Prognostic signature of early lung adenocarcinoma based on the expression of ribonucleic acid metabolism–related genes

Ruben Pio, PharmD, PhD,[a,b,c] Jackeline Agorreta, PhD,[a,c,d] and Luis M. Montuenga, PhD[a,c,d]

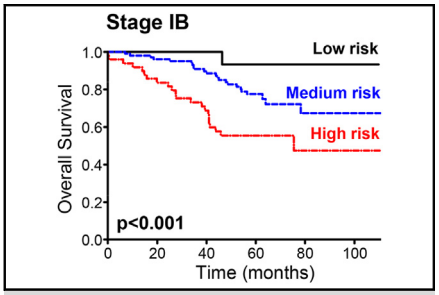
ABSTRACT

Objective: The current staging system for lung cancer is not sufficient to accurately identify those patients with early-stage tumors who would benefit from postsurgery chemotherapy. The objective of this study was to validate a prognostic signature based on the expression of 5 RNA (ribonucleic acid) metabolism-related genes.

Methods: Five lung cancer microarray datasets, 3 from adenocarcinomas and 2 from squamous cell carcinomas, were analyzed. Kaplan-Meier survival curves and Cox proportional hazards models were used to evaluate the relationship between the classifier and recurrence and survival.

Results: Statistically significant differences in relapse-free survival and overall survival were observed when lung adenocarcinoma patients were divided into 3 risk groups. The prognostic information provided by the signature was independent from other demographic and disease variables, including stage. Significant differences in survival were observed between low- and high-risk groups in stage-IB patients: 5-year survival rates ranged from 83% to 100% in the low-risk groups, and from 30% to 71% in the high-risk groups, depending on the dataset. The RNA metabolism score additionally displayed an association with the benefit of adjuvant chemotherapy ($P < .001$), suggesting that those patients in the low-risk group are not good candidates for this treatment.

Conclusions: The RNA metabolism signature is a prognostic marker that may be useful for predicting survival and optimizing the benefit of adjuvant chemotherapy in patients with lung adenocarcinoma. (J Thorac Cardiovasc Surg 2015;150:986-92)



The RNA score accurately predicts overall survival in stage IB lung adenocarcinoma patients.

Central Message

Expression of 5 RNA metabolism–related genes predicts survival, and benefit of adjuvant therapy in resectable lung adenocarcinomas.

Perspective

More than half of patients with early-stage lung cancer have a recurrence after surgery, and prognostic factors are needed to select those patients who would benefit from adjuvant therapy. We describe a prognostic signature, based on the expression of RNA metabolism–related genes, that accurately classifies adenocarcinoma patients, independently of stage, and predicts benefit from adjuvant chemotherapy.

See Editorial Commentary page 993.

Supplemental material is available online.

Lung cancer is the leading cause of cancer mortality, representing >25% of all cancer deaths.[1] Non–small cell lung cancer (NSCLC) accounts for approximately 80% of all lung cancer cases. Currently, only 15% to 20% of NSCLC cancer patients present with localized, potentially curable disease.[2] These numbers are expected to increase in the near future, owing to the implementation of screening programs based on the results of the National Lung Screening Trial.[3] Surgical resection is the treatment of choice for early-stage NSCLC, but several studies have demonstrated that adjuvant chemotherapy can improve survival.[4-6] Based on these studies, adjuvant chemotherapy has become the standard of care for resected stage II, but its use is controversial in stage IB, and is not recommended in stage IA.[7,8] Nevertheless, data suggest that adjuvant chemotherapy is effective for selected patients with stage-I NSCLC.[9]

From the [a]Program in Solid Tumors and Biomarkers, Center for Applied Medical Research (CIMA); [b]Department of Biochemistry and Genetics, School of Science, University of Navarra; [c]Navarra's Health Research Institute (IDISNA); and [d]Department of Histology and Pathology, School of Medicine, University of Navarra, Pamplona, Spain.

This work was supported by the Foundation for Applied Medical Research (FIMA), Instituto de Salud Carlos III-Fondo Europeo de Desarrollo Regional (RD12/0036/0040, PI11/00618, PI13/00806, and PI14/01686), and AECC Scientific Foundation (GCB14-2170).

Received for publication March 17, 2015; revisions received May 18, 2015; accepted for publication June 2, 2015; available ahead of print July 7, 2015.

Address for reprints: Ruben Pio, PharmD, PhD, Program in Solid Tumors and Biomarkers, Center for Applied Medical Research (CIMA), Pio XII Avenue 55, 31008 Pamplona, Spain (E-mail: rpio@unav.es).

0022-5223/$36.00

http://dx.doi.org/10.1016/j.jtcvs.2015.06.001

Abbreviations and Acronyms
CI = confidence interval
DCC = Director's Challenge Consortium (for the Molecular Classification of Lung Adenocarcinoma)
NSCLC = non–small cell lung cancer
RNA = ribonucleic acid

Several studies have reported clinicopathologic and molecular prognostic factors in patients with early-stage NSCLC.[10-12] In search of RNA (ribonucleic acid) metabolism–related genes with altered expression in lung adenocarcinoma, we previously identified a 5-gene categoric classifier with prognostic potential.[13] The biological importance of RNA metabolism in lung adenocarcinomas is evidenced by the identification of aberrant RNA transcripts associated with somatic mutations in genes encoding splicing factors.[14] The main objective of the present study was to validate the prognostic performance of the RNA metabolism signature in independent cohorts of patients with early-stage NSCLC. Our study reveals a substantial capacity of the signature to classify adenocarcinoma patients into various risk groups. In addition, the signature shows an association with the benefit of adjuvant chemotherapy.

METHODS

Patient Cohorts

Five independent NSCLC microarray datasets were used to assess the prognostic capacity of the RNA metabolism signature: 3 with expression data from adenocarcinoma patients, and 2 from patients with squamous-cell carcinoma. Tumor specimens were collected under approval from the respective institutional review boards. All datasets were retrieved from the Gene Expression Omnibus repository (http://www.ncbi.nlm.nih.gov/geo/), except for the files from the Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma (DCC) (https://array.nci.nih.gov/caarray/home.action).

Characteristics of the datasets used in this study are shown in Table 1. Briefly, the DCC dataset was based on expression data from 302 lung adenocarcinoma patients.[15] Four institutions formed the DCC: the University of Michigan Cancer Center, Moffitt Cancer Center, Memorial Sloan-Kettering Cancer Center, and Dana-Farber Cancer Institute. None of the patients received preoperative chemotherapy or radiation. We divided the patients in the DCC database into those who did not receive any adjuvant therapy (n = 213), and those who received adjuvant chemotherapy (n = 89). The GSE31210 dataset contained data from 204 lung adenocarcinoma patients who did not receive postoperative therapy from the National Cancer Center Hospital in Japan.[16] The GSE13213 dataset was obtained from 117 lung adenocarcinoma patients who underwent potential curative resection at Aichi Cancer Center in Japan.[17] None of them received adjuvant chemotherapy.

The GSE17710 dataset contained information on 56 primary tumors surgically extracted at the University of North Carolina from treatment-naive patients with lung squamous cell carcinomas.[18] The fifth NSCLC dataset (GSE4573) contained expression data of resected lung squamous cell carcinomas collected from 129 patients at the University of Michigan Hospital.[19] No information was found about the use of adjuvant therapy in these patients.

Description of the RNA Metabolism Signature

The RNA metabolism signature was derived from a comparison of normal lung tissues with lung adenocarcinomas.[13] Five genes were selected for the prognostic classifier: *ADAR2*, *MARS*, *RAE1*, *SNRPB*, and *SNRPE*. The median expression of each of these genes was used to dichotomize tumors as high or low expressers (the probesets used to assess gene expression are specified in Table E1). *MARS*, *RAE1*, *SNRPB*, or *SNRPE* are overexpressed in lung adenocarcinomas[13]; therefore, tumors with high expression of these 4 genes received 1 risk point for each elevated gene. In the case of *ADAR2*, this gene is down-regulated in lung tumors and may function as a tumor suppressor.[13] Consequently, 1 risk point was assigned to those tumors with low expression of *ADAR2*. The combined risk score was calculated as the sum of the 5 individual scores (ranging from 0 to 5). Patients were divided into 3 risk groups based on the combined risk score: 0: low risk; 1 to 3: medium risk; 4 or 5: high risk.

Statistical Analysis

The clinical endpoints of the cohorts are shown in Table 1. Overall survival was estimated by the Kaplan-Meier method, and differences between curves were compared with the log-rank test. Kaplan-Meier estimates of survival, with their 95% confidence intervals (CIs), are shown in Tables E2-E6. Multivariate analyses were performed using the Cox proportional hazards model. Based on univariate Cox significance levels ($P < .2$), variables were incorporated into Cox models for multivariate analysis. The assumption of proportional hazard was tested by introducing time-dependent covariates into the models. The interaction between the RNA metabolism signature and adjuvant chemotherapy was additionally evaluated, using the Cox proportional hazards model. All statistical analyses were 2-sided. All tests were performed using STATA/IC 12.1 software (Stata Corporation, College Station, Tex).

RESULTS

Lung adenocarcinoma patients included in the DCC dataset showed important differences in prognosis when divided into low-, medium-, and high-risk groups based on a 5-gene RNA metabolism signature.[13] To validate this observation, we used 2 independent microarray datasets of lung adenocarcinoma patients: GSE31210 and GSE13213, based on Affymetrix, Inc and Agilent technologies (both Santa Clara, Calif), respectively. Demographics and clinical characteristics of these cohorts are summarized in Tables E7 and E8.

In the GSE31210 dataset (204 lung adenocarcinoma patients), the RNA metabolism score was associated with risk of recurrence and death (Figure 1, *A* and *B*). Cox regression analyses for relapse-free survival and overall survival are shown in Tables E9 and E10, respectively. The prognostic information provided by the RNA metabolism score was independent of other demographic and disease variables, including stage (Table 2). The 5-year survival rates were 95.8% (95% CI: 73.9%-99.4%) for the low-risk group; 89.6% (95% CI: 81.4%-94.3%) for the medium-risk group; and 73.8% (95% CI: 60.7%-83.2%) for the high-risk group.

Differences in survival were additionally observed among the 117 lung adenocarcinoma patients from the GSE13213 dataset (Figure 1, *C*; and Table E11). A multivariate analysis showed the prognostic importance of the score