

Integrated classification of lung tumors and cell lines by expression profiling

Carl Virtanen*, Yuichi Ishikawa†, Daisuke Honjoh*, Mami Kimura*, Miyuki Shimane*, Tatsu Miyoshi†, Hitoshi Nomura*, and Michael H. Jones**

*Chugai Pharmaceuticals, 3-41-8 Takada, Toshima-ku, Tokyo 171-8545, Japan; and †Department of Pathology, Cancer Institute, Japanese Foundation for Cancer Research, 1-37-1 Kami-Ikebukuro, Toshima-ku, Tokyo 170-8455, Japan

Edited by Albert de la Chapelle, Ohio State University, Columbus, OH, and approved July 30, 2002 (received for review April 23, 2002)

The utility of cancer cell lines depends largely on their accurate classification, commonly based on histopathological diagnosis of the cancers from which they were derived. However, because cancer is often heterogeneous, the cell line, which also has the opportunity to alter *in vitro*, may not be representative. Yet without the overall architecture used in histopathological diagnosis of fresh samples, reclassification of cell lines has been difficult. Gene-expression profiling accurately reproduces histopathological classification and is readily applicable to cell lines. Here, we compare the gene-expression profiles of 41 cell lines with 44 tumors from lung cancer. These profiles were generated after hybridization of samples to four replicate 7,685-element cDNA microarrays. After removal of genes that were uniformly up- or down-regulated in fresh compared with cell-line samples, cluster analysis produced four major branch groups. Within these major branches, fresh tumor samples essentially clustered according to pathological type, and further subclusters were seen for both adenocarcinoma (AC) and small cell lung carcinoma (SCLC). Four of eight squamous cell carcinoma (SCC) cell lines clustered with fresh SCC, and 11 of 13 SCLC cell lines grouped with fresh SCLC. In contrast, although none of the 11 AC cell lines clustered with AC tumors, three clustered with SCC tumors and six with SCLC tumors. Although it is possible that preexisting SCC or SCLC cells are being selected from AC tumors after establishment of cell lines, we propose that, even *in situ*, AC will ultimately progress toward one of two poorly differentiated phenotypes with expression profiles resembling SCC or SCLC.

Lung cancer is the leading cause of cancer death in men and women worldwide and continues to rise in frequency. In Japan alone, lung cancer claims 50,000 people annually (1). Lung carcinomas are generally classified as either small cell lung carcinoma (SCLC) or non-SCLC (NSCLC). SCLC is grouped with carcinoid tumors on the basis of neuroendocrine features. Together these neuroendocrine tumors account for 15–25% of all lung cancers. SCLCs are generally inoperable, because most patients present at an advanced stage, and although initially responsive to chemotherapy, prognosis is poor. Typical carcinoid tumors, on the other hand, are usually resectable and have much better prognosis. The remaining lung cancers, including adenocarcinoma (AC), squamous cell carcinoma (SCC), and large cell carcinoma (LCC) are classified as NSCLCs. More commonly than SCLC, NSCLCs are localized at the time of diagnosis and surgically resected. The prognosis for NSCLCs is variable.

In lung carcinomas, heterogeneity is noted to varying degrees, and although the majority of NSCLCs contain mixtures of different cancer cell types, they are usually classified as a mixed carcinoma only where the minority cell type exceeds an arbitrary proportion of the tumor cells. This is the case in adenosquamous carcinoma, which is composed of a mixture of at least 10% AC and squamous cells (2). Combined small cell carcinomas contain both SCLC and NSCLC components, and their existence indicates that all lung carcinomas are derived from a common progenitor cell type.

The degree of identity between lung cancer cell lines and lung tumors is not well defined; beyond the initial characterization of the cancer from which they were derived, cell lines lack the overall tissue architecture needed for accurate histopathological classification. Although it is likely that the most easily established cell lines such as SCLC will resemble cancer *in situ* to the greatest extent, the degree to which particular cell lines diverge during establishment and propagation *in vitro* is unknown. Furthermore, cell lines isolated from heterogeneous cancer may be assumed wrongly to be representative. Expression profiling, which is applicable to both cell lines and tumors, provides the opportunity to verify cell-line classification.

Here, we report microarray-based expression profiling of lung tumors and cell lines. Our results for fresh tumors follow histopathological classification very closely. Furthermore, our analysis shows that SCC and SCLC cell lines can follow predictable paths toward molecular pathologies, retaining similarity to their corresponding well differentiated tumors. In marked contrast, AC cell lines lose their original identity and move to molecular pathologies more closely related to either SCCs or SCLCs, which may reflect the progression of tumors *in situ*.

Materials and Methods

Cell Lines and Clinical Sample. Details of the 41 cell lines, 43 tumors, and 6 normal samples are given in Tables 1 and 2, which are published as supporting information on the PNAS web site, www.pnas.org. Surgical samples were from patients who had undergone surgery at the Cancer Institute (Tokyo) with informed consent and ethical committee approval. The reference sample, designed to be representative of all the tissues and cell types tested, contained 20% fresh normal lung mixed with 8% each from cell lines NCI-H226, NCI-H522, EKVX, NCI-H460, Lu-134A, MRC5, SQ5, H23, PC14, and MS1.

Array Preparation. First, 6,734 of the 7,685 elements were derived from IMAGE clones purchased from Research Genetics (Huntsville, AL). In addition, 699 proprietary clones with a known or suspected role in cancer were used. Then, 252 control genes used for quality assurance were removed from the final analysis. All IMAGE clones that are referred to by name in this paper were resequenced. The preparation of slides was performed essentially as described (3). Two 7,685-element arrays were printed in duplicate on each slide by using a MicrogridII (BioRobotics, Cambridge, U.K.).

Probe Preparation. Before harvesting, the cell lines were resuspended in RNAlater (Ambion, Austin, TX). Surgical samples were snap-frozen in liquid nitrogen. RNA was extracted by using Isogen (Nippon Gene, Toyama, Japan). cRNA was prepared essentially as described (4). cRNA (2 μ g) was used as a template

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

Abbreviations: SCLC, small cell lung carcinoma; NSCLC, non-SCLC; AC, adenocarcinoma; SCC, squamous cell carcinoma; LCC, large cell carcinoma.

†To whom reprint requests should be addressed. E-mail: mike@chugai-pharm.co.jp.

for probe synthesis as described (5). The reference sample was labeled with Cy3, and the test sample was labeled with Cy5. Hybridization and washing were performed as described (3). Slides were scanned and quantified by using a Scannaray 4000 scanner and Quantarray (Packard Bioscience, Billerica, MA).

Statistical Analysis, Filtering, and Clustering. All cRNA samples were labeled in duplicate. Each duplicate sample was applied to two separate arrays, giving four measurements for each gene. Normalization within an array experiment was performed by using a lowess normalization (6) and by dividing the normalized signal intensity by the control channel intensity. The four repeat measurements of each gene then were examined for outliers as described (7). Outliers were flagged for exclusion from further analysis. Data were imported into GENESPRING (Silicon Genetics, Redwood City, CA) for clustering and visualization. An initial scan of the gene set using a binary search algorithm (described in *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site) was performed to remove genes differentially expressed in fresh samples compared with cell lines. Genes then were filtered according to their expression in the control channel. Genes in which more than half of the 90 samples had an absorption measurement less than 1,000 were removed from analysis. A Welch's ANOVA was performed by using a Bonferonni correction at the 0.05% significance level to find genes that varied significantly across samples. Two-way hierarchical clustering of genes and samples was achieved by using similarity measurements based on Pearson correlations around 0. Standard Pearson correlations were also used to make sure that choice of similarity measurement had no marked effect on the final tree structure. Dendrograms displaying the degree of similarity between adjacent clusters of genes or samples were created to visualize the results. All cluster diagrams are presented in Microsoft EXCEL spreadsheet format and are published as supporting information on the PNAS web site for ease of access.

Results

Preliminary Clustering. The microarray used here incorporated an estimated 6,671 unique genes. The raw, background-subtracted data are shown in Data Sets 1 and 2, which are published as supporting information on the PNAS web site. Included in our negative controls were constructs containing *AluSq*, *AluSx*, and firefly *luciferase* gene. We were unable to completely compete away hybridization to these elements in all experiments. However, they proved useful in identifying a group of 285 genes that had similar expression patterns to all three control elements across the data set (>0.5 Pearson similarity in 90% of the group). These genes were regarded as suspect and removed from further consideration. This initial filtering reduced the number of unique genes from 6,671 to 6,141.

Two-way hierarchical clustering (Fig. 1) of these data produced a dendrogram with tumor samples and cell lines falling into two distinct groups. Within the tumor group, all the SCLCs are found in one branch and NSCLCs in the other. Subbranches were formed that contained samples essentially clustering according to pathological type. For cell lines, a general distinction could be made between SCLC and NSCLC, and within the NSCLC branch, SCC cell lines produced a distinct cluster. Two of the three fibroblasts also produced a distinct cluster. However, the AC cell lines failed to cluster separately, tending to cluster with SCLC and SCC cell lines. Isolated clustering of cell lines only (Fig. 5, which is published as supporting information on the PNAS web site) produced similar results.

Generation of a Tree with Integrated Cell Lines and Tumors. To better understand the relationship between cell-line and tumor samples, we decided to filter the data to promote integration. Visual

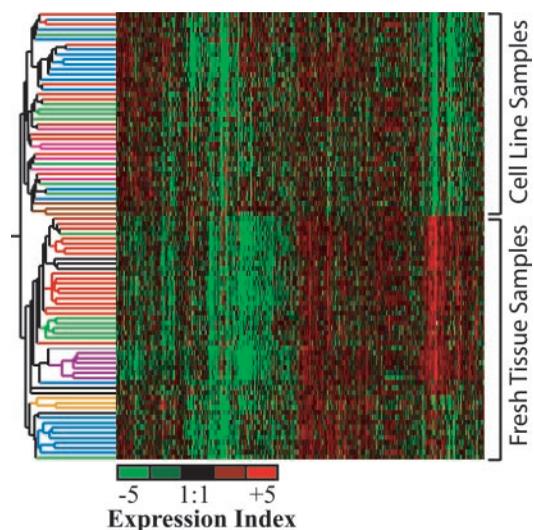


Fig. 1. Dendrogram of a two-way hierarchical clustering of 6,141 genes. Samples, colored according to type, are indicated to the left: red, AC; green, SCC; purple, normal; blue, SCLC; black, LCC; orange, carcinoids; brown, fibroblasts; and pink, NSLC. Each column represents a particular gene. Squares are colored according to the log mean expression ratios across four replicates. Red indicates expression ratios greater than 1 (overexpression), green less than 1 (underexpression), and black roughly equal to 1 (no expression change) in relation to the reference sample.

inspection of the initial clustering (Fig. 1) revealed large groups of genes that were overexpressed or underexpressed systematically in fresh tissue relative to the cell lines. Many of these genes will be derived from the component of “contaminating” normal tissue that is only present in the fresh samples. Other genes may be proliferation-associated and therefore more highly expressed in the cell lines. We used a binary search algorithm to identify 2,103 genes differentially expressed (1,209 overexpressed and 894 underexpressed) in tumor compared with cell lines. Plotting the expression ratios for these genes (Fig. 2*A* and *B*) illustrates the ability of the binary algorithm to identify genes that are generally differentially regulated between tumors and cell lines. Visual inspection after hierarchical clustering of these 2,103 genes (Fig. 6, which is published as supporting information on the PNAS web site) revealed a distinct group of 299 genes without marked contrast in expression between fresh tissue and cell lines. These 299 genes were not filtered and were used in the final clustering (Fig. 3). Hierarchical clustering of the remaining 1,804 genes selected by the algorithm (Fig. 2*C*) shows the clear distinction they make between fresh and cell-line samples. The expression of these genes varies slightly within the fresh tumor samples such that they retained some ability to classify the various histopathological types. These 1,804 genes were filtered out from further analysis.

After all the filtering steps, 4,253 genes remained and were used for subsequent analysis. Hierarchical clustering of these genes (Fig. 3) resulted in a large degree of integration of cell lines into the tumor branches. The four main branches produced were an SCLC branch, an SCC branch, a cell-line branch, and a branch containing normal tissue, AC and LCC. A small cluster with two NSCLC cell lines breaks off at a similar level to these four main branches. Although none of the major branches exclusively contain a particular pathological type, with the exception of two NSCLCs clustering with the SCLCs, all the SCLC and NSCLC tumors cluster separately.

Clustering of Fresh Samples. At least eight different gene groups can be defined that are regulated distinctly in various tumor categories (Figs. 3 and 4).

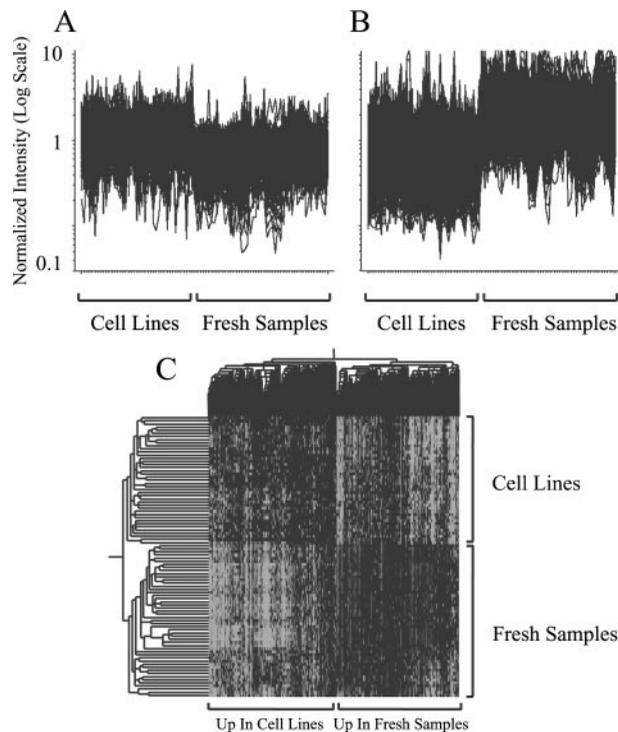


Fig. 2. Identification of genes differentially regulated between tumors and cell lines. Expression change across all samples for 905 genes, the expression of which was generally overexpressed in cell-line samples compared with fresh samples (A), and 1,313 genes, the expression of which was generally underexpressed in cell-line samples compared with fresh samples (B). (C) Two-way hierarchical clustering of 1,804 genes that are differentially regulated between fresh samples and cell lines. The left side of the dendrogram shows the clustering of fresh samples with cell-line samples. Branch colors are as indicated for Fig. 1. The upper dendrogram shows the clustering of genes.

Within the SCLC branch, three subclusters are defined. These include a group made up solely of cell lines (SCLC-cell lines) discussed below, a group of SCLC tumor samples (SCLC-fresh), and a group containing both fresh and cell-line SCLC (SCLC-mixed). The SCLC-fresh group also contains three NSCLC samples. Comparison of survival times for SCLC patients in each of the two branches containing tumor samples did not reveal any association between cluster group and survival, although sample sizes were small, and prognosis for SCLC patients is almost invariably poor. At least three groups of genes (Fig. 4 F–H), could be identified that contributed to the clustering of the three SCLC groups. Cluster F contains genes that are up-regulated only in the two SCLC groups containing tumors. Many of the genes in these clusters have known roles in tumorigenesis. For example, SOX4 acts as a transcriptional regulator and has been implicated in colon cancer (8). The second SCLC group, represented by genes in Fig. 4G, contains genes that are only up-regulated in the cluster containing cell lines and tumors (SCLC-mixed). A number of the genes found here are associated with neuroendocrine differentiation. These include reticulon 1 (9) and reelin, which is up-regulated in esophageal carcinoma (10). A third cluster (Fig. 4H) contains genes that are generally overexpressed in all three SCLC groups. Two of these genes, kinesin heavy chain member 2 (11) and stathmin (12), are related to microtubule function, predominantly in neural tissue.

The major branch containing the SCLCs also has a subbranch containing all three carcinoids. Although they have different treatment regimes and much better prognosis than SCLC, carcinoids sometimes are grouped pathologically with SCLCs on

the basis of shared neuroendocrine characteristics. Our classification is consistent with this and supports the molecular classification produced by Bhattacharjee *et al.* (13). Amongst several notable genes in the carcinoid cluster (Fig. 4A), vitronectin functions in cell attachment (14).

The SCC branch contains five of the eight SCC tumors. Two of the remaining three SCCs clustered with the AC group despite expressing many genes that are characteristic of the SCC group (Fig. 4D). Genes predominantly overexpressed in tumors within the SCC group (Fig. 4D and E) include a number of detoxifying genes such as the glutathione *S*-transferases, glutathione reductase, and glutathione peroxidase (15). SCCs exhibit squamous epithelia features, and a number of genes related to keratinization, most obviously the keratins, are present in both clusters of genes. Also of note, bullous pemphigoid antigen 1 is a component of the hemidesmosome, to which keratin is anchored and, unsurprisingly, also is up-regulated here and in squamous carcinomas of the head and neck (16).

The fourth branch contains fresh AC, LCC, and normal tissue. Although four of the six LCCs cluster together, a specific group of genes that are clearly differentially regulated for these cancers could not be identified readily. More subtle variations in expression from large numbers of genes will have formed the LCC cluster. Alongside the six normal samples was a single AC tumor. This sample was in fact a bronchioalveolar carcinoma, a subclass of AC that is slower growing, is less well demarcated, and has large amounts of normal tissue. Many genes that regulate angiogenesis including ADAMTS1 (17), Tie2 (TEK), and VEGF-D (18) are overexpressed in normal tissue compared with cancers (Fig. 4B).

The AC samples grouped into three distinct clusters: AC group one primarily contains poorly differentiated carcinomas, whereas groups two and three contain much lower proportions of poorly differentiated samples. These clusters are defined by genes, for example TTF1, in common with AC clusters reported previously (13, 20), indicating they define common molecular pathologies. However, the small sample size precluded any correlation between these groupings and prognosis. Fig. 4C shows genes up-regulated in AC groups two and three. Carcinoembryonic antigen (CEA), a known serum marker for AC, was up-regulated in these AC samples. However, CEA is not expressed at high levels in many of the poorly differentiated AC samples found in AC group one or in some of the AC cell lines. Also up-regulated is TTF1, which regulates lung-specific gene activity and is overexpressed in AC (19).

Cell-Line Integration and Classification. Of the 38 cancer cell lines, 24 integrated with tumor samples, with many clustering according to type (Fig. 3). In general, the degree of conservation between cell lines and tumors was not as great as within the tumors themselves; genes that were overexpressed consistently in tumors were not necessarily overexpressed for the cell lines falling within the same branch (Fig. 4). However, this may be consistent with a poor level of differentiation for cell lines. As with LCC tumor samples, distinct clusters of genes that caused the cell lines to cluster could not always be identified. Although the fresh samples provided a frame for the classification of the cell-line samples, the cell lines also had a significant effect on tumor-sample clustering. For example, clustering only tumors using the same 4,253 genes still resulted in two distinct SCLC branches, but the sorting of SCLC samples between the branches was different to that shown in Fig. 3 (data not shown).

Those cell lines that clustered away from tumors formed two branches. The first branch only contains two samples, both simply classified as “non-small cell lung cancer” (NS), the second branch contains 10 cancer cell lines and all three embryonic fibroblasts. With the exception of two of these fibroblast cell lines, there is no evident clustering of the various pathological

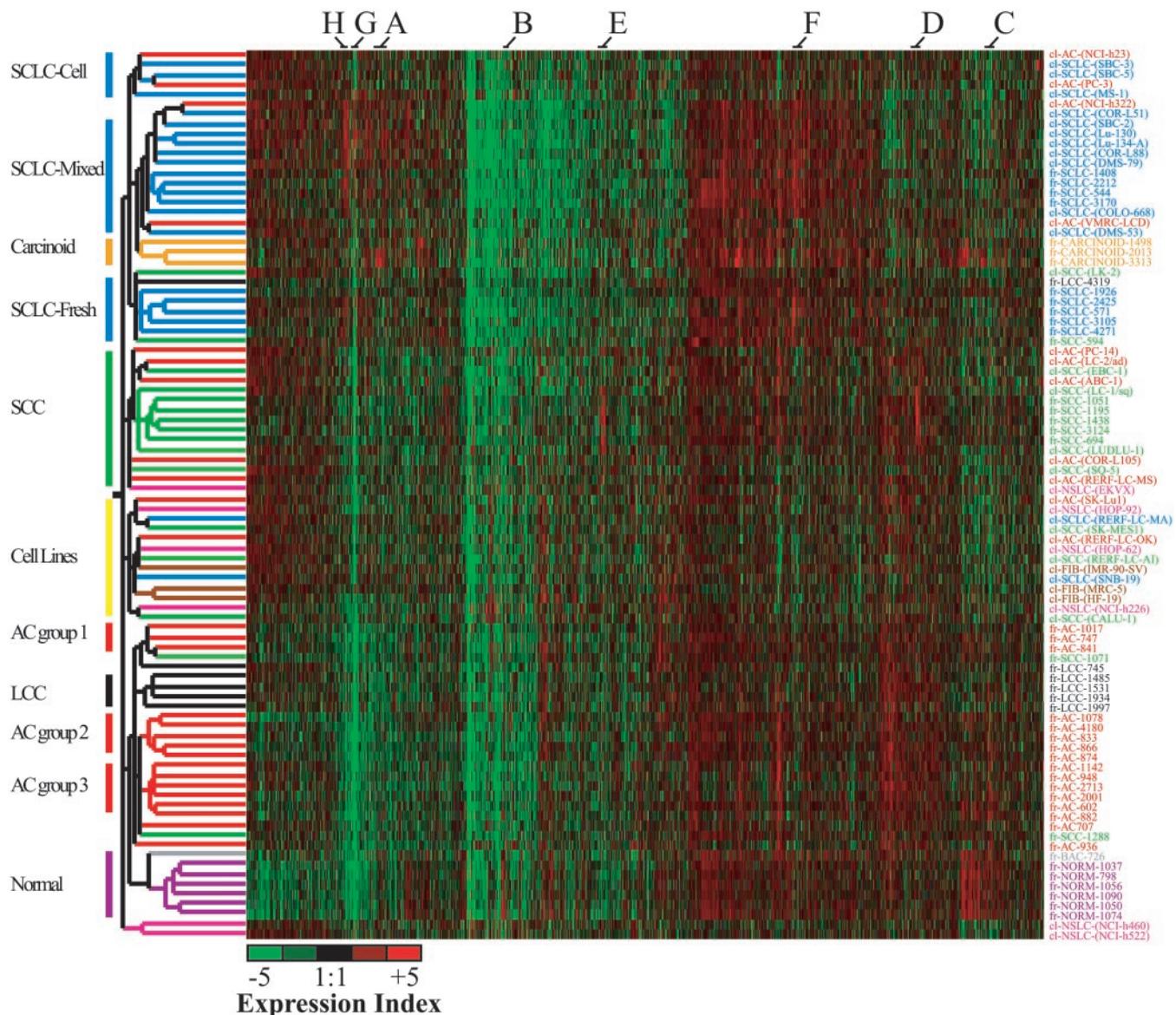


Fig. 3. Dendrogram of the reduced data set of 4,253 genes after filtering for commonly regulated genes in either fresh or cell-line samples. Samples are colored as described for Fig. 1. Groupings indicated on the left represent distinct clusters of particular carcinoma types. cl, cell line sample; fr, fresh tumor sample.

types within this cell-line branch. This branch also contains all four of the remaining NS cell lines. We had anticipated that our study might help better classify the six NS cell-line samples. However, in contrast to the other 32 cancer cell lines, 24 of which integrated with tumors, only one of these NS cell lines, EKVX, integrated. Given that AC cell lines also integrate with the SCC tumors (see below), it is difficult to improve on the original classification beyond saying that this cell line is likely to be derived from an SCC or AC.

Of the 13 SCLC cell lines that were analyzed, 2 are retained in the cell-line branch, with the remaining 11 all clustering with the SCLC tumors (Fig. 3). Of these, eight were in a subbranch with a subset of the SCLCs (SCLC-mixed), and the remaining three clustered in a separate subbranch (SCLC-cell) but still within the major SCLC branch. Many of the genes that were differentially regulated between the SCLC-mixed and SCLC-cell groups are down-regulated in the SCLC-cell group (Fig. 4G). These include insulinoma-associated 1, which has been used as a marker for SCLC (20).

From eight SCC cell lines analyzed, four integrated with the tumors; these four all clustered with the SCC group, although they were clustered less tightly than the tumors themselves. Some of the genes that were up-regulated in fresh SCCs were also up-regulated in these cell lines (Fig. 4D). However, as with the SCLC cell lines in the SCLC-cell group, many of these samples did not show such strong overexpression of genes that characterized the tumors (Fig. 4D and E).

Whereas SCLC and SCC cell lines had clustered with tumors consistently according to type, none of the 11 AC cell lines did so: 4 clustered with the SCLC group, 5 clustered with SCCs, and 2 were retained in the cell-line branch (Fig. 3). The overall level of similarity between AC cell lines and tumors was similar to that between other cell lines and tumors. For example, of the four AC cell lines that clustered with the SCLCs, two were found in the SCLC-cell cluster and two in the SCLC-mixed cluster. The branch lengths for these AC cell-line samples, a measure of the degree of identity, also were similar to the branch lengths for the SCLC cell lines. As we discuss below, these results have two

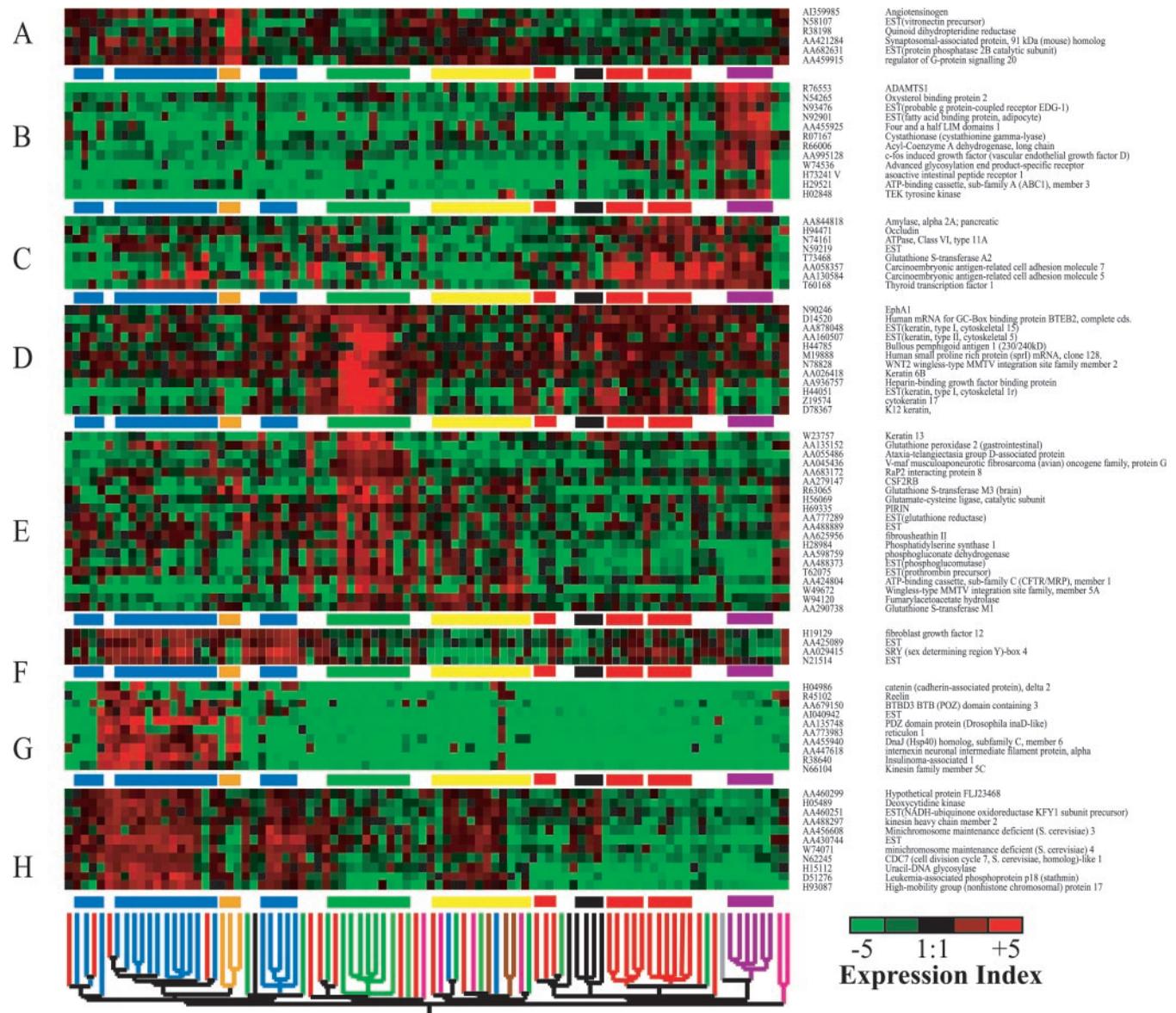


Fig. 4. Expression profiles across all samples of eight selected gene clusters from Fig. 3 showing distinct patterns of expression. The dendrogram and highlighted clusters from Fig. 3 are shown along the bottom and between individual gene clusters for reference. (A) Genes up-regulated in carcinoids. (B) Genes up-regulated in normal tissue. (C) Genes up-regulated in AC and LCC. (D and E) Genes up-regulated in SCC. (F–H) Genes up-regulated in SCLC.

possible explanations. They may indicate that subcomponents of SCLC or SCC within AC samples are selected when cell lines are established. Alternatively, they indicate that AC cells are established in culture but then progress toward a pathology resembling either SCLC or SCC.

Discussion

The utility of cancer cell lines depends on how closely they equate to the progenitor tumor. Here, we have sought to resolve this question by comparing the gene-expression profiles of a group of cell lines and fresh tumor samples.

Previous work in this area has been focused on other types of cancer. In a study of ovarian cancer (21), all four cancer cell lines analyzed clustered together with poorly differentiated cancer, primarily on the basis of proliferation rates. Similarly, breast cancer cell lines and tumors have been compared (22), but clustering resulted in isolated groups of cell lines and tumors.

Here, we were able to make comparisons after integrating cell lines with their fresh counterparts by using a filtering algorithm.

Our classification of fresh lung cancers closely parallels histopathological classification. Previous classifications of lung cancer (13, 20) placed various NSCLC and SCLC samples within the same major branch groups. Our initial clusterings essentially separated all fresh NSCLC from SCLC. The selection of tumors, the genes represented, the choice of reference sample, the filtering steps performed before clustering, and the clustering technique itself can account for some of these differences. In particular, the number of AC samples was larger in the earlier studies, and most AC samples that clustered near SCLC or SCC were poorly differentiated. We had fewer poorly differentiated AC samples in our study, which may explain why none of our fresh ACs clustered near SCC or SCLC.

Two distinct clusters of fresh SCLC were identified. Although no prognostic significance can be attached to these two groups because of small sample size and the almost invariably poor

prognosis for SCLC, the SCLC-mixed group had elevated expression levels of a large group of genes that may allow useful markers for the development of targeted treatment regimes.

One of the difficulties in classifying tumors by expression profiling is the presence of varying amounts of normal tissue. Although cancerous tissue can be isolated by microdissection (4), this is laborious. In our study, removing genes differentially regulated between fresh and cell-line samples allowed us to partially dissect away the influence of normal tissue, which may have helped improve the overall classification as well as allowed integration of cell lines with tumors.

Many genes that are known markers for various lung cancers also were identified here. These include carcinoembryonic antigens, a widely used marker of AC, and TTF1, which is commonly used to distinguish metastases of small-cell cancers from other organs to the lung and AC. We found that most of the AC cell lines have significantly reduced levels of both TTF1 and other markers of fresh AC, illustrating how far removed these AC cell lines are from their fresh counterparts.

Even after filtering to encourage clustering of cell lines with tumors, some cell lines formed isolated distinct groups. These cell lines may well have integrated with fresh samples if a different subset of genes had been selected before clustering. However, the same cell lines formed a separate subbranch within the cell-line group when 6,141 genes were used in the initial clustering. This indicates that this group of cell lines is significantly distinct from the cell lines that integrated with the tumors, and it questions their use as models for lung cancer.

The expression profiles of most SCLC and half the SCC cell lines resembled their fresh counterparts sufficiently to allow integration after clustering. As expected, the more easily established SCLC cell lines were most likely to resemble their tumor

counterparts. In contrast, although none of the 10 AC cell lines integrate with AC tumors, 5 resemble SCC and 3 resemble SCLC tumors. Because none of the AC tumors grouped with SCLC or SCC, this may imply that an SCC or SCLC component of heterogeneous AC tumors is selected and propagated when the cell lines are established. This may also occur *in situ* as AC progresses; both previous classifications of lung cancer (13, 20) clustered significant numbers of AC samples with SCC and SCLC. In both reports, these were predominantly poorly differentiated cancers. Indeed, in our study and the study of Bhattacharjee *et al.* (13), several AC tumors can be identified that express genes otherwise up-regulated only in SCC. Moreover, some AC tumors express markers of neuroendocrine differentiation seen in SCLC (20).

This change in expression patterns as AC tumors progress may reflect, alternatively, the progression of individual cells. Although there is no histopathological evidence that AC progresses to SCC or SCLC, this is true only at the time of surgery. At autopsy, some tumor cells become very anaplastic and impossible to categorize. Histopathological examination may reflect only a subset of gene expression, and thus there may be genes very important in terms of biological behavior not detected by histology. The SCLC cell lines examined here provide another line of evidence for the insufficiency of histology: Although lack of prominent nucleoli characterizes SCLC histologically, most of the cell lines used here, despite having prominent nucleoli, were classified correctly with the SCLC tumors.

Our observations suggest that AC cell lines either dedifferentiate toward molecular pathologies resembling SCLC or SCC, or clonal expansion of SCC or SCLC subcomponents of AC tumors occurs frequently. Analysis of larger numbers of AC samples taken at the time of surgery and autopsy will be required to verify that AC develops similarly *in situ*.

1. Statistics and Information Department, Minister's Secretariat (2002) *Ministry of Health, Labour, and Welfare Vital Statistics of Japan* (Ministry of Health, Labour, and Welfare, Tokyo).
2. Stancu, M. & Libbey, P. N. (2002) in *Cancer of the Lung: From Molecular Biology to Treatment Guidelines*, ed. Weitberg, A. B. (Humana, Totowa, NJ), pp. 35–80.
3. Hegde, P., Qi, R., Gaspard, R., Abernathy, K., Dharap, S., Earle-Hughes, J., Gay, C., Nwokekeh, N. U., Chen, T., Saeed, A. I., *et al.* (2001) *Cancer Res.* **61**, 7792–7797.
4. Luo, L., Salunga, R. C., Guo, H., Bittner, A., Joy, K. C., Galindo, J. E., Xiao, H., Rogers, K. E., Wan, J. S., Jackson, M. R., *et al.* (1999) *Nat. Med.* **5**, 117–122.
5. Hughes, T. R., Mao, M., Jones, A. R., Burchard, J., Marton, M. J., Shannon, K. W., Lefkowitz, S. M., Ziman, M., Schelter, J. M., Meyer, M. R., *et al.* (2001) *Nat. Biotechnol.* **19**, 342–347.
6. Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J. & Speed, T. P. (2002) *Nucleic Acids Res.* **30**, e15.
7. Tseng, G. C., Oh, M. K., Rohlin, L., Liao, J. C. & Wong, W. H. (2001) *Nucleic Acids Res.* **29**, 2549–2557.
8. McCracken, S., Kim, C. S., Xu, Y., Minden, M. & Miyamoto, N. G. (1999) *Oncogene* **15**, 2929–2937.
9. Roebroek, A. J., van de Velde, H. J., Van Bokhoven, A., Broers, J. L., Ramaekers, F. C. & Van de Ven, W. J. (1993) *J. Biol. Chem.* **268**, 13439–13447.
10. Wang, Q., Lu, J., Yang, C., Wang, X., Cheng, L., Hu, G., Sun, Y., Zhang, X., Wu, M. & Liu, Z. (2001) *Cancer Lett. (Shannon, Irel.)* **179**, 71–77.
11. Kanai, Y., Okada, Y., Tanaka, Y., Harada, A., Terada, S. & Hirokawa, N. (2000) *J. Neurosci.* **20**, 6374–6384.
12. Charbaut, E., Curmi, P. A., Ozon, S., Lachkar, S., Redeker, V. & Sobel, A. (2001) *J. Biol. Chem.* **276**, 16146–16154.
13. Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98**, 13790–13795.
14. Felding-Habermann, B. & Cheresch, D. A. (1993) *Curr. Opin. Cell. Biol.* **5**, 864–868.
15. 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.
16. Herold-Mende, C., Kartenbeck, J., Tomakidi, P. & Bosch, F. X. (2001) *Cell. Tissue Res.* **306**, 399–408.
17. Vazquez, F., Hastings, G., Ortega, M. A., Lane, T. F., Oikemus, S., Lombardo, M. & Iruela-Arispe, M. L. (1999) *J. Biol. Chem.* **274**, 23349–23357.
18. Niki, T., Iba, S., Tokunou, M., Yamada, T., Matsuno, Y. & Hirohashi, S. (2000) *Clin. Cancer Res.* **6**, 2431–2439.
19. Pelosi, G., Frassetta, F., Pasini, F., Maisonneuve, P., Sonzogni, A., Iannucci, A., Terzi, A., Bresaola, E., Valduga, F., Lupo, C., *et al.* (2001) *Am. J. Surg. Pathol.* **25**, 363–372.
20. Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G. D., Perou, C. M., Whyte, R. L., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98**, 13784–13789.
21. Welsh, J. B., Zarrinkar, P. P., Sapinoso, L. M., Kern, S. G., Behling, C. A., Monk, B. J., Lockhart, D. J., Burger, R. A. & Hampton, G. M. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 1176–1181.
22. Ross, D. T. & Perou, C. M. (2001) *Dis. Markers* **17**, 99–109.