Combining Gene Expression Signature With Clinical Features for Survival Stratification of Gastric Cancer

Qiang Sun
Zhejiang University School of Medicine
Dongyang Guo
Zhejiang University School of Medicine
Shuang Li
Zhejiang University School of Medicine
Yanjun Xu
Zhejiang Cancer Hospital
Mingchun Jiang
Zhejiang University School of Medicine
Yang Li
Zhejiang University School of Medicine
Huilong Duan
Zhejiang University College of Biomedical Engineering and Instrument Science
Wei Zhuo
Zhejiang University School of Medicine
Wei Liu
Zhejiang University School of Medicine
Shankuan Zhu
Zhejiang University School of Medicine
Xiangrui Liu
Zhejiang University School of Medicine
Liangjing Wang
Zhejiang University School of Medicine Second Affiliated Hospital
Tianhua Zhou (tzhou@zju.edu.cn)
Zhejiang University https://orcid.org/0000-0002-1791-2124

Research

Keywords: Gastric Cancer, Prognosis, CGRS, Nomogram, Web application

Posted Date: September 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-74747/v1

Version of Record: A version of this preprint was published at Genomics on July 1st, 2021. See the published version at https://doi.org/10.1016/j.ygeno.2021.06.018.

Abstract

Background: The AJCC staging system is considered as the golden standard in clinical practice. However, it remains some pitfalls in assessing the prognosis of gastric cancer (GC) patients with similar clinicopathological characteristics. We aim to develop a new clinic and genetic risk score (CGRS) to improve the prognosis prediction of GC patients.

Methods: The gene expression profiles of the training set from the Asian Cancer Research Group (ACRG) cohort were used for developing genetic risk score (GRS) by LASSO-Cox regression algorithms. CGRS was established by integrating GRS with clinical risk score (CRS) derived from Surveillance, Epidemiology, and End Results (SEER) database. GRS and CGRS were validated in ACRG validation set and other four independent GC cohorts with different data types, such as microarray, RNA sequencing, and qRT-PCR. Multivariable Cox regression was adopted to evaluate the independence of GRS and CGRS in prognosis evaluation.

Results: We established GRS based on a nine-gene signature including *APOD, CCDC92*, *CYS1*, *GSDME*, *ST8SIA5*, *STARD3NL*, *TIMEM245*, *TSPYL5*, and *VAT1*. GRS and CGRS dichotomized GC patients into high and low risk groups with significantly different prognosis in four independent cohorts, including our Zhejiang cohort (all HR > 1, all $P < 0.001$). Both GRS and CGRS were prognostic signatures independent of the AJCC staging system. Receiver operating characteristic (ROC) analysis showed that area under ROC curve of CGRS was larger than that of the AJCC staging system in most cohorts we studied. Nomogram and web tool (http://39.100.117.92/CGRS/) based on CGRS were developed for clinicians to conveniently assess GC prognosis in clinical practice.

Conclusions: CGRS integrating genetic signature with clinical features shows strong robustness in predicting GC prognosis, and can be easily applied in clinical practice through the web application.

Background

Gastric cancer (GC) is one of the most commonly diagnosed cancers and the third leading cause of cancer-related death around the world [1, 2]. Despite the improvement of diagnosis, surgical and other treatment approaches in the past few decades, the prognosis of patients with advanced GC that accounts for approximately 65% of GC cases remains very poor [3]. The AJCC staging system based on clinical and pathological characteristics has been considered as the golden standard for predicting GC prognosis; however, it remains a big challenge to stratify GC patients with similar clinical and pathological characteristics.

Emerging studies show that gene expression profiles of tumor tissues based on microarray or RNA sequencing have provided prognostic information [4, 5]. Successful applications of gene expression profiles have yielded many tools with potential prognostic value for clinicians in a variety of cancers, such as lung cancer [6, 7], breast cancer [8, 9], and large B-cell lymphoma [10, 11]. Large scale studies such as The Cancer Genome Atlas (TCGA) and Asian Cancer Research Group (ACRG) have produced a variety of publicly available expression profiles of GC tissues [12, 13], while researchers have developed various approaches for survival stratification for GC patients [14-16]. However, model overfitting, lack of adequate validation, and failure to be applied across different data platforms hinder their clinical application. Even though clinical [17] and genetic [14-16, 18] models for risk stratification in GC patients have been established, the tool that integrates clinic with genetic information of GC patients has yet to be developed.

In this study, we established a new prognostic signature, clinic and genetic risk score (CGRS), by integrating gene expression profiles with clinical characteristics. CGRS has been confirmed in four different cohorts for accurately predicting GC prognosis, and significantly stratified stage III GC patients into high and low risk groups with different survival time. Furthermore, an easy-to-use nomogram and web application based on CGRS were developed to facilitate its application in clinical practice.

Methods

Patients included in this study

Four cohorts of publicly available GC gene expression profiles were included: ACRG cohort (GSE62254, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62254) [13], TCGA cohort (http://firebrowse.org/?cohort=STAD) [12], Singapore cohort (GSE15459, https://www.ncbi.nlm.nih. gov/geo/query/acc.cgi?acc=GSE15459) [18] and Korea cohort (GSE84437, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84437). SEER database was used to generate the clinical risk score (https://seer.cancer.gov/) [19]. An additional validation set of archived fresh frozen tumor specimens from GC patients who underwent surgery from 2008 to 2013 were obtained from Zhejiang Cancer Hospital. All aspects of this study were approved by the ethics committee of Zhejiang University School of Medicine. All participants gave written, informed consent. Research was conducted in accordance with the Declaration of Helsinki guidelines for the ethical conduct of research in 1975. For all patients, detailed clinical and pathological information can be found in Table 1.

Gastric cancer gene expression datasets and preprocessing

We collected four cohorts (ACRG, TCGA, Singapore, and Korea cohorts) comprising gene expression profiles of GC patients for which survival data were available online. ACRG cohort was randomly split into training and validation sets. Other cohorts based on different platforms were used as additional validation sets. Gene mutational statuses were obtained from TCGA and ACRG cohorts. For Affymetrix microarray data, CEL files were downloaded and normalized with MAS5 algorithm using Custom chip Definition Files (Brainarray v.22, http://brainarray.mbni.med. umich.edu/), followed by log2 transformation and quantile normalization [20]. For Illumina microarray data, the IDAT files were downloaded and normalized by Illumina Genomestudio software (https://www.illumina.com), followed by log2 transformation and quantile normalization. For RNASeq data, RSEM data were downloaded and log2 transformed shift by 0.001. Additional details are included in the Supplementary Materials and Methods.

RNA extraction, amplification, and real-time quantitative RT-PCR

Total RNA was extracted from fresh frozen tissues and 1 μg of total RNA was reverse-transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). qRT-PCR was performed in a Roche Real-Time PCR System using ChamQ Universal SYBR qPCR Master Mix Kit (Vazyme Biotech) and their specific primers (Supplementary Table 1). $\Delta\Delta$Ct method was used to assess the relative value of gene expression. Additional details are included in the Supplementary Materials and Methods.

Functional annotation and pathway enrichment analysis

The correlations between genes and GRS were assessed by the Pearson correlation test. The genes with absolute R greater than 0.4 and *P* values under 0.05 were considered as statistically significant. The genes co-expressing

with GRS in the training set were clustered using AutoSOME [21]. The genes in the top two clusters were assessed for enrichment analysis with curated gene sets from the Molecular Signatures Database (MsigDB, http://software.broadinstitute.org/gsea/msigdb) by the R clusterProfiler package (version 3.6.0) [22, 23]. Gene Set Enrichment Analysis (GSEA) was performed by the JAVA program using gene sets collection from the MsigDB [24]. The Enrichment Map was used to visualize networks discriminating low CGRS Group from high CGRS Group [25]. Additional details are included in the Supplementary Materials and Methods.

Statistical analysis

All statistical tests performed were two-sided, except for one-sided hypergeometric tests. *P* values under 0.05 were considered as statistically significant. All genes were fitted in the univariate Cox proportional hazards regression, and those with P values that were assessed by 10000 permutations under 0.01 (likelihood ratio test) were considered as prognostic genes. Those prognostic genes were then fitted into a multivariate Cox model adjusted with patients' clinical characteristics. The remaining genes ($P < 0.01$) were considered as independent prognostic factors for GC patients. To obtain the minimal set of genes, the LASSO penalty algorithm was carried out for selecting features that passed 10-fold internal cross-validation. The remaining nine prognostic genes were integrated into GRS. CRS based on age and the AJCC stages was developed from SEER database. CGRS was defined as the integration of GRS and CRS weighted by their coefficient in the multivariate Cox model. Receiver operating characteristic and prediction error curves were produced using the survcomp (version 1.28.5) and (version 1.4.18) packages, respectively [26, 27]. The nomogram and calibration plots were generated by rms (version 5.1.2) package [28]. The decision curve was generated by rmda (version 1.6) package [29]. All the analysis was conducted by R software (version 3.4.4). Additional details are provided in the Supplementary Materials and Methods.

Results

Identification of prognostic genes from the training set

To develop a new model for precisely predicting GC prognosis, we selected the ACRG dataset that has detailed clinical information and gene expression profiles (Fig. 1a). We evaluated the impact of sample size on prognostic power for two genes, *TEAD1* and *GZMB* [30, 31], which have been reported as prognostic factors for GC patients previously. The data showed that about 150 patients were required for reliable assessment of prognostic power (Supplementary Fig. 1a and b). Therefore, we randomly split 300 GC patients from ACRG cohort into the training (n = 150) and validation sets (n = 150) (Table 1). The univariate Cox proportional regression analysis was used to identify prognostic genes in the training set. As a result, 2069 genes were considered as survival associated genes ($P < 0.01$, 10000 permutations; Supplementary Table 2). To eliminate the noise caused by other factors, these genes were further fitted into a multivariate Cox proportional regression model, adjusted by patients' clinical characteristics including the AJCC stages, age, gender, and Lauren's subtypes. Finally, 558 genes whose expression was significantly associated with survival independently were identified in the training set ($P < 0.01$; Supplementary Table 2).

GRS for GC patients' prognosis prediction

To obtain the minimal set of genes to build GRS for GC prognosis prediction, we applied the LASSO penalty algorithm to assess the prognostic value of previously identified 558 genes (Supplementary Fig. 1c). After LASSO

selection, nine genes were retained (Fig. 1b, Supplementary Table 2). One gene whose expression was significantly associated with favorable prognosis was *ST8SIA5* (ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 5). The remaining eight genes whose expression were significantly associated with adverse outcomes were *STARD3NL* (STARD3 N-terminal like), *GSDME* (gasdermin E), *TMEM245* (transmembrane protein 245), *VAT1* (vesicle amine transport 1), *CCDC92* (coiled-coil domain containing 92), *TSPYL5* (TSPY like 5), *APOD* (apolipoprotein D), and *CYS1* (cystin 1). Coefficients for these nine genes were determined by the multivariate Cox regression model, and GRS was then calculated in terms of the normalized expression levels of these nine genes (Fig. 1c).

The patients from the training set were assigned to high (n = 75) and low GRS groups (n = 75) using the median GRS value as the cutoff. Kaplan-Meier analysis showed there existed a significant difference in 5-year overall survival between high and low GRS groups (HR = 2.70, 95% CI = 2.07 to 3.52, P = 2.35e-12) (Fig. 1d). Further univariate Cox analysis revealed that GRS remained prognostic in each subgroup (Fig. 1e). Moreover, after calculating the correlations between GRS and global gene expression profiles in the training set, we found that 1205 genes were found to be significantly correlated with GRS (absolute R > 0.4, P < 0.01) (Supplementary Table S4). These genes were clustered into two largest clusters in the training set, which were further compared with gene sets from Molecular Signatures Database to assess the enrichment of biological pathways and processes. The results indicated that cluster 1 shared genes associated with extracellular matrix and genes expressed in stem cells, and cluster 2 was overlapped with cell cycle related genes and genes highly expressed in the early stage of cancer (Fig. 1f; Supplementary Table 3).

To further validate the prognostic value of GRS, GC patients from ACRG validation set were stratified into high and low GRS groups by the cutoff (the median GRS value of the training set, the same below). GRS was significantly associated with overall survival (HR = 1.49, 95% CI = 1.21 to 1.83, P = 1.66e-4), which was further confirmed in whole ACRG cohort (HR = 1.89, 95% CI = 1.61 to 2.22, P = 3.70e-14) (Fig. 2a and b). To further evaluate the performance of GRS, 192 patients from the Singapore cohort, 433 patients from the Korea cohort, and 388 patients from TCGA cohort were stratified into high and low GRS groups according to the cutoff, respectively. GRS remained significantly associated with GC prognosis in all the cohorts (HR = 1.31, 95% CI = 1.09 to 1.57, P = 4.77e-3 in Korea cohort; HR = 1.46, 95% CI = 1.20 to 1.77, P = 1.40e-4 in Singapore cohort; HR = 1.29, 95% CI = 1.10 to 1.52, P = 2.21e-3 in TCGA cohort) (Fig. 2c-e). The multivariate Cox analysis showed that GRS was a prognostic signature independent of the AJCC staging system, age, gender, and Lauren's subtypes (Supplementary Table 4). Moreover, GRS was also prognostic within subgroups of patients harboring wild-type or mutant forms of *TP53*, *MUC16*, *ARID1A*, or *PIK3CA* in ACRG and TCGA cohorts whose gene mutation statuses were available (Supplementary Table 5).

To conveniently apply GRS in clinical practice, we employed qRT-PCR assays on fresh frozen tumor specimens for the nine GRS genes and one addition housekeeping gene *RNU6-1* that lacks prognostic association and displays stable expression [32]. Specimens were obtained from 109 patients with GC who underwent gastrectomy from 2008 to 2013 at Zhejiang Cancer Hospital, termed as the Zhejiang cohort (Table 1). GRS was shown to remain significantly prognostic in Zhejiang cohort (Table 2; Supplementary Table 4). The patients with low GRS had longer survival time than that of high GRS patients (HR = 1.40, 95% CI = 1.12 to 1.75, P = 2.93e-3) (Fig. 2f). Further multivariate Cox analysis showed that GRS was associated with GC prognosis independent of age, gender, and the AJCC stage in Zhejiang cohort (Supplementary Table 4). Taken together, these data suggest that GRS may be

applied for GC prognosis prediction in clinical practice across different platforms, such as microarray, RNA sequencing, and qRT-PCR.

CGRS for prognosis prediction of GC patients

Given that the AJCC stages and age are significantly associated with GC prognosis, and GRS is a prognostic factor independent of the AJCC staging system and age. We integrated GRS with clinical variables to create CGRS for predicting GC survival. First, SEER database that contains 33250 GC patients was used to determine coefficients for different AJCC stages and age by the multivariate Cox regression model. The data showed that clinical risk score (CRS) for each patient could be calculated by the following formula, CRS = 0.021*Age (years) + AJCC stage, where the values for different stages are 0 (stage I), 0.31 (stage II), 0.75 (stage III), and 1.56 (stage IV), respectively (Fig. 2g). The univariate and multivariate Cox analyses, as well as the Kaplan-Meier curve, showed that CRS was significantly associated with GC prognosis in all cohorts we studied (Table 2; Supplementary Fig. 2).

Since there was no significant difference in patients' distribution between ACRG training set and SEER set (Supplementary Table 6), we integrated CRS with GRS into CGRS through the formula determined by multivariate Cox regression model (CGRS = 1.25*CRS + 0.88*GRS) in ACRG training set (Fig. 2h). CGRS was validated to be significantly associated with GC prognosis when stratified GC patients into high and low CGRS groups according to the median value from the ACRG training set (HR = 2.70, 95% CI = 1.57 to 2.16, P = 2.53e-19) (Fig. 2i). Moreover, CGRS showed strong robustness in predicting overall survival of GC patients in internal (HR = 1.80, 95% CI = 1.51 to 2.16, P = 3.21e-10 in ACRG validation set; HR = 2.12, 95% CI = 1.85 to 2.42, P = 1.27e-27 in whole ACRG cohort) and external validation sets (HR = 2.10, 95% CI = 1.71 to 2.57, P = 2.33e-13 in Singapore cohort; HR = 1.72, 95% CI = 1.44 to 2.05, P = 1.61e-9 in TCGA cohort; HR = 2.72, 95% CI = 1.71 to 4.33, P = 2.00e-5 in Zhejiang cohort) (Fig. 2j-n). Further univariate and multivariate Cox analyses confirmed the survival prediction power of CGRS (Supplementary Table 7). Additionally, CGRS remained prognostic within subgroups of patients harboring wild-type or mutant forms of *TP53*, *MUC16*, *ARID1A*, or *PIK3CA* in ACRG and TCGA cohorts whose gene mutation statuses were available (Supplementary Table 8). Together, these results reveal that CGRS can be used to assess GC prognosis independent of other clinical characteristics including AJCC stages, age, gender, and Lauren's subtypes across different platforms.

The prognosis prediction of GRS and CGRS in different AJCC stages

The AJCC staging system is generally considered as the golden standard for evaluating GC prognosis in current clinical practice [33]; however, it remains some deficiencies in predicting patients with similar clinical and pathological characteristics [34, 35]. In this study, we applied our GRS and CGRS in GC patients within the same stage. Due to the small population of stage I GC patients, the performance of GRS and CGRS was fluctuated in different cohorts (Supplementary Fig. 3). For stage II GC patients, GRS and CGRS were significantly associated with GC prognosis in several independent cohorts when stratified the stage II GC patients into high and low risk groups, however they failed in Singapore cohort and ACRG validation sets due to relatively fewer patients (Supplementary Fig. 4). Patients with GC are often diagnosed at advanced stage, and stage III accounts for about 35% of GC cases [36, 37]. Both GRS and CGRS were able to classify stage III patients into high and low risk groups with statistically significantly different survival time in all of the training and validation sets (all HRs > 1, all P < 0.05; Fig. 3). Further multivariate Cox analysis confirmed the robustness of GRS and CGRS in stage III GC patients (Table 2; Supplementary Table 9). Finally, we examined the prediction power of GRS and CGRS in stage

IV GC patients, the performance of GRS and CGRS were unstable in different cohorts because of relatively small population size (Supplementary Fig. 5). Together, these data indicate that both GRS and CGRS are able to predict the prognosis of stage III GC patients, and can be important complements for the AJCC staging system.

The association between GRS, CGRS and molecular subtypes

Emerging studies have established several molecular subtype systems of GC in the past few years [12, 13, 18]. Here, we systematically analyzed the association between our risk scores and molecular subtypes of GC. In TCGA study, GC can be divided into four molecular subtypes. Though there is no significant relevance between clinical outcome and TCGA subtypes, the microsatellite instability (MSI) group that has relatively favorable outcome exhibited lower value of GRS and CGRS (Supplementary Fig. 6a-c). Further analysis indicated that CGRS and GRS were negatively correlated with the levels of mutation load and DNA methylation (Supplementary Fig. 6g-i), which was consistent with GC patients with high mutation or methylation loads tend to have better prognosis (Supplementary Fig. 6d-f). According to Singapore study, the metabolic subtype of GC patients that have relatively longer survival time acquired lower value of GRS and CGRS than other subtypes. The invasive subtype of GC patients that showed relatively poor prognosis displayed high value of GRS and CGRS (Supplementary Fig. 7a-c). In ACRG cohort, GC patients have been classified into four molecular subtypes with different clinical outcomes. The MSS/EMT subtype that has the poorest outcome acquired relatively higher value of GRS and CGRS (Supplementary Fig. 7d-f). Taken together, these results suggest that our CGRS and GRS are significantly associated with molecular subtypes with significant survival differences.

Comparisons with other established GC signatures

To investigate the prediction accuracy of GRS and CGRS in GC prognosis, we compared the prediction power of GRS and CGRS with other three published gene signatures [15, 16, 38]. All of the three signatures were significantly associated with GC prognosis in multiple cohorts (Supplementary Table 10). Since GRS and CGRS contained no overlap genes with other signatures, we computed ROC of signatures and the AJCC staging system in four cohorts. GRS had larger area under the curve (AUC) according to ROC analysis compared with published signatures (Fig 4a and b). Further prediction error curve analysis also indicated that GRS showed lower prediction error rate in evaluating GC prognosis (Supplementary Fig. 8). However, GRS showed no advantages in predicting GC prognosis compared with the AJCC staging system. Moreover, CGRS that integrated GRS with clinical characteristics could predict GC prognosis with more sensitivity and specificity according to the ROC analysis (Fig. 4a and b). The prediction error curve analysis also revealed that CGRS had relatively lower prediction error rate in four independent cohorts (Supplementary Fig. 8). The above results demonstrate that CGRS has more advantages in predicting GC prognosis compared with the AJCC staging system and several published signatures in most cohorts we obtained.

Potential clinical application of CGRS

To facilitate the clinical applications of CGRS, we generated an easy-to-use nomogram for predicting the 1-, 3- and 5-year overall survival probability of GC patients using CGRS (Fig. 5a). The nomogram was evaluated for its calibration by plotting predicted probabilities at 1, 3, and 5 years, respectively. The overall survival probability predicted by nomogram was close to the observed probability at these three thresholds (Fig. 5b). Furthermore, the decision curve analysis showed that CGRS could bring more benefits for high risk GC patients in clinical applications (Fig. 5c). Moreover, we developed an online tool for conveniently applying CGRS in clinical practice

(http://39.100.117.92/CGRS/). In the web application, the oncologists only need to select the data type, and then input age, the AJCC stage, and nine gene expression values of an individual GC patient. When clicking the Calculate button, 1-, 3- and 5-year overall survival predicted probabilities will be calculated for the patient. These findings indicate that the easy-to-use nomogram and web application may accelerate the application of CGRS in predicting GC prognosis in clinical practice.

Biological pathways involved in GC prognosis

To investigate the biological processes and pathways involved in GC prognosis, we dichotomized the patients from ACRG and TCGA cohorts into high and low CGRS groups according to the median CGRS value of ACRG training set, respectively. GSEA was subsequently performed to identify prognostic biological processes and pathways. Functional networks based on significantly enriched gene sets were built by enrichment map (FDR < 0.05) (Fig. 6a and b; Supplementary Table. 11). Intriguingly, the cell cycle, RNA transcription, apoptosis and cell metabolism pathways were significantly enriched in low CGRS patients from ACRG (Fig. 6c-f) and TCGA cohorts (Figure. 6i-6l). However, extracellular matrix pathways that play important roles in tumor invasion and metastasis were significantly enriched in high CGRS patients (Fig. 6g and m). Furthermore, T cell receptors were also significantly enriched in high CGRS patients, which indicated that high CGRS patients might have more neoantigens for immunotherapy (Fig.6h and n). Taken together, these data suggest that genes correlated with cell cycle and tumor microenvironment might be involved in GC prognosis.

Discussion

The current assessment of GC prognosis is mainly based on the AJCC staging system [33, 37]. However, the AJCC staging system is not sensitive and accurate enough in predicting the survival of GC patients with similar clinical and pathological characteristics [34, 35]. Previous signatures based on genetic or clinicopathological features for GC have been built to solve this problem in previous studies [14-16]. Meanwhile, model overfitting and lack of adequate validation largely hinder their clinical applications [39]. As far as we know, no signature has been established by integrating clinical data with gene expression profiles in GC patients. Here, we developed CGRS by integrating genetic signature with clinical risk score for GC patients. CGRS was validated in four independent cohorts to ensure its robustness in evaluating GC prognosis (Fig. 2). CGRS showed more sensitivity and specificity than previously published prognostic signatures according to both ROC and PEC analyses (Fig.4; Supplementary Fig. 8), which have been widely used to assess the prediction power in survival analysis. Moreover, CGRS showed stronger robustness than the AJCC staging system in TCGA, ACRG, and Zhejiang cohorts (Fig. 4). In further subset analysis, CGRS was able to stratify stage III GC patients into high and low risk groups with significantly different overall survival rates. Therefore, our results indicate that CGRS shows strong robustness in predicting GC prognosis.

Another major challenge for prognostic risk scores is the complex calculation procedure in clinical practice. Given that the nomogram has been developed for predicting prognosis in cancer [40, 41], we tried to establish an easy-to-use nomogram based on CGRS for predicting the overall survival probability of GC patients at 95% confidence interval, which was confirmed by the calibration plot and decision curve (Fig. 5). Furthermore, we developed a web-based tool (http://39.100.117.92/CGRS/) to facilitate the clinical application of CGRS. The users only need to select the data type (RNASeq, microarray, or qRT-PCR), and input nine gene expression values, age, and the AJCC stage of the individual patient and press the Calculate button, 1-, 3- and 5-year predicted overall survival

probabilities will be calculated for the patient. Thus, our nomogram and web application based on CGRS can be easily applied in evaluating the prognosis of GC patients in clinical practice.

In line with our expectations, several genes in GRS are also present in other prognostic signatures [42-44]. For example, APOD gene encoding a component of high-density lipoprotein has been reported to promote cell migration through interacting with growth factors [45], and higher APOD mRNA levels indicate poor survival in breast or colorectal cancer patients [46, 47]. TSPYL5 gene contributes to breast cancer progression by reducing p53 protein levels and inhibiting the expression of p53-target genes [48]. TSPYL5 has been also documented as an independent prognostic factor in breast or liver cancer patients [43, 49]. GSDME gene plays an important role in pyroptosis, and is reported to be used as a prognostic factor in oesophageal squamous cell carcinoma [50, 51]. For the other genes incorporated into GRS system so far, no association with prognosis has been reported yet. The biological function and potential mechanism of the nine genes in GC need to be further investigated.

Similar to other prognostic signatures, one limitation of our signature is the high heterogeneity of GC [52]. Patients of different races and different regions may have inconsistent gene expression patterns [53]. Although we have validated our CGRS on patients of different races and areas, including ACRG (Asian, Korea), Singapore (Asian, Singapore), TCGA (Caucasian, USA) and Zhejiang (Asian, China) cohorts retrospectively, the prospective, multicenter clinical trials with large population are still necessary for validating the robustness of CGRS. In addition, due to the small population size of early stage and metastatic GC samples, CGRS is not robust enough for predicting the prognosis of early stage and metastatic GC patients.

Conclusion

Our CGRS integrated genetic signature with clinical features shows great robustness in predicting prognosis of GC patients. CGRS has good performance in stratifying stage III GC patients. CGRS-based nomogram and web application have been developed to conveniently predict the prognosis of GC patients in clinical practice.

Abbreviations

GC: Gastric Cancer; TCGA: The Cancer Genome Atlas; ACRG: Asian Cancer Research Group; CGRS: Clinic and Genetic Risk Score; GRS: Genetic Risk Score; CRS: Clinic Risk Score; SEER: Surveillance, Epidemiology, and End Results database; ROC: Receiver operating characteristic; AUC: Area Under the Curve; PEC: Prediction error curve; AJCC: American Joint Committee on Cancer; OS: Overall survival; HR: Hazard ratio; CI: Confidence interval; LASSO: Least Absolute Shrinkage and Selection Operator; RSEM: RNA-Seq by Expectation-Maximization; GSDME: gasdermin E; VAT1: vesicle amine transport 1; APOD: apolipoprotein D ST8SIA5: ST8alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 5; STARD3NL: STARD3 N-terminal like; TMEM245: transmembrane protein 245; CCDC92: coiled-coil domain containing 92; TSPYL5: TSPY like 5;; CYS1: cystin 1; qRT-PCR: Real-Time Quantitative Reverse Transcription PCR; RNU6-1: RNA, U6 Small Nuclear 1; MSI: Microsatellite Instability; MSS: Microsatellite Stability; EMT: Epithelial-to-mesenchymal transition; FDR: False Discovery Rate; MsigDB: Molecular Signatures Database; GSEA: Gene Set Enrichment Analysis.

Declarations

Acknowledgments

The authors are grateful to Richard Yang and Xuefei Gao (Princeton University, Princeton, NJ, USA) for help with editing the language of our manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (91740205, 31620103911, 31771540, 31301149), the 111 Project (B13026) and Fundamental Research Funds for the Central Universities (2017QNA7005).

Availability of data and materials

All data are available within the article and supplementary files, or available from the author upon reasonable request.

Ethics approval and consent to participate

All aspects of this study were approved by the ethics committee of Zhejiang University School of Medicine. All participants gave written, informed consent. Research was conducted in accordance with the Declaration of Helsinki guidelines for the ethical conduct of research in 1975.

Consent for publication

All authors agree to submit the article for publication.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

Study concept and design: QS, LW, TZ. Acquisition of data: QS, DG, SL, YX, MJ, YL. Analysis and interpretation of data: QS, DG, SL, CH, HD, SZ. Improvements of the project design and interpreted the results: WZ, WL. Drafting of the manuscript: QS, LW, TZ. Critical revision of the manuscript: all authors. Study supervision: LW, TZ.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel R L, Torre L A, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
2. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. Lancet. 2016;388(10060):2654-2664.
3. Allemani C, Weir H K, Carreira H, Harewood R, Spika D, Wang X S, Bannon F, Ahn J V, Johnson C J, Bonaventure A, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet. 2015;385(9972):977-1010.
4. Arpino G, Generali D, Sapino A, Del Matro L, Frassoldati A, de Laurentis M, Pronzato P, Mustacchi G, Cazzaniga M, De Placido S, et al. Gene expression profiling in breast cancer: a clinical perspective. Breast.

2013;22(2):109-120.

5. Bild A H, Yao G, Chang J T, Wang Q, Potti A, Chasse D, Joshi M B, Harpole D, Lancaster J M, Berchuck A, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature. 2006;439(7074):353-357.
6. Gentles A J, Bratman S V, Lee L J, Harris J P, Feng W, Nair R V, Shultz D B, Nair V S, Hoang C D, West R B, et al. Integrating Tumor and Stromal Gene Expression Signatures With Clinical Indices for Survival Stratification of Early-Stage Non-Small Cell Lung Cancer. J Natl Cancer Inst. 2015;107(10):34-45.
7. Guo N L, Wan Y W, Tosun K, Lin H, Msiska Z, Flynn D C, Remick S C, Vallyathan V, Dowlati A, Shi X, et al. Confirmation of gene expression-based prediction of survival in non-small cell lung cancer. Clin Cancer Res. 2008;14(24):8213-8220.
8. Azim H A, Jr., Michiels S, Zagouri F, Delaloge S, Filipits M, Namer M, Neven P, Symmans W F, Thompson A, Andre F, et al. Utility of prognostic genomic tests in breast cancer practice: The IMPAKT 2012 Working Group Consensus Statement. Ann Oncol. 2013;24(3):647-654.
9. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner F L, Walker M G, Watson D, Park T, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351(27):2817-2826.
10. Alizadeh A A, Gentles A J, Alencar A J, Liu C L, Kohrt H E, Houot R, Goldstein M J, Zhao S, Natkunam Y, Advani R H, et al. Prediction of survival in diffuse large B-cell lymphoma based on the expression of 2 genes reflecting tumor and microenvironment. Blood. 2011;118(5):1350-1358.
11. Lossos I S, Czerwinski D K, Alizadeh A A, Wechser M A, Tibshirani R, Botstein D, Levy R. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med. 2004;350(18):1828-1837.
12. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202-209.
13. Cristescu R, Lee J, Nebozhyn M, Kim K M, Ting J C, Wong S S, Liu J, Yue Y G, Wang J, Yu K, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21(5):449-456.
14. Busuttil R A, George J, Tothill R W, Ioculano K, Kowalczyk A, Mitchell C, Lade S, Tan P, Haviv I, Boussioutas A. A signature predicting poor prognosis in gastric and ovarian cancer represents a coordinated macrophage and stromal response. Clin Cancer Res. 2014;20(10):2761-2772.
15. Cho J Y, Lim J Y, Cheong J H, Park Y Y, Yoon S L, Kim S M, Kim S B, Kim H, Hong S W, Park Y N, et al. Gene expression signature-based prognostic risk score in gastric cancer. Clin Cancer Res. 2011;17(7):1850-1857.
16. Chen C N, Lin J J, Chen J J, Lee P H, Yang C Y, Kuo M L, Chang K J, Hsieh F J. Gene expression profile predicts patient survival of gastric cancer after surgical resection. J Clin Oncol. 2005;23(29):7286-7295.
17. Goseki N, Takizawa T, Koike M. Differences in the mode of the extension of gastric cancer classified by histological type: new histological classification of gastric carcinoma. Gut. 1992;33(5):606-612.
18. Tan I B, Ivanova T, Lim K H, Ong C W, Deng N, Lee J, Tan S H, Wu J, Lee M H, Ooi C H, et al. Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. Gastroenterology. 2011;141(2):476-485.
19. Doll K M, Rademaker A, Sosa J A. Practical Guide to Surgical Data Sets: Surveillance, Epidemiology, and End Results (SEER) Database. JAMA Surg. 2018;153(6):588-589.

20. Dai M, Wang P, Boyd A D, Kostov G, Athey B, Jones E G, Bunney W E, Myers R M, Speed T P, Akil H, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res. 2005;33(20):175-180.
21. Newman A M, Cooper J B. AutoSOME: a clustering method for identifying gene expression modules without prior knowledge of cluster number. BMC Bioinformatics. 2010;11117(10):568-575.
22. Yu G, Wang L G, Han Y, He Q Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-287.
23. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov J P, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-425.
24. Subramanian A, Tamayo P, Mootha V K, Mukherjee S, Ebert B L, Gillette M A, Paulovich A, Pomeroy S L, Golub T R, Lander E S, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-15550.
25. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019;14(2):482-517.
26. Mogensen U B, Ishwaran H, Gerds T A. Evaluating Random Forests for Survival Analysis using Prediction Error Curves. J Stat Softw. 2012;50(11):1-23.
27. Schroder M S, Culhane A C, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics. 2011;27(22):3206-3208.
28. Walz J, Gallina A, Saad F, Montorsi F, Perrotte P, Shariat S F, Jeldres C, Graefen M, Benard F, McCormack M, et al. A nomogram predicting 10-year life expectancy in candidates for radical prostatectomy or radiotherapy for prostate cancer. J Clin Oncol. 2007;25(24):3576-3581.
29. Fitzgerald M, Saville B R, Lewis R J. Decision curve analysis. JAMA. 2015;313(4):409-410.
30. Cheong J H, Yang H K, Kim H, Kim W H, Kim Y W, Kook M C, Park Y K, Kim H H, Lee H S, Lee K H, et al. Predictive test for chemotherapy response in resectable gastric cancer: a multi-cohort, retrospective analysis. Lancet Oncol. 2018;19(5):629-638.
31. Zhou Y, Huang T, Zhang J, Wong C C, Zhang B, Dong Y, Wu F, Tong J H M, Wu W K K, Cheng A S L, et al. TEAD1/4 exerts oncogenic role and is negatively regulated by miR-4269 in gastric tumorigenesis. Oncogene. 2017;36(47):6518-6530.
32. Lamba V, Ghodke-Puranik Y, Guan W, Lamba J K. Identification of suitable reference genes for hepatic microRNA quantitation. BMC Res Notes. 2014;7129(23):6134-6139.
33. Reim D, Loos M, Vogl F, Novotny A, Schuster T, Langer R, Becker K, Hofler H, Siveke J, Bassermann F, et al. Prognostic implications of the seventh edition of the international union against cancer classification for patients with gastric cancer: the Western experience of patients treated in a single-center European institution. J Clin Oncol. 2013;31(2):263-271.
34. Rizk N P, Venkatraman E, Bains M S, Park B, Flores R, Tang L, Ilson D H, Minsky B D, Rusch V W, American Joint Committee on C. American Joint Committee on Cancer staging system does not accurately predict survival in patients receiving multimodality therapy for esophageal adenocarcinoma. J Clin Oncol. 2007;25(5):507-512.
35. Thompson S K, Ruszkiewicz A R, Jamieson G G, Esterman A, Watson D I, Wijnhoven B P, Lamb P J, Devitt P G. Improving the accuracy of TNM staging in esophageal cancer: a pathological review of resected specimens.

Ann Surg Oncol. 2008;15(12):3447-3458.

36. Catalano V, Labianca R, Beretta G D, Gatta G, de Braud F, Van Cutsem E. Gastric cancer. Crit Rev Oncol Hematol. 2009;71(2):127-164.
37. In H, Solsky I, Palis B, Langdon-Embry M, Ajani J, Sano T. Validation of the 8th Edition of the AJCC TNM Staging System for Gastric Cancer using the National Cancer Database. Ann Surg Oncol. 2017;24(12):3683-3691.
38. Zhu X, Tian X, Sun T, Yu C, Cao Y, Yan T, Shen C, Lin Y, Fang J Y, Hong J, et al. GeneExpressScore Signature: a robust prognostic and predictive classifier in gastric cancer. Mol Oncol. 2018;12(11):1871-1883.
39. Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? J Natl Cancer Inst. 2010;102(7):464-474.
40. Filipits M, Dubsky P, Rudas M, Greil R, Balic M, Bago-Horvath Z, Singer C F, Hlauschek D, Brown K, Bernhisel R, et al. Prediction of Distant Recurrence Using EndoPredict Among Women with ER(+), HER2(-) Node-Positive and Node-Negative Breast Cancer Treated with Endocrine Therapy Only. Clin Cancer Res. 2019;25(13):3865-3872.
41. Mell L K, Shen H, Nguyen-Tan P F, Rosenthal D I, Zakeri K, Vitzthum L K, Frank S J, Schiff P B, Trotti A M, 3rd, Bonner J A, et al. Nomogram to Predict the Benefit of Intensive Treatment for Locoregionally Advanced Head and Neck Cancer. Clin Cancer Res. 2019;25(23):7078-7088.
42. Patsialou A, Wang Y, Lin J, Whitney K, Goswami S, Kenny P A, Condeelis J S. Selective gene-expression profiling of migratory tumor cells in vivo predicts clinical outcome in breast cancer patients. Breast Cancer Res. 2012;14(5):139-146.
43. Qiu X, Hu B, Huang Y, Deng Y, Wang X, Zheng F. Hypermethylation of ACP1, BMP4, and TSPYL5 in Hepatocellular Carcinoma and Their Potential Clinical Significance. Dig Dis Sci. 2016;61(1):149-157.
44. van 't Veer L J, Dai H, van de Vijver M J, He Y D, Hart A A, Mao M, Peterse H L, van der Kooy K, Marton M J, Witteveen A T, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415(6871):530-536.
45. Leung W C, Lawrie A, Demaries S, Massaeli H, Burry A, Yablonsky S, Sarjeant J M, Fera E, Rassart E, Pickering J G, et al. Apolipoprotein D and platelet-derived growth factor-BB synergism mediates vascular smooth muscle cell migration. Circ Res. 2004;95(2):179-186.
46. Soiland H, Skaland I, Varhaug J E, Korner H, Janssen E A, Gudlaugsson E, Baak J P, Soreide J A. Co-expression of estrogen receptor alpha and Apolipoprotein D in node positive operable breast cancer--possible relevance for survival and effects of adjuvant tamoxifen in postmenopausal patients. Acta Oncol. 2009;48(4):514-521.
47. Bajo-Graneras R, Crespo-Sanjuan J, Garcia-Centeno R M, Garrote-Adrados J A, Gutierrez G, Garcia-Tejeiro M, Aguirre-Gervas B, Calvo-Nieves M D, Bustamante R, Ganfornina M D, et al. Expression and potential role of apolipoprotein D on the death-survival balance of human colorectal cancer cells under oxidative stress conditions. Int J Colorectal Dis. 2013;28(6):751-766.
48. Epping M T, Meijer L A, Krijgsman O, Bos J L, Pandolfi P P, Bernards R. TSPYL5 suppresses p53 levels and function by physical interaction with USP7. Nat Cell Biol. 2011;13(1):102-108.
49. Li J, Liu C, Chen Y, Gao C, Wang M, Ma X, Zhang W, Zhuang J, Yao Y, Sun C. Tumor Characterization in Breast Cancer Identifies Immune-Relevant Gene Signatures Associated With Prognosis. Front Genet. 2019;101119(10):168-179.

50. Wang Y, Gao W, Shi X, Ding J, Liu W, He H, Wang K, Shao F. Chemotherapy drugs induce pyroptosis through caspase-3 cleavage of a gasdermin. Nature. 2017;547(7661):99-103.
51. Wu M, Wang Y, Yang D, Gong Y, Rao F, Liu R, Danna Y, Li J, Fan J, Chen J, et al. A PLK1 kinase inhibitor enhances the chemosensitivity of cisplatin by inducing pyroptosis in oesophageal squamous cell carcinoma. EBioMedicine. 2019;12(3):41244-255.
52. Ziogas D E, Roukos D H. Limitations of isolated tumor cells in gastric cancer: heterogeneity requests systems biology approaches towards personalized medicine. Ann Surg Oncol. 2010;17(1):343-344.
53. Boussioutas A, Li H, Liu J, Waring P, Lade S, Holloway A J, Taupin D, Gorringe K, Haviv I, Desmond P V, et al. Distinctive patterns of gene expression in premalignant gastric mucosa and gastric cancer. Cancer Res. 2003;63(10):2569-2577.

Tables

Table1. Clinical features of the cohorts of GC patients included for training and validation of GRS and CGRS

Variables	ACRG training set No.(%)	ACRG validation set No.(%)	Singapore cohort No.(%)	Korea cohort No.(%)	TCGA cohort No.(%)	Zhejiang cohort No.(%)
Tissue Type	Fresh/Frozen	Fresh/Frozen	Fresh/Frozen	Fresh/Frozen	Fresh/Frozen	Fresh/Frozen
Platform	Affymetrix microarrays	Affymetrix microarrays	Affymetrix microarrays	Illumina microarrays	RNASeq	qRT-PCR
No. of samples	150	150	192	433	388	109
Median age, y (range)	65 (28-86)	63 (24-84)	67 (23-92)	62 (27-36)	67 (30-90)	65 (28-85)
Male	96 (64.0)	103 (68.7)	125 (65.1)	296 (68.4)	252 (64.9)	76 (69.7)
Female	54 (36.0)	47 (31.3)	67 (34.9)	137 (31.6)	136 (35.1)	33 (30.3)
Stage I	20 (13.3)	11 (7.3)	31 (16.1)	NA	51 (13.1)	6 (5.5)
Stage II	60 (40.0)	37 (24.7)	29 (15.1)	NA	122 (31.4)	17 (15.6)
Stage III	34 (22.7)	61 (40.7)	72 (37.5)	NA	167 (43.0)	79 (72.5)
Stage IV	35 (23.3)	40 (26.7)	60 (31.3)	NA	40 (10.3)	5 (4.6)
Intestinal	82 (54.7)	64 (42.7)	99 (51.6)	NA	306 (78.9)	NA
Diffuse	58 (38.7)	76 (50.7)	75 (39.1)	NA	67 (17.3)	NA
Median follow-up, mo (range)	59.8 (2.7-104.9)	52.6 (1.0-105.7)	19.0 (0-157.8)	69.0 (0-161.0)	15.6 (0-124.0)	40.5 (0.9-119.4)
No. of deaths	66 (44)	86 (57.3)	95 (49.5)	209 (48.3)	158 (40.7)	68 (62.4)

ACRG = Asian cancer research group, data from GSE62254; GC = Gastric cancer; Korea cohort data from GSE84437; Singapore cohort data from GSE15459; TCGA = The cancer genome atlas, data from firebrowse. qRT-PCR = quantitative real time polymerase chain reaction; y = years; mo = months.

Table 2. Univariate and multivariable analysis of GRS, CRS, and CGRS in multiple cohorts.

Variables	ACRG cohort		Singapore cohort		TCGA cohort		Zhejiang cohort	
	HR (95.0% CI)	*P*-value	HR (95.0% CI)	*P*-value	HR (95.0% CI)	*P*-value	HR (95.0% CI)	*P*-value
Univariate								
GRS (all stages)	1.850 (1.579-2.168)	<0.001	1.476 (1.225-1.778)	<0.001	1.289 (1.092-1.520)	0.002	1.384 (1.114-1.721)	0.002
GRS (stage III)	1.838 (1.376-2.453)	<0.001	1.571 (1.411-2.162)	<0.001	1.234 (0.984-1.548)	0.063	1.310 (1.023-1.675)	0.028
CRS (age, stage)	3.745 (2.813-4.986)	<0.001	4.113 (2.865-5.905)	<0.001	2.994 (2.115-4.239)	<0.001	2.923 (1.539-9.844)	<0.001
Multivariate (all stages)								
GRS	1.699 (1.455-1.985)	<0.001	1.549 (1.272-1.885)	<0.001	1.373 (1.155-1.633)	<0.001	1.460 (1.164-1.830)	<0.001
CRS (age, stage)	3.382 (2.526-4.528)	<0.001	4.511 (3.077-6.614)	<0.001	3.178 (2.248-4.493)	<0.001	3.723 (1.840-7.533)	<0.001
CGRS (H vs L, all stages)	4.304 (2.909-6.368)	<0.001	4.539 (2.637-7.815)	<0.001	2.707 (1.872-3.913)	<0.001	2.689 (1.507-4.798)	<0.001
CGRS (H vs L, stage III)	2.522 (1.322-4.812)	0.003	3.064 (1.335-7.032)	0.003	2.511 (1.425-4.423)	<0.001	1.994 (1.088-3.654)	0.019
CGRS (all stages)	2.104 (1.846-2.398)	<0.001	2.123 (1.742-2.588)	<0.001	1.710 (1.435-2.038)	<0.001	1.664 (1.293-2.141)	<0.001
CGRS (stage III)	2.373 (1.646-3.420)	<0.001	1.861 (1.290-2.685)	<0.001	1.498 (1.141-1.965)	0.003	1.439 (1.098-1.886)	0.006

P values were calculated using two sided likelihood ratio test. CI = confidence interval; HR = hazard ratio; GRS = Genetic risk score; CRS =Clinical risk score; CGRS = Combined genetic and clinical risk score; H: high risk; L: low risk.

Figures

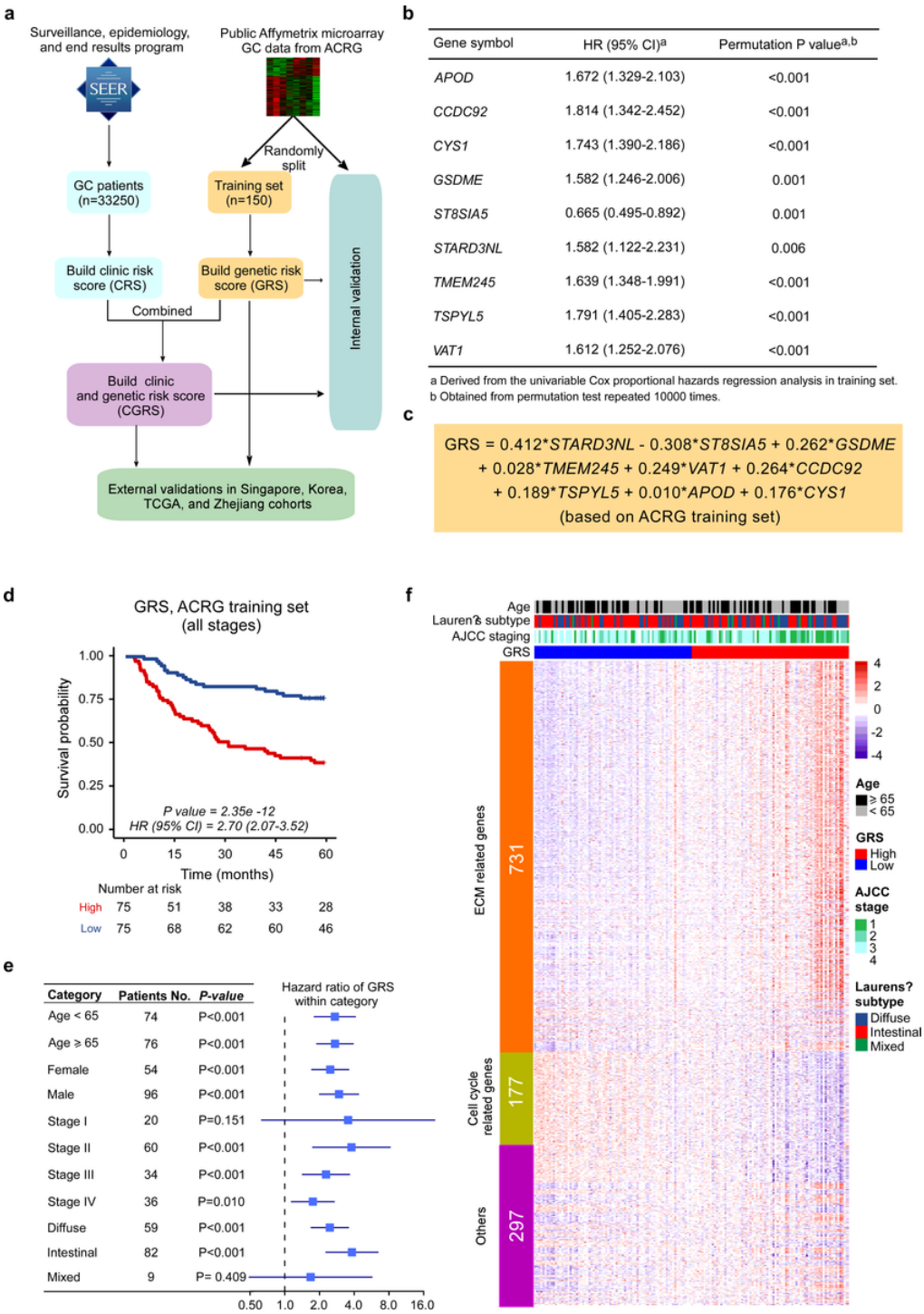


Figure 1

Construction of prediction signatures for GC prognosis. a Schematic representation of the study design. b Univariate Cox regression results of the nine genes in the prognosis prediction signature from ACRG training set. c The formula used to calculate GRS for each GC patient. d Kaplan-Meier analysis of 5-year overall survival in ACRG training set. P values were calculated by using the log-rank test. e Forest plot was used to depict the prediction performance of GRS in ACRG training set within each subgroup. P values were calculated by two-sided likelihood ratio test. f The heatmap reflects the GRS-related genes expression levels of two dominant clusters revealed by AutoSOME in ACRG training set.

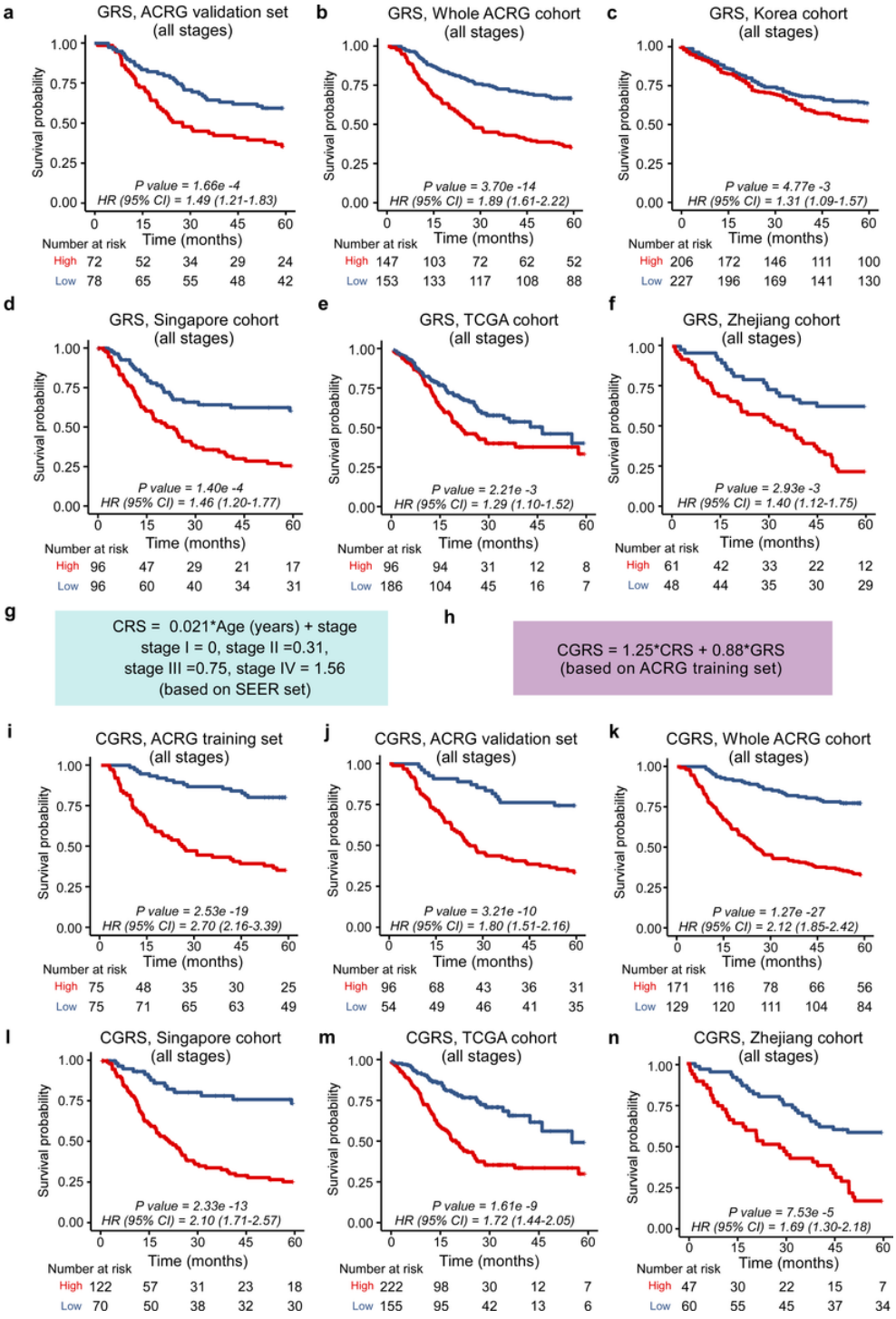


Figure 2

Kaplan-Meier analysis of 5-year overall survival of GC patients based on GRS or CGRS. a-f 5-year overall survival prediction of GC patients according to GRS in ACRG validation set (a), ACRG (b), Singapore (c), Korea (d), TCGA (e), and Zhejiang (f) cohorts. g, h The formulas used to calculate CRS and CGRS for each GC patients. i-n 5-year overall survival prediction of GC patients according to the CGRS in ACRG training set (i), ACRG validation set (j), ACRG (k), Singapore (l), TCGA (m), and Zhejiang (n) cohorts. The median value from the ACRG training set was used as the cut-off to classify patients to high and low risk groups. HRs and 95% CIs were calculated using the Cox regression method. P values were calculated using the log-rank test. Tick marks on curves represent censoring.

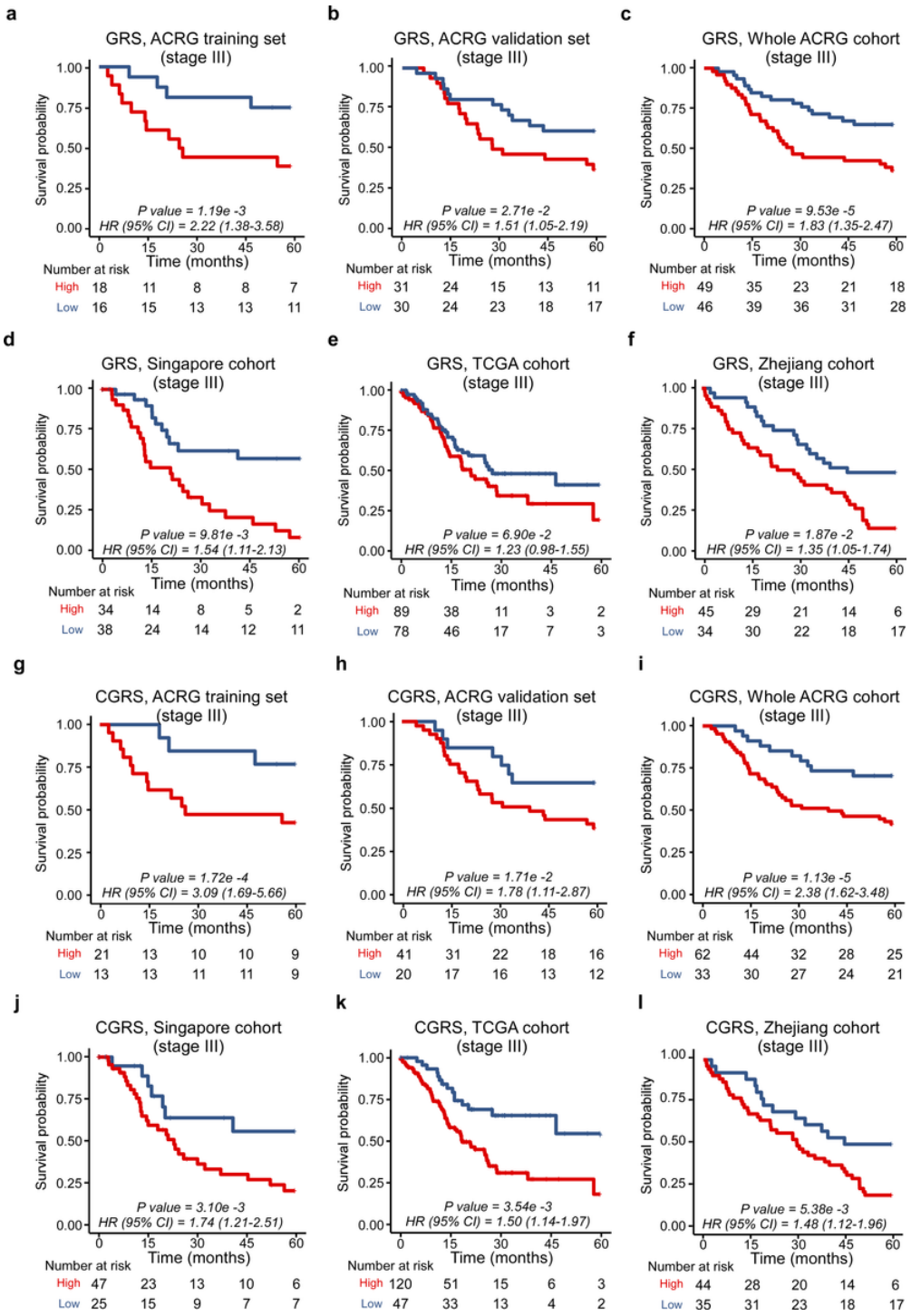
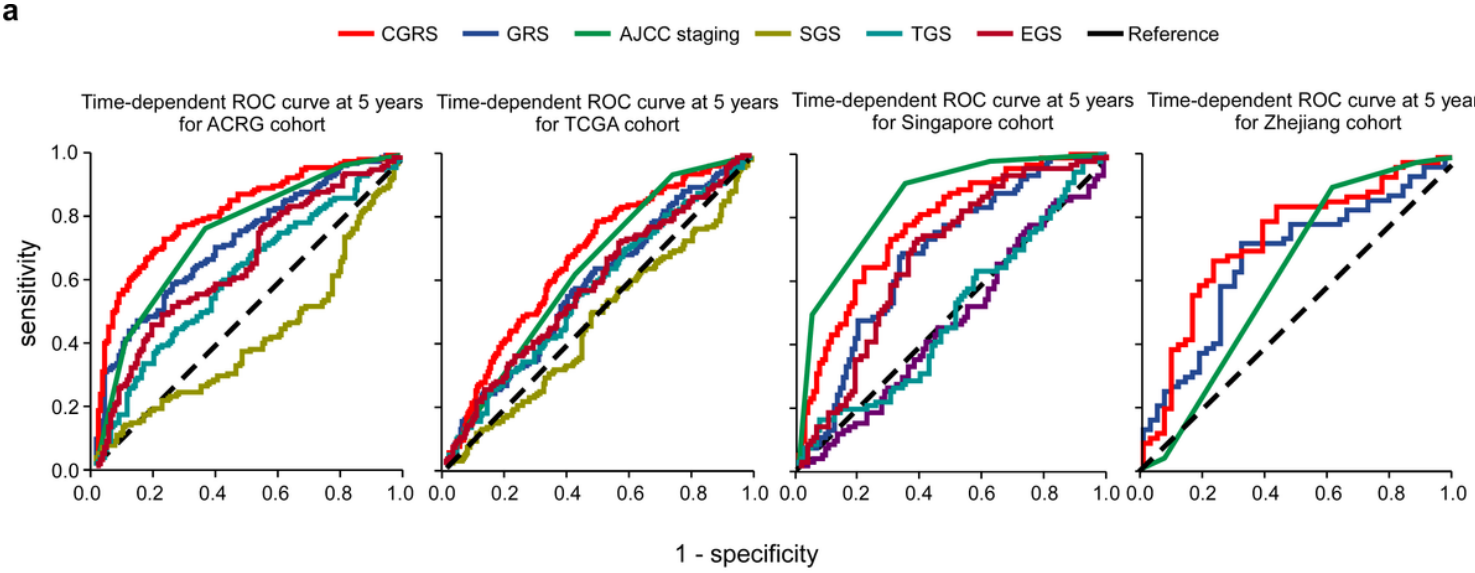


Figure 3

Kaplan-Meier analysis of 5-year overall survival of Stage III GC patients based on GRS or CGRS. a-f 5-year overall survival prediction of Stage III GC patients according to GRS in ACRG training set (a), ACRG validation set (b), ACRG (c), Singapore (d), TCGA (e), and Zhejiang (f) cohorts. g-l 5-year overall survival prediction of Stage III GC patients according to CGRS in ACRG training set (g), ACRG validation set (h), ACRG (i), Singapore (j), TCGA (k), and Zhejiang (l) cohorts. The median value from the ACRG training set was used as the cut-off to classify patients to high and low risk groups. HRs and 95% CIs were calculated using the Cox regression method. P values were calculated using the log-rank test. Tick marks on curves represent censoring.

a



b

The ROC analysis at 5 years for each cohort

	ACRG				TCGA				Singapore				Zhejiang			
	AUC	P^b	P^c	P^d	AUC	P^b	P^c	P^d	AUC	P^b	P^c	P^d	AUC	P^b	P^c	P^d
CGRS	0.819	-	1.000	1.000	0.671	-	1.000	1.000	0.773	-	1.000	<0.001	0.737	-	1.000	1.000
GRS	0.729	<0.001	-	<0.001	0.577	<0.001	-	<0.001	0.680	<0.001	-	<0.001	0.675	<0.001	-	<0.001
Stage[a]	0.740	<0.001	1.000	-	0.614	<0.001	1.000	-	0.855	1.000	1.000	-	0.627	<0.001	<0.001	-
SGS	0.424	<0.001	<0.001	<0.001	0.466	<0.001	<0.001	<0.001	0.463	<0.001	<0.001	<0.001	-	-	-	-
TGS	0.620	<0.001	<0.001	<0.001	0.556	<0.001	0.285	<0.001	0.487	<0.001	<0.001	<0.001	-	-	-	-
EGS	0.655	<0.001	<0.001	<0.001	0.569	<0.001	0.321	<0.001	0.674	<0.001	1.000	<0.001	-	-	-	-

ROC = receiver operating characteristic curve; AUC = area under the curve; SGS (six gene signature, obtained from Cho et.al). TGS (three gene signature, obtained from Chen et.al). EGS (eight gene signature, obtained from Zhu et.al). a: AJCC staging system. b: P values were calculated if CGRS is greater than other signatures. c: P values were calculated if GRS is greater than other signatures. d: P values were calculated if AJCC staging is greater other than signatures.

Figure 4

Time-dependent receiver operating curve analysis of different signatures in multiple cohorts. a Time-dependent receiver operating curves at 5 years for ACRG, TCGA, Singapore, and Zhejiang cohorts. b The area under the curve for different signatures in each cohort.

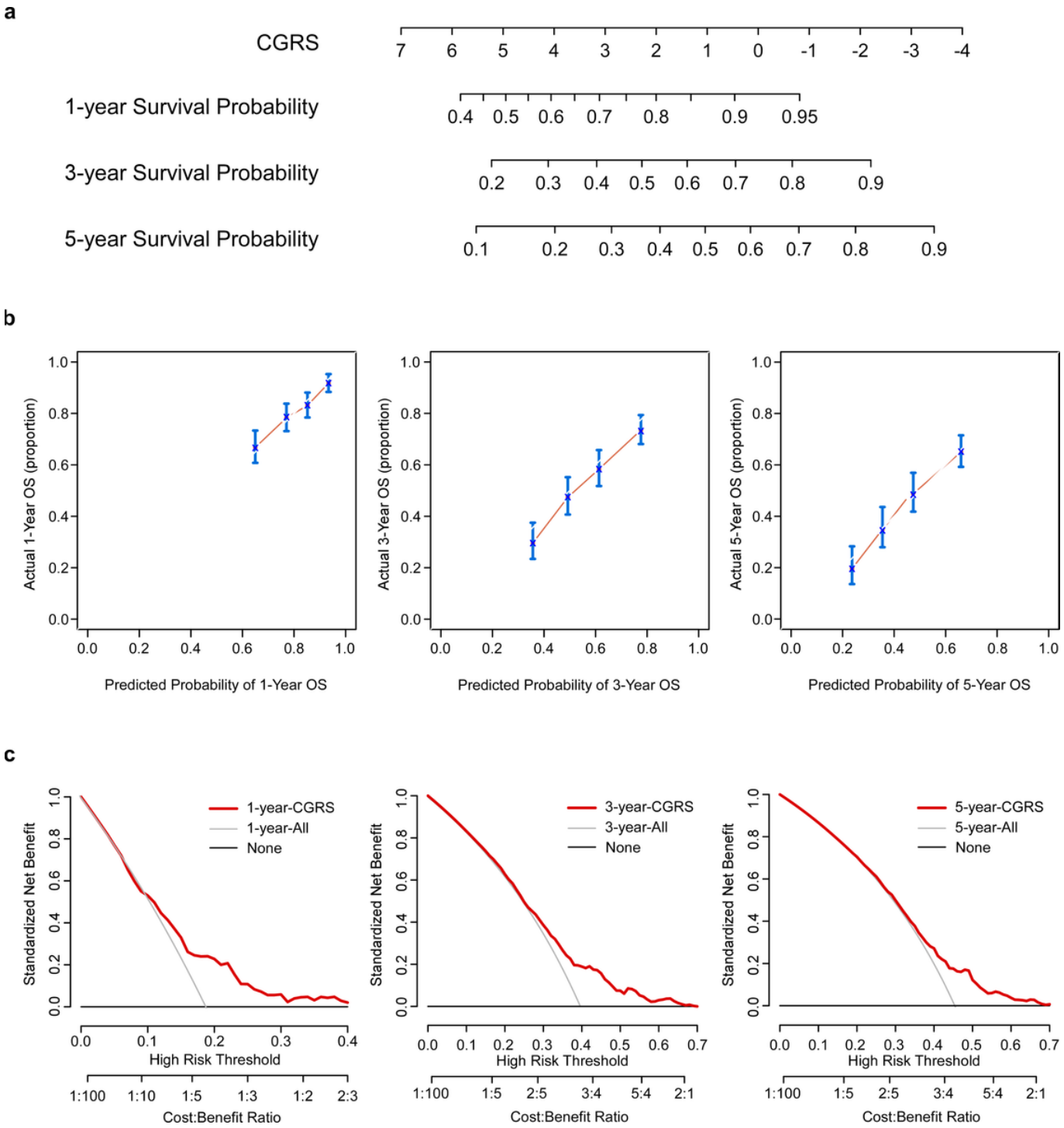


Figure 5

Nomogram based on CGRS predicts 1-, 3-, and 5-year overall survival probability of GC patients. a The nomogram was generated using data from all platforms. The overall survival probability of 1-, 3- and 5-year can be acquired based on CGRS. b The calibration curve for predicting 1-, 3- and 5-year overall survival probability of GC patients. The y-axis represents the actual overall survival probability of the GC patients, and the x-axis represents the nomogram-predicted survival probability. The solid vertical lines represent the 95% CIs, the gray line indicates an ideal nomogram that has 100% accuracy. c Decision Curve Analysis (DCA) of the nomogram at 1-, 3- and 5-year. The curves could assess the clinical benefits of the nomogram. The y-axis represents the net benefit, and the x-axis represents the threshold probability. Gray solid lines assume that all patients will die at 1-, 3- and 5-year

("treat all"). Solid horizontal lines assume that no one will die at 1-, 3- and 5-year ("treat none"). The red solid lines indicate the prediction model of the nomogram.

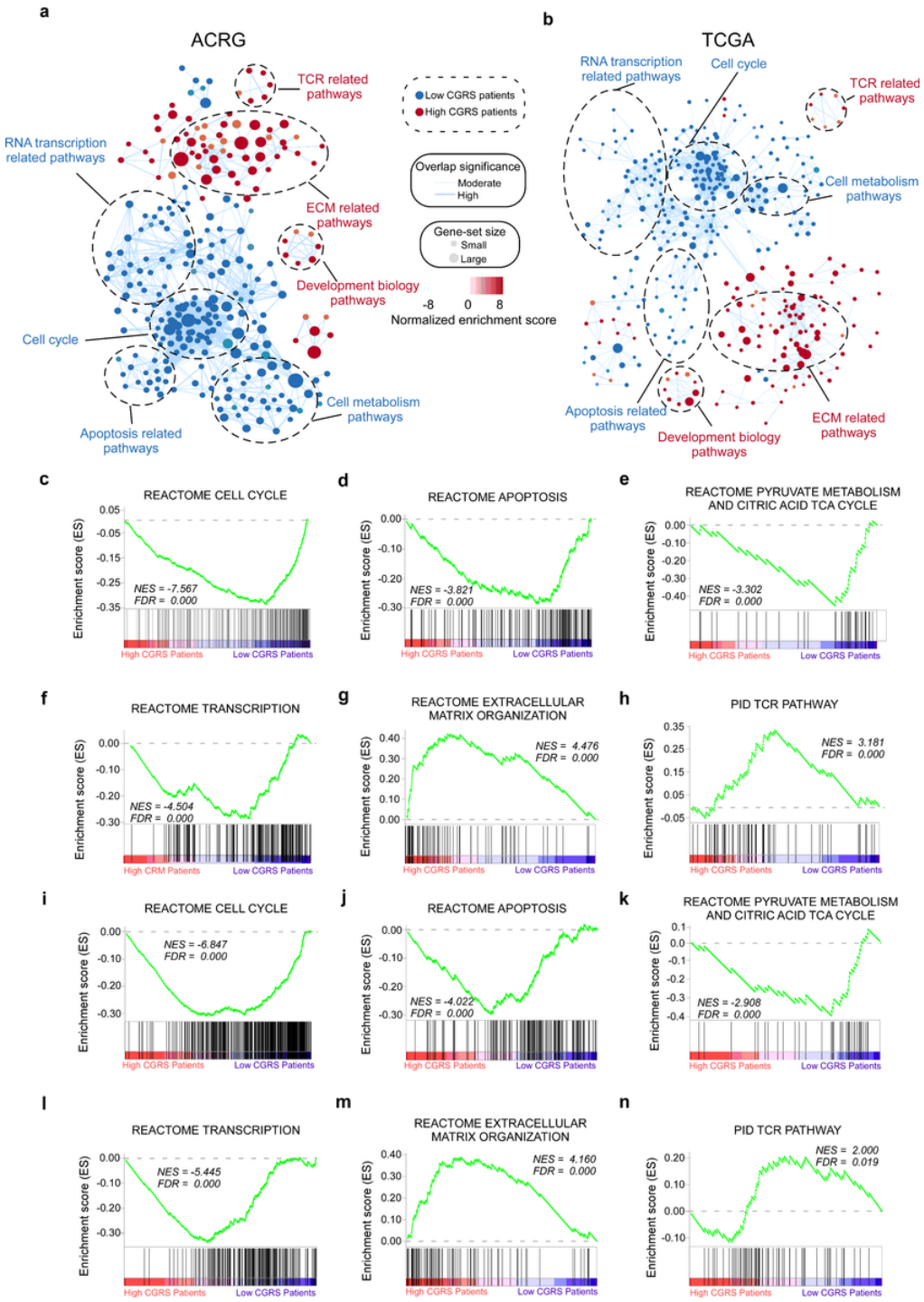


Figure 6

The association between CGRS and biological pathways and processes is evaluated via gene set enrichment analysis (GSEA). a, b Networks of biological pathways and processes between low and high CGRS GC patients in ACRG (a) and TCGA (b) cohorts. Nodes indicate enriched gene sets; similar nodes are grouped and annotated. Node size was proportional to the number of genes within the gene set. The thickness of blue lines was proportional to shared genes between gene sets. Low connective and uninformative sub-networks and nodes

were removed from the network map. c-n Representative GSEA results for cell cycle, apoptosis, cell metabolism, transcription, and immune related pathways in ACRG (c-h) and TCGA (i-n) cohorts.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable11.xlsx
- SupplementaryTable11.xlsx
- SupplementaryTable3.xlsx
- SupplementaryTable3.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable2.xlsx
- JECCRSupplementaryinformation.docx
- JECCRSupplementaryinformation.docx