



## OPEN ACCESS

## EDITED BY

Mehdi Pirooznia,  
Johnson & Johnson, United States

## REVIEWED BY

Gagandeep Gahlay,  
Guru Nanak Dev University, India  
Modjtaba Emadi-Baygi,  
Shahrood University, Iran

## \*CORRESPONDENCE

Alexis Germán Murillo Carrasco,  
✉ agmurillo@usp.br  
César Alexander Ortiz Rojas,  
✉ ortizr@alumni.usp.br

<sup>†</sup>These authors share last authorship

RECEIVED 16 April 2023

ACCEPTED 14 August 2023

PUBLISHED 12 September 2023

## CITATION

Velásquez Sotomayor MB,  
Campos Segura AV, Asurza Montalva RJ,  
Marín-Sánchez O, Murillo Carrasco AG  
and Ortiz Rojas CA (2023), Establishment  
of a 7-gene expression panel to improve  
the prognosis classification of gastric  
cancer patients.

*Front. Genet.* 14:1206609.

doi: 10.3389/fgene.2023.1206609

## COPYRIGHT

© 2023 Velásquez Sotomayor, Campos  
Segura, Asurza Montalva, Marín-Sánchez,  
Murillo Carrasco and Ortiz Rojas. This is  
an open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Establishment of a 7-gene expression panel to improve the prognosis classification of gastric cancer patients

Mariana Belén Velásquez Sotomayor<sup>1,2</sup>,  
Anthony Vladimir Campos Segura<sup>1,3,4</sup>,  
Ricardo José Asurza Montalva<sup>1,2</sup>, Obert Marín-Sánchez<sup>1,5</sup>,  
Alexis Germán Murillo Carrasco<sup>1,6\*†</sup> and  
César Alexander Ortiz Rojas<sup>1,7\*†</sup>

<sup>1</sup>Immunology and Cancer Research Group (IMMUCA), Lima, Peru, <sup>2</sup>Escuela de Medicina Humana, Facultad de Ciencias de la Salud, Universidad Científica del Sur, Lima, Perú, <sup>3</sup>Biochemistry and Molecular Biology Research Laboratory, Faculty of Natural Sciences and Mathematics, Universidad Nacional Federico Villarreal, Lima, Peru, <sup>4</sup>Laboratory of Genomics and Molecular Biology, International Center of Research CIPE, A.C. Camargo Cancer Center, São Paulo, Brazil, <sup>5</sup>Departamento Académico de Microbiología Médica, Facultad de Medicina, Universidad Nacional Mayor de San Marcos, Lima, Peru, <sup>6</sup>Centro de Investigação Translacional em Oncologia (LIM24), Departamento de Radiologia e Oncologia, Faculdade de Medicina da Universidade de São Paulo e Instituto do Câncer do Estado de São Paulo, São Paulo, Brazil, <sup>7</sup>Laboratório de Investigação Médica (LIM) 31, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil

Gastric cancer (GC) ranks fifth in incidence and fourth in mortality worldwide. The high death rate in patients with GC requires new biomarkers for improving survival estimation. In this study, we performed a transcriptome-based analysis of five publicly available cohorts to identify genes consistently associated with prognosis in GC. Based on the ROC curve, patients were categorized into high and low-expression groups for each gene using the best cutoff point. Genes associated with survival ( $AUC > 0.5$ ; univariate and multivariate Cox regressions,  $p < 0.05$ ) were used to model gene expression-based scores by weighted sum using the pooled Cox  $\beta$  regression coefficients. Cox regression ( $p < 0.05$ ),  $AUC > 0.5$ , sensitivity  $> 0.5$ , and specificity  $> 0.5$  were considered to identify the best scores. Gene set enrichment analysis (KEGG, REACTOME, and Gene Ontology databases), as well as microenvironment composition and stromal cell signatures prediction (CIBERSORT, EPIC, xCell, MCP-counter, and quanTseq web tools) were performed. We found 11 genes related to GC survival in the five independent cohorts. Then, we modeled scores by calculating all possible combinations between these genes. Among the 2,047 scores, we identified a panel based on the expression of seven genes. It was named GES7 and is composed of *CCDC91*, *DYNC1I1*, *FAM83D*, *LBH*, *SLTRK5*, *WTIP*, and *NAP1L3* genes. GES7 features were validated in two independent external cohorts. Next, GES7 was found to recategorize patients from AJCC TNM stages into a best-fitted prognostic group. The GES7 was associated with activation of the TGF- $\beta$  pathway and repression of anticancer immune cells. Finally, we compared the GES7 with 30 previous proposed scores, finding that GES7 is one of the most robust scores. As a result, the GES7 is a reliable gene-expression-based signature to improve the prognosis estimation in GC.

## KEYWORDS

prognosis, gastric cancer, score, risk classification, gene expression

## 1 Introduction

Gastric cancer (GC) is responsible for over one million new cases in 2020 and an estimated 769,000 deaths, ranking fifth for incidence and fourth for mortality globally (Sung et al., 2021). Gastric adenocarcinoma is the more common subtype, comprising more than 95% of cases, and is a highly heterogeneous group concerning anatomical location, histological subtypes, and molecular features (Kupfer, 2017; Yan et al., 2022). Although a variety of new cancer treatments have been introduced that have led to improvement of life expectancy in GC, the standard treatment continues to be surgical resection, chemotherapy, and/or radiotherapy, with five-year survival rates from 90% to 98% for pT1 stage tumors and from 38% to 59% for pT2–4 tumors (Yan et al., 2022).

High-throughput sequencing efforts have been made to characterize the genomic landscape of gastric cancer, allowing better categorization of the heterogeneity among patients. The Cancer Genome Atlas (TCGA) developed a robust molecular classification and identified dysregulated pathways and candidate drivers of distinct classes of gastric cancer (Bass et al., 2014). TCGA classification divides GC into four subtypes: tumors positive for Epstein–Barr virus (EBV); microsatellite instability (MSI); genetically stable tumors (GS); and tumors with chromosomal instability (CIN) (Bass et al., 2014). Similarly, the Asian Cancer Research Group (ACRG) established clinically relevant molecular subtypes based on gene expression analysis and mutations (Cristescu et al., 2015). ACRG classification describes four molecular subtypes linked to disease progression and prognosis: microsatellite instability (MSI), microsatellite instability (MSS)/Epithelial–Mesenchymal Transition (EMT), MSS/TP53 mutated, and MSS/TP53 wild-type (Cristescu et al., 2015).

Although these classifications could be valuable for clinical decisions, the American Joint Committee on Cancer (AJCC) TNM staging (Liu et al., 2018) remains the primary prognostic determinant in GC. Since the disease can sometimes progress unexpectedly to what is established by the TNM classification, the clinical application of biological and molecular markers is pending as a complement to AJCC TNM staging to improve the prognosis. A possible reason for this pending is the need for more consensus on how efficiently to include in the laboratory routine the determination of *TP53* mutations (Malcikova et al., 2018), CIN (McGranahan et al., 2012), and EMT (Cristescu et al., 2015).

Some studies have proposed gastric tumor biomarkers from tissue or blood samples of GC patients through different analytic strategies (Wang et al., 2016; Wang et al., 2017; Cheong et al., 2018; Shi et al., 2018; Zhu et al., 2020; Li et al., 2021) with clinically relevant results. Although their relevant contribution to the field, some of these panels face some limitations, like the high number of component genes, the low number of cohorts evaluated (generally 1–3 datasets), and the geographic origin of these cohorts. In the present study, we aim to improve the current GC biomarker context by using a robust algorithm to evaluate all available genes in over one thousand samples from five different cohorts. After detecting the most relevant genes for assessing the

prognosis of GC patients in the five cohorts, we performed mathematical modeling to get a better gene-expression panel. Thus, we found a score based on the expression of seven genes, herein referred to as GES7. This score was validated in two independent cohorts and demonstrated to be robust in comparison with previous proposed scores/indexes.

## 2 Materials and methods

### 2.1 Patient cohorts and gene expression profiling

This study analyzed publicly available clinical and transcriptomic data of five GC cohorts, as discovery datasets. First, The Cancer Genome Atlas (TCGA) gastric cancer cohort ( $n = 345$ ) (Bass et al., 2014) with RNA-seq data was included (Table 1). Transcriptome data of the TCGA cohort were generated using HiSeq 2000 (Illumina) and included information on 20,508 genes. Clinical and genetic information were retrieved from the Firebrowse data portal site (<http://firebrowse.org/>). Next, four transcriptome datasets based on the microarray were included: Asian Cancer Research Group (ACRG, [GSE66229](#) dataset,  $n = 297$ ) (Cristescu et al., 2015), Yonsei University Severance Hospital (YUSH, [GSE84437](#) dataset,  $n = 433$ ) (Kim et al., 2019), Korea University Guro Hospital (KUGH, [GSE26899](#) data set,  $n = 93$ ) (Oh et al., 2018) and the National Cancer Centre of Singapore (NCCS, [GSE15459](#) data set,  $n = 192$ ) (Ooi et al., 2009) (Table 1). Microarray data were retrieved from the Gene Expression Omnibus (GEO) database ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)). Data from ACRG and NCCS cohorts were generated using Human Genome U133 Plus 2.0 (Affymetrix) which contained probes for 23,520 genes. In comparison, YUSH and KUGH cohorts were generated using HumanHT-12 V3.0 (Illumina) which contained probes for 25,124 genes. As external and independent validation datasets, we used additional tumor specimens collected from the Kosin University College of Medicine (KUCM, [GSE26901](#) dataset,  $n = 109$ ) and The University of Texas MD Anderson Cancer Center (MDACC, [GSE28541](#) data set,  $n = 40$ ), described in a previous study (Oh et al., 2018).

### 2.2 Identification of genes associated with prognosis

After collecting gene expression data, we used survival receiver operating characteristic (ROC) curves to calculate the optimal gene expression cutoff value for each gene stratifying patients into high and low-expression level groups. About the sample size of high/low expression groups, we considered a cutoff value to preserve at least 20% of the total sample size in these groups avoiding possible overfitting, groups with small sample sizes, and biased selection of cutoff value. If multiple cutoffs were retrieved for each gene, the cutoff value closer to the median was selected. Overall survival (OS) events were used in this calculation. Then, these groups were

**TABLE 1** Clinical and biological information of gastric cancer cohorts.

Level		TCGA	ACRG	YUSH	KUGH	NCCS
n		345	297	433	93	192
Age [median (IQR)]		67.00 [58.00, 72.00]	64.00 [55.00, 70.00]	62.00 [53.00, 68.00]	60.00 [50.00, 69.00]	66.55 [56.80, 73.00]
Sex, n (%)	Female	122 (35.4)	101 (34.0)	137 (31.6)	20 (21.5)	67 (34.9)
	Male	223 (64.6)	196 (66.0)	296 (68.4)	73 (78.5)	125 (65.1)
Race category, n (%)	Asian	73 (21.2)	—	—	—	—
	Black or African American	9 (2.6)	—	—	—	—
	White	218 (63.2)	—	—	—	—
	Unknown	45 (13)	—	—	—	—
Histologic grade, n (%)	G1	6 (1.7)	—	—	—	—
	G2	116 (33.6)	—	—	—	—
	G3	215 (62.3)	—	—	—	—
	GX	8 (2.3)	—	—	—	—
Tumor weight (median [IQR])		300.00 [135.50, 821.50]	—	—	—	—
Tumor stage T, n (%)	T1/T2	88 (25.5)	185 (62.3)	49 (11.3)	—	—
	T3/T4	257 (74.5)	112 (37.7)	384 (88.7)	—	—
Nodal status, n (%)	N0	107 (31.0)	38 (12.8)	80 (18.5)	—	—
	N1/N2/N3	238 (69.0)	259 (87.2)	353 (81.5)	—	—
Metastasis status, n (%)	M0	320 (92.8)	270 (90.9)	—	85 (92.4)	—
	M1	25 (7.2)	27 (9.1)	—	7 (7.6)	—
AJCC TNM staging, n (%)	I	—	—	—	11 (12.0)	31 (16.1)
	IA	9 (2.6)	0 (0)	—	—	—
	IB	3 (0.9)	31 (10.4)	—	—	—
	II	54 (15.7)	94 (31.6)	—	18 (19.6)	29 (15.1)
	III	—	—	—	27 (29.3)	72 (37.5)
	IIIA	188 (54.5)	69 (23.2)	—	—	—
	IIIB	31 (9.0)	26 (8.8)	—	—	—
	IV	60 (17.4)	77 (25.9)	—	36 (39.1)	60 (31.2)
Lauren classification, n (%)	Intestinal	152 (44.1)	144 (48.5)	—	59 (64.1)	75 (39.1)
	Diffuse	73 (21.2)	134 (45.1)	—	31 (33.7)	99 (51.6)
	Mixed	—	19 (6.4)	—	2 (2.2)	18 (9.4)
	Unknown	120 (34.8)	—	—	—	—
Primary tumor site, n (%)	Fundus/Body	123 (35.7)	—	—	31 (33.3)	—
	Cardia/Proximal	45 (13.0)	—	—	7 (7.5)	—
	Antrum/Distal	131 (38.0)	—	—	55 (59.1)	—
	Gastroesophageal junction	35 (10.1)	—	—	—	—
	Unknown	11 (3.2)	—	—	—	—
	Body	—	107 (36.0)	—	—	—

(Continued on following page)

**TABLE 1 (Continued)** Clinical and biological information of gastric cancer cohorts.

Level		TCGA	ACRG	YUSH	KUGH	NCCS
	Cardia or cardia + body	—	31 (10.4)	—	—	—
	Antrum	—	149 (50.2)	—	—	—
	Body + antrum or fundus to antrum	—	6 (2.0)	—	—	—
	Entire stomach	—	4 (1.3)	—	—	—
TCGA molecular subtype, n (%)	MSI	66 (19.1)	—	—	—	—
	EBV	27 (7.8)	—	—	—	—
	CIN	198 (57.4)	—	—	—	—
	POLE	7 (2.0)	—	—	—	—
	GS	47 (13.6)	—	—	—	—
ACRG molecular subtype, n (%)	MSI	—	68 (22.9)	—	—	—
	TP53 mut	—	77 (25.9)	—	—	—
	TP53 wt	—	106 (35.7)	—	—	—
	EMT	—	46 (15.5)	—	—	—
NCCS molecular subtype, n (%)	Invasive	—	—	—	—	51 (26.6)
	Metabolic	—	—	—	—	40 (20.8)
	Proliferative	—	—	—	—	70 (36.5)
	Unstable	—	—	—	—	31 (16.1)

compared by univariate and multivariate Cox proportional hazard regression for each gene in each cohort. Possible confounding variables were included in the multivariate analysis depending on data availability. Thus, we adjusted prognosis prediction with the following potential confounders: age (as continuous), sex, Lauren classification, T stage, N and M status, primary tumor site, and molecular subtype. We considered that a gene is associated with prognosis if the area under the survival ROC curve (AUC) was >0.5, and a *p*-value < 0.05 for Cox regressions, in the five cohorts in an independent analysis for each cohort.

### 2.3 Modeling of gene-expression-based index

To improve the prognosis prediction of the gene signatures, we modeled gene expression-based scores using the genes identified as described in the previous section. The scores were calculated by the weighted sum of gene expression values, using the  $\beta$  coefficient of Cox regression, according to the next formula:

$$\text{Score} = \text{Gene}_1\beta_1 + \text{Gene}_2\beta_2 + \text{Gene}_3\beta_3 + \dots + \text{Gene}_i\beta_i$$

The  $\beta$  value corresponds to pooled  $\beta$  coefficient calculated from the five cohorts by gene and estimated by a fixed- and random-effects model according to the heterogeneity between studies. Heterogeneity was assessed using Higgins'  $I^2$  statistic and

Cochran's Q-test (Higgins and Thompson, 2002). A score was considered highly predictive when univariate HR > 2 (*p* < 0.05), multivariate HR > 1.5 (*p* < 0.05), AUC > 0.5, sensitivity > 0.5, and specificity > 0.5, in the five cohorts in an independent analysis for each cohort. Then, we selected the best gene expression-based score after considering the HR values and the number of genes constituting the panel. Finally, to homogeneously apply the score to any cohort, we normalized the score by dividing the score values by the median of the score calculated for each discovery dataset. Then, we established three groups according to the score values: low (score < 0.9), intermediate (0.9 ≤ score < 1.1), and high (score ≥ 1.1) groups.

### 2.4 Biological pathways enrichment

We performed the Gene Set Enrichment Analysis (GSEA) using the Broad Institute software (<http://software.broadinstitute.org/gsea/index.jsp>) to find biological processes associated with the score. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Reactome databases were included in our analysis. Enrichment scores were calculated based on Kolmogorov-Smirnov statistics tested for significance using 1,000 permutations. Additionally, Pearson correlation was used as a metric for ranking genes. A pathway was considered enriched when a nominal *p*-value and FDR *q*-value were <0.05.

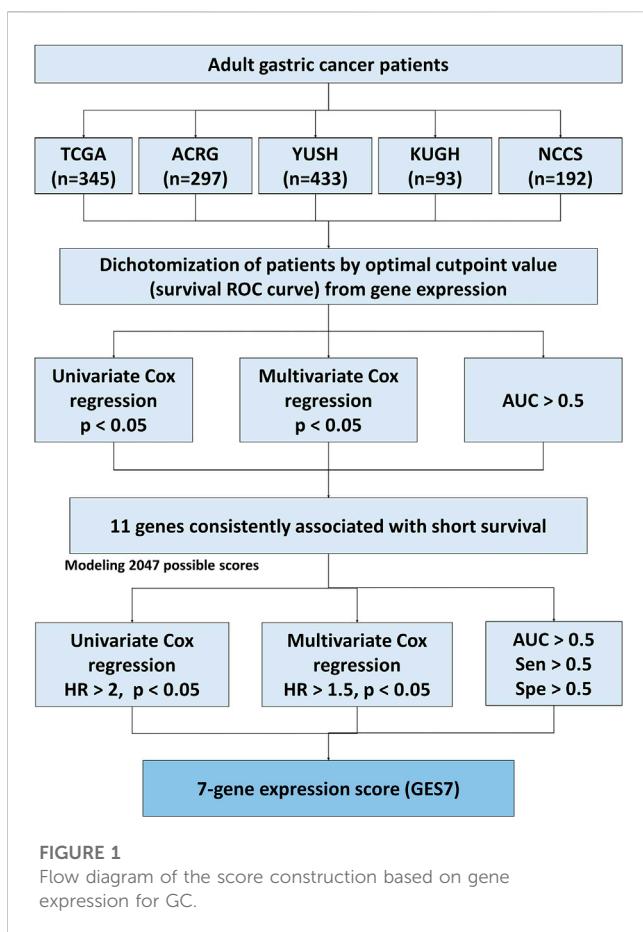


FIGURE 1

Flow diagram of the score construction based on gene expression for GC.

## 2.5 Enrichment of cells signatures and drug sensitivity evaluation

The immune stromal cell signature of tumor samples was evaluated by using the following algorithms: CIBERSORT (Newman et al., 2015), EPIC (Racle and Gfeller, 2020), xCell (Aran et al., 2017), MCP-counter (Becht et al., 2016), and quanTIseq (Finotello et al., 2019); from the TIMER web tool (<http://timer.cistrome.org/>). Next, we explored the drug response in gastric cancer cells with different score values. The gene expression score was calculated in the following gastric cancer cell lines: AGS, FU97, GCIY, HGC-27, Hs-746T, IM-95, MKN45, MKN7, NCI-N87, NUGC-3, RERF-GC-1B, SNU-1, SNU-16, SNU-5, TBC11TKB, and 23132-87. Gene expression data were retrieved from the Expression Atlas database ([www.ebi.ac.uk/gxa/experiments/E-MTAB-2770/](http://www.ebi.ac.uk/gxa/experiments/E-MTAB-2770/)). Then, we correlate the scores with the inhibitory concentration 50 (IC<sub>50</sub>) for 448 drugs (<https://www.cancerrxgene.org/>). Spearman rho coefficients >0.25 or < -0.25 with  $p < 0.05$  were considered to identify drug responses.

## 2.6 Statistical analysis

The OS was defined as the time from diagnosis to death for any cause; those alive or lost to follow-up were censored. Relapse-free survival (RFS) or disease-free survival (DFS) was defined as the time from the achievement of response to the first adverse event. That is relapse or death from any cause, whichever occurs first.

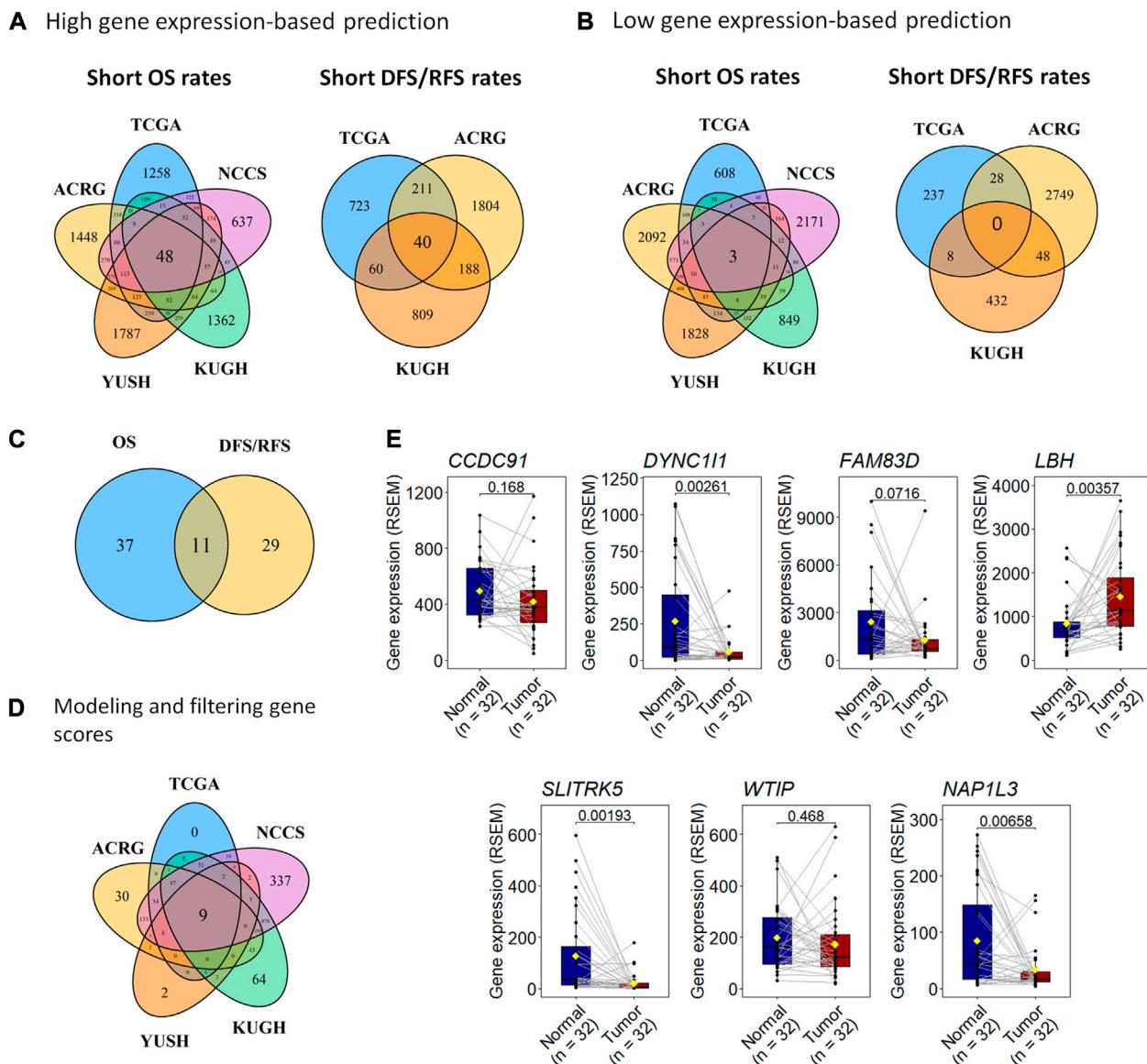
Kaplan-Meier (KM) plots were performed to demonstrate the power of stratification. The Mann-Whitney test and Fisher's exact test were used to compare clinical and biological characteristics between patients. All  $p$ -values were two-sided with a significance level of 0.05. All calculations were performed using R 4.1.1 (The CRAN project, [www.r-project.org](http://www.r-project.org)) software.

## 3 Results

### 3.1 Survival-related genes identified a 7-gene expression-based score applied to risk stratification in gastric cancer

To identify survival-related genes in gastric cancer (GC), we analyzed the transcriptome data of five GC cohorts (Table 1). We dichotomized cohorts according to gene expression using the survival ROC curve. Then, for prognosis evaluation, high and low expression categories were established for each gene in each cohort (Figure 1). We identified 48 genes whose high expression was associated with short overall survival rates. Furthermore, 40 genes were consistently associated with disease-free and relapse-free survival in TCGA, ACRG, and KUGH cohorts (Figure 2A). On the other hand, low expression of three genes was associated with short OS rates, although none of these could predict DFS or RFS (Figure 2B); consequently, these genes were not considered in further analysis. By merging these results, we identified 11 genes associated with prognosis (Figure 2C). Supplementary Tables S1–S5 show the association between gene expression and overall survival for all genes included in the platforms for each cohort. Next, based on the 11 genes we found, we modeled 2,047 gene expression scores, which refer to all unique combinations among these 11 genes. To uniformize our score calculations, we pooled the  $\beta$  coefficients of the five cohorts for each gene (Supplementary Figures S1–S3). After applying Cox regression and AUC discrimination, 9 scores showed an improved power of survival prediction in the five cohorts (Figure 2D; Supplementary Figure S4). Next, we selected the most parsimonious score, based on seven genes, here referred to as GES7. The GES7 is composed of CCDC91, DYNC1I1, FAM83D, LBH, SLTRK5, WTIP, and NAPIL3. Expression levels of these genes and the comparisons with their normal counterparts are shown in Figure 2E. Also, Kaplan-Meier plots showing the stratification power of these genes are shown in Supplementary Figures S5–S8 (Logrank,  $p < 0.05$ ).

The GES7 was highly efficient in stratifying overall survival in GC patients (Figure 3A). Thus, GES7<sup>high</sup> patients had inferior survival rates, with median OS of 19.9 vs. 69 months for TCGA, 27.5 months vs. not reached for ACRG, 44 months vs. not reached for YUSH, 22.1 months vs. not reached for KUGH, and 20.3 months vs. not reached for NCCS. In all cases, the AUC values of the GES7<sup>high</sup> patients were higher than 0.5 (Figure 3B). Consistent with these results, we found that the GES7 score covariate with pointwise estimates of the overall survival hazard ratios (HRs) (Figure 3C). Also, GES7<sup>high</sup> patients were more likely to relapse according to the disease- and relapse-free survival analysis (Figure 3D). Furthermore, after including potential confounding variables in our Cox regression analysis, GES7 remained as a relevant prognostic marker with HRs of 1.75 ( $p = 0.0061$ ), 1.88 ( $p = 0.0014$ ), 2.01 ( $p < 0.0001$ ), 3.8 ( $p = 0.001$ ), and



**FIGURE 2**

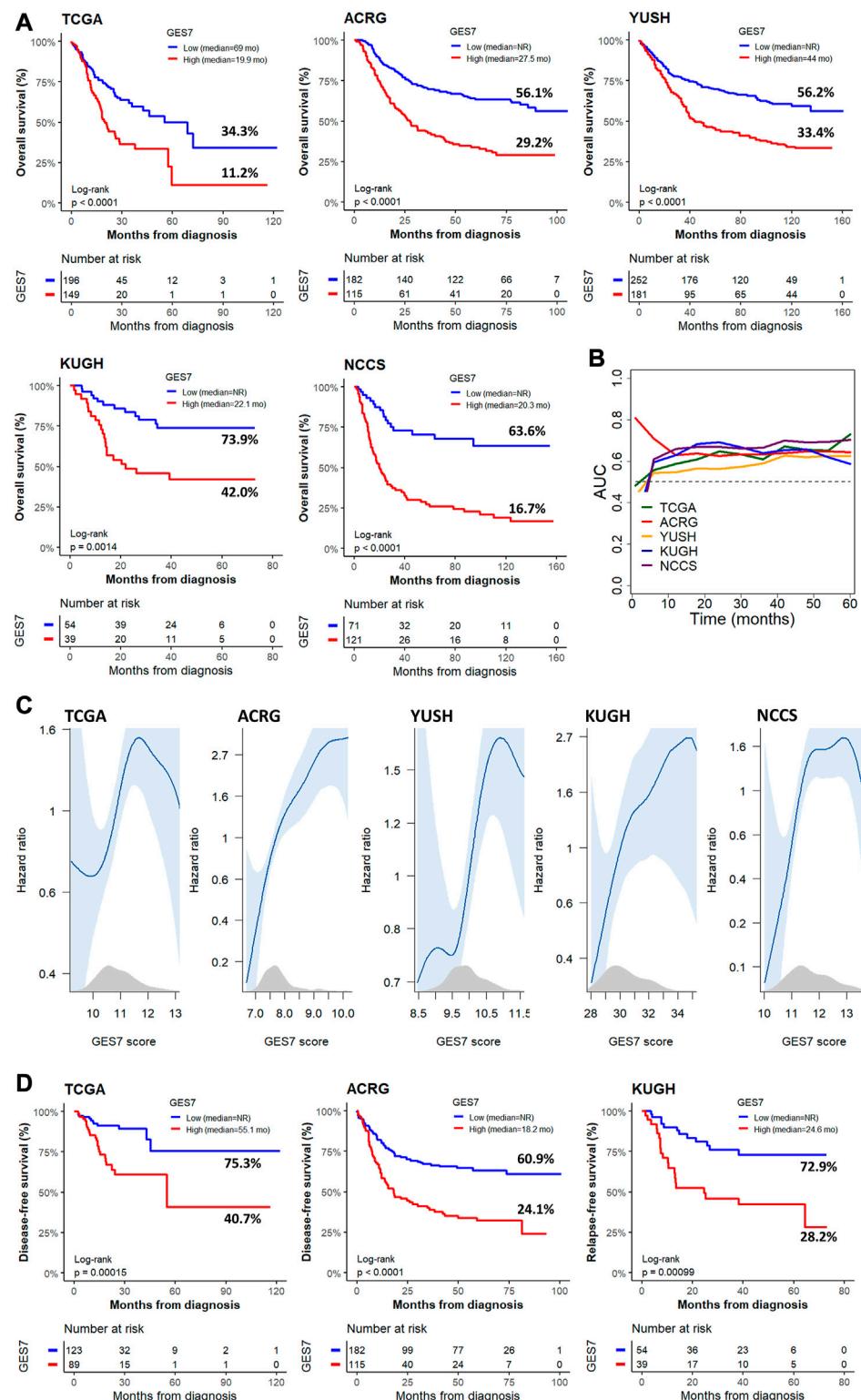
A 7-gene expression score, GES7, can predict prognosis in gastric cancer. **(A)** We used five transcriptome datasets to find 48 genes with high expression associated with unfavorable OS and 40 genes related to unfavorable DFS (TCGA and ACRG) or RFS (KUGH). **(B)** Also, the high expression of three genes was associated with favorable OS, but neither predicts DFS (TCGA and ACRG) or RFS (KUGH). **(C)** By merging these results, 11 genes were associated with prognosis in all datasets. **(D)** Nine scores based on gene expression were highly predictive of prognosis, where a 7-gene expression panel (GES7) was the most parsimonious. **(E)** Comparison of gene expression in tumor versus healthy tissues of genes composing GES7.

3.59 ( $p < 0.0001$ ), for TCGA, ACRG, YUSH, KUGH, and NCCS, respectively (Figures 4A–E). Although GES7<sup>low</sup> patients included more MSI cases and GES7<sup>high</sup> patients showed more CIN, GS, and EMT cases (Supplementary Tables S6, S7), our multivariate regression demonstrates that the molecular subtypes are not a confounder variable. Interestingly, we found that GES7<sup>high</sup> was associated with an invasive phenotype as described for the NCCS cohort (Supplementary Table 10). No other clinical variable was consistently associated with the GES7 (Supplementary Tables S6–S10). Finally, we validated our GES7 in two independent cohorts, KUCM and MDACC. We found that GES7<sup>high</sup> patients had inferior survival rates, with median OS of 17.5 months vs. not reached (log-rank  $p < 0.0001$ ), and 9.3 vs. 24.6 months (log-rank  $p = 0.0013$ ), respectively (Figure 4F). Also, ROC curve analysis demonstrated

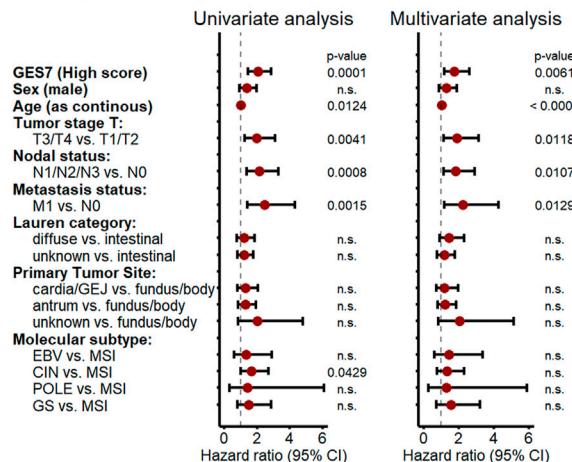
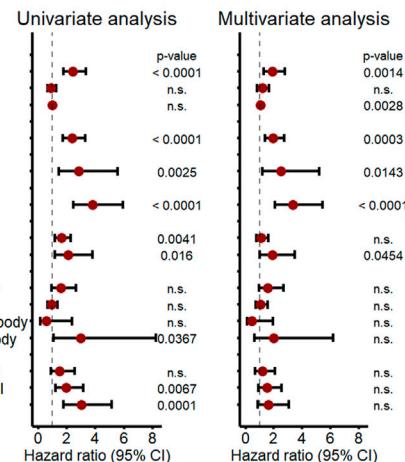
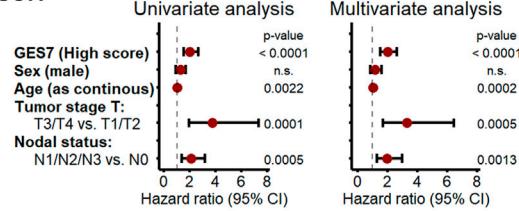
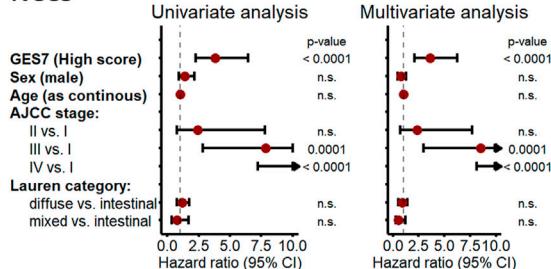
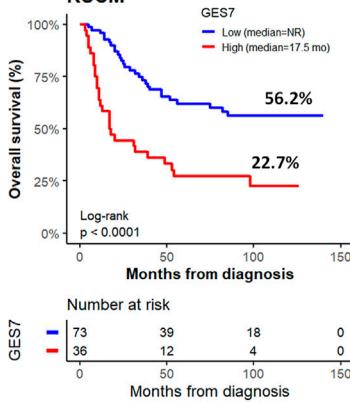
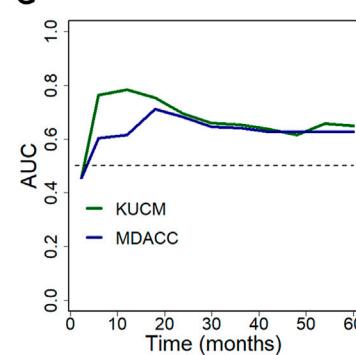
that the area under the curve values remained consistently high, exceeding 0.6, for up to 5 years of follow-up in both validation cohorts (Figure 4G). Therefore, the GES7 gene signature proved to be highly efficient in stratifying overall survival in gastric cancer patients across multiple cohorts.

### 3.2 GES7 is a promissory tool for recategorizing AJCC TNM subgroups

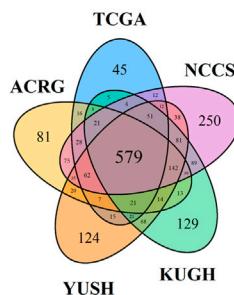
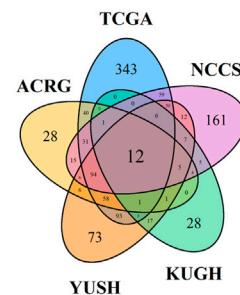
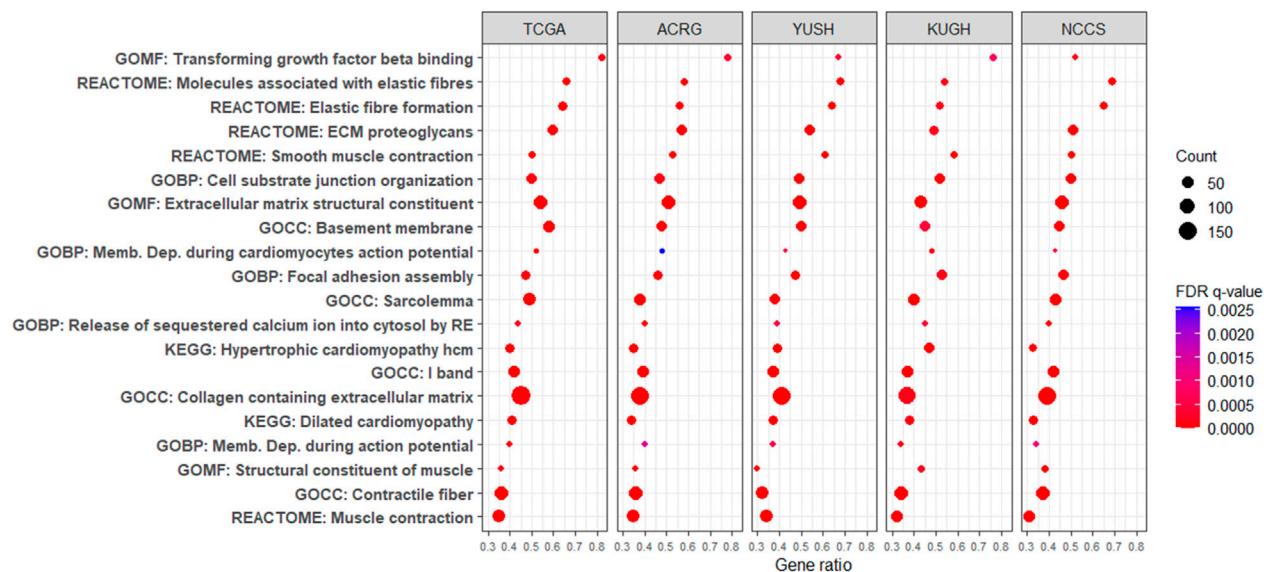
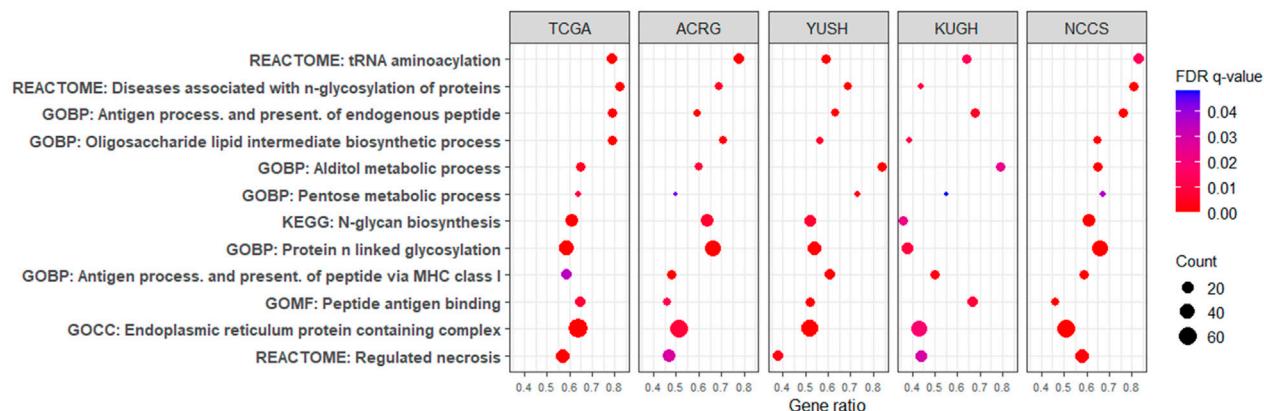
Next, we combined the traditional anatomical staging system with the molecular signature of GES7. For this purpose, we normalized the GES7 values across the five cohorts to find

**FIGURE 3**

Survival of GC patients according to the GES7 values. **(A)** Kaplan-Meier plots show the survival stratification power of GES7 for OS. **(B)** Survival AUC analysis of GES7 until 5 years of follow-up. **(C)** Smooth plot showing the relation between the risk of death and the GES7 values. **(D)** Kaplan-Meier plots show the survival stratification power of GES7 for DFS and RFS.

**A TCGA****B ACRG****C YUSH****E NCCS****F****G****FIGURE 4**

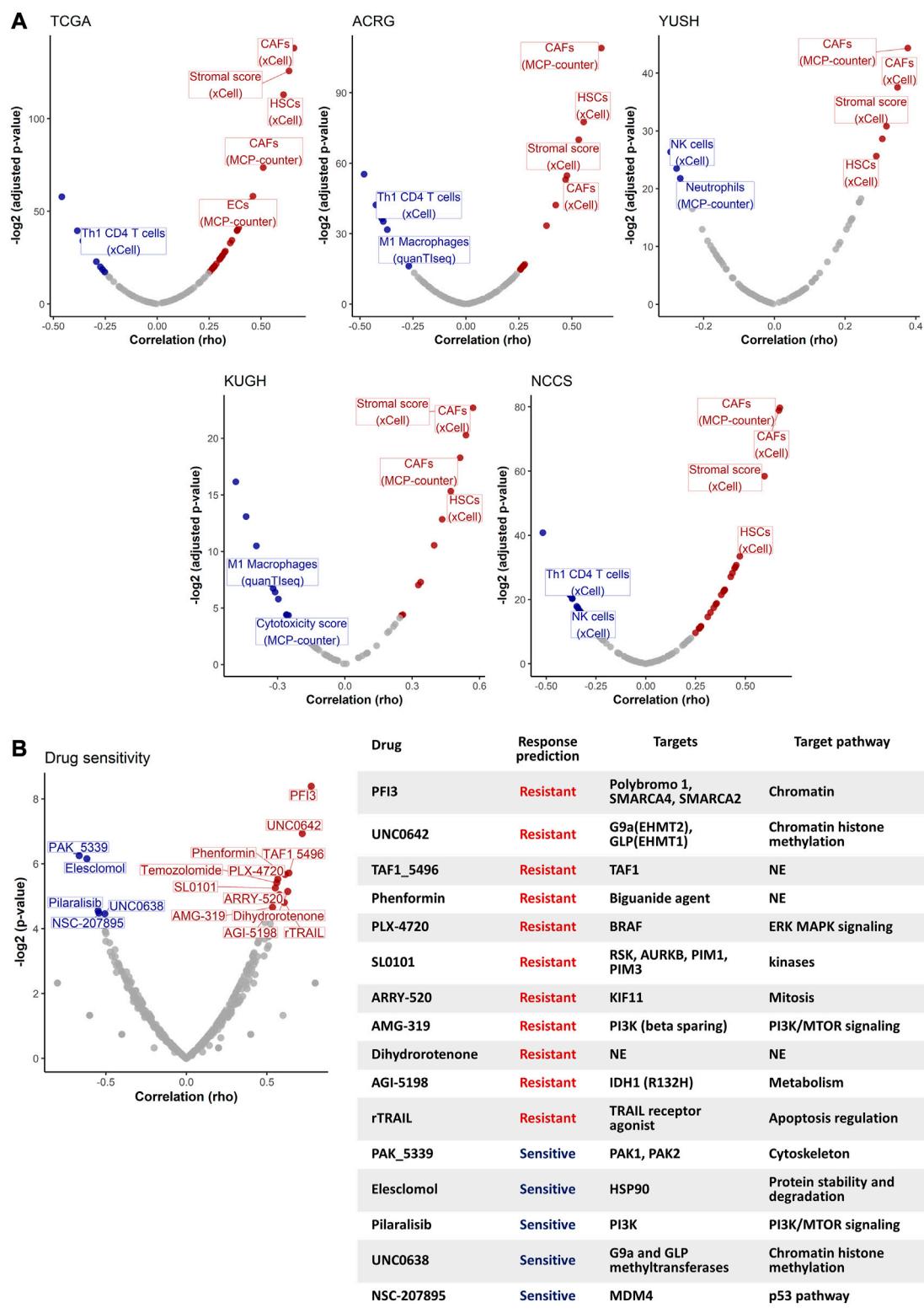
Cox regression for overall survival according to GES7 and validation analysis. Univariate and multivariate analyses were performed for the TCGA (A), ACRG (B), YUSH (C), KUGH (D), and NCCS (E) cohorts, showing that GES7 is an independent prognostic factor in gastric cancer. Also, Kaplan-Meier plots (F) and ROC curve analysis (G) in KUCM and MDACC validation datasets confirm the clinical applicability of GES7.

**A Positively enriched pathways****B Negatively enriched pathways****C Positively enriched pathways****D Negatively enriched pathways****FIGURE 5**

Enrichment analysis of biological pathways according to GES7. (A) Number of positive and (B) negative correlated paths with GES7. (C) Gene set enrichment analysis demonstrated an invasive and motility phenotype in tumors with high GES7, while (D) antigen-presenting processes were diminished.

widely applicable cutoffs despite the techniques used to measure gene expression (RNA-seq or microarray) (Supplementary Figure S9). Next, we established three categories, GES7 low, intermediate, and high, applying cut-off points of 0.9 and

1.1 for all cohorts. By considering these categories, we proposed a risk classification system that integrated the AJCC TNM staging and the GES7 (Supplementary Table 11). As seen in Supplementary Figure 10A, the three categories of

**FIGURE 6**

Non-tumor cell type abundance and drug sensitivity analysis. (A) Volcano plots show the correlation between GES7 values and non-tumor cell type abundance. (B) Drug sensitivity analysis of GC cell lines. The volcano plot (left) shows the correlation between the GES7 and the IC<sub>50</sub> for each drug. The table (right) shows the description for each drug that was correlated with GES7.

GES7 identified patients with different survival rates in the ACRG cohort. As expected, the AJCC TNM staging (7th edition) also efficiently distributed patients into other prognostic groups (Supplementary Figure 10B). After integrating both systems, we observed a reclassification of patients in a best-fitted risk group (Supplementary Figures 10C–D). Considering the IIIA category as intermediate for prognosis (median OS<sub>IIIA</sub> = 68.1 months) and IIIB and IV categories as adverse (median OS<sub>IIIB</sub> = 28.2 months and median OS<sub>IV</sub> = 17.4 months), we had that the GES7 restructured these groups by the formation of an adverse-risk group constituting of patients from IIIA, IIIB and IV (median OS<sub>adverse</sub> = 25.2 months), leaving the remaining patients of IV group forming the very-adverse risk group (median OS<sub>adverse</sub> = 12.7 months) (Supplementary Figures 10D, E). These results were validated in the TCGA cohort, where the adverse and very-adverse risk groups differed from the intermediate and favorable groups compared to the AJCC TNM system alone (Supplementary Figures 11A–E). Therefore, the integration of our gene signature with the classic anatomical staging system improves prognostication accuracy and provides clinicians with a more comprehensive assessment of the patient's disease.

### 3.3 GES7 predicts stromal cells-enriched microenvironment

To explore GES7-related biological pathways, we performed a GSEA with the transcriptome from the five discovery datasets. Figures 5A, B show that 579 pathways were consistently enriched in GES7 high samples, while 12 were for GES7<sup>low</sup>. The top enriched pathway was the transforming growth factor beta (TGF-β) activation (Figure 5C). In line with this result, pathways related to elastic fiber were enriched, indicating a motility phenotype. On the other hand, antigen processing and presentation-related pathways were negatively enriched, indicating low activity of the innate immune system in these tumors (Figure 5D). Then, we tested the hypothesis of non-tumor cell infiltrations being associated with GES7. We found that GES7<sup>high</sup> tumors are enriched with stromal cells like cancer-associated fibroblasts and have a reduced presence of innate immune cells (Figure 6A). Finally, we found that GC cell lines with high GES7 respond well to *in vitro* treatment with PAK\_5339. This drug inhibits the p21-activated kinases 1 (PAK1) and 2 (PAK2), Elesclomol (anti-HSP90), Pilaralisib (anti-PI3K), UNC0628 (anti-G9a and GLP methyltransferases), and NSC-207895 (anti-MDM4) (Figure 6B), that could be valuable treatment strategies for GES7<sup>high</sup> patients.

### 3.4 Comparison of scores/indexes reveals GES7 as a robust biomarker of prognosis

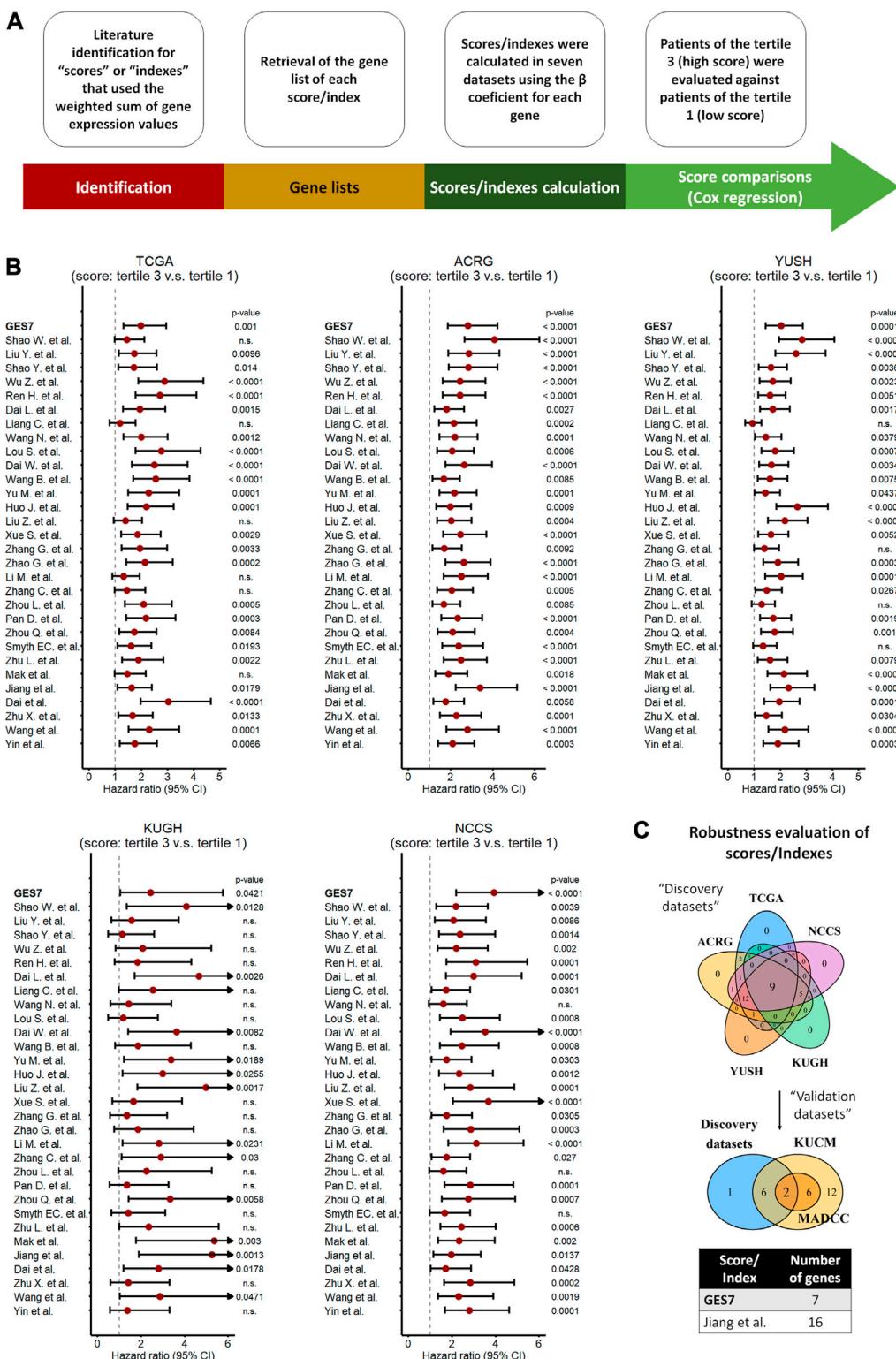
Previous gene expression-based prognostic indexes have been proposed as predictors of survival in GC. Here, we compared the GES7 with 30 prognostic scores published between 2013 and 2023 (Supplementary Table 12). After retrieving the gene list of each

score, we calculated them using the weighted sum of gene expression values, the base strategy of all the studies. To homogenize the comparison between scores, we divided the cohorts into tertiles, and considered tertile 3 as the high score group, while tertile 1 as the low score group (Figure 7A). After applying the scores to our discovery datasets by univariate Cox regression analysis (Figure 7B), we found that 9 of 31 indexes were able to discriminate between patients of different risk groups. Finally, when we added the evaluation of our validation datasets, only two scores, within which is the GES7, were shown as the robust indexes (Figure 7C; Supplementary Figure 12).

## 4 Discussion

Gene expression signatures are a valuable tool for risk categorization in GC (Chen et al., 2005; Shi et al., 2018). In this study, our primary objective was to develop a robust and interpretable genetic index for predicting prognosis in gastric cancer. We opted for a methodology that is straightforward, easy to interpret, and based on widely accepted statistical techniques. We utilized survival ROC curves to determine the optimal gene expression cutoff value for each gene, which allows us to dichotomize patients into high and low-expression level groups based on prognostic relevance. Furthermore, we performed univariate and multivariate Cox proportional hazard regression analyses to identify genes significantly associated with prognosis in each cohort while adjusting for potential confounders. The multivariate analysis allowed us to consider the influence of several relevant clinical variables, enhancing the clinical applicability of our findings. Regarding the modeling of gene expression-based scores, we calculated the scores using the weighted sum of gene expression values based on the β coefficient from Cox regression. This approach allows us to assign appropriate weights to individual genes based on their prognostic significance, making the index a reflection of gene contributions to overall survival outcomes. As result of this process, a gene expression signature based on seven genes (GES7) was established. For four of them, WTIP, NAP1L3, CCDC91, and SLTRK5, this is the first study in which an association between their expression levels and the prognosis of GC was found. Interestingly, studies in other cancers had already identified WTIP, NAP1L3, and CCDC91, as relevant genes for tumor progression (Zeng et al., 2016; Heshmati et al., 2018; Wu et al., 2019; Lv et al., 2021). In contrast, mutations in SLTRK5 have been described as predictors of prolonged survival in GC patients with TP53 wild-type status (Park et al., 2016). In contrast, the high expression of DYNC1I1, FAM83D, and LBH, have been previously associated with poor prognosis and an aggressive tumor phenotype in GC (Huang et al., 2017; Deng et al., 2018; Chonov et al., 2019; Gong et al., 2019; Yu et al., 2019; Zhang et al., 2019).

Considering the expression status of these seven genes, we found that GES7<sup>high</sup> tumors are characterized by the activation of the transforming growth factor beta (TGF-β), an essential fibrogenic agent that can promote an invasive phenotype, immunosuppression, and an interplay between tumor and stromal cells (Balkwill et al., 2012; Colak and ten Dijke, 2017; Henke et al., 2020). These results

**FIGURE 7**

Comparison of scores/indexes based on gene expression to evaluate prognosis in GC. (A) Strategy used to calculate and apply the scores to our datasets. (B) Univariate Cox regression analysis of each score in our discovery datasets. (C) Merged results between discovery and validation datasets indicating the most robust scores, which included the GES7.

align with the association between GES7<sup>high</sup> and the high number of EMT cases (ACRG dataset) or invasive (NCCS) subtypes. Also, a high stromal cell abundance was associated with GES7<sup>high</sup> after applying CIBERSORT-based algorithms. In addition, the endogenous antigen presentation process pathway, M1-macrophages, and Th1 CD4 T cell abundance were negatively correlated with the GES7. Usually, antigens derived from tumor-specific mutations should be displayed on the surface of cells by MHC-I, processed by cells like macrophages, and presented to T lymphocytes. However, tumor cells can exploit multiple escape mechanisms to evade immune recognition (Jhunjhunwala et al., 2021). Thus, our GES7 score has the potential to identify tumors that evade the immune system. Other suppressed pathways in GES7<sup>high</sup> tumors were the DNA replication and cell cycle. The suppression of DNA repair induces an accumulation of somatic mutation, which is usually related to a worse prognosis (Turgeon et al., 2018), consistent with our results.

Next, we found that GES7<sup>high</sup> GC cells could respond to treatment with PAK\_5339, Elesclomol, Pilaralisib, UNC0638, and NSC-207895. Among therapeutic proposals with current tests in cancer patients, we have PAK\_5339, which is a drug that inhibits PAK1 and PAK2, proteins related to the activation of cell survival by different mechanisms, such as the metabolism of Bcl-2 in the mitochondria, the modulation of the cell motility, and the activation of the PI3K-AKT/mTOR pathway (Yao et al., 2020; Liu et al., 2021). Consequently, these PAK-promoted pathways could be targets of putative drugs for treating GES7<sup>high</sup> patients. On the other hand, Elesclomol, an inducer of oxidative stress (Qu et al., 2009), and Pilaralisib, an inhibitor of PI3K (Hashemzadeh et al., 2018), have been tested in combination with conventional treatments in other types of cancers (Qu et al., 2009; Tolane et al., 2014; Bechter et al., 2016; Hashemzadeh et al., 2018; Buccarelli et al., 2021). Therefore, our results and the previous evaluation of these drugs in brain and breast tumors warrant for further experiments.

Several previous gene expression-based studies have proposed biomarkers for GC (Szász et al., 2016; Wang et al., 2016; Wang et al., 2017; Cheong et al., 2018; Shi et al., 2018; Li et al., 2020; Liu et al., 2020; Zhu et al., 2020; Li et al., 2021; Lou et al., 2021; Zhang et al., 2021; Chang et al., 2022; Rothwell et al., 2022). Here, we tested 30 gene expression-based scores in the seven cohorts used in this study (Figure 7; Supplementary Table 12), and the results were compared with our GES7. Only two scores, which include the GES7, were the robust for overall survival prediction, highlighting its potential clinical use. Possible explanations can be raised up for this finding. First, here we tested our GES7 in cohorts from different centers around the world, which means that our score is widely applicable independent of ethnicity. Secondly, we performed an independent analysis in seven GC datasets, which describes our score as robust. Thirdly, we used genes that predict deaths and relapse events by themselves. In addition, we found an improvement in the TNM risk stratification. The TNM system has been criticized for its poor stratification ability (Dikken et al., 2012; Marrelli et al., 2012; Yoon et al., 2012; Shu et al., 2017). Although in the most recent update (8th edition), some modifications were made, such as the division of the

pN3 stage into pN3a and pN3b (Amin et al., 2017; Lu et al., 2017; Ji et al., 2018), one study concluded that this latest version has similarities to the previous one in terms of its predictive capabilities (Abdel-Rahman, 2017). Then, the inclusion of our GES7 could be a multigenic option to improve GC prognosis. Finally, the GES7 signature is associated with sensitivity to specific drugs. This last result still needs to be validated but opens the possibility of applying potential drugs for GES7<sup>high</sup> patients.

The discovery process for genes associated with prognosis and for building scores used in this study relies on classic statistics. Nowadays, more sophisticated modeling approach, like machine learning algorithms, can be used to capture and model complex non-linear relationships in the data, which may be challenging for classic statistical models (Zhang et al., 2023). However, the implementation of these algorithms can sometimes introduce challenges in interpretability and may not always guarantee better performance, especially in situations with limited sample sizes or noisy data. The simplicity of our index is advantageous, as it enables easier clinical translation and adoption in routine practice. On the other hand, we recognize limitations to our study, such as the use of previously published data and reliance on self-reported participant information. Furthermore, the inclusion of a prospective validation cohort will strengthen the findings. Also, it will be important to explore more affordable methods for measuring GES7, such as QPCR, to facilitate its future use in routine clinical practice.

Taken together, our findings provide compelling evidence for the relevance of GES7 as a prognostic marker in gastric cancer, and we hope that our results will inspire further research in this area.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The Cancer Genome Atlas (TCGA) was obtained from <http://firebrowse.org/> (TCGA). The microarray data from Asian Cancer Research Group (ACRG), Yonsei University Severance Hospital (YUSH), Korea University Guro Hospital (KUGH), National Cancer Centre of Singapore (NCCS), Kosin University College of Medicine (KUCM), and The University of Texas MD Anderson Cancer Center (MDACC), was obtained from <https://www.ncbi.nlm.nih.gov/geo/>.

## Ethics statement

Ethical approval was not required for the studies involving humans because our study describes analysis done on publicly available data of tumor human transcriptomes. The links of the public databases, or their references, were provided in the manuscript and in the “Data availability” request. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from as explained in previous answers, we do not use human samples. Written informed consent to participate in this study was not required from the participants or the participants’ legal guardians/next of kin in accordance with

the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

MV, RA, and CO, conceived and designed the study. MV, AM, and CO performed the statistical and bioinformatic analyses. MV, RA, AC, OM-S, AM, and CO, reviewed the statistical analysis and reviewed the paper. AM and CO, gave final approval of the submitted version. All authors contributed to the article and approved the submitted version.

## Funding

This investigation was supported by São Paulo Research Foundation (FAPESP) through CO (Grant #2017/08430-5) and AM (Grant #2019/05583-0).

## References

- Abdel-Rahman, O. (2017). Validation of the 8th AJCC staging system for gastric cancer in a population-based setting. *Expert Rev. Gastroenterol. Hepatol.* 12, 525–530. doi:10.1080/17474124.2018.1413348
- Amin, M. B., Frederick, Greene, L., Edge, S. B., Compton, C. C., Gershenwald, J. E., et al. (2017). The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA Cancer J. Clin.* 67, 93–99. doi:10.3322/CAAC.21388
- Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18, 220–314. doi:10.1186/s13059-017-1349-1
- Balkwill, F. R., Capasso, M., and Hagemann, T. (2012). The tumor microenvironment at a glance. *J. Cell Sci.* 125, 5591–5596. doi:10.1242/jcs.116392
- Bass, A. J., Thorsson, V., Shmulevich, I., Reynolds, S. M., Miller, M., Bernard, B., et al. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nat* 513, 202–209. doi:10.1038/nature13480
- Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17, 218–220. doi:10.1186/s13059-016-1070-5
- Bechter, O. E., Dumez, H., Costermans, J., Punie, K., Hsu, K., Dedieu, J. F., et al. (2016). Phase I safety and pharmacokinetic dose-escalation study of pilaralisib polymorph E, a phosphoinositide 3-kinase inhibitor in tablet formulation, in patients with solid tumors or lymphoma. *Cancer Chemother. Pharmacol.* 781 (78), 83–90. doi:10.1007/S00280-016-3056-0
- Buccarelli, M., D’Alessandris, Q. G., Matarrese, P., Mollinari, C., Signore, M., Cappannini, A., et al. (2021). Elesclomol-induced increase of mitochondrial reactive oxygen species impairs glioblastoma stem-like cell survival and tumor growth. *J. Exp. Clin. Cancer Res.* 40, 228–317. doi:10.1186/s13046-021-02031-4
- Chang, J-J., Wang, X-Y., Zhang, W., Tan, C., Sheng, W-Q., and Xu, M-D. (2022). Comprehensive molecular characterization and identification of prognostic signature in stomach adenocarcinoma on the basis of energy-metabolism-related genes. *World J. Gastrointest. Oncol.* 14, 478–497. doi:10.4251/wjgo.v14.i2.478
- Chen, C. N., Lin, J. J., Chen, J. J. W., Lee, P. H., Yang, C. Y., Kuo, M. L., et al. (2005). Gene expression profile predicts patient survival of gastric cancer after surgical resection. *J. Clin. Oncol.* 23, 7286–7295. doi:10.1200/JCO.2004.00.2253
- Cheong, J. H., Yang, H. K., Kim, H., Kim, W. H., Kim, Y. W., Kook, M. C., et al. (2018). Predictive test for chemotherapy response in resectable gastric cancer: a multi-cohort, retrospective analysis. *Lancet Oncol.* 19, 629–638. doi:10.1016/S1470-2045(18)30108-6
- Chonov, D. C., Ignatova, M. M. K., Ananiev, J. R., and Gulubova, M. V. (2019). IL-6 activities in the tumour microenvironment. Part 1. *Open Access Maced. J. Med. Sci.* 7, 2391–2398. doi:10.3889/OAMJMS.2019.589
- Colak, S., and ten Dijke, P. (2017). Targeting TGF- $\beta$  signaling in cancer. *Trends Cancer* 3, 56–71. doi:10.1016/j.trecan.2016.11.008
- Cristescu, R., Lee, J., Nebozhyn, M., Kim, K. M., Ting, J. C., Wong, S. S., et al. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.* 21 (21), 449–456. doi:10.1038/nm.3850
- Deng, M., Yu, R., Wang, S., Zhang, Y., Li, Z., Song, H., et al. (2018). Limb-Bud and heart attenuates growth and invasion of human lung adenocarcinoma cells and predicts survival outcome. *Cell Physiol. Biochem.* 47, 223–234. doi:10.1159/000489801
- Dikken, J. L., Van De Velde, C. J. H., Gönen, M., Verheij, M., Brennan, M. F., and Coit, D. G. (2012). The new American Joint committee on cancer/international union against cancer staging system for adenocarcinoma of the stomach: increased complexity without clear improvement in predictive accuracy. *Ann. Surg. Oncol.* 19, 2443–2451. doi:10.1245/S10434-012-2403-6
- Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder Di, Hackl, H., et al. (2019). Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* 11, 34–20. doi:10.1186/s13073-019-0638-6
- Gong, L. B., Wen, T., Li, Z., Xin, X., Che, X. F., Wang, J., et al. (2019). DYNC1I1 promotes the proliferation and migration of gastric cancer by up-regulating IL-6 expression. *Front. Oncol.* 9, 491. doi:10.3389/fonc.2019.00491
- Hashemzadeh, K., Jokar, M. H., Sedighi, S., and Moradzadeh, M. (2018). Therapeutic potency of PI3K pharmacological inhibitors of gastrointestinal cancer. *Middle East J. Dig. Dis.* 11, 5–16. doi:10.15171/mejjd.2018.122
- Henke, E., Nandigama, R., and Ergün, S. (2020). Extracellular matrix in the tumor microenvironment and its impact on cancer therapy. *Front. Mol. Biosci.* 6, 160. doi:10.3389/fmbo.2019.00160
- Heshmati, Y., Kharazi, S., Türköz, G., Chang, D., Kamali Dolatabadi, E., Boström, J., et al. (2018). The histone chaperone NAPIL3 is required for haematopoietic stem cell maintenance and differentiation. *Sci. Rep.* 8, 11202–11217. doi:10.1038/s41598-018-29518-z
- Higgins, J. P. T., and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558. doi:10.1002/SIM.1186
- Huang, M., Ma, X., Shi, H., Hu, L., Fan, Z., Pang, L., et al. (2017). FAM83D, a microtubule-associated protein, promotes tumor growth and progression of human gastric cancer. *Oncotarget* 8, 74479–74493. doi:10.18632/ONCOTARGET.20157
- Jhunjhunwala, S., Hammer, C., and Delamarre, L. (2021). Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nat. Rev. Cancer* 21 (21), 298–312. doi:10.1038/s41568-021-00339-z
- Ji, X., De Bu, Z., Yan, Y., Li, Z. Y., Wu, A. W., Zhang, L. H., et al. (2018). The 8th edition of the American Joint committee on cancer tumor-node-metastasis staging system for gastric cancer is superior to the 7th edition: results from a Chinese mono-institutional study of 1663 patients. *Gastric Cancer* 21, 643–652. doi:10.1007/s10120-017-0779-5
- Kim, D. G., An, J. Y., Kim, H., Shin, S. J., Choi, S., Seo, W. J., et al. (2019). Clinical implications of microsatellite instability in early gastric cancer. *J. Gastric Cancer* 19, 427–437. doi:10.5230/JGC.2019.19.E38

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1206609/full#supplementary-material>

- Kupfer, S. S. (2017). Gaining ground in the genetics of gastric cancer. *Gastroenterology* 152, 926–928. doi:10.1053/j.gastro.2017.02.026
- Li, C., Zheng, Y., Pu, K., Zhao, D., Wang, Y., Guan, Q., et al. (2020). A four-DNA methylation signature as a novel prognostic biomarker for survival of patients with gastric cancer. *Cancer Cell Int.* 20, 88. doi:10.1186/S12935-020-1156-8
- Li, J., Pu, K., Li, C., Wang, Y., and Zhou, Y. (2021). A novel six-gene-based prognostic model predicts survival and clinical risk score for gastric cancer. *Front. Genet.* 12, 109. doi:10.3389/fgene.2021.615834
- Liu, H., Liu, K., Dong, Z., Cao, Y., Zhang, C., Du, J., et al. (2021). The role of p21-activated kinases in cancer and beyond: where are we heading? *Front. Cell Dev. Biol.* 9, 325–336. doi:10.1080/03008207.2020.1736054
- Liu, J. Y., Peng, C. W., Yang, X. J., Huang, C. Q., and Li, Y. (2018). The prognosis role of AJCC/UICC 8th edition staging system in gastric cancer, a retrospective analysis. *Am. J. Transl. Res.* 10, 292–303.
- Liu, Y., Wu, J., Huang, W., Weng, S., Wang, B., Chen, Y., et al. (2020). Development and validation of a hypoxia-immune-based microenvironment gene signature for risk stratification in gastric cancer. *J. Transl. Med.* 18, 201. doi:10.1186/S12967-020-02366-0
- Lou, S., Meng, F., Yin, X., Zhang, Y., Han, B., and Xue, Y. (2021). Comprehensive characterization of RNA processing factors in gastric cancer identifies a prognostic signature for predicting clinical outcomes and therapeutic responses. *Front. Immunol.* 12, 719628. doi:10.3389/fimmu.2021.719628
- Lu, J., Zheng, C. H., Long, C. L., Li, P., Xie, J. W., Wang, J. B., et al. (2017). The effectiveness of the 8th American Joint committee on cancer TNM classification in the prognosis evaluation of gastric cancer patients: a comparative study between the 7th and 8th editions. *Eur. J. Surg. Oncol.* 43, 2349–2356. doi:10.1016/j.ejso.2017.09.001
- Lv, S., Li, Y., Ning, H., Zhang, M., Jia, Q., and Wang, X. (2021). CircRNA GFRA1 promotes hepatocellular carcinoma progression by modulating the miR-498/NAP1L3 axis. *Sci. Rep.* 111 (11), 386–410. doi:10.1038/s41598-020-79321-y
- Malcikova, J., Tausch, E., Rossi, D., Sutton, L. A., Soussi, T., Zenz, T., et al. (2018). ERIC recommendations for TP53 mutation analysis in chronic lymphocytic leukemia—Update on methodological approaches and results interpretation. *Leuk 325* (32), 1070–1080. doi:10.1038/s41375-017-0007-7
- Marrelli, D., Morgagni, P., De Manzoni, G., Coniglio, A., Marchet, A., Saragoni, L., et al. (2012). Prognostic value of the 7th AJCC/UICC TNM classification of noncardia gastric cancer: analysis of a large series from specialized western centers. *Ann. Surg.* 255, 486–491. doi:10.1097/SLA.0B013E3182389B1A
- McGranahan, N., Burrell, R. A., Endesfelder, D., Novelli, M. R., and Swanton, C. (2012). Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep.* 13, 528–538. doi:10.1038/EMBOR.2012.61
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (12), 453–457. doi:10.1038/nmeth.3337
- Oh, S. C., Sohn, B. H., Cheong, J. H., Kim, S. B., Lee, J. E., Park, K. C., et al. (2018). Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. *Nat. Commun.* 91 (9), 1777–1814. doi:10.1038/s41467-018-04179-8
- Ooi, C. H., Ivanova, T., Wu, J., Lee, M., Tan, I. B., Tao, J., et al. (2009). Oncogenic pathway combinations predict clinical prognosis in gastric cancer. *PLOS Genet.* 5, e1000676. doi:10.1371/JOURNAL.PGEN.1000676
- Park, S., Lee, J., Kim, Y. H., Park, J., Shin, J. W., and Nam, S. (2016). Clinical relevance and molecular phenotypes in gastric cancer, of TP53 mutations and gene expressions, in combination with other gene mutations. *Sci. Rep.* 61 (6), 34822–34910. doi:10.1038/srep34822
- Qu, Y., Wang, J., Sim, M. S., Liu, B., Giuliano, A., Barsoum, J., et al. (2009). Elesclomol, counteracted by Akt survival signaling, enhances the apoptotic effect of chemotherapy drugs in breast cancer cells. *Breast Cancer Res. Treat.* 1212 (121), 311–321. doi:10.1007/S10549-009-0470-6
- Racle, J., and Gfeller, D. (2020). Epic: a tool to estimate the proportions of different cell types from bulk gene expression data. *Methods Mol. Biol.* 2120, 233–248. doi:10.1007/978-1-0716-0327-7\_17
- Rothwell, J. A., Jenab, M., Karimi, M., Truong, T., Mahamat-Saleh, Y., Ferrari, P., et al. (2022). Metabolic syndrome and risk of gastrointestinal cancers: an investigation using large-scale molecular data. *Clin. Gastroenterol. Hepatol.* 20, e1338–e1352. doi:10.1016/j.cgh.2021.10.016
- Shi, J., Cheng, C., Ma, J., Liew, C. C., and Geng, X. (2018). Gene expression signature for detection of gastric cancer in peripheral blood. *Oncol. Lett.* 15, 9802–9810. doi:10.3892/ol.2018.8577
- Shu, P., Qin, J., Shen, K., Chen, W., Liu, F., Fang, Y., et al. (2017). The IGCA staging system is more accurate than AJCC7 system in stratifying survival of patients with gastric cancer in stage III. *BMC Cancer* 17, 238–246. doi:10.1186/s12885-017-3235-3
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi:10.3322/CAAC.21660
- Szász, A. M., Lánczky, A., Nagy, Á., Förster, S., Hark, K., Green, J. E., et al. (2016). Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget* 7, 49322–49333. doi:10.18632/ONCOTARGET.10337
- Tolaney, S., Burris, H., Gartner, E., Mayer, I. A., Saura, C., Maurer, M., et al. (2014). Phase I/II study of pilaralisib (SAR245408) in combination with trastuzumab or trastuzumab plus paclitaxel in trastuzumab-refractory HER2-positive metastatic breast cancer. *Breast Cancer Res. Treat.* 149, 151–161. doi:10.1007/S10549-014-3248-4
- Turgeon, M. O., Perry, N. J. S., and Poulogiannis, G. (2018). DNA damage, repair, and cancer metabolism. *Front. Oncol.* 8, 15. doi:10.3389/fonc.2018.00015
- Wang, P., et al. Wang, P., Wang, Y., Hang, B., Zou, X., Mao, J.-H. (2016). A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. *Oncotarget* 7, 55343–55351. doi:10.18632/ONCOTARGET.10533
- Wang, X., Liu, Y., Niu, Z., Fu, R., Jia, Y., Zhang, L., et al. (2017). Prognostic value of a 25-gene assay in patients with gastric cancer after curative resection. *Sci. Rep.* 71 (7), 7515–7611. doi:10.1038/s41598-017-07604-y
- Wu, Z., Qiu, M., Mi, Z., Meng, M., Guo, Y., Jiang, X., et al. (2019). WT1-interacting protein inhibits cell proliferation and tumorigenicity in non-small-cell lung cancer via the AKT/FOXO1 axis. *Mol. Oncol.* 13, 1059–1074. doi:10.1002/1878-0261.12462
- Yan, X., Wang, J., and Bai, Y. (2022). Potentials of circSOBP in the diagnosis and prognosis of gastric cancer. *Scand. J. Gastroenterology* 57, 1344–1348. doi:10.1080/00365521.2022.2088246
- Yao, D., Li, C., Rajkota, M. S. R., He, Z., Huang, J., Wang, J., et al. (2020). P21-Activated kinase 1: emerging biological functions and potential therapeutic targets in cancer. *Theranostics* 10, 9741–9766. doi:10.7150/THNO.46913
- Yoon, H. M., Ryu, K. W., Nam, B. H., Cho, S. J., Park, S. R., Lee, J. Y., et al. (2012). Is the new seventh AJCC/UICC staging system appropriate for patients with gastric cancer? *J. Am. Coll. Surg.* 214, 88–96. doi:10.1016/J.JAMCOLLSURG.2011.09.018
- Yu, R., Li, Z., Zhang, C., Song, H., Deng, M., Sun, L., et al. (2019). Elevated limb-bud and heart development (LBH) expression indicates poor prognosis and promotes gastric cancer cell proliferation and invasion via upregulating Integrin/FAK/Akt pathway. *PeerJ* 2019, e6885. doi:10.7717/peerj.6885
- Zeng, C., Guo, X., Long, J., Kuchenbaecker, K. B., Droit, A., Michailidou, K., et al. (2016). Role of CSL-dependent and independent Notch signaling pathways in cell apoptosis. *Breast Cancer Res.* 18, 1–12. doi:10.1007/s10495-015-1188-z
- Zhang, B., Shi, H., and Wang, H. (2023). Machine learning and ai in cancer prognosis, prediction, and treatment selection: a critical approach. *J. Multidiscip. Healthc.* 16, 1779–1791. doi:10.2147/JMDH.S410301
- Zhang, T., Lai, S., Cai, Y., Huang, Z., Li, Y., Chen, S., et al. (2021). Comprehensive analysis and identification of prognostic biomarkers and therapeutic targets among FAM83 family members for gastric cancer. *Front. Cell Dev. Biol.* 9, 3122. doi:10.3389/fcell.2021.719613
- Zhang, Y., Yu, R., Li, Z., Zhang, C., Song, H., Deng, M., et al. (2019). Upregulation of family with sequence similarity 83 member D expression enhances cell proliferation and motility via activation of Wnt/β-catenin signaling and predicts poor prognosis in gastric cancer. *Cancer Manag. Res.* 11, 6775–6791. doi:10.2147/CMAR.S203082
- Zhu, L., Wang, H., Jiang, C., Li, W., Zhai, S., Cai, X., et al. (2020). Clinically applicable 53-gene prognostic assay predicts chemotherapy benefit in gastric cancer: a multicenter study. *EBioMedicine* 61, 103023. doi:10.1016/j.ebiom.2020.103023