Rapid Communication

Gene Expression in Lung Adenocarcinomas of Smokers and Nonsmokers

Charles A. Powell,* Avrum Spira,* Adnan Derti, Charles DeLisi, Gang Liu, Alain Borczuk, Steve Busch, Sudhir Sahasrabudhe, Yangde Chen, David Sugarbaker, Raphael Bueno, William G. Richards, and Jerome S. Brody

Department of Medicine, Columbia College of Physicians and Surgeons, New York, New York; Pulmonary Center and Department of Medicine, Boston University School of Medicine, and Bioinformatics Program, College of Engineering, Boston University, Boston, Massachusetts; Department of Pathology, Columbia University School of Medicine, New York, New York; Aventis Pharmaceuticals, Bridgewater, New Jersey; and Division of Thoracic Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts

Adenocarcinoma (AC) has become the most frequent type of lung cancer in men and women, and is the major form of lung cancer in nonsmokers. Our goal in this paper was to determine if AC in smokers and nonsmokers represents the same genetic disease. We compared gene expression profiles in resected samples of nonmalignant lung tissue and tumor tissue in six never-smokers with AC and in six smokers with AC, who were matched for clinical staging and histologic criteria of cell differentiation. Results were analyzed using a variety of bioinformatic tools. Four times as many genes changed expression in the transition from noninvolved lung to tumor in nonsmokers as in smokers, suggesting that AC in nonsmokers evolves locally, whereas AC in smokers evolves in a field of genetically altered tissue. There were some similarities in gene expression in smokers and nonsmokers, but many differences, suggesting different pathways of cell transformation and tumor formation. Gene expression in the noninvolved lungs of smokers differed from that of nonsmokers, and multidimensional scaling showed that noninvolved lungs of smokers groups with tumors rather than noninvolved lungs of nonsmokers. In addition, expression of a number of genes correlated with smoking intensity. Our findings, although limited by small sample size, suggest that additional studies comparing noninvolved to tumor tissue may identify pathogenetic mechanisms and therapeutic targets that differ in AC of smokers and nonsmokers.

Lung cancer is the leading cause of cancer deaths in both men and women in the United States (1). Adenocarcinoma (AC) is the most common form of lung cancer and has increased relative to other histologic types of lung cancer over the last several decades. Although cigarette smoking is the principle causal agent of lung cancer, 10–15% of patients with lung cancer have no history of smoking. The majority of these individuals have AC of the lung and are women (Report of the Lung Cancer Progress Review Group, National Cancer Institute, August 2001. http://osp.nci.nih.gov/prg_assess/prg/lungprg/lung_rpt.htm). The changing pattern of lung cancer, and the rising incidence of lung cancer in nonsmokers, prompted us to ask whether AC in nonsmokers is the same genetic disease as in smokers.

DNA microarray technology is capable of producing large gene expression data sets that can provide novel insights into fundamental cancer biology at the molecular level. Microarrays have been applied to a number of different cancers, and several studies comparing different histologic types of lung cancer and lung cancer with different outcomes have appeared recently (2–6). However, none of these studies have explored the specific differences between lung cancer in smokers and nonsmokers, and none have compared nonmalignant to tumor tissue in the same subject. In this study we used gene expression arrays to compare AC with nonmalignant lungs of six smokers and six nonsmokers who were matched for clinical and histologic features of cancer. Although the small numbers in our study limit conclusions regarding specific genes, our findings suggest AC in smokers arises in a field of genetically altered lung tissue, whereas AC in nonsmokers arises locally in relatively normal lung tissue. Although changes in a number of genes are shared in the transition from nonmalignant lung to tumor in smokers and nonsmokers, the majority of genes differ. In addition, there were a number of genes whose expression differed between the nonmalignant lung of smokers compared with nonsmokers, and many of these genes correlated with smoking history. Our studies, although only suggestive because of small sample size, point to the need for further studies of adenocarcinomas of smokers and nonsmokers.

Materials and Methods

Patients and Tissue Specimens

We collected primary lung adenocarcinomas and histologically nonmalignant lung tissue from patients undergoing lung cancer resection at Brigham and Women's Hospital (Boston, MA). Samples were snap-frozen in liquid nitrogen and stored at −140°C. We chose six cases of AC in never-smokers from the database, and six cases of adenocarcinoma from cigarette smokers, selected by matching for the following criteria in a descending order of priority: (*i*) cell type; (*ii*) histologic stage of differentiation; (*iii*) clinical stage; and (*iv*) patient age. Each sample was accompanied by an adjacent section for histologic confirmation. A pathologist

(*Received in original form September 6, 2002 and in revised form February 14, 2003*)

* These authors contributed equally to the work presented in this article.

Address correspondence to: Jerome S. Brody, M.D., Pulmonary Center R3, Boston University Medical Center, 715 Albany St., Boston, MA 02118. E-mail: jbrody@bu.edu

Abbreviations: adenocarcinoma, AC; multidimensional scaling, MDS; normal, N; nonsmoker normal, Nns; smoker normal, Ns; real-time polymerase chain reaction, QRT-PCR; tumor, T; transforming growth factor, TGF; nonsmoker tumor, Tns; smoker tumor, Ts.

Am. J. Respir. Cell Mol. Biol. Vol. 29, pp. 157–162, 2003
Originally Published in Press as DOI: 10.1165/rcmb.2002-0183RC on February 21, 2003
Internet address: www.atsjournals.org

determined the proportion of tumor cells and emphysema in each section. Nonmalignant samples contained no recognizable tumor cells; tumor cells made up on average 60% of cells in tumor tissues. (There were no significant differences in amount of emphysema between nonmalignant samples.) The study was approved by the Human Studies Committees of Brigham and Women's Hospital and Boston University Medical Center.

Microarray Data Acquisition and Normalization

RNA was extracted using RNAeasy kits (Qiagen, Valencia, CA). The HuGene FL array (Affymetrix, Santa Clara, CA), which contains $\sim$ 6,800 human genes, was used to generate gene expression profiles for the 24 tissue samples. Six to eight micrograms of total RNA from each of the 24 samples was prepared, hybridized onto the HuGene FL array, and scanned following the Affymetrix protocol as described (7). 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).

To filter out arrays of poor quality, we assessed several quality control measures and excluded from further analysis two arrays that differed significantly from the others. We scaled the data from each array to normalize the results for inter-array comparisons. We filtered out genes whose median detection *P* value was not less than 0.05 in at least one of the four comparative groups, leaving 2,047 genes for which data was available in 22 samples. Many of the genes traditionally linked with cancer (i.e., p53, retinoblastoma etc.) were not found on the array or were filtered (*see* www.netaffx.com for list of all genes on the HU6800 array).

Statistical Analysis

Following data normalization and filtering, we used several methods to generate lists of genes with different levels of expression among four comparative groups of samples. We performed a paired *t* test to identify genes with statistically significant differences in expression levels between three groups of matched experiments (each patient matched to themselves): smoker tumor (Ts) versus smoker normal (Ns), nonsmoker tumor (Tns) versus nonsmoker normal (Nns), and all tumor (T) versus all normal (N). An unpaired *t* test was used to identify genes with differences in expression levels between Ns versus Nns tissue. We selected a *P* value threshold of 0.01 for statistical significance. Due to the presence of multiple comparisons, there was the potential problem of finding genes differentially expressed between two groups when no difference actually exists (8). Current methods for adjustments for multiple comparisons break down quickly, as they are too conservative (9). In addition, they assume independence of the different tests, which is unlikely to hold true in the microarray setting, where multiple genes are co-regulated (10). We first compared the number of genes at our *P* value threshold with the expected number of false positives under the null hypothesis to estimate the overabundance of information in the analyzed dataset (11). In addition, we employed a permutation test to assess the significance of our *P* value threshold for any given gene's comparison between two groups (9).

We also performed a permutation-based neighborhood analysis to select the top 50 genes that distinguish the comparative groups (12). The permutation-test neighborhood analysis was used to select marker genes for each of the classes and to assess their statistical significance with respect to a reference empirical distribution obtained by permuting the phenotype class labels (12). The neighborhood analysis using the *t* test distance function and 1,000 permutations was performed with Genecluster 2.0b software (http://www.genome.wi.mit.edu/cancer/software/software.html).

Following the application of a gene variation filter (*see* www.bubiopulmatics.org for details), hierarchical clustering of the genes and samples using a Pearson correlation (uncentered) similarity metric and average linkage clustering was performed using CLUSTER and TREVIEW software programs obtained at http://rana.lbl.gov/EisenSoftware.htm. Multidimensional scaling (MDS) of samples according to the expression of the top 30 paired *t* test genes was performed using the Partek 5.0 software (www.partek.com). In addition, MDS was performed on all tumor samples according to the expression of top 20 *t* test genes that distinguished Ns versus Nns. We also calculated a correlation coefficient (r value ranging from –1 to 1) for each gene, to quantify how well a given gene's expression level in smoker normal lung correlates with pack-years of smoking (*see* www.bubiopulmatics.org for details).

Confirmatory Array Studies

To confirm genes differentially expressed, we analyzed a second set of unmatched samples, one each of Ts, Ns, Tns, and Nns, using an Affymetrix U95A array that included most of the genes present on the HU6800 array. Sample collection, microarray data acquisition, and data normalization were performed as described above. We performed multidimensional scaling of these four new samples with the initial samples for three comparisons (T versus N, Ts versus Ns, and Tns versus Nns), using the top 30 genes from the paired *t* test comparisons.

Real-time polymerase chain reaction (QRT-PCR) was used to confirm the differential expression of a select number of genes that changed in transition from normal lung to tumor among nonsmokers. Primer sequences were designed with Primer Express software (Applied Biosystems, Foster City, CA). Forty cycles of amplification, data acquisition, and data analysis were performed in an ABI Prism 7,700 Sequence Detector (Applied Biosystems).

In addition, we compared our results with data generated by Bhattacharjee and coworkers (2), to confirm our findings among a larger number of lung adenocarcinoma samples. Microarray Suite 5.0 software (Affymetrix) was used to analyze Affymetrix U95A Genechip image files from the 139 adenocarcinoma and 17 normal lung samples from that study. Given that the smoking status of all patients (including all normal samples) in that study was unavailable, we limited our comparison to the genes that distinguish all tumors from all normal lung samples regardless of smoking status. Hierarchical clustering of all samples from the Bhattacharjee dataset was performed using genes identified as differentialy expressed ($P < 0.01$) between T versus N in our study.

Supplemental Information

Additional information on the statistical methods, microarray protocol, expression levels for all genes in all samples (stored in a relational database), and data analysis is available at www.bubiopulmatics.org.

Results

Table 1 presents demographic data for matched pairs of smokers and nonsmokers. Figure 1 shows that four times as many genes differed at the $P < 0.01$ level in Tns versus Nns than in Ts versus Ns. The strength of this type of analysis lies in the fact that each subject's own tissue served as the control for tumor gene expression. The small sample size in our study limits the conclusions one can draw about specific gene changes because of the false discovery rate inherent in the multiple comparisons. However, the overabundance of information in the analyzed dataset demon-

TABLE 1

Subject demographics

Patient No.	Age	Sex	Smoke	Stage	Grade
1	60–65	M	Yes	T3N0	Mod/poor
2	70–75	F	No	T2N0	—
3	45–50	F	Yes	T2N2	Poor
4	40–45	F	No	T2N2	Poor
5	65–70	F	Yes	T1N0	Poor
6	60–65	F	No	T1N0	Mod/poor
7	—	F	Yes	—	Mod
8	65–70	F	No	T1N0	Well
9	50–55	F	Yes	T2N0	Well/BAC
10	40–45	F	No	T2N0	Well/BAC
11	65–70	F	Yes	T2N2	Mod
12	65–70	F	No	T2N3	Well/BAC

Subjects 1 and 2, 3 and 4, etc. were matched for various clinical and histologic criteria (*see* MATERIALS AND METHODS). Clinical stage and histologic state of differentiation are listed.

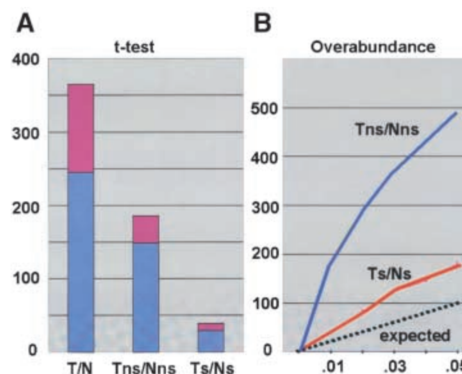

Figure 1. Number of genes that differ between groups by *t* test. (*A*) Total number of genes that differ at the $P < 0.01$ level between T and N, Tns and Nns, and Ts and Ns. *Bar* for each group includes genes that decreased (*bottom*) and genes that increased (*top*). (*B*) Overabundance graph: a comparison of the number of genes at various *P* value thresholds with the expected number of false positives under the null hypothesis reveals that 20 differences would be expected at a $P < 0.01$ and 102 at a $P < 0.05$. The differences for Tns versus Nns were more than four times that expected, and for Ts versus Ns were twice that expected.

strates that more than four times as many differences in gene expression occurred than would be expected as a result of multiple comparisons for Tns versus Nns, and twice as many changes for Ts versus Ns (Figure 1). Although a number of changes were shared in smokers and nonsmokers, there were many differences in the Ts and Tns transition.

Table 2 shows selected changes that occurred in Tns/Nns but not in Ts/Ns, changes that occurred in both Tns/Nns and Ts/Ns, and changes in Tns/Nns that also appeared in Ns/Nns comparisons. A complete list of genes that changed in each setting can be found on the website (www.bubiopulmatics.org). Tns was characterized by decreased expression of two transforming growth factor (TGF)-β receptors and of several components of TGF-β signaling such as dematopoetin and endoglin, and downstream TGF-β genes such as connective tissue growth factor. These microarray findings were validated by QRT-PCR, which revealed downregualtion of TGF-βR2, TGF-βR3, endoglin, and CTGF in transition from normal lung to adenocarcinoma in nonsmokers (see www.bubiopulmatics.org for QRT-PCR data). TGF-β has been shown to be a growth suppression gene in a number of epithelial cells. Tns was also characterized by increased expression of several potential oncogenes and decreased expression of several putative tumor suppressor genes, e.g., FEZ1 and BTG1. In addition, a number of cell–cell genes and cell–matrix genes decreased and matrix invasion genes increased, suggesting movement of tumor cells through matrix in Tns. Table 2 shows some of the genes that were shared in the *t* test comparison between Tns/Nns and Ns/Nns. Several of

TABLE 2

Genes that change in transition from Nns to Tns

Unique to Tns versus Nns			Shared with Ts versus Ns			Shared with Ns versus Nns		
TGFbR2	D50683	0.23	AEGR	U89336	0.05	Mig2	Z24725	0.38
TGFbR3	L07594	0.35	GLP3	L47125	0.07	Id3	X69111	0.40
DPT	Z22865	0.32	ABC3	U78735	0.53	Id1	HG3342	0.23
Endoglin	X72012	0.27				DTR	M60278	0.29
CTGF	M92934	0.32	FoxF1	U13219	0.16			
			SPC	J03890	0.06	TIMP3	U14394	0.30
Muc1	HG371	2.74				AQP1	U41518	0.25
CSNK1	U29174	1.44	GAPDH		2.37	Elastin	HG2994	0.48
Pik	U01038	1.41	CAT	X04085	0.31	GPC3	L47125	0.07
BCL3	U05681	0.69	GPX3	D00632	0.21	A2M	M11313	0.21
Axl	HG162	0.62						
Fez1	U60060	0.31	Tie-2	L06139	0.08			
BTG1	X61123	0.57	CDH5	X79981	0.31			
ICAM	M24283	0.28	Cav-1	Z18951	0.07			
PCAM	L34657	0.21						
TNA	X64559	0.02						
MDK	M92450	3.91						

Gene symbols, accession numbers, and fold change for genes that changed between Tns and Nns in first column only in Tns, in second column also in Ts, and in third column also in Ns.

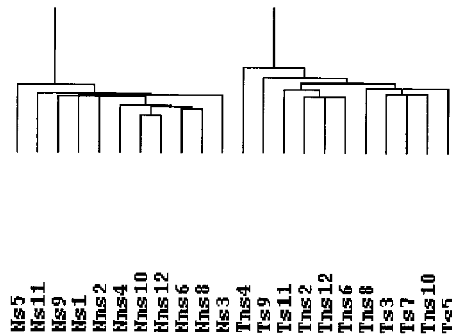

Figure 2. Dendrogram of unsupervised hierarchical clustering. Unsupervised clustering of all samples according to the expression of all filtered genes. T and N samples separated into two groups. Within N samples, four of the five Ns clustered together. Within the T group, four of six Tns clustered together.

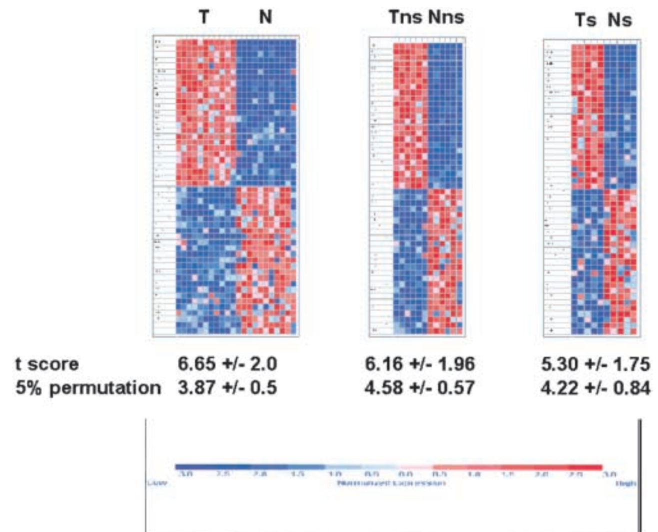

Figure 3. Permutation-based neighborhood analysis. Top 50 genes (by *t* score) differentially expressed between three comparative groups. Genes in *red* were increased, and genes in *blue* decreased compared with the mean expression level (*see legend*). On top are individual samples; on side are genes (list of genes on website). Genes that distinguish between T and N, Tns and Nns, and Ts and Ns are fairly consistent within each class. However, the top genes in the Tns versus Nns neighborhood analysis provide a stronger and more consistent distinction between classes as compared with genes differentially expressed in Ts versus Ns. This is confirmed by the higher *t* scores, which represent the mean difference between groups for each of the 50 genes divided by the sum of their standard variations, for Tns versus Nns. Class labels were permuted 1,000 times, and the 5% permutation score represents the *t* score that would be expected to occur by chance in less than 5% of permutations if there were no true difference between the classes (*see* website for further detail).

the genes that decreased (e.g., Id1, Id3, and mig2) are putative tumor suppressor genes, and one, DTR or heparin-binding epidermal growth factor–like gene, is a known oncogene. Many of these genes were also identified in the neighborhood analysis (*see* below).

We also used more robust methods designed to show relations between subject groups. However, these methods do not allow comparisons between tumor and nontumor sample from the same subject. Unsupervised hierarchical clustering (Figure 2) separated tumor and nontumor samples, and grouped 4/5 noninvolved smokers tissue together. However, Ts and Tns were not completely separated by this method. Neighborhood analysis, an algorithm that selects those genes that best distinguish classes and quantifies their ability to distinguish classes (as compared with what would be expected by chance), showed clear differences between T and N, Tns and Nns, and Ts and Ns (Figure 3); the mean *t* score, a measure of difference in gene expression between groups, was greater than that expected by chance (5% permutation) in all three groups. A complete list of the top 50 genes in each category can be found on the website. Sixteen of the 50 top genes changed in a similar fashion in the transition from noninvolved lung to tumor in smokers and nonsmokers. These included decreased expression of cell–cell and cell–matrix-associated genes such as CDH5, A2M, and GPC3; the latter two have been shown to bind and inactivate growth factors. There was altered expression of hypoxia/stress-related genes such as GAPDH and PKM2 (increased) and GPX and Cox7A (decreased), and changes in cancer-related genes such as RAGE, Cav1, EMP2, and KAL1, all decreased. In addition, two endothelial genes, PCAM and TEK, decreased, and α1Col increased, suggesting a change in lung interstitial tissue components. Differential expression of RAGE and GPC3 was confirmed by QRT-PCR (see www.bubiopulmatics.org for data). Differences between noninvolved lung and tumor in smokers and nonsmokers included, in smokers, several myc-associated genes (APEX and PHB); and ras-associated genes (GFR, GADD45, and PCTK1); and in nonsmokers, expression of several cell and matrix adhesion molecules (e.g., TIMP3 and MDK) and several genes associated with poor prognosis in other tumors (e.g., CEACAM1, LGAL3BP, and MUC1).

Although Ns versus Nns were also separated by the neighborhood analysis method, the differences were not as dramatic as the first three groups, indicated by the mix of *red* and *blue* in Figure 4A and by the 5% permutation score exceeding the *t* score (Figure 4). The Ns group contained individuals who smoked between 15 and 150 pack-years. If gene expression was related to amount of smoking, one might expect highly variant patterns of gene expression in Ns. Indeed, 166 genes had a correlation coefficient of > 0.8 when plotted against pack-years of smoking. Genes that correlated with pack-years (*see* Figure 4 and website) included a number of immunomodulatory genes (e.g., butyrophilin; $r = 0.95$), apoptosis-related genes (SLC25A; $r = -0.92$), and signaling genes (zynin; $r = 0.91$, Pax6; $r = 0.99$), as well as oxidative stress genes (mitochondrial NADH; $r = 0.93$). In addition, hnRNP, a gene that has been studied as an early diagnostic marker of lung cancer in sputum and in alveolar lavage (13), was positively correlated with pack-years of smoking.

Multidimensional scaling using the top 30 genes from the paired *t* tests effectively distinguished between T and N,

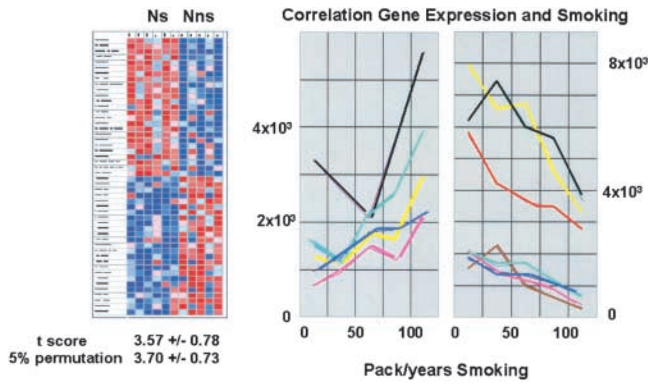

Figure 4. Neighborhood analysis and correlation with smoking history in Ns. Neighborhood analysis (on left) reveals that the top 50 genes (by *t* score) separating Ns and Nns show considerable variability within each class, as indicated by the overlap of *red* and *blue* expression values and by the mean *t* score being less than 5% permutation score (more than 5% of permutations would be expected to have a higher *t* score by chance). Pearson correlation for pack-years and gene expression is shown, on the *right*, for select genes with R value > 0.8 or < −0.8. Genes positively correlated with pack-years include, top to bottom, BRD2, DXS59928, TGM1, PAX6, and BTN2A2. Genes negatively correlated with pack-years, from top to bottom, are SFTPB, COX4I1, CAPZA1, FXR1, RAF1, NFI/B, and RAP1A.

Ts and Ns, and Tns and Nns (Figure 5). New unpaired samples were assigned to the correct categories. In addition, MDS and hierarchcal clustering of all tumor samples according to the expression of the top 20 *t* test genes that distinguish Ns versus Nns revealed clustering of the tumor samples closer to the Ns samples (Figures 6A and 6B). Hierarchical clustering of all adenocarcinoma and normal lung tissue samples from the study by Bhattacharjee and colleagues, using the genes that distinguished all T versus N in Table 1, resulted in all but one of the normal lung samples clustering separately from the tumors (*see* supplementary figure at www.bubiopulmatics.org).

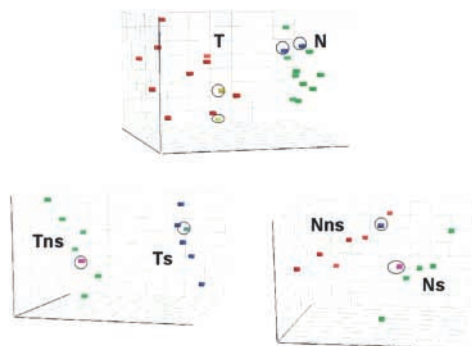

Figure 5. Multidimensional Scaling (MDS) with addition of four new samples. MDS plot of samples in three-dimensional space according to the expression of top 30 genes from the paired *t* test comparisons. The four new samples (*circled*) group with their appropriate class from the initial sample set. Ns, normal smoker sample; Nns, normal nonsmoker sample; Ts, tumor smoker sample; Tns, Tumor nonsmoker sample.

Discussion

Access to nonmalignant lung tissue and lung tissue with AC from the same individual in nonsmokers and smokers who were matched for tumor cell type, histologic state of differentiation, clinical stage, and age, has provided a unique opportunity to determine whether AC of the lung in smokers and nonsmokers evolve by similar mechanisms. This feature of our study design enabled us to use each subject's noninvolved lung tissue as a control, and thus enhance our

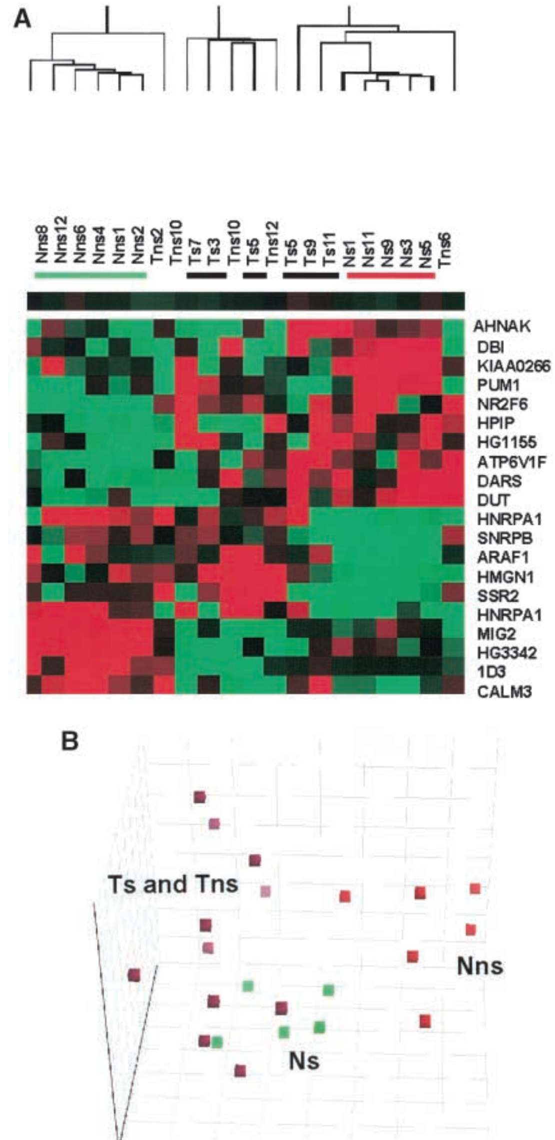

Figure 6. "Normal" lungs of smokers group with tumor lungs. Hierarchical clustering and MDS plot using top 20 *t* test genes that distinguish Ns from Nns show that Ns samples group closer to tumor samples than do Nns samples, consistent with the concept that Ns have premalignant changes associated with a smoking-induced "field defect". (*A*) In cluster figure, *green* represents low level of expression, and *red* high level of expression. *Green bar* above cluster identifies Ns, and *red bar* identifies Nns samples. (*B*) MDS in two-dimensional space shows that Ns group with all tumors rather than with Nns lungs.

ability to detect gene changes that occur in transition from normal lung tissue to AC of the lung. Four times as many genes changed in the transition between nonmalignant lung and tumor in nonsmokers as in smokers, suggesting, because the tumors were matched for clinical and histologic stage, that the noncancerous lungs of smokers already had many alterations in gene expression. This is confirmed by comparison of noncancerous lungs of smokers and nonsmokers, and is consistent with the concept that smoking causes widespread field defects in lungs and airways of smokers, and that tumors in smokers arise within a field of genetically abnormal cells (14, 15). In contrast, AC in nonsmokers likely arises in a field of relatively normal cells, as might be the case with prior infection, such as seen in "scar carcinomas."

The asymmetry between the number of genes measured and the number of samples limits the analysis of specific gene changes identified by paired *t* tests; therefore, we used alternative analytic approaches that compared groups of individuals but eliminated the comparison of normal and tumor tissue in the same individual. Several different analytic methods confirmed the conclusion that the transition between N and T differed in smokers and nonsmokers, and that the noncancerous lungs of smokers differed from those of nonsmokers. This conclusion was confirmed by multidimensional scaling, which assigned both original samples and several new samples not used to establish the original groupings, to the correct class in multidimensional space.

Although these methods identified genes that changed in a similar fashion and thus were characteristic of AC in smokers and nonsmokers, a number of genes changed only in Ts or in Tns. Tns is characterized by decreased expression of many genes associated with TGF-β signaling and by changes in cell and matrix genes, suggesting major changes in cell–cell and cell–matrix interactions. In addition, there were differences in gene expression between Ns and Nns, and expression levels of many genes correlated positively or negatively with the intensity or duration of smoking. Many of these latter genes likely represent premalignant changes in the noncancerous lungs of smokers. This is supported by the fact that tumor samples group together with Ns samples according to the expression of the top 20 genes that distinguish Ns versus Nns (Figure 5B). However, it is not known whether these changes are unique to smokers who develop lung cancer or are present in all smokers. Similarly, it is not know whether any of these changes are associated with the presence of concomitant COPD, although only one of the 166 genes whose expression correlates with pack-years of smoking in this study also correlated with pack-years of smoking in a study of lung tissue gene expression in COPD that is underway in our laboratory (data not published).

This study, although preliminary, suggests that AC of smokers and of nonsmokers have different etiologies, and involve different pathways of cell transformation. This implies that optimal approaches to treatment might differ in AC of smokers and nonsmokers. Our study also suggests that a number of premalignant changes occur in the noninvolved lungs of smokers; these changes in gene expression might represent targets for preventative therapy. The possibilities raised in this article suggest that a larger study comparing both noncancerous and tumor tissue in smokers and nonsmokers, and examining gene expression samples in tissue from smokers with and without cancer, would be of considerable clinical importance.

Acknowledgments: This work was supported in part by the Doris Duke Charitable Foundation Clinical Scientist Development Award (A.S.).

References

1. Greenlee, R. T., M. B. Hill-Harmon, T. Murray, and M. Thun. 2001. Cancer statistics, 2001. *CA Cancer J. Clin.* 51:15–36.
2. Bhattacharjee, A., W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson. 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* 98:13790–13795.
3. Garber, M. E., O. G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyna-Gengelbach, R. M. van de Rijn, G. D. Rosen, C. M. Perou, R. I. Whyte, R. B. Altman, P. O. Brown, D. Botstein, and I. Petersen. 2001. Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. USA* 98:13784–13789.
4. Giordano, T. J., K. A. Shedden, D. R. Schwartz, R. Kuick, J. M. Taylor, N. Lee, D. E. Misek, J. K. Greenson, S. L. Kardia, D. G. Beer, G. Rennert, K. R. Cho, S. B. Gruber, E. R. Fearon, and S. Hanash. 2001. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. *Am. J. Pathol.* 159:1231–1238.
5. Beer, D. G., S. L. Kardia, C. C. Huang, T. J. Giordano, A. M. Levin, D. E. Misek, L. Lin, G. Chen, T. G. Gharib, D. G. Thomas, M. L. Lizyness, R. Kuick, S. Hayasaka, J. M. Taylor, M. D. Iannettoni, M. B. Orringer, and S. Hanash. 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* 8:816–824.
6. Anbazhagan, R., T. Tihan, D. M. Bornman, J. C. Johnston, J. H. Saltz, A. Weigering, S. Piantadosi, and E. Gabrielson. 1999. Classification of small cell lung cancer and pulmonary carcinoid by gene expression profiles. *Cancer Res.* 59:5119–5122.
7. Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander, and T. R. Golub. 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* 96:2907–2912.
8. Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. (Ser. B)* 57:289–300.
9. MacDonald, T. J., K. M. Brown, B. LaFleur, K. Peterson, C. Lawlor, Y. Chen, R. J. Packer, P. Cogen, and D. A. Stephan. 2001. Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat. Genet.* 29:143–152.
10. Tusher, V. G., R. Tibshirani, and G. Chu. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* 98:5116–5121.
11. Kaminski, N., and N. Friedman. 2002. Practical approaches to analyzing results of microarray experiments. *Am. J. Respir. Cell Mol. Biol.* 27:125–132.
12. Pomeroy, S. L., P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. Kim, L. C. Goumnerova, P. M. Black, C. Lau, J. C. Allen, D. Zagzag, J. M. Olson, T. Curran, C. Wetmore, J. A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D. N. Louis, J. P. Mesirov, E. S. Lander, and T. R. Golub. 2002. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 415:436–442.
13. Wistuba, I. I., S. Lam, C. Behrens, A. K. Virmani, K. M. Fong, J. LeRiche, J. M. Samet, S. Srivastava, J. D. Minna, and A. F. Gazdar. 1997. Molecular damage in the bronchial epithelium of current and former smokers. *J. Natl. Cancer Inst.* 89:1366–1373.
14. Powell, C. A., S. Klares, G. O'Connor, and J. S. Brody. 1999. Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. *Clin. Cancer Res.* 5:2025–2034.
15. Sueoka, E., N. Sueoka, Y. Goto, S. Matsuyama, H. Nishimura, M. Sato, S. Fujimura, H. Chiba, and H. Fujiki. 2001. Heterogeneous nuclear ribonucleoprotein B1 as early cancer biomarker for occult cancer of human lungs and bronchial dysplasia. *Cancer Res.* 61:1896–1902.