Protein profiles associated with survival in lung adenocarcinoma

Guoan Chen*, Tarek G. Gharib*, Hong Wang[†], Chiang-Ching Huang[‡], Rork Kuick[†], Dafydd G. Thomas[§], Kerby A. Shedden[¶], David E. Misek[†], Jeremy M. G. Taylor[‡], Thomas J. Giordano[§], Sharon L. R. Kardia[∥], Mark D. Iannettoni*, John Yee*, Philip J. Hogg**, Mark B. Orringer*, Samir M. Hanash[†], and David G. Beer*[††]

Departments of *Surgery, [†]Pediatrics, [‡]Biostatistics, [§]Pathology, [¶]Statistics, and [∥]Epidemiology, University of Michigan, Ann Arbor, MI 48109; and **Center for Vascular Research, University of New South Wales, Sydney, NSW 2052, Australia

Morphologic assessment of lung tumors is informative but insufficient to adequately predict patient outcome. We previously identified transcriptional profiles that predict patient survival, and here we identify proteins associated with patient survival in lung adenocarcinoma. A total of 682 individual protein spots were quantified in 90 lung adenocarcinomas by using quantitative two-dimensional polyacrylamide gel electrophoresis analysis. A leave-one-out cross-validation procedure using the top 20 survival-associated proteins identified by Cox modeling indicated that protein profiles as a whole can predict survival in stage I tumor patients ($P = 0.01$). Thirty-three of 46 survival-associated proteins were identified by using mass spectrometry. Expression of 12 candidate proteins was confirmed as tumor-derived with immunohistochemical analysis and tissue microarrays. Oligonucleotide microarray results from both the same tumors and from an independent study showed mRNAs associated with survival for 11 of 27 encoded genes. Combined analysis of protein and mRNA data revealed 11 components of the glycolysis pathway as associated with poor survival. Among these candidates, phosphoglycerate kinase 1 was associated with survival in the protein study, in both mRNA studies and in an independent validation set of 117 adenocarcinomas and squamous lung tumors using tissue microarrays. Elevated levels of phosphoglycerate kinase 1 in the serum were also significantly correlated with poor outcome in a validation set of 107 patients with lung adenocarcinomas using ELISA analysis. These studies identify new prognostic biomarkers and indicate that protein expression profiles can predict the outcome of patients with early-stage lung cancer.

Lung cancer is the leading cause of cancer death for both men and women in the United States, with ≈160,000 new cases each year. The 5- and 10-year overall survival rates for patients receiving treatment were 14% and 8%, respectively. Nonsmall cell lung cancer accounts for almost 80% of lung cancers, of which 40% are adenocarcinomas. Although patients diagnosed with stage I adenocarcinoma have an overall 5-year survival rate of 63%, nearly 35% will relapse after surgical resection, thus portending a poor prognosis (1, 2). Identification of these high-risk patients with resectable early-stage disease would provide the opportunity for adjuvant therapy, possibly leading to increased survival.

The relationship of many clinical, pathological, and molecular factors to patient survival in nonsmall cell lung cancer has been investigated. Although prognosis in general correlates with clinical variables, such as stage, it is currently difficult to predict the clinical outcome for individual patients with stage I tumors (3). Several molecular prognostic factors have been proposed, such as K-ras or p53 mutational status, Bcl-2 and c-erbB-2 overexpression (4), and DNA replication errors manifested as microsatellite instability (5). Given the known morphologic and molecular heterogeneity of lung carcinomas and the complex nature of treatment responses, analysis of multiple biologic or molecular markers may be more informative than any single marker (6, 7). We have previously demonstrated the use of mRNA-based gene expression profiles to predict survival in lung adenocarcinoma (7). To date, however, very few prognostic protein markers have been accepted for routine clinical use, either because of conflicting reports or because associations are insufficient for formulation of clinical treatment plans (8).

High-resolution 2D PAGE analysis allows the simultaneous assessment of hundreds of known and unknown polypeptides. In previous studies, we have identified individual proteins and specific protein isoforms that were increased in lung cancer (9, 10). Here, among 682 total proteins quantified, we have identified those that can, when used together in protein expression profiles or as individual protein candidates, predict patient survival in lung adenocarcinomas. Confirmation of candidates was performed by using both mRNA microarrays and tissue microarray (TMA). Phosphoglycerate kinase 1 (PGK1) was found to be strongly predictive of patient survival both as measured in primary tumors or in patients' sera.

Materials and Methods

The procedures that are only briefly described here are detailed in *Supporting Text*, which is published as supporting information on the PNAS web site.

Patient Population. A total of 62 stage I lung adenocarcinomas, 28 stage III lung adenocarcinomas, and 10 nonneoplastic lung tissues were examined by 2D PAGE. A validation set of 90 lung adenocarcinomas and 27 squamous lung carcinomas were examined by using TMAs and immunohistochemistry. An overlapping series of 107 lung adenocarcinoma sera were used for PGK1 ELISA analysis.

Analytical 2D PAGE. All details of the 2D gel methods, protein spot detection, spot quantitation as well as mass spectrometry are published as supporting information on the PNAS web site.

Oligonucleotide Arrays. Expression values for mRNA from 86 lung adenocarcinomas have been previously reported (7). Of these, 76 tumors had available quantitative 2D protein data allowing for analysis of the correlation between mRNA and protein for individual candidates.

Two-Dimensional Westerns and Immunohistochemistry of Tissue Arrays. A detailed description of all antibodies and methods used are published as supporting information as described above.

Serum PGK1 ELISA Analysis. Sera from 107 lung adenocarcinomas, including 66 stage I, 15 stage II, 20 stage III and 6 stage IV, were

---

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

Abbreviations: TMA, tissue microarray; PGK1, phosphoglycerate kinase 1.

[††]To whom correspondence should be addressed at: Department of Surgery, University of Michigan, B560 MSRB II, Box 0686, Ann Arbor, MI 48109. E-mail: dgbeer@umich.edu.

MEDICAL SCIENCES

used for ELISA analysis. Of these 107 patients, 46 tissues were also used in the 2D PAGE analysis set and 17 tissues were used in the TMA validation set in this study. Details of the assay are published as supporting information.

Statistical Analysis. Gels were prepared in batches of 20, and methods used for batch adjustment are described in detail in the supporting information. Survival time was defined as the interval between the day of operation for the lung adenocarcinoma and the date of lung-cancer-related death. Cox proportional hazards regression methods (11) were used to investigate the relationship between survival, clinical–pathological variables, and protein quantities. A leave-one-out cross-validation procedure (7, 12) was used to test the predictive ability of the top 20 ranked proteins, based on P values from the Cox model, to predict low and high risk subsets. For each of the 90 tumors, the remaining 89 were used to select 20 top-ranked proteins and to create a risk index. The risk index is a linear combination of the top 20 protein values multiplied by the coefficient betas from the univariate Cox models. The index was then used to classify the left-out tumor as either low or high risk, depending on whether it was above or below the median (50th percentile cutoff point) of the risk indices of the other 89 tumors. We found that using 20 proteins and a 50th percentile cutoff point provided a conservative estimate of the predictive ability of proteins relative to other numbers of proteins examined (15–25 proteins; see Table 2, which is published as supporting information on the PNAS web site).

Results

Association Between Clinical Variables and Patient Survival. Univariate Cox proportional hazards regression analysis was performed by using 15 separate clinical–pathological variables to determine potential relationships between these variables and overall survival among 90 patients with lung adenocarcinoma (Table 3, which is published as supporting information on the PNAS web site). Six variables were observed to have a significant ($P < 0.05$) relationship with survival: stage ($P < 0.0001$), tumor size ($P = 0.007$), nodal status ($P < 0.0001$), pleural surface involvement ($P < 0.0001$), positive lymphocytic response ($P = 0.01$), and the presence of angiolymphatic invasion ($P = 0.001$). χ^2 test showed tumor size, nodal status, pleural surface involvement, and the presence of angiolymphatic invasion were significantly associated with tumor stage ($P = 0.003$), but lymphocytic response was not (data not shown). Multivariate Cox analysis showed that both stage and lymphocytic response were independently predictive of patient outcome ($P = 0.00007$ and 0.024, respectively). Patient's age, gender, race, smoking status, tumor classification (bronchioloalveolar vs. bronchial-derived), differentiation, tumor location (left lobe vs. right lobe), p53 nuclear accumulation, and K-*ras* mutational status were not correlated with survival (Table 3).

Protein Expression Profiles Predict Survival in Stage I. Univariate Cox proportional hazards regression analysis using all 90 samples and 682 protein spots indicated 46 proteins were associated with patient survival ($P < 0.05$, Table 1). The location of some of these proteins on a representative 2D gel is shown in Fig. 1. We chose to examine the predictive potential of protein profiles by using a leave-one-out cross-validation procedure. In this approach, each possible set of 89 tumors was used to select the top 20 ranked proteins based on P value, and to create a risk index that was used to classify the left-out sample as either low or high risk. By using 20 proteins and a 50th percentile cutoff value, a significantly better survival ($P = 0.005$, log-rank test) for the low-risk group compared with the high-risk group is observed among all samples (Fig. 2*A*). We observed similar significant predictive abilities by using 15–20 proteins at the 40th, 45th or 50th percentile cutoff points (Table 2). The data presented represent a conservative result using 20 proteins and a 50th percentile cutoff point. Among all tumors, there were 7 of 39

(17.9%) deaths among the low-risk group of patients, and 24 of 51 (47.1%) deaths among the high-risk group. In this same procedure, survival was significantly worse in the high-risk stage I group ($P = 0.01$, Fig. 2*B*). Stage III tumors could not be separated into significantly different subgroups as most of these tumors resulted in a poor outcome. These results demonstrate that protein profiles as a whole were predictive of patient outcome in the entire tumor set as well as in stage I patients alone.

Identification of Proteins Associated with Survival. Of the 46 proteins found to be associated with patient survival, 33 were identified by mass spectrometry (Table 1). Thirteen protein spots could not be definitively identified because of very low abundance (Table 4, which is published as supporting information on the PNAS web site). A negative coefficient beta indicates that the protein expression values are associated with a favorable prognosis, whereas a positive coefficient beta is associated with an unfavorable prognosis. For spots with $P < 0.05$, we expected 34 by chance alone and obtained 46. For spots with $P < 0.01$, we obtained 14 and expected 6.8 by chance. Thus, we found more significant proteins than would be expected by chance alone. Table 5, which is published as supporting information on the PNAS web site, lists the proteins showing associations ($P < 0.05$) with clinical–pathological variables as determined by using F test statistics.

Validation of Protein Expression with mRNA Levels. Although protein expression levels predictive of patient outcome may not imply that mRNA measures are similarly associated with outcome, we investigated mRNA levels in both the same set of tumors as well as in an independent set of mRNA and survival data (13). For the samples from the present protein study, oligonucleotide array data (7) was available for 27 genes coding for survival related proteins. Correlation coefficients between protein and mRNA levels are given in Table 1. Of the 27 genes available in the mRNA data, 6 were associated with survival ($P < 0.05$), and in an independent data set from another laboratory of 84 lung adenocarcinomas (13), 7 genes were associated with survival ($P < 0.05$, Table 1). *Phosphoglycerate kinase 1* (*PGK1*) and *triosephosphate isomerase* (*TPI*) mRNA were significant for patient outcome in both data sets. Comparisons of the mRNA values to the protein values from each tumor sample indicate that only PGK1, and specific isoforms of cytokeratins (KRT) 7, 8, and 19 as previously reported (14), were found to be significantly correlated ($P < 0.05$) with their respective mRNA expression levels (Table 1). The lack of correlation among the other protein isoforms may reflect regulation by posttranscriptional or translational mechanisms (15).

Two-Dimensional Western Blot Analysis. To confirm the results from mass spectrometry and to potentially identify additional isoforms of the survival-related proteins, 2D Western blot analysis of A549 lung adenocarcinoma cell lines or primary tumor tissues were performed for the 12 proteins for which antibodies were available (Table 1). One predominant isoform of GRK4 and GRP78, four isoforms of PGAM1, and eight isoforms of PGK1 (Fig. 1*C*) from the tumor 2D gels were detected, including those isoforms identified by mass spectrometry. The GRP78 antibody reacted slightly with three heat shock proteins of the same family (GRP75, HSPA10, and HSPA1b) that were confirmed by mass spectrometry, but their expression levels did not correlate with survival (data not shown). The other immunoreactive PGAM1 protein isoforms had not been quantitatively analyzed because of low abundance. Of the eight PGK1 isoforms detected by 2D Western blot analysis (Fig. 1*C*), three had been analyzed quantitatively and one isoform (no. 826) was significantly correlated with survival (Fig. 1*D* and Table 1).

Antibodies that primarily react with the specific protein isoforms associated with survival were in most instances not available. Antibodies reactive to GFAP and PBP were found to
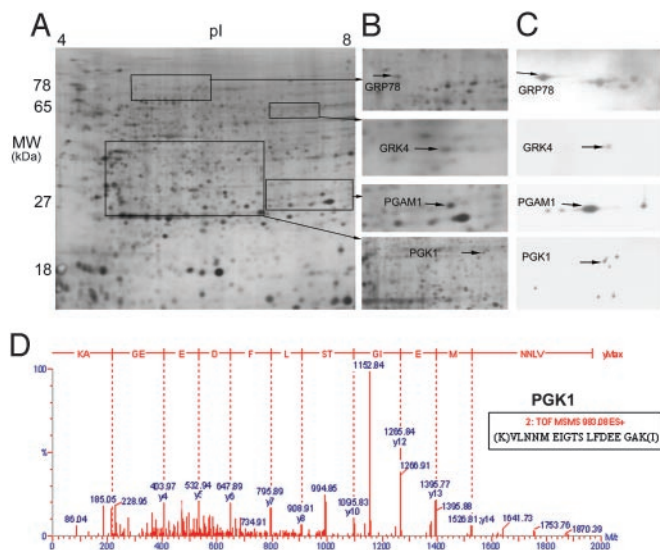
Table 1. Survival-associated proteins and correlation with mRNA levels

Gene name*	Protein spot no.	Coefficient β†	Cox's P value†	Protein name	Estimated protein MW/PI	Native protein MW/PI	Identified protein isoform	Protein function	mRNA_UM Cox's P value‡	mRNA_HV Cox's P value§	Protein–mRNA r¶
GRP78	152	−0.49	0.020	Glucose-regulated protein 78	72.1/5.0	72.3/5.1	Native	Chaperone	0.790	0.573	−0.06
PLCB3	1463	0.43	<0.001	Phospholipase C-β-3	10.9/7.9	138.8/5.7	Degradation	Hydrolase	0.426	0.105	−0.15
ATP5D	1354	−0.40	0.027	F1F0-type ATP synthase subunit d	18/5.4	17.5/5.4	Native	Hydrolase	0.403	0.327	0.09
PGAM1	1160	0.72	<0.001	Phosphoglycerate mutase	28.8/7.0	28.8/6.7	Native	Hydrolase	0.140	**<0.001**	0.20
EPHX1	891	−0.46	0.022	Microsomal epoxide hydrolase	34.2/5.3	52.9/6.8	Degradation	Hydrolase	0.622	**0.026**	−0.06
TPI	1204	−0.38	0.042	Triosephosphate isomerase	26.3/6.5	26.7/6.5	Native	Isomerase	**0.020**	**0.002**	0.02
ENO1B	627	0.32	0.043	α enolase, lung specific	49.4/5.8	49.5/5.8	Native	Lyase	0.722	0.979	0.01
SOD2	1328	0.59	0.003	Superoxide dismutase(MN), mitochondria	24.7/7.6	24.7/7.8	Native	Oxidoreductase	0.239	0.519	0.18
P4HB	320	0.42	0.009	PDI (proly-4-OH-B)	57.8/4.1	57.1/4.8	Phosphorylation	Oxidoreductase	0.606	0.132	0.00
PDIR	513	0.41	0.021	Protein disulfide isomerase related A5	48.4/4.7	59.6/8.1	Degradation	Oxidoreductase	0.740	0.671	−0.03
GSTP1	1298	0.40	0.022	Glutathione S-transferase π	23.7/5.4	23.4/5.4	Native	Transferase	**0.047**	0.069	0.18
HPRT	1205	0.28	0.016	HG phosphoribosyltransferase	24.8/6.2	24.6/6.2	Native	Transferase	0.518	**<0.001**	0.16
PGK1	826	0.45	0.007	Phosphoglycerate kinase 1	41.7/6.3	44.6/8.3	Phosphorylation	Transferase	**0.018**	**<0.001**	**0.37**
PKM2	468	0.35	0.018	Pyruvate kinase, M2	57.8/6.5	57.9/7.9	Phosphorylation	Transferase	0.210	0.083	0.12
GRK4	477	−0.62	0.001	G protein-coupled receptor kinase 4	66.9/7.9	66.9/7.8	Native	Transferase	0.451	0.149	0.19
RAB14	1262	−0.50	0.025	Ras-associated protein RAB14	23.7/6.1	23.9/5.9	Native	GTP-binding protein	NA	0.176	NA
FGF4	1331	0.39	0.021	Fibroblast growth factor 4 (precursor)	22/7	22.1/9.7	Phosphorylation	Growth factor	0.119	0.364	0.15
MYCN	1268	−0.52	0.033	N-myc proto-oncogene protein	23/6.5	49.6/5.5	Degradation	Transcription factor	0.170	0.135	0.22
PBP	1546	−0.52	0.010	Phosphatidylethanolamine-binding protein	13/7.5	21.1/7.0	Degradation	RAF kinase inhibitor	0.514	**0.014**	0.18
AAT	1413	0.46	0.010	α-1-antitrypsin (precursor)	20/6.1	46.7/5.4	Degradation	Proteinase inhibitor	0.215	0.621	0.13
KRT1	1888	0.35	0.030	Cytokeratin 1	28/5.1	66.1/8.2	Degradation	Structural protein	0.868	0.148	0.06
KRT7	871	−0.66	0.005	Cytokeratin 7	41.7/4.7	51.3/5.4	Degradation	Structural protein	**<0.001**	0.234	−0.18
KRT7	1968	0.52	0.014	Cytokeratin 7	41/4.6	51.3/5.4	Degradation	Structural protein	**<0.001**	0.234	0.14
KRT7	2165	−0.49	0.019	Cytokeratin 7	36.7/4.7	51.3/5.4	Degradation	Structural protein	**<0.001**	0.234	**0.35**
KRT7	691	0.41	0.025	Cytokeratin 7	41.7/4.9	51.3/5.4	Degradation	Structural protein	**<0.001**	0.234	0.10
KRT7	2091	0.36	0.029	Cytokeratin 7	40.4/4.6	51.3/5.4	Degradation	Structural protein	**<0.001**	0.234	**0.29**
KRT8	439	0.38	0.033	Cytokeratin 8	50.7/5.4	53.7/5.5	Degradation	Structural protein	0.093	0.368	**0.30**
KRT19	1955	0.43	0.026	Cytokeratin 19	43.6/4.8	44.1/5.1	Native	Structural protein	**0.001**	0.130	**0.39**
GFAP	2336	0.33	0.041	Glial fibrillary acidic protein	28/5.5	49.9/5.4	Degradation	Structural protein	0.597	0.503	−0.10
ACTB/G	634	0.39	0.044	Actin, β/γ	41/5.6	41.7/5.3	Native	Structural protein	0.306‖	0.294‖	0.21‖
ANXA1	1245	0.39	0.026	Annexin I	26/6	38.7/6.6	Degradation	Calcium binding	0.126	0.571	−0.02
ANXA8	923	−0.80	0.006	Annexin VIII	35.9/6.1	36.9/5.6	Native	Calcium binding	**0.009**	0.806	−0.23
CALU	1738	0.51	0.004	Calumenin	38/4.1	37.1/4.5	Native	Calcium binding	NA	**<0.001**	NA

MW/PI, molecular weight†/isoelectric point; NA, not applicable.

*Bold indicates proteins that were verified by 2D Western blot and immunohistochemistry; alternate gene name: GRP78, HSPA5; GRK4, GPRK2L.

†The value of P and coefficient β were made from univariate Cox analysis based on all 90 samples.

‡P values were made from univariate Cox analysis based on 86 samples from University of Michigan; bold values indicate $P < 0.05$.

§P values were made from univariate Cox analysis based on 84 adenocarcinomas from Harvard University; bold values indicate $P < 0.05$.

¶Spearman correlation coefficient based on 76 tumors, bold values indicate $P < 0.05$.

‖Data from one of two ACTB mRNA probe sets were used. Thirteen unidentified protein isoforms are provided in Table 4.

MEDICAL SCIENCES

Fig. 1. Two-dimensional PAGE image and 2D Western blots for selected survival-related proteins. (*A*) Two-dimensional gel image showing protein separation by molecular mass (MW) and isoelectric point (pI). (*B*) Two-dimensional separation of the regions that include GRP78, GRK4, PGAM1, and PGK1 isoforms. (*C*) Two-dimensional Western blot showing GRP78, PGAM1, GRK4, or PGK1 immunoreactive protein spots. The same isoforms for each protein are indicated with arrows for *B* and *C*. (*D*) Tandem mass spectrometry (ESI MS/MS) confirmation for the PGK1 protein spot shown in *B*.

recognize the native-sized proteins, but did not recognize the smaller molecular weight isoforms identified by MS and related to patient survival. Antibodies for FGF4, GST-π, and mEH either reacted nonspecifically or insufficiently to the survival-related protein isoforms on 2D Western blot. Analysis of KRT7, KRT8, and KRT19 isoform expression and relationship to survival were described elsewhere (14).

Confirmation of Tumor-Specific Expression and Relationship to Risk by Using an Independent Set of Tumors. To determine whether the survival-related proteins identified by mass spectrometry and 2D Western blot were expressed in lung tumors, all 12 antibodies were probed on a TMAs containing the 90 lung adenocarcinomas and normal lung samples from this study. Immunohistochemical analysis with these antibodies revealed expression primarily in tumor cells. Among the 12 antibodies, those for PGK1, GRK4, PGAM1, GRP78, and KRT19 recognized at least one of the survival-related isoforms of each protein. As further validation, these five antibodies were used on the TMAs representing an independent set of 90 lung adenocarcinomas, including 69 stage I, 12 stage II, and 9 stage III, and 27 squamous lung tumors. Importantly, PGK1 showed a graded staining pattern with some tumors expressing little, and others expressing very abundant immunoreactivity (Fig. 3 *A–E*). Strong nuclear PGK1 staining was detected in a subset of tumors (Fig. 3*D*). Moreover, abundant PGK1 immunoreactivity was associated with reduced survival in this independent set of lung adenocarcinomas (Fig. 2*C*; *P* = 0.04) as well as in the squamous lung tumors (*P* = 0.04, data not shown). Squamous lung carcinomas showed abundant cytoplasmic and nuclear PGK1 expression similar to adenocarcinomas (Fig. 3*E*).

The antibody for the GRK4 protein recognized one survival-associated isoform by 2D Western blot (Fig. 1*C*) and was found to react at low levels with tumor cytoplasm. However, the strongest immunoreactivity was to neutrophils in normal lung, adenocarcinomas and squamous tumors (Fig. 3 *F–H*). The increased expression of GRK4 protein as determined by 2D PAGE was found to be associated with a favorable outcome
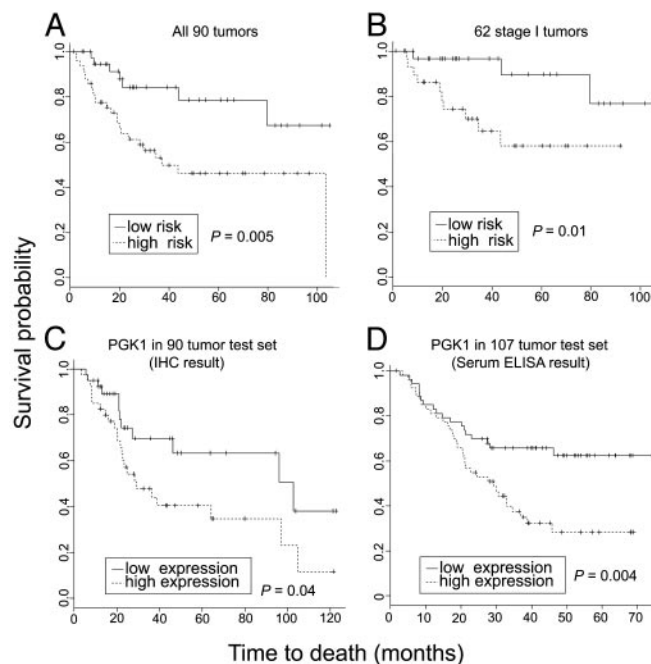
Fig. 2. Protein expression profiles and patient survival. (*A*) Kaplan–Meier survival plots showing the relationship between patient survival and the risk index based on the leave-one-out cross-validation procedure using the top 20 survival-associated proteins among all 682 proteins using all 90 tumors. The high- and low-risk groups differ significantly (*P* = 0.005). (*B*) Relationship between patient survival and the risk index based on the leave-one-out cross-validation procedure using the top 20 survival-associated proteins among the 62 stage I tumors. The high- and low-risk groups differ significantly (*P* = 0.01). (*C*) Relationship between patient survival and PGK1 protein expression in an independent validation set of 90 lung adenocarcinomas. PGK1 immunohistochemical analysis of a tissue array indicates that increased PGK1 is associated with a reduced survival (*P* = 0.04). (*D*) Relationship between patient survival and serum PGK1 levels (ratio of PGK1/total serum protein) by using ELISA analysis with 107 lung adenocarcinomas (*P* = 0.004).

(Table 1), and analysis of GRK4 immunoreactivity in the independent set of tumors also showed a trend for favorable outcome, although it did not reach statistical significance (*P* = 0.1). Interestingly, the mRNA levels of myeloperoxidase, a gene abundantly expressed in neutrophils, was also favorable for survival in the adenocarcinomas (*P* = 0.03). Both PGAM1 and GRP78 were observed to show strong immunoreactivity in tumor cells, but were only marginally predictive of patient outcome in the validation set of lung tumors (*P* = 0.07 for both, Fig. 3 *I* and *J*) possibly because of the difficulty in obtaining precise quantitative data with immunohistochemistry.

PGK1 Protein Is Detected in Patients' Sera. ELISA analysis showed that high levels of serum PGK1 were also significantly correlated with poor outcome in an overlapping series of 107 lung adenocarcinomas (*P* = 0.004, Fig. 2*D*). The level of PGK1 in the serum did not correlate with tumor stage and tumor size (data not shown). Multivariate Cox proportional hazards regression analysis of PGK1 and stage showed that serum PGK1 predicts survival independent of stage (*P* = 0.01, Table 6, which is published as supporting information on the PNAS web site).

Increased Glycolysis Pathway Components Are Associated with Poor Survival. We observed that at least four of the proteins (PGK1, phosphoglycerate mutase, α enolase, and pyruvate kinase M1) that are increased in expression and associated with poor survival (Table 1) are components of the glycolysis pathway (Fig. 4; www.
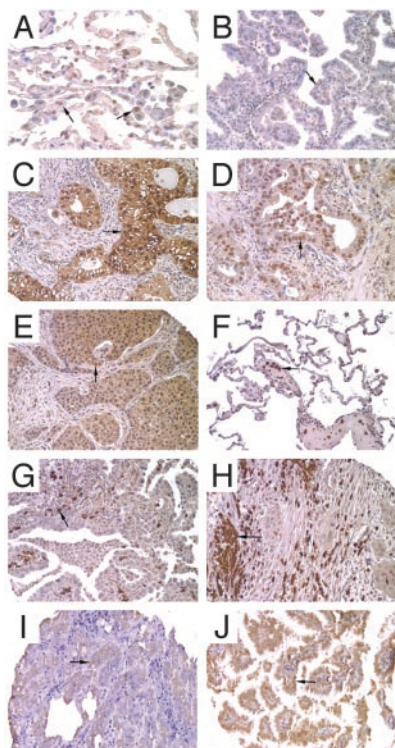
Fig. 3. Immunohistochemical analysis of survival-associated proteins PGK1, GRK4, GRP78, and PGAM1. (*A*) PGK1 expression in normal lung. Low-level staining of alveloar macrophages (arrow) and parenchymal lung cells is observed. (*B*) An adenocarcinoma demonstrating low-level staining of PGK1 (arrow). (*C*) Adenocarcinoma showing abundant cytoplasmic and/or nuclear staining of PGK1 within the tumor cells (arrow). (*D*) Relatively abundant nuclear localization (arrow) of PGK1 immunoreactivity in an adenocarcinoma. (*E*) Both cytoplasmic and nuclear PGK1 staining of a squamous cell carcinoma of the lung. (*F*) GRK4 expression in normal lung is primarily observed in neutrophils as shown located within a large blood vessel (arrow). (*G*) Abundant GRK4 staining in neutrophils within an adenocarcinoma. (*H*) GRK4 staining in neutrophils and in the tumor cells (arrow) in a squamous lung carcinoma. (*I*) GRP78 positive staining in an adenocarcinoma (arrow). (*J*) PGAM1 positive staining (arrow) in an adenocarcinoma. All original magnifications, ×200.

Fig. 4. Proteins and mRNAs significantly correlated with patient survival in lung adenocarcinomas representing components in the glycolysis pathway. Increased expression in the glycolysis pathway enzymes is shown for tumors showing poor patient outcome. The figure is based on results using GENMAPP (www.GenMAPP.org).

GenMapp.org and www.genome.ad.jp/kegg/kegg2.html contain ≈150 pathways). In our mRNA analyses using these same tumors (7), there were 45 probe sets (42 distinct genes) in the glycolysis/gluconeogenesis pathway. Of 538 probe sets significantly associated with survival ($P < 0.05$ in Cox models of 4,966 probe sets), 15 genes were in the glycolysis pathway. The association between survival and this pathway is significant ($P = 0.00004$, Fisher's exact test), and furthermore, 13 of these 15 genes were increased in high-risk patients. When an independent set of oligonucleotide array data of lung adenocarcinoma (13) containing 12,600 probe sets was used, 56 were in the pathway (50 distinct genes), and we found 16 of 1,066 probe sets associated with survival ($P < 0.05$). Here the enrichment of survival-associated genes for pathway members was significant ($P = 0.00001$, Fisher's exact test), and 13 of 16 genes were associated with increased patient risk (see http://dot.ped.med.umich.edu:2000/pub/Lung_prot/index.html for additional information).

Discussion

Tumor stage is an important predictor of patient outcome; however, nearly a third of stage I lung adenocarcinoma patients will not survive 5 years (16, 17). Because these patients receive only surgical resection as the current standard of care, it is important to identify the high-risk stage I patients who may
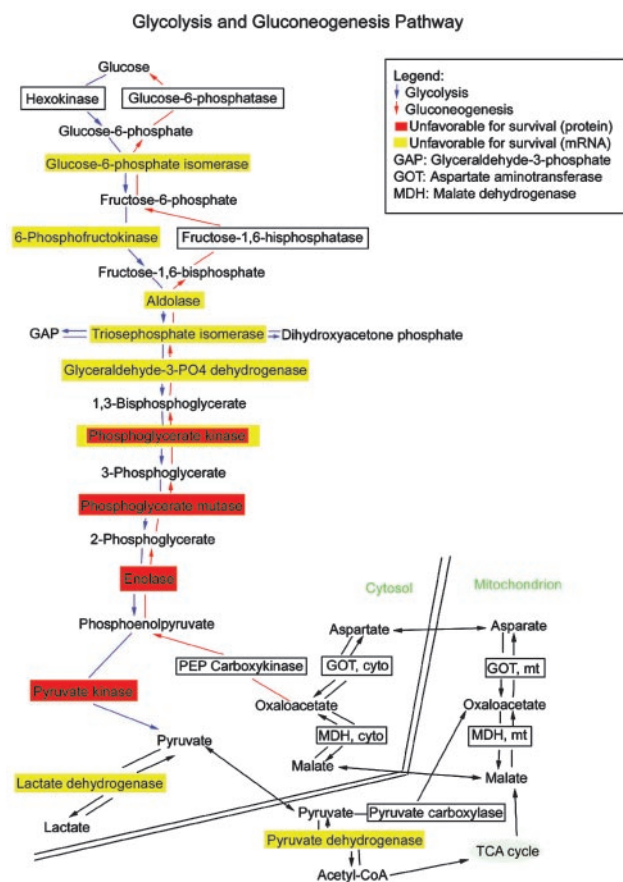
benefit from additional therapy (16, 18). To assess the ability of protein expression profiles to predict patient outcome, we used a leave-one-out cross-validation procedure in which a risk index based on the top 20 ranked proteins determined from each possible set of 89 tumors was used to classify the left-out sample as either low or high risk. Samples classified as low ($n = 39$) and high ($n = 51$) risk by this method differed significantly in their survival, as did low ($n = 32$) and high ($n = 30$) risk stage I tumors (Fig. 2 *A* and *B*). This indicates the potential survival-relatedness of these proteins, and that protein expression profiles in addition to gene expression profiles (7) can identify a subgroup containing high-risk stage I patients.

Based on the quantitative 2D PAGE analysis of 90 lung adenocarcinomas, we found 46 proteins were significantly associated with survival by using a univariate Cox hazard regression analysis. Of the 33 survival-related protein isoforms identified by MS, 14 are enzymes, 10 represent structural proteins, and 8 include proteins with chaperone, growth factor, potential oncogenic, proteinase inhibitory, or calcium-binding properties. GRP78, also known as heat shock 70kD protein 5, was associated with a favorable outcome in our analysis. Although this is consistent with reports of GRP78 expression in lung cancers having a low microvessel density (19), a feature of the less aggressive tumors, and GRP78 association with increased sensitivity to clinically useful chemotherapy agents (20), many reports suggest that this protein is associated with antiapoptosis, cancer progression, and drug resistance (21). A favorable rela-

tionship to survival was also observed for the G protein-coupled-receptor kinase 4 (GRK4). Interestingly, this appears to react primarily with neutrophils. The exact role of neutrophils in these tumors is uncertain; however, this demonstrates the potential for proteomic-based studies using heterogeneous tissue to uncover processes that may be relevant to tumor growth or progression.

In this study, both an individual protein isoform and the mRNA level for PGK1 were significantly associated with an unfavorable survival in lung cancers. We identified multiple isoforms of PGK1 (Fig. 1C) based on immunoreactivity. This may indicate multiple posttranslational or degradation events influencing the expression of PGK1 isoforms in lung adenocarcinomas. PGK1 expression likely reflects increased glycolysis (Fig. 4) in the tumor cells and acts to catalyze the reversible conversion of 1,3, diphosphoglycerate to 3-phosphoglycerate with the generation of one molecule of ATP. PGK1 has also been reported to induce a multidrug-resistant (MDR) phenotype through an MDR-1-independent mechanism (22). The increased expression of genes involved in glycolysis has been associated with the metastatic potential of colon carcinoma cell lines (23). Stable overexpression of human PGK1 in rat mammary adenocarcinoma cells significantly increases tumorigenesis in Balb/C nude mice (E. B. Daly, personal communication, and P. J. Hogg, personal communication). *PGK1* mRNA in our series of lung adenocarcinomas (7) was increased in stage III tumors, and in tumors showing nodal involvement (N1–2), larger tumor size (T2–4), and poor differentiation (data not shown). Increased *PGK1* mRNA significantly associated with poor survival in lung adenocarcinomas from both this and a Harvard University-based study (13), and we observe that *PGK1* mRNA levels are significantly correlated with the expression of the PGK1 isoform predictive of survival. Importantly, levels of immunoreactive PGK1 measured on TMAs and in serum were both predictive of reduced survival in two independent series of lung adenocarcinomas (Fig. 2 C and D). Multivariate Cox proportional hazards regression analysis showed that PGK1 predicts survival independent of tumor differentiation, pleural surface involvement, lymphocytic response, angiolymphatic invasion, and T status in both protein and mRNA data sets. The serum PGK1 predicts survival independent of stage (Table 6).

A highlight of our findings was that both proteins and mRNA studies relate genes responsible for increased glycolysis to poor patient outcome (Fig. 4). Importantly, four proteins and seven mRNAs encoding enzymes in the glycolysis pathway were all increased in expression and associated with poor survival, demonstrating the potential of integrating both types of expression

analyses. Only PGK1 was significantly related at both mRNA and protein levels, possibly reflecting regulation differences for each enzyme in this pathway. PGK1 is controlled by oxygen tension (24) and increased expression may reflect faster growing and more hypoxic tumors. Consistent with induction of the glycolysis pathway, we observed that both vascular endothelial growth factor and IGFBP3 are increased in these tumors (7), supporting a hypoxia-inducible factor 1α-mediated pathway (24, 25) as related to poor patient outcome in lung adenocarcinomas. Proteins such a phosphoglycerate mutase (PGAM1), an enzyme that catalyzes the reversible reaction of 3-phosphoglycerate to 2-phosphoglycerate in the glycolytic pathway, was elevated in tumors showing pleural surface involvement and angiolymphatic invasion (Table 5), consistent with unfavorable prognosis. PGAM1 is expressed in multiple cancer types including lung cancer (26). PGAM1 immunoreactivity was observed in lung tumor cells, and it was marginally predictive of survival in our validation lung tumor series.

We have used 2D PAGE and mass spectrometry to identify proteins associated with both patient survival and with specific clinical–pathological features of lung adenocarcinomas. Many of these proteins represent specific isoforms of known proteins and reflect posttranslational modifications and potential degradation forms. Additional research is required to understand the underlying basis for cleavage or degradation of these proteins. A number of these proteins appear to reflect the increased metabolic properties of aggressive lesions. A risk index derived from the 20 proteins most associated with survival identified a subset of stage I patients with a significantly worse clinical outcome. We have identified individual survival-related proteins such as PGK1 that predict patient outcome by using serum ELISA with an independent series of tumors. Development of reagents for accurate quantification of specific protein isoforms associated with survival could provide useful predictive biomarkers as well as indicate potential targets for therapy.

1. Hoffman, P. C., Mauer, A. M. & Vokes, E. E. (2000) *Lancet* **355,** 479–485.
2. Fry, W. A., Philips, J. L. & Menck, H. R. (1999) *Cancer* **86,** 1867–1876.
3. Martins, S. J. & Pereira, J. R. (1999) *Am. J. Clin. Oncol.* **22,** 453–457.
4. Brundage, M. D., Davies, D. & Mackillop, W. J. (2002) *Chest* **122,** 1037–1057.
5. Rosell, R., Pifarre, A., Monzo, M., Astudillo, J., Lopez-Cabrerizo, M. P., Calvo, R., Moreno, I., Sanchez-Cespedes, M., Font, A. & Navas-Palacios, J. J. (1997) *Int. J. Cancer (Pred. Oncol.)* **74,** 330–334.
6. Schneider, P. M., Praeuer, H. W., Stoeltzing, O., Boehm, J., Manning, J., Metzger, R., Fink, U., Wegerer, S., Hoelscher, A. H. & Roth, J. A. (2000) *Br. J. Cancer* **83,** 473–479.
7. Beer, D. G., Kardia, S. L., Huang, C. C., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G., *et al.* (2002) *Nat. Med.* **8,** 816–824.
8. Alaiya, A. A., Franzen, B., Auer, G. & Linder, S. (2000) *Electrophoresis* **21,** 1210–1217.
9. Chen, G., Gharib, T. G., Huang, C. C., Thomas, D. G., Shedden, K. A., Taylor, J. M., Kardia, S. L., Misek, D. E., Giordano, T. J., Iannettoni, M. D., *et al.* (2002) *Clin. Cancer Res.* **8,** 2298–2305.
10. Chen, G., Wang, H., Gharib, T. G., Huang, C. C., Thomas, D. G., Shedden, K. A., Kuick, R., Taylor, J. M., Kardia, S. L., Misek, D. E., *et al.* (2003) *Mol. Cell. Proteomics* **2,** 107–116.
11. Cox, D. R. (1972) *J. R. Stat. Soc.* **34,** 187–220.
12. Stone, M. (1977) *Biometrika* **64,** 29–38.
13. Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 13790–13795.
14. Gharib, T. G., Chen, G., Wang, H., Huang, C. C., Prescott, M. S., Shedden, K., Misek, D. E., Thomas, D. G., Giordano, T. J., Taylor, J. M., *et al.* (2002) *Neoplasia* **4,** 440–448.
15. Chen, G., Gharib, T. G., Huang, C. C., Taylor, J. M., Misek, D. E., Kardia, S. L., Giordano, T. J., Iannettoni, M. D., Orringer, M. B., Hanash, S. M., *et al.* (2002) *Mol. Cell. Proteomics* **1,** 304–313.
16. Williams, D. E., Pairolero, P. C., Davis, C. S., Bernatz, P. E., Payne, W. S., Taylor, W. F., Uhlenhopp, M. A. & Fontana, R. S. (1981) *J. Thorac. Cardiovasc. Surg.* **82,** 70–76.
17. Pairolero, P. C., Williams, D. E., Bergstralh, E. J., Piehler, J. M., Bernatz, P. E. & Payne, W. S. (1984) *Ann. Thorac. Surg.* **38,** 331–336.
18. Naruke, T., Goya, T., Tsuchiya, R. & Suemasu, K. (1988) *J. Thorac. Cardiovasc. Surg.* **96,** 440–447.
19. Koomagi, R., Mattern, J. & Volm, M. (1999) *Anticancer Res.* **19,** 4333–4336.
20. Belfi, C. A., Chatterjee, S., Gosky, D. M., Berger, S. J. & Berger, N. A. (1999) *Biochem. Biophys. Res. Commun.* **257,** 361–368.
21. Lee, A. S. (2001) *Trends Biochem. Sci.* **26,** 504–510.
22. Duan, Z., Lamendola, D. E., Yusuf, R. Z., Penson, R. T., Preffer, F. I. & Seiden, M. V. (2002) *Anticancer Res.* **22,** 1933–1941.
23. Lay, A. J., Jiang, X. M., Kisker, O., Flynn, E., Underwood, A., Condron, R. & Hogg, P. J. (2000) *Nature* **408,** 869–873.
24. Lu, H., Forbes, R. A. & Verma, A. (2002) *J. Biol. Chem.* **277,** 23111–23115.
25. Giatromanolaki, A., Koukourakis, M. I., Sivridis, E., Turley, H., Talks, K., Pezzella, F., Gatter, K. C. & Harris, A. L. (2001) *Br. J. Cancer.* **85,** 881–890.
26. Durany, N. Joseph, J., Campo, E., Molina, R. & Carreras, J. (1997) *Br. J. Cancer* **75,** 969–977.