Research Paper

Prognostic and Predictive Value of p21-activated Kinase 6 Associated Support Vector Machine Classifier in Gastric Cancer Treated by 5-fluorouracil/Oxaliplatin Chemotherapy

Yuming Jiang [a,b,1], Wei Liu [b,c,1], Tuanjie Li [a,1], Yanfeng Hu [a,1], Sile Chen [d], Sujuan Xi [b,e], Yajia Wen [b,e], Lei Huang [f], Liying Zhao [a], Cuicui Xiao [b,c], Xiaohui Huang [b], Zhen Han [a], Hao Liu [a], Xiaolong Qi [a], Yang Yang [b,c], Jiang Yu [a,***], Shirong Cai [d,**], Guoxin Li [a,*]

[a] Department of General Surgery, Nanfang Hospital, Southern Medical University, 1838 North Guangzhou Avenue, Guangzhou, China
[b] Guangdong Key Laboratory of Liver Disease Research, The 3rd Affiliated Hospital of Sun Yat-sen University, Guangzhou 510630, China
[c] Department of Hepatic Surgery, The 3rd Affiliated Hospital of Sun Yat-sen University, Guangzhou, China
[d] Department of Gastrointestinal Surgery of the First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510700, Guangdong, China
[e] Department of Infectious disease, The 3rd Affiliated Hospital of Sun Yat-sen University, Guangzhou, China
[f] German Cancer Research Center (DKFZ), Heidelberg 69120, Germany

ARTICLE INFO

ABSTRACT

To determine whether p21-activated Kinase (PAK) 6 is a prognostic and predictive marker in gastric cancer (GC) and to construct a classifier that can identify a subset of patients who are highly sensitive to 5-fluorouracil/oxaliplatin chemotherapy. We retrospectively analyzed the expression levels of PAK6, cyclooxygenase 2, $p21^{WAF1}$, Ki-67, excision repair cross-complementing gene 1, and thymidylate synthase in 242 paraffin-embedded GC specimens of the training cohort by immunohistochemistry. Then, we used support vector machine (SVM)–based methods to develop a predictive classifier for chemotherapy (chemotherapy score – CS-SVM classifier). Further validation was performed in an independent cohort of 279 patients. High PAK6 expression was associated with poor prognosis and increased chemoresistance to 5-FU/oxaliplatin chemotherapy. The CS-SVM classifier distinguished patients with stage II and III GC into low- and high-CS-SVM groups, with significant differences in the 5-year disease-free survival (DFS) and overall survival (OS) in chemotherapy patients. Moreover, chemotherapy significantly prolonged the DFS and OS of the high CS-SVM patients in the training and validation cohorts. In conclusion, PAK6 was an independent prognostic factor and increased chemoresistance. The CS-SVM classifier distinguished a subgroup of stage II and III patients who would highly benefit from chemotherapy, thus facilitating patient counseling and individualizing the management.

1. Introduction

Gastric cancer (GC) is the fourth most common human malignant disease and the second leading cause of cancer-related deaths worldwide. (Torre et al., 2015) In recent years, significantly improved outcomes have been reported for patients with GC, largely thanks to improved drug therapy. (Noh et al., 2014) According to new guidelines, 5-fluorouracil/oxaliplatin chemotherapy is recommended as the standard postoperative chemotherapy regimen for advanced GC. (Noh et al., 2014) Significant survival benefits of 5-fluorouracil/oxaliplatin chemotherapy have been reported in patients with metastatic gastric cancer as well as those who have undergone surgery. (Longley et al., 2003; McLean and El-Omar, 2014; Noh et al., 2014; Razzak, 2014) However, the efficacy of 5-fluorouracil/oxaliplatin chemotherapy greatly differs among individuals, with local recurrence or distant metastasis occurring in approxiamately 40% of GC patients during the course of the disease. (McLean and El-Omar, 2014; Razzak, 2014) Therefore, the accurate prediction of chemotherapy efficacy is clinically important to guide individualized regimens and improve treatment outcomes.(Jiang et al., 2016; McLean and El-Omar, 2014) Several potential molecular predictors (e.g., p53, cyclooxygenase 2 (COX2), $p21^{WAF1}$, Ki-67, excision repair cross-complementing gene 1 (ERCC1), thymidylate synthase (TS), and thymidine phosphorylase (TP)) of chemosensitivity have

been investigated in colon cancer, breast cancer and GC,(Kwon et al., 2007; Metzger et al., 1998; Sulzyc-Bielicka et al., 2011; Xu et al., 2016) but these biomarkers still require validation and are not a part of the standard clinical practice in GC evaluations. Therefore, prognostic and predictive tools are urgently required to identify those patients likely to benefit from 5-fluorouracil/oxaliplatin chemotherapy.(Longley et al., 2003).

The PAKs (p21-activated kinases) belong to a highly conserved family of serine/threonine protein kinases that are important mediators of Rac and Cdc42 function.(Radu et al., 2014) PAKs have been implicated in the regulation of multiple cellular functions, including cell motility, actin reorganization, gene transcription, cell transformation, and apoptotic signaling, and more recently in radiotherapy and chemotherapy resistance signaling.(Radu et al., 2014; Zhang et al., 2010) PAK6 is the most recently identified and the least well known member of the PAK family. Recent studies have reported that PAK6 expression is increased and indicated poor prognosis in hepatocellular carcinoma (HCC), colorectal cancer (CRC) and prostate cancer (PCa).(Chen et al., 2015; Wen et al., 2009) Furthermore, PAK6 may be a good predictor of 5-FU response in colon cancer.(Chen et al., 2015) Knockdown of PAK6 expression inhibits PCa growth and enhances chemosensitivity to docetaxel (Wen et al., 2009). When combined with irradiation, inhibition of PAK6 leads to significantly decreased PCa cell survival.(Zhang et al., 2010) Another study reported that PAK6 expression was reduced and associated with good prognosis in clear cell renal cell carcinoma (ccRCC).(Liu et al., 2014) However, the molecular and prediction (prognosis and chemosensitivity) functions of PAK6 in GC are not yet well known.

Recently, several supervised learning methods, such as decision trees, have been applied to the analysis of cDNA microarrays to refine prognosis in nasopharyngeal carcinoma, breast cancer and non–small-cell lung cancer.(Chen et al., 2007) State-of-the-art classification algorithms such as support vector machines (SVMs) can be used to select a small subset of highly discriminating markers and patients or disease attributes to build reliable cancer classifiers.(Wang et al., 2011).

In this study, we analyzed PAK6 expression in GC specimens by immunohistochemical (IHC) analysis and assessed its relationship with disease-free survival (DFS), overall survival (OS) and chemosensitivity. We further assessed the predictive value of COX2, p21^{WAF1}, Ki-67, TS and ERCC1 and then constructed a SVM classifier – chemotherapy score (CS-SVM) – integrating PAK6, Ki-67, TS, ERCC1, COX2, and p21^{WAF1}, that could effectively identify a small subset of GC patients that would highly benefit from 5-FU/oxaliplatin chemotherapy.

2. Materials and Methods

2.1. Patients and Tissue Specimens

We enrolled two independent panels of GC patients, including a training cohort of 241 patients with incident, primary, biopsy-confirmed GC diagnosed between January 2005 and June 2007 at Nanfang Hospital of Southern Medical University (Guangzhou, China); an independent validation cohort of 171 patients with incident, primary, biopsy-confirmed GC diagnosed between July 2007 to April 2009, at Nanfang Hospital of Southern Medical University and 171 patients with incident, primary, biopsy-confirmed GC diagnosed between January 2005 to December 2007 at the First Affiliated Hospital of Sun Yat-sen University (SYSU). Detailed information of the inclusion criteria was provided in Supplementary materials. Two pathologists reassessed all of these samples. All of the patients underwent standard resection operation with or without chemotherapy, with similar treatment protocols across study sites in accordance with the US National Comprehensive Cancer Network guidelines. This study was approved by the research ethics committees at all of the participating centers and the need to obtain informed consent was waived. The quality of the study was ensured

by following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

2.2. Immunohistochemistry and Selection of Cutoff Score

Formalin-fixed paraffin-embedded (FFPE) samples were cut into 4-μm sections, which were then processed for IHC as previously described (Fig. S1). (Li et al., 2016; Li et al., 2017) Detailed information was provided in Supplementary materials. We selected the optimum cutoff score for every marker using the X-tile software version 3.6.1 (Yale University School of Medicine, New Haven, CT, USA) based on the association with the patients' OS.

2.3. Prognosis Prediction Using SVM-based Methods

SVM was introduced by Vapnik (Vapnik, 1999) for data classification and function approximation. An SVM is a binary classifier trained on a set of labeled patterns called training samples. The objective of training an SVM is to find a hyperplane that divides these samples into two sides so that all the points with the same label will be on the same side of the hyperplane (Choi et al., 2011; Vapnik, 1999; Wang et al., 2011; Xu et al., 2012; Zhu et al., 2009). In this study, SVM was used to predict whether a patient died as a result of GC within 5 years. We adopted the SVM–recursive feature elimination algorithm to select and rank useful features.(Wang et al., 2011) To investigate the possibility of identifying different prognostic subgroups of stage II and III GC based on six immunomarkers using SVM, we performed a set of experiments in the training cohort of 203 patients with stage II and III GC; the developed CS-SVM classifier was further validated in 279 stage II and III GC patients from an independent cohort of 342 GC patients. In the training cohort, patients on the side of the hyperplane who had better survival were classified into high CS-SVM group. The SVM data processing methods were conducted as previously described. (Wang et al., 2011; Xu et al., 2012; Zhu et al., 2009) The programs were coded using R software; scripts are available on request.

2.4. Construction, Validation and Calibration of the Nomogram

In the training cohort, on the basis of the results of the multivariable analysis, a nomogram was formulated by R 3.0.1 (http://www.r-project.org) with the survival and rms package. Backward step-wise selection was applied by using the likelihood ratio test with Akaike's information criterion as the stopping rule.(Collins et al., 2015; Sauerbrei et al., 2011) Time-dependent receiver operating characteristic (ROC) analysis was used to compare the accuracy of the prediction of clinical outcome by the nomograms. The performance of the developed nomogram was tested in the external validation cohort. Calibration of the nomogram for 1-, 3-, 5-year OS was performed by comparing the predicted survival with the observed survival after bias correction.

2.5. Clinical Use

Decision curve analysis was conducted to determine the clinical usefulness of the nomograms by quantifying the net benefits at different threshold probabilities.(Localio and Goodman, 2012; Vickers et al., 2008).

2.6. Statistical Analysis

We compared two groups using the t-test for continuous variables and χ^2 test for categorical variables. Survival curves were generated according to the Kaplan-Meier method and compared by the log-rank test. Univariate and multivariate analyses of the prognostic factors were performed using the Cox proportional hazards model. All of statistical tests were conducted using R software (version 3.0.1) and SPSS software (version 19.0). Statistical significance was set at 0.05.

3. Results

3.1. Patient Characteristics and IHC Findings

Table 1 lists the detailed clinicopathological characteristics of the training and validation cohorts. The specific expression of cytoplasmic PAK6 was observed in both nontumoral and intratumoral tissues (Fig. 1A–B). Compared with the nontumoral PAK6 density, intratumoral PAK6 expression was higher ($P < 0.0001$; Fig. 1C). The expression of PAK6 was much higher in advanced stage GC [stages I–II (n = 46) vs. stages III–IV (n = 166), $P = 0.0007$]. Furthermore, the percentage of patients with high intratumoral PAK6 expression increased moderately accompanying by disease progression from TNM stage I to IV (Fig. 1D, Table 1).

3.2. Prognostic and Predict Value of PAK6 Expression for GC Outcome

In the training cohort, the 5-year DFS and OS were 16.7% and 24.6% for the high PAK6 group, respectively, and were 41.7% and 47.2% for the low PAK6 group, respectively (hazard ratio (HR) 1.777 (1.287–2.452), 1.608 (1.149–2.250), respectively; all $P < 0.01$; Fig. S2A). To confirm that the PAK6 expression had an excellent prognostic value in different populations, we further applied it to the validation cohort, and obtained similar results (Fig. S2A). Univariate and multivariate analysis showed that PAK6 was a prognostic factor in the prediction of patient outcomes (Table 2 and S1–2).

A subset analysis using chemotherapy indicated that PAK6 expression had an excellent prognostic value in the chemotherapy patients (DFS: HR, 2.447 (1.536–3.899), $P < 0.0001$; OS: 2.512 (1.681–3.753), $P < 0.0001$; Fig. 2), but not in the no chemotherapy patients of the training cohort. Similar results were observed in the validation cohort. In a subset analysis according to stage, for stage III cancer patients, PAK6 expression had an excellent prognostic value in patients receiving chemotherapy, but not in patients without chemotherapy (Fig. S3–4). For stage II tumor patients, PAK6 expression was not significantly associated with survival (Fig. S3–4). To investigate the relationship of intratumoral PAK6 expression and 5-FU/oxaliplatin chemotherapy response, a subset analysis using PAK6 expression revealed that chemotherapy significantly prolonged DFS and OS in the low-PAK6 group (training cohort: $P = 0.007$, $P = 0.001$; validation cohort: $P = 0.009$, $P = 0.008$; Fig. S5), but did not yield a survival benefit in the high-PAK6 group. We also noted similar results in a subset analysis of stage II and stage III cancer patients (Fig. S6). These results suggest that high PAK6 expression indicated increased chemoresistance in 5-FU/oxaliplatin chemotherapy GC patients.

We further explored the predictive value of TS, ERCC1, COX2, p21^{WAF1} and Ki-67 in 5-FU/oxaliplatin chemotherapy for Stage II and III GC patients in the training cohort. Stage II and III GC patients with low expression of TS, ERCC1, COX2, and Ki-67 showed a favorable

Table 1
Clinical characteristics of patients according to PAK6 in the training and validation cohorts.

Variables	Training cohort (n = 241)				Validation cohort (n = 342)			
	N	Low PAK6 (%)	High PAK6 (%)	p-Value	N	Low PAK6 (%)	High PAK6 (%)	p-Value
Gender				0.874				0.693
Male	173	91(52.6%)	82(47.4%)		139	52(37.4%)	87(62.6%)	
Female	68	35(51.5%)	33(48.5%)		89	31(34.8%)	58(65.2%)	
Age(years)				0.969				0.579
<60	138	72(52.2%)	66(47.8%)		140	49(35%)	91(65%)	
≥60	103	54(52.4%)	49(47.6%)		88	34(38.6%)	54(61.4%)	
Tumor size(cm)				0.341				0.391
<4	123	68(55.3%)	55(44.7%)		88	29(33%)	59(67%)	
≥4	118	58(49.2%)	60(50.8%)		140	54(38.6%)	86(61.4%)	
Tumor location				0.547				0.474
Cardia	49	29(59.2%)	20(40.8%)		85	39(45.9%)	46(54.1%)	
Body	39	17(43.6%)	22(56.4%)		82	39(47.6%)	43(52.4%)	
Antrum	119	62(52.1%)	57(47.9%)		145	75(51.7%)	70(48.3%)	
Whole	34	18(52.9%)	16(47.1%)		30	11(36.7%)	19(63.3%)	
Differentiation status				0.111				0.057
Well + Moderate	124	71(57.3%)	53(42.7%)		90	26(28.9%)	64(71.1%)	
Poor and undifferentiated	117	55(47.0%)	62(53.0%)		138	57(41.3%)	81(58.7%)	
Lauren type				0.137				0.838
Intestinal type	194	106(54.6%)	88(45.4%)		153	55(35.9%)	98(64.1%)	
Diffuse type	47	20(42.6%)	27(57.4%)		75	28(37.3%)	47(62.7%)	
CEA				0.282				0.096
Elevated	73	42(57.5%)	31(42.5%)		47	22(52.7%)	25(53.2%)	
Nomal	168	84(50.0%)	84(50.0%)		181	61(33.7%)	120(66.3%)	
CA199				0.847				0.004
Elevated	74	38(51.4%)	36(48.6%)		55	29(52.7%)	26(47.3%)	
Normal	167	88(52.7%)	79(47.3%)		173	54(31.2%)	199(68.8%)	
Depth of invasion				0.347				0.619
T1 + T2	50	29(58.0%)	21(42.0%)		48	16(33.3%)	32(66.7%)	
T3 + T4	190	96(50.5%)	94(49.5%)		180	67(37.2%)	113(62.8%)	
Lymph node metastasis				0.055				0.455
N0	55	35(63.6%)	20(36.4%)		53	17(32.1%)	36(67.9%)	
N1 + N2 + N3	186	91(48.9%)	95(51.1%)		165	66(37.7%)	109(62.3%)	
TNM stage				0.013				0.009
I	12	7(58.3%)	5(41.7%)		30	21(70.0%)	9(30.0%)	
II	68	46(67.6%)	22(32.4%)		70	40(57.1%)	30(42.9%)	
III	135	65(48.1%)	70(51.9%)		209	91(43.5%)	118(56.5%)	
IV	26	9(34.6%)	17(65.4%)		33	12(36.4%)	21(63.6%)	
Chemotherapy				0.231				0.687
No	107	61(57.0%)	46(43.0%)		138	68(49.3%)	70(50.7%)	
Yes	134	66(49.3%)	68(50.7%)		204	96(47.1%)	108(52.9%)	
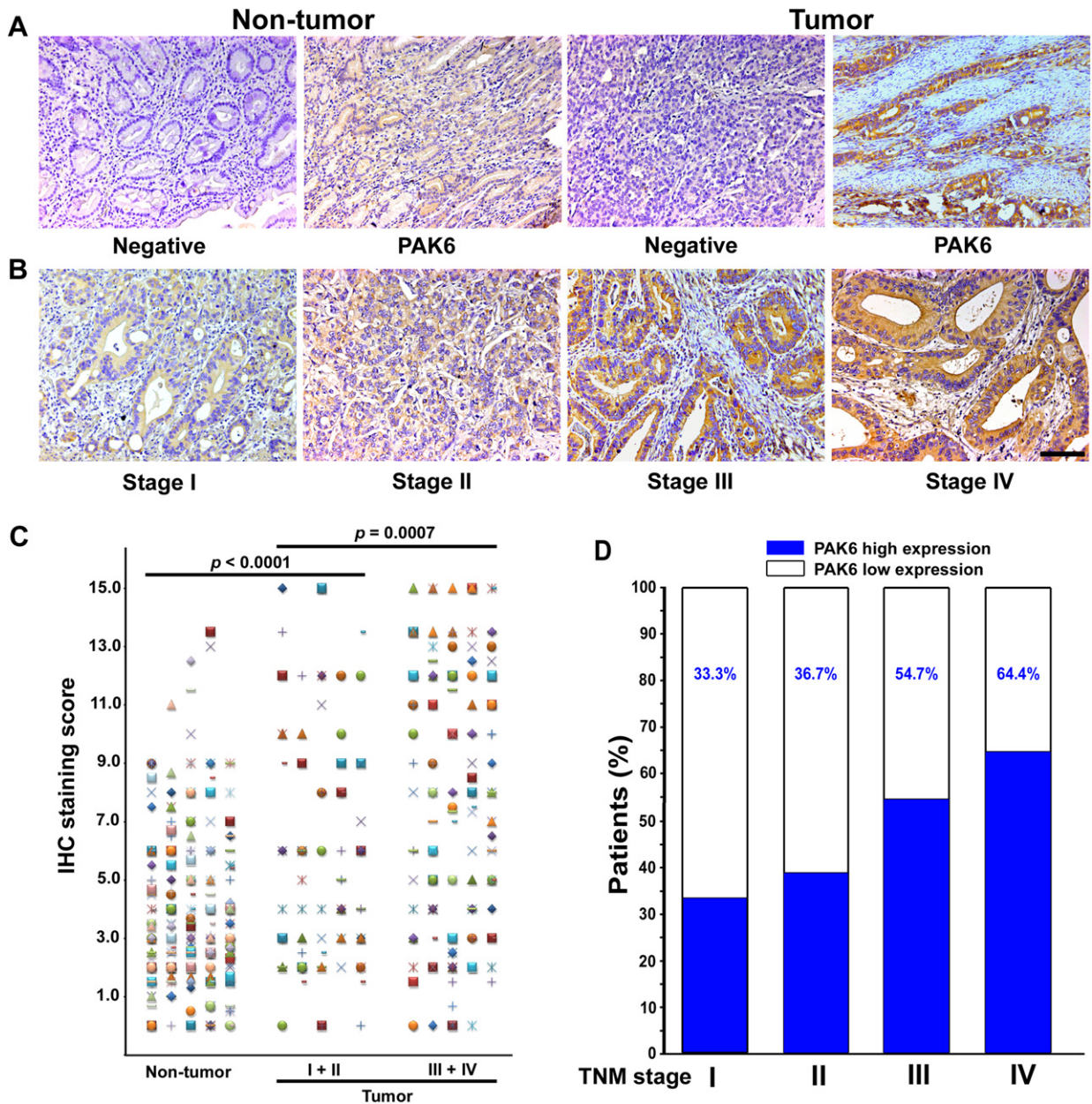
Fig. 1. PAK6 expression in GC tissues. Representative IHC photographs reveal high PAK6 density in tumor tissue, low density in nontumor tissue (A), and density from TNM stage I–IV (B). (C) Scatter plots for IHC staining score in unpaired nontumor tissue (n = 242) and tumor tissue (n = 242) from the training cohort. (D) Percentage of patients with high intratumoral PAK6 expression increased moderately accompanied by disease progression from TNM stage I–IV (data from the trainging and validation cohort). Scale bar, 100 μm.

response to chemotherapy, whereas patients with high p21^{WAF1} expression patients had a favorable response to chemotherapy (Fig. 3A and S7A). Similar results were observed in the validation cohort (Fig. 3B and S7B).

We then performed a prediction accuracy analysis to evaluate the potential of combining multimarkers as a biomarker. The prediction accuracy reached a maximum, when the six features were combined (Fig. S8B). We then developed a SVM classifier – CS-SVM using PAK6, TS, ERCC1, COX2, p21^{WAF1} and Ki-67. In addition, the AUC of CS-SVM was higher than TNM stage and PAK6 (Fig. S8A).

3.3. Prognostic and Predictive Value of the CS-SVM Classifier for Chemotherapy Benefit

The characteristics of stage II and III patients received and not received adjuvant chemotherapy were similar in our study (Table S3). The CS-SVM classifier was not significantly associated with clinical characteristics of stage II and III patients (Table S4). Multivariate analysis showed that the CS-SVM classifier was an independent prognostic factor in Stage II and III GC patients (Table S5–7). For the chemotherapy patients, the 5-year DFS and OS were 27.8% and 35.4% for the low CS-SVM group, respectively, and were 75.9% and 86.2% for the high CS-SVM group, respectively (HR 0.239 (0.108–0.528), P = 0.0001; 0.236 (0.100–0.556), P = 0.0003; Fig. 4) in the training cohort. However, for the no chemotherapy patients, the difference was no such significant (HR 0.786 (0.440–1.404), P = 0.409; 0.770 (0.430–1.380), P = 0.373; Fig. 4). We also noted similar results in the validation cohort (Fig. 4). These results suggest that the CS-SVM classifier was strongly associated with chemotherapy.

For the high CS-SVM patients, the 5-year DFS and OS were 31.8% and 40.9% for the no chemotherapy patients, respectively, and were 75.9% and 86.2% for the chemotherapy patients, respectively (HR 0.256 (0.104–0.630), P = 0.001; 0.203 (0.078–0.526), P = 0.0003; Fig. 5), in the training cohort. However, the low CS-SVM patients did not obtain

Table 2
Multivariable Cox regression analysis of the PAK6 and survival in the training cohort and validation cohorts.

Variables	HR (95% CI)	p-Value
Disease-free survival		
Training cohort (n = 241)		
PAK6 (high vs. low)	1.467 (1.057–2.035)	0.022
TNM stage (III + IV vs. I + II)	2.880 (2.209–3.756)	<0.0001
Validation cohort (n = 342)		
PAK6 (high vs. low)	1.740 (1.297–2.333)	0.0002
TNM stage (III + IV vs. I + II)	1.496 (1.310–1.709)	<0.0001
Overall survival		
Training cohort (n = 241)		
PAK6 (high vs. low)	1.442 (1.011–2.057)	0.043
TNM stage (III + IV vs. I + II)	1.804 (1.475–2.207)	<0.0001
Validation cohort (n = 342)		
PAK6 (high vs. low)	1.813 (1.347–2.441)	<0.0001
TNM stage (III + IV vs. I + II)	1.486 (1.300–1.697)	<0.0001

CEA: carcino-embryonic antigen; CA199: carbohydrate antigen 199.

survival benefits from chemotherapy (HR 0.933 (0.632–1.376), $P = 0.722$; 0.732 (0.486–1.100), $P = 0.128$). We then performed the same subset analysis in the validation cohort. Chemotherapy significantly prolonged the DFS and OS of the high CS-SVM patients (HR 0.455 (0.264–0.786), $P = 0.004$; 0.418 (0.239–0.728), $P = 0.001$; respectively, Fig. 5) but was not beneficial and was potentially even detrimental to the low CS-SVM patients (HR 1.119 (0.761–1.634), $P = 0.562$; 1.061 (0.722–1.558), $P = 0.761$; Fig. 5). Then we performed a subset analysis for patients with stage II or stage III GC, and similar results were observed both in stage II and III cancer patients (Fig. S9–11).

3.4. Development of an Individualized Prediction Model

A Cox regression analysis identified the CS-SVM classifier, location, lauren type, depth of invasion, lymph nodes metastasis, and the chemotherapy status as independent predictors for OS (Table S7). To provide the clinician with a quantitative method to predict stage II and III GC patients' probability of 3- and 5-year OS, the model that incorporated the above independent predictors was developed and presented as the nomogram (Fig. 6A). To use the nomogram, first draw a vertical line up to the top Points row to assign points for each variable. Then, add up the total points and drop a vertical line from the Total Points row to obtain the 3-year OS and 5-year OS. To estimate the net survival benefit from adjuvant chemotherapy, first using the nomogram estimates the predicted survival without adjuvant chemotherapy, and then using the nomogram estimates survival with adjuvant chemotherapy. The difference between the two estimates is the expected net survival benefit from the addition of adjuvant chemotherapy.

The calibration plots showed that the nomogram performed well compared with an ideal model (Fig. 6B). The actual survival corresponded closely with the predicted survival and was always within the 7.5% margin of error in the validation cohort. Time-dependent ROC analysis was performed to assess the predictive accuracy of the nomogram. In the training cohort, the area under the curves (AUCs) at 1, 3 and 5 years were 0.751, 0.755 and 0.749 respectively (Fig. 6C). In the validation cohort, the AUCs at 1, 3 and 5 years were 0.780, 0.763 and 0.749 respectively (Fig. 6C).

3.5. Clinical Use

The decision curve analysis for the nomogram is presented in Fig. 6D–E. The decision curve showed that if the threshold probability of a patient or doctor is >10%, using the nomogram to predict 3 and 5 years OS adds more benefit than either the treat-all-patients scheme or the treat-none scheme. Within this range, net benefit was comparable, with several overlaps, on the basis of the nomogram.
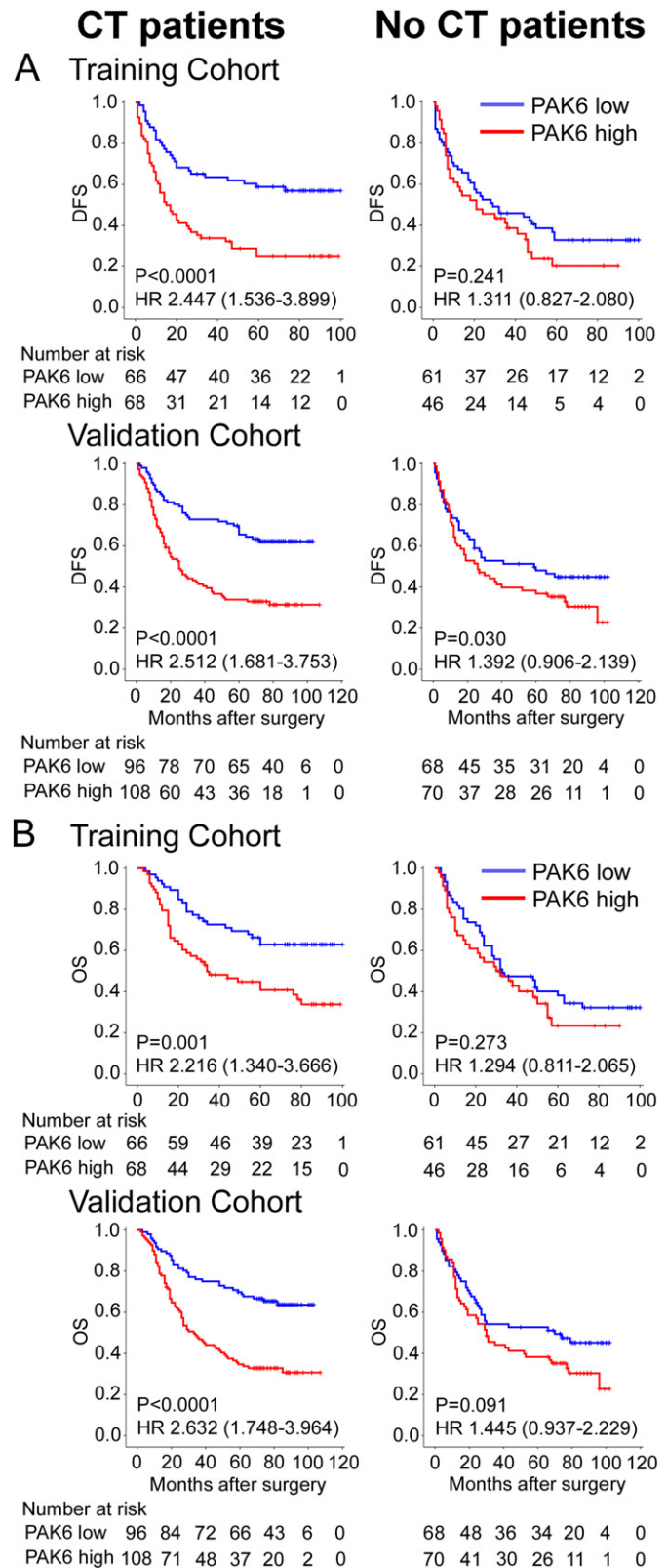
Fig. 2. Kaplan–Meier analysis of DFS and OS according to intratumoral PAK6 expression in GC patients. Left panel: CT patients; right panel: no CT patients. Training cohort: n = 241, validation cohort: n = 342.

4. Discussion

To our knowledge, this is the first study to identify the potential prognostic and predictive value of PAK6 in GC. Increased PAK6
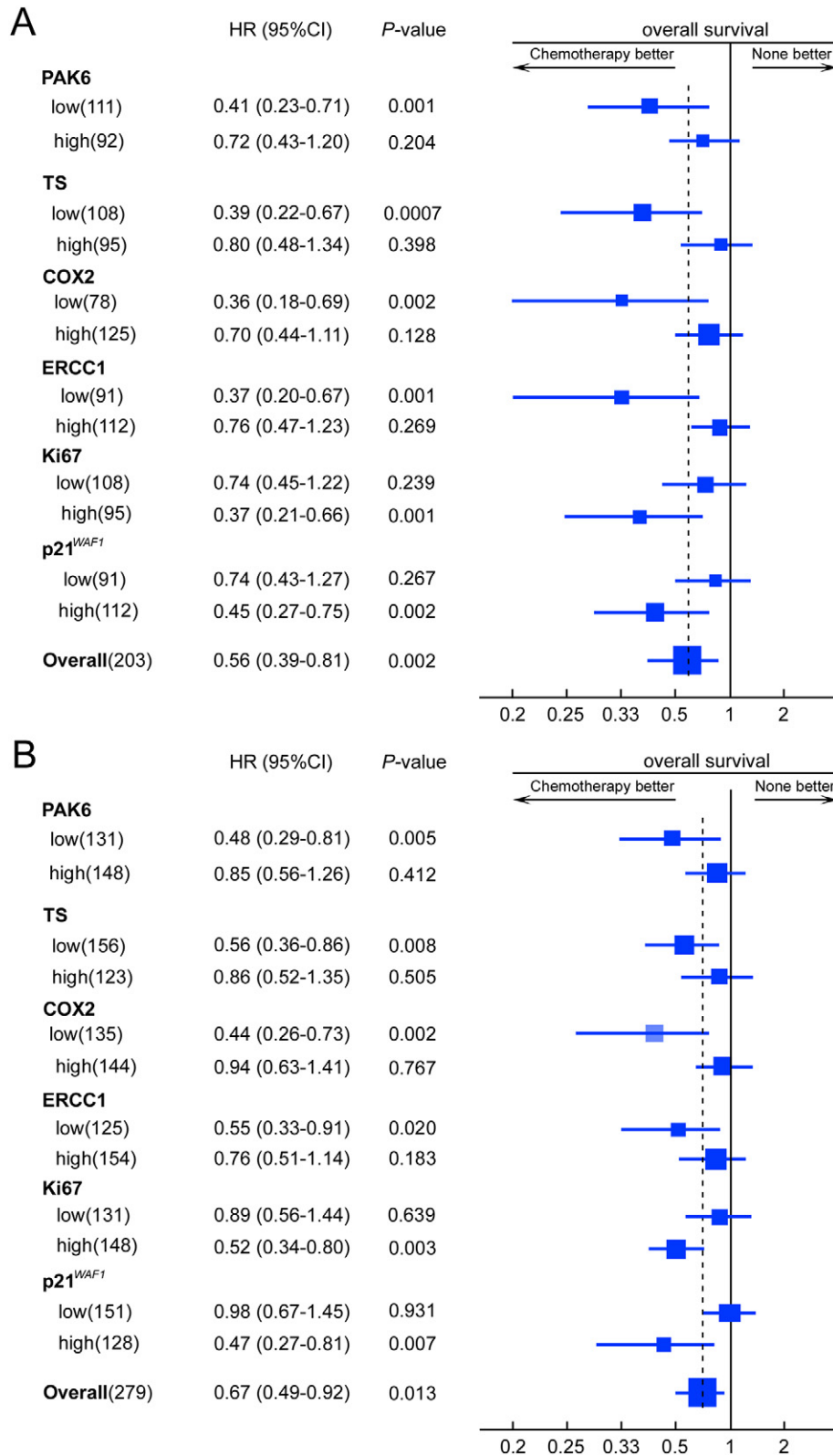
Fig. 3. Effect of adjuvant chemotherapy on overall survival (OS) in different subgroups of stage II and III GC patients. A, training cohort; B, validation cohort.

expression indicated a more aggressive phenotype and chemoresistance to 5-FU/oxaliplatin chemotherapy. We further assessed the predictive value of PAK6, COX2, p21^{WAF1}, Ki-67, TS, and ERCC1 in stage II and III GC and then constructed a SVM classifier integrating PAK6, Ki-67, TS, COX2, ERCC1, and p21^{WAF1}. The classifier effectively helped to identify a subgroup of stage II and III patients who

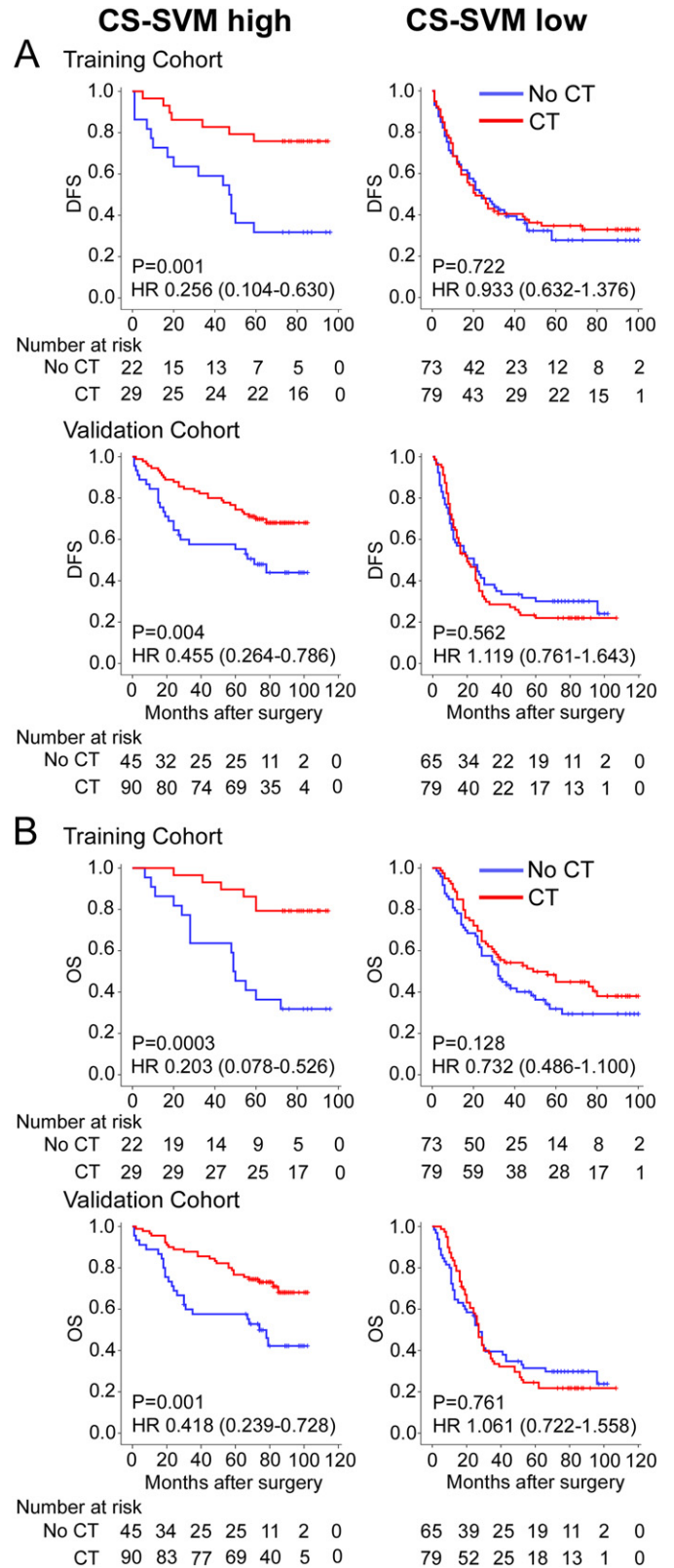would highly benefit from 5-FU/oxaliplatin chemotherapy, thus leading to a more personalized therapy.

Overexpression of PAK6 has previously been reported in primary and metastatic pancreatic cancer and has been shown to be further increased in recurrence tumors after androgen deprivation therapy.(Kaur et al., 2008) In addition, PAK6 expression has been found to be increased
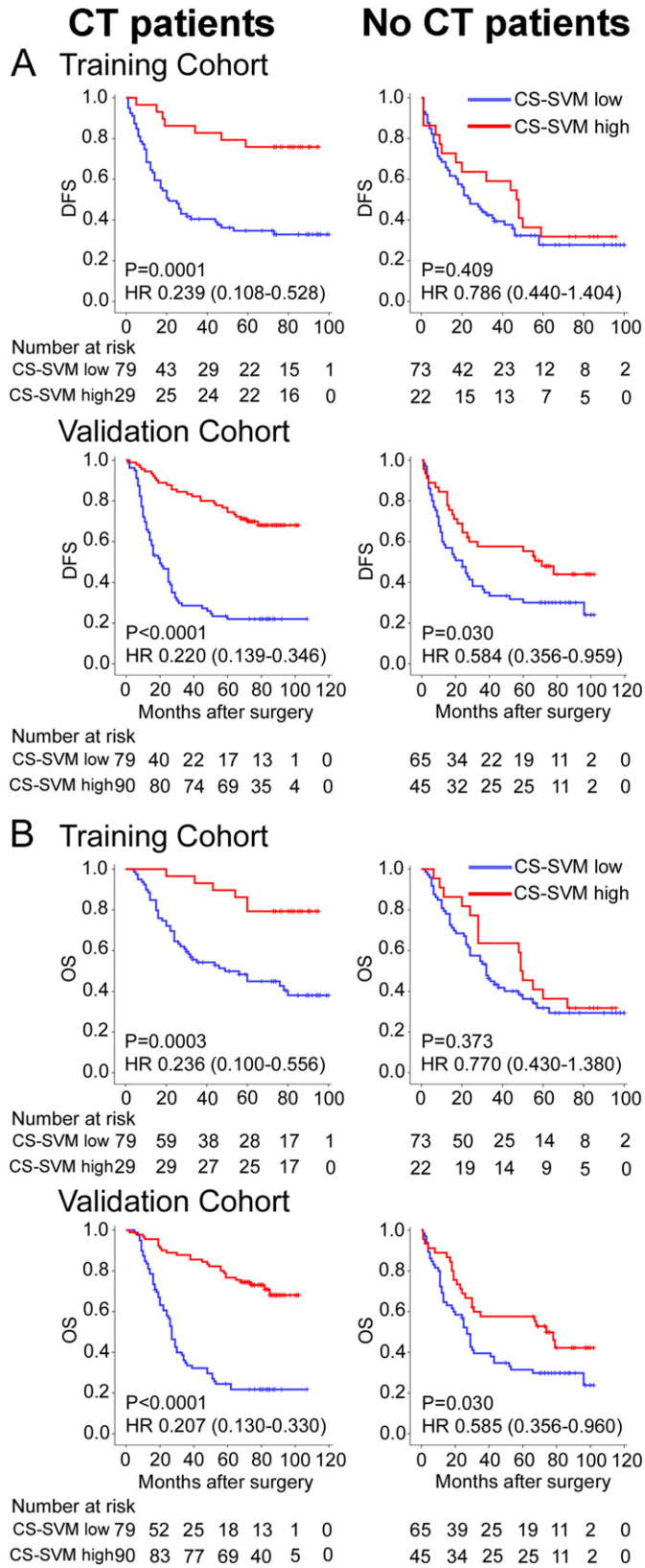
Fig. 4. Kaplan–Meier analysis of DFS and OS according to CS-SVM signature in stage II and III GC patients. Left panel: CT patients; right panel: no CT patients. Training cohort (n = 203), validation cohort (n = 279).

Fig. 5. Kaplan–Meier analysis of DFS and OS according to chemotherapy in stage II and III GC patients. Training cohort (n = 203), validation cohort (n = 279). Left panel: CS-SVM high patients; right panel: CS-SVM low patients.

in PCa, CRC and HCC (Chen et al., 2014, 2015) but decreased in ccRCC (Liu et al., 2014). PAK6 expression was also regarded as a potentially predictor for the differentiation of human uterine cervical

adenocarcinoma from squamous cell carcinoma (Lee et al., 2010). Furthermore, PAK6 decreased 5-FU drug susceptibility in colon cancer cells and was an independent prognostic factor for adjuvant 5-FU-based chemotherapy in patients with stage II and III colon cancer
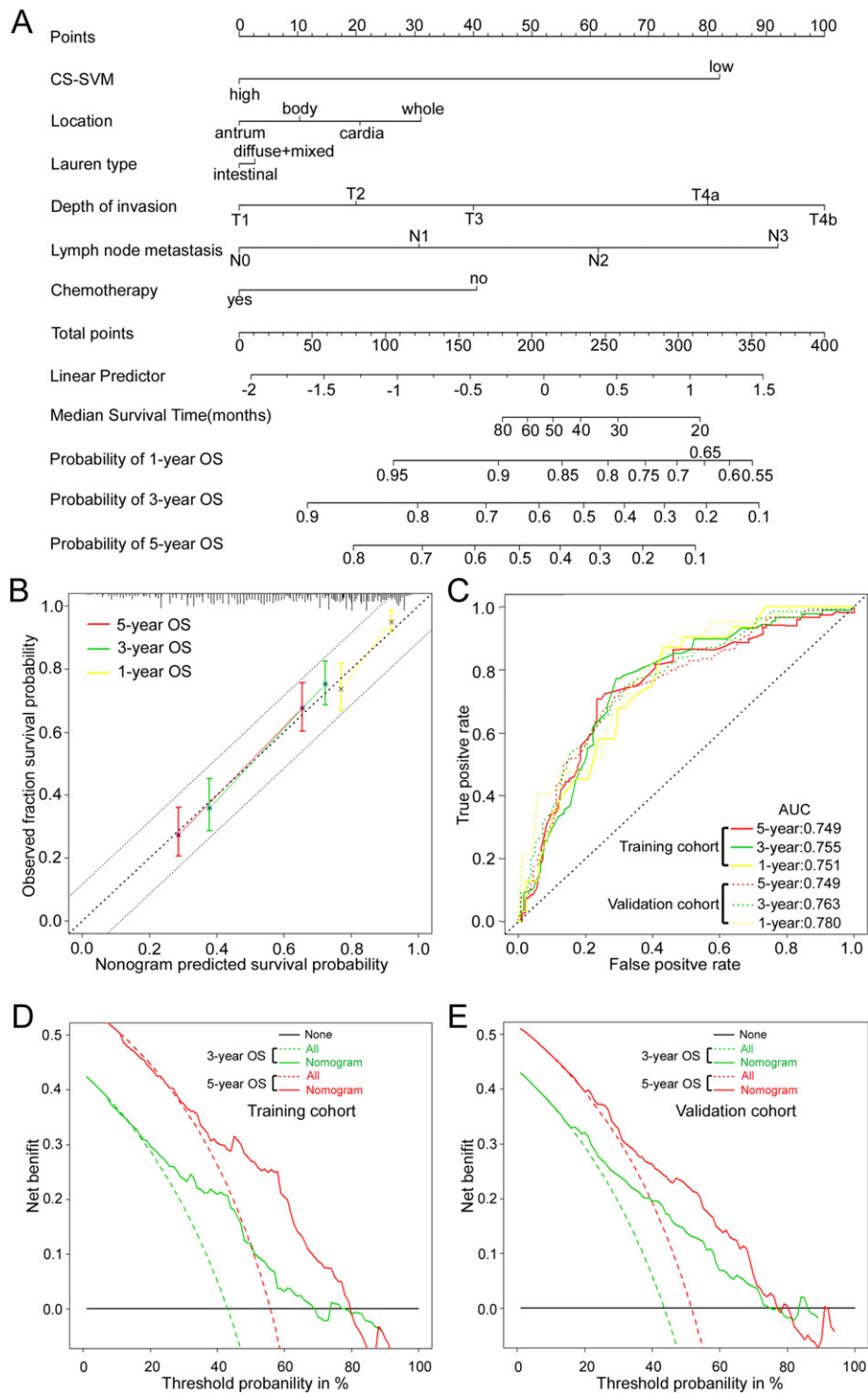
Fig. 6. Nomogram to predict 1-, 3- and 5-year survival probability in gastric cancer. (A) Nomogram for predicting proportion of GC patients with OS after surgery. (B) Plots depict the calibration of each model in terms of agreement between predicted and observed outcomes. Model performance is shown by the plot, relative to the 45-degree line, which represents perfect prediction. All predictions lie within a 7.5% margin of error (within the dashed line). (C) Time-dependent ROC curves by nomogram for 1-, 3- and 5-year OS probability in the training cohort and validation cohort. (D) (E) Decision curve analysis for the nomogram. The y-axis measures the net benefit. The red and blue solid line represents the nomogram for 3, 5-year OS. The dotted line represents the assumption that all patients have 3, 5-year OS. Thin black line represents the assumption that no patients have 3, 5-year OS. The net benefit was calculated by subtracting the proportion of all patients who are false positive from the proportion who are true positive, weighting by the relative harm of forgoing treatment compared with the negative consequences of an unnecessary treatment.(Vickers et al., 2008) Here, the relative harm was calculated by (p_t/[1 - p_t]). "p_t" (threshold probability) is where the expected benefit of treatment is equal to the expected benefit of avoiding treatment; at which time a patient will opt for treatment informs us of how a patient weighs the relative harms of false-positive results and false-negative results ([a–c]/[b–d] = [1 - pt]/pt); a - c is the harm from a false-negative result; b–d is the harm from a false-positive result. a, b, c and d give, respectively, the value of true positive, false positive, false negative, and true negative.(Vickers et al., 2008).

(Chen et al., 2015). High PAK6 expression in PCa may reduce chemosensitivty and decrease survival (Wen et al., 2009; Zhang et al., 2010). In the present study, we initially measured PAK6 expression by IHC in FFPE GC specimens and found that PAK6 had higher expression in GC tissues than in the surrounding nontumor mucosa. PAK6 expression was associated with poor prognosis of GC patients with chemotherapy, but was not significantly associated with prognosis in the no chemotherapy patients. Furthermore, high PAK6 expression increased chemoresistance in 5-FU/oxaliplatin chemotherapy patients.

Microsatellite instability (MSI) is a tumor phenotype linked to somatic or germline (Lynch syndrome) inactivating alterations of DNA mismatch repair genes. Results from a large cohort showed that MSI-H tumors were associated with a good prognosis in Stage II and III gastric cancer when patients were treated by surgery alone, and the benefits of MSI-H status were attenuated by chemotherapy (Kim et al., 2015). In the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC) trial, mismatch repair (MMR) deficiency and high MSI were associated with a positive prognostic effect in patients treated with surgery alone and a differentially negative prognostic effect in patients treated with chemotherapy (Smyth et al., 2017). According to a phase II study, MMR deficiency renders different solid tumors highly sensitive to immune checkpoint blockade with the PD-1 inhibitor pembrolizumab, and these tumors contain prominent immune infiltrates (Le et al., 2015; Lee et al., 2016). However, there were not studies about the relationship of PAK6 with MSI expression or immunotherapy. In the future, we will study the association of PAK6 with MSI expression and immunotherapy.

Until now, no valid prognostic biomarkers for GC chemotherapy have been established, although several potential molecular predictors of recurrence risk and chemotherapy benefit have been investigated (e.g., Ki-67, TS, TP, ERCC1, p21^{WAF1}, and immune cell infiltration) (Jiang et al., 2016; Kwon et al., 2007; Metzger et al., 1998; Xu et al., 2016; Zhang et al., 2016). So far, none has been used clinically as a predictive marker for the efficacy of chemotherapy, because they are less sensitivity or specificity (Kwon et al., 2007; Metzger et al., 1998; Xu et al., 2016). SVMs can combine clinicopathological features with independently informative markers and be used to select a small subset of patients attributes to build reliable cancer classifiers, which may significantly improve the predictive accuracy (Wang et al., 2011). GC patients are considered for adjuvant chemotherapy if they are deemed to be high risk of relapse on the basis of clinical and pathological evaluation (Razzak, 2014). However, these clinicopathological factors do not clearly identify patients who have a high or low risk of disease recurrence and do not predict which patients are likely to specifically benefit from chemotherapy (Jiang et al., 2016; McLean and El-Omar, 2014; Wadhwa et al., 2013). The administration of 5-FU/oxaliplatin chemotherapy to all of the stages II and III GC patients is unnecessary (Noh et al., 2014), so selecting the subset of patients that is sensitive to chemotherapy is meaningful.

5-FU and oxaliplatin are key drugs for the GC management. Pharmacogenetic variability in metabolizing enzymes of 5-FU and folate is a major determinant of the sensitivity to 5-FU and prognosis in GC (Longley et al., 2003). TS is a target enzyme of 5-FU; overexpression of TS at the protein and messenger RNA (mRNA) levels has been found to be correlated with resistance to 5-FU chemotherapy in CRC and GC patients (Choi et al., 2001). Oxaliplatin/DNA adducts are repaired by the nucleotide excision repair (NER) system, and two enzymes, XPA and ERCC1, have been found to be essential for this repair process (Raymond et al., 2002). Moreover, low ERCC1 expression is known to be correlated with prolonged survival after oxaliplatin or cisplatin-based chemotherapy in non-small-cell lung cancer, CRC and GC (Kwon et al., 2007; Lord et al., 2002). In addition, ERCC1 and TS mRNA levels have been shown to predict the response rate and prognosis of GC patients on combination cisplatin and fluorouracil chemotherapy (Metzger et al., 1998). Furthermore, irinotecan or taxane-based regimens has been shown to be useful to treat ERCC1-positive advanced

GC patients (Morabito et al., 2009). COX-2 plays an important role in prostaglandin synthesis and mediates angiogenesis, tumor growth, tumor invasiveness, and metastasis (Li et al., 2016). Low COX-2 expression may predict the tumor response to CRT (Xu et al., 2016). The total cell proliferative activity can be evaluated by assessing the Ki-67 expression, which is the strongest in the S phase, but is observed in all cell cycle phases other than G0. Increased expression of Ki-67 is an indicator of poor prognosis in patients who are not responding to chemotherapy but is also an indicator of good prognosis in patients with response to chemotherapy (Yerushalmi et al., 2010). The p21 protein is a cyclin-dependent kinase inhibitor that is a downstream effector of p53-dependent cell cycle regulation. Expression of p21^{WAF1} is associated with the good prognosis from 5-FU-based adjuvant chemotherapy in CRC (Sulzyc-Bielicka et al., 2011). In some way, the p21^{WAF1} may be associated with PAK6, but their responses to chemotherapy were different. The reasons are not clear now and would be explored in further studies. Whereas, the predictive values of TS, ERCC1, COX2, Ki-67 and p21^{WAF1} for 5-FU/oxaliplatin chemotherapy in stage II and III GC are still controversial. In the present study, we found that the increased expression levels of TS, ERCC1, COX2, and Ki-67 were related to chemoresistance of 5-FU/oxaliplatin chemotherapy in GC patients, whereas a high expression level of p21^{WAF1} indicated increased chemosensitivity to 5-FU/oxaliplatin chemotherapy.

The CS-SVM classifier can select a small subset of stage II and III patients who can specifically benefit from 5-FU/oxaliplatin chemotherapy. Patients identified as high CS-SVM benefited significantly from chemotherapy in the training cohort (DFS, $P = 0.001$; OS, $P = 0.0003$; Fig. 5). In contrast, low CS-SVM GC patients did not benefit from chemotherapy (DFS, $P = 0.722$; OS, $P = 0.128$). For patients with low CS-SVM, more effective systemic approaches to improve treatment outcomes may need to be identified. We validated the predictive value of the CS-SVM classifier in different populations of the validation cohort (Fig. 5). Then, to provide the clinician and patients with a quantitative method to predict stage II and III GC patients' probability of 3- and 5-year OS, the model that incorporated the CS-SVM classifier and five clinicopathological risk factors was developed and presented as the nomogram. Validation of the nomogram was performed by using calibration plots and values of AUC. The nomogram performed well with a good calibration. Besides, the AUCs were satisfactory. Using the nomogram could estimate the expected net survival benefit from the addition of adjuvant chemotherapy for every stage II and III GC patient. Thus, the models may be useful for both clinicians and patients in patient counseling and individualized adjuvant treatment decision-making, as well as follow-up scheduling. Whereas, validation by other cohorts is required for the generalized use of the nomogram as the basis for postoperative treatment recommendations.

The AUCs of our CS-SVM and nomogram were decent but not great, slightly lower than those reported in previous studies (He et al., 2013). One possible explanation is the difference in patient populations. Previous studies included many patients with stage I GC whose prognosis was excellent and/or with stage IV GC whose prognosis was poor (He et al., 2013). This study, on the other hand, included only patients with stage II and III GC whose prognoses varied widely. When patients with stage I and/or IV GC were included in this study, the AUCs of our models were much higher (data not shown). The AUCs of our CS-SVM and nomogram were superior to that of the TNM stage grouping (Fig. 6 and S5). The calibration plots indicated that actual survival corresponded closely with predicted survival and was always within a 7.5% margin from the ideal reference line; this suggests that the nomogram was well predictive.

There are some limitations to our study. First, our nomogram was developed and validated using data from almost exclusively Chinese patients. Our current study is also limited in that it was conducted retrospectively, making it susceptible to the inherent biases of such a study format. The use of adjuvant chemotherapy was not within a randomized comparison and the decision to treat or not treat patients after surgery

was met by the patients and/or clinicians. Nevertheless, the clinical characteristics of patients with and without adjuvant chemotherapy were similar in the training and validation cohorts (Table S5). Clearly, our results should be further validated by prospective studies in multi-center clinical trials. Other predictive variables may be included to improve performance of this model. As more specific patients and tumor information becomes routinely collected in the future, such as genetic information and other molecular tumor markers, use of these types of predictive models will become increasingly important. In addition, the application of the CS-SVM classifier depends on IHC results that are only available after surgery. Therefore, the CS-SVM classifier has limited impact on alternative treatments before surgery, including the use of neoadjuvant chemotherapy.

Future efforts will explore to test our model performance in external validation of other population. We will also seek the possibility of including additional prognostic and predictive variables to further improve the predictive accuracy of the model. Other regression modeling techniques will also be exploited to determine whether predictive accuracy could be further improved.

In summary, our study indicates that PAK6 was a predictive biomarker and increased chemoresistance to 5-FU/oxaliplatin chemotherapy. The CS-SVM classifier distinguished a subgroup of stage II and III patients who would highly benefit from chemotherapy.

Funding

Conflicts of Interest

The authors have declared that no competing interests exist.

Guarantor of the Article

G.-X. Li, W. Liu, S.-R. Cai and J. Yu.

Author Contributions

G.-X. Li, W. Liu, S.-R. Cai, Y. Yang and J. Yu conceived and designed the research. Y.-M. Jiang, T.-J. Li, Y.-F. Hu, S.-L. Chen, S.-J. Xi, Y.-J. Wen, X.-H. Huang, Z. Han, L.-Y. Zhao and C.-C. Xiao performed the research. Y.-F. Hu, L.-Y.Z., X.-H. Huang, H. Liu, X.-L. Qi, S.-R. Cai and J. Yu. provided samples and clinical data. Y.-M. Jiang, T.-J. Li, Y.-F. Hu, S.-J. Xi, and J. Yu analyzed the data. Y.-M. Jiang, S.-R. Cai and J. Yu participated in evaluation of results. Y.-M. Jiang, W. Liu, T.-J. Li and Y.-F. Hu wrote the manuscript, and all authors reviewed and approved the manuscript for publication.

Supplementary Materials

Supplemental material includes Supplementary methods, eleven supplementary figures, and eight supplementary tables.

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.ebiom.2017.06.028.

References

Chen, H.Y., Yu, S.L., Chen, C.H., et al., 2007. A five-gene signature and clinical outcome in non-small-cell lung cancer. N. Engl. J. Med. 356, 11–20.

Chen, H., Miao, J., Li, H., et al., 2014. Expression and prognostic significance of p21-activated kinase 6 in hepatocellular carcinoma. J. Surg. Res. 189, 81–88.

Chen, J., Lu, H.J., Yan, D.W., et al., 2015. PAK6 increase chemoresistance and is a prognostic marker for stage II and III colon cancer patients undergoing 5-FU based chemotherapy. Oncotarget 6, 355–367.

Choi, J., Lim, H., Nam, D.K., et al., 2001. Expression of thymidylate synthase in gastric cancer patients treated with 5-fluorouracil and doxorubicin-based adjuvant chemotherapy after curative resection. Br. J. Cancer 84, 186–192.

Choi, H., Yeo, D., Kwon, S., Kim, Y., 2011. Gene selection and prediction for cancer classification using support vector machines with a reject option. Comput. Stat. Data Anal. 55, 1897–1908.

Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 350, g7594.

He, H.Y., Wang, C., Shen, Z.B., et al., 2013. Upregulated expression of C-X-C chemokine receptor 4 is an independent prognostic predictor for patients with gastric cancer. PLoS One 8.

Jiang, Y., Zhang, Q., Hu, Y., et al., 2016. ImmunoScore signature: a prognostic and predictive tool in gastric cancer. Ann. Surg. http://dx.doi.org/10.1097/SLA.0000000000002116 [Epub ahead of print].

Kaur, R., Yuan, X., Lu, M.L., Balk, S.P., 2008. Increased PAK6 expression in prostate cancer and identification of PAK6 associated proteins. Prostate 68, 1510–1516.

Kim, S.Y., Choi, Y.Y., An, J.Y., et al., 2015. The benefit of microsatellite instability is attenuated by chemotherapy in stage II and stage III gastric cancer: results from a large cohort with subgroup analyses. Int. J. Cancer 137, 819–825.

Kwon, H.C., Roh, M.S., Oh, S.Y., et al., 2007. Prognostic value of expression of ERCC1, thymidylate synthase, and glutathione S-transferase P1 for 5-fluorouracil/oxaliplatin chemotherapy in advanced gastric cancer. Ann. Oncol. 18, 504–509.

Le, D.T., Uram, J.N., Wang, H., et al., 2015. PD-1 blockade in tumors with mismatch-repair deficiency. N. Engl. J. Med. 372, 2509–2520.

Lee, E.J., McClelland, M., Wang, Y.P., et al., 2010. Distinct DNA methylation profiles between adenocarcinoma and squamous cell carcinoma of human uterine cervix. Oncol. Res. 18, 401–408.

Lee, V., Murphy, A., Le, D.T., Diaz Jr., L.A., 2016. Mismatch repair deficiency and response to immune checkpoint blockade. Oncologist 21 (10):1200–1211. http://dx.doi.org/10.1634/theoncologist.2016-0046 [Epub 2016 Jul 13].

Li, T., Zhang, Q., Jiang, Y., et al., 2016. Gastric cancer cells inhibit natural killer cell proliferation and induce apoptosis via prostaglandin E2. Oncoimmunology 5, e1069936.

Li, T.J., Jiang, Y., Hu, Y.F., et al., 2017. Interleukin-17-producing neutrophils link inflammatory stimuli to disease progression by promoting angiogenesis in gastric cancer. Clin. Cancer Res. 23 (6):1575–1585. http://dx.doi.org/10.1158/1078-0432.CCR-16-0617 [Epub 2016 Sep 12].

Liu, W., Liu, H., Liu, Y., et al., 2014. Prognostic significance of p21-activated kinase 6 expression in patients with clear cell renal cell carcinoma. Ann. Surg. Oncol. 21 (Suppl. 4), S575–S583.

Localio, A.R., Goodman, S., 2012. Beyond the usual prediction accuracy metrics: reporting results for clinical decision making. Ann. Intern. Med. 157, 294–295.

Longley, D.B., Harkin, D.P., Johnston, P.G., 2003. 5-fluorouracil: mechanisms of action and clinical strategies. Nat. Rev. Cancer 3, 330–338.

Lord, R.V., Brabender, J., Gandara, D., et al., 2002. Low ERCC1 expression correlates with prolonged survival after cisplatin plus gemcitabine chemotherapy in non-small cell lung cancer. Clin. Cancer Res. 8, 2286–2291.

McLean, M.H., El-Omar, E.M., 2014. Genetics of gastric cancer. Nat. Rev. Gastroenterol. Hepatol. 11, 664–674.

Metzger, R., Leichman, C.G., Danenberg, K.D., et al., 1998. ERCC1 mRNA levels complement thymidylate synthase mRNA levels in predicting response and survival for gastric cancer patients receiving combination cisplatin and fluorouracil chemotherapy. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 16, 309–316.

Morabito, A., Carillio, G., Longo, R., 2009. Systemic treatment of gastric cancer. Crit. Rev. Oncol. Hematol. 70, 216–234.

Noh, S.H., Park, S.R., Yang, H.K., et al., 2014. Adjuvant capecitabine plus oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): 5-year follow-up of an open-label, randomised phase 3 trial. Lancet Oncol. 15, 1389–1396.

Radu, M., Semenova, G., Kosoff, R., Chernoff, J., 2014. PAK signalling during the development and progression of cancer. Nat. Rev. Cancer 14, 13–25.

Raymond, E., Faivre, S., Chaney, S., Woynarowski, J., Cvitkovic, E., 2002. Cellular and molecular pharmacology of oxaliplatin. Mol. Cancer Ther. 1, 227–235.

Razzak, M., 2014. Genetics: new molecular classification of gastric adenocarcinoma proposed by The Cancer Genome Atlas. Nat. Rev. Clin. Oncol. 11, 499.

Sauerbrei, W., Boulesteix, A.L., Binder, H., 2011. Stability investigations of multivariable regression models derived from low- and high-dimensional data. J. Biopharm. Stat. 21, 1206–1231.

Smyth, E.C., Wotherspoon, A., Peckitt, C., et al., 2017. Mismatch repair deficiency, microsatellite instability, and survival: an exploratory analysis of the medical research council adjuvant gastric infusional chemotherapy (MAGIC) trial. JAMA Oncol. http://dx.doi.org/10.1001/jamaoncol.2016.6762 [Epub ahead of print].

Sulzyc-Bielicka, V., Domagala, P., Urasinska, E., et al., 2011. Expression of p21WAF1 in Astler-Coller stage B2 colorectal cancer is associated with survival benefit from 5FU-based adjuvant chemotherapy. Virchows Arch. 458, 431–438.

Torre, L.A., Bray, F., Siegel, R.L., et al., 2015. Global cancer statistics, 2012. CA Cancer J. Clin. 65, 87–108.

Vapnik, V.N., 1999. An overview of statistical learning theory. IEEE Trans. Neural Netw. 10, 988–999.

Vickers, A.J., Cronin, A.M., Elkin, E.B., Gonen, M., 2008. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med. Inform. Decis. Making 8, 53.

Wadhwa, R., Song, S., Lee, J.S., et al., 2013. Gastric cancer-molecular and clinical dimensions. Nat. Rev. Clin. Oncol. 10, 643–655.

Wang, H.Y., Sun, B.Y., Zhu, Z.H., et al., 2011. Eight-signature classifier for prediction of nasopharyngeal [corrected] carcinoma survival. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 29, 4516–4525.

Wen, X., Li, X., Liao, B., et al., 2009. Knockdown of p21-activated kinase 6 inhibits prostate cancer growth and enhances chemosensitivity to docetaxel. Urology 73, 1407–1411.

Xu, J., Ding, T., He, Q., et al., 2012. An in situ molecular signature to predict early recurrence in hepatitis B virus-related hepatocellular carcinoma. J. Hepatol. 57, 313–321.

Xu, H.B., Shen, F.M., Lv, Q.Z., 2016. Celecoxib enhanced the cytotoxic effect of cisplatin in chemo-resistant gastric cancer xenograft mouse models through a cyclooxygenase-2-dependent manner. Eur. J. Pharmacol. 776, 1–8.

Yerushalmi, R., Woods, R., Ravdin, P.M., Hayes, M.M., Gelmon, K.A., 2010. Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol. 11, 174–183.

Zhang, M., Siedow, M., Saia, G., Chakrayarti, A., 2010. Inhibition of p21-activated kinase 6 (PAK6) increases radiosensitivity of prostate cancer cells. Prostate 70, 807–816.

Zhang, H., Liu, H., Shen, Z., et al., 2016. Tumor-infiltrating neutrophils is prognostic and predictive for postoperative adjuvant chemotherapy benefit in patients with gastric cancer. Ann. Surg. http://dx.doi.org/10.1097/SLA.0000000000002058 [Epub ahead of print].

Zhu, Z.H., Sun, B.Y., Ma, Y., et al., 2009. Three immunomarker support vector machines-based prognostic classifiers for stage IB non-small-cell lung cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 27, 1091–1099.