

*Biomarkers, Genomics, Proteomics, and Gene Regulation*

# Protein Signatures for Classification and Prognosis of Gastric Cancer

## A Signaling Pathway-Based Approach

Daguang Wang,\* Fei Ye,<sup>†</sup> Yabin Sun,<sup>‡</sup> Wei Li,\* Hongyi Liu,\* Jing Jiang,<sup>§</sup> Yang Zhang,\* Chengkui Liu,\* Weihua Tong,\* Ling Gao,<sup>¶</sup> Yezhou Sun,<sup>||</sup> Weijia Zhang,<sup>||</sup> Terry See-<sup>†</sup> Peng Lee,\*\* Jian Suo,\* and David Y. Zhang\*<sup>†</sup>

From the Departments of General Surgery,\* Ophthalmology,<sup>‡</sup> BioBank,<sup>§</sup> and Hematology and Oncology;<sup>¶</sup> The First Hospital, Jilin University, Changchun, China; the Departments of Pathology<sup>†</sup> and Medicine,<sup>||</sup> Mount Sinai School of Medicine, New York, New York; and the Departments of Pathology, Urology, and the New York University Cancer Institute,\*\* New York University, School of Medicine, New York, New York

**Current methods have limited accuracy in predicting survival and stratifying patients with gastric cancer for appropriate treatment. We sought to identify protein signatures of gastric cancer for classification and prognostication. The Protein Pathway Array (initial study) and Western blot (confirmation) were used to assess the protein expression in a total of 199 fresh frozen gastric samples. There were 56 paired samples divided into a training set ( $n = 37$ ) and a validation set ( $n = 19$ ) for the identification of differentially expressed proteins between tumor and normal tissues. There were 56 tumor samples used to identify proteins correlating with tumor and nodal staging. All 93 tumor samples were used to identify candidate proteins for predicting survival. We confirmed the survival prediction of the candidate proteins by using an additional cohort of gastric cancer samples ( $n = 50$ ). There were 22 proteins differentially expressed between normal and tumor tissues. Nine proteins were selected to build the predictor to classify normal and tumor samples. Ten proteins were differentially expressed among different T stages and four of these were associated with invasive behavior. An additional four proteins were associated with lymph node metastasis. Two proteins were identified as independent risk factors for overall survival. This study indicated that some dysregulated signaling proteins could be selected**

**as useful biomarkers for tumor classification and predicting outcome in gastric cancer patients. (Am J Pathol 2011, 179:1657–1666; DOI: 10.1016/j.ajpath.2011.06.010)**

Gastric cancer is the fourth most common malignancy and ranked as the second leading cause of cancer death worldwide.<sup>1</sup> The geographic distribution of incidence and mortality of gastric cancer varies remarkably worldwide. Areas with high incidence include Japan, Korea, China, Eastern Europe, and parts of Latin America. The mortality of gastric cancer has declined in past decades, mainly due to early detection by gastric endoscopy.<sup>2</sup> However, unlike that of other common cancers, the prognosis for most gastric cancer is poor and has improved little for the past several decades. Despite recent advances in chemotherapy and surgical techniques, the overall 5-year survival rate is lower than 40%.<sup>3</sup> Perplexingly, the prognosis varies widely in patients with stage II or III disease for undetermined biological reasons.

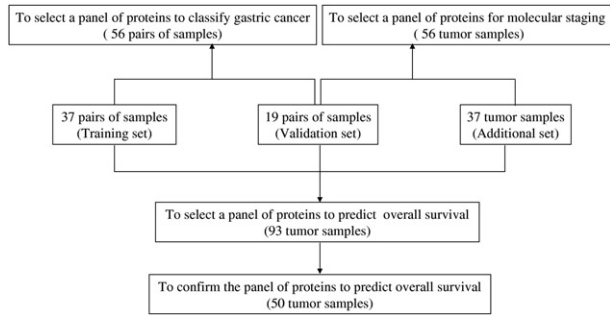
This study was partially supported by funding from The First Hospital, Jilin University, Changchun, China.

Accepted for publication June 27, 2011.

D.Z., J.S., and F.Y. contributed equally by conceiving the hypothesis and overall study design. D.W., D.Z., and F.Y. contributed equally in the experimental design, and D.W. and Y.S. contributed equally by performing the Protein Pathway Array assay. Y.S., H.L., and W.T. contributed equally to protein preparation. J.J. and J.S. contributed equally to the specimen collection and, together with Y.Z. and C.L., contributed equally by collecting the clinical data. Y.S., L.G., W.Z., and T.S. contributed equally to statistical analysis and D.W. and F.Y. contributed equally to summarizing the final data. D.W. and D.Z. contributed equally to preparing the manuscript.

Supplemental material for this article can be found on <http://ajp.amjpathol.org> or at doi: 10.1016/j.ajpath.2011.06.010.

Address reprint request to Jian Suo, M.D., Ph.D., Department of General Surgery, The First Hospital, Jilin University, Changchun, Jilin, 130021 China; or David Y. Zhang, M.D., Ph.D., Department of Pathology, Mount Sinai School of Medicine, New York, NY 10029. E-mail: [suojian42@gmail.com](mailto:suojian42@gmail.com) or [david.zhang@mssm.edu](mailto:david.zhang@mssm.edu).



**Figure 1.** Overall study design. Three sets of samples were used to select panels of proteins for classifying normal and tumor samples, staging the invasion and lymph node metastasis, and predicting overall survival.

Currently, prognosis of gastric cancer is based on pathology (ie, histology type, invasion, and metastasis), radiological imaging (for staging), and other clinical factors (age and comorbidity), which all determine how patients should be managed (surgery and subsequent chemotherapy). However, these traditional clinicopathological factors have significant limitations. Therefore, a large effort has been made to search for molecular markers for diagnosis, classification, and prognosis of gastric cancer.<sup>4-7</sup> For example, cell cycle regulation factors (p27 and cyclin E),<sup>4,5</sup> cell adhesion molecules (E-cadherin),<sup>6</sup> oncogenes (c-erbB2 and c-myc),<sup>7</sup> and tumor suppressor genes (p53)<sup>7</sup> have been reported to correlate with the prognosis of gastric cancer patients. Despite these reports, inconsistent results exist among the different studies, and the reported parameters provide limited information on the prognosis of individual patients because of the complex biology of the disease.<sup>8</sup>

In this study, we attempted to screen for proteins that can be used for diagnosis and prognosis of gastric cancer using the Protein Pathway Array method, a multiplex immunoblot-based assay combined with computational analysis.<sup>9</sup> The Protein Pathway Array is a novel proteomic method that can characterize hundreds of proteins in clinical samples and identify alterations in protein expression or abundance with biomarker potential. We applied this unique approach to identify differentially expressed signal transduction proteins in gastric cancer tissue. Because the dysregulation of signal transduction proteins is responsible for cancer development, these proteins can be used as a signature for the diagnosis and prognosis of gastric cancer. Using this approach, we successfully identified a panel of nine proteins for distinguishing gastric cancer, four proteins associated with invasion, and two proteins for prognosis of survival.

**Materials and Methods**

*Patients and Tissue Samples*

Fifty-six pairs of gastric cancer and adjacent nontumor mucosa (37 in the training set and 19 in the validation set), and an additional 87 cancer tissues (37 in the additional set and 50 in the second cohort) (Figure 1) were obtained after informed consent from patients who un-

derwent D2 gastrectomy (ie, radical gastrectomy with level 2 extended lymphadenectomy) between February 2008 and June 2009 at The First Hospital of Jilin University, Jilin, China. This study was reviewed and approved by The First Hospital of Jilin University's Institution Ethical Review Boards.

The representative tumors and adjacent normal tissues of these patients were dissected and frozen within 30 minutes of removal in a liquid nitrogen tank after immediate pathological examination. Tumor samples of 3 × 3 × 5 mm<sup>3</sup> were taken from areas without gross necrosis. Adjacent nontumor mucosa samples of 3 × 3 × 5 mm<sup>3</sup> were taken from the same patient at 3 cm away from the tumor margin. The tumor samples did not contain normal mucosal tissue, except for occasional entrapped gastric glands. The mucosa samples contained mucosa and a part of adherent submucosa; neither tumor nor dysplasia was included.<sup>9</sup>

The clinicopathological data of the patients are summarized in Table 1. A total of 143 patients (137 advanced and 6 early gastric cancers) were included (93 initial

**Table 1.** Patient Demographics and Gastric Cancer Characteristics

Clinicopathological characteristics	Patient number (%)	
	First cohort (n = 93)	Second cohort (n = 50)
Age		
≤60 years	38 (41)	17 (34)
>60 years	55 (59)	33 (66)
Sex		
Male	73 (78)	24 (48)
Female	20 (22)	26 (52)
Family history		
Yes	12 (13)	13 (26)
No	81 (87)	37 (74)
Histology		
Histological grade		
Moderately differentiated adenocarcinoma	30 (32)	15 (30)
Poorly differentiated adenocarcinoma	63 (68)	35 (70)
Vascular invasion		
Yes	60 (65)	31 (62)
No	33 (35)	19 (38)
AJCC TNM stage*		
I	15 (16)	0 (0)
II	16 (17)	10 (20)
III	39 (42)	39 (78)
IV	23 (25)	1 (2)
Primary tumor		
T1	6 (6)	0 (0)
T2	19 (20)	0 (0)
T3	64 (69)	50 (100)
T4	4 (4)	0 (0)
Node status		
N0	23 (24)	0 (0)
N1	26 (28)	5 (10)
N2	22 (24)	35 (70)
N3	22 (24)	10 (20)
Metastasis		
M0	92 (99)	50 (100)
M1 <sup>†</sup>	1 (1)	0 (0)

\*According to the American Joint Committee on Cancer (AJCC).<sup>10</sup>  
<sup>†</sup>Metastasis to liver.

samples and 50 second cohort samples). One hundred and twenty patients had regional lymph node metastasis and one patient had distant metastasis (liver) at the surgery. The TNM stage of the tumor was done according to the American Joint Committee on Cancer.<sup>10</sup>

### Summary of Experimental Design

The proteins from 199 samples (56 paired samples and 87 unpaired tumors) were extracted, with 149 of them being used to assess the level of protein expression and phosphorylation using the Protein Pathway Array, and 50 of them being used to detect the expression levels of two candidate proteins using Western blot (Figure 1). Fifty six paired tumors and adjacent normal tissues were used to select the protein panel to distinguish between normal and tumor tissues of gastric cancer. These 56 pairs of samples were divided into a training set ( $n = 37$ ) and a validation set ( $n = 19$ ). A total of 56 tumor samples (including 19 tumor samples from the validation set and additional 37 new tumor samples) were used to identify the protein panel to distinguish different TNM stages. All 93 tumor samples were used to assess the candidate proteins for predicting survival. An additional cohort of gastric cancer samples ( $n = 50$ ) were used to confirm the ability of candidate proteins to predict survival.

### Protein Pathway Array Analysis

Total proteins were extracted from the 149 fresh frozen gastric samples using 1× sample lysis buffer (Cell Signaling Technology, Danvers, MA) containing 20 mmol/L Tris-HCL (pH 7.5), 150 mmol/L NaCl, 1 mmol/L Na<sub>2</sub>EDTA, 1 mmol/L EGTA, 1% Triton, 2.5 mmol/L sodium pyrophosphate, 1 mmol/L β-glycerophosphate, 1 mmol/L Na<sub>3</sub>VO<sub>4</sub>, and 1 μg/mL leupeptin in the presence of 1× proteinase inhibitor cocktail (Roche Applied Science, Indianapolis, IN) and 1× phosphatase inhibitor cocktail

(Roche Applied Science). The lysate was sonicated 3 times for 15 seconds each, and then centrifuged at 14,000 rpm for 30 minutes at 4°C. The protein concentration was determined with the BCA Protein Assay kit (Pierce, Rockford, IL). Approximately 300 μg of protein lysate was loaded in one well across the entire width of 10% SDS polyacrylamide and separated by electrophoresis, as previously described.<sup>11</sup> After electrophoresis, the proteins were transferred electrophoretically to a nitrocellulose membrane (Bio-Rad, Hercules, CA), which was then blocked for 1 hour with blocking buffer including either 5% milk or 3% bovine serum albumin in 1× Tris-HCl, NaCl, and Tween 20 (TBST) containing 20 mmol/L Tris-HCl (pH 7.5), 100 mmol/L NaCl, and 0.1% Tween-20. Next, the membrane was clamped on a Western blotting manifold (Mini-PROTEAN II Multiscreen Apparatus, Bio-Rad, Hercules, CA) that isolates 20 channels across the membrane. The multiplex immunoblot was performed using a total of 142 protein-specific or phosphorylation site-specific antibodies (Table 2). Four sets of antibodies (a total of 36 to 38 protein-specific or phosphorylation site-specific antibodies per set) were individually used for each membrane, and all of the antibodies (from various companies) were validated independently before inclusion in the Protein Pathway Array. For the first set of 36 primary antibodies, a mixture of two antibodies in the blocking buffer were added to each channel and then incubated at 4°C overnight. The membrane was then washed with 1× Tris-buffered saline and 1× TBST, and was further incubated with secondary anti-rabbit or anti-mouse antibody conjugated with horseradish peroxidase (Bio-Rad) for 1 hour at room temperature. The membrane was developed with chemiluminescence substrate (Immun-Star HRP Peroxide Buffer/Immun-Star HRP Luminol Enhancer, Bio-Rad), and chemiluminescent signals were captured using the ChemiDoc XRS System (Bio-Rad). The same membrane was then stripped off using stripping buffer (Restore Western Blot Stripping

**Table 2.** List of Antibodies Included in the Protein Pathway Array

Antibodies specific for phosphorylation

p-PKCα(Ser657), p-EGFR (Tyr1068), p-HER2/ERBB2 (Tyr1221/1222), p-PDK1 (Ser241), p-PKCα/βII (Thr638/641), p-p53 (Ser392), p-Akt (Ser473), p-PTEN (Ser380), p-Rb (Ser780), p-survivin (Thr 34), p-beta-catenin (Ser33/37/Thr41), p-STAT5 (Tyr694), p-STAT3 (Ser727), p-ERK (Thr202/Tyr204), p-p70 S6 kinase (Thr389), p-VEGFR-2 (Tyr951), p-FGFR (Tyr653/654), p-EIF4B (Ser422), p-HGFR/C-Met (Y1234/Y1235), p-Smad (Ser463/465), p-ERK5 (Thr218/Tyr220), p-p90RSK (Ser380), p-CREB (Ser133), p-mTOR (Ser2448), p-PKCδ(Thr505), p-CDC2 (Tyr15), p-c-Jun (Ser73), p-SAPK/JNK (Thr183/Tyr185), p-FLT3 (Tyr 591), p-p38 (Thr180/Tyr182), p-GSK-3α/β(Ser21/9), p-FAK (Tyr397), p-RB (Ser807/811), p-HGFR/C-Met (Y1003).

Antibodies for signal transduction proteins

CyclinB1, cyclinD1, CDK6, CDC25B, cyclinE, CDK2, p27, BRCA1, CDK4, neu, 14-3-3 beta, cPKCα, ERK, EGFR, WEE1, CDC25C, HSP90, CHK1, MDM2, CDC2 p34, E2F-1, PCNA, c-myc, Notch1, beta-catenin, Akt, Trap, XIAP, Bcl-2, ETS1, HIF-1α, HIF-2α, TTF-1, p53, Notch4, PTEN, SRC-1, p300, c-Kit, Bax, N-cadherin, Raf-1, CDC42, EIF4B, TNF-α, vimentin, OPN, survivin, E-cadherin, TGF-β, p16, p27, WT1, Mesothelin, Cleaved Caspase-3, COX2, ATF-1, CREB, p21, NF-κB52, NF-κB50, calretinin, H-Ras, Bcl-6, K-Ras, alpha-tubulin, NF-κB p65, Myf-6, p15, ATR, Fas, SUMO-1, MetRS, Ep-CAM, FOXM1, Era, SYK, STAT1, Eg5, HIF-3α, RAD52, ATM, ABCG2, Bad-7, KLF6, CaMKKα, Topo IIa, p38, IL-1β, TERT, Ub, PR, Rap1, HCAM, Lyn, twist, TAP, patched, Erb, VEGF, GLI-3, FGF-7, p63, SK3, rhoB, WNT-1, TDP1, SLUG.

Underlines indicate detectable expression in either tumor or normal tissues.

All phosphorylation state-specific antibodies were obtained from Cell Signaling Technology (Danvers, MA), except p-HGFR/C-Met (Y1234/Y1235) and p-HGFR/C-Met (Y1003), which were purchased from R&D Systems (Minneapolis, MN).

All non-phospho-antibodies were obtained from Santa Cruz Biotechnology (Santa Cruz, CA), except the following antibodies: 1) ERK, Akt, beta-catenin, Notch4, CREB, Cleaved Caspase-3, EIF4B, NF-κB52, NF-κB50, and STAT1 were obtained from Cell Signaling Technology (Danvers, MA); 2) XIAP was obtained from BD Biosciences (San Jose, CA); 3) TGF-β was obtained from R&D Systems (Minneapolis, MN).

TFG, transforming growth factor.

Buffer, Thermo Scientific, Rockford, IL) and then used to detect a second set of 36 primary antibodies as previously described. The signal of each protein were determined by densitometric scanning (Quantity One software package, Bio-Rad).

The background was locally subtracted from raw protein signal and the background subtracted intensity was normalized by the "global median subtraction" normalization method to reduce the variations arising from experimental results derived from different runs (such as transferring and blotting efficiency, total protein loading amount, and exposure density). In detail, for each protein, its intensity was divided by total intensities of all proteins from each sample, and then multiplied by average intensities of all proteins in all samples. The normalized data were transformed to log<sub>2</sub> and were used in the subsequent statistical analysis.

### Western Blot Analysis

Total proteins were extracted from 50 fresh frozen gastric cancer samples, as previously described. There were 20 μg of proteins that were fractionated by electrophoresis through a 10% SDS-polyacrylamide gel electrophoresis, and then the proteins were transferred onto a nitrocellulose membrane. The membrane was incubated with the primary antibodies, including Akt (1:1000 dilution) and cyclin-dependent kinase 2 (1:1000 dilution) (both from Santa Cruz Biotechnology, Santa Cruz, CA) at 4°C overnight. The membrane was then incubated with a secondary anti-rabbit antibody conjugated with horseradish peroxidase (Amersham, Arlington Height, IL). The protein was detected using chemiluminescence method and chemiluminescent signals were captured using the ChemiDoc XRS System (Bio-Rad), as previously described. The same membrane was then blotted using a monoclonal anti-β-actin antibody (1:10,000 dilution; Sigma, St. Louis, MO). The signal of each protein was determined by densitometric scanning (Quantity One software package, Bio-Rad).

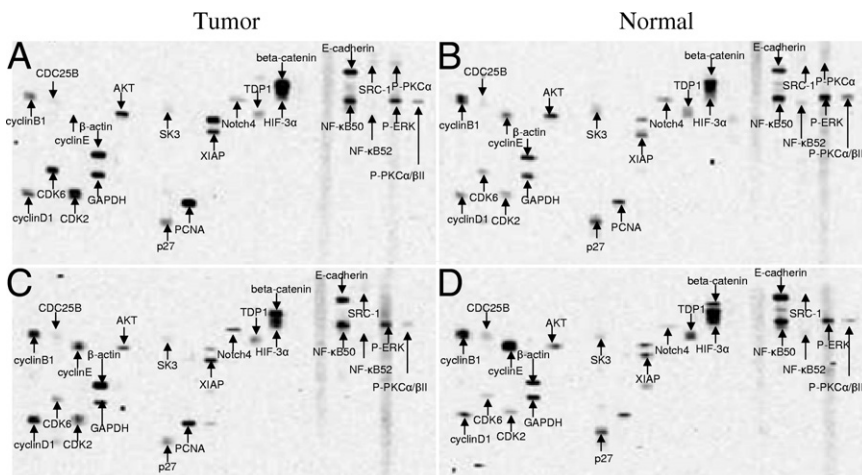
### Statistical Analysis

Paired Student's *t*-test and Significant Analysis of Microarray (SAM) tool (<http://www-stat.stanford.edu/~tibs/SAM>) were used to select the proteins differentially expressed between tumors and normal tissues. K-fold cross validation (K = 10) was used to select those proteins with a great discriminating power to distinguish tumors from normal tissues. K-fold cross validation and unsupervised hierarchical clustering analysis were performed using BRB Array Tools software v.3.3.0 (<http://linus.nci.nih.gov/BRB-ArrayTools.html>). SPSS v.17.0 software (SPSS Inc., Chicago, IL) was used for Cox proportional hazard regression analysis to correlate the Protein Pathway Array data with the clinical data (TNM and survival), as well as for Kaplan-Meier and log-rank analysis of overall survival.

## Results

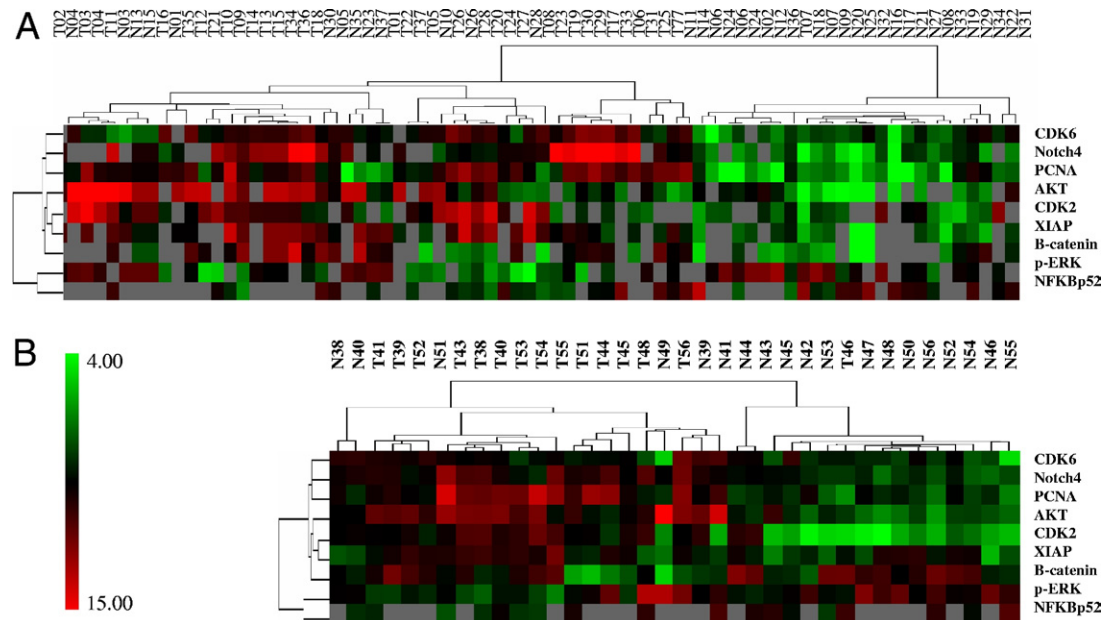
### Identification of Differentially Expressed Signaling Proteins in Gastric Cancer

There were 22 (of 142) proteins found to be differentially expressed between tumors and normal tissues in the training set (37 paired samples) using paired *t*-test and SAM analysis ( $P < 0.05$  or  $q < 5\%$ ) (see Supplemental Table S1 at <http://ajp.amjpathol.org>) (Figure 2). Among them, 9 proteins and phosphoproteins were up-regulated in tumors, including proliferating cell nuclear antigen (PCNA), Notch4, CDK4, CDK6, XIAP, p-protein kinase C (PKC)α/βII, Akt, β-catenin, and p-PKCα, and 13 proteins were down-regulated in tumors, including p-ERK, cyclin B1, cyclin E, p27, E-cadherin, Hypoxia-inducible factor (HIF)-3α, Cdc25B, NF-κB52, TDP1, SK3, NF-κB50, SRC-1, and cyclin D1. To identify a robust set of proteins for classification, we carried out supervised K-fold cross validation (K = 10) using two class prediction models, including a support vector machine (SVM) and 3-nearest neighbor (3NN). Nine proteins (PCNA, Notch 4, p-ERK, CDK6, X-linked inhibitor of apoptosis (XIAP), CDK2, Akt, β-catenin, and NF-κB52) with the value of  $P < 0.01$  were



**Figure 2.** Representative autoradiographs showing expression and phosphorylation of the signaling-related proteins detected by Protein Pathway Array. (A and B) Paired tumor and normal tissues. (C and D) Paired tumor and normal tissues.





**Figure 3.** Hierarchical clustering analysis of differentially expressed proteins in paired tumor and normal samples. The expression profile of nine proteins between the paired tumor and normal samples in training set ( $n = 37$ ) (A) and validation set ( $n = 19$ ) (B). The color scale showed the level of expression. Red indicates overexpression and green indicates underexpression, black indicates no change, and gray no expression. The number in each column represents the sample number. Each row represents a protein.

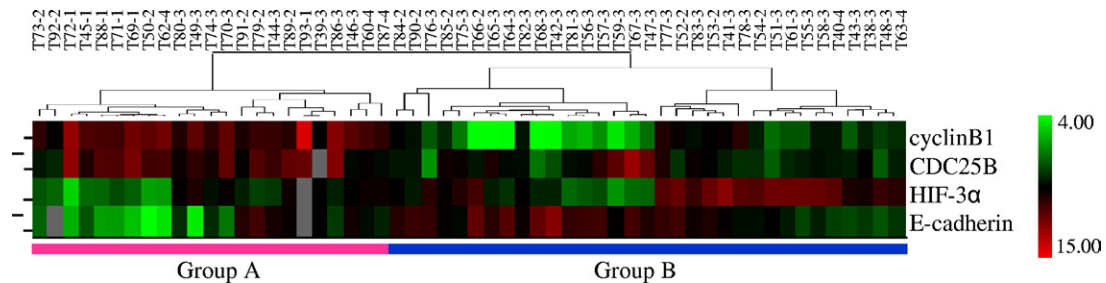
selected to build the SVM predictor. Five proteins (PCNA, Notch 4, p-ERK, CDK6, and XIAP) were selected to build the 3NN predictor. Ninety seven percent of the samples in the training set were correctly classified by either SVM or 3NN modeling. Only two samples (1 pair) in the training set were misclassified with this model. To further confirm the ability of these nine proteins to classify gastric cancer, we tested these proteins using a separate validation set of specimens (19 pairs) by 3NN and SVM models as previously described. All samples in the validation set were correctly classified by 3NN modeling (100% sensitivity and specificity), but 1 pair of the samples were misclassified by SVM modeling (95% sensitivity and specificity). A two-way hierarchical clustering analysis was performed for both sets of samples and revealed distinct patterns for both training sets (Figure 3A) and validation set (Figure 3B), although several samples were misclassified. We also compared the protein expression pattern between two histology grades (moderately and poorly differentiated tumors), and no significant difference was found.

### Correlation of Protein Expression with Clinical Behaviors of Gastric Cancer

To identify molecular markers to predict gastric cancer behaviors (ie, invasion and lymph node metastasis), we applied the SAM tool to identify proteins differentially expressed among different tumor groups. Based on the pathology report, we classified the level of tumor invasion into four (T stage) groups: 1) T1 stage group (mucosa/submucosa), 2) T2 stage group (muscularis propria/subserosa), 3) T3 stage group (serosa without invasion of

adjacent structures), and 4) T4 stage group (adjacent structures). For the ability of lymph node metastasis, we classified the tumors into N0 (no lymph node involvement), N1 ( $\leq 6$  positive nodes), N2 (7 to 15 positive nodes), and N3 ( $\geq 16$  positive nodes) (N stage).

Among different levels of invasion, 10 differentially expressed proteins were identified by SAM analysis ( $P < 0.05$ ). Five proteins (E-cadherin, NFKB50, HIF-3 $\alpha$ , cyclin B1, and cyclin E) were differentially expressed between T1 and T2, and 10 proteins (E-cadherin,  $\beta$ -catenin, NF- $\kappa$ B50, HIF-3 $\alpha$ , cyclin B1, cyclin E, XIAP, TDP1, SK3, and CDC25B) were differentially expressed between T1 and T3, and T1 and T4, as well as between T1 and T2, T1 and T3, and T1 and T4. No differentially expressed proteins were identified between T2 and T3, and T2, and T4. Among these, 10 proteins (E-cadherin, beta-catenin, NF- $\kappa$ B50, HIF-3 $\alpha$ , cyclin B1, cyclin E, XIAP, TDP1, SK3, and CDC25B) were differentially expressed between T1 and combined T3 and T4 as well as between T1 and combined T2, T3, and T4. No differentially expressed proteins were identified between T2 and combined T3 and T4. Among these 10 proteins, 4 proteins (E-cadherin, CDC25B, HIF-3 $\alpha$ , and cyclin B1) were selected as the best predictors by K-fold cross-validation (K=10) analysis (with  $p < 0.05$ ) to distinguish T1 (early cancer) and combined T2, T3, and T4 (advanced cancer). Two-way hierarchical clustering analysis by BRB Array Tools software using these four proteins separated 56 tumors into two main groups: 23 tumor samples into group A and the 33 rest of the samples into group B (Figure 4). It is worthy to note that all six T1 tumors and six T2 tumors (of 12) were classified into group A and five of six T1 tumors were clustered into one subgroup. Twenty-six T3 and T4 tumors (78.8%) were clustered into group B, but only 11



**Figure 4.** Two way hierarchical clustering analysis of tumor samples. The expression profile of four proteins separates the tumors ( $n = 56$ ) into two large groups. All six T1 tumors and six T2 tumors were classified in Group A ( $n = 23$ ) and the most T3 and T4 tumors were classified in Group B ( $n = 33$ ). The numerical number after each case represents the level of tumor invasion. The color scale showed the level of expression. Red indicates overexpression and green indicates underexpression, black indicates no change, and gray no expression. The number in each column represents the sample number. Each row represents a protein.

T3 and T4 tumors (47.8%) were clustered into group A ( $\chi^2 = 5.796$ ;  $P = 0.016$ ). The results suggest that group A tumors represent a biologically less invasive cancer. Of these four proteins, two (cyclin B1 and CDC25B) were up-regulated and two (HIF-3 $\alpha$  and E-cadherin) were down-regulated in group A tumors, suggesting these proteins are associated with invasive behavior of the gastric cancer.

Among different N stages, four differentially expressed proteins were identified by SAM analysis ( $P < 0.05$ ), including PCNA, NF- $\kappa$ B50, Notch 4, and CDK6. PCNA was down-regulated in N1 tumors when compared with N0 tumors. NF- $\kappa$ B50 was down-regulated in N2 tumors when compared with N1 tumors. Notch4 and CDK6 were down-regulated in N3 tumors and NF- $\kappa$ B50 was up-regulated in N3 tumors when compared with N2 tumors. These data suggest that these four proteins may be associated with lymph node metastasis.

In addition, two proteins (HIF-3 $\alpha$  and p-PKC  $\alpha/\beta$  II) were found to be associated with vascular invasion of gastric cancer. Of these, HIF-3 $\alpha$  was up-regulated in the tumors with vascular invasion ( $P = 0.042$ ), whereas p-PKC  $\alpha/\beta$  II was down-regulated in the tumors with vascular invasion ( $P = 0.042$ ).

### Correlation of Protein Expression with Overall Survival

To identify proteins that may predict overall survival, a univariate Cox proportion hazard regression analysis was performed on the 22 differentially expressed proteins in gastric cancer in a cohort of 93 patients (Table 1 and Figure 1). Two proteins (CDK2 and Akt) were found to correlate with overall survival with hazard ratios of 1.293 [ $P = 0.036$ ; 95% confidence interval (CI): 1.017 to 1.644] and 1.431 ( $P = 0.028$ ; 95% CI: 1.039 to 1.971), respectively. To determine whether these proteins can be independent prognostic markers, a multivariate analysis was performed taking into consideration other clinical parameters, such as age, sex, family history, histology grade, vascular invasion, and TNM stage (Table 3). The data showed that CDK2 and Akt still stood as independent predictors with hazard ratios of 1.289 ( $P = 0.044$ , 95% CI: 1.011 to 1.644) and 1.572 ( $P = 0.011$ , 95% CI: 1.111 to 2.224), respectively. In addition, age at surgery ( $P = 0.008$ ) and TNM staging ( $P = 0.011$ ) were also indepen-

dent predictors of survival (Table 3). Based on the hierarchical clustering analysis of CDK2 and Akt expression, the tumor samples were separated into either high or low expression groups. The group with high level expression of CDK2 or Akt associated with a poorer prognosis according to Kaplan-Meier and log-rank survival analysis ( $P = 0.01$  and  $P = 0.03$ , respectively) (see Supplemental Figure S1 at <http://ajp.amjpathol.org>).

Two-way hierarchical clustering analysis using BRB Array Tools software classified 93 tumors into three groups based on expression patterns of CDK2 and Akt (Figure 5A). Tumors ( $n = 21$ ) in group 1 had a decreased expression of both CDK2 and Akt. Tumors ( $n = 42$ ) in group 2 had a decreased expression of either CDK2 or Akt. Tumors ( $n = 30$ ) in group 3 had an increased expression of both CDK2 and Akt. The survival among these groups were also different with group 1 being the best (median survival, 464.1 days), group 2 being intermediate (median survival, 451.2 days), and group 3 being the worst (median survival, 361 days) (Figure 5A). Kaplan-Meier and log-rank survival analysis showed that the patients in these groups had a different prognosis ( $P = 0.024$ ) with group 3 being the worst as compared to group 1 ( $P = 0.008$ ) (Figure 5B). However, there was no difference between groups 1 and 2 ( $P = 0.143$ ) and between groups 2 and 3 ( $P = 0.098$ ), which may be due to the limited number of cases in each group.

To improve the prognostic capability, a risk score was calculated based on expression level of CDK2 and Akt

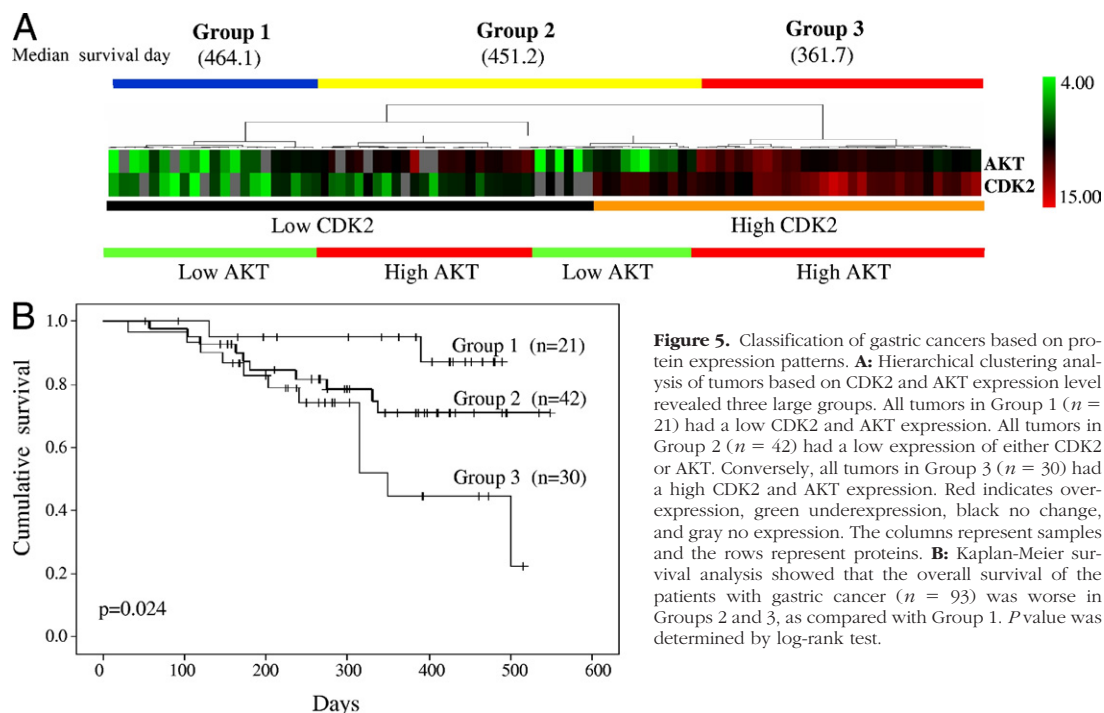
**Table 3.** Hazard Ratios of Overall Survival Determined by Multivariate Analysis

	HR	95% CI		P value
		Lower	Upper	
Age	4.799	1.506	15.288	0.008*
Sex	0.953	0.249	3.641	0.944
Family history	1.634	0.458	5.822	0.449
Histology grade	0.829	0.318	2.158	0.700
Vascular invasion	3.790	0.712	20.177	0.118
TNM	2.417	1.224	4.769	0.011*
CDK2	1.289	1.011	1.644	0.040*
Akt	1.572	1.111	2.224	0.011*

Note that the multivariate Cox regression analysis was performed against each of the variables.

Statistically significant as an independent predictor of survival (\* $P < 0.05$ ).

95% CI, 95% confidence interval; HR, hazard ratio.



**Figure 5.** Classification of gastric cancers based on protein expression patterns. **A:** Hierarchical clustering analysis of tumors based on CDK2 and AKT expression level revealed three large groups. All tumors in Group 1 ( $n = 21$ ) had a low CDK2 and AKT expression. All tumors in Group 2 ( $n = 42$ ) had a low expression of either CDK2 or AKT. Conversely, all tumors in Group 3 ( $n = 30$ ) had a high CDK2 and AKT expression. Red indicates over-expression, green underexpression, black no change, and gray no expression. The columns represent samples and the rows represent proteins. **B:** Kaplan-Meier survival analysis showed that the overall survival of the patients with gastric cancer ( $n = 93$ ) was worse in Groups 2 and 3, as compared with Group 1.  $P$  value was determined by log-rank test.

and corresponding regression coefficients. A patient's risk score was calculated as the sum of the expression values of CDK2 and Akt as determined by the Protein Pathway Array multiplied by the corresponding regression coefficients ( $CDK2 = 0.257$  and  $Akt = 0.359$ ), which were determined by univariate Cox regression analysis of  $\log_2$  transformed expression level. The risk scores for entire patient cohort ( $n = 93$ ) ranged from 5.2 to 11.08, as shown in Figure 6A. Based on the risk score curve, we separated the patients into two groups: low-risk score (5.2 to 9.03) and high-risk score (9.04 to 11.08) (Figure 6A). The patients in the low-risk score group had a better prognosis than those in the high-risk score group ( $P = 0.007$ ) (Figure 6B). To eliminate the confounding effect of invasion levels, we analyzed the survival rate of patients with T3 invasion only using risk scores. Our results showed that the high-risk group (risk score, 9.04 to 11.08) had a worse prognosis compared with the low-risk group (risk score, 5.2 to 9.03) (Figure 6C).

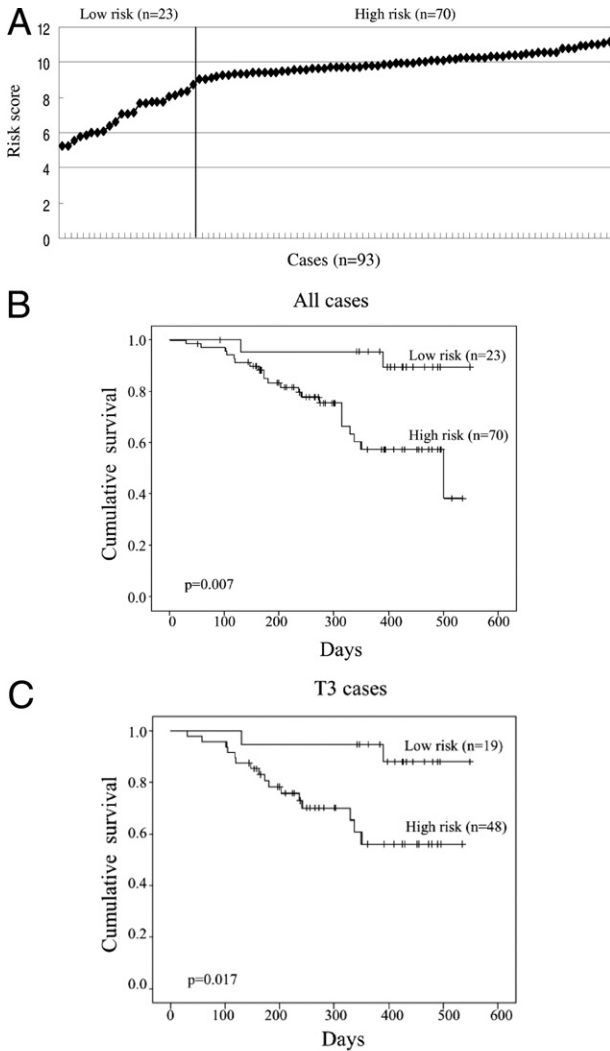
To further confirm the ability of the risk scores to predict survival, the expression levels of CDK2 and Akt in an additional cohort of gastric cancer tissues ( $n = 50$ ) were assessed using another method, conventional Western blot. The patients' risk scores were calculated as previously described after adjustment of the expression levels of CDK2 and Akt, based on  $\beta$ -actin level. Also as previously described, the patients were separated into two groups: low-risk score (6.5 to 9.03) and high-risk score (9.04 to 12.3) (Figure 7A). The patients with a high-risk score group had a lower median overall survival (median survival, 430 days) than those with a low-risk score group (median survival, 512 days;  $P = 0.039$ ) (Figure 7B). The results confirmed that the patients in the low-risk score group also had a better prognosis than those in the high-risk score group in the validation set. These data

suggest that risk score based on CDK2 and Akt can be used to stratify patients with gastric cancer.

### Discussion

Major efforts are being made to develop molecular signature-based methods to complement the traditional histopathological diagnosis and prognosis in gastric cancer and also to understand the biology of gastric cancer at a molecular level. Currently, majority of molecular signatures for gastric cancer have been derived from gene expression microarray studies.<sup>12-14</sup> However, we take a different approach to identify protein signatures with a focus on signaling transduction-related proteins using the novel Protein Pathway Array technology.<sup>9</sup> Our hypothesis is that the changes at the genetic and epigenetic levels in cancer cells will affect the expression and activation of signaling transduction proteins, which ultimately controls cancer cell function, such as proliferation, invasion, and metastasis. Therefore, the pattern of signal transduction protein expression and phosphorylation can serve as protein signatures for classification and prognostication of gastric cancer, which should more closely reflect the behavior of gastric cancers.

In this study, we initially screened 142 proteins and phosphoproteins in 37 pairs (the training set) of gastric cancers and adjacent nontumor mucosa tissues using the Protein Pathway Array. Among 142 proteins and phosphoproteins tested, 83 were detected and 22 were differentially expressed with  $P < 0.05$  or  $q < 0.05$  as determined by paired  $t$ -test and SAM analysis. Among these 22 proteins, 9 proteins could distinguish tumors from normal tissues in 19 pairs of gastric cancer samples (the validation set) with 100% sensitivity and specificity



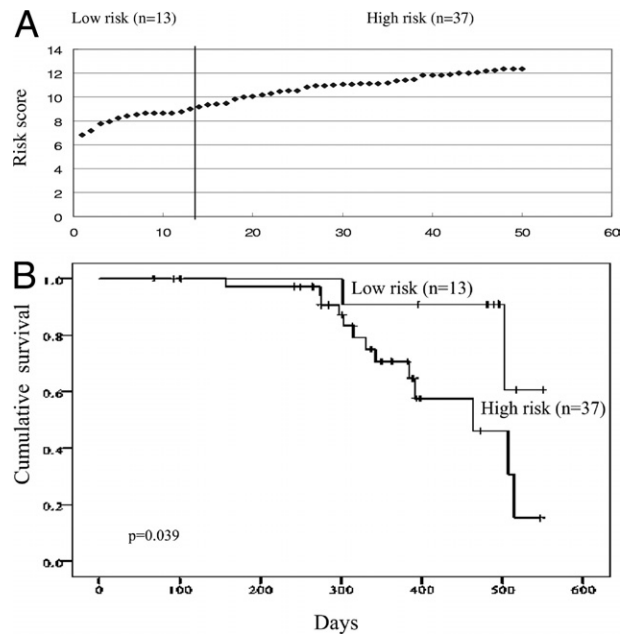
**Figure 6.** Kaplan-Meier survival analysis of patients with gastric cancer based on the risk scores. **A:** The cases ( $n = 93$ ) were ranked according to the risk scores calculated, based on the expression level of CDK2 and AKT of each sample. The line divided the cases into low- and high-risk groups. **B:** The overall survival rate of all gastric cancer cases ( $n = 93$ ) was correlated with risk scores as shown in (A). **C:** The overall survival rate of the cases with a T3 gastric cancer ( $n = 73$ ) was determined based on their risk scores. *P* values were determined by log-rank test.

based on three-nearest neighbor model, suggesting that signaling related proteins can be used as a molecular signature to classify tumor and normal tissue.

The proteins dysregulated in gastric cancer are those important in diverse cellular processes, including cell proliferation/differentiation, cell cycle regulation, DNA replication/recombination/repair, cell death, cell adhesion/migration, stress response, and organ development (see Supplemental Table S1 at <http://ajp.amjpathol.org>). The dysregulation of some of these proteins was reported in gastric cancer in previous studies, consistent with our findings. For example, PCNA, XIAP, and p-PKC $\alpha$  were up-regulated in gastric cancer<sup>15,16</sup>, and E-cadherin and p27 were down-regulated in gastric cancer.<sup>17,18</sup> However, the expression of some of the proteins differed from that described in previous publications in different can-

cers. For example, cyclin D1, cyclin E, p-ERK, and CDC25B were down-expressed in our study, but reportedly over-expressed in breast (cyclin D1), pancreas (cyclin E), and lung (ERK1/2 and CDC25B) cancers by immunohistochemical staining.<sup>19–22</sup> These contradictory findings may be the result of different methodologies used in these studies (ie, quantitative Protein Pathway Array in our study versus qualitative immunohistochemical staining in other studies). It is noteworthy that several proteins found in this study have not been reported previously in gastric cancer, such as Notch4, p-ERK, CDK6, CDK2, NF- $\kappa$ B52, SK3, TDP1, SCR-1, and HIF-3 $\alpha$ . These findings demonstrate the possibility of using Protein Pathway Array technology to identify new roles of these proteins in gastric cancer, even though these proteins have been previously reported to have many different physiological functions (see Supplemental Table S1 at <http://ajp.amjpathol.org>). Our results also demonstrated a broad disturbance of cell regulatory systems in gastric cancer (ie, 22 proteins with different cellular functions) (see Supplemental Table S1 at <http://ajp.amjpathol.org>) and suggest a complicated process of gastric cancer development. Furthermore, not surprisingly, changes in these proteins do not occur in all tumor samples, suggesting heterogeneity of gastric cancer. In addition, it is suggested that single targeted therapy may not be effective in gastric cancer based on these findings.

Gastric cancer is an aggressive malignancy with a great tendency of deep invasion, lymph node metastasis, and distance metastasis.<sup>10</sup> Therefore, we sought to identify the proteins that may be involved in invasion and lymph node metastasis of gastric cancer by comparing



**Figure 7.** Kaplan-Meier survival analysis of additional patients with gastric cancer based on the risk scores. **A:** The additional cases ( $n = 50$ ) were ranked according to the risk scores calculated based on the expression level of CDK2 and AKT of each sample, as determined by Western blot. The line divided the cases into low- and high-risk groups. **B:** The overall survival rate of the additional gastric cancer cases ( $n = 50$ ) was correlated with risk scores as shown in (A). *P* values were determined by log-rank test.



the protein expression pattern based on the different pathological features. Our results showed two distinct protein expression patterns exist in gastric cancer, which separate gastric cancer into two groups (Figure 4). In group A, cancers that may represent a more indolent gastric cancer, such as cell proliferation-related proteins (cyclin B1 and CDC25B) were up-regulated. In contrast, the adhesion (ie, E-cadherin) and hypoxia (ie, HIF-3 $\alpha$ ) related proteins were up-regulated in group B cancers, which may represent a more aggressive gastric cancer. Furthermore, up-regulation of HIF-3 $\alpha$  was associated with tumor vascular invasion, supporting its involvement in aggressive behavior. Up-regulation of E-cadherin in advanced gastric cancer and down-regulation in early gastric cancer has been previously reported.<sup>17</sup> Abnormal expression of E-cadherin and  $\beta$ -catenin was suggested to be molecular markers for submucosal invasion and lymph node metastasis in early gastric cancer.<sup>18</sup> However, there are no reports on the association of cyclin B1, CDC25B, and HIF-3 $\alpha$ , with invasion of gastric cancer in the literature.

When correlating with lymph node status (N stage), four proteins (PCNA, NF- $\kappa$ B50, Notch4, and CDK6) were associated with lymph node metastasis. Czyzewska et al<sup>23</sup> reported a statistically significant correlation of PCNA in advanced gastric cancer with local lymph node involvement. Katarzyna et al reported that the expression of PCNA was found to correlate with the presence of lymph node metastases in colon cancer.<sup>24</sup> However, to our knowledge, there are no reports on the association of Notch4, CDK6, and NF- $\kappa$ B52 with lymph node metastasis in any cancer. It is suggested that these results indicate different sets of proteins are involved in invasion (ie, T stage) and lymph node metastasis (ie, N stage).

Overall survival of gastric cancer is poor and only 40% of patients survive more than 5 years after initial surgery.<sup>3</sup> It is well known that even with patients who have similar clinical and pathological features, the outcome varies, suggesting that current staging systems for gastric cancer that are based on clinical and pathological findings may have reached their limit of usefulness for predicting outcomes. Therefore, the addition of molecular biomarkers may add value for more accurate staging. In this study, we demonstrated that the expression level of CDK2 and Akt are independent markers, together with age and TNM stage, to predict the overall survival of gastric cancer patients. Akt is a serine/threonine-specific protein kinase and it is a critical protein in regulating many different cellular functions, including proliferation, survival, apoptosis, and angiogenesis (see Supplemental Table S1 at <http://ajp.amjpathol.org>). CDK2 is a cyclin-dependent serine/threonine-specific protein kinase and it is a critical protein in controlling G1 to S transition via interaction with cyclin E (see Supplemental Table S1 at <http://ajp.amjpathol.org>). Previous reports demonstrated that the expression of Akt and CDK2 is associated with poor prognosis of many cancers. Capodanno et al<sup>25</sup> reported that dysregulated PI3K/Akt/phosphatase and tensin homolog (PTEN) pathway is associated with a short disease-free survival in node-negative breast carcinoma. Mihara et al<sup>26</sup> reported that elevated expression of CDK2 is a negative predictive marker of the patient prognosis

with oral cancer. However, to our knowledge, this is the first study demonstrating that CDK2 and Akt are dysregulated in gastric cancer and correlate significantly with patient survival. To predict the risk of each patient, we further developed prognostic risk scores based on the combined expression level of CDK2 and Akt. Our second cohort of patients confirmed that the risk score can be used for more accurate patient stratification for surgical and chemotherapeutic treatment in combination with TNM stage.

In conclusion, our data show a broad dysregulation of signaling proteins in gastric cancer, suggesting the important roles of these signaling proteins in carcinogenesis. The altered expression of some of the proteins correlated with invasion and metastasis, whereas other proteins correlated with overall survival, indicating that different sets of signaling proteins associate with different tumor behaviors and clinical outcomes. Future studies will be focused on understanding the roles of these proteins in controlling tumor behavior and confirming the ability of these proteins to classify gastric cancer and predict survival in a different cohort of patients.

### Acknowledgments

We thank Yang Li for his excellent technical support in protein preparation, Dr. Wei Li for collecting the clinical information, and Drs. Steven Itzkowitz and Michelle Kim for valuable comments on the manuscript.

### References

1. Parkin DM, Bray F, Ferlay J, Pisani P: Global cancer statistics, 2002. *CA Cancer J Clin* 2005, 55:74–108
2. Tsugane S, Sasazuki S: Diet and the risk of gastric cancer: review of epidemiological evidence. *Gastric Cancer* 2007, 10:75–83
3. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ: Cancer statistics, 2008. *CA Cancer J Clin* 2008, 58:71–96
4. Mori M, Mimori K, Shiraishi T, Tanaka S, Ueo H, Sugimachi K, Akiyoshi T: p27 expression and gastric carcinoma. *Nat Med* 1997, 3:593
5. Akama Y, Yasui W, Yokozaki H, Kuniyasu H, Kitahara K, Ishikawa T, Tahara E: Frequent amplification of the cyclin E gene in human gastric cancer. *Jpn J Cancer Res* 1995, 86: 617–621
6. Graziano F, Mandolesi A, Ruzzo A, Bearzi I, Testa E, Arduini F, Silva R, Muretto P, Mari D, Berardi R, Scartozzi M, Lai V, Cascinu S, Magnani M: Predictive and prognostic role of E-cadherin protein expression in patients with advanced gastric carcinomas treated with palliative chemotherapy. *Tumour Biol* 2004, 25:106–110
7. Sanz-Ortega J, Steinburg SM, Moro E, Saez M, Lopez JA, Sierra E, Sanz-Esponera J, Merino MJ: Comparative study of tumor angiogenesis and immunohistochemistry for p53, c-erbB2, c-myc and EGFr as prognostic factors in gastric cancer. *Histol Histopathol* 2000, 15:455–462
8. Zheng L, Wang L, Ajani J, Xie K: Molecular basis of gastric cancer development and progression. *Gastric Cancer* 2004, 7:61–77
9. Zhang DY, Ye F, Gao L, Liu X, Zhao X, Che Y, Wang H, Wang L, Wu J, Song D, Liu W, Xu H, Jiang B, Zhang W, Wang J, Lee P: Proteomics, pathway array and signaling network-based medicine in cancer. *Cell Div* 2009, 4:20–36
10. Mullaney PJ, Wadley MS, Hyde C, Wyatt J, Lawrence G, Hallissey MT, Fielding JW: Appraisal of compliance with the UICC/AJCC staging system in the staging of gastric cancer. *Union International Contra la Cancrum/American Joint Committee on Cancer. Br J Surg* 2002, 89:1405–1408

11. Ye F, Che Y, McMillen E, Gorski J, Brodman D, Saw D, Jiang B, Zhang DY: The effect of *Scutellaria Baicalensis* on the signaling network in hepatocellular carcinoma cells. *Nutr Cancer* 2009, 61:530–537
12. Chen CN, Lin JJ, Chen JJ, Lee PH, Yang CY, Kuo ML, Chang KJ, Hsieh FJ: Gene expression profile predicts patient survival of gastric cancer after surgical resection. *J Clin Oncol* 2005, 23:7286–7295
13. Takeno A, Takemasa I, Doki Y, Yamasaki M, Miyata H, Takiguchi S, Fujiwara Y, Matsubara K, Monden M: Integrative approach for differentially overexpressed genes in gastric cancer by combining large-scale gene expression profiling and network analysis. *Br J Cancer* 2008, 99:1307–1315
14. Zheng L, Wang L, Ajani J, Xie K: Molecular basis of gastric cancer development and progression. *Gastric cancer* 2004, 7:61–77
15. Lee KE, Lee HJ, Kim YH, Yu HJ, Yang HK, Kim WH, Lee KU, Choe KJ, Kim JP: Prognostic significance of p53, nm23, PCNA and c-erbB-2 in gastric cancer. *Jpn J Clin Oncol* 2003, 33:173–179
16. Shibata T, Noguchi T, Takeno S, Gabbert HE, Ramp U, Kawahara K: Disturbed XIAP and XAF1 expression balance is an independent prognostic factor in gastric adenocarcinomas. *Ann Surg Oncology* 2008, 15:3579–3587
17. Joo YE, Rew JS, Kim HS, Choi SH, Park CS, Kim SJ: Changes in the E-cadherin-catenin complex expression in early and advanced gastric cancers. *Digestion* 2001, 64:111–119
18. Ohene-Abuakwa Y, Noda M, Perenyi M, Kobayashi N, Kashima K, Hattori T, Pignatelli M: Expression of the E-cadherin/catenin (alpha-, beta-, and gamma-) complex correlates with the macroscopic appearance of early gastric cancer. *J Pathol* 2000, 192:433–439
19. Gillett C, Smith P, Gregory W, Richards M, Millis R, Peters G, Barnes D: Cyclin D1 and prognosis in human breast cancer. *Int J Cancer* 1996, 69:92–99
20. Umekita Y, Ohi Y, Sagara Y, Yoshida H: Overexpression of cyclinD1 predicts for poor prognosis in estrogen receptor-negative breast cancer patients. *Int J Cancer* 2002, 98:415–418
21. Yue H, Jiang HY: Expression of cell cycle regulator p57kip2, cyclin E protein and proliferating cell nuclear antigen in human pancreatic cancer: an immunohistochemical study. *World J Gastroenterol* 2005, 11:5057–5060
22. Boldrini L, Gisfredi S, Ursino S, Lucchi M, Mussi A, Fontanini G: CDC25B: relationship with angiogenesis and prognosis in non-small cell lung carcinoma. *Hum Pathol* 2007, 38:1563–1568
23. Czyżewska J, Guzińska-Ustymowicz K, Pryczynicz A, Kemon A, Bandurski R: Immunohistochemical evaluation of Ki-67, PCNA and MCM2 proteins proliferation index (PI) in advanced gastric cancer. *Folia Histochem Cytobiol* 2009, 47:289–296
24. Guzińska-Ustymowicz K, Stepień E, Kemon A. MCM-2, Ki-67 and PCNA protein expressions in pT3G2 colorectal cancer indicated lymph node involvement. *Anticancer Res* 2008, 28:451–458
25. Capodanno A, Camerini A, Orlandini C, Baldini E, Resta ML, Bevilacqua G, Collecchi P: Dysregulated PI3K/Akt/PTEN pathway is a marker of a short disease-free survival in node-negative breast carcinoma. *Hum Pathol* 2009, 40:1408–1417
26. Mihara M, Shintani S, Nakahara Y, Kiyota A, Ueyama Y, Matsumura T, Wong DT: Overexpression of CDK2 is a prognostic indicator of oral cancer progression. *Jpn J Cancer Res* 2001, 92(3):352–360