A gene expression-based risk model reveals prognosis of gastric cancer

Xiaorong Deng[1,*], Qun Xiao[2,*], Feng Liu[3,*] and Cihua Zheng[4]

[1] Department of Gastrointestinal Surgery, The Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China
[2] Department of Hepatobiliary Surgery, The First Affiliated Hospital of Hunan College of Traditional Chinese Medicine, Zhuzhou, Hunan, China
[3] Department of General Surgery, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, China
[4] Jiangxi Medical College of Nanchang University, Nanchang, Jiangxi, China
[*] These authors contributed equally to this work.

ABSTRACT

Background. The prognosis of gastric cancer is difficult to determine, although clinical indicators provide valuable evidence.

Methods. In this study, using screened biomarkers of gastric cancer in combination with random forest variable hunting and multivariable Cox regression, a risk score model was developed to predict the survival of gastric cancer. Survival difference between high/low-risk groups were compared. The relationship between risk score and other clinicopathological indicators was evaluated. Gene set enrichment analysis (GSEA) was used to identify pathways associated with risk scores.

Results. The patients with high risk scores (median overall survival: 20.2 months, 95% CI [16.9–26.0] months) tend to exhibit early events compared with those with low risk scores (median survival: 70.0 months, 95% CI [46.9–101] months, $p = 1.80e-5$). Further validation was implemented in another three independent datasets (GSE15459, GSE26253, GSE62254). Correlation analyses between clinical observations and risk scores were performed, and the results indicated that the risk score was not significantly associated with gender, age and primary tumor size but was significantly associated with grade and tumor stage. In addition, the risk score was also not influenced by radiation therapy. Cox multivariate regression and three-year survival nomogram suggest that the risk score is an important indicator of gastric cancer prognosis. GSEA was used to identified KEGG pathways significantly associated with risk score, and signaling pathways involved in focal adhesion and the TGF-beta signaling pathway were identified.

Conclusion. The risk score model successfully predicted the survival of 1,294 gastric cancer samples from four independent datasets and is among the most important indicators in clinical clinicopathological information for the prognosis of gastric cancer. To our knowledge, it is the first report to predict the survival of gastric cancer using optimized expression panel.

Subjects Bioinformatics, Gastroenterology and Hepatology, Oncology
Keywords Gatric cancer, Prognosis, Model

INTRODUCTION

Gastric cancer is among the most lethal of cancers worldwide. According to most recent statistical reports in 2015, in China, 679,100 new cases and 498,000 deaths were estimated (*Chen et al., 2016*). Clinical indicators, including TNM staging, were proven to be effective indicators of prognosis (*Wittekind, 2015*). Additionally, the molecular classification also plays a powerful role in prognosis (*Chen, Xu & Zhou, 2016*). However, the classification effect of the staging system is still unfavorable. Thus, molecular biomarkers were needed to predict the survival of gastric cancer patients.

In recent decades, molecular biomarkers for gastric prognosis have been widely reported (*Arigami et al., 2013*; *Guo et al., 2013*; *Liu et al., 2015*; *Rachidi et al., 2013*). PD-L1 and MET1 co-expression predicted a poor survival of gastric cancer, with shorter overall survival rate and disease-free survival rate (*Kwon et al., 2017*). The low expression of BUB1 also suggested a poor prognosis (*Stahl et al., 2017*), and EPHB4 showed a similar pattern for prognosis (*Yin et al., 2017*). In addition, lncRNAs including SNHG and PCAT-1 were also reported to be associated with the proliferation, migration and prognosis of gastric cancer (*Cui, Wu & Qu, 2017*; *Hu et al., 2017*). However, single molecular biomarkers often fail to predict the survival of gastric cancer due to their heterogeneity, while transcriptome-based classification includes redundant information. However, the multiple molecular biomarker-based model has been proven to be robust across datasets and has been implemented in cancer (*Bou Samra et al., 2014*; *Chang et al., 2014*; *Kim et al., 2014*; *Salazar et al., 2011*; *Wu et al., 2012*).

In this work, genes significantly associated with survival were identified. Using these genes, the machining learning (random forest) method and Cox regression, a risk score model was developed. The risk score successfully divided the patients with good and poor prognosis. The robustness of the model was further validated in another three independent datasets. Clinical correlation analysis has shown that the score is not associated with other clinical information but was significantly correlated with primary tumor stage. Additionally, the score was effective for patients who underwent radiotherapy or not. KEGG pathway analysis showed that various cancer-related signaling pathways and focal adhesion pathways were significantly enriched.

MATERIAL AND METHODS

Data manipulation

The raw microarray data files were downloaded as GEO (https://www.ncbi.nlm.nih.gov/geo) according to the provided accession numbers. After pre-processing, including background correction and Robust Multichip Average (RMA) normalization using the R package ''affy'', probes in each dataset and platform were matched to HUGO gene symbols using the manufacturer's provided annotation files. If a single gene matched multiple probes, the average value of the probes was calculated as the relative expression of the corresponding gene. Clinical observations, including survival information, were downloaded from the same website along with the raw data. The TCGA dataset was downloaded from the UCSC Xena website (https://tcga.xenahubs.net/) and further

converted to log 2 transformed RPKM values according to the website's provided protocol. Clinical information was also downloaded via UCSC Xena. To normalize the data among batches and platforms, z-scores were calculated for each patient in each dataset.

Gene selection and model development

Correlation analyses between overall survival and the relative expression value of each gene were evaluated with Cox univariate regression with function "coxph" in the R package "survival", and genes significantly associated with overall survival ($p < 0.01$) in both TCGA and GSE15459 were retained for further analysis. Genes not significantly different in normal and tumor tissues ($p > 0.05$) in the TCGA dataset were excluded. Afterwards, random forest variable hunting was used to optimize the panel content to develop the prediction model. After 100 repeats and 100 iterations, thirteen genes were selected. Based on the relative expression of these genes, a Cox multivariate model was carried out to develop the risk score model, and the risk score formula was calculated as follows:

$$\text{Risk score} = \sum_{i}^{n} x_i * \beta_i$$

where x_i indicates the z-score transformed relative expression level of gene i, and β_i refers to coefficient of gene i.

Statistical analysis and Gene Set Enrichment Analysis (GSEA)

The Cox multivariate and univariate regression was carried out with the R package "survival", and random forest variable hunting was implemented using the R package "randomForestSRC" (*Ishwaran et al., 2014*) with 100 repeats and 100 iterations. The clinical correlation between risk score and clinical observations was calculated with Student's t-test. The nomogram for the one-year survival rate was calculated using the R package "rms". Three-year survival ROC was plotted using R functions in the package "pROC" (*Robin et al., 2011*). GSEA analysis was implemented based on TCGA dataset using Gene Set Enrichment Analysis (*Subramanian et al., 2005*) (GSEA) java software. Differential gene identification in the TCGA dataset was implemented with the R package "limma" using log 2 transformed RPKM values.

RESULTS

Gene selection and model development

To identify the survival-related genes, Cox univariate regression between overall survival and gene expression was implemented in both the TCGA dataset ($N = 380$) and GSE15459 datasets to remove the false discovery. Genes significantly associated with overall survival ($p < 0.01$) in both datasets were considered survival-related genes (termed gene list 1). Differential genes between normal and tumor tissues were identified, and the expression levels of genes that were not significantly different between normal and tumor tissues were excluded from gene list 1 (termed gene list 2). Considering that redundant information exists in these genes and excessive genes may hinder the utilization of the model, a machine learning method called random forest variable hunting was
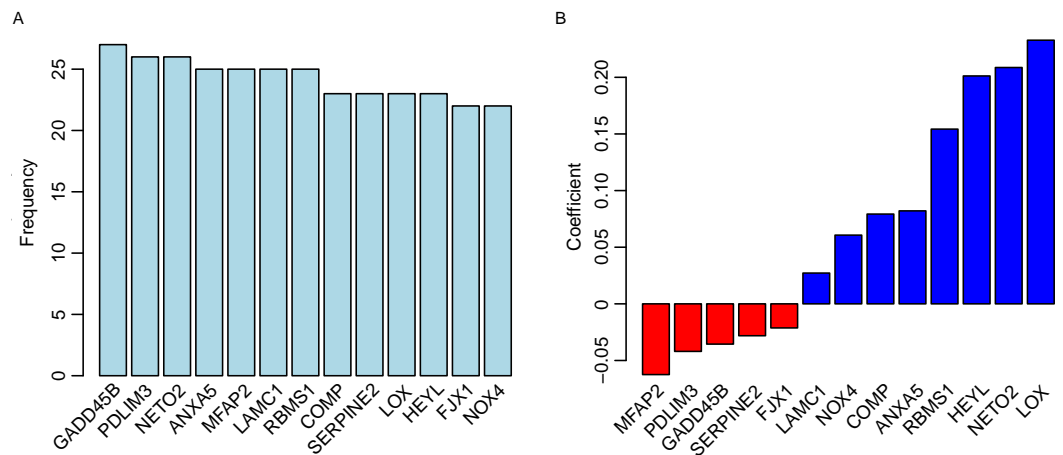
Deng et al. (2018), *PeerJ*, DOI 10.7717/peerj.4204

3/12

Figure 1 Gene selection and model development. The frequency of genes presented in random forest variable hunting (A) and the coefficient for each gene (B).

Table 1 Parameters of variables. Hazard ratio, 95% confidence interval, and *p* values of candidate genes according to Cox univariate and multivariate regression.

	Univariate Cox regression			Multivariate Cox regression		
	HR	95% CI	*p*value	HR	95% CI	*p*value
NOX4	1.3	1.1–1.5	0.00312	0.96	0.71–1.28	0.76338
FJX1	1.2	1.1–1.5	0.00854	0.96	0.78–1.19	0.71577
HEYL	1.4	1.2–1.6	0.00028	1.3	0.99–1.71	0.05764
LOX	1.3	1.1–1.5	0.0015	1.1	0.86–1.42	0.44066
SERPINE2	1.2	1.1–1.5	0.00701	0.91	0.73–1.14	0.42652
COMP	1.2	1.1–1.4	0.00696	1.08	0.84–1.38	0.53894
RBMS1	1.3	1.1–1.6	0.00147	1.16	0.91–1.48	0.21607
LAMC1	1.3	1.1–1.5	0.00175	1.07	0.86–1.34	0.53129
MFAP2	1.3	1.1–1.5	0.00222	0.96	0.72–1.26	0.75653
ANXA5	1.3	1.1–1.5	0.00456	1.18	0.95–1.46	0.13056
NETO2	1.3	1.1–1.5	0.00697	1.35	1.11–1.64	0.00245
PDLIM3	1.2	1.1–1.4	0.005	0.96	0.74–1.25	0.78218
GADD45B	1.3	1.1–1.5	0.00428	1.06	0.85–1.33	0.59772

employed to reduce the complexity and optimize the gene combination. Thirteen genes were selected for further analysis (Fig. 1A, Table 1). The risk score was calculated as follows: Risk score $= (0.060675302 * \text{NOX4}) + (-0.021259171 * \text{FJX1}) + (0.20119841 * \text{HEYL}) + (0.23276666 * \text{LOX}) + (-0.028145979 * \text{SERPINE2}) + (0.079260655 * \text{COMP}) + (0.154255568 * \text{RBMS1}) + (0.027185616 * \text{LAMC1}) + (-0.062461521 * \text{MFAP2}) + (0.082089956 * \text{ANXA5}) + (0.208657253 * \text{NETO2}) + (-0.041982925 * \text{PDLIM3}) + (-0.035559668 * \text{GADD45B})$, where the gene symbol represents the gene expression values.

The coefficients of genes are shown in Fig. 1B. High expression of genes with positive coefficients positively correlated the risk score value, thus, these genes are tumor genes.
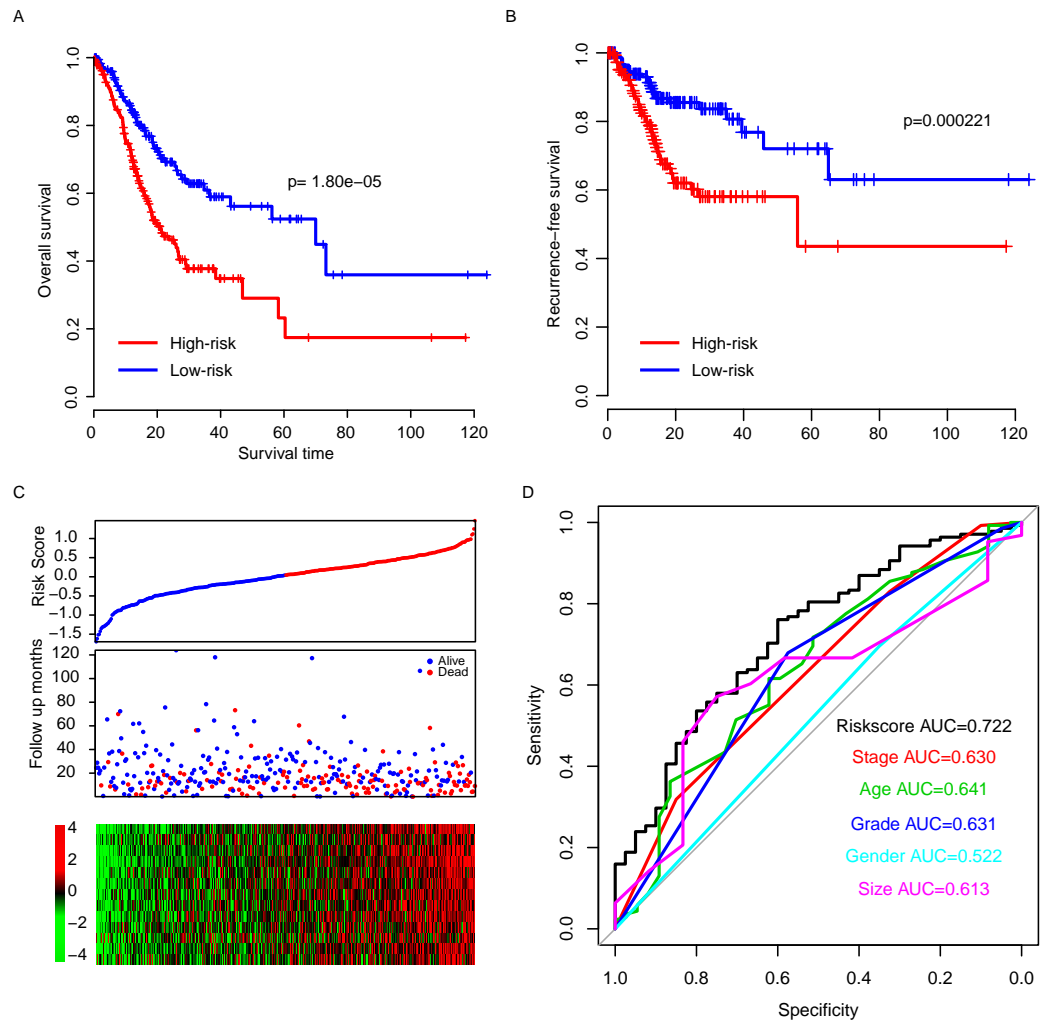
Figure 2 Risk score in the TCGA dataset. The high-risk group had a significantly longer overall survival (OS) time than low risk group (A), and a similar pattern was observed for recurrence-free survival (RFS, B). The detailed survival information of samples, risk score and gene expression (C) and three-year survival ROC were also calculated (D).

Full-size 🖼 DOI: 10.7717/peerj.4204/fig-2

While high expression of genes with negative coefficients negatively correlated the risk score value, thus these genes are tumor suppressor genes.

Risk score predicts the survival of the TCGA dataset

The performance of the risk score was evaluated in the training datasets by dividing the samples in the TCGA dataset into two subgroups, high-risk and low-risk, using the median risk score as a cutoff (0.00436). The survival time of the low risk score is 70.0 (95% CI [46.9–101]) months, which is significantly longer ($p = 1.80e–5$, Fig. 2A) than the high-risk group (20.2 months, 95% CI [16.9–26.7]). The recurrence-free survival (RFS) was also compared between the two groups, and the RFS of the low-risk group is also significantly longer than the high-risk group ($p = 0.000221$, Fig. 2B). As shown in Fig. 2C, along with an
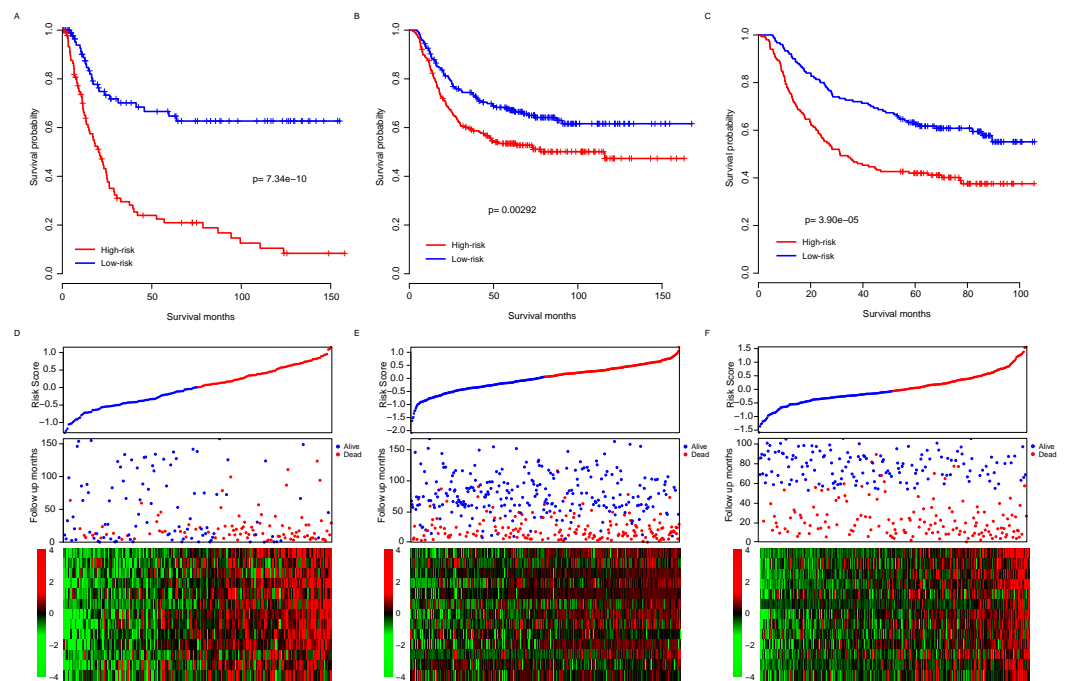
Figure 3 Risk score performance validation. The performance of risk in predicting survival was validated in the GSE15459 (A), GSE26253 (B) and GSE62254 (B) datasets. The detailed survival information and gene expression of the three datasets (D–F) also resembled the profile of the training dataset (TCGA).

Full-size 🖼 DOI: 10.7717/peerj.4204/fig-3

increase of risk score, patients tend to exhibit early events, a high expression of oncogenes and a low expression of tumor repressor genes. The three-year survival area under the receiving operating characteristic (AUROC) curve was calculated, and the AUROCs of risk score, stage, age, grade, gender and primary tumor size were 0.722, 0.630, 0,641, 0.631, 0.522 and 0.613 (Fig. 2D), suggesting that risk score is an important indicator of the survival of gastric patients.

Risk score performance validation

The observed prognostic performance of the risk score in the training dataset (TCGA) may have resulted from over-fitness between the data and model. To test the robustness of the model, after locking the coefficient of each gene, the risk score of each sample in each dataset was evaluated. The validation datasets include another three independent datasets, GSE15459 ($N = 192$), GSE26253 ($N = 422$) and GSE62254 ($N = 300$). By dividing the patients of each dataset into high-risk and low-risk groups according to the median risk score as the cutoff in each dataset, the survival difference of these two subgroups was evaluated. The survival time in the high-risk group was significantly shorter than the low-risk group in all three datasets ($p = 7.34e–10$, 0.00292 and 3.90e–5 for GSE15459, GSE26253 and GSE62254, respectively, Figs. 3A–3C). Similar to the training dataset, along with the increase in the risk score, early death was detected in patients with a high risk score in each sample (Figs. 3D–3F). In addition, the gene expression patterns in these three datasets of these thirteen genes also resembles those in the training dataset (Figs. 3D–3F).
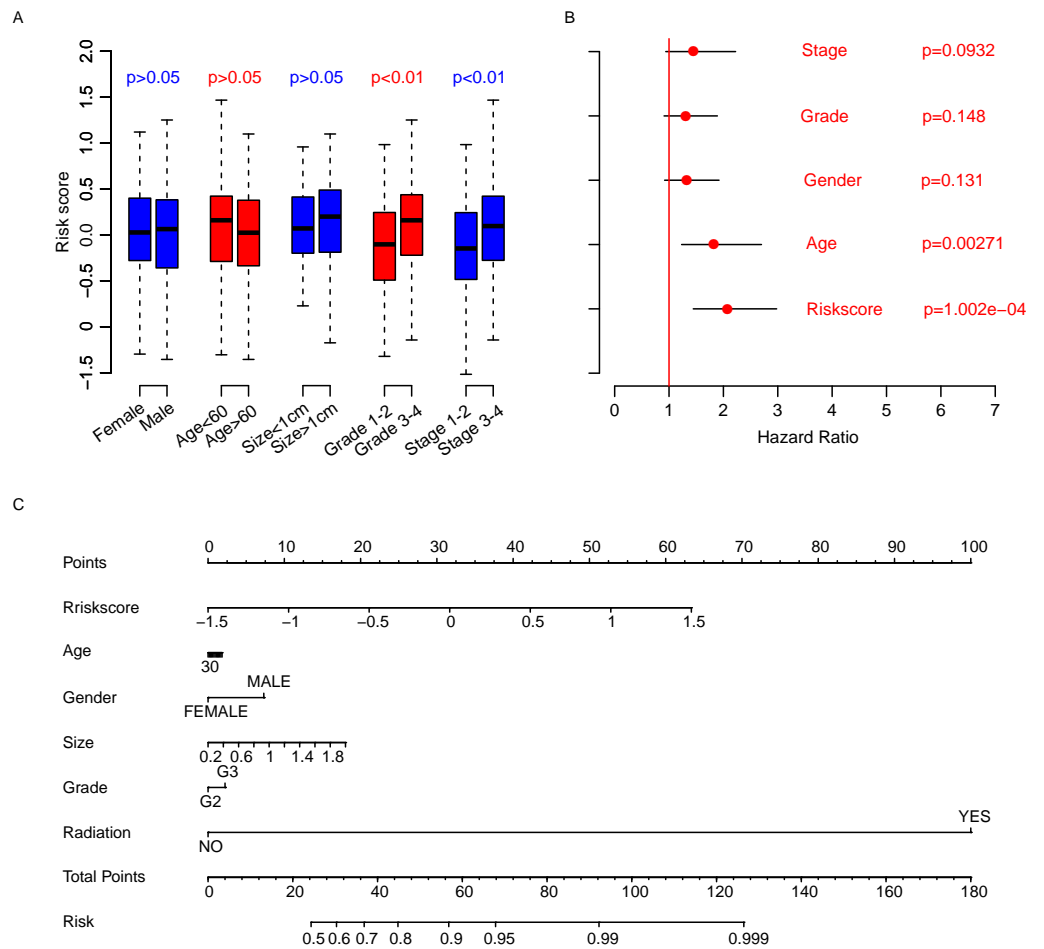
Figure 4 Clinical observations and risk score. The risk score is not associated with other clinical observations (A) besides stage and grade, and it is an important indicator of survival (B) according to Cox multivariate regression. The three-year survival nomogram was plotted to facilitate the utilization of the risk score (C).

Collectively, these results indicate that the risk score model is robust in predicting the survival of gastric patients across datasets and platforms.

Risk score and clinicopathological information

The correlation analyses between clinicopathological information and risk score were also performed. First, we compared the risk score values in the clinical observation categories. It was noted that age (<60, >60), gender, and primary tumor size (>1 cm, <1 cm) were not significantly associated with risk score (Fig. 4A), while the risk score was significantly associated with higher grade and stage ($p < 0.01$). Subsequently, Cox multivariate regression was implemented to evaluate the significance of age, gender, stage, grade and risk score (Fig. 4B). The results showed that the risk score is one of the most important clinical indicators of prognosis. To facilitate the utilization of the risk score, a nomogram for three-year overall survival using the aforementioned clinical information
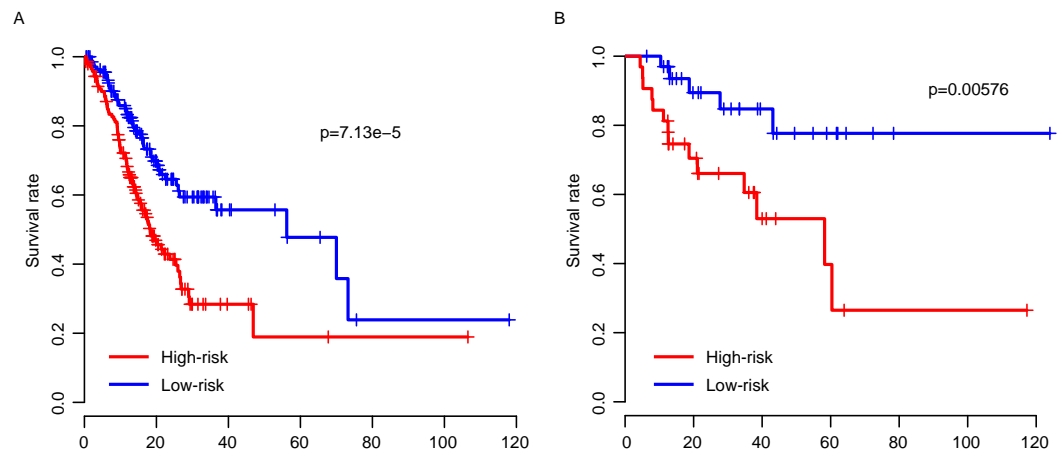
Figure 5 Risk score and radiotherapy. The risk score successfully predicted the survival of patients who received radiotherapy (A) or not (B).

was plotted (Fig. 4C). All these results indicate that the risk score is an important clinical indicator of gastric cancer prognosis.

Risk score and radiation

Radiation is among the most important adjuvant therapy methods in gastric treatment. Thus, the risk score performance in patients who underwent radiation or not was investigated to test whether it was effective in these sub-categories. The patients who did not receive radiation were divided into high-risk and low-risk groups according to the median risk score value in these samples. As expected, patients who did not receive radiation and who had higher risk scores had significantly poorer survival than those with a low risk score (Fig. 5A, $N = 289$). The survival pattern of patients who received radiation also resembled that of those without radiation (Fig. 5B, $N = 69$). These results indicate that the risk score is robust and not effected by radiation therapy.

KEGG pathways associated with risk score

To investigate how the risk score predicted the survival of gastric cancer, we divided the samples in TCGA datasets into high-risk and low-risk groups according to the median risk score values, as previously described. GSEA was carried out to investigate the pathways that were significantly different between the high/low risk groups. Multiple cancer-related KEGG signaling pathways, including the TGF-beta signaling pathway, focal adhesion, gap junction, regulation of actin cytoskeleton and MAPK signaling pathway, were significantly enriched (Fig. 6A; false discovery rate (FDR) < 0.01). Of these pathways, focal adhesion, the regulation of the actin cytoskeleton and the MAPK signaling pathway were noted (Figs. 6B–6D). These results suggest that the risk score reflects multiple cancer statuses of gastric cells and thus predicts the survival.
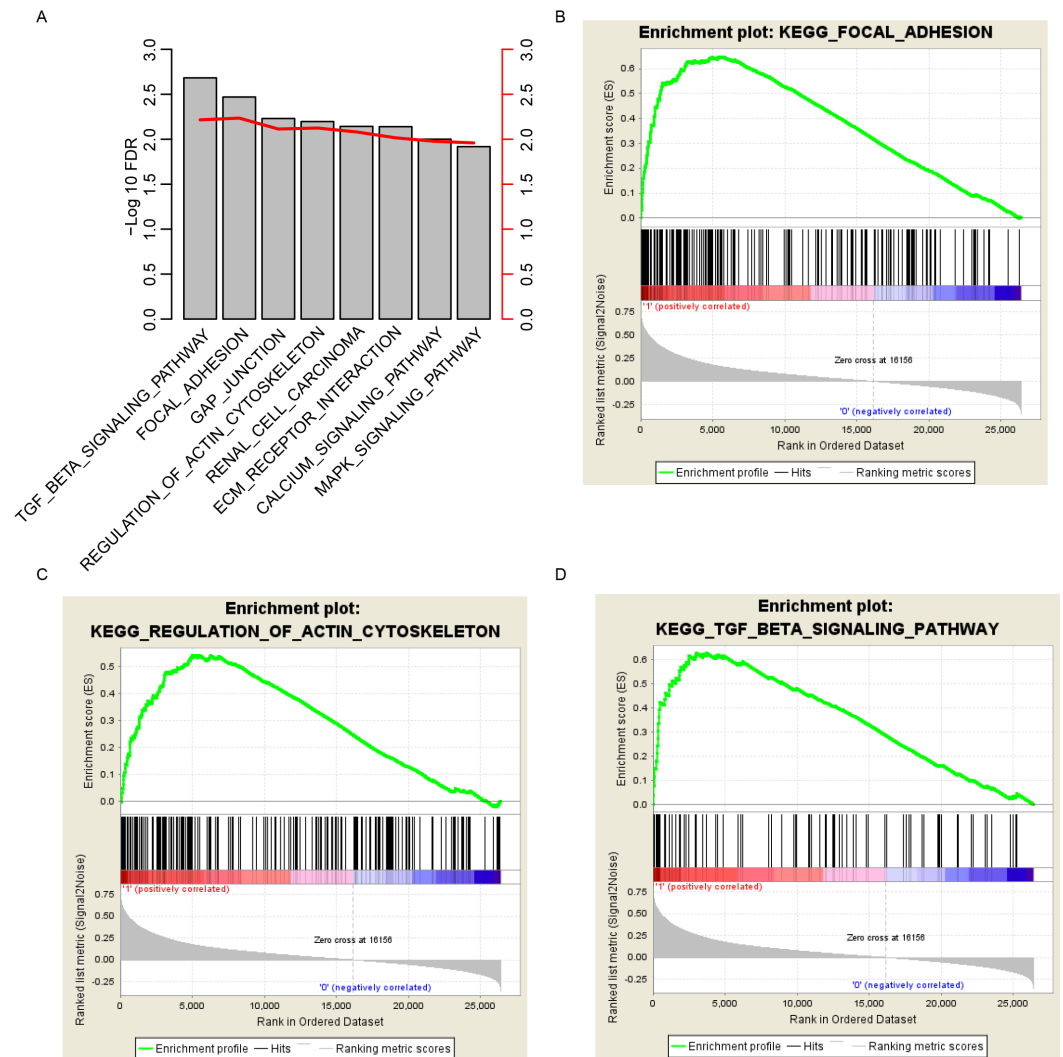
Figure 6 **KEGG pathways associated with risk score.** GSEA according to the expression of the TCGA dataset revealed a significant pathway associated with risk score (A), including focal adhesion (B), the regulation of the actin cytoskeleton (C) and the TGF-beta signaling pathway (D).

Full-size ⊡ DOI: 10.7717/peerj.4204/fig-6

DISCUSSION

The prognosis of gastric cancer varies due to many reasons. First, the progression status evaluated by the clinical and pathological indicators explain the prognosis, to some extent (*Wittekind, 2015*). Second, the treatment method, including the surgery (R0/R1/R2) and treatment method (adjuvant therapy and targeted therapy), also influence the outcome of gastric cancer patients (*Chan et al., 2016*; *Song et al., 2017*). The third reason is the biological heterogeneity of gastric cancer, which has an important impact on carcinogenesis and development. This is the reason why biomarkers are needed for gastric cancer.

Although single biomarkers have been reported in recent years (*Arigami et al., 2013*; *Chan et al., 2016*; *Guo et al., 2013*; *Hu et al., 2017*; *Liu et al., 2015*; *Rachidi et al., 2013*; *Stahl et al., 2017*), the performance of a single biomarker is not robust across datasets, which results from the biological heterogeneity of gastric cancer. One gene was detected to be significantly associated with survival in all four datasets. However, the multiple gene-based model utilized the complement of genetic information and effectively removed the redundancy of the genome. Thus, the multiple gene-based model is effective in determining the prognosis of multiple cancer types (*Kim et al., 2014*; *Massari et al., 2015*; *Zhang et al., 2017*).

One of the most important limitations of this study is that all samples involved in this study were retrospectively obtained, and clinicopathological indicators were not available. For example, time to metastasis, molecular subtypes including HER2 status, and anatomical location were not available for most datasets. Another important limitation of this study is that the relative expression values were z-score transformed; thus, a pooled dataset is needed to facilitate the utilization of this model.

CONCLUSION

The risk score model is robust and useful in predicting the survival of gastric cancer.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests
The authors declare there are no competing interests.

Author Contributions
- Xiaorong Deng conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Qun Xiao conceived and designed the experiments, contributed reagents/materials/-analysis tools, wrote the paper, reviewed drafts of the paper.
- Feng Liu analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Cihua Zheng performed the experiments, wrote the paper, reviewed drafts of the paper.

Data Availability
The following information was supplied regarding data availability:

The raw data is publicly available in GEO with accession numbers GSE15459, GSE26253 and GSE62254.

The other data were downloaded from Xena (xena.ucsc.edu):

https://xenabrowser.net/datapages/?hub=https://tcga.xenahubs.net:443.

REFERENCES

Arigami T, Uenosono Y, Ishigami S, Yanagita S, Hagihara T, Haraguchi N, Matsushita D, Hirahara T, Okumura H, Uchikado Y, Nakajo A, Hokita S, Natsugoe S. 2013. Clinical significance of stanniocalcin 2 expression as a predictor of tumor progression in gastric cancer. *Oncology Reports* **30**:2838–2844 DOI 10.3892/or.2013.2775.

Bou Samra E, Klein B, Commes T, Moreaux J. 2014. Identification of a 20-gene expression-based risk score as a predictor of clinical outcome in chronic lymphocytic leukemia patients. *BioMed Research International* **2014**:Article 423174 DOI 10.1155/2014/423174.

Chan BA, Jang RW, Wong RK, Swallow CJ, Darling GE, Elimova E. 2016. Improving outcomes in resectable gastric cancer: a review of current and future strategies. *Oncology* **30**:635–645.

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, Cao G. 2014. Gene expression profiling-derived immunohistochemistry signature with high prognostic value in colorectal carcinoma. *Gut* **63**:1457–1467 DOI 10.1136/gutjnl-2013-305475.

Chen T, Xu XY, Zhou PH. 2016. Emerging molecular classifications and therapeutic implications for gastric cancer. *Chinese Journal of Cancer* **35**:49 DOI 10.1186/s40880-016-0111-5.

Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. 2016. Cancer statistics in China, 2015. *CA: A Cancer Journal for Clinicians* **66**:115–132 DOI 10.3322/caac.21338.

Cui WC, Wu YF, Qu HM. 2017. Up-regulation of long non-coding RNA PCAT-1 correlates with tumor progression and poor prognosis in gastric cancer. *European Review for Medical and Pharmacological Sciences* **21**:3021–3027.

Guo W, Dong Z, Guo Y, Chen Z, Kuang G, Yang Z. 2013. Methylation-mediated repression of GADD45A and GADD45G expression in gastric cardia adenocarcinoma. *International Journal of Cancer* **133**:2043–2053 DOI 10.1002/ijc.28223.

Hu Y, Ma Z, He Y, Liu W, Su Y, Tang Z. 2017. LncRNA-SNHG1 contributes to gastric cancer cell proliferation by regulating DNMT1. *Biochemical and Biophysical Research Communications* **491(4)**:926–931 DOI 10.1016/j.bbrc.2017.07.137.

Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. 2014. Random survival forests for competing risks. *Biostatistics* **15**:757–773 DOI 10.1093/biostatistics/kxu010.

Kim SK, Kim SY, Kim JH, Roh SA, Cho DH, Kim YS, Kim JC. 2014. A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Molecular Oncology* **8**:1653–1666 DOI 10.1016/j.molonc.2014.06.016.

Kwon MJ, Kim KC, Nam ES, Jin Cho S, Park HR, Min SK, Seo J, Choe JY, Lee HK, Kang HS, Min KW. 2017. Programmed death ligand-1 and MET co-expression is

a poor prognostic factor in gastric cancers after resection. *Oncotarget* **8**:82399–82414 DOI 10.18632/oncotarget.19390.

Liu YF, Yang A, Liu W, Wang C, Wang M, Zhang L, Wang D, Dong JF, Li M. 2015. NME2 reduces proliferation, migration and invasion of gastric cancer cells to limit metastasis. *PLOS ONE* **10**:e0115968 DOI 10.1371/journal.pone.0115968.

Massari F, Bria E, Ciccarese C, Munari E, Modena A, Zambonin V, Sperduti I, Artibani W, Cheng L, Martignoni G, Tortora G, Brunelli M. 2015. Prognostic value of beta-tubulin-3 and c-Myc in muscle invasive urothelial carcinoma of the bladder. *PLOS ONE* **10**:e0127908 DOI 10.1371/journal.pone.0127908.

Rachidi SM, Qin T, Sun S, Zheng WJ, Li Z. 2013. Molecular profiling of multiple human cancers defines an inflammatory cancer-associated molecular pattern and uncovers KPNA2 as a uniform poor prognostic cancer marker. *PLOS ONE* **8**:e57911 DOI 10.1371/journal.pone.0057911.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**:77 DOI 10.1186/1471-2105-12-77.

Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, Lopez-Doriga A, Santos C, Marijnen C, Westerga J, Bruin S, Kerr D, Kuppen P, Van de Velde C, Morreau H, Van Velthuysen L, Glas AM, Van't Veer LJ, Tollenaar R. 2011. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *Journal of Clinical Oncology* **29**:17–24 DOI 10.1200/jco.2010.30.1077.

Song Z, Wu Y, Yang J, Yang D, Fang X. 2017. Progress in the treatment of advanced gastric cancer. *Tumour Biology* **39**(7):1–7 DOI 10.1177/1010428317714626.

Stahl D, Braun M, Gentles AJ, Lingohr P, Walter A, Kristiansen G, Gutgemann I. 2017. Low BUB1 expression is an adverse prognostic marker in gastric adenocarcinoma. *Oncotarget* **8**:76329–76339 DOI 10.18632/oncotarget.19357.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**:15545–15550 DOI 10.1073/pnas.0506580102.

Wittekind C. 2015. The development of the TNM classification of gastric cancer. *Pathology International* **65**:399–403 DOI 10.1111/pin.12306.

Wu X, Weng L, Li X, Guo C, Pal SK, Jin JM, Li Y, Nelson RA, Mu B, Onami SH, Wu JJ, Ruel NH, Wilczynski SP, Gao H, Covarrubias M, Figlin RA, Weiss LM, Wu H. 2012. Identification of a 4-microRNA signature for clear cell renal cell carcinoma metastasis and prognosis. *PLOS ONE* **7**:e35661 DOI 10.1371/journal.pone.0035661.

Yin J, Cui Y, Li L, Ji J, Jiang WG. 2017. Overexpression of EPHB4 is associated with poor survival of patients with gastric cancer. *Anticancer Research* **37**:4489–4497 DOI 10.21873/anticanres.11845.

Zhang ZL, Zhao LJ, Chai L, Zhou SH, Wang F, Wei Y, Xu YP, Zhao P. 2017. Seven LncRNA-mRNA based risk score predicts the survival of head and neck squamous cell carcinoma. *Scientific Reports* **7**:309 DOI 10.1038/s41598-017-00252-2.