGFB1 were significantly induced in emphysema lung fibroblasts. COL1A1 was highly expressed and detectable at almost equal levels in both emphysema and normal lung fibroblasts. CX3CL1 was not expressed in cultured lung fibroblasts.

CSE-Induced Egr-1 Protein Expression and DNA-Binding Activity in Human Lung Fibroblasts.

To gain further insight into the mechanism of COPD, we investigated the changes in the expression of these genes by exposing primary fibroblasts from lung tissue of non-COPD smokers (NL9) to the CSE. We selected Egr-1, which has been reported to be highly expressive in late-stage emphysema (24), for further study. As shown in Fig. 3A, NL9 fibroblasts exhibited increased Egr-1 expression in a dose-dependent manner (10–50%). Similar induction of Egr-1 expression was also observed in another human lung fibroblasts (MRC-5) (Fig. 3B). Time course experiments showed that peak CSE-induced Egr-1 expression occurred at 1 h, with decreased expression to control levels by 24 h (Fig. 4A). This induction of Egr-1 protein expression by CSE was accompanied by an increase in the binding activity of Egr-1 to its consensus oligonucleotide sequences in a time-dependent manner (Fig. 4B).

Differential Regulation of MMPs by EGR1. It is well known that TNF- α is crucial to the acute response to smoke (28). TNF- α , together with other cytokines, may activate inflammatory cells and structural cells to produce matrix metalloproteinases and induce matrix breakdown, one of the hallmarks of emphysema (29). EGR1 is considered a major transcription factor for TNF- α (30). To

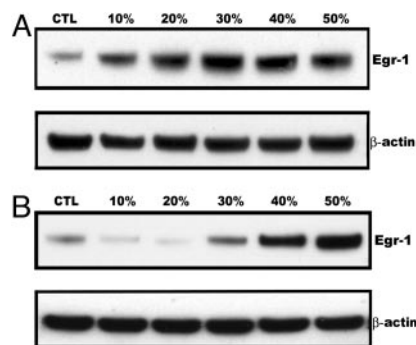


Fig. 3. Dose–response of Egr-1 protein expression in human lung fibroblast cells after exposure to CSE. (A) CSE induces Egr-1 expression in a dose-dependent manner in non-COPD smoker fibroblast cells (NL9). (B) CSE induces Egr-1 expression in a dose-dependent manner in normal fibroblast cells (MRC-5). Fibroblasts were incubated in the absence (CTL) or in the presence of CSE at varying concentrations for 1 h, then whole cell lysates were subjected to Western blot analysis. The same membrane was probed with an anti- β -actin antibody to assess equal loading of the gel.

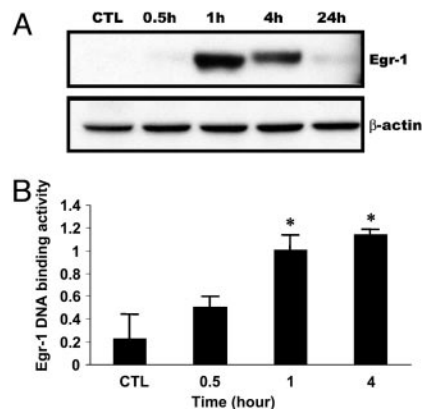


Fig. 4. Time course induction of Egr-1 protein expression and Egr-1 DNA-binding activity in human lung fibroblast cells after exposure to CSE. (A) CSE induces Egr-1 expression in a time-dependent manner in NL9 cells. NL9 cells were incubated in the absence (CTL) or in the presence of CSE (50%) and whole-cell lysates were subjected to Western blot analysis. The same membrane was probed with an anti- β -actin antibody to assess equal loading of the gel. (B) Nuclear extracts were assayed for the activity of transcription factor Egr-1 by an ELISA-based TransFactor kit. Significant difference between 1 and 4 h CSE (50%) treatment and CTL is shown (*, $P < 0.05$).

examine whether EGR1 is involved in the regulation of MMPs activity, we exposed lung fibroblasts from EGR1 null ($-/-$) and EGR1 control ($+/+$) mice to cytomix, and we measured the activities of MMP-2 and MMP-9 that contribute to the degradation of the cell basement membrane (31) with gelatin zymography. As shown in Fig. 5A, cytomix increased the activity of MMP-2 in EGR1 control fibroblasts in a time-dependent manner. However, EGR1 null fibroblasts exhibited significantly less MMP-2 activity after cytomix treatment than the EGR1 fibroblasts controls and this expression did not change after cytomix treatment. In contrast, cytomix increased the activity of MMP-9 in EGR1 null fibroblasts in a time-dependent manner, whereas it had no effect on EGR1 control fibroblasts (Fig. 5B).

Discussion

Recent renewed interests in the candidate genes that may contribute to the pathogenesis of COPD have provided insightful clues to the molecular basis of this irreversible lung disease. To better characterize candidate genes, we have conducted a global comprehensive analysis of gene expression associated with smoking-related COPD.

Several classes of genes were identified, many of which have not been previously associated with COPD, whose expressions are altered in COPD. Aside from the issue of COPD-related differences, the database of >60,000 SAGE tags may be a useful resource for investigators interested in the relative expression levels of mRNAs in human lung. A complete searchable list of all transcripts

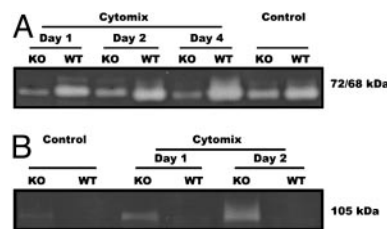


Fig. 5. Gelatin zymography. Gelatin zymography was performed to identify matrix MMP-2 (A) and MMP-9 (B) in the condition media of mouse lung EGR1 null ($-/-$) and $+/+$ fibroblasts. Media were harvested after 1–4 days of culturing in the presence of cytomix. KO, knockout.

is published on the Internet (see *Results*). As can be predicted from statistical calculations, some of these tags may represent previously unidentified genes. Gene identification by generation of longer cDNA fragments from SAGE tags for gene identification (GLGI) (32) may lead to the discovery of genes relevant for COPD.

Among the genes differentially expressed between GOLD-2 and GOLD-0 smoker, SAGE analysis showed a significant up-regulation of many candidate genes. It is beyond the scope of this article to confirm both technical and biological validation of all 327 such candidate genes identified by SAGE analysis. Undoubtedly, many of these candidate genes can represent functionally significant genes in the pathogenesis of COPD. The six genes confirmed by QRT-PCR in this study were chosen primarily because they were replicated consistently by both SAGE and microarray experiments and also because they represent functional classes of genes with important roles in the pathogenesis of COPD. As it turned out, our reliance on repeatable data and insights into COPD pathology were justified by the subsequent functional data. For example, EGR1 mRNA was one of the significantly up-regulated transcripts identified by both SAGE and microarray analysis with confirmation by QRT-PCR and immunohistochemical approaches with the same human samples or emphysematous fibroblasts. Egr-1, a zinc finger transcription factor, has been termed an immediate-early response protein (33, 34) and is potentially activated by a variety of cellular stressors. As a major transcription factor it alters gene expression of EGR1 target genes, including repair enzyme systems, angiogenic factors, cytokines (TNF- α), apoptotic factors (Fas), cell cycle factors (p21, p53), metabolic factors, proteases (MMP-1) (35), and HO-1 (36).

We also observed that CSE treatment could stimulate Egr-1 expression and transcriptional activity. The activity of MMP-2 and MMP-9 of mouse lung fibroblast cells changed in an Egr-1-dependent manner when treated with cytomix. Although there are no Egr-1-binding sites reported in either MMP-2 or MMP-9 promoters, studies showed that the activation of pro-MMP-2 is well correlated with the expression of MIT-MMP (31), whose promoter has a binding site for Egr-1 (35). Our data suggest that EGR1 may serve as a potentially important molecule, which could play a key role in development of cigarette smoke-related COPD by regulating MMP activity and then affect the turnover of ECM proteins during pathogenesis of COPD.

Although little is known regarding the role of MMPs in fibroblasts in the pathogenesis of COPD, our data and recent reports describing the production of MMPs by lung fibroblasts (28) suggest that useful information may be obtained from rigorous investigations in the study of MMPs and lung fibroblasts. Our results support

the protease-antiprotease hypotheses of the development of cigarette smoke-related COPD, which theorizes that the elastolytic proteases overwhelm the antiprotease systems and lead to destruction of alveolar septal architecture (37).

A variety of proteolytic enzymes participate in the alveolar destruction that leads to COPD. Uncertainty exists whether inflammatory cells are the only source of destructive enzymes in the emphysematous lungs. Senior (37) assumed that structural cells of the lung were possible producers of enzymes that degrade lung tissue in emphysema. Recent reports demonstrate that MMP-2 plays a much longer and more important role in the degradation and fragmentation of internal elastic lamina (IEL) than MMP-9, even though both MMP-2 and MMP-9 had contributed to the degradation of the cell basement membrane in chronic flow-induced arterial enlargement (31). We hypothesize here that the production of MMP2 and MMP9 by lung fibroblasts may be involved in the pathogenesis of COPD, which is regulated by EGR1.

Cigarette-smoking-induced chronic inflammation has also long been viewed as central to the pathogenesis of COPD. Airway inflammation could result in the oxidant-antioxidant imbalance and protease-antiprotease imbalance (38). Recently, apoptosis of pulmonary microvascular cells and the resulting lung destruction is suggested to precede the development of emphysema (37, 39). In this study, SAGE analysis identified a group of stress response genes, many genes that regulate inflammation such as growth factors, cytokines, and chemokines, a number of pro-apoptotic and anti-proliferation genes among thousands of genes expressed differentially between GOLD-2 and GOLD-0 smokers (as shown in Table 8).

In summary, we have used SAGE and microarray analysis to identify gene expression profiles of GOLD-2 and GOLD-0 smokers. More studies are needed to determine the specific roles of these genes in the pathogenesis of COPD and to establish whether DNA sequence variation within these genes causes or predicts COPD. We showed that SAGE analysis provides a sensitive and efficient means to study genes associated with this disease and our current study represents a beginning in the systematic analysis of transcripts that are expressed in the lungs of GOLD-2 and GOLD-0 smokers. The list of differentially expressed genes that we have described will provide a foundation for the development of new molecular targets for improved diagnosis, prognosis, and therapy for COPD.

We thank Emeka Ifedigbo for his enthusiasm for all aspects of this work, and Liqiang Xi and Paul R. Reynolds for assistance with the QRT-PCR technique. This work was supported by National Institutes of Health Grants R01-HL60234, R01-AI42365, and R01-HL55330 (to A.M.K.C.).

1. Voelkel, N. F. & MacNee, W. (2002) in *Chronic Obstructive Lung Diseases* (BC Decker, Hamilton, ON, Canada), pp. 90–113.
2. Petty, T. L. (2002) *Chest* **121**, 116S–120S.
3. Sethi, J. M. & Rochester, C. L. (2000) *Clin. Chest Med.* **21**, 1–26.
4. Mayer, A. S. & Mewman, L. S. (2001) *Respir. Physiol.* **128**, 3–11.
5. Pauwels, R. A., Buist, A. S., Calverley, P. M. A., Jenkins, C. R. & Jurd, S. S. (2001) *Am. J. Respir. Crit. Care Med.* **163**, 1256–1276.
6. Velculescu, V. E., Zhong, L., Vogelstein, B. & Kinzler, K. W. (1995) *Science* **270**, 484–487.
7. Kagnoff, M. F. & Eckmann, L. (2001) *Curr. Opin. Microbiol.* **4**, 246–250.
8. Nacht, M., Dracheva, T., Gao, Y., Fujii, T., Chen, Y., Player, A., Akmaev, V., Cook, B., Dufault, M., Zhang, M., et al. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 15203–15208.
9. Boheler, K. R. & Stern, M. D. (2003) *Trends Biotechnol.* **21**, 55–57.
10. Ning, W., Chu, T. J., Li, C. J., Choi, A. M. & Peters, D. G. (2004) *Physiol. Genomics* **18**, 70–78.
11. Mirmics, K., Middleton, F. A., Marquez, A., Lewis, D. A. & Levitt, P. (2000) *Neuron* **28**, 53–67.
12. Kaminski, N. & Friedman, N. (2002) *Am. J. Respir. Cell Mol. Biol.* **27**, 1–8.
13. Benjamini, Y. & Hochberg, Y. (1995) *J. R. Stat. Soc. B* **57**, 289–300.
14. Godfrey, T. E., Kim, S. H., Chavira, M., Ruff, D. W., Warren, R. S., Gray, J. W. & Jensen, R. H. (2000) *J. Mol. Diagn.* **2**, 84–91.
15. Zhou, X., Tan, F. K., Xiong, M., Milewicz, D. M., Feghali, C. A., Fritzier, M. J., Reveille, J. D. & Arnett, F. C. (2001) *J. Immunol.* **167**, 7126–7133.
16. Clark, R. S. B., Kochanek, P. M., Watkins, S. C., Chen, M., Dixon, C. E., Seidberg, N. A., Mellick, J., Loeffert, J. E., Nathaniel, P. D., Jin, K. L., et al. (2000) *J. Neurochem.* **74**, 740–753.
17. Ishii, T., Matsuse, T., Igarashi, H., Masuda, M., Teramoto, S. & Ouchi, Y. (2001) *Am. J. Physiol.* **280**, L1189–L1195.
18. Ning, W., Song, R., Li, C., Park, E., Mohsenin, A., Choi, A. M. & Choi, M. E. (2002) *Am. J. Physiol.* **283**, L1094–L1102.
19. Kleiner, D. E. & Stettler-Stevenson, W. G. (1994) *Anal. Biochem.* **218**, 325–329.
20. Hastie, N. D. & Bishop, J. O. (1976) *Cell* **9**, 761–744.
21. de Waard, V., van den Berg, B. M., Veken, J., Schultz-Heienbrok, R., Pannekoek, H. & van Zonneveld, A. J. (1999) *Gene* **226**, 1–8.
22. Draghici, S., Khatri, P., Bhavsar, P., Shah, A., Kravetz, S. A. & Tainsky, M. A. (2003) *Nucleic Acids Res.* **31**, 3775–3781.
23. Al-Shahrour, F., Diaz-Urriarte, R. & Dopazo, J. (2004) *Bioinformatics* **20**, 578–580.
24. Zhang, W., Yan, S. D., Zhu, A., Zou, Y. S., Williams, M., Godman, G. C., Thomashow, B. M., Ginsburg, M. E., Stern, D. M. & Yan, S. F. (2000) *Am. J. Pathol.* **157**, 1311–1320.
25. DiCamillo, S. J., Carreras, I., Panchenko, M. V., Stone, P. J., Nugent, M. A., Foster, J. A. & Panchenko, M. P. (2002) *J. Biol. Chem.* **277**, 18938–18946.
26. Takizawa, H., Tanaka, M., Takami, K., Ohtoshi, T., Ito, K., Satoh, M., Okada, Y., Yamasawa, F., Nakahara, K. & Umeda, A. (2001) *Am. J. Respir. Crit. Care Med.* **163**, 1476–1483.
27. Knight, D. (2001) *Immunol. Cell Biol.* **79**, 160–164.
28. Zhu, Y. K., Liu, X., Ertl, R. F., Kohyama, T., Wen, F. Q., Wang, H., Spurzem, J. R., Romberger, D. J. & Rennard, S. I. (2001) *Am. J. Respir. Cell Mol. Biol.* **25**, 620–627.
29. Churg, A., Zay, K., Shay, S., Xie, C., Shapiro, S. D., Hendricks, R. & Wright, J. L. (2002) *Am. J. Respir. Cell Mol. Biol.* **27**, 368–374.
30. Shi, L., Kishore, R., McMullen, M. R. & Nagy, L. E. (2002) *Am. J. Physiol.* **282**, C1205–C1211.
31. Sho, E., Sho, M., Singh, T. M., Nanjo, H., Komatsu, M., Xu, C., Masuda, H. & Zarins, C. K. (2002) *Exp. Mol. Pathol.* **73**, 142–153.
32. Chen, J., Rowley, J. D. & Wang, S. M. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 349–353.
33. Lee, S. L., Sadovsky, Y., Swirnow, A. H., Polish, J. A., Goda, P., Gavriliina, G. & Milbrandt, J. (1996) *Science* **273**, 1219–1221.
34. Yan, S. F., Pinsky, D. J., Mackman, N. & Stern, D. M. (2000) *J. Clin. Invest.* **105**, 553–554.
35. Haas, T. L., Stitelman, D., Davis, S. J., Apte, S. S. & Madri, J. A. (1999) *J. Biol. Chem.* **274**, 22679–22685.
36. Yang, G., Nguyen, X., Ou, J., Rekulapelli, P., Stevenson, D. K. & Dennery, P. A. (2001) *Blood* **97**, 1306–1313.
37. Senior, R. M. (2000) *Chest* **117**, 320S–323S.
38. MacNee, W. (2001) *Eur. J. Pharmacol.* **429**, 195–207.
39. Shin, H. J., Park, K. K., Lee, B. H., Moon, C. K. & Lee, M. O. (2003) *Toxicology* **191**, 121–131.