Feasibility of Using Low-Volume Tissue Samples for Gene Expression Profiling of Advanced Non-Small Cell Lung Cancers

Elaine H. Lim,[1] Amit Aggarwal,[2]
Thirugananam Agasthian,[2] Poo-Sing Wong,[1]
Christie Tan,[1] Eugene Sim,[1] Lenny Tan,[1]
Poh-Sun Goh,[1] Shih-Chang Wang,[1]
Kay-Leong Khoo,[1] Amartya Mukherjee,[1]
See-Meng Khoo,[1] Gerald Chua,[1] Barbro Nilsson,[1]
Kang-Hoe Lee,[1] and Patrick Tan[2]

[1]National University Hospital and [2]National Cancer Centre of Singapore, Singapore

ABSTRACT

Purpose: The majority of patients with non-small cell lung cancer (NSCLC) present at an advanced clinical stage, when surgery is not a recommended therapeutic option. In such cases, tissues for molecular research are usually limited to the low-volume samples obtained at the time of diagnosis, usually via fine-needle aspiration (FNA). We tested the feasibility of performing gene expression profiling of advanced NSCLCs using amplified RNA from lung FNAs.

Experimental Design and Results: A total of 46 FNAs was tested, of which 18 yielded RNA of sufficient quality for microarray analysis. Expression profiles of these 18 samples were compared with profiles of 17 pairs of tumor and normal lung tissues that had been surgically obtained. Using a variety of unsupervised and supervised analytical approaches, we found that the FNA profiles were highly distinct from the normal samples and similar to the tumor profiles.

Conclusions: We conclude that when RNA amplification is successful, gene expression profiles from NSCLC FNAs can determine malignancy and suggest that with additional refinement and standardization of sample collection and RNA amplification protocols, it will be possible to conduct additional and more detailed molecular analysis of advanced NSCLC using lung FNAs.

INTRODUCTION

Lung cancer is a major cause of cancer mortality, accounting for ~20% of cancer deaths worldwide (1). Survival statistics are dismal with an average 5-year survival of 14% in the United States and <10% in Europe, India, China, and the developing countries. Lung cancers are traditionally subdivided by histology into SCLC[3] and NSCLC. NSCLC is the more common variant (~85% of lung cancers) and, unlike SCLC, is less sensitive to chemotherapeutic agents (response rate 20 *versus* 70%; Ref. 2, 3). Thus, it is crucially important to develop better diagnostic and therapeutic strategies for the management of NSCLC.

In recent years, there has been an explosion in the application of gene expression profiling to study various tumor types, which has provided valuable insights into the pathways of cancer development and progression (4). One important clinical aspect of this technology lies in the identification of novel molecular markers for disease detection, prognostication, and treatment selection (5, 6). In the case of lung cancer, several groups have recently reported microarray gene expression analysis of lung cancers (7–12). One potential limitation of these previous studies, however, has been their reliance on surgical specimens because these large-volume tissue samples typically yield sufficient RNA for microarray analyses. As a result, these studies have primarily focused on early-stage NSCLCs, when surgical resection is the treatment of choice. Unfortunately, many NSCLC cases present at a late clinical stage (stage IV), when surgery is not recommended, and chemotherapy is commonly undertaken on a palliative basis (2). For these late-stage NSCLCs, tissue samples for analysis are typically limited to low-volume samples such as FNAs guided by endoscopy, CT, or fluoroscopy, and the RNA extracted from these FNA samples is usually insufficient for gene expression profiling. Because of this limitation, the molecular exploration of late stage NSCLCs has, at present, been comparatively underaddressed.

Recently, a high-fidelity RNA amplification protocol has been previously described (13) that has allowed analyzable gene expression profiles to be obtained from FNAs of melanomas (14) and breast cancers (15). Compared with the skin and breast, the lung represents a more challenging organ with regards to accessibility because of its greater risk of procedure-related complications (see "Discussion"). In this article, we examined the feasibility of using a similar amplification procedure on lung FNA samples for gene expression profiling of advanced

---

[3] The abbreviations used are: SCLC, small cell lung cancer; NSCLC, non-small cell lung cancer; FNA, fine-needle aspiration; LOOCV, leave-one-out cross validation; CT, computed tomography; SVM, support vector machine.

NSCLC. We successfully generated gene expression profiles for a series of surgical and image-guided FNAs and compared the FNA profile data to that obtained from surgical specimens. Using a variety of unsupervised and supervised analytical approaches, we found that the FNA profiles were highly distinct from the normal samples and similar to the tumor profiles. We conclude that when RNA amplification is successful, gene expression profiles from lung FNAs can determine malignancy and suggest that upon refinement of sample collection and RNA amplification, it will be possible to conduct additional and more detailed molecular profiling of advanced NSCLC using lung FNAs.

MATERIALS AND METHODS

Patients and Sample Collection. Approvals for this study were obtained from the Institutional Review Board of the National University Hospital, Singapore, and samples were obtained from patients with informed consent. For patients undergoing surgery, a sample of tumor tissue (1 cm^3), a sample of adjacent normal lung (1 cm^3), and a FNA of the tumor using a 23-gauge needle (Becton Dickinson, Singapore) were obtained from each patient. Endoscopic FNAs were obtained with a 22-gauge Wang cytology needle (Bard Endoscopic Technologies, Billerica, MA), whereas CT- or fluoroscopic-guided FNAs were obtained with a 22-gauge Chiba needle (Boston Scientific, Natick, MA) or 21-gauge Sonopsy C1 needle (Hakko Medical, Nagano, Japan). NSCLCs of various histological subtypes were included, *i.e.,* adenocarcinoma, squamous cell carcinoma, and large cell carcinoma (Supplementary Information Table A). All samples obtained were from patients with lung primaries, *i.e.,* no known primary malignancy elsewhere. Each aspirate was collected in 80 μl of RNAlater (Ambion, Austin, TX), whereas each surgical sample was collected in 1 ml of RNAlater. All samples were stored at −80°C before processing.

Sample Processing and Microarray Hybridization. Total RNA was extracted from all samples using RNeasy kits (Qiagen, Valencia, CA) and subjected to linear amplification as described by Wang *et al.* (13). The reproducibility of the RNA amplification process was confirmed by subjecting various surgical samples to independent replicate amplifications, *i.e.,* independent amplifications of the same starting material (see Supplementary Information Table D and "Discussion"). Cy3 and Cy5 fluorescently labeled cDNAs were prepared from amplified RNA and hybridized to an 18,000 element human cDNA microarray (clones obtained from Incyte, Palo Alto, CA, and Research Genetics, Carlsbad, CA) printed using an OmniGrid arrayer (GeneMachines, San Carlos, CA). Hybridizations were performed as indirect comparisons (*i.e.,* sample *versus* reference) using commercially available reference RNA (Universal Human Reference, Stratagene, La Jolla, CA).

Data Acquisition and Preprocessing. Raw scans of individual microarrays were acquired using a 10-μm resolution GenePix 4000 scanner (Axon Instruments, Union City, CA). Fluorescence data corresponding to each array element was obtained using GenePix 4.0 analysis software and uploaded into a centralized Oracle 8i database, which is accessible via a user interface on line.[4] Individual array targets that were well measured across 85% of all arrays were selected, corresponding to spots exhibiting a foreground to background ratio of >2 for at least one of the two wavelengths. This dataset, comprising 12,329 array elements, was then internally normalized by median centering each sample (array) and is available on line.[5]

Identification of Differentially Expressed Genes, Class Prediction, and Other Data Analysis. A combination of two-sample *t* tests and fold change ratios (1.5- and 2-fold) were used to derive gene sets for discriminating between normal and tumor (surgical) samples at high confidence. These gene sets are downloadable (Supplementary Information Table B). Supervised class predictions were performed using SVMs (16, 17), a classification algorithm that is capable of handling sparse data and which has been previously used in several microarray analyses for cancer class prediction (18) and in the functional classification of genes (19). The SVM separates a given set of binary-labeled training data (normal *versus* tumor in our case) with a hyperplane that is maximally distant from the two classes. The hyperplane can then be used to predict the classes of unknown samples. In this study, the SVM was trained to segregate normal tissues from tumors based on the gene set selected by the fold change/*t* test analysis. The system, having learned from the expression features of normal and tumor classes, was then used to classify samples in the blind set, consisting of FNAs and surgical samples from patients A and B. Classification accuracies of the various gene sets were assessed using LOOCV or independent testing. In LOOCV, for each discriminator gene set, each sample in the training set was left out once and a maximum margin hyperplane constructed using the remaining samples as training samples. The sample left out was then used as the test case and its class predicted using the output of the decision function. This was then repeated for all samples in the training set. The samples whose distances from the hyperplane fell between +0.1 and −0.1 were considered no-call cases. Independent testing was performed on the FNAs and surgical samples from patients A and B. Average linkage hierarchical clustering (20) and principal component analysis (21–23) using GeneData Analyst version 4 software (GeneData, Basel, Switzerland) were performed on all 52 samples to assess sample similarities and to visualize the variance across samples.

RESULTS

Gene Expression Profiles Can Be Successfully Obtained from Lung FNAs

We collected a total of 89 lung samples, corresponding to 17 tumor samples (surgical), 17 normal samples (surgical), 17 FNAs corresponding to the surgical tumor samples, and 38 image-guided FNAs. Nine image-guided FNAs were excluded because of these patients exhibiting non-NSCLC malignancies (SCLC and one case of carcinoid). When the remaining 80 samples were processed, we found that significant RNA degradation had occurred in 7 surgical FNAs and 21 image-guided

---

[4] Internet address: http://www.omniarray.com.
[5] Internet address: http://www.omniarray.com/lungFNA.html.

FNAs, rendering them unsuitable for additional analysis. In summary, 52 of 89 samples, corresponding to 17 tumor and normal paired samples (surgical) and 18 FNAs (10 surgical, 8 image-guided), were additionally analyzed (Supplementary Information Table A). We amplified total RNA from the 18 FNAs and obtained gene expression profiles for all 18, thus indicating a success rate for FNAs of ~39% (*i.e.,* 18 of 46 FNAs). This figure is comparable with the published article from Wang *et al.* (14), suggesting that in cases where RNA isolation is successful, the linear amplification protocol described in Ref. 13 can be performed on these samples to create an analyzable gene expression profile.

Identification of a Gene Set to Discriminate between Normal and Tumor Lung Samples

To determine whether FNA gene expression profiles can be used to determine malignancy, we first identified genes that were differentially expressed between malignant and nonmalignant tissues at high confidence. Of the 17 patients from whom surgical samples had been obtained, one (A) had received preoperative anticancer treatment, whereas another patient's (B) resected tumor was contaminated with a large proportion of pericardium. To avoid potential confounding factors, we excluded these samples (patients A and B) from the initial analysis, focusing on the remaining 15 pairs of tumor and normal surgical samples. First, a two-sample *t* test was performed to identify genes in which the average expression levels were significantly different between tumor and normal samples at various levels of confidence ($P < 0.05$, $P < 0.01$, and $P < 0.001$). Second, an intergroup median comparison was performed to identify genes varying by at least 1.5- or 2-fold between the tumor and normal groups. In total, 6 gene sets comprising 656, 449, and 257 (1.5-fold, from $P < 0.05$, $P < 0.01$, and $P < 0.001$) and 133, 115, and 92 genes (2-fold, from $P < 0.05$, $P < 0.01$ and $P < 0.001$) were identified (Table 1). These gene sets were then compared with a random perturbation assay whereby samples of normal and tumor chunks were randomly selected to form two fictitious groups, each group comprising 45–55% of tumor (or normal) samples. From a total of 300 randomly generated groupings, the numbers of genes regulated by >1.5- and >2-fold at three different Ps ($P < 0.05$, $P < 0.01$, and $P < 0.001$) based on *t* tests between the two fictitious groups were calculated (Table 1). In all six cases, the genes identified as differentially expressed in the *bona fide*

tumor *versus* normal comparison strongly exceeded what one would expect on the basis of chance alone, suggesting that the genes in the various gene sets are reflective of a true biological distinction (*i.e.,* tumor *versus* normal; Table 1).

Genes that were differentially expressed between tumors and normals could be broadly classified into a number of functional groups such as cell signaling, cell cycle regulation, apoptosis, cell adhesion, angiogenesis, immune system, cell trafficking, cytoskeletal components, enzymes in cellular metabolism, transcription, translation, and unknown function. Consistent with other studies (12), significantly more genes were downregulated in tumors than up-regulated (~85% in 5 gene sets, 70% in 656-gene set), which may be a reflection of tumor heterogeneity as compared with normal lung tissue. Table 2 lists selected examples from the 656-gene set (complete list in Supplementary Information Table B), and we briefly mention a few:

Lung Differentiation. Giordano *et al.* (9) compared gene expression profiles of lung, colon, and ovarian cancers and found overexpression in the lung tumors of pulmonary-associated surfactant, SFTPA1, and thyroid transcription factor, TITF1, which is implicated in surfactant gene expression (24). When compared with normal lung tissue, however, we found that our tumors exhibited features suggestive of lung dedifferentiation because they exhibited down-regulation of SFTPA1, TITF1, and pronapsin A, an aspartate protease involved in proteolytic processing of surfactant precursors (25). This may indicate the aggressive nature of our tumors, consistent with Garber *et al.* (7) who showed that lung adenocarcinomas with down-regulation of pulmonary-specific genes exhibited a worse clinical outcome compared with other lung adenocarcinomas in which these genes were highly expressed. We also found downregulation of forkhead box F1, a transcription factor implicated in lung differentiation in mice (26). The clinical outcomes of these patients are being closely followed.

Cell Cycle and Apoptosis. Our tumors showed downregulation of several cell signaling factors implicated in cell cycle regulation and apoptosis pathways, *e.g.,* protein kinase C, protein phosphatases, the p21 cell cycle inhibitor (27), ARF (28), gravin (29), dual specificity phosphatase 1 (30, 31), and up-regulation of apoptosis inhibitors, *e.g.,* tumor necrosis factor receptor-associated factor 1 and tumor necrosis factor receptor-associated factor interacting protein (32).

Angiogenesis. Vascular endothelial growth factor, a target for antiangiogenic cancer therapy, is not always overex-

Table 1 Significant genes (between normal and tumor surgical samples) at various levels of fold change and *P*, compared with genes called significant under similar conditions in a random labeling assay

The median value of significant genes is shown from 300 randomly generated groups. The values in brackets represent the mean and standard deviation (rounded to the nearest integer), respectively.

Case	*P*	Fold change (median)	Down-regulated in tumors	Up-regulated in tumors	Total significant	Random assay significant
1	0.05	1.5	453	203	656	18 (28, 27)
2	0.01	1.5	355	94	449	5 (8, 12)
3	0.001	1.5	222	35	257	0 (1, 3)
4	0.05	2	118	15	133	1 (2, 3)
5	0.01	2	100	15	115	0 (1, 3)
6	0.001	2	80	12	92	0 (0, 0)

Table 2 Selected examples from the tumor-normal differential gene set

↓, down-regulated in tumors; ↑, up-regulated in tumors.

Gene Name	Description	Regulation
Lung specific		
SFTPA1	Surfactant, pulmonary-associated protein A1	↓
TITF1	Thyroid transcription factor 1	↓
NAP1	Pronapsin A	↓
Cell adhesion/extracellular matrix		
CDH5	Cadherin 5	↓
FGFR1	Fibroblast growth factor receptor 1	↓
ITGA3/5	Integrin α 3/5	↓
LAMA3	Laminin α 3	↓
DSG2	Desmoglein 2	↓
ITM2A/2B	Integral membrane protein 2A/2B	↓
ICAM2	Intercellular adhesion molecule 2	↓
MATN2	Matrilin 2	↓
TIMP3	Tissue inhibitor of metalloproteinase 3	↓
Metastatic/aggressive phenotype		
SPP1	Secreted phosphoprotein 1 (osteopontin)	↑
MLN51	Metastatic lymph node 51	↑
ERO1L	ERO1-like	↑
Cell cycle regulation/apoptosis		
TRIP	TRAF interacting protein	↑
TRAF	TNF receptor-associated factor 1	↑
CDKN1A	p21, cyclin-dependent kinase inhibitor 1A	↓
PAWR	Protein kinase C, apoptosis WT1 regulator	↑
PRKCM	Protein kinase C, mu	↑
AKAP12	Gravin	↑
DUSP1	Dual specificity phosphatase 1	↑
PPP2CB/PPP2R1B/PPPR5B/PTPN12/PTPRB	Protein phosphatase 2 (catalytic subunit, regulatory subunits A/B), protein tyrosine phosphatase 12/B	↑
Angiogenesis		
VEGF	Vascular endothelial growth factor	↑
CYR61	Cysteine-rich angiogenic inducer 61	↑

pressed in NSCLC (33–35). Explanations for this include the high vascularity of normal lung, therefore nullifying the need for new blood vessels for continued tumor growth. We found that vascular endothelial growth factor was down-regulated in our study. Another angiogenic factor, CYR61, implicated in carcinogenesis (36), was similarly down-regulated in our tumors.

Cell Adhesion. A number of cell adhesion molecules and matrix proteins (*e.g.,* cadherin 5, intercellular adhesion molecule 2, integrin 3, integrin 5, desmoglein 2, fibroblast growth factor receptor 1, laminin, and matrilin 2) were generally down-regulated in our tumors, as well as tissue inhibitor of metalloproteinase 3 (37). This is likely to reflect tumor aggression and invasive/metastatic potential. Supporting this idea, osteopontin (38), which is associated with the metastatic phenotype, was up-regulated in our tumors. MLN51 (39) was also up-regulated, a gene previously isolated from differential screening of a human breast cancer metastasis cDNA library. Finally, ERO1-like, involved in oxidative protein folding in the endoplasmic reticulum (40), was strongly expressed in the poor prognosis group of lung adenocarcinomas mentioned above (7) and was also up-regulated in our tumors.

Classification of FNA Expression Profiles by a Supervised Learning Methodology

Using the 30 surgical samples (15 tumor and 15 normal) as a training set, we then trained an SVM classification algorithm to discriminate tumors from normals based upon the gene sets

defined by the *t* test/fold change assay in the previous section. Classification accuracy was assessed using LOOCV, and the results are shown in Table 3. Across the various tumor/normal discriminator gene sets, two cases, HU02151 (normal) and HU02164 (tumor), were consistently no-called or misclassified, yielding a training classification accuracy of 93.4%. As HU02164 was sampled from the edge of the tumor and HU02151 was sampled from a resected lung specimen containing a very large tumor, it is possible that these samples were microscopically contaminated with normal and tumor elements, respectively, accounting for the frequent no-calls and misclassifications. We then proceeded to classify a series of independent samples, which had been isolated and thus blinded from the SVM during the training process.

Table 3 LOOCV results for the normal and tumor data set using various sets of differential genes from Table 1. The values in brackets in the column "Gene set" represent the median fold change and *P* from *t* test, respectively. Sample identities of the no-calls and misclassifications are shown in brackets.

Case	Gene set	No-calls	Misclassifications
1	656	1 (HU02151)	1 (HU02164)
2	449	0	2 (HU02151, HU02164)
3	257	0	2 (HU02151, HU02164)
4	133	1 (HU02151)	1 (HU02164)
5	115	1 (HU02151)	1 (HU02164)
6	92	1 (HU02151)	1 (HU02164)

Table 4 Linear SVM classification errors using the 18 FNAs as a blind data set with all 6 gene sets. The values in brackets in the column "Gene set" represent the median fold change and *P* from *t* test, respectively. Sample identities of the no-calls and misclassifications are shown in brackets.

Case	Gene set	No-calls	Misclassifications	Accuracy
1	656	2 (HU02181, HU02158)	3 (HU02182, HU02183, HU02176)	13/18
2	449	4 (HU02181, HU02182, HU02183, HU02176)	0	14/18
3	257	1 (HU02182)	1 (HU02183)	16/18
4	133	0	1 (HU02176)	17/18
5	115	1 (HU02176)	0	17/18
6	92	0	0	18/18

First, the tumor and normal surgical samples from patient A (preoperatively treated) and patient B (tumor contaminated with pericardium) were classified, patient A's tumor and normal samples were classified as tumor and normal, respectively, whereas patient B's tumor and normal samples were both classified as normal. This result was consistent in all six cases using different gene sets (data not shown).

Second, the 18 FNA tumor profiles were classified. The results of the misclassifications and no-calls are given in Table 4 along with the sample identities. The classification accuracy varied from 72% (13 of 18) to 100% (18 of 18) depending on the gene set used. The best accuracy (100%) was obtained using the gene set obtained under the most stringent selection criteria (2-fold, $P < 0.001$). In general, we observed that 2 of 4 (50%) CT- or fluoroscopic-guided FNAs were frequently no-calls or misclassified, whereas this was the case in 1 of 4 (25%) endoscopic-guided FNAs. This could be related to the nature of cells obtained because the needle route in CT- or fluoroscopic-guided FNAs is percutaneous. Of the surgical FNAs, 2 of 10 (20%) were no-calls or misclassified, both are from the two special cases, patients A and B. Patient A's FNA (HU02158) was classified with the tumors, except when using the 656-gene set where it was a no-call. Patient B's FNA (HU02176) was misclassified using two of the 6 gene sets and a no-call using the 449-gene set. This result suggests that with the appropriate gene set, the majority of lung FNA tumor profiles can be correctly classified as malignant, in a similar fashion to surgical samples.

Similarity Assessment of Surgical FNAs with Parent Tumors

The lung samples in this report were obtained by several different clinicians (4 cardiothoracic surgeons, 6 respiratory physicians, and 4 interventional radiologists), all of whom are likely to vary to some degree in their procedural technique and expertise. To better visualize potential similarities and differences among these samples, average hierarchical cluster analysis and principal component analysis) were performed. In the unsupervised clustering analysis, the normal samples formed a tight, highly correlated subgroup with one to two tumor samples clustering with them. Figs. 1 and 2 depict the results based upon the 257-gene set from Table 1 ($P < 0.001$, 1.5-fold change), with similar results obtained for other gene sets (data not shown). In addition, the principal component analysis and hierarchical cluster analysis revealed that FNA tumor profiles mostly clustered with the tumors, distinct from the normal lung samples. As a more stringent comparison, we then compared the similarity of the FNA tumor profiles obtained from the surgical

specimens to the profiles of their parent tumors because in the ideal setting, one might expect that the expression profiles of the two would be extremely highly correlated. Pearson's correlation coefficient was calculated for the 10 surgical FNAs and 17 tumor chunks, using the global normalized dataset of 12,329 genes. As a negative control, the 17 normal profiles were added in the analysis. We found that 3 of the 10 FNAs showed the highest correlation with their parent tumors, whereas the remaining 7 did not (Supplementary Information Table C). These results indicate that although the FNA tumor profiles do resemble surgical tumor profiles (as shown by the supervised analysis), the former does contain features that render them distinct from the latter, if compared on a global scale.

DISCUSSION

The successful management and treatment of NSCLC remains one of the key challenges in oncology today. Although early-stage NSCLCs can be treated surgically, most NSCLC cases present at an advanced stage, when surgical resection is not a recommended therapeutic option. The optimal management of locally advanced disease (stage III) is controversial, often involving a combination of chemotherapy, radiotherapy, with or without surgery (41, 42). In this study, we assessed the feasibility of generating gene expression profiles from lung FNAs because molecular data from late-stage NSCLCs may be invaluable for addressing important clinical questions. Unlike breast and skin tumors, the lung is a relatively difficult organ to investigate, primarily because of the risks involved in obtaining tissue such as the induction of a life-threatening pneumothorax and procedure-related hypoxia, as well as requiring a greater level of patient cooperation, *e.g.,* breath-holding and the ability to tolerate endoscopy. In contrast with melanoma and breast tumors, where needle size and number of needle passes used are of little consequence in terms of medical risks because of their anatomical sites, the risk of creating a pneumothorax from CT- or fluoroscopic-guided lung FNAs is proportional to these factors. Hence, the FNA samples used in this study were often the remains of a single pass, and the larger Sonopsy needle was used only for peripherally located tumors. Of note, there were no procedure-related deaths in our study.

Given these challenges, the results presented in this article represent what we were able to achieve at a practical level. Among the FNAs we collected, there was a high incidence of RNA degradation: 80% of CT- or fluoroscopic-guided FNAs, 50% of endoscopic-guided FNAs, and 40% of surgical FNAs. Contributing factors would include operator-dependent technique in procuring tissue (several different clinicians were in-

Fig. 1 Average linkage hierarchical clustering of 52 samples (using 257 gene set from Table 1). Positive correlation was used as the similarity metric.

volved in our study), RNA processing technique (time to freezer), number of cells obtained (remains of a single pass *versus* a fresh pass), and contamination with blood affecting the quality of RNA (the lung is a highly vascular organ). The failure rates for the CT- or fluoroscopic-guided FNAs were the highest and may be because these are biopsies obtained precutaneously, *i.e.,* traversing through the skin, s.c. tissue, normal lung tissue, before hitting the tumor. In contrast, in surgical and endoscopic-guided FNAs, the needle is directly inserted into the tumor. As a comparison, the melanoma FNA study (14) reported a failure rate of ~90% because of RNA quality and availability of clinical outcome data, whereas the breast FNA study (15) was ~15%. Our success rate lies in between these two studies.

Although the majority of samples in our study were classified correctly by the molecular data, there were a few exceptions. For example, patient A's histology report showed no evidence of malignancy, but the surgical tumor sample and FNA corresponding to this patient were classified as malignant by the various classifier gene sets. It remains to be seen whether this molecular assessment of tumor response to preoperative treat-

ment is clinically significant, and this patient's clinical progress will be followed closely. Patient B's histology report showed malignancy with a significant amount of pericardial tissue, and the surgical tumor sample was classified as normal in our study. The FNA was classed as normal using 3 of the 6 gene sets. These results most likely reflect the presence of contaminating normal tissue in the tumor sample, which can affect the resultant gene expression profile.

Three of the 10 surgical FNAs exhibited the highest correlation to their parent tumor but the remaining 7 did not. This might possibly reflect variations incurred during the processes of sample collection or RNA processing. We note, however, that the FNA samples exhibiting highest correlation to their parent tumors were obtained at the end of the 7-month sample collection period, when sample collection protocols became more standardized. In addition, we (Supplementary Information Table D) and others (14, 15) have also found the RNA amplification protocol (13) to be highly consistent in generating reproducible expression profiles. Thus, we currently lean toward the hypothesis that variations in sample collection, rather than RNA am-
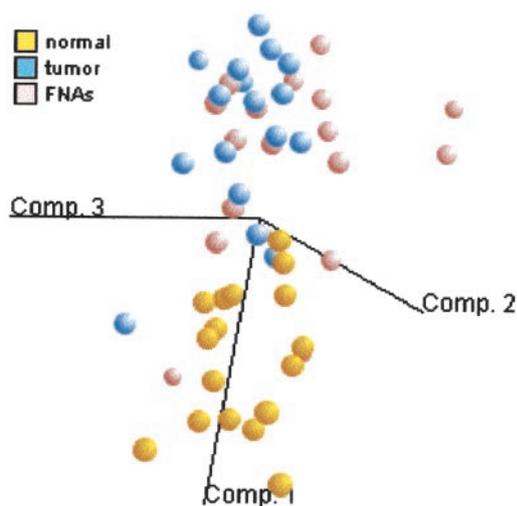
Fig. 2 Principal component analysis to visualize the variance of all of the samples. The 257-gene set (from Table 1) was used to generate the above figure.

plification, represent the major contributing factor for the overall low correlation between the FNAs and their parent surgical tumors. However, further optimization of both is being pursued.

In conclusion, performing molecular genetic analysis on advanced NSCLC cases has always been historically difficult, primarily because of the limited amount of tissue available. We believe that our results indicate that within the daily clinical constraints and variables associated with a busy clinical environment, it is nevertheless possible to use lung FNAs obtained at the time of diagnosis to generate gene expression profiles that determine malignancy. It will be important to optimize the procedures described here so that these profiles can ultimately be used to impact the clinical management of advanced NSCLC patients. Data on correlations of these profiles with stage of disease, histological type, smoking status, ethnic group, and gender is not presented here but is being collected. Our local population comprises mainly Chinese individuals, and we have particular interest in Chinese nonsmoking women with lung adenocarcinoma (43). Epidemiological studies have shown smoking rates of 16–52% in Chinese female lung cancers in Singapore, China, San Francisco, and Hawaii in contrast with 77–90% in Caucasian women in North America and the United Kingdom. It would be fascinating to see if a specific gene expression profile characterizes this patient group.

ACKNOWLEDGMENTS

REFERENCES

1. Parkin, D. M., Pisani, P., and Ferlay, J. Global cancer statistics. CA - Cancer J. Clin., 49: 33–64, 1999.

2. Schiller, J. H., Harrington, D., Belani, C. P., Langer, C., Sandler, A., Krook, J., Zhu, J., and Johnson, D. H. Comparison of four chemotherapy regimens for advanced non-small cell lung cancer. N. Engl. J. Med., 346: 92–98, 2002.

3. Fukuoka, M., Furuse, K., Saijo, N., Nishiwaki, Y., Ikegami, H., Tamura, T., Shimoyama, M., and Suemasu, K. Randomized trial of cyclophosphamide, doxorubicin, and vincristine *versus* cisplatin and etoposide *versus* alternation of these regimens in small-cell lung cancer. J. Natl. Cancer Inst. (Bethesda), 83: 855–861, 1991.

4. DeRisi, J., Penland, L., and Brown, P. O. (Group 1), Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su Y. A., and Trent, J. M. (Group 2). Use of a cDNA microarray to analyze gene expression patterns in human cancer. Nat. Genet., 14: 457–460, 1996.

5. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, W. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature (Lond.), 403: 503–511, 2000.

6. van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., and Bernards, R. A gene expression signature as a predictor of survival in breast cancer. N. Engl. J. Med., 347: 1999–2009, 2002.

7. 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. I., Altman, R. B., Brown, P. O., Botstein, D., and Petersen, I. Diversity of gene expression in adenocarcinoma of the lung. Proc. Natl. Acad. Sci. USA, 98: 13784–13789, 2001.

8. Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., and Meyerson, M. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc. Natl. Acad. Sci. USA, 98: 13790–13795, 2001.

9. Giordano, T. J., Shedden, K. A., Schwartz, D. R., Kuick, R., Taylor, J. M. G., Lee, N., Misek, D. E., Greenson, J. K., Kardia, S. L. R., Beer, D. G., Rennert, G., Cho, K. R., Gruber, S. B., Fearon, E. R., and Hanash, S. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. Am. J. Pathol., 159: 1231–1238, 2001.

10. Nacht, M., Dracheva, T., Gao, Y., Fuji, T., Chen, Y., Player, A., Akmaev, V., Cook, B., Dufault, M., Zhang, M., Zhang, W., Guo, M., Curran, J., Han, S., Sidransky, D., Buetow, K., Madden, S. L., and Jen, J. Molecular characteristics of non-small cell lung cancer. Proc. Natl. Acad. Sci. USA, 98: 15203–15208, 2001.

11. Beer, D. G., Kardia, S. L. R., Huang, C., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G., Lizyness, M. L., Kuick, R., Hayasaka, S., Taylor, J. M. G., Iannettoni, M. D., Orringer, M. B., and Hanash, S. Gene expression profiles predict survival of patients with lung adenocarcinoma. Nat. Med., 8: 816–824, 2002.

12. Heighway, J., Knapp, T., Boyce, L., Brennand, S., Field, J. K., Betticher, D. C., Ratschiller, D., Gugger, M., Donovan, M., Lasek, A., and Rickert, P. Expression profiling of primary non-small cell lung cancer for target identification. Oncogene, 21: 7749–7763, 2002.

13. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T., and Marincola, F. M. High-fidelity mRNA amplification for gene profiling. Nat. Biotech., 18: 457–459, 2000.

14. Wang, E., Miller, L. D., Ohnmacht, G. A., Mocellin, S., Perez-Diez, A., Petersen, D., Zhao, Y., Simon, R., Powell, J. I., Asaki, E., Alexander, H. R., Duray, P. H., Herlyn, M., Restifo, N. P., Liu, E. T., Rosenberg, S. A., and Marincola, F. M. Prospective molecular profiling of melanoma metastases suggests classifiers of immune responsiveness. Cancer Res., 62: 3581–3586, 2002.

15. Sotiriou, C., Powles, T. J., Dowsett, M., Jazaeri, A. A., Feldman, A. L., Assersohn, L., Gadisetti, C., Libutti, S. K., and Liu, E. T. Gene expression profiles derived from fine needle aspiration correlate with

response to systemic chemotherapy in breast cancer. Breast Cancer Res., *4:* R3, 2002.

16. Vapnik, V. Statistical Learning Theory. New York: Wiley, 1998.

17. Cristianini, N., and Shawe-Taylor, J. An Introduction to Support Vector Machines. Cambridge, United Kingdom: Cambridge University Press, 2000.

18. Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C-H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J. P., Poggio, T., Gerald, W., Loda, M., Lander, E. S., and Golub, T. R. Multiclass cancer diagnosis using tumor gene expression signatures. Proc. Natl. Acad. Sci. USA, *98:* 15149–15154, 2001.

19. Brown, M. P. S., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., Ares, M., Jr., and Haussler, D. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc. Natl. Acad. Sci. USA, *97:* 262–267, 2000.

20. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA, *95:* 14863–14868, 1998.

21. Jolliffe, I. T. Principal Component Analysis. New York: Springer Verlag, 2002.

22. Basilevsky, A. Statistical Factor Analysis and Related Methods: Theory and Applications. New York: John Wiley and Sons, 1994.

23. Crescenzi, M., and Giuliani, A. The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. FEBS Lett., *507:* 114–118, 2001.

24. Pelosi, G., Fraggetta, F., Pasini, F., Maisonneuve, P., Sonzogni, A., Iannucci, A., Terzi, A., Bresaola, E., Valduga, F., Lupo, C., and Viale, G. Immunoreactivity for thyroid transcription factor-1 in stage I non-small cell carcinomas of the lung. Am. J. Surg. Pathol., *25:* 363–372, 2001.

25. Cook, M., Buhling, F., Ansorge, S., Tatnell, P. J., and Kay, J. Pronapsin A and B gene expression in normal and malignant human lung and mononuclear blood cells. Biochim. Biophys. Acta, *1577:* 10–16, 2002.

26. Kalinichenko, V. V., Zhou, Y., Shin, B., Stolz, D. B., Watkins, S. C., Whitsett, J. A., and Costa, R. H. Wild-type levels of the *mouse Forkhead Box f1* gene are essential for lung repair Am. J. Physiol. Lung Cell Mol. Physiol., *282:* L1253–L1265, 2002.

27. LaBaer, J., Garrett, M. D., Stevenson, L. F., Slingerland, J. M., Sandhu, C., Chou, H. S., Fattaey, A., and Harlow, E. New functional activities for the p21 family of CDK inhibitors. Genes Dev., *11:* 847–862, 1997.

28. Eymin, B., Leduc, C., Coll, J. L., Brambilla, E., and Gazzeri, S. p14 (ARF) induces G_2 arrest and apoptosis independently of p53 leading to regression of tumours established in nude mice. Oncogene, *22:* 1822–1835, 2003.

29. Gelman, I. H. The role of SSeCKS/gravin/AKAP12 scaffolding proteins in the spaciotemporal control of signaling pathways in oncogenesis and development. Front Biosci. *7:* 1782–1797, 2002.

30. Suzuki, C., Unoki, M., and Nakamura, Y. Identification and allelic frequencies of novel single-nucleotide polymorphisms in the *DUSP1* and *BTG1* genes. J. Hum. Genet., *46:* 155–157, 2001.

31. Davies, M. A., Kim, S. J., Parikh, N. U., Dong, Z., Bucana, C. D., and Gallick, G. E. Adenoviral-mediated expression of MMAC/PTEN

inhibits proliferation and metastasis of human prostate cancer cells. Clin. Cancer Res., *8:* 1904–1914, 2002.

32. Baker, S. J., and Reddy, E. P. Transducers of life and death: TNF receptor superfamily and associated proteins. Oncogene, *12:* 1–9, 1996.

33. Pezzella, F., Pastorino, U., Tagliabue, E., Andreola, S., Sozzi, G., Gasparini, G., Menard, S., Gatter, K. C., Harris, A. L., Fox, S., Buyse, M., Pilotti, S., Pierotti, M., and Rilke, F. Non-small cell lung carcinoma tumor growth without morphological evidence of neo-angiogenesis. Am. J. Pathol., *151:* 1417–1423, 1997.

34. Offersen, B. V., Pfeiffer, P., Hamilton-Dutoit, S., and Overgaard, J. Patterns of angiogenesis in non-small cell lung carcinoma. Cancer (Phila.), *91:* 1500–1509, 2001.

35. Passalidou, E., Trivella, M., Singh, N., Ferguson, M., Hu, J., Cesario, A., Granone, P., Nicholson, A. G., Goldstraw, P., Ratcliffe, C., Tetlow, M., Leigh, I., Harris, A. L., Gatter, K. C., and Pezzella, F. Vascular phenotype in angiogenic and non-angiogenic lung non-small cell carcinomas. Br. J. Cancer, *86:* 244–249, 2002.

36. Tsai, M. S., Bogart, D. F., Castaneda, J. M., Li, P., and Lupu, R. Cyr61 promotes breast tumorigenesis and cancer progression. Oncogene, *21:* 8178–8185, 2002.

37. Apte, S. S., Olsen, B. R., and Murphy, G. The gene structure of tissue inhibitor of metalloproteinases (TIMP)-3 and its inhibitory activities define the distinct TIMP gene family. J. Biol. Chem., *270:* 14313–14318, 1995.

38. Denhardt, D. T., Mistretta, D., Chambers, A. F., Krishna, S., Porter, J. F., Raghuram, S., and Rittling, S. R. Transcriptional regulation of osteopontin and the metastatic phenotype: evidence for a ras-activated enhancer in the human OPN promoter. Clin. Exp. Metastasis, *20:* 77–84, 2003.

39. Degot, S., Regnier, C. H., Wendling, C., Chenard, M. P., Rio, M. C., and Tomasetto, C. Metastatic lymph node 51, a novel nucleo-cytoplasmic protein overexpressed in breast cancer. Oncogene, *21:* 4422–4434, 2002.

40. Cabibbo, A., Pagani, M., Fabbri, M., Rocchi, M., Farmery, M. R., Bulleid, N. J., and Sitia, R. ERO-L, a human protein that favors disulfide bond formation in the endoplasmic reticulum. J. Biol. Chem., *275:* 4827–4833, 2000.

41. Schaake-Koning, C., van den Bogaert, W., Dalesio, O., Festen, J., Hoogenhout, J., van Houtte, P., Kirkpatrick, A., Koolen, M., Maat, B., Nijs, A., Renaud, A., Rodrigus, P., Schuster-Uitterhoeve, L., Sculier, J-P., van Zandwijk, N., and Bartelink, H. Effects of concomitant cisplatin and radiotherapy on inoperable non-small cell lung cancer. N. Engl. J. Med., *326:* 524–530, 1992.

42. Rosell, R., Gomez-Codina, J., Camps, C., Maestre, J., Padille, J., Canto, A., Mate, J. L., Li, S., Roig, J., Olazabal, A., Canela, M., Ariza, A., Skacel, Z., Morera-Prat, J., and Abad, A. A randomized trial comparing preoperative chemotherapy plus surgery with surgery alone in patients with non-small cell lung cancer. N. Engl. J. Med., *330:* 153–158, 1994.

43. Seow, A., Duffy, S. W., Ng, T. P., McGee, M. A., and Lee, H. P. Lung cancer among Chinese females in Singapore 1968–1992: time trends, dialect group differences and implications for aetiology. Int. J. Epidemiol., *27:* 167–172, 1998.