

A novel molecular and clinical staging model to predict survival for patients with esophageal squamous cell carcinoma

Wei Wang^{1,*}, Zhiwei Wang^{2,3,*}, Jun Zhao⁴, Min Wei², Xinghua Zhu⁵, Qi He², Tianlong Ling³, Xiaoyan Chen⁶, Ziang Cao³, Yixin Zhang¹, Lei Liu¹, Minxin Shi¹

¹Department of Surgery, The Affiliated Tumor Hospital of Nantong University, Nantong, Jiangsu Province, China

²Department of Breast, International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University, Shanghai, China

³Department of Thoracic Surgery, Shanghai Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

⁴Department of Thoracic and Cardiovascular Surgery, The First Affiliated Hospital of Soochow University, Medical College of Soochow University, Suzhou, Jiangsu Province, China

⁵Department of Pathology, The Affiliated Tumor Hospital of Nantong University, Nantong, Jiangsu Province, China

⁶Department of Pathology, Shanghai Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

*These authors have contributed equally to this work

Correspondence to: Lei Liu, **email:** liulei5602@126.com
Minxin Shi, **email:** Docshimx@126.com

Keywords: esophageal squamous cell carcinoma, UBE2C, MGP, survival

Received: December 08, 2015

Accepted: July 26, 2016

Published: August 18, 2016

ABSTRACT

Current prognostic factors fail to accurately determine prognosis for patients with esophageal squamous cell carcinoma (ESCC) after surgery. Here, we constructed a survival prediction model for prognostication in patients with ESCC. Candidate molecular biomarkers were extracted from the Gene Expression Omnibus (GEO), and Cox regression analysis was performed to determine significant prognostic factors. The survival prediction model was constructed based on cluster and discriminant analyses in a training cohort (N=205), and validated in a test cohort (N=207). The survival prediction model consisting of two genes (UBE2C and MGP) and two clinicopathological factors (tumor stage and grade) was developed. This model could be used to accurately categorize patients into three groups in the test cohort. Both disease-free survival and overall survival differed among the diverse groups (P<0.05). In summary, we have developed and validated a predictive model that is based on two gene markers in conjunction with two clinicopathological variables, and which can accurately predict outcomes for ESCC patients after surgery.

INTRODUCTION

Esophageal carcinoma ranks among the top ten most common malignancies in the world [1]. In China, more than 90% of esophageal cancer cases are esophageal squamous cell carcinoma (ESCC), which is the sixth most prevalent cancer in the country [2]. Despite recent improvements in surgical techniques and medical treatments, the overall prognosis for patients with ESCC remains poor [3–6]. Similarly as in other solid cancers, TNM staging has been widely used to estimate survival and make clinical decisions about adjuvant therapy

for ESCC patients [7]. Indeed, this staging system has improved survival of ESCC patients in the past. However, current staging methods and therapeutic decisions remain suboptimal.

ESCC patients with the same stage and similar treatment may have quite different clinical outcomes. Importantly, patients with early-stage cancer and low risk of recurrence could be spared the toxicity of systemic treatment if clearly distinguished, while others at high risk of distant recurrence could get maximal benefit from adjuvant therapy. It is increasingly recognized that a tremendous heterogeneity between patients exists in

the biology underlying ESCC. Hence, the ideal staging system would take into account the biology and molecular features of each individual tumor and correlate prognosis with patient-specific tumor biomarkers.

Recently, gene expression profiles enabled to predict outcomes and select patients for adjuvant therapies in colon cancer, lung cancer, hepatocellular carcinoma, breast cancer, and esophageal adenocarcinoma [8–13]. Unfortunately, practical predictive models for the outcomes of patients with ESCC after curative resection are not currently available. In this study, we investigated the prognostic value of the traditional clinicopathological factors and selected protein expression in a training cohort. We aimed to develop a novel predictive model, which would be capable to predict outcomes of ESCC patients after surgery.

RESULTS

Candidate gene selection

During data analysis, we found that 8171 genes were differentially expressed. Among these genes, 3972 genes were up-regulated and 4199 genes were down-regulated. From 75 microarray samples, 115 genes were identified, and 8 genes were randomly selected for further analysis: MGP, SPP1, SULF1, NECAB1, SPARC, GLIPR1, UBE2C and KCNB1 (Supplementary Table S1).

Patients and follow-up

Patients' follow-up was performed until the end of September 2014. The median follow-up time was 48 months (range 2 to 105 months) in the training cohort and 50 months (range 2 to 93 months) in the validation cohort. During the follow-up period, a total of 153 all-cause deaths were observed in the training cohort and 115 in the validation cohort. 160 patients had recurrence or metastasis in the training cohort and 140 in the validation cohort. No differences in the distributions of basic characteristics were observed between training and validation cohorts ($P>0.05$) (Table 1).

Gene expression analysis

Quantitative RT-PCR was carried out to analyze expression of the eight selected genes in 100 tumor and non-tumor tissues. Consistent with the results from microarray data analysis, two of the eight genes, UBE2C and MGP, were highly expressed in tumor tissues (Figure 1A). However, there was no significant difference between tumor and non-tumor tissues in the mRNA levels of the other six genes (SPP1, SULF1, NECAB1, SPARC, GLIPR1 and KCNB1) (Supplementary Figure S1). Hence, UBE2C and MGP genes were selected for survival model training.

UBE2C and MGP protein expression analysis

To evaluate the potential use of UBE2C and MGP as biomarkers for ESCC, we analyzed the protein levels of UBE2C and MGP in 205 training ESCC tumor tissues by immunohistochemistry (IHC). For both proteins examined, we scored nuclear, cytoplasmic, and membrane staining. In the carcinoma cells, UBE2C was found both in the nuclei and in the cytoplasm. In contrast, MGP was found exclusively in the cytoplasm in the carcinoma cells. The positive rates of UBE2C and MGP in ESCC tumor tissues were up to 43.4% (89/205) and 41.5% (85/205), respectively (Figure 1B). The UBE2C protein expression levels of five human ESCC cell lines (KYSE-70, KYSE-110, KYSE-140, KYSE-220, KYSE-510) and a human normal esophageal squamous epithelial cell line (Het-1A) were assessed by western blot (Figure 1C). The KYSE-70, KYSE-110, KYSE-140 and KYSE-220 ESCC cells showed high-UBE2C expression, as compared to normal epithelial cell with low UBE2C expression.

Construction of survival prediction model for training cohort

For training cases, univariate analysis showed that the factors of T stage, lymph nodes status, TNM stage, tumor grade, UBE2C expression and MGP expression were significantly associated with disease-free survival (DFS) ($P<0.05$) (Table 2). These prognostic factors were used to perform unsupervised hierarchical clustering analysis. The tree-view program displays the data with the prognostic factors on the horizontal axis and the cases on the vertical axis. The cases and prognostic factors are arranged in such a way that the cases with the most similar expression profiles are placed next to each other. The length of the dendrogram branches connecting the cases or prognostic factors, is inversely proportional to the similarity of their profiles. Clustering of 205 training ESCC cases produced three groups, groups A, B, and C, which can be discerned from the dendrogram (Figure 2A). Groups A, B and C contained 52, 85 and 68 cases, respectively.

Next, discriminant analysis was used to establish the survival prediction model by clustering results and to evaluate the efficiency of classification. The model was obtained as follows:

$$Y1=5.400T-5.763N+6.115TNM+4.066Grade+0.01UBE2C+0.024MGP-14.340;$$

$$Y2=4.334T-4.804N+5.726TNM+4.205Grade-0.026UBE2C+0.034MGP-12.399;$$

$$Y3=5.078T-4.984N+5.964TNM+3.869Grade+0.026UBE2C-0.024MGP-13.085;$$

$Y(1,2,3)$ is the probability of groups A, B and C. Classifying rule: if $Y1>Y2$ and $Y3$, group A case; if $Y2>Y1$ and $Y3$, group B case; if $Y3>Y1$ and $Y2$, group C case. Correct classification rate reached 84.9%. Using the results from discriminant analysis, Kaplan–Meier curves

Table 1: Characteristics of study participants in the training and validation datasets

Variables	Training (n=205)		Validation (n=207)		P-value
	No.	%	No.	%	
Age (years)					0.107
<60	78	38.0	95	45.1	
≥60	127	62.0	112	54.9	
Gender					0.263
Male	149	72.7	140	67.6	
Female	56	27.3	67	32.4	
Tumor size					0.139
<5cm	93	45.4	109	52.7	
≥5cm	112	54.6	98	47.3	
T stage					0.311
pT1	31	15.1	43	20.8	
pT2	55	26.8	61	29.5	
pT3	104	50.7	88	42.5	
pT4	15	7.3	15	7.2	
Lymph nodes status					0.084
Negative	126	61.5	144	69.6	
Positive	79	38.5	63	30.4	
TNM stage					0.183
I	53	25.9	72	34.1	
II	68	33.1	63	29.9	
III	84	41.0	76	36.0	
Grade					0.750
Well-differentiated	30	14.6	34	16.4	
Moderately-differentiated	102	49.8	106	51.2	
Poorly-differentiated	73	35.6	67	32.4	
Vascular invasion					0.505
Positive	16	7.8	20	9.7	
Negative	189	92.2	187	90.3	

showed that group C cases had a better DFS, whereas group A cases had a worse DFS. The 5-year DFS rate for groups A, B and C patients was 17.9%, 31.5% and 43.2%, respectively (P <0.001) (Figure 3A).

The Cox regression analysis showed that T stage, lymph nodes status, TNM stage, tumor grade, vascular invasion, UBE2C expression and MGP expression were significantly associated with overall survival (OS) (p<0.05) (Table 2). These prognostic factors were used to perform unsupervised hierarchical clustering analysis. Two hundred and seven training cases were divided into

three groups. Groups A, B and C included 53, 69 and 83 cases, respectively (Figure 2B). Discriminant analysis showed the prediction model for OS as follows:

$$Y1=4.326T-4.608N+6.007 \quad TNM+4.058Grade-0.025UBE2C+0.034MGP-12.468;$$

$$Y2=5.138T-5.077N+5.671TNM+4.001Grade+0.026UBE2C-0.025MGP-12.955;$$

$$Y3=5.379T-5.394N+6.497TNM+3.839Grade+0.004UBE2C+0.019MGP-14.312;$$

Correct classification rate reached 82.9%. Using the results from discriminant analysis, the survival results

showed that group C cases had a better OS, whereas group A cases had a worse OS. The 5-year OS rate for groups A, B and C patients was 22.2%, 37.5% and 47.8%, respectively ($P < 0.001$) (Figure 3B).

Survival prediction for testing patients

The survival prediction model was applied to an independent test group of 207 ESCC patients (Figure 4A). The positive rates of UBE2C and MGP in ESCC tumor tissues were up to 56.5% (117/207) and 53.6% (111/207), respectively (Figure 4B). For DFS, 65, 94 and 48 cases were classified as groups A, B and C respectively. As shown in Figure 4C, group C cases had a significantly better DFS, whereas group A cases had a significantly worse DFS. The 5-year DFS rate for groups A, B and C patients was 28.7%, 38.5% and 57.1%, respectively ($P = 0.006$). With regard to overall survival, the predictive results showed that 55, 99 and 53 cases were classified as groups A, B and C respectively. As shown in Figure 4D, there was a significant difference in OS among diverse groups. The 5-year OS rate for groups A, B and C patients was 32.1%, 47.1% and 58.6%, respectively ($P = 0.002$).

DISCUSSION

The clinical outcomes of ESCC are heterogeneous, and the available prognostic markers are limited. Currently, only the stage based on TNM classification is widely accepted as a prognostic factor, which is still far from an accurate prediction. To improve prognosis, we hypothesized that a combination of clinical features and well-validated molecular markers would allow stratification of ESCC patients into clinically meaningful prognostic subgroups. In our study, eight genes were initially selected from microarrays. After confirming their increased expression in tumor tissues by quantitative RT-PCR, two of the eight genes were included in constructing the survival prediction model. Finally, a predictive model containing two proteins and traditional staging was developed in a training cohort. Examination of training samples resulted in a predictive accuracy of more than 80%. Both training cases and testing cases had a worse survival. It is possible that higher tumor stage and overexpression of UBE2C and MGP in group A have contributed to the observed trend of worse outcome, because both factors have been associated with poor survival in ESCC.

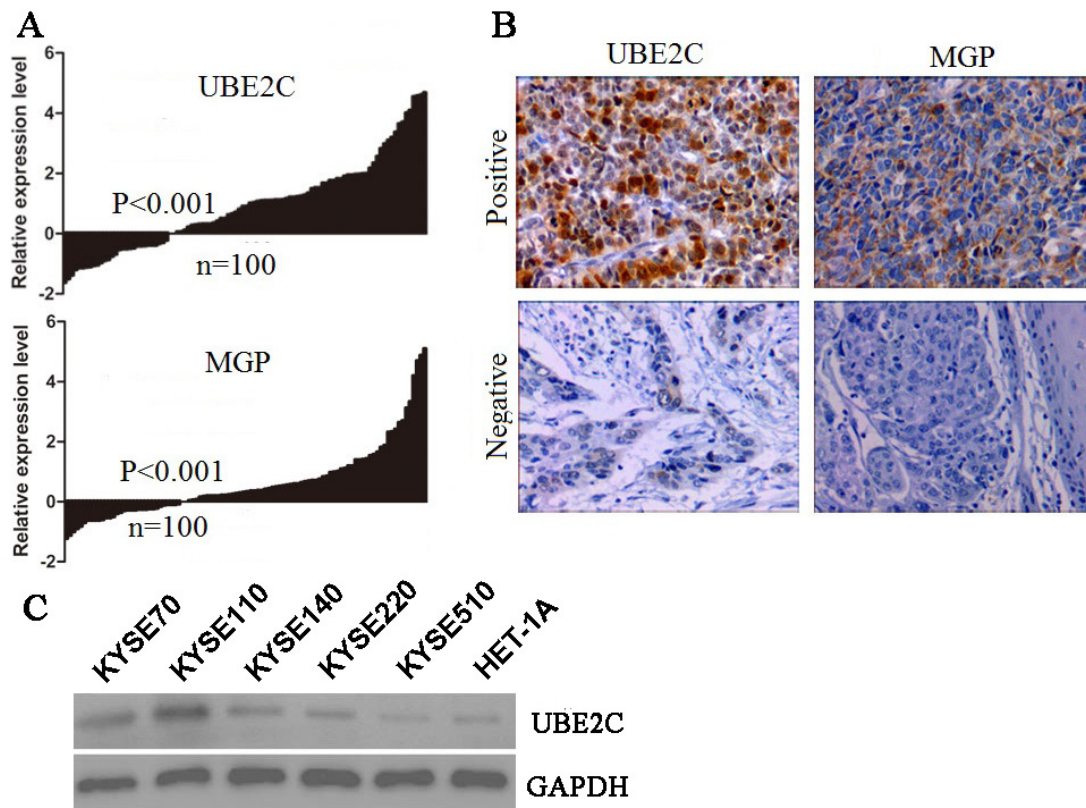


Figure 1: The candidate gene expression in the training cohort and cell lines. A. Quantitative RT-PCR of two selected genes, UBE2C and MGP. B. Representative IHC staining showing protein expression in the invasive tumors ($\times 200$). C. The western blot analysis of UBE2C protein in five ESCC cell lines (KYSE) and human normal esophageal squamous epithelial cell line (Het-1A).

Table 2: Univariate Cox proportional hazards regression for disease-free survival and overall survival in training cohort

Variables	Category	Disease-free survival			Overall survival		
		HR	95%CI	P-value	HR	95%CI	P-value
T stage	pT1+ pT2	1.00			1.00		
	pT3+ pT4	1.82	1.30-2.55	<0.001	1.86	1.33-2.60	<0.001
Lymph nodes status	N0	1.00			1.00		
	N1	1.75	1.32-2.16		1.80	1.29-2.31	
	N2	2.18	1.63-2.72		2.62	2.13-3.10	
	N3	2.43	1.98-2.89	<0.001	2.94	2.27-369	<0.001
TNM stage	I	1.00			1.00		
	II	1.56	1.30-1.82		1.61	1.42-1.86	
	III	2.18	1.74-2.62	<0.001	2.33	1.78-2.68	<0.001
Grade	Well-differentiated	1.00			1.00		
	Moderately-differentiated	1.33	0.85-2.07		1.34	0.86-2.09	
	Poorly-differentiated	2.23	1.39-3.58	0.001	2.30	1.43-3.69	0.001
Vascular invasion	No	1.00			1.00		
	Yes	1.67	0.98-2.81	0.057	1.71	1.02-2.92	0.046
UBE2C	Negative	1.00			1.00		
	Positive	2.50	1.81-3.46	<0.001	2.55	1.84-3.52	<0.001
MGP	Negative	1.00			1.00		
	Positive	1.95	1.42-2.67	<0.001	1.96	1.43-2.68	<0.001

Ubiquitin conjugating enzyme E2 C, UBE2C, is required for degradation of mitotic cyclins, regulation of anaphase-promoting complex, and for cell cycle progression [14, 15]. Overexpression of UBE2C causes chromosome missegregation and alters the cell cycle profile, thus facilitating cell proliferation [16, 17]. Increased expression of UBE2C has been observed in various solid cancers including breast cancer, non-small cell lung cancer, bladder cancer, and ovarian cancer [18–22]. As for ESCC, a recently published study has indicated that elevated level of UBE2C is associated with a more advanced clinical stage and worse prognosis [23]. We have discovered that UBE2C was up-regulated in group A, which was characterized by worse outcome. In esophageal adenocarcinoma (EA), the expression of UBE2C was elevated compared to Barrett's metaplasia, and was suppressed by proteasome inhibition, resulting in suppressed cell proliferation and cell cycle [24].

Matrix Gla protein, MGP, is a calcium-binding and vitamin K-dependent protein, which is found in bone matrix and cartilage [25]. The MGP gene is overexpressed in different types of cancer, including ovarian, lung, urogenital, skin, and glioblastomas [26–30]. In glioblastoma, MGP promotes tumor growth and angiogenesis [31]. Up-regulation

of MGP mRNA correlates with poor prognosis for breast cancer and gastric cancer [32, 33]. However, conflicting findings exist in colorectal cancer, where MGP mRNA does not seem to correlate with histopathologic features, such as tumor progression, size and cell differentiation [34]. Up to now, there was no study on MGP in ESCC. Our study has found that ESCC patients with higher MGP expression in group A have poor clinical outcome.

Previous studies have used expression profiles to predict prognosis in EC, where overall survival is the main outcome indicator [35–37]. In our study, not only overall survival but also disease-free survival were considered as end points. Although tumor recurrence and overall survival are closely related, it is not uncommon for some patients with localized recurrence to benefit from radical treatment resulting in relatively prolonged survival. Furthermore, the overall endpoint is affected by competing risks due to comorbidities. We considered prediction of tumor recurrence as a superior end point to overall survival for our purpose of developing survival predictive model to identify patients who would benefit from adjuvant treatment.

Several biases can be associated with the patient cohort included in this study. The patients included in our

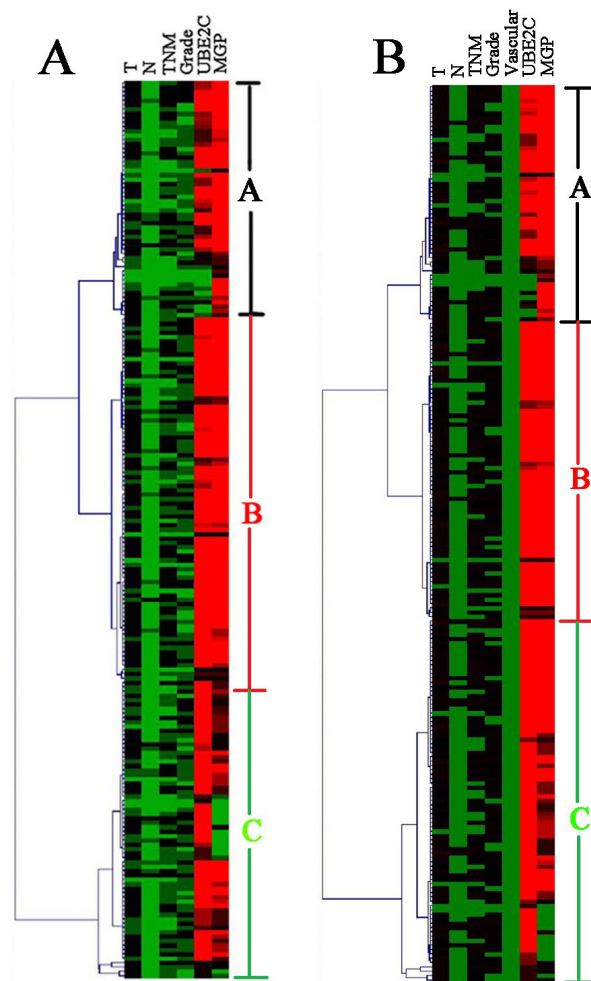


Figure 2: Unsupervised cluster analysis of the training set of 205 ESCC genes identified three clinically relevant subsets (group A, B and C). Each row is a sample, and each column is a prognostic factor. High value is depicted as red, low value as green and median value as black. **A.** The clustering results by the prognostic factors for disease-free survival. **B.** The clustering results by the prognostic factors for overall survival.

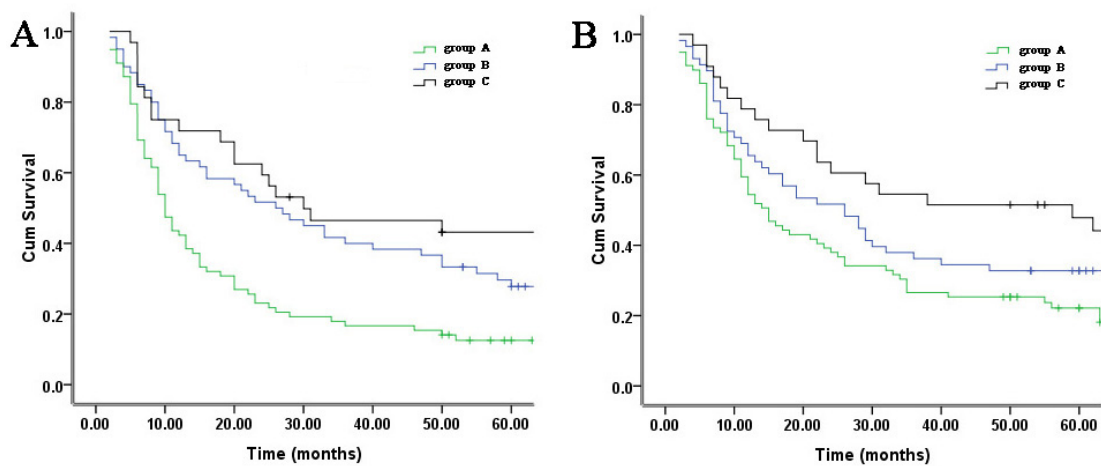


Figure 3: Kaplan-Meier analysis of three clusters of 205 training set identified by hierarchical clustering (groups A, B, C). Differences in survival between subgroups are assessed by log-rank tests. **A.** Disease-free survival ($P < 0.001$). **B.** Overall survival ($P < 0.001$).

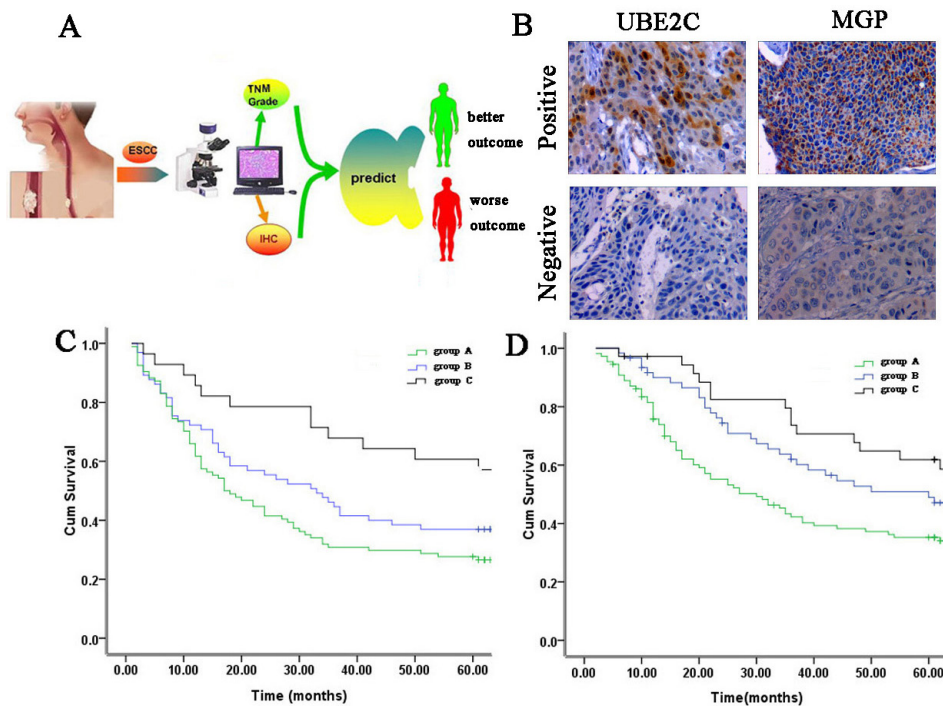


Figure 4: Survival prediction in the testing cohort. **A.** Schematic layout of survival prediction. **B.** Representative IHC staining showing protein expression in the invasive tumors in testing set ($\times 200$). **C.** Postoperative survival curves of disease-free survival in the testing cohort based on the survival prediction model by the Kaplan-Meier analysis and log-rank test ($P=0.006$). **D.** Postoperative survival curves of overall survival in the testing cohort based on the survival prediction model by the Kaplan-Meier analysis and log-rank test ($P=0.002$).

study were treated at two hospitals for several years. The training cohort has a relative overrepresentation of late progressing tumors compared with the test cohort. This may complicate a comparison of the training cohort with the validation cohort. Therefore, a multicenter study under standardized conditions needs to be performed before a possible clinical use. In addition, although IHC can be easily performed in standard clinical pathology laboratories, the IHC scoring is subjective; however, this problem can be minimized using standard-intensity pictures to allow for accurate classification of the staining pattern.

In summary, this study demonstrates that the survival of ESCC patients can be predicted with a reasonable accuracy by using a simple discriminant function based on UBE2C and MGP gene expression and clinical stage. Our methodology may be important to both patients and physicians in terms of life planning and therapeutic strategies.

MATERIALS AND METHODS

Microarray data analysis

Publicly available microarray data in Gene Expression Omnibus (GEO, under the accession number GSE 13898) were used to extract the candidate genes. R/BioConductor [38] was used for processing the microarray text data from BeadStudio. The expression level of each gene was transformed into a log 2 base before further analysis. The log

2 fold change (FC) of each probe on the array within each tissue pair was then calculated. Rank product testing was performed to test whether the differential expression between tumor tissue and matched normal mucosa was significant. The differential expression was considered significant if the adjusted p-value, i.e. the FDR q-value, was less than 0.05.

Study population

A total of 412 patients who had undergone esophagectomy for ESCC at Nantong tumor hospital (between January 2007 and July 2008) and Renji hospital, Shanghai (between January 2006 and September 2008) were retrospectively enrolled. ESCC was confirmed by postoperative histologic pathology in all cases. The recruited cases met the following eligibility criteria: a minimum of 50% tumor cells present in the tissue section, R0 resection, no previous malignancy. Resected tumor and paired non-tumor tissue specimens were immediately cut from the resected esophagus and placed in RNA-Later (TaKaRa, Japan), frozen in liquid nitrogen and kept at -80°C until RNA extraction. One hundred and twenty-two patients received adjuvant chemotherapy and eighty-four patients received adjuvant radiotherapy (50 Gy, 2 Gy/day for five days/week for five weeks) after surgery. The most common chemotherapy regimen consisted of paclitaxel plus cisplatin for a mean of 3 cycles after surgery, depending on clinical response or the occurrence of adverse effect.

This study was approved by the institutional review board and ethics committee at Nantong tumor hospital and Renji hospital. The written informed consents were obtained from all patients.

Patient follow-up

Uniform follow-up protocol was carried out at two centers. The patients were examined every 3 months for the first 2 years after operation, every 6 months for the following 3 years, and yearly thereafter. After being discharged from the hospital, recording of medical history, physical examination, endoscopic biopsy, CT, and PET were performed during the follow-up time. The endpoints included disease free survival and overall survival. The disease free survival was defined as the interval from the date of surgery to the date of local or regional disease recurrence, distant metastasis, or to the last follow-up date. The overall survival was calculated from the time of surgery to the time of death from any cause, or to the time of last follow-up. Some patients received palliative treatment to improve their quality of life during disease recurrence. When local or regional disease recurrence was observed, patients were treated with chemoradiotherapy or chemotherapy. If distant metastasis was found, chemotherapy or best supportive care were given.

Quantitative real-time PCR (qRT-PCR) and western blot

2 mg of mRNA from each sample was reverse transcribed to single stranded cDNA using an Advantage RT for PCR kit (Clontech). Gene expression was assessed by qRT-PCR using an Applied Biosystems 7500 Fast Sequence Detection System (Life Technologies Corp, CA, USA). The PCR reaction mixture consisted of QuantiTect SYBR Green PCR master mix (2X QuantiTect SYBR Green kit, which contains HotStart Taq® DNA polymerase, QuantiTect SYBR Green PCR buffer, dNTP mix, SYGB I, Rox passive reference dye and 5 mM MgCl₂) (Qiagen), 0.5 μmol/l of each primer and cDNA. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene was used as an endogenous control to normalize the expression data. The primers used for qRT-PCR analysis are summarized in Supplementary Table S2. The comparative Ct (threshold cycle) method was used to calculate the relative changes in gene expression.

The five human esophageal squamous carcinoma cell lines (KYSE-70, KYSE-110, KYSE-140, KYSE-220, KYSE-510) and a human normal esophageal squamous epithelial cell line (Het-1A) were purchased from the DSMZ-German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). The cells were cultured in RPMI 1640 supplemented with 10% fetal bovine serum (FBS) at 37°C in a humidified atmosphere with 5% CO₂. Total proteins from cells were extracted using RIPA buffer (Solarbio, Beijing, China) and the protein concentration was measured by BCA Protein Assay Kit (Pierce, Rockford, USA). Western blot was performed

according to the standard protocol, with primary antibody against UBE2C (1:1000, Abcam) GAPDH was used as a control to confirm equal loading of protein. Tannon 2500 imaging system was used to analyze the band intensities.

Immunohistochemical analysis

Paraffin-embedded samples (4 mm thickness) of ESCC tumors were used for histopathological analysis after antigen retrieval. Briefly, endogenous peroxidase was blocked by incubating the sections in a 0.3% solution of hydrogen peroxide (in PBS) for 10 min. Antigen retrieval was performed by heating the sections for 10 min at 100°C in 10 mM citrate buffer (pH 6.0). Sections were incubated overnight at 4°C with rabbit anti-UBE2C (1:300 dilution, ab12290, Abcam, Cambridge, UK) and mouse anti-MGP (1:50 dilution, SC-81546, Santa Cruz, Texas, USA) antibodies.

Immunostaining scoring was performed by using the semiquantitative H-score method to calculate the sum of the percentage and intensity of positively stained invasive tumor cells. The staining intensity was divided into four grades: 0=negative staining; 1=weak staining; 2=moderate staining; 3= strong staining. Staining score was determined by the following formula: H score = percentage score x intensity score ((1*%1+)(2 *%2+)(3*%3+)), which ranged from 0 to 300 [39]. The X-tile software (version 3.6.1) was used to determine H-score cutoffs by dichotomizing patients according to H-score value and clinical outcome [40]. Positive staining was interpreted as H-score>45 for UBE2C and >50 for MGP. The cut points for each protein were defined in the Shanghai patient cohort (training set) and directly applied to Nantong patient cohort (validation set).

Construction of survival prediction model and statistical analysis

The prognostic significance of clinical and pathologic characteristics was determined using Cox regression analysis. The significant prognostic factors for training samples were selected for establishing survival prediction model. All training samples were categorized by unsupervised cluster analysis (Multi Experiment Viewer, MEV4-6-1). Following log transformation and centre to median calculations, average hierarchical clustering was performed using the Spearman rank correlation. We assigned one of three categories to cluster analysis data for each sample. Then, the Categorical data and significant prognostic factors were applied to construct survival prediction model by Bayer's discriminant analysis. The significance of discriminating variables was estimated using one-factor ANOVA F statistics. The set of optimal features was constructed using stepwise selection of variables minimizing overall Wilk's lambda statistic. Finally, we applied this prediction model to test a group of 205 samples. For the sample testing, each predictor was assigned one of three categories as those obtained in the training group. Figure 5 outlines the experimental flow of the study.

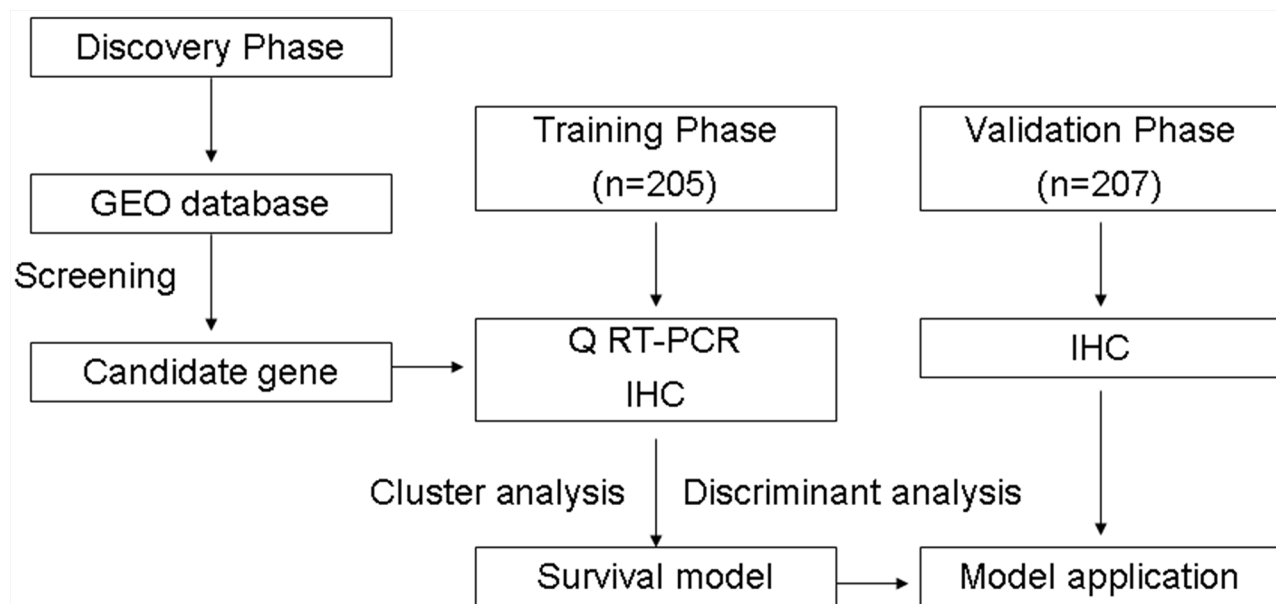


Figure 5: Study design.

Statistical analysis

The Kaplan-Meier method was used to estimate the survival of patients in different groups, and the two-side log-rank test was used to determine the statistical significance. Categorical data were presented as counts and group comparisons were made with the chi-squared test or the Fisher's exact test. All data were processed using SPSS 15.0 software package. The P values less than 0.05 were considered as significant.

ACKNOWLEDGMENTS

This study was supported in part by funding from NSFC (Natural Science Foundation of China, 81101847), Shanghai Committee of Science and Technology, China (124119a4801), Nantong Science and Technology Commission (grant no.: HS149127), and Scientific Fund of International Peace Maternity and Child Health Hospital (2014).

CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65:87-108.
2. Chen W, Zheng R, Zeng H, Zhang S and He J. Annual report on status of cancer in China, 2011. *Chinese journal of cancer research.* 2015; 27:2-12.
3. Castro C, Bosetti C, Malvezzi M, Bertuccio P, Levi F, Negri E, La Vecchia C and Lunet N. Patterns and trends in esophageal cancer mortality and incidence in Europe (1980-2011) and predictions to 2015. *Ann Oncol* 2014; 25:283-290.
4. Trivers KF, Sabatino SA and Stewart SL. Trends in esophageal cancer incidence by histology, United States, 1998-2003. *Int J Cancer.* 2008; 123:1422-1428.
5. Gavin AT, Francisci S, Foschi R, Donnelly DW, Lemmens V, Brenner H and Anderson LA. Oesophageal cancer survival in Europe: a EURO CARE-4 study. *Cancer epidemiology.* 2012; 36:505-512.
6. Chen MF, Yang YH, Lai CH, Chen PC and Chen WC. Outcome of patients with esophageal cancer: a nationwide analysis. *Annals of surgical oncology.* 2013; 20:3023-3030.
7. LH. S, MK. G and C. W. International Union Against Cancer (UICC):TNM Classification of Malignant Tumours. 7th edition. New York: Wiley-Liss. 2010.
8. Chang W, Gao X, Han Y, Du Y, Liu Q, Wang L, Tan X, Zhang Q, Liu Y, Zhu Y, Yu Y, Fan X, Zhang H, Zhou W, Wang J, Fu C, et al. Gene expression profiling-derived immunohistochemistry signature with high prognostic value in colorectal carcinoma. *Gut.* 2014; 63:1457-1467.
9. Liang W, Zhang L, Jiang G, Wang Q, Liu L, Liu D, Wang Z, Zhu Z, Deng Q, Xiong X, Shao W, Shi X and He J. Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *Journal of clinical oncology.* 2015; 33:861-869.
10. Sveen A, Agesen TH, Nesbakken A, Meling GI, Rognum TO, Liestol K, Skotheim RI and Lothe RA. ColoGuidePro: a prognostic 7-gene expression signature for stage III colorectal cancer patients. *Clinical cancer research* 2012; 18:6001-6010.

11. Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, Gupta S, Moore J, Wrobel MJ, Lerner J, Reich M, Chan JA, Glickman JN, Ikeda K, Hashimoto M, Watanabe G, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *The New England journal of medicine*. 2008; 359:1995-2004.
12. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J and Wolmark N. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine*. 2004; 351:2817-2826.
13. Ong CA, Shapiro J, Nason KS, Davison JM, Liu X, Ross-Innes C, O'Donovan M, Dinjens WN, Biermann K, Shannon N, Worster S, Schulz LK, Luketich JD, Wijnhoven BP, Hardwick RH and Fitzgerald RC. Three-gene immunohistochemical panel adds to clinical staging algorithms to predict prognosis for patients with esophageal adenocarcinoma. *Journal of clinical oncology* 2013; 31:1576-1582.
14. Summers MK, Pan B, Mukhyala K and Jackson PK. The unique N terminus of the UbcH10 E2 enzyme controls the threshold for APC activation and enhances checkpoint regulation of the APC. *Molecular cell*. 2008; 31:544-556.
15. Williamson A, Wickliffe KE, Mellone BG, Song L, Karpen GH and Rape M. Identification of a physiological E2 module for the human anaphase-promoting complex. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:18213-18218.
16. van Ree JH, Jeganathan KB, Malureanu L and van Deursen JM. Overexpression of the E2 ubiquitin-conjugating enzyme UbcH10 causes chromosome missegregation and tumor formation. *The Journal of cell biology*. 2010; 188:83-100.
17. Fujita T, Ikeda H, Taira N, Hatoh S, Naito M and Doihara H. Overexpression of UbcH10 alternates the cell cycle profile and accelerate the tumor proliferation in colon cancer. *BMC cancer*. 2009; 9:87.
18. Parris TZ, Kovacs A, Aziz L, Hajizadeh S, Nemes S, Semaan M, Forssell-Aronsson E, Karlsson P and Helou K. Additive effect of the AZGP1, PIP, S100A8 and UBE2C molecular biomarkers improves outcome prediction in breast carcinoma. *International journal of cancer* 2014; 134:1617-1629.
19. Pallante P, Malapelle U, Berlingieri MT, Bellevisine C, Sepe R, Federico A, Rocco D, Galgani M, Chiariotti L, Sanchez-Cespedes M, Fusco A and Troncone G. UbcH10 overexpression in human lung carcinomas and its correlation with EGFR and p53 mutational status. *Eur J Cancer*. 2013; 49:1117-1126.
20. Frstrup N, Birkenkamp-Demtroder K, Reinert T, Sanchez-Carbayo M, Segersten U, Malmstrom PU, Palou J, Alvarez-Mugica M, Pan CC, Ulhoi BP, Borre M, Orntoft TF and Dyrskjot L. Multicenter validation of cyclin D1, MCM7, TRIM29, and UBE2C as prognostic protein markers in non-muscle-invasive bladder cancer. *The American journal of pathology*. 2013; 182:339-349.
21. Berlingieri MT, Pallante P, Guida M, Nappi C, Masciullo V, Scambia G, Ferraro A, Leone V, Sboner A, Barbareschi M, Ferro A, Troncone G and Fusco A. UbcH10 expression may be a useful tool in the prognosis of ovarian carcinomas. *Oncogene*. 2007; 26:2136-2140.
22. Okamoto Y, Ozaki T, Miyazaki K, Aoyama M, Miyazaki M and Nakagawara A. UbcH10 is the cancer-related E2 ubiquitin-conjugating enzyme. *Cancer research*. 2003; 63:4167-4173.
23. Matsumoto A, Ishibashi Y, Urashima M, Omura N, Nakada K, Nishikawa K, Shida A, Takada K, Kashiwagi H and Yanaga K. High UBCH10 protein expression as a marker of poor prognosis in esophageal squamous cell carcinoma. *Anticancer research*. 2014; 34:955-961.
24. Lin J, Raouf DA, Wang Z, Lin MY, Thomas DG, Greenon JK, Giordano TJ, Orringer MB, Chang AC, Beer DG and Lin L. Expression and effect of inhibition of the ubiquitin-conjugating enzyme E2C on esophageal adenocarcinoma. *Neoplasia*. 2006; 8:1062-1071.
25. Schurgers LJ, Cranenburg EC and Vermeer C. Matrix Gla-protein: the calcification inhibitor in need of vitamin K. *Thrombosis and haemostasis*. 2008; 100:593-603.
26. Hough CD, Cho KR, Zonderman AB, Schwartz DR and Morin PJ. Coordinately up-regulated genes in ovarian cancer. *Cancer research*. 2001; 61:3869-3876.
27. Bianchi F, Hu J, Pelosi G, Cirincione R, Ferguson M, Ratcliffe C, Di Fiore PP, Gatter K, Pezzella F and Pastorino U. Lung cancers detected by screening with spiral computed tomography have a malignant phenotype when analyzed by cDNA microarray. *Clinical cancer research* 2004; 10:6023-6028.
28. Levedakou EN, Strohmeyer TG, Effert PJ and Liu ET. Expression of the matrix Gla protein in urogenital malignancies. *International journal of cancer* 1992; 52:534-537.
29. Micke P, Kappert K, Ohshima M, Sundquist C, Scheidl S, Lindahl P, Heldin CH, Botling J, Ponten F and Ostman A. In situ identification of genes regulated specifically in fibroblasts of human basal cell carcinoma. *The Journal of investigative dermatology*. 2007; 127:1516-1523.
30. Mertsch S, Schurgers LJ, Weber K, Paulus W and Senner V. Matrix gla protein (MGP): an overexpressed and migration-promoting mesenchymal component in glioblastoma. *BMC cancer*. 2009; 9:302.
31. Kuzontkoski PM, Mulligan-Kehoe MJ, Harris BT and Israel MA. Inhibitor of DNA binding-4 promotes angiogenesis and growth of glioblastoma multiforme by elevating matrix GLA levels. *Oncogene*. 2010; 29:3793-3802.
32. Yoshimura K, Takeuchi K, Nagasaki K, Ogishima S, Tanaka H, Iwase T, Akiyama F, Kuroda Y and Miki Y. Prognostic value of matrix Gla protein in breast cancer. *Molecular medicine reports*. 2009; 2:549-553.

33. Guo L, Guo XB, Jiang JL, Zhang JN, Ji J, Liu BY, Zhu ZG and Yu YY. Discovery and verification of matrix gla protein, a TNM staging and prognosis-related biomarker for gastric cancer [Article in Chinese]. *Zhonghua bing li xue za zhi*. 2010; 39:436-441.
34. Fan C, Sheu D, Fan H, Hsu K, Allen Chang C and Chan E. Down-regulation of matrix Gla protein messenger RNA in human colorectal adenocarcinomas. *Cancer letters*. 2001; 165:63-69.
35. Peng ZM, Yu W, Xie Y, Peng WH, Cao HH, Shen JH, Wu ZY, Li EM and Xu LY. A four actin-binding protein signature model for poor prognosis of patients with esophageal squamous cell carcinoma. *International journal of clinical and experimental pathology*. 2014; 7:5950-5959.
36. Cao HH, Zheng CP, Wang SH, Wu JY, Shen JH, Xu XE, Fu JH, Wu ZY, Li EM and Xu LY. A molecular prognostic model predicts esophageal squamous cell carcinoma prognosis. *PloS one*. 2014; 9:e106007.
37. Sun LL, Wu JY, Wu ZY, Shen JH, Xu XE, Chen B, Wang SH, Li EM and Xu LY. A three-gene signature and clinical outcome in esophageal squamous cell carcinoma. *International journal of cancer* 2015; 136:E569-577.
38. Kim SM, Park YY, Park ES, Cho JY, Izzo JG, Zhang D, Kim SB, Lee JH, Bhutani MS, Swisher SG, Wu X, Coombes KR, Maru D, Wang KK, Buttar NS, Ajani JA, et al. Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome. *PloS one*. 2010; 5:e15074.
39. McCarty KS, Jr., Miller LS, Cox EB, Konrath J and McCarty KS, Sr. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *Archives of pathology & laboratory medicine*. 1985; 109:716-721.
40. Camp RL, Dolled-Filhart M and Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clinical cancer research* 2004; 10:7252-7259.