Advances in Brief

Classification of Small Cell Lung Cancer and Pulmonary Carcinoid by Gene Expression Profiles[1]

Ramaswamy Anbazhagan, Tarik Tihan, Daniel M. Bornman, James C. Johnston, Joel H. Saltz, Anthony Weigering, Steven Piantadosi, and Edward Gabrielson[2]

Divisions of Molecular Pathology [R. A., D. M. B., J. C. J., E. G.], Neuropathology [T. T.], Pathology Informatics [J. H. S., A. W.], and Oncology Biostatistics [S. P.], Johns Hopkins University School of Medicine, Baltimore, Maryland 21224

Abstract

Small cell lung cancer is a common type of lung cancer that is generally classified within the spectrum of neuroendocrine lung neoplasms. Using high-density cDNA arrays, we profiled gene expression of small cell lung cancers and compared these expression profiles to those of normal bronchial epithelial cells and pulmonary carcinoids, which are classified as benign neuroendocrine tumors. We found the overall expression profiles of two small cell lung cancer cell lines, two microdissected tissue samples of primary small cell lung cancer, and cultured bronchial epithelial cells to be relatively similar to one another, with an average Pearson correlation coefficient for these comparisons of 0.63. However, we found the expression profiles of small cell lung cancers (and bronchial epithelial cells) to be surprisingly dissimilar to those of two samples of pulmonary carcinoid tumors, with an average correlation coefficient for these comparisons of 0.20. We then compared the pulmonary carcinoid expression profiles to those of two samples of infiltrating astrocytic brain cancers (oligodendroglioma and high-grade astrocytoma) and found similarity of gene expression among these four samples (average correlation coefficient, 0.57). These gene expression profiles suggest that small cell lung cancers are closely related to (and possibly derived from) epithelial cells, and that pulmonary carcinoids are related to neural crest-derived brain tumors. More generally, our results suggest that broad profiles of gene expression may reveal similarities and differences between tumors that are not apparent by traditional morphological criteria.

Introduction

SCLC[3] accounts for ~20% of all lung cancers, and >100,000 new cases of this cancer occur each year worldwide (1). SCLC is a highly aggressive type of cancer that has distinctive clinical manifestations, including frequent and widespread metastases and high sensitivity to chemotherapy. This tumor also has unique pathological features, including scant cytoplasm, finely granular chromatin, and by ultrastructural studies, scattered dense core neuroendocrine granules (2). SCLC tumors also stain frequently for neuroendocrine markers such as neuron-specific enolase, chromogranin, and synaptophysin (3), and as a result of this evidence for neuroendocrine differentiation, SCLC is commonly regarded as part of the spectrum of neuroendocrine lung tumors (4).

We initiated a study of gene expression of SCLC using cDNA arrays with a goal to identify specific gene expression changes related to the pathogenesis of this disease. For purposes of comparison, we measured gene expression patterns in cultured human bronchial epi-

thelial cells, a normal lung cell type that is thought to represent the progenitor cell for many bronchogenic lung cancers. Because we had concerns that bronchial epithelial cells would not accurately reflect features of neuroendocrine differentiation, we also measured gene expression patterns in pulmonary carcinoid tumors. Carcinoid tumors are relatively benign neuroendocrine tumors with neurosecretory granules and immunochemical staining for the same neuroendocrine markers that stain SCLC tumors (5) and are thus often considered to be benign counterparts to SCLC in the spectrum of neuroendocrine tumors (4).

Although our initial goal in these studies was to identify specific genes involved in the pathogenesis of SCLC, we recognized that we might be able to determine how similar these tumors are to one another by comparing the gene expression profiles of the different tumor samples. Encouraging results for grouping of tumors by gene expression arrays were reported recently for alveolar rhabdomyosarcoma cell lines, which have patterns of gene expression that are relatively similar to one another and relatively different from those of other human tumor cell lines (6). Therefore, using our gene expression array data as a broad representation of gene expression for each sample, we calculated correlation coefficients for each sample-to-sample comparison to estimate how different samples are from one another. We then performed a hierarchical clustering analysis to determine groupings among the tumor expression profiles.

Materials and Methods

Cell Lines and Tissue Samples. Two SCLC cell lines (HUT 69 and HUT 209) were purchased from the American Type Culture Collection (Gaithersburg, MD), cultured using RPMI 1640 supplemented with 10% fetal bovine serum (Biofluids, Gaithersburg, MD), and harvested in log growth for extraction of RNA as described below. Primary human bronchial epithelial cells were removed from bronchi of two surgical specimens and cultured for two passages using fully supplemented bronchial epithelial basal medium (Clonetics, San Diego, CA), prior to harvesting for RNA. mRNAs from two patients' bronchi were combined together for a single sample. The primary tumor tissue samples (two samples of small cell lung carcinoma, two samples of pulmonary carcinoid, one sample of astrocytoma grade 4, and one sample of oligodendroglioma) were collected and frozen at the time of surgery at the Johns Hopkins Hospital or the Johns Hopkins Bayview Medical Center. Frozen sections of these tissues were placed on Rnase-Zap (Ambion, Austin, TX)-treated slides and microdissected as described previously (7) prior to extraction of RNA.

Analysis of Gene Expression by cDNA Arrays. From each tissue or cell culture sample, total RNA was extracted using Trizol reagent (Life Technologies, Gaithersburg, MD), and mRNA was isolated from total RNA using the Micro Poly(A) Pure mRNA Isolation kit (Ambion) using manufacturers' protocols. High-density cDNA filters (Gene Discovery Array Human I version 1.2) were purchased from Genome Systems (St. Louis, MO). Using the Genome Systems protocols, [^{33}P]dCTP-labeled cDNA probe was prepared using 2.5-μg aliquots of mRNA and hybridized to filters. A new array was used for each sample. After being washed, filters were imaged on a Molecular Dynamics Storm phosphorimager after a 48-h exposure. Digital images were

processed (Genome Systems) to quantitatively measure hybridization intensities for each spot.

Analysis and Comparisons of Gene Expression Data. To compare expression levels among the different samples, we first normalized the level of each gene (both spots) to the total of all genes measured for that sample. The expression level of each gene was thus expressed as a fraction of the total of all genes measured. We then calculated all 36 pair-wise Pearson correlation coefficients for each sample to sample comparison using all 18,210 normalized gene expression measurements. Methodological limitations of using the correlation coefficients in this setting include assumptions of independence for each measurement, and these assumptions were not verified. Furthermore, correlation coefficients represent a single-dimensional measure of similarity between two samples, and estimating relationships of multiple samples to one another represents a complex, multidimensional statistical problem that may not be accurately reflected by correlation coefficients. However, considering the relatively small numbers of samples studied, hierarchical clustering analysis using the correlation coefficients can reasonably represent these relationships. The final step in the analysis, therefore, was to construct hierarchical clustering dendograms with the Statistical Analysis Systems[4] statistical package, using each of the nine samples as a reference.

Results and Discussion

Radiolabeled cDNA samples from each of the nine samples were incubated with an array containing 18,210 cDNA clones, and hybridization to the arrays was measured using a Molecular Dynamics Storm phosphorimager. Representative arrays hybridized to samples are shown in Fig. 1.

These arrays contain duplicate spots for each gene represented, and in general, these duplicate measurements were consistent with one another with ~80% of paired results showing <35% variation. We found reasonable reproducibility when one sample (HUT 209) was hybridized to two different array membranes, with a Pearson correlation coefficient for this comparison of 0.82. Furthermore, we observed expected patterns of expression in the various samples for known gene markers such as chromogranin A (accession number W23477 in Table 1). Finally, for several of the genes with large differences in gene expression among the different samples, we validated these differences by RNase protection analysis using expressed sequence tags corresponding to genes differentially expressed on the arrays. For some of these tested genes, however, clones proposed by the manufacturer to correspond to specific array spots did not contain the specified expressed sequence tag insert, and we were unable to validate the expression values measured by arrays. Expression levels of representative genes that have significantly different levels among the various samples are summarized in Table 1. The entire data files for all samples and summaries of specific validation experiments are released separately on the World Wide Web.[5]

Although we noted several apparent manufacturing defects in the arrays,[6] these defects appear to be generally consistent among different arrays within the same manufacturer version. We therefore reasoned that the arrays should be useful for comparing overall gene expression profiles from one sample to another, similar to the approach used in a recent study that compared gene expression profiles of alveolar rhabdomyosarcoma cell lines to those of other types of human cancer cell lines (6). Sample-to-sample comparisons of overall expression profiles can be effectively displayed using scatterplots, and the extent of similarity between any two samples can be estimated by a bivariate correlation coefficient. Representative scatterplots and the correlation coefficients for all comparisons are shown in Fig. 2.

Using these correlation coefficients as a simple measure of relative similarity between two samples, it became readily evident to us that
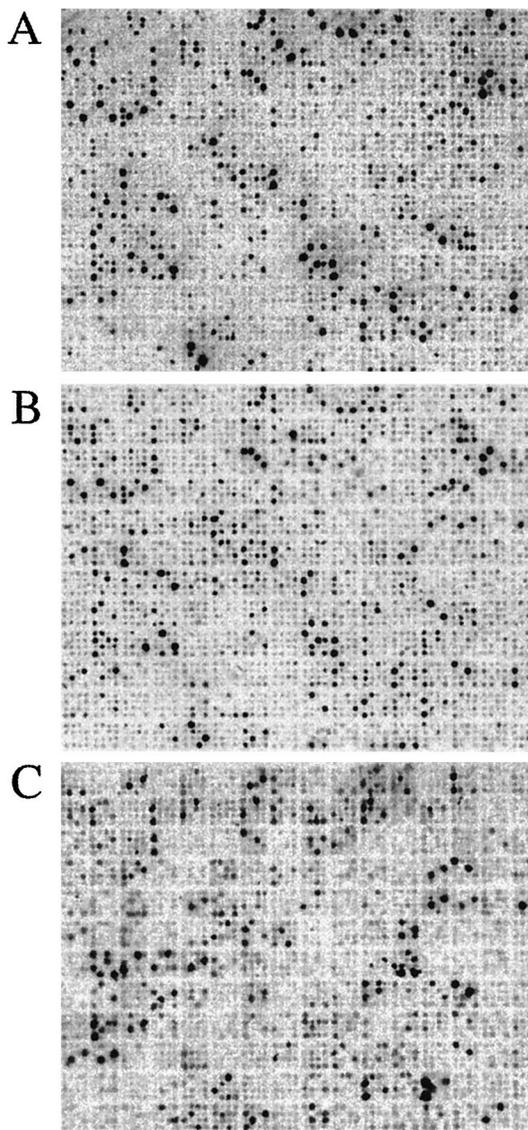

Fig. 1. Representative portions of arrays hybridized with radiolabeled cDNA from bronchial epithelium (*A*), primary SCLC #1 (*B*), and carcinoid sample #1 (*C*). The area depicted for each array represents 8.33% of the entire array.

gene expression profiles are relatively similar among the samples of SCLCs and bronchial epithelial cells. Pearson correlation coefficients for comparisons of samples within this group ranged from 0.36 to 0.80, with a mean of 0.63 (95% confidence interval, 0.34–0.92). In addition, we recognized the gene expression profiles of pulmonary carcinoid tumors to be relatively similar to those of the oligodendroglioma and high-grade astrocytoma samples, with a mean of correlation coefficients for comparisons of samples within this group of 0.57 (95% confidence interval, 0.36–0.79). Remarkably, we noted a relative dissimilarity of gene expression for comparisons of samples between groups (SCLCs and bronchial epithelial cells *versus* carcinoids and brain tumors) with a mean of correlation coefficients for these comparisons of only 0.20 (95% confidence interval, 0.08–0.33). We did not perform a formal *t* test analysis of this data because the correlation coefficients are not independent variables. However, based on the 95% confidence intervals for the within-group comparisons not overlapping with those for the between-group comparisons, we found a convincing degree of dissimilarity between the two groups. Notably, the consistency of our observations suggests that the differences we observed between SCLC and carcinoid were not random fluctuations

---

[4] Internet address: www.sas.com.
[5] Internet address: http://128.220.85.49/genomics.
[6] Described at Internet address: http://128.220.85.49/genomics.

Table 1 *Expression levels of selected genes with differential levels of expression among five of the samples studied*

GenBank ID	Fraction of all genes expressed ($\times 10^6$)[a]				
	BE	HUT 209	SCLC-1	CAR-1	OLIGO
R90793	4660	1557	6754	38	30
H16193	6182	2053	11815	78	42
R88587	12437	4815	7938	72	31
R48041	4812	5314	7037	99	41
R39478	111	47	16	1289	124
AA069527	4	6	4	1745	516
R68506	163	16	9	2407	345
N94259	4	4	5	566	513
H86896	159	126	467	7696	1266
W23477[b]	14	46	27	419	35

[a] All expression values are normalized and expressed as a fraction of total ($\times 10^6$) of all genes measured for the particular sample. BE, bronchial epithelial cells; H209, a small cell lung cancer cell line; SCLC-1, a primary small cell lung cancer sample; CAR-1, a carcinoid sample; OLIGO, the oligodendroglioma sample. All expression data is released in an expanded table on the World Wide Web (http://128.220.85.49/genomics).

[b] Accession number W23477 designates the *chromogranin A* gene.

A
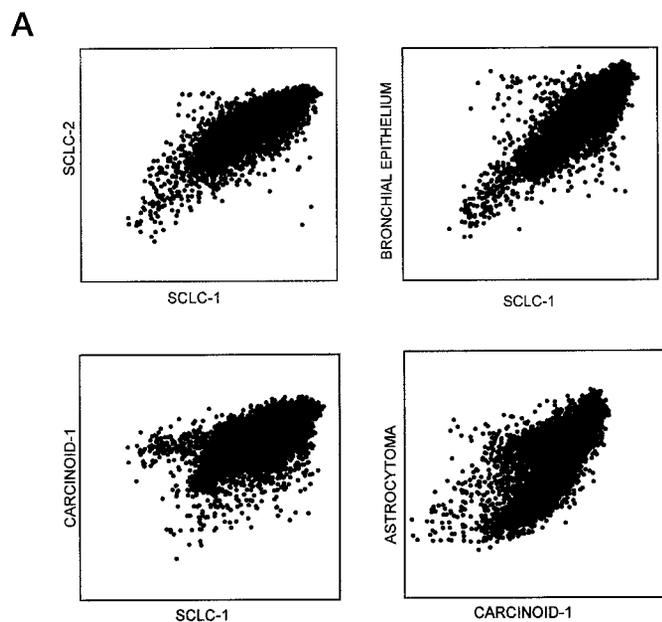
B

	BE	H69	H209	SCLC1	SCLC2	CAR1	CAR2	OLIGO
H69	0.526	·						
H209	0.799	0.513	·					
SCLC1	0.801	0.504	0.760	·				
SCLC2	0.649	0.355	0.659	0.701	·			
CAR1	0.109	0.099	0.129	0.146	0.253	·		
CAR2	0.149	0.204	0.196	0.226	0.208	0.534	·	
OLIGO	0.127	0.229	0.239	0.201	0.197	0.533	0.598	·
ASTRO	0.193	0.302	0.332	0.283	0.254	0.429	0.588	0.760

Fig. 2. *A*, scatterplots for comparisons of selected samples. In each scatterplot, the $\log_{10}$ of the normalized expression level of each gene was plotted for both samples designated. *B*, correlation coefficients for all pairwise comparisons of samples are tabulated. Samples are bronchial epithelial cells (*BE*), H69 and H209 (SCLC cell lines), *SCLC1* and *SCLC2* (primary SCLC samples), *CAR1* and *CAR2* (carcinoid samples), *OLIGO* (oligodendroglioma sample), and *ASTRO* (astrocytoma grade 4 sample).

in global gene expression or results of experimental artifacts but rather reflect intrinsic biological differences between these two neoplastic types.

Using the correlation coefficients in Fig. 2, the nine possible dendograms generated with each sample as a reference consistently clustered the small cell cancer samples together with the sample of bronchial epithelial cells but separated them from the carcinoid samples. The resulting dendograms did not depend strongly on which sample was chosen as a reference. The dendogram shown in Fig. 3 represents the groupings obtained using the oligodendroglioma sam-

ple as the reference sample and represents the best groupings by both statistical and biological considerations.

SCLC and typical carcinoids have long been regarded to both be within the spectrum of neuroendocrine tumors of the lung, a category that also includes atypical pulmonary carcinoids and large cell neuroendocrine carcinoma (4). This classification is primarily based on all of these tumors sharing features of neuroendocrine differentiation, although these tumors have also been shown recently to share loss of heterozygosity at several chromosomal regions including 11q13 and 3p14.3–21 (8). Our broad measurements of gene expression using arrays allow us to reconsider the relationship of SCLC to pulmonary carcinoid by comparing gene expression on a more comprehensive level than previously possible. It is notable that the arrays did confirm high expression of individual neuroendocrine markers, such as neuron-specific enolase, in both SCLC and carcinoid samples. However, the overall patterns of gene expression would suggest that SCLCs are closely related to, and possibly derived from, bronchial epithelial cells but distantly related to pulmonary carcinoids. Carcinoids, in turn, appear to be more closely related to neural crest-derived brain tumors. We therefore propose that SCLCs be considered a part of the spectrum of epithelial lung cancers rather than a part of the spectrum of pulmonary neuroendocrine tumors.

Our proposed alignment of SCLC as an epithelial neoplasm distantly related to carcinoid actually fits well with a number of clinical and pathological features of this disease. For example, many lung cancers have histological heterogeneity, and SCLC exists frequently with an admixture of adenocarcinoma, squamous cell carcinoma, or undifferentiated large cell patterns of lung cancer (9). These findings would be compatible with a common histogenesis for these different forms of lung cancer. In contrast, the coexistence of carcinoid and non-SCLC has not, to our knowledge been reported. Furthermore, although atypical carcinoids with infiltrative and metastatic properties are recognized (5), pulmonary carcinoids have not been observed to progress to classical SCLC. Thus, there is clinical evidence for a link between SCLC and non-small cell epithelial lung cancer but not between SCLC and carcinoid tumors.

Although we found that carcinoid tumors had gene expression profiles dissimilar to SCLC, the carcinoid tumors were found by our measurements to have a surprising similarity to two types of glial brain cancer, oligodendroglioma and high-grade astrocytoma. Glial cells are considered to be derived from neural crest cells, as are possibly the Kulchinstky cells of the bronchi (10). The gene expression profiles measured in our samples, therefore, support a concept that pulmonary carcinoid tumors are derived from these Kulchinsky cells (5) and thus ultimately from neural crest.

In summary, our measurements of gene expression profiles provide strong evidence to classify SCLCs as epithelial-derived neoplasms and pulmonary carcinoid tumors as being related to neural crest-derived brain tumors. Moreover, an important general implication of
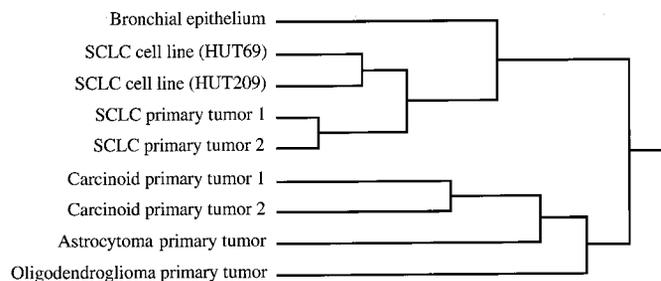

Fig. 3. Hierarchical clustering dendogram constructed using the oligodendroglioma sample as a reference. Lengths of *horizontal lines* are proportional to dissimilarity between samples, as estimated by 1 − Pearson correlation coefficient.

5121

this study is that gene expression profiles may help us to recognize similarities and differences among tumors that cannot be recognized by traditional morphological examination. Notably, we demonstrated a high degree of dissimilarity between two types of tumors thought previously to be related, which significantly extends the previously reported demonstration of alveolar rhabdomyosarcoma cell lines having gene expression profiles distinct from those of tumors with obviously very different histogenesis (6). With refinements and improvements of gene expression array technology, there will likely be increasingly prominent roles for gene expression profiles in classification and characterization of tumors.

References

1. Parkin, D. M., Pisani, P., and Ferlay, J. Estimates of the worldwide incidence of eighteen major cancers in 1985. Int. J. Cancer, *54:* 594–606, 1993.
2. Mackay, B., Ordonez, N. G., Bennington, J. L., and Dugan, C. C. Ultrastructural and morphometric features of poorly differentiated and undifferentiated lung tumors. Ultrastruct. Pathol., *13:* 561–571, 1989.
3. Guinee, D. G., Jr., Fishback, N. F., Koss, M. N., Abbondanzo, S. L., and Travis, W. D. The spectrum of immunohistochemical staining of small-cell lung carcinoma in specimens from transbronchial and open-lung biopsies. Am. J. Clin. Pathol., *102:* 406–414, 1994.
4. Travis, W. D., Linnoila, R. I., Tsokos, M. G., Hitchcock, C. L., Cutler, Jr., Nieman, L., Chrousos, G., Pass, H., and Doppman, J. Neuroendocrine tumors of the lung with proposed criteria for large-cell neuroendocrine carcinoma. An ultrastructural, immunohistochemical, and flow cytometric study of 35 cases. Am. J. Surg. Pathol., *15:* 529–553, 1991.
5. Colby, T. V., Koss, M. N., and Travis, W. D. Carcinoid and other neuroendocrine tumors. *In:* Atlas of Tumor Pathology, pp. 287–317. Washington, DC: Armed Forces Institute of Pathology, 1994.
6. Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S. B., Pohida, T., Smith, P. D., Jiang, Y., Gooden, G. C., Trent, J. M., and Meltzer, P. S. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res., *58:* 5009–5013, 1998.
7. Fujii, H., Biel, M. A., Zhou, W., Weitzman, S. A., Baylin, S. B., and Gabrielson, E. Methylation of the *HIC-1* candidate tumor suppressor gene in human breast cancer. Oncogene, *16:* 2159–2164, 1998.
8. Onuki, N., Wistuba, I. I., Travis, W. D., Virmani, A. K., Yashima, K., Brambilla, E., Hasleton, P., and Gazdar, A. F. Genetic changes in the spectrum of neuroendocrine lung tumors. Cancer (Phila.), *85:* 600–607, 1999.
9. Roggli, V. L., Vollmer, R. T., Greenberg, S. D., McGavran, M. H., Spjut, H., and Yesner, R. Lung cancer heterogeneity: a blinded and randomized study of 100 consecutive cases. Hum. Pathol., *16:* 569–579, 1985.
10. Cutz, E. Neuroendocrine cells of the lung: an overview of morphologic characteristics and development. Exp. Cell Res., *3:* 185–208, 1982.

Classification of Small Cell Lung Cancer and Pulmonary Carcinoid by Gene Expression Profiles

Ramaswamy Anbazhagan, Tarik Tihan, Daniel M. Bornman, et al.

Updated version	Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/59/20/5119

Cited articles	This article cites 8 articles, 1 of which you can access for free at: http://cancerres.aacrjournals.org/content/59/20/5119.full#ref-list-1
Citing articles	This article has been cited by 11 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/59/20/5119.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/59/20/5119. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.