Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfma-online.com

ORIGINAL ARTICLE

CPEB4 and *IRF4* expression in peripheral mononuclear cells are potential prognostic factors for advanced lung cancer

Yi-Ying Wu [a,b], Yi-Ting Hwang [c], Wann-Cherng Perng [d],
Chih-Feng Chian [d], Ching-Liang Ho [b], Shih-Chun Lee [e],
Hung Chang [e], Harn-Jing Terng [f], Tsu-Yi Chao [a,g,*]

[a] *Graduate Institute of Life Sciences, National Defense Medical Center, Taipei, Taiwan*
[b] *Division of Hematology and Oncology, Department of Internal Medicine, Tri-Service General Hospital, Taipei, Taiwan*
[c] *Department of Statistics, National Taipei University, New Taipei City, Taiwan*
[d] *Division of Pulmonary Medicine, Department of Internal Medicine, Tri-Service General Hospital, Taipei, Taiwan*
[e] *Division of Thoracic Surgery, Department of Surgery, Tri-Service General Hospital National, Taipei, Taiwan*
[f] *Advpharma, Inc., New Taipei City, Taiwan*
[g] *Division of Hematology and Oncology, Department of Internal Medicine, Taipei Medical University—Shuang Ho Hospital, New Taipei City, Taiwan*

Background/purpose: Lung cancer is a heterogeneous disease with varied outcomes. Molecular markers are eagerly investigated to predict a patient's treatment response or outcome. Previous studies used frozen biopsy tissues to identify crucial genes as prognostic markers. We explored the prognostic value of peripheral blood (PB) molecular signatures in patients with advanced non-small cell lung cancer (NSCLC).
Methods: Peripheral blood mononuclear cell (PBMC) fractions from patients with advanced NSCLC were applied for RNA extraction, cDNA synthesis, and real-time polymerase chain reaction (PCR) for the expression profiling of eight genes: *DUSP6, MMD, CPEB4, RNF4, STAT2, NF1, IRF4,* and *ZNF264*. Proportional hazard (PH) models were constructed to evaluate the association of the eight expressing genes and multiple clinical factors [e.g., sex, smoking status, and Charlson comorbidity index (CCI)] with overall survival.

Results: One hundred and forty-one patients with advanced NSCLC were enrolled. They included 109 (77.30%) patients with adenocarcinoma, 12 (8.51%) patients with squamous cell carcinoma, and 20 (14.18%) patients with other pathological lung cancer types. A PH model containing two significant survival-associated genes, *CPEB4* and *IRF4,* could help in predicting the overall survival of patients with advanced stage NSCLC [hazard ratio (HR) = 0.48, $p < 0.0001$]. Adding multiple clinical factors further improved the prediction power of prognosis (HR = 0.33; $p < 0.0001$).

Conclusion: Molecular signatures in PB can stratify the prognosis in patients with advanced NSCLC. Further prospective, interventional clinical trials should be performed to test if gene profiling also predicts resistance to chemotherapy.

Introduction

Lung cancer is the leading cause of cancer deaths worldwide with non-small cell lung cancer (NSCLC) being the most common type.[1] Clinical trials for new treatment modalities for NSCLC have increased in recent decades; however, the outcome remains dismal for patients with advanced disease. Accurate staging is important for treatment selection and for estimating prognosis. However, many variables between individuals such as age, performance status, smoking, and ethnicity may contribute significantly to the outcome of patients with NSCLC.[2] The TNM staging system is the standard prognostic assessment in which tumor size, lymph node involvement, and distant metastasis are evaluated.[3] Lung cancer is a heterogeneous disease; therefore, personalized treatment has become a trend for maximizing treatment efficacy and minimizing adverse effects.[4]

Gene expression profiling for prognosis prediction or personalized medicine is emerging and most strategies utilize tissue section as the specimen.[5] Previous studies identified transcriptional levels of several genes that were correlated with clinical outcome in lung cancer patients, based on resected lung tissues. However, this approach is only feasible in patients with early stage lung cancer.[5–10] Patients with advanced stage disease typically undergo a biopsy to confirm the diagnosis. A limited amount of clinical specimen may be inadequate or inappropriately processed for extensive gene expression analysis. Suwinski et al[11] proposed using specimens from bronchoscopic examinations. In a small cohort, they found that *ERCC1* and *CA9* had an impact on NSCLC prognosis. In current clinical settings, a peripheral blood (PB) sample would be more feasible for patients diagnosed with advanced stage disease. Cancer cells enter blood circulation by invading through tumor-associated blood vessels.[12] It is reasonable to propose that gene expression in circulating tumor cells may predict the outcome of patients with advanced disease. Only a few available reports present data on applying PB mononuclear cells (PBMCs) containing circulating tumors cells (CTC) for gene expression profiling.[13–21]

Using tumor tissue samples as a source, Chen et al[5] defined 16 prognosis-associated mRNA markers a Taiwan cohort of non-small cell lung cancer (NSCLC) patients. In the previous study,[5] 12 mRNA markers with increased expression level were poor prognostic factors, while four other markers with reduced expression were protective factors. In our study, PB was collected from patients with advanced NSCLC and processed for the gene expression analysis of 12 poor prognosis risk markers via real-time polymerase chain reaction. The four protective markers with reduced expression were not studied since they may not be consistently detectable in PB on account of the potentially low numbers of CTC. There may be a high probability of identifying tumor cell-derived markers in PB because these mRNA markers are well documented as being lung tumor tissue-specific. By combining clinical risk factors with gene expression signatures, a correlation was established between survival and gene expression in these patients. The predictive value of PB gene expression on the survival of patients with advanced NSCLC was validated.

Methods

Patients and blood samples

One hundred and forty-one patients with clinically confirmed NSCLC were enrolled (June 2006–April 2013) in a prospective investigational protocol that was approved by the Institutional Review Board at Tri-Service General Hospital (Taipei, Taiwan). The disease stage was classified in accordance with the TNM system utilized by the *AJCC Cancer Staging Manual,*[22] 7th edition by the American Joint Committee on Cancer. Forty-two (29.79%) patients had Stage IIIB disease and 99 (70.21%) patients had Stage IV disease. The pathological types of cancer were adenocarcinoma in 109 patients (77.30%), squamous cell carcinoma in 12 (8.51%) patients, and other pathological types in 21 (14.18%) patients. Data on the vital status were available through August 13, 2013.

Before any treatment, patients provided written informed consent. Samples of PB (6–8 mL) were then drawn. Blood samples were stored at 4°C until the isolation of PBMC fraction within 3 hours. The PBMC fraction was used for further preparations, as described by Huang

et al,[23] which included RNA extraction, reverse transcription, and real-time PCR analysis. All RNA and cDNA samples were stored at $-80°C$ before analysis.

Cell cultures

A human lung adenocarcinoma cell line, A549, was obtained from Bioresourse Collection and Research Center (Hsinchu, Taiwan; catalog number, BCRC60074) and maintained in Ham's F-12K medium with 2mM of glutamine adjusted to contain 1.5 g/L sodium bicarbonate supplemented with 10% fetal bovine serum. The cells were incubated in 5% carbon dioxide and humidified at $37°C$ for growth. Cell cultures were split 10-fold every 3—4 days.

Validation of the reference genes and real-time PCR sensitivity assay

Human lung adenocarcinoma cell line A549 was spiked into blood samples from healthy volunteers at concentrations of 0 cells/mL, 50 cells/mL, 100 cells/mL, 300 cells/mL, 1000 cells/mL, and 3000 cells/mL. The samples were processed for isolation of the PBMC fraction. Total RNA extraction, cDNA synthesis, and real-time PCR assay were performed using the predesigned gene-specific amplification primers HK-HPRT1 primer (Advpharma, New Taipei City, Taiwan) and HK-ACTB primer (Advpharma). The average of five repeated measurements (cycle number; Cp-values), which was derived from the graded A549 spike-in samples, was 28.84 ± 0.57 for *HPRT1* and 18.50 ± 0.77 for *ACTB*. The Cp-values of *HPRT1* [i.e., Cp(*HPRT1*)] had a medium-low expression level, which corresponded to the findings by Dheda et al.[24] Because the Cp-values of the investigated genes in this study had a similar expression level, all test results confirmed that *HPRT1* was a reproducible and proper reference gene for real-time PCR assays for our study. The expression level of each investigated gene in a sample or cell was normalized to that of *HPRT1* and is presented as the delta-Ct (ΔCt) value [i.e., Ct(*HPRT1*) $-$ Ct(test)], which is inversely correlated with the gene expression level.[25]

For assessing the sensitivity of real-time PCR, all blood samples (with and without the spike-in) were simultaneously processed for the isolation of PBMC fraction and total RNA extraction. Quantitative real-time PCR of A549 cell numbers in the spike-in blood sample using keratin 19-primers (KRT19 Primer; Advpharma) was performed. The *HPRT1* mRNA level was used as the reference for normalization using real-time PCR assay. A reaction mixture without cDNA was used as the negative control to confirm PCR assay quality during each analytic batch.

Each reaction mixture contained 1/20 volume of cDNA derived from 500—20 cells/mL blood as a dilute percentage for the experiment. Therefore, the cell number equivalents were 25 cells/assay, 15 cells/assay, 5 cells/assay, 2.5 cells/assay, 2 cells/assay, 1.5 cells/assay, and 1 cell/assay. To assess the detection performance, the Cp(*KRT19*) of each sample was first normalized with the Cp(*HPRT1*), and obtained $\triangle Cp =$ Cp(*HPRT1*) $-$ Cp(*KRT19*). The relative change of each spiked sample was then calculated as $\triangle Cp = \triangle Cp$(Reference) $- \triangle Cp$(spike-in sample).

Statistical analysis

The Kaplan—Meier (KM) method was used to estimate the overall survival (i.e., death from all causes). Differences in survival between risk factors were analyzed with the log-rank test. The univariate Cox model was used to assess the association between each gene expression and overall survival. Furthermore, the multivariate Cox proportional method was used to evaluate the association between survival and the independent influence of prognostic factors, including all gene expressions, sex, smoking status, and Charlson comorbidity index (CCI).[11] The stepwise procedure was chosen to determine the most significant risk factors in each model in which the inclusion and exclusion criteria were set to 0.2. The patient's risk score was computed by summing the estimated coefficients times the corresponding predictor values. Using the median risk score as the cutoff value, patients were classified into the low-risk group or high-risk group. A high-risk score meant a poorer prognosis for patients. The hazard ratio (HR) between the high-risk and low-risk groups was again obtained by the univariate Cox model to assess the predicative power for the model. Furthermore, the likelihood displacement statistic, Martingale residuals, and the score residuals were used to inspect the model adequacy.[26] In particular, the influential and poorly fit subjects were identified by the likelihood displacement statistic and Martingale residuals. The unusual influential subjects for each covariate were inspected through the score residuals.

Results

Assessment of the feasibility and sensitivity of the PBMC-based molecular assay

We used A549 cell-spiked whole blood obtained from healthy volunteers to mimic blood sample containing CTCs for the feasibility test. To determine in which subfraction these spiked A549 tumor cells may be located after the blood sample preparation, the Ficoll-based method was applied to separate plasma, mononuclear cells, granulocytes, and red blood cells. Total RNA extraction of these subfractions and cDNA synthesis were performed on the patients' blood samples. The following PCR tests using A549 cell-specific KRT19 Primer (Advpharma) were performed to detect the spiked A549 cells since *KRT19* is a A549-specific molecular marker.[20,27] The test results showed that the A549 cell-specific PCR product was mostly observed for the mononuclear cell fraction, a weak signal for the granuocyte fraction and no signal for plasma and red blood cells. Different concentrations of A549 cell-spiked blood samples were used to evaluate the detection sensitivity of the molecular assay. Blood samples without the spike-in tumor cells were the reference.

Assays with the serial spiked A549 cells showed very good linearity ($R^2 = 0.9934$) between the spike-in A549 cell numbers and the relative quantitative change measured by real-time PCR assay. In summary, the spike-in test results showed that the spiked A549 cells at a concentration of > 5 cells/mL blood could be consistently detected using the assay method described.

Selection of potentially prognostic genes in PB samples

Twelve tumor tissue-specific prognosis risk genes with increased expression in NSCLC (i.e., HR > 1), based on previously published findings,[5] were selected as investigating genes at the beginning of this study. High-risk genes were *ERBB3, DUSP6, MMD, CPEB4, RNF4, STAT2, NF1, DLG2, IRF4, HMMR, HGF,* and *ZNF264*.

Real-time PCR assays for the expression of these 12 genes were performed initially with a small cohort of 10 clinical samples to evaluate PCR specificity and reproducibility of measurements. After completing this learning set, eight genes—*DUSP6, MMD, CPEB4, RNF4, STAT2, NF1, IRF4,* and *ZNF264*—could be reproducibly assayed using the PB samples and were selected for further investigation in all patients.

Demographics and histopathology as risk factors for survival

The average age for our cohort was 65.5 ± 12.3 years. Among them, 63.12% were male, 54.61% were older than 65 years, and 41.84% had never smoked. Characteristics such as sex, age, smoking status, and pathological type, were significantly associated with overall survival, based on the log-rank test (Table 1). In particular, the median survival time for female patients was 2.28 times longer than that for male patients. Patients older than 65 years had a shorter median survival time than younger patients (9.23 months vs. 22.6 months). Furthermore, patients who were nonsmokers had stage IIIB cancer or who had adenocarcinoma had longer survival times, compared to smokers, and patients with stage IV and other NSCLC tumor types. In addition, the CCI was inversely correlated with overall survival. Patients whose CCI value was 3—5 had a median survival time longer than 3 years. The median survival time decreased dramatically to 1.6 years for patients with a CCI value of 6—8 and less than 1 year for patients with a CCI value greater than 8.

Univariate Cox analysis of expressing genes and survival

Table 2 lists the HR from the univariate Cox regression analysis of the association between each expressing gene and survival. Risk genes were associated with a HR more than 1 for death, whereas protective genes had a HR less than 1. Among the eight genes, only two genes, *CPEB4* and *IRF4*, were significantly associated with survival, based on PB samples; however, the other six genes were significant prognostic factors, based on lung tumor tissue samples.[5] Depending on the HR, for every unit increase in *CPEB4* gene expression, the hazard rate increased 31%, whereas for every unit increase in *IRF4* gene expression, the hazard rate declined 28%.

Multivariate Cox proportional hazards models and survival

The multivariate Cox proportional hazards (PH) method was implemented to evaluate the association between eight genes and three clinical factors—sex, smoking status, and CCI—with overall survival. Gene signatures were included in the PH models without clinical factors (Model 1) and with clinical factors (Model II).

The stepwise approach using the multivariate PH model was computed, and selected the same two significant genes for Model I and Model II as those obtained from univariate regression analysis (Table 3). Model I contained two significant survival-associated genes: *CPEB4* and *IRF4*. Model II included the same gene signature and two additional clinical factors: sex and CCI. The levels of significance and HR for *CPEB4* and *IRF4* remained similar to those in Model I. In addition to gene expression, men had a higher HR than women (HR = 1.91; *p* = 0.0023). For every unit increase in CCI, the HR increased 18%.

Risk scores and overall survival

For genes and clinical factors that were significantly associated with overall survival, we used a linear combination

Table 1 Medium survival time for each demographic characteristic.

Characteristic	Categories	Patient No.	%	Median survival, no. of days (range)	*p*
Sex	Female	52	36.88	861 (631—1275)	<0.0001
	Male	89	63.12	378 (269—458)	
Age	≤65	64	45.39	678 (505—961)	0.0001
	>65	77	54.61	277 (240—400)	
Smoking status	Nonsmoker	59	41.84	739 (430—1143)	0.0002
	Smoker	82	58.16	400 (282—490)	
Clinical Stage	IIIB	42	29.79	875 (478—1275)	0.0027
	IV	99	70.21	424 (276—517)	
Type	Adenocarcinoma	109	77.30	588 (433—739)	0.0014
	Other	20	14.18	292.5 (159—564)	
	Squamous cell carcinoma	12	8.51	196 (21—431)	
Charlson Comorbidity Index	3—5	23	16.31	1401 (564—1988)	0.0001
	6—8	43	30.50	588 (431—678)	
	>8	75	53.19	307 (240—429)	

Table 2 Univarite Cox model for genes.

Gene ID	Gene name	Hazard ratio (95% CI)	p
DUSP6	Dual specificity phosphatase 6	1.05 (0.83—1.32)	0.6822
MMD	Monocyte to macrophage differentiation-associated	1.10 (0.93—1.31)	0.2536
CPEB4	Cytoplasmic polyadenylation element binding protein 4	1.31 (1.12—1.55)	0.0011
RNF4	Ring finger protein 4	0.80 (0.59—1.09)	0.1571
STAT2	Signal transducer and activator of transcription 2 (113 kDa)	0.90 (0.66—1.22)	0.4924
NF1	Neurofibromin 1	1.39 (0.98—1.97)	0.0657
IRF4	Interferon regulatory factor 4	0.72 (0.55—0.94)	0.0160
ZNF264	Zinc finger protein 264	0.93 (0.70—1.24)	0.6217

CI = confidence interval.

of the gene expression values weighted by the regression coefficients to calculate a risk score for each patient. Patients with a risk score greater than the median risk score (as the cutoff) were classified into the high-risk group, and patients with risk scores less than the cutoff value were classified into the low-risk group. Figures 1A and 1B display the KM curves for the stratified risk groups using Models I and II, respectively. Patients in the high-risk group had a significant lower median overall survival than patients in the low-risk group (9.9 months vs. 20.75 months), based on Model I (i.e., gene signature only). In Model II, which combines gene signature, sex, and CCI, the median overall survival of high-risk group was also significantly lower, compared to patients in the low-risk group (8.68 months vs. 28.31 months). The distinction between the KM curves for the low and high-risk groups for Model I was not as evident as in Model II. The HR for Model I and Model II for median survival were 0.48 and 0.33, respectively. These results indicated that Model I (i.e., gene signature only) was able to stratify low risk and high risk; the inclusion of clinical factors in Model II provided even better stratification.

The model adequacy was assessed by the likelihood displacement statistic, Martingale residuals, and the score residuals. After examining the three residuals (data not shown), nine patients (010, 026, 028, 043, 100, 112, 117, 128, and 140; Table 4) were classified as potential outliers or influential points. To understand the discrepancy of the nine patients from all other patients, the scatter plot of IRF4 expression level versus the CCI of all patients, stratified by sex is presented in Figure 2. The potential outliers, except patient 140, had a relatively higher expression level of IRF4 at each CCI unit. Patients 043, 112, 128, and 140 were female. Patients 028, 043, 117, and 140 were younger than 65 years. The survival time of these outliers ranged 17.13—77.7 months and their CCI values ranged 4—13. Patient 117 had a strong influence on the significance of IRF4 for Model II. In particular, when this patient was censored from the analysis, IRF4 became insignificant ($p = 0.0768$). On censoring all nine outliers, IRF4 had a weak association with the survival ($p = 0.4632$).

Discussion

A certain proportion of cancer patients do not respond to treatment and have unexpected disease progression during treatment. In the current era, molecular markers have been investigated to predict this poor prognosis in NSCLC. For patients with advanced stage NSCLC for whom there are inadequate tissue samples, gene testing or gene expression profiling may not be feasible. Circulating tumor cells have been associated with a shorter survival in patients with advanced lung cancer.[17] There is no published data on CTC-derived gene expression and its correlation with cancer diagnosis or progression. The approach using PB as assay material is more feasible for most patients and the methodology of gene expression measurement is clinically applicable. In this study, we intended to quantify valid cancer gene transcriptional products derived from PBMC-fractions from NSCLC patients. We have demonstrated that this approach could be performed to predict the patients' outcomes. We began by conducting a pilot study in a small cohort and measuring the transcription level of 12 tumor tissue-specific genes, based on previous results reported by Chen et al.[5] Only eight genes were repeatedly detected in the PBMCs samples so these were applied to all clinical samples in this study. Model I, the PH model with a two-gene signature (i.e., CPEB4 and IRF4), demonstrated a significant predictive parameter for survival in our cohort. Additional clinical variables for PH Model II did not influence the predictive genes, which showed the validity of the same two-gene signature.

Our sample size was not sufficient; therefore, we could not check the validation cohort. We instead checked many types of residuals such as likelihood displacement

Table 3 Proportional hazards model for genes only and for genes plus risk factors.

Characteristic	Categories	PH Model I (genes)		PH Model II (genes and risk factors)	
		Hazard ratio (95% CI)	p	Hazard ratio (95% CI)	p
Sex	Male	None	None	1.91 (1.26—2.90)	0.0023
Charlson Comorbidity Index		None	None	1.18 (1.08—1.28)	0.0001
Gene expression	CPEB4	1.37 (1.15—1.63)	0.0004	1.30 (1.09—1.56)	0.0037
	IRF4	0.68 (0.52—0.89)	0.0046	0.72 (0.55—0.95)	0.0185

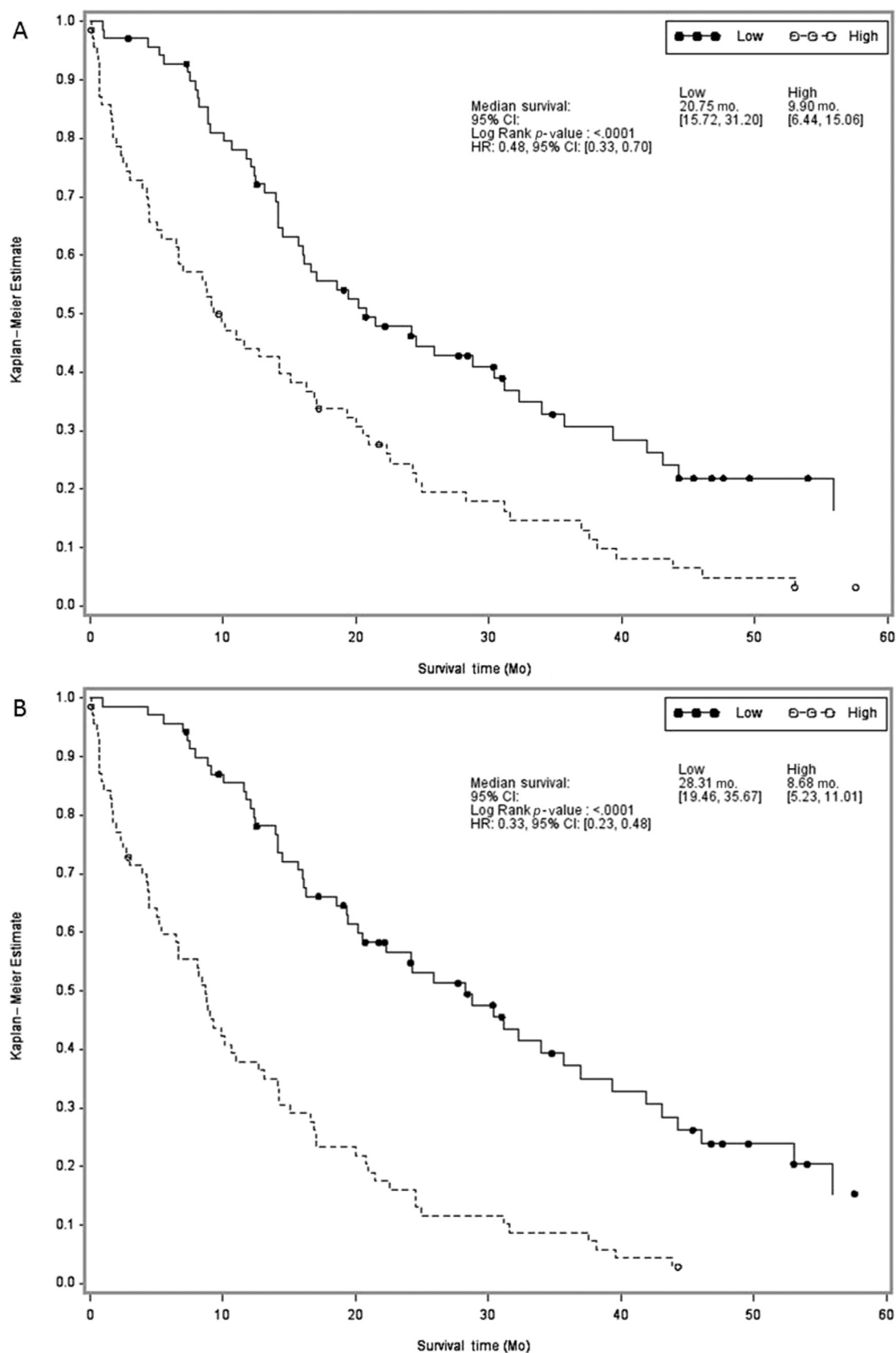CI = confidence interval; PH = proportional hazards.

Figure 1   The Kaplan—Meier (KM) method was used to estimate the overall survival, according to the proportional hazards model. (A) Model I (i.e., gene signature only). (B) Model II [i.e., gene signature, sex and Charlson comorbidity index (CCI)]. Differences in survival between risk factors were analyzed with the log-rank test. In the upper right corner of the figure is information on the median survival time (in months); 95% confidence interval (CI); *p*-value, based on the log-rank test; and hazard ratio (HR) between low-risk and high-risk group.

Table 4 Clinical characteristics, CCI, and survival status of nine potential outliers.

No.	Patient ID	Stage	CCI	Survival status	Time (mo)	Treatment
1	010	IV	9	D	25.27	None
2	026	IV	9	D	17.13	C/T
3	028	IV	10	D	38.70	C/T follow by afatinib
4	043	IIIB	4	D	56.70	C/T follow by getitinib
5	100	IV	13	D	21.03	C/T follow by getitinib
6	112	IV	13	D	38.10	Getitinib follow by C/T
7	117	IV	8	A	65.07	C/T
8	128	IV	11	D	44.43	C/T follow by getitinib
9	140	IIIB	4	A	58.43	C/T follow by getitinib

A = remained alive until August 13, 2013; C/T = chemotherapy; CCI = Charlson Comorbidity Index; D = death.

statistics, Martingale residuals, and score residuals to assess the model adequacy and influential sample points. Nine potential outliers for Model II, which were identified by Martingale residuals, all showed relatively better prognosis with a longer survival time (17.13–77.7 months; Table 4). They all had the adenocarcinoma type of disease, which was confirmed as a significant survival-associated factor.[28] These outliers unexpectedly had a better survival among the late-stage study population, although not all of them had other favorable prognostic factors such as female sex, young age, and lower CCI. However, we found eight of nine patients could tolerate, and were treated with, more than two lines of chemotherapy, including targeted therapy. In addition to certain regulated expression of genes, the combination of the histologic subtype (i.e., adenocarcinoma) and targeted therapy may be contributing factors for a better clinical outcome.

Our results demonstrated that among patients with advanced disease, the elevated transcription of *CPEB4* in PB was associated with shorter survival. *CPEB4* activation promotes protein translation by inducing cytoplasmic polyadenylation. Some clinical studies report increased *CPEB4* transcription in patients with pancreatic and metastatic prostate cancer.[29] In cell-based experiments, CPEB4 was elevated in pancreatic, lung, and ovarian cell lines compared to nontransformed control cells.[30] Huang et al[23] also found elevated *CPEB4* in PBMC of colon cancer
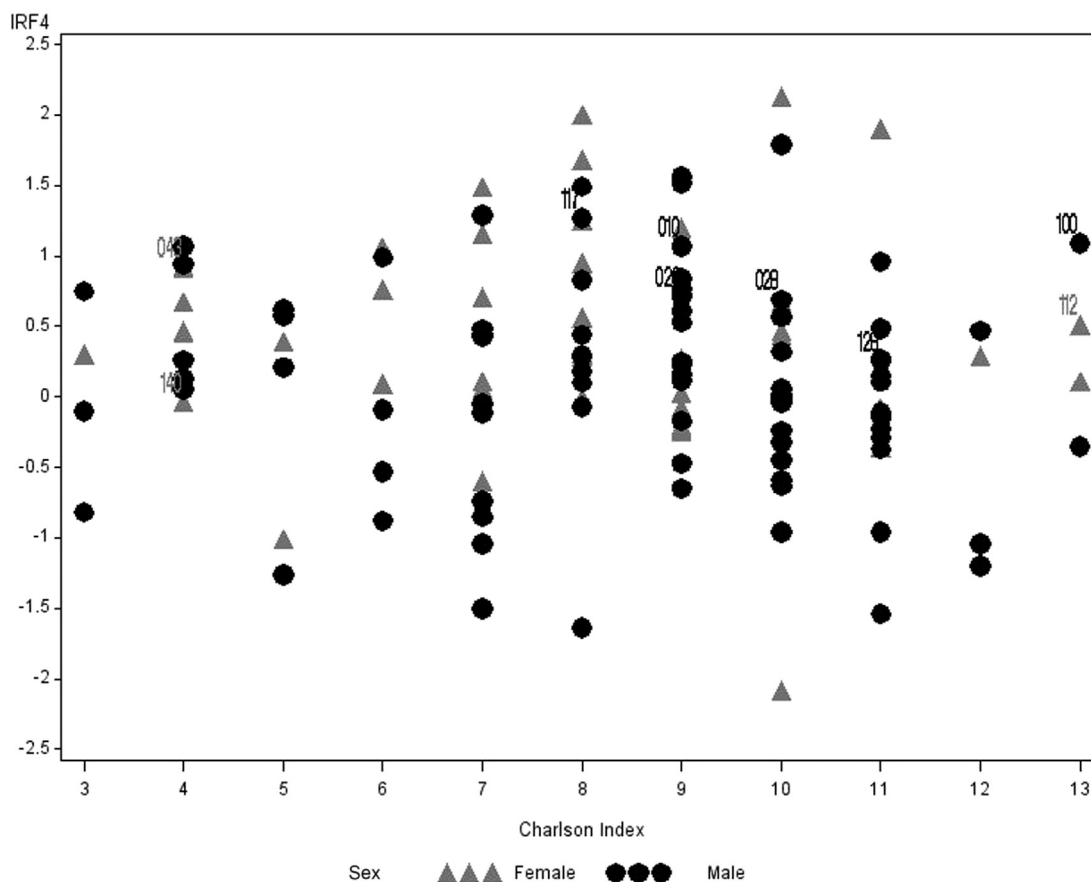

Figure 2 The scatter plot shows the IRF4 expression level versus CCI. Outliers identified by Martingale residuals—patients 010, 026, 028, 043, 100, 112, 117, 140, and 128—are specifically marked.

patients, compared to that of a healthy cohort. *CPEB4* expression also facilitates transcription factors of the SMAD family to regulate transforming growth factor-beta, which activates epithelial—mesenchymal transition and promotes further metastasis.[31] Our finding that *CPEB4* represents a negative prognostic risk factor in NSCLC patients is consistent with all of the aforementioned evidence.

IRF4, also known as MUM1, is a transcriptional factor that participates in the immune regulation of lymphoid, myeloid, and dendritic-cell differentiation.[32–34] Lymphoid malignancy and multiple myeloma have reportedly increased IRF4, which promotes downstream nuclear factor kappa B or MYC.[35] IRF4 primarily regulates immune response against infection and regulates interferon-inducible genes.[36] In this study, we demonstrated that *IRF4* was a protective prognostic factor in NSCLC patients in both PH models.

Compared to PH Model I, Model II included two additional risk factors: CCI and sex. To understand the interaction of these variables, the scatter plot for *IRF4* versus CCI stratified by sex is displayed in Figure 2. Female patients seemed to have higher *IRF4*. As CCI increased, *IRF4* declined slightly. The influential male patients seemed to have slightly higher *IRF4*. Under the same CCI, patients 043 and 140 seemed to have a lower expression of *IRF4* than other male patients. When deleting these influential patients, the effect of *IRF4* on survival could be replaced by the CCI and sex. *IRF4* behaves as a tumor suppressor in B-cell malignancy, and we assume that it may also be a protective factor in lung cancer pathogenesis.[37] Spitz et al[38] has shown that single-nucleotide polymorphisms of *IRF4* may have a role in lung adenocarcinoma patients who had never smoked.[12] Our study results are compatible with Spitz et al. Since all nine *IRF4*-dependent cases were adenocarcinoma. In a previous study by Chen et al,[5] *IRF4* was shown to predict poorer survival of NSCLC patients. By contrast, our study demonstrated that *IRF4* has a protective role. These contradictory data suggest that *IRF4* may reflect tumor-infiltrating lymphocyte activity in tissue section, but reflect different lymphocyte population in PB.

There are several limitations of this study. First, it was a noninterventional cohort study in a single institute. Candidate gene survey was limited to a small group of the Taiwanese population. Other ethnicities may have different genetic influences on lung cancer; lung cancer incidence, and demographics are significantly different in Asia than in Western countries.[39] Second, patients in our study underwent treatment as clinically indicated with targeted agents and chemotherapy with or without radiation therapy. Because these treatment factors were not controlled, the final results of survival analysis may have been influenced. To capture as many patients as possible, we enrolled all NSCLC patients without previous balancing of tumor type. Most patients had adenocarcinoma. The potential impact of tumor type was therefore not adjusted. Moreover, the types of gene mutations in different pathological types may have differed. Third, our study started from 2006, 3 years before tyrosine kinase inhibitor (TKI) was approved (in 2009) to treat patients with advanced adenocarcinoma.[40] Epidermal growth factor receptor (EGFR) mutation predicts a survival benefit in patients with advanced adenocarcinoma of the lung who undergo TKI treatment.[41] In a

subgroup analysis, we also found that our patients with adenocarcinoma had better prognosis, compared to other pathological subtypes. However, we did not analyze the difference between NSCLC with and without mutated *EGFR*. Whether this mutation could have an impact on the prognosis of our patients is unknown and will be the subject of further study.

In conclusion, we have shown that risk assessment using a model that includes selected gene expression alone or in combination with clinical risk factors could predict the clinical outcome of patients with advanced NSCLC. Most importantly, we showed that PBMC is a convenient and adequate resource for gene signature testing that would be available for all patients. Larger cohorts that include other ethnic groups and studies controlling for treatment regimen and pathological subtype are necessary to validate our current results.

Acknowledgments

References

1. Travis WD, Brambilla E, Riely GJ. New pathologic classification of lung cancer: relevance for clinical practice and clinical trials. *J Clin Oncol* 2013;**31**:992–1001.
2. Lo YL, Hsiao CF, Chang GC, Tsai YH, Huang MS, Su WC, et al. Risk factors for primary lung cancer among never smokers by gender in a matched case-control study. *Cancer Causes Control* 2013;**24**:567–76.
3. Mirsadraee S, Oswal D, Alizadeh Y, Caulo A, van Beek Jr E. The 7th lung cancer TNM classification and staging system: review of the changes and implications. *World J Radiol* 2012;**4**:128–34.
4. Salgia R, Hensing T, Campbell N, Salama AK, Maitland M, Hoffman P, et al. Personalized treatment of lung cancer. *Semin Oncol* 2011;**38**:274–83.
5. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 2007;**356**:11–20.
6. Tang H, Xiao G, Behrens C, Schiller J, Allen J, Chow CW, et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin Cancer Res* 2013;**19**:1577–86.
7. Park YY, Park ES, Kim SB, Kim SC, Sohn BH, Chu IS, et al. Development and validation of a prognostic gene-expression signature for lung adenocarcinoma. *PLoS One* 2012;**7**:e44225.
8. Skrzypski M, Dziadziuszko R, Jassem E, Szymanowska-Narloch A, Gulida G, Rzepko R, et al. Main histologic types of

non-small-cell lung cancer differ in expression of prognosis-related genes. *Clin Lung Cancer* 2013;**14**. 666—73.e2.

9. Melguizo C, Prados J, Luque R, Szymanowska-Narloch A, Gulida G, Rzepko R, et al. Modulation of multidrug resistance gene expression in peripheral blood mononuclear cells of lung cancer patients and evaluation of their clinical significance. *Cancer Chemother Pharmacol* 2013;**71**:537—41.

10. Li Y, Tang H, Sun Z, Bungum AO, Edell ES, Lingle WL, et al. Network-based approach identified cell cycle genes as predictor of overall survival in lung adenocarcinoma patients. *Lung Cancer* 2013;**80**:91—8.

11. Suwinski R, Klusek A, Tyszkiewicz T, Kowalska M, Szczesniak-Klusek B, Gawkowska-Suwinska M, et al. Gene expression from bronchoscopy obtained tumour samples as a predictor of outcome in advanced inoperable lung cancer. *PloS ONE* 2012;**7**: e41379.

12. Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. *Nature* 2000;**407**:249—57.

13. Zander T, Hofmann A, Staratschek-Jox A, Classen S, Debey-Pascher S, Maisel D, et al. Blood-based gene expression signatures in non-small cell lung cancer. *Clin Cancer Res* 2011;**17**:3360—7.

14. Devriese LA, Bosma AJ, van de Heuvel MM, Heemsbergen W, Voest EE, Schellens JH. Circulating tumor cell detection in advanced non-small cell lung cancer patients by multi-marker QPCR analysis. *Lung Cancer* 2012;**75**:242—7.

15. Hou JM, Krebs M, Ward T, Sloane R, Priest L, Hughes A, et al. Circulating tumor cells as a window on metastasis biology in lung cancer. *Am J Pathol* 2011;**178**:989—96.

16. Liu L, Liao GQ, He P, Zhu H, Liu PH, Qu YM, et al. Detection of circulating cancer cells in lung cancer patients with a panel of marker genes. *Biochem Biophys Res Commun* 2008;**372**:756—60.

17. Krebs MG, Sloane R, Priest L, Lancashire L, Hou JM, Greystoke A, et al. Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer. *J Clin Oncol* 2011;**29**:1556—63.

18. Showe MK, Vachani A, Kossenkov AV, Yousef M, Nichols C, Nikonova EV, et al. Gene expression profiles in peripheral blood mononuclear cells can distinguish patients with non-small cell lung cancer from patients with nonmalignant lung disease. *Cancer Res* 2009;**69**:9202—10.

19. Chen YC, Hsiao CC, Chen KD, Hung YC, Wu CY, Lie CH, et al. Peripheral immune cell gene expression changes in advanced non-small-cell lung cancer patients treated with first line combination chemotherapy. *PloS ONE* 2013;**8**:e57053.

20. Sher YP, Shih JY, Yang PC, Roffler SR, Chu YW, Wu CW, et al. Prognosis of non-small cell lung cancer patients by detecting circulating cancer cells in the peripheral blood with multiple marker genes. *Clin Cancer Res* 2005;**11**:173—9.

21. Allard WJ, Matera J, Miller MC, Repollet M, Connelly MC, Rao C, et al. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin Cancer Res* 2004;**10**:6897—904.

22. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th Edition of the AJCC Cancer Staging Manual and the Future of TNM. *Ann Surg Oncol* 2010;**17**(6):1471—4.

23. Huang CS, Terng HJ, Chou YC, Su SL, Terng HJ, Chou HL, et al. A gene expression profile of peripheral blood in colorectal cancer. *J Microb Biochem Technol* 2014;**6**:102—9.

24. Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques* 2004;**37**:112—4. 116, 118—9.

25. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;**25**:402—8.

26. David W, Hosmer SL, Stanley L, Susanne M. *Applied survival analysis: regression modeling of time-to-event data*. Hoboken, NJ, USA: John Wiley and Sons, Inc.; 2008.

27. Croce MV, Colussi AG, Price MR, Segal-Eiras A. Identification and characterization of different subpopulations in a human lung adenocarcinoma cell line (A549). *Pathol Oncol Res* 1999;**5**: 197—204.

28. Cetin K, Ettinger DS, Hei YJ, O'Malley CD. Survival by histologic subtype in stage IV nonsmall cell lung cancer based on data from the Surveillance, Epidemiology and End Results Program. *Clin Epidemiol* 2011;**3**:139—48.

29. Ortiz-Zapater E, Pineda D, Martinez-Bosch N, Fernández-Miranda G, Iglesias M, Alameda F, et al. Key contribution of CPEB4-mediated translational control to cancer progression. *Nat Med* 2012;**18**:83—90.

30. Xu H, Liu B. CPEB4 is a candidate biomarker for defining metastatic cancers and directing personalized therapies. *Med Hypotheses* 2013;**81**:875—7.

31. D'Ambrogio A, Nagaoka K, Richter JD. Translational control of cell growth and malignancy by the CPEBs. Nature reviews. *Cancer* 2013;**13**:283—90.

32. Ozato K, Tailor P, Kubota T. The interferon regulatory factor family in host defense: mechanism of action. *J Biol Chem* 2007; **282**:20065—9.

33. Lu R. Interferon regulatory factor 4 and 8 in B-cell development. *Trends Immunol* 2008;**29**:487—92.

34. Tamura T, Yanai H, Savitsky D, Taniguchi T. The IRF family transcription factors in immunity and oncogenesis. *Annu Rev Immunol* 2008;**26**:535—84.

35. Shaffer AL, Emre NC, Romesser PB, Staudt LM. IRF4: Immunity. Malignancy! Therapy? *Clin Cancer Res* 2009;**15**:2954—61.

36. Pena-Chilet M, Blanquer-Maceiras M, Ibarrola-Villava M, Martinez-Cadenas C, Martin-Gonzalez M, Gomez-Fernandez C, et al. Genetic variants in PARP1 (rs3219090) and IRF4 (rs12203592) genes associated with melanoma susceptibility in a Spanish population. *BMC Cancer* 2013;**13**:160.

37. Acquaviva J, Chen X, Ren R. IRF-4 functions as a tumor suppressor in early B-cell development. *Blood* 2008;**112**: 3798—806.

38. Spitz MR, Amos CI, Land S, Wu X, Dong Q, Wenzlaff AS, et al. Role of selected genetic variants in lung cancer risk in African Americans. *J Thorac Oncol* 2013;**8**(4):391—7.

39. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev* 2010;**19**:1893—907.

40. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009;**361**:947—57.

41. Takano T, Fukui T, Ohe Y, Tsuta K, Yamamoto S, Nokihara H, et al. EGFR mutations predict survival benefit from gefitinib in patients with advanced lung adenocarcinoma: a historical comparison of patients treated before and after gefitinib approval in Japan. *J Clin Oncol* 2008;**26**:5589—95.