Validation of a 16-Gene Expression Signature in Non-Small-Cell Lung Cancers from FFPE Samples

Yong Wu[1], Lily Nan[1], Hsuan-Yu Chen[2], Bassam El-Fahmawi[3], Kathryn Sciabica[1], Handy Yowanto[1], Sung-Liang Yu[4]

[1] Discovery Products, Beckman Coulter, Inc., 4300 N Harbor Blvd, Fullerton, CA 92834
[2] Institute of Statistical Science, Academia Sinica; Taipei, Taiwan
[3] Agencourt Bioscience Corporation, 500 Cummings Center, Beverly, MA, 01915
[4] Department of Clinical Laboratory Sciences and Medical Biotechnology, College of Medicine, National Taiwan University; Taipei, Taiwan

Abstract

Non-small-cell lung cancer (NSCLC) causes more deaths than any other type of cancer worldwide. The recurrence rate among patients with early-stage NSCLC is about 40% within 5 years even after receiving combined surgical and chemo therapies. The current tumor progression classification system is not an adequate measurement for treatment prognosis. Here we developed a multiplexed, quantitative biomarker assay method to validate the association between a 16-gene expression signature and the risk of overall or relapse-free survival. A multiplex of 16 NSCLC biomarkers, previously identified from a microarray study, and a reference gene was developed and evaluated. This multiplex was used to analyze 77 surgically resected, formalin-fixed, paraffin-embedded (FFPE) samples of NSCLC containing major histology types of adenocarcinoma (62%) and squamous-cell carcinoma (26%). Using as little as 5 ng of total RNA extracted from FFPE slices, we quantified the expression level of each gene with a linear correlation coefficient > 0.99. Risk scores were calculated based on the expression levels of the 16 genes. Patients with a low-risk, 16-gene signature had longer survival time than those with a high-risk, 16-gene signature (p=0.001, log-rank test). In addition, patients with high-risk, 16-gene signature had shorter relapse-free survival time than those with a low-risk, 16-gene signature (p=0.006, log-rank test). As a measure of the multiplex assay as an independent prognostic indicator, the hazard ratios (adjusted by age, gender, stage, and histology) were calculated as 2.2 (95% CI= 1.31-3.66, p=0.003) and 1.8 (95% CI= 1.11-2.92, p=0.0183) for overall survival and relapse-free survival, respectively.

Introduction

Gene-expression profiling by means of microarray and real-time PCR analysis has showed potential for classifying tumors and formulating a prognosis for patients with various types of cancer, including lung cancer. However, the use of microarray in practice is limited by complicated methods and the lack of reproducibility. On the other hand, the one-gene-at-a-time approach of real-time PCR is labor intensive and time consuming. XP-PCR is multiplexed quantitative method for analysis of gene expression that offers high sensitivity and excellent reproducibility. In the current study, we examined gene expression in 77 surgical specimens of NSCLC, using the XP-PCR method to identify a 16-gene signature that is correlated with the survival data.

Materials and Methods

Patients and Tissue Specimens. Computer-generated identification numbers were assigned to FFPE specimens of lung-cancer tissue from 77 randomly selected patients who underwent surgical resection of NSCLC at the Taichung Veterans General Hospital and didn't receive adjuvant chemotherapy between December 1999 and December 2003. Of these 77 specimens, 48 were adenocarcinomas, 20 were squamous-cell carcinomas, and 9 were other types of cancer (Table 1). The study was approved by the institutional review board of the hospital. Written informed consent was obtained from all patients.

Total RNA extraction from FFPE samples. RNA from FFPE samples was purified using FormaPure™ and the RNA only extraction protocol per the manufacturer recommendations (Agencourt® Bioscience, Beverly, MA). The deparafinization and proteinase K lysis steps were performed manually in 1.2 mL tubes by using 70˚C for deparafinization and 55˚C for proteinase K lysis, using a water bath incubator. After lysis, 200 µL of lysate was transferred to a 96 well plate, and then extracted using a Biomek® NXP SPAN 8 Automation Workstation* (Beckman Coulter®, Fullerton) with the FormaPure isolation method vs.3.01-3.3. RNA was then eluted from the magnetic beads using 80 µL nuclease free H_2O, 70 µL were transferred for further analysis. The quantification of total RNA isolated was determined by Nanodrop 260nm absorbance measurement.

Genes and Primer design. The 16 functional genes and a reference gene were selected from previous reported microarray results[1]. Primers were designed using the eXpress Design module of the GenomeLab™ GeXP Genetic Analysis System**. The gene-specific amplicon sizes used for this multiplex ranged from 63 nt to 95 nt, Gene names, accession numbers and the respective PCR product size are listed in Table 2.

Multiplexed quantitative gene expression analysis. The multiplexed reverse transcription and PCR amplification, which incorporates a universal priming strategy (XP-PCR), were conducted using the GeXP Start Kit** (Beckman Coulter) in accordance with the manufacturer's instruction. Specificity and identification of each gene peak is based on the differential migration rate of each gene fragment. Quantitation is based on the signal level of each peak relative to the standard curve with the Quant Tool** software (Beckman Coulter). The linear correlation coefficient (R^2) of each gene was shown in Table 2.

Statistical Analysis. Risk scores were calculated for 16 genes. A patient's risk score was calculated as the sum of the levels of expression of each gene. Patients were classified as having a high-risk gene signature or a low-risk gene signature, with the 50th percentile (median) of the risk score as the threshold value (median, 4.9; range, 1.3 to 21.9). The median risk score was chosen as the threshold value to reflect the fact that almost half of patients with early stage NSCLC relapse within 5 years after potentially curative surgery and also in order to eliminate the effect of extreme values in the training cohort by ensuring that there were equal numbers of patients in the high-risk and low-risk groups. The risk scores and the threshold value derived from the training cohort were not re-estimated but were applied directly to the testing cohort.

Table 1. Characteristics of the samples

Characteristic	Number	Percentage (%)
Age	63.5 (mean)	10.04 (sd)
Gender		
Male	47	61
Female	30	39
Cell Type		
Adenocarcinoa	48	62
Squamous cell carcinoma	20	26
others	9	12
Stage		
Stage I and II	24	31
Stage III and IV	53	69

Figure 1. The detection of the 16-gene panel with XP-PCR. A) All the genes were detected with 5 ng of total RNA from FFPE blocks. Red arrow indicates the internal reaction control for quantitative analysis. B) A representative linearity result shows the R^2 = 0.9983.

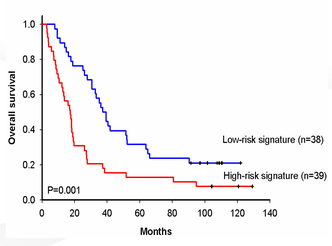

Figure 2. Kaplan–Meier estimates of overall survival according to the 16-gene signature

Table 3. Hazard Ratios for overall survival according to Multivariate Cox Regression Analysis

Variable	Hazard ratio	P value
Gene-signature	2.2	0.003
Stage	1.4	0.251
Age	0.99	0.633
Gender	1.02	0.93
Cell type	1	0.996

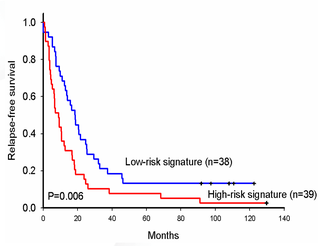

Figure 3. Kaplan–Meier estimates of relapse-free survival according to the 16-gene signature

Table 4. Hazard Ratios for relapse-free survival according to Multivariate Cox Regression Analysis

Variable	Hazard ratio	P value
Gene-signature	1.8	0.018
Stage	1.4	0.235
Age	0.99	0.680
Gender	0.9	0.781
Cell type	0.8	0.443

Table 2. The Non-small-cell lung cancer panel. The reference gene is highlighted in yellow. The correlation coefficient (R^2) of each gene was determined by the quantitative analysis.

Gene Symbol	Accession #	PCR Size	R^2
DLG2	NM_001364	100	0.9996
ZNF264	NM_003417	103	0.9999
NF1	NM_000267	106	0.9988
HMMR	NM_012484	109	0.9993
LCK	NM_005356	112	0.9996
IRF4	NM_002460	115	0.9992
CPEB4	NM_030627	118	0.9996
TBP	NM_003194	121	0.9991
HGF	NM_000601	124	0.9984
ERBB3	NM_001005915	127	0.9995
STAT1	NM_139266	130	0.9993
DUSP6	NM_001946	133	0.9998
RNF4	NM_002938	136	0.9989
FRAP1	NM_004958	139	0.9989
MMD	NM_012329	142	0.9992
STAT2	NM_005419	145	0.9984
ANXA5	NM_001154	148	0.9999

Conclusions

- Gene expression analysis from the XP-PCR results demonstrate a high degree of linearity for the quantitation of gene expression for each gene of interest.
- The 16-gene signature of non-small cell lung cancer panel correlates well with the overall survival data and the relapse-free survival data.
- Together, Agencourt FormaPure isolation and GeXP multiplex XP-PCR amplification were able to profile a 16-gene signature from 5ng of partially degraded FFPE samples.

Reference: 1. Chen *et al.* The New England Journal of Medicine. 2007, Vol 356: 12-20

*For Laboratory Use Only; not for use in diagnostic procedures.

** For research use only; not for use in diagnostic procedures.