Genome-wide identification of CpG island methylator phenotype related gene signature as a novel prognostic biomarker of gastric cancer

Zhuo Zeng[1,2], Daxing Xie[1,2] and Jianping Gong[1,2]

[1] Molecular Medicine Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China
[2] Department of GI Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

ABSTRACT

Background: Gastric cancer (GC) is one of the most fatal cancers in the world. Results of previous studies on the association of the CpG island methylator phenotype (CIMP) with GC prognosis are conflicting and mainly based on selected CIMP markers. The current study attempted to comprehensively assess the association between CIMP status and GC survival and to develop a CIMP-related prognostic gene signature of GC.

Methods: We used a hierarchical clustering method based on 2,082 GC-related methylation sites to stratify GC patients from the cancer genome atlas into three different CIMP subgroups according to the CIMP status. Gene set enrichment analysis, tumor-infiltrating immune cells, and DNA somatic mutations analysis were conducted to reveal the genomic characteristics in different CIMP-related patients. Cox regression analysis and the least absolute shrinkage and selection operator were performed to develop a CIMP-related prognostic signature. Analyses involving a time-dependent receiver operating characteristic (ROC) curve and calibration plot were adopted to assess the performance of the prognostic signature.

Results: We found a positive relationship between CIMP and prognosis in GC. Gene set enrichment analysis indicated that cancer-progression-related pathways were enriched in the CIMP-L group. High abundances of CD8+ T cells and M1 macrophages were found in the CIMP-H group, meanwhile more plasma cells, regulatory T cells and CD4+ memory resting T cells were detected in the CIMP-L group. The CIMP-H group showed higher tumor mutation burden, more microsatellite instability-H, less lymph node metastasis, and more somatic mutations favoring survival. We then established a CIMP-related prognostic gene signature comprising six genes (*CST6, SLC7A2, RAB3B, IGFBP1, VSTM2L* and *EVX2*). The signature was capable of classifying patients into high-and low-risk groups with significant difference in overall survival (OS; $p < 0.0001$). To assess performance of the prognostic signature, the area under the ROC curve (AUC) for OS was calculated as 0.664 at 1 year, 0.704 at 3 years and 0.667 at 5 years. When compared with previously published gene-based signatures, our CIMP-related signature was comparable or better at predicting prognosis. A multivariate Cox regression analysis indicated the CIMP-related prognostic gene signature was an independent

Corresponding author
Jianping Gong,
jpgong@tjh.tjmu.edu.cn

prognostic indicator of GC. In addition, Gene ontology analysis indicated that keratinocyte differentiation and epidermis development were enriched in the high-risk group.

Conclusion: Collectively, we described a positive association between CIMP status and prognosis in GC and proposed a CIMP-related gene signature as a promising prognostic biomarker for GC.

INTRODUCTION

Gastric cancer (GC) is responsible for over 1,000,000 new cases and around 783,000 deaths in the world annually, making it the 5th most frequently diagnosed cancer and the third leading cause of cancer-related death (*Bray et al., 2018*). Surgery with subsequent adjuvant chemoradiotherapy remains the only treatment with curative potential (*Bang et al., 2012*), and the prognosis for gastric adenocarcinoma is primarily determined by the TNM classification of staging system (*Warneke et al., 2011*). The clinical outcome, nonetheless, is notably variable and erratic in individual patient, which firmly implies that a few of the biological determinants of tumor behavior are unidentified. Thus, advances in molecular insight into GC are critically required for improved prognostic stratification and new targeted therapeutic strategies.

Recently, with the progress of high-throughput screening, sequencing has enabled a more thorough insight into the molecular identity of GC. An updated classification scheme has been introduced based on comprehensive molecular characterization including tumors infected with Epstein–Barr virus, tumors with microsatellite instability (MSI), and tumors with a distinct degree of aneuploidy, which were termed genomic stability and chromosomal instability. Each subgroup shows peculiar genetic and clinical characteristics (*Cancer Genome Atlas Research, 2014*).

Alterations of DNA methylation is a vital event during tumorigenesis, and gastrointestinal cancers show the highest frequency of DNA methylation alterations among the reported tumor types (*Cancer Genome Atlas Research, 2014*). Methylation of the dinucleotides of CpG islands throughout the genome is mediated by DNA methyltransferases (*Craig & Bickmore, 1994*), and commonly results in gene silencing. Disorder of DNA methylation in cancer affects gene expression and results in the cancer progression (*Vaissiere, Sawan & Herceg, 2008*).

CpG island methylator phenotype (CIMP) in tumors, which has been initially described and broadly debated in colorectal cancer (*Hughes et al., 2013*). Lately CIMP has been described in other tumor types including bladder, breast, glioblastoma, pancreatic and prostate cancers, as well as for gastric adenocarcinomas and is considered to be helpful for predicting prognosis (*Jia et al., 2019*; *Moarii, Reyal & Vert, 2015*; *Ueki et al., 2000*). In GC, conflicting conclusions regarding the prognostic association of CIMP have been scattered among previous studies, owing to the limitation of selected DNA

methylation markers and the presence of multiple confounding factors in these studies (*An et al., 2005*; *Ben Ayed-Guerfali et al., 2011*; *Park et al., 2010*). Although a meta-analysis of the prognostic value of CIMP status in GC has been performed, an explicit conclusion was not reached (*Powell et al., 2018*).

In this study, we aimed to use publicly available data to comprehensively analyze CIMP in GC, and to develop a CIMP-related prognostic gene signature.

MATERIALS AND METHODS

Data acquisition

We downloaded methylation data, which has 408,376 probes and 397 samples, including 395 GC samples and two normal samples, measured by the Illumina HumanMethylation450 platform, from the cancer genome atlas (TCGA)-STAD project (https://portal.gdc.cancer.gov/) by using the *TCGA-Assembler 2* package (*Wei et al., 2018*). RNA-Seq profiles were obtained from TCGA by virtue of *GDC Data Transfer Tool*. We downloaded two verified microarrays with matched clinical information from the GEO GC database: GSE13861 (*Cho et al., 2011*) (65 GC samples; platform: GPL6884 Illumina HumanWG-6 v3.0 expression beadchip), GSE62254 (*Cristescu et al., 2015*) (300 GC samples; platform: GPL570 Affymetrix Human Genome U133 Plus 2.0 Array). We used the *TCGAbiolinks* package to acquire the mutation data of GC samples (*Colaprico et al., 2016*). We obtained complete and matched clinical information on GC patients from cBioportal, including sex, age, histologic features, pathologic stage, family history and, infection status for *Helicobacter pylori* and Epstein–Barr virus.

Our study was performed according to the publication guidelines required by TCGA.

Data analysis

The *Minfi* package was adopted to analyze methylation data (*Aryee et al., 2011*). In view of the distribution of CpG islands and the technical limitations of sequencing, we filtered the probes from the X and Y chromosomes or probes that are known to have common SNPs at the CpG site, and cross-reactive probes. The *DESeq2* package was adopted to analyze the differentially expressed genes (DEGs) between CIMP-related subgroups (*Love, Huber & Anders, 2014*). The criteria to determine DEGs were an adjusted p-value < 0.05 and an absolute value of log2 fold change >2, and BH method was used for adjustment for multiple testing. In order to identify the different pathways between GC samples with specific CIMP status, we performed GSEA analysis (*Subramanian et al., 2005*). The mRNA expression data downloaded from the Gene Expression Omnibus (GEO) database were normalized and analyzed by the *limma* package (*Ritchie et al., 2015*). The mutation data were summarized and analyzed by the *maftools* package (*Mayakonda et al., 2018*). Raw code for analyzing was uploaded in Data S1.

Identification of CIMP in GC samples

To assess the CIMP feature in GC, CpG methylation sites with a relatively high variability of β-values in tumor samples (SD > 0.2) and relatively low β-values in normal

samples (mean β value < 0.05), were chosen as the representative CpG methylation sites for subsequent clustering analysis, following a previous study (*Li et al., 2019*). The *ConsensusClusterPlus* package was adopted to run unsupervised clustering analysis based on M value of the selected 2,082 probes by means of the *K*-means algorithm (*Wilkerson & Hayes, 2010*).

Analysis of tumor-infiltrating immune cells

The proportions of the 22 types of tumor-infiltrating immune cells were counted by CIBERSORT (*Newman et al., 2015*). CIBERSORT is a tool to provide an estimation of cell composition of mixed tissues based on gene expression profiles. We uploaded the modified gene expression data and standard annotation to the CIBERSORT portal and ran the LM22 signature, which contains 547 genes distinguishing 22 human immune cell types, and 1,000 permutations. Final results were normalized to sum up to one and could be assessed straightforwardly as cell fractions for contrast (Table S1).

Development and validation of a CIMP-related prognostic signature

Using the *DESeq2* package, 1,072 DEGs were calculated between the CIMP-H and CIMP-L samples (Table S2), which were defined by relative methylation level. We then performed Cox regression to assess the prognostic significance of the DEGs. A suitable prognostic model is assumed to identify a smaller number of genes favorable for clinical practice. We therefore used the least absolute shrinkage and selection operator (LASSO) in combination to diminish the number of CIMP-related prognostic genes (*Gui & Li, 2005*). The *glmnet* package was used to perform the penalized Cox regression model with the LASSO penalty, and 1,000-times cross-validations were applied to determine the optimal values of the penalty parameter lambda. We selected lambda.min to get six CIMP-related prognostic genes. We then extracted the coefficients from multivariate Cox regression to build a gene signature. We adopted the *survminer* package to determine the cut-off value of the risk score. Then, the patients were divided into high-and low-risk subgroups according to their risk score. We used Kaplan–Meier analysis to compare OS rates between the high- and low-risk group. To identify whether the risk score was an independent factor, we conducted univariate Cox regression and multivariate Cox regression analyses. Statistical significance was inferred where $p < 0.05$.

Validation in GEO dataset

To confirm the performance of our prognostic signature, we applied it to two GEO databases, GSE13861 ($n = 65$) and GSE62254 ($n = 300$). The mRNA expression data were prepared using the *limma* package. Scale, a generic R function including centering and scaling, was used to scale the GEO mRNA expression data to common range. Next we used a developed prognostic signature to calculate the risk score of every sample and divided patients by the cut-off value. Kaplan–Meier analysis was conducted between the high- and low-risk groups for overall survival. In addition, we conducted
receiver operating characteristic (ROC) curve analysis and calculated an AUC for every database.

Functional enrichment analysis

We applied the *clusterProfiler* package for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis between different risk score-related subgroups (*Yu et al., 2012*). Then we adopted the *GOplot* package to illustrate our GO and KEGG results (Tables S3 and S4). Protein–protein interaction analysis was carried out based on the DEGs of high-and low-risk subgroups using the STRING portal (https://string-db.org) and Cytoscape software (*Shannon et al., 2003*). Functional annotation of genes in the module was perform by DAVID database (*Huang et al., 2007*).

RESULTS

Methylation landscape of GC sample

In this study, we utilized DNA methylation profiles from TCGA database to perform a comprehensive analysis of DNA methylation in GC. We adopted 2,082 methylation sites with high variability as our CIMP signature for downstream analysis. Unsupervised hierarchical clustering analysis of 395 GC samples based on our specific CIMP signature was performed and all the patients were separated into three subgroups as CIMP-L, CIMP-M and CIMP-H (Figs. 1A and 1B; Table S5). The CIMP-L subgroup had the lowest methylation level, while the CIMP-H subgroup had broad hypermethylation across these sites. In addition, we plotted the Delta area and consensus CDF to verify our clustering pattern (Figs. S1A and S1B). To assess the performance of our classification based on CIMP signature, we reclassified GC samples according to a previous DNA methylation clustering analysis (*Cancer Genome Atlas Research, 2014*). A strong concordance was exhibited between these two classification systems (Fig. 1C). The C1 cluster, representing an EBV-associated DNA methylation signature with extreme hypermethylation, consisted primarily of CIMP-H samples, while the C4 cluster, representing a hypomethylated subgroup, consisted mainly of CIMP-L. Importantly, to evaluate the correlation between CIMP and prognosis, the overall survival of each subgroup was assessed by the Kaplan–Meier method. The result indicated a significant difference in prognosis among the different CIMP-related subgroups, with the CIMP-H group showing better prognosis and the CIMP-L group showing worse prognosis (Fig. 1D). In addition, we investigated the relationship between CIMP and progression-free survival (PFS). However, we found no significant differences in PFS existed among the CIMP-related subgroups (Fig. S1C).

The clinical characteristics of patients with different CIMP statuses were summarized (Table 1). Clinical features, including lymph node metastasis, MSI status and EBV infection, had significant differences between CIMP-related subgroups. No significant difference was found in aspects of age, gender, pathologic tumor classification, Lauren classification, grade or *Helicobacter pylori* infection. Specifically, within the CIMP-H subgroup more patients suffered MSI gastric adenocarcinoma and EBV infection, and less
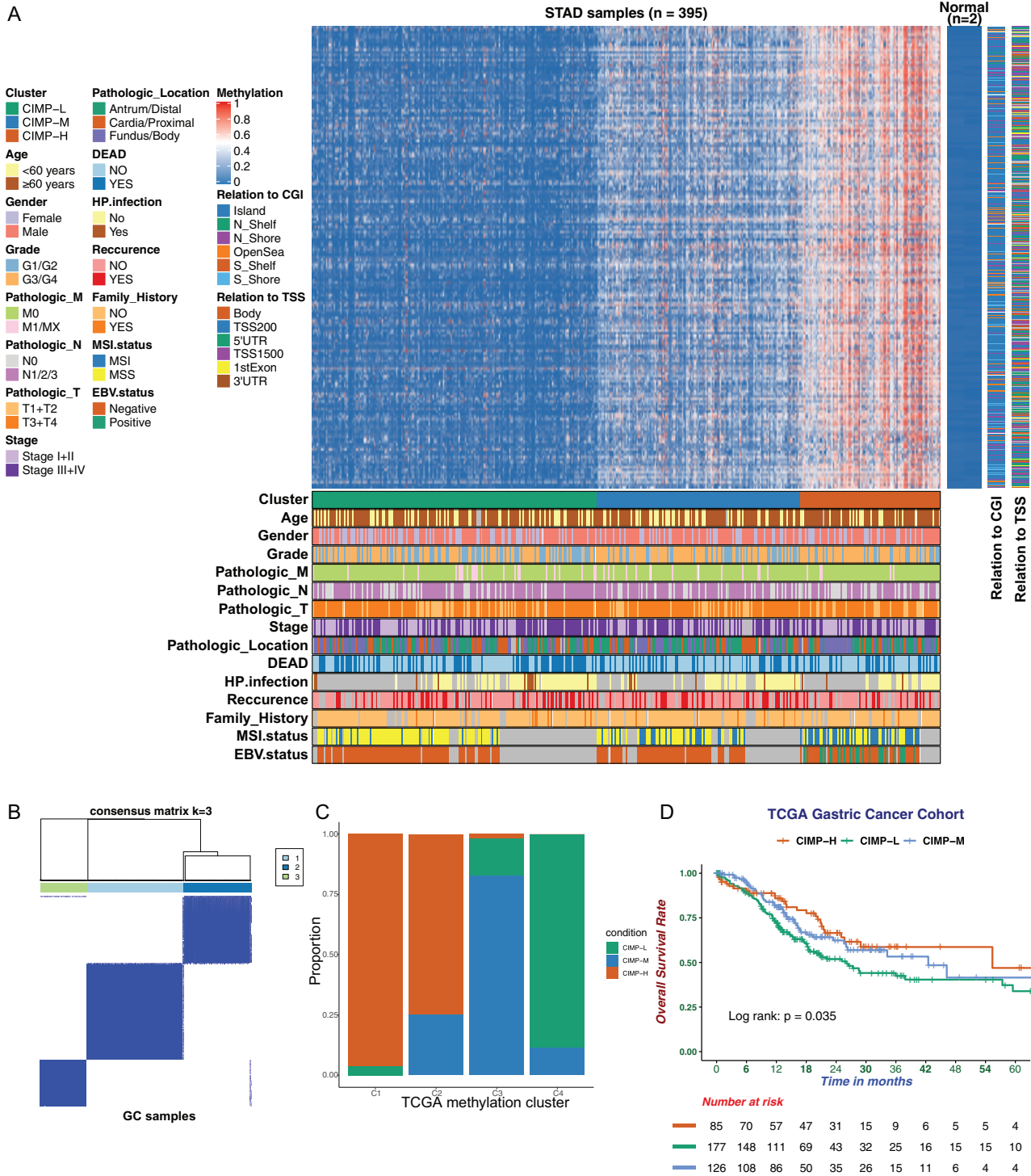
Figure 1 The landscape of CpG island methylator phenotype in gastric cancer. (A) Unsupervised hierarchical clustering of GC samples. The rows represent 2082 CpG methylation sites for clustering. Green, blue, and red cluster represents CIMP-Low (CIMP-L) subgroup, CIMP-Medium (CIMP-M) subgroup and CIMP-High (CIMP-H) subgroup respectively. The CIMP-L subgroup had the lowest methylation level. Clinical information is marked with different colors, and missing information is marked with gray. (B) Clustering result of K-means algorithm by *ConsensusClusterPlus*. (C) Comparison of CMIP-related subgroups with TCGA methylation cluster. (D) Kaplan-Meier survival curves of CIMP-related subgroups. The CIMP-H subgroup had a better OS than other subgroups.

Table 1 Clinical and demographic characteristics of GC patients in CIMP-related subgroups.

	CIMP-L	CIMP-M	CIMP-H	*p*-Value
Number of patients	179	128	88	
Age (Mean (SD))	64.3 (10.5)	65.6 (10.5)	66.4 (11.3)	0.191
Gender				0.194
Female	70 (39.1%)	38 (29.7%)	28 (31.8%)	
Male	109 (60.9%)	90 (70.3%)	60 (68.2%)	
Pathologic_T				0.724
T1	7 (3.9%)	7 (5.5%)	7 (8.0%)	
T2	39 (21.8%)	24 (18.8%)	15 (17.0%)	
T3	85 (47.5%)	63 (49.2%)	38 (43.2%)	
T4	48 (26.8%)	34 (26.6%)	28 (31.8%)	
Pathologic_N				0.044
N0	51 (28.5%)	36 (28.1%)	37 (42.6%)	
N1	49 (27.4%)	31 (24.2%)	22 (25.3%)	
N2	36 (20.1%)	34 (26.6%)	9 (10.3%)	
N3	38 (21.2%)	27 (21.1%)	18 (20.7%)	
NX	5 (2.8%)	0 (0%)	1 (1.1%)	
Pathologic_M				0.115
M0	155 (86.6%)	115 (89.8%)	83 (94.3%)	
M1	11 (6.1%)	10 (7.8%)	2 (2.3%)	
MX	13 (7.3%)	3 (2.3%)	3 (3.4%)	
Pathologic_Stage				0.225
Stage I	23 (12.8%)	14 (10.9%)	16 (18.2%)	
Stage II	58 (32.4%)	42 (32.8%)	32 (36.4%)	
Stage III	78 (43.6%)	59 (46.1%)	38 (43.2%)	
Stage IV	20 (11.2%)	13 (10.2%)	2 (2.3%)	
Lauren.Class				0.112
Diffuse	49 (27.4%)	36 (28.1%)	18 (20.5%)	
Intestinal	115 (64.2%)	71 (55.5%)	61 (69.3%)	
Mixed	15 (8.4%)	21 (16.4%)	9 (10.2%)	
Grade				0.264
G1	4 (2.2%)	3 (2.3%)	2 (2.3%)	
G2	76 (42.5%)	41 (32.0%)	25 (28.4%)	
G3	94 (52.5%)	81 (63.3%)	60 (68.2%)	
GX	5 (2.8%)	3 (2.3%)	1 (1.1%)	
H. pylori infection				0.779
No	79 (87.8%)	56 (90.3%)	33 (91.7%)	
Yes	11 (12.2%)	6 (9.7%)	3 (8.3%)	
MSI.status				<0.001
MSI-H	4 (4.0%)	11 (14.4%)	34 (47.9%)	
MSI-L	11 (10.9%)	18 (23.7%)	8 (11.3%)	
MSS	86 (85.1%)	47 (61.9%)	29 (40.8%)	
EBV.positive				<0.001
Negative	101 (100%)	76 (100%)	46 (64.8%)	
Positive	0 (0%)	0 (0%)	25 (35.2%)	
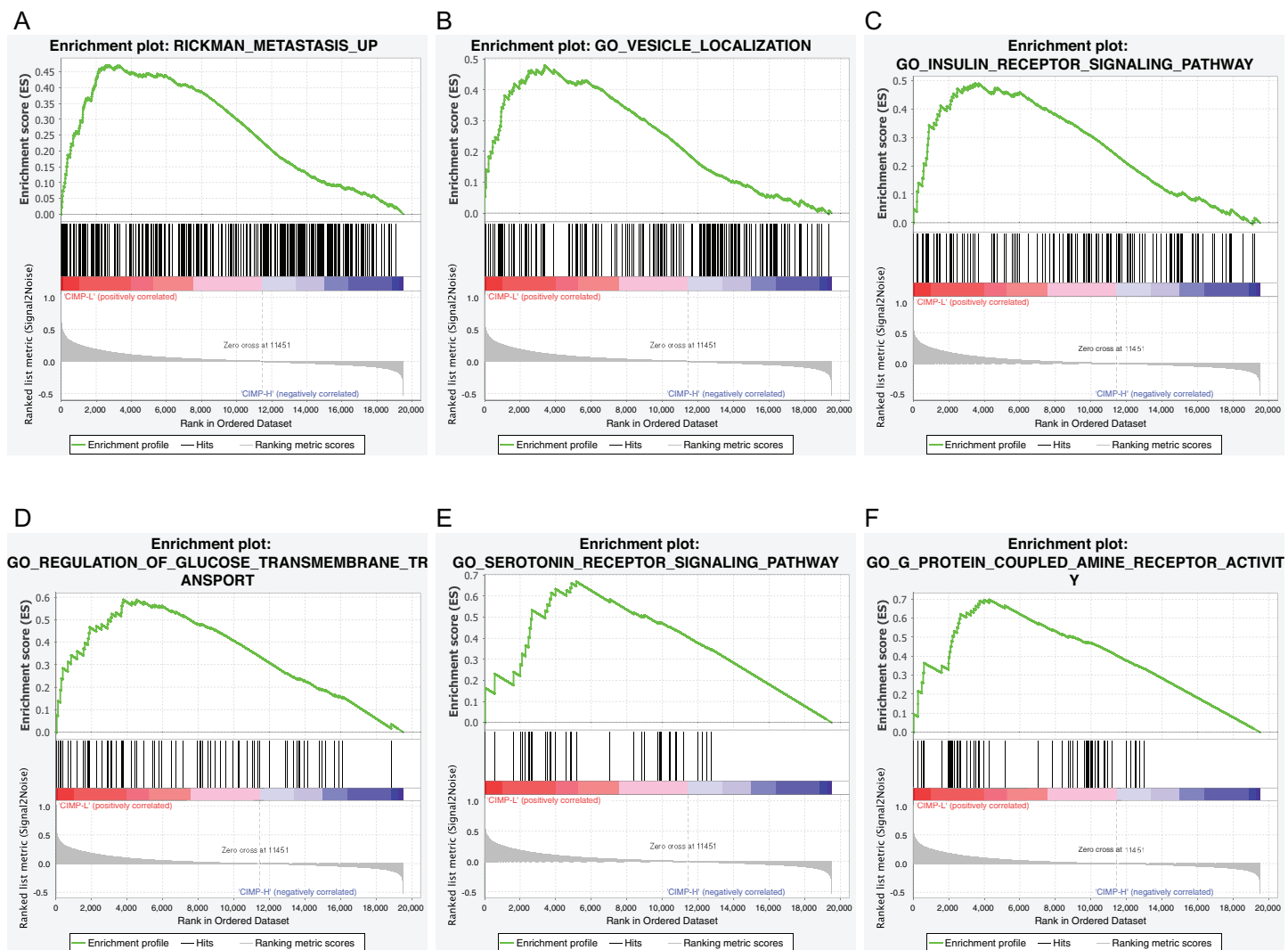
Figure 2 Gene set enrichment analysis of CIMP-related subgroups in the TCGA dataset. Significant enrichment in the CIMP-L subgroup compared with the CIMP-H subgroup. (A) RICKMAN metastasis up; (B) vesicle localization; (C) insulin receptor signaling pathway; (D) glucose transmembrane transport; (E) serotonin receptor signaling pathway; (F) G-protein coupled amine receptor activity.

had lymph node metastasis. In addition, we found that in the CIMP-H subgroup, patients with EBV could be well distinguished from patients with MSI (Fig. S1D).

Gene set enrichment analysis in different CIMP-related subgroups

To identify the biological processes or pathways potentially regulated by the CpG island methylation signature, we applied the GSEA analysis between different CIMP-related subgroups based on RNA-seq profiles. We found that the gene signatures of "Rickman metastasis up", "Vesicle localization", "Insulin receptor signaling pathway", "Regulation of glucose transmembrane transport", "Serotonin receptor signaling pathway", "G-protein coupled amine receptor activity" were enriched in CIMP-L subgroup (Figs. 2A–2F). Importantly, all the pathways have been linked to GC progression.
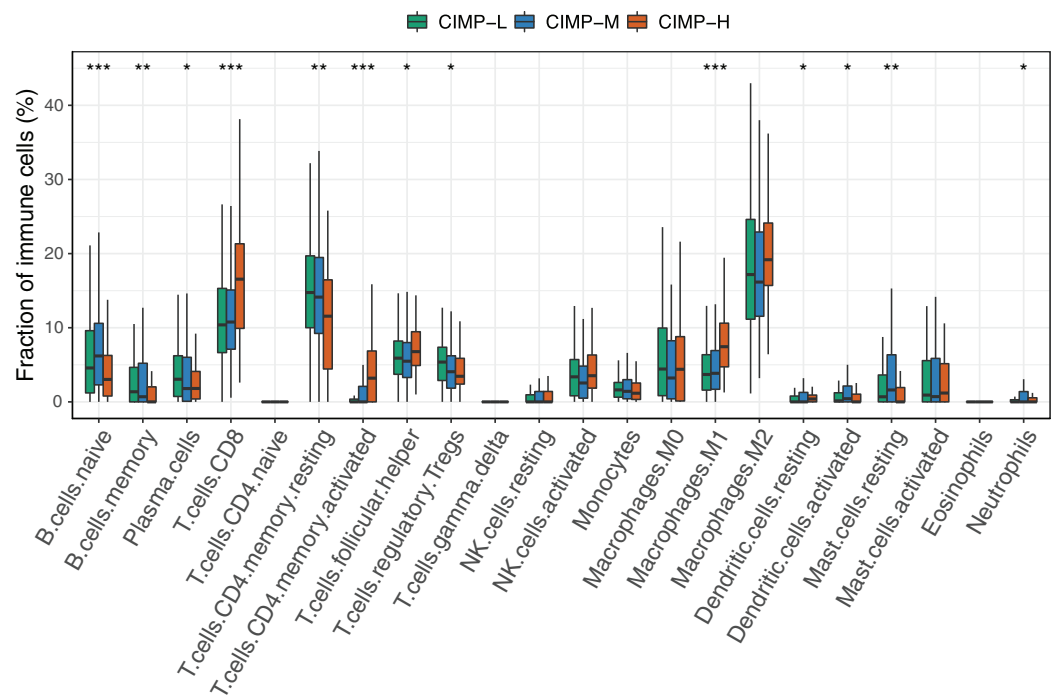
Figure 3 The comparison of fractions of tumor-infiltrating immune cells between CIMP-related subgroups in GC. (Kruskal–Wallis test was used, * represents for $p < 0.05$, ** represents for $p < 0.01$, *** represents for $p < 0.001$).

Tumor-infiltrating immune cells in different CIMP-related subgroups

We assessed the presence of tumor-infiltrating immune cells (TIICs) in CIMP-related subgroups by using CIBERSORT (Fig. 3). Obviously, immune cells showed differential infiltration pattern between CIMP-related subgroups. The proportions of B cells, plasma cells, T cells CD4 memory resting, regulatory T cells and resting mast cells were significantly higher in the in the CIMP-L subgroup. Meanwhile, CD8+ T cells, T cells CD4 memory activated, T follicular helper cells, M1 macrophages, and Dendritic cells resting were higher in the CIMP-H subgroup. Other immune cells, including NK cells and monocytes, didn't show significant differences. In addition, we found in most samples no apparent T cells CD4 naïve, T cells gamma delta and Eosinophils were infiltrated. These results indicated that CIMP-L subgroup have a distinct immune phenotype, which is considered to impair and suppress antitumor immunity.

Analysis of DNA somatic mutations in patients with distinct CIMP status

A distinct set of genetic aberrations between CIMP-related subgroups was evident in our study. We found 350 samples with mutations in a total of 391samples (89.51%), with TTN and TP53 ranking as the most common mutation gene (Fig. 4A). Mutation of the TP53 gene was found enriched in the CIMP-L subgroup. At the same time, mutations of TTN and MUC16 were higher in the CIMP-H subgroup. The most common mutations in GC samples were missense mutations, comprising the majority of SNPs, the main
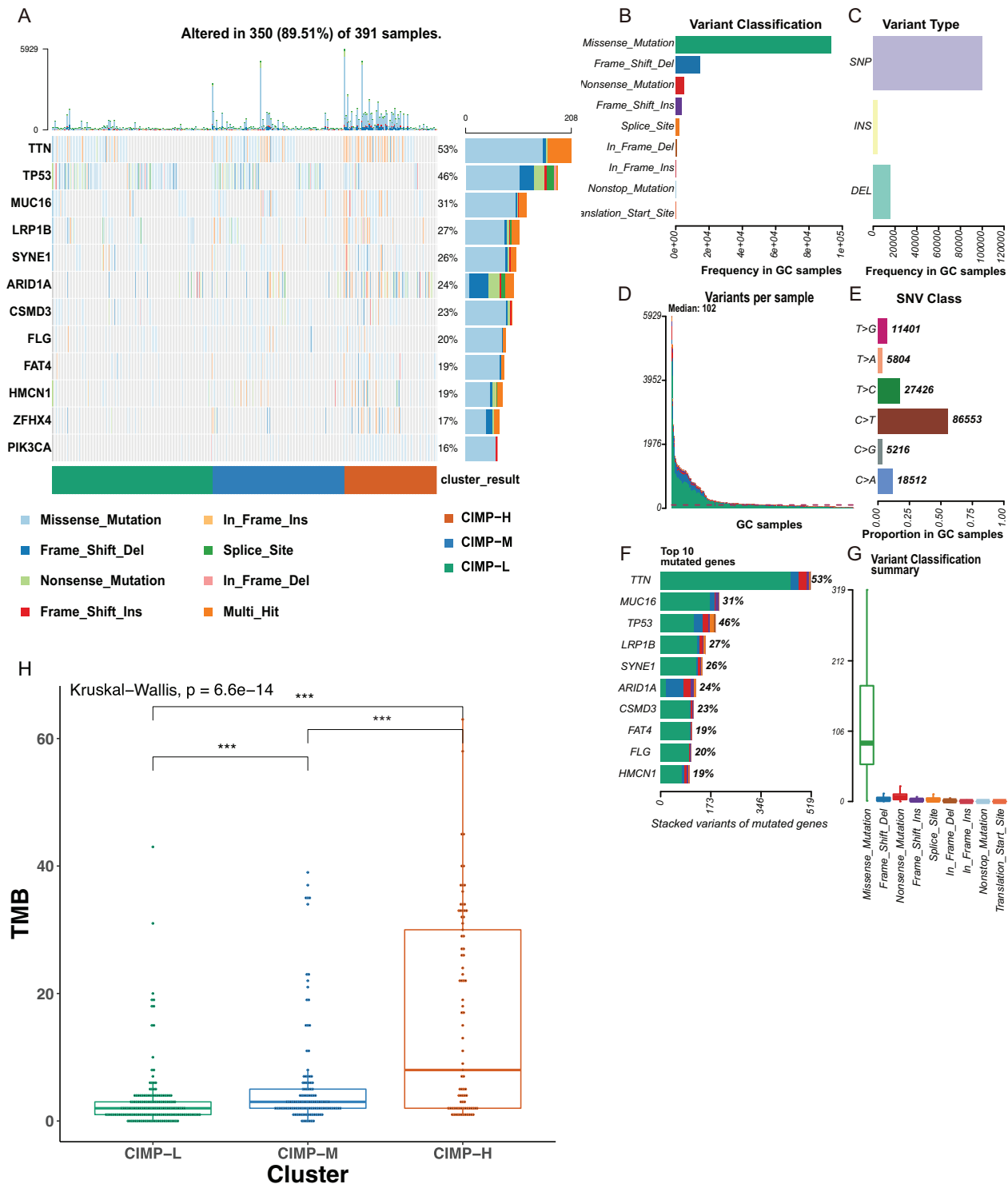
Figure 4 The mutational signatures in CIMP-related subgroups. (A) Waterfall plots showed mutation information of each gene in GC subgroups stratified by CIMP status, and various colors with annotations at the bottom represented the different mutation types. (B) Frequency of variant classifications. (C) Summary of variant types. (D) Summary of variants in per sample. (E) Summary of SNV classes. (F) Top ten mutated genes. (G) Summary of variant classifications. (H) Tumor mutation burden (TMB) of CIMP-related subgroups. The CIMP-H subgroup had higher TMB than the other subgroups. (Kruskal–Wallis test was used, SteelDwass test was used for post-hoc test, * represents for $p < 0.05$, ** represents for $p < 0.01$, *** represents for $p < 0.001$).

SNV classification was C > T transition, and the number of altered bases in each sample was counted (Figs. 4B–4E). We then showed the top 10 mutated genes in GC with ranked percentages, including TTN (53%), MUC16 (31%), TP53 (46%), LRP1B (27%), SYNE1 (26%), ARID1A (24%), CSMD3 (23%), FAT4 (19%), FLG (20%), HMCN1 (19%), and summarized the mutation types in GC (Figs. 4F and 4G). We then calculated the tumor mutation burden (TMB), which is considered to correlate with enhanced clinical response to immunotherapy and superior OS. We found TMB was higher in the CIMP-H subgroup (Fig. 4H).

Establishment of a CIMP-related prognostic gene signature

To screen differentially expressed genes (DEGs) in CIMP subgroups, we downloaded RNA-seq data for 208 samples defined as CIMP-H or CIMP-L subgroups from the TCGA database and analyzed the data by using the *DEseq2* package. We identified 1,072 DEGs, which we narrowed down to 147 genes highly associated with OS using univariate Cox regression. To obtain the genes with the highest potential prognostic values, we used least absolute selection and shrinkage operator (LASSO) regression analysis. A prognostic signature comprising six genes, including *cystatin E/M (CST6), solute carrier family 7 member 2 (SLC7A2), RAB3B, member RAS oncogene family (RAB3B), insulin like growth factor binding protein 1 (IGFBP1), V-set and transmembrane domain containing 2 like (VSTM2L)* and *even-skipped homeobox 2 (EVX2)*, was developed (Figs. 5, 1A and 1B; Table 2). The risk score was calculated as follows: risk score = (0.230 × the normalized expression of *CST6*) + (0.257 × the normalized expression of *SLC7A2*) + (0.156 × the normalized expression of *RAB3B*) + (0.114 × the normalized expression of *IGFBP1*) + (0.024 × the normalized expression of *VSTM2L*) + (0.187 × the normalized expression of *EVX2*). The cutoff value (0.235) was counted by the *survminer* package (Fig. 5C).

The patients were then divided into high- and low-risk subgroups according to their risk score. We found high-risk patients had more deaths and higher expression levels of CIMP-related prognostic genes (Figs. 5D and 5E). We then found those in the high-risk had a worse OS than those in the low-risk group (Fig. 5F). To access the performance of the prognostic signature, time-dependent ROC curves and AUC were printed and counted (Fig. 5G). The AUC was 0.664 at 1 year, 0.704 at 3 years and 0.667 at 5 years.

The univariate and multivariate Cox regression analyses indicated that the predictive value of risk score for overall survival was independent of CIMP status (Figs. 5H and 5I).

Validation and evaluation of the CIMP-related signature in the GEO cohort

To further verify the robustness of the six-genes prognostic signature in GC, two verified microarrays with matched clinical information from the GEO GC database were analyzed. In every dataset, patients were stratified into high-or low-risk group according to the cutoff point calculated following the prognostic signature. Consistent with the results from the TCGA cohort, the high-risk group had a worse survival outcome in two datasets (Figs. 6A and 6B). Indicating favorable performance of our prognostic signature, the AUC of the GSE13861 dataset was 0.638 at 1 year, 0.777 at 3 years, 0.745 at 5 years (Figs. 6C).
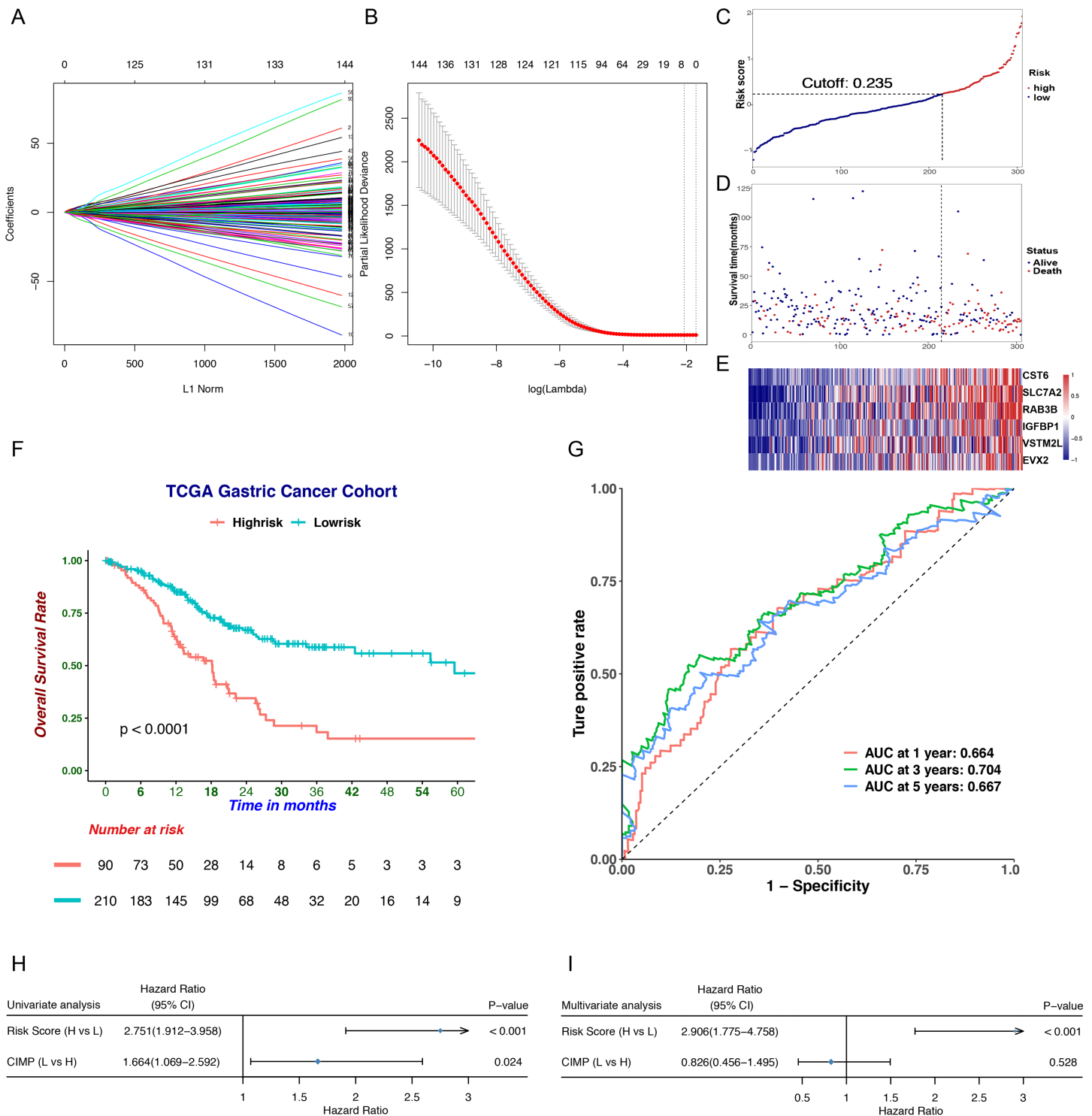
Figure 5 Prognostic analysis of the CIMP-related prognostic gene signature in TCGA cohort. (A) LASSO coefficient profiles of candidate genes. Each curve indicated one gene. (B) Cross-validation in the LASSO model. (C) The distribution of risk score. (D) The distribution of survival status. (E) The distribution of expression levels of the six genes in TCGA cohort. (F) The Kaplan–Meier curve for patients divided into high-and low-risk based on CIMP-related prognostic gene signature. (G) Receiver operating characteristic curve of CIMP-related prognostic signature at different years. (H) and (I) Univariate and multivariate regression analysis of the CIMP status and risk score calculated based on CIMP-related prognostic signature.

Table 2 The CIMP-related prognostic gene signature based on six genes in TCGA cohort. HR: hazard ratio, CI: confidential interval.

Symbol	Multivariate Cox regression coefficient	Univariate Cox regression			
		HR	95%CI	*p*-Value	z score
CST6	0.230	1.327	[1.139–1.547]	2.80E−04	3.633
SLC7A2	0.257	1.405	[1.192–1.656]	5.01E−05	4.055
RAB3B	0.156	1.361	[1.152–1.607]	2.76E−04	3.637
IGFBP1	0.114	1.244	[1.070–1.448]	4.63E−03	2.832
VSTM2L	0.024	1.233	[1.053–1.444]	9.44E−03	2.596
EVX2	0.187	1.221	[1.084–1.376]	1.02E−03	3.284
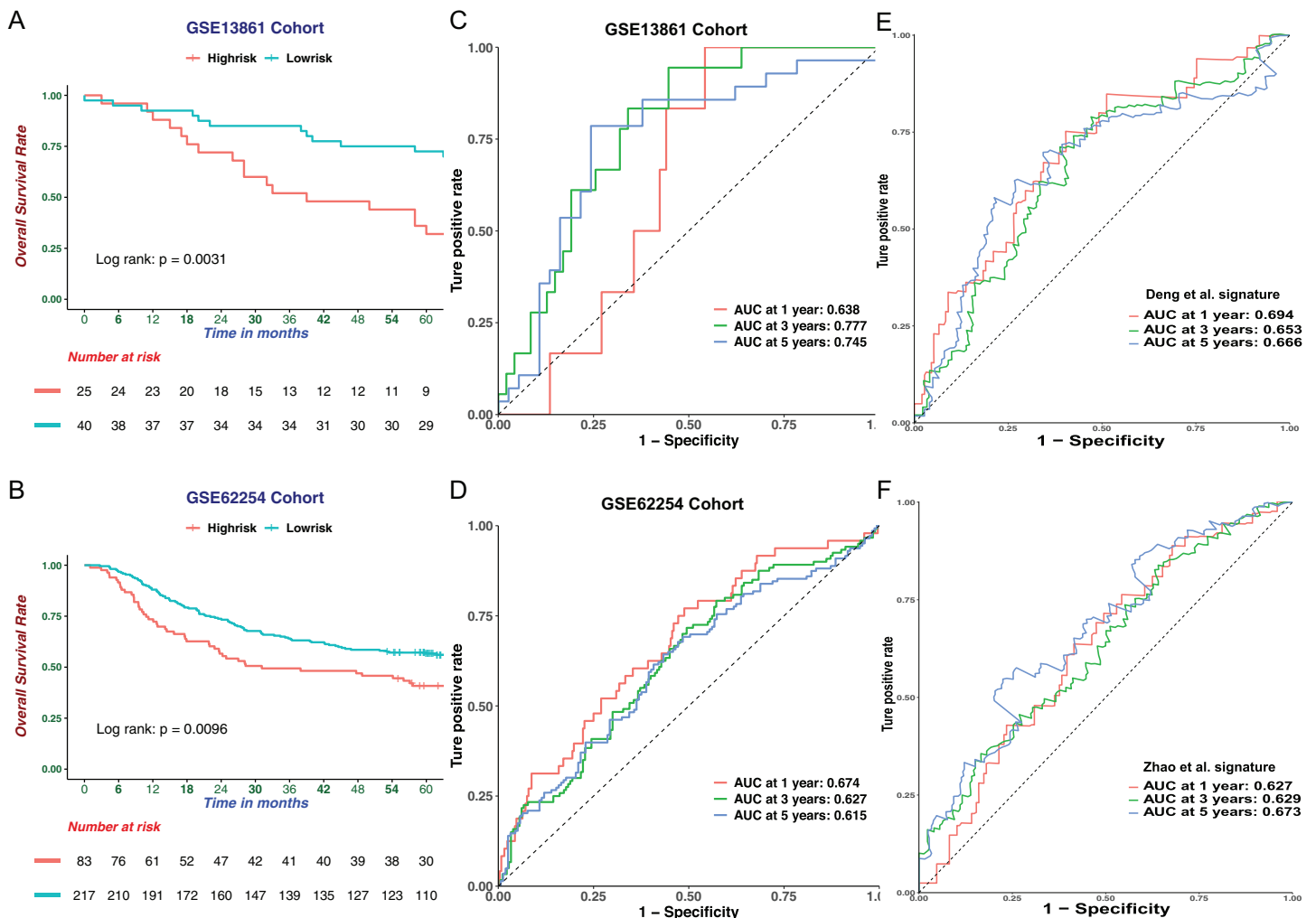
Figure 6 Validation of the CIMP-related prognostic gene signature in independent cohorts. The Kaplan–Meier curve for patients divided into high-and low-risk based on CIMP-related prognostic gene signature in (A) GSE13861 (*n* = 65) and (B) GSE62254 (*n* = 300) cohorts. Receiver operating characteristic curve of CIMP-related prognostic signature at different years in (C) GSE13861 (*n* = 65) and (D) GSE62254 (*n* = 300) cohorts. (E) and (F) Receiver operating characteristic curves of the other signatures reported in previous studies in the prediction of OS for TCGA cohort.

The AUC of GSE62254 dataset was 0.674 at 1 year, 0.627 at 3 years and 0.615 at 5 years (Fig. 6D). We then compared our CIMP-related prognostic signature with two prognostic signatures published previously. We extracted formulae from each study, and the results of ROC curve analysis implied that the CIMP-related prognostic signature was comparable or better at predicting the prognosis in the TCGA cohort (Figs. 6E and 6F). Taken together, the prognostic signature based on CIMP was a reliable prognostic marker in GC.

The risk score developed from the six genes signature as an independent prognostic factor

We contrasted the prognostic value of the risk score was contrasted with clinical parameters by univariate and multivariate analyses. Clinical parameters included diagnostic age, gender, pathologic TNM, pathologic stage, pathologic grade, Lauren classification, status of *H. pylori* and EB virus infection and MSI status. We found that risk score acted as an independent prognostic factor and had significant effects in both the univariate analysis and the multivariate analysis, with *p* values < 0.05 (Fig. 7A). Furthermore, the risk score had robust prognostic value (with HR = 3.364, 95% CI [1.906–5.937]). We then used the risk score as a nomogram to predict patients' outcome (Fig. 7B). The Calibration plot indicated that predicted OS and the actual OS rates at 1,3 and 5 years were similar (Figs. 7C–7E). To verify the role of methylation in the expression of prognostic signature gene, we used Pearson correlation to evaluate the relationship between the methylation levels of the *CST6, SLC7A2, RAB3B, IGFBP1, VSTM2L* and *EVX2* promoters and their expression levels. Consistent results were found among these six genes (Figs. 8A–8F). Moreover, the expression of signature genes was consistent among the CIMP-related subgroups. Expression levels of the signature genes were higher in the CIMP-L subgroup than those in the other subgroups (Figs. 8G–8L).

Distinct biological processes in risk score stratified subgroups

We identified 382 DEGs between the high-risk and low-risk subgroups in GC samples. Then, we carried out GO and KEGG analyses to identify the molecular mechanisms associated with these DEGs. For GO analysis, the top five enriched terms were "cornification", "keratinocyte differentiation", "epidermis development", "keratinization" and "epidermal cell differentiation" (Fig. 9A). In KEGG analysis four pathways were enriched, including "neuroactive ligand-receptor interaction", "complement and coagulation cascades", "staphylococcus aureus infection" and "cholesterol metabolism". The "neuroactive ligand-receptor interaction" was shown to be the main associated pathway with 14 genes involved (Fig. 9B). In addition, STRING was used to draw 382 DEGs into a PPI network complex, which contained 366 nodes and 1,176 interactions (Fig. S3). Then, Cytoscape was used to identify the most significant module in the PPI network. The most significant module (score = 17.2) recognized by MCODE, a plug-in of Cytoscape, contained 38 nodes and 318 interactions (Fig. 9C). Consistent with the results of GO analysis, the genes in the module were found to be related to "keratinocyte
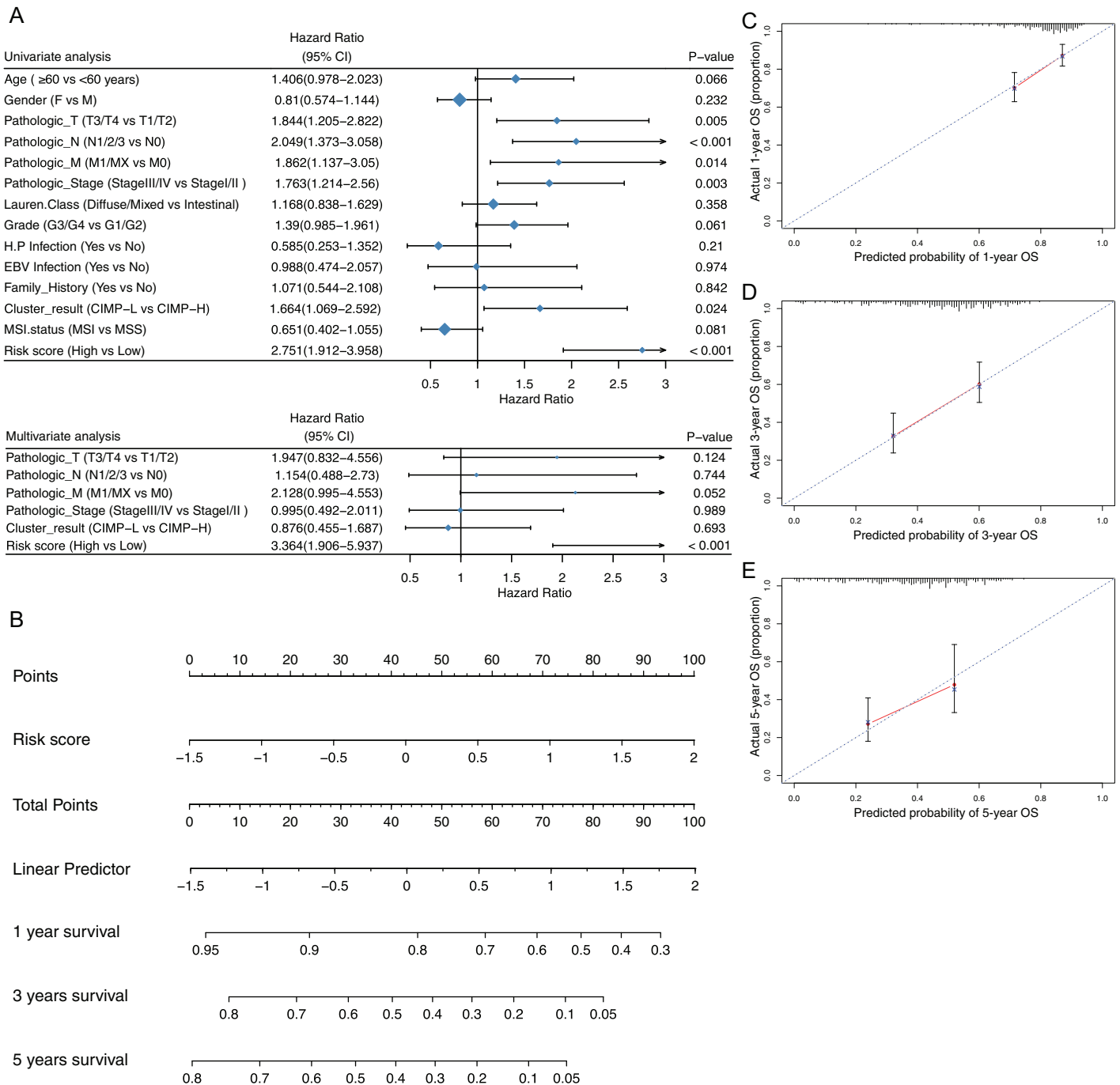
Figure 7 Prediction of risk score for overall survival (OS). (A) Univariate and multivariate regression analysis of the relation between the CIMP-related prognostic risk score and clinicopathological characters regarding OS (CI, confidential interval). (B) The nomogram for predicting probabilities of overall survival at 1,3 and 5 years. Calibration plot of predicted survival and actual survival at (C) 1 year, (D) 3 years, (E) 5 years.

Figure 8 **Correlations of CIMP-related prognostic gene signature.** Correlations between signature gene promoter methylation and signature gene expression (A) CST6; (B) SLC7A2; (C) RAB3B; (D) IGFBP1; (E) VSTM2L; (F) EVX2 (Pearson's rank correlation analysis was used). Expression of prognostic signature genes in CIMP-related subgroups (G) CST6; (H) SLC7A2; (I) RAB3B; (J) IGFBP1; (K) VSTM2L; (L) EVX2 (Kruskal-Wallis test was used, SteelDwass test was used for post-hoc test, * represents for $p < 0.05$, ** represents for $p < 0.01$, *** represents for $p < 0.001$, ns represents no significance).

Full-size DOI: 10.7717/peerj.9624/fig-8
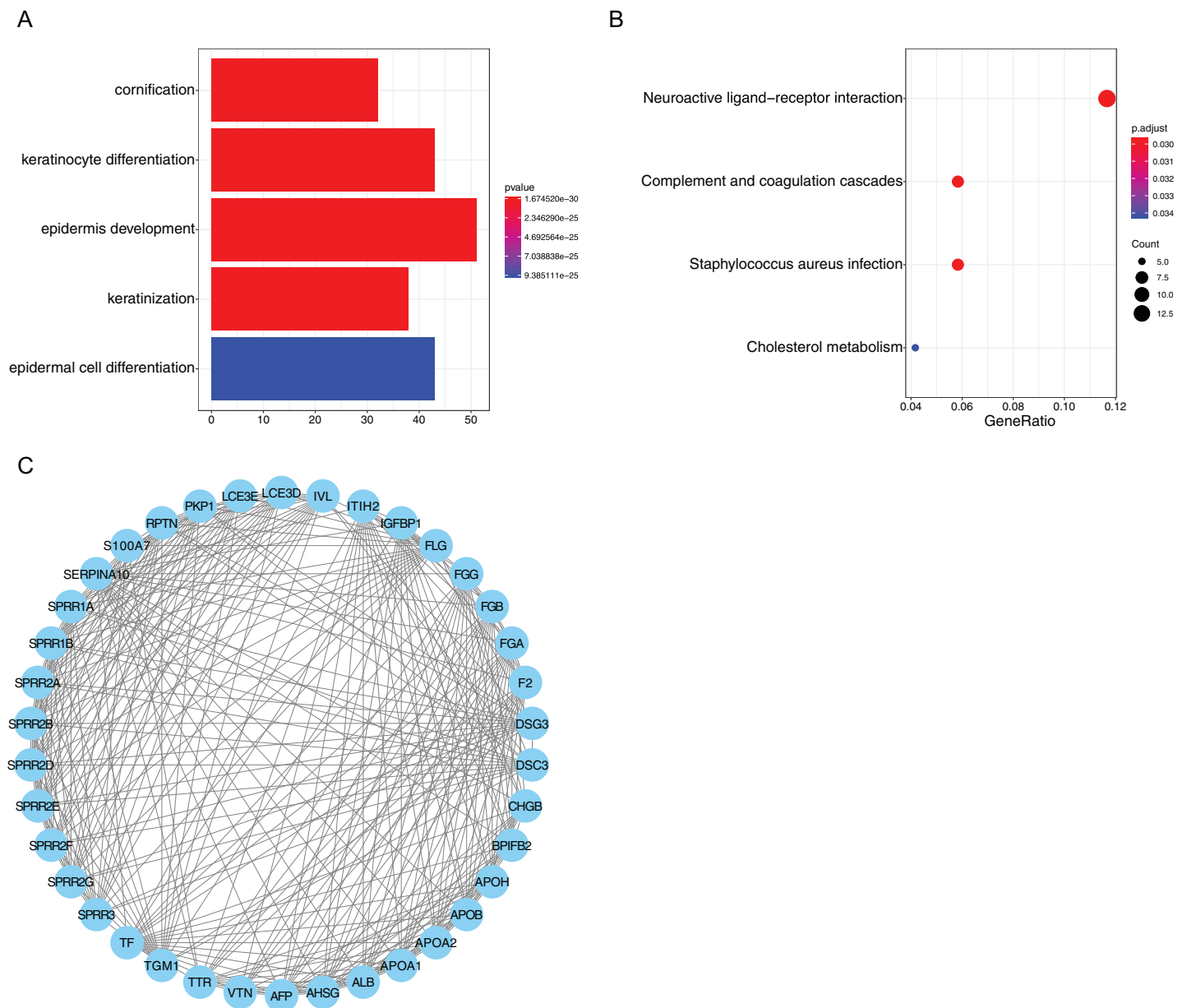
Figure 9 Functional enrichment analysis of risk score related genes. (A) GO analysis of differentially expressed genes between risk score stratified subgroups. (B) KEGG analysis of differentially expressed genes between risk score stratified subgroups. (C) The most significant module identified in the PPI network of differentially expressed genes between risk score stratified subgroups.

differentiation", "keratinization", "peptide cross-linking" and "epidermis development" (Table 3).

DISCUSSION

In the field of cancer research, increasing attention has focused on DNA methylation. Patterns of DNA methylation can predict prognosis and survival of human cancers (*Hao et al., 2017*). CIMP refers to promoter CpG island hypermethylation and is well characterized in colorectal cancers. In contrast, the relationships between CIMP and

Table 3 The enriched GO terms of genes in the most significant module.

ID	Term	Count	p-Value
GO:0030216	keratinocyte differentiation	15	5.49E−25
GO:0031424	keratinization	13	3.49E−23
GO:0018149	peptide cross-linking	13	6.06E−23
GO:0008544	epidermis development	11	7.52E−16
GO:0002576	platelet degranulation	8	1.25E−09

clinicopathological features are controversial in GC. Previous studies on CIMP used different methods to measure methylation, like measuring across several CpG sites of a gene or across several genes. Conflicting conclusions may be drawn, due to variation among studies in methodologies of DNA methylation analysis and CIMP marker panels, which brought the bias of chosen marker panels (*An et al., 2005*; *Oue et al., 2003*; *Park et al., 2010*; *Powell et al., 2018*). In the present study, we used the methylation data measured by the Illumina HumanMethylation450K platform to assess the DNA methylation status of 40,8376 CpG sites including CpG sites located at the promoter regions of protein-coding genes in multiple samples simultaneously. We adopted methylation sites with high variability as our CIMP signature for unsupervised hierarchical clustering to comprehensively assess CIMP status in GC. In addition, our study included more patients with methylation data from the same platform than previous study (*Cancer Genome Atlas Research, 2014*). We firstly used this methodology to divide GC samples into three distinct subgroups according to their levels of methylation at selected methylation sites and describe the positive relation between CIMP and prognosis at the global methylation level, in contrast to analyses using chosen markers. Consistent with this, our CIMP-H subgroup showed more favorable clinical characteristics including less lymph node metastasis and lower metastasis status. Promoter hypermethylation is a prominent feature of EBV-associated GC, and we found that samples with EBV infection were enriched in the CIMP-H subgroup in this study (*Kang et al., 2002*). Previous studies indicated that GC patients with MSI show a significant longer overall survival compared with those who have MSS and assumed that MSI GC has a better prognosis because of its earlier stage at diagnosis, less lymph node metastasis and intestinal histological type (*Mathiak et al., 2017*; *Polom et al., 2018*). Consistent with these researches, samples with MSI were enriched in the CIMP-H subgroup which showed a better prognosis in our study. We then plotted the Kaplan–Meier survival curves according to previous DNA methylation clustering. However, we found no significant differences in this clustering (Fig. S2A). We believed that it was caused by too few cases in the C1 group, which was corresponding to our CIMP-H group. Therefore, we combined C2 and C3 groups, and plotted Kaplan-Meier survival curves between C2 + C3 and C4 groups. We found that the difference in survival was still not statistically significant, but we could see the difference in survival between these two groups (Fig. S2B). We believed it was due to the fact that the previous DNA methylation clustering was based on the merger of two methylation

platforms, and that the number of samples included was not as large as ours. In our study, we found that there was no significant relationship between CIMP and PFS in GC. However, considering that some patients lack PFS information, we need to include more patients with complete PFS information to clarify the role of CIMP in PFS of GC in future research.

Our study provided insight into the landscapes of molecular features in patients with distinct CIMP statuses. Based on the gene set enrichment analysis, we identified cancer-related oncogenic pathways enriched in the CIMP-L subgroup, including metastasis, vesicle transfer, G-protein coupled receptor, energy transfer. Cancer cell-derived vesicles serve as intercellular communication vehicles and carry pathogenic components, such as proteins, mRNA, miRNA, DNA, lipids and transcriptional factors, that can mediate paracrine signaling in the tumor microenvironment (*Fujita, Yoshioka & Ochiya, 2016*). Vesicles mediate the formation of pre-metastatic niches to promote metastasis in tumors, including GC (*Deng et al., 2017*; *Jung et al., 2009*; *Peinado et al., 2012*). *Li et al. (2011)* reported that Insulin-like growth factor-I (IGF1) regulated the expression of the VEGF ligand to facilitate angiogenesis and lymphangiogenesis in GC cell lines, and blocking IGF1 could enhance the effectiveness of bevacizumab. High glucose conditions were shown to promote GC cell proliferation and reduce susceptibility to chemotherapy (*Zhao et al., 2015*). In addition, serotonin-induced signaling pathways promoted tumor progression (*Sarrouilhe & Mesnil, 2019*). This suggests these pathways could have potential as novel drug targets.

Immunotherapy is becoming a routine cancer treatment option, and disparate tumor-infiltrating immune cells profiles were observed among CIMP-related subgroups. CD8+ (cytotoxic) T cells are very important for immune defense and tumor surveillance, and are known to correlate with more favorable outcome in GC (*He et al., 2017*). In GC, tumor-associated plasma cells are polarized to produce IgG4 and associated with tumor progression and poor prognosis (*Miyatani et al., 2016*). Regulatory T cells (Tregs) are T cells which have a role in regulating or suppressing other cells in the immune system, leading to limiting excessive immune responses. Tregs suppress activation, proliferation and cytokine production of CD4+ T cells and CD8+ T cells. Tregs are thought to suppress B cells and dendritic cells. *Liu et al. (2019)* revealed that Tregs promoted Lgr5 expression in GC cells via TGF-β1 signaling pathway and was negatively associated with survival. Differences in the levels of the M1 macrophages were also observed. M1 macrophages are an integral cellular component of the immune system, and play a critical role in protection against intracellular pathogens and cancer cells (*Yin et al., 2017*). M1 macrophages have been reported to inhibit tumor growth in GC (*Liu et al., 2013*).

Recently, TMB has been increasingly accepted as a biomarker of response to immunotherapy. High TMB contributes to the synthesis of aberrant and potentially immunogenic mutation-associated neoantigens by the cancer cells, which attract CD8+ CTLs and activated Th1 cells to the tumor microenvironment. In this study, we found higher TMB in the CMIP-H subgroup, potentially indicating a better response to immunotherapy in this group. Consistent with the distribution of TMB, more MSI samples, especially MSI-H, were detected in the CIMP-H subgroup. The prevalence of MSI

in GC is relatively high, and as MSI-H GCs are strongly associated with PD-L1 positivity, they could be applicable targets of anti-PD-1 therapies (*Kim et al., 2018*). The mutations of TTN and MUC16 were announced to be correlated with better survival result in lung and GCs (*Cheng et al., 2019*; *Li et al., 2018a*).

Recent researches have indicated that clinicopathological parameters such as tumor depth, lymph node metastasis, margin status, and metastatic condition are unsatisfactory for accurately predicting patient prognosis. Outcome varies tremendously among patients with comparable clinicopathological features. Therefore, with the advantage of high-throughput sequencing technologies, mRNAs have been used as molecular biomarkers of the cancer diagnosis and prognosis and shown their critical clinical application potential. For examples, *Zhao et al. (2019)* investigated genes relevant to the cell cycle from the TCGA database and described a set of five genes (*MARCKS, CCNF, MAPK14, INCENP* and *CHAF1A*), which were significantly associated with OS. They used this signature to stratify GC patients into two groups with significantly different survival outcomes. Distinct clinical features were also demonstrated between the two groups. In another research, the predictive value of DNA methylation gene for prognosis was determined in GC, and different pathways and biological processes associated with tumorigenesis were found in groups with distinct gene methylation levels (*Hu et al., 2019*).

In our study, a gene signature including *CST6, SLC7A2, RAB3B, IGFBP1*, *VSTM2L* and *EVX2*, was developed based on CIMP. *CST6* has been reported to play a role in the progression of triple-negative breast cancer (TNBC) and may act as a tumor-promoter gene. High *CST6* expression was also associated with a higher rate of lymph node metastasis (*Li et al., 2018b*). *SLC7A2* is essential for transport of L-arginine, lysine and ornithine and genetic polymorphisms in the *SLC7A2* gene are associated with colorectal cancer progression (*Sun et al., 2017*). *SLC7A2* has also been found to play a role in radio-resistance of non-small cell lung cancer. *RAB3B*, a member of RAS oncogene family, is shown to be a target of miR-200b, which is supposed to be tumor suppressor in GC (*Tang et al., 2013*; *Ye et al., 2014*). *RAB3B*, has been shown to be overexpressed in prostate cancer patients and promote prostate cancer cell survival (*Tan et al., 2012*). *IGFBP1*, an insulin-like growth factor binding protein, is revealed to be associated with hematogenous metastasis and poor survival in GC. And the expression of *IGFBP1* is positively associated with tumor invasion, lymph node metastasis and vascular invasion (*Sato et al., 2019*). *VSTM2L* is reported to be downregulated in the *H. pylori*-positive GC samples (*Hu et al., 2018*). However, the role of *VSTM2L* in tumors is rarely reported, and its role requires further studies. *EVX2* is recently revealed to be regulated by methylation and serve as a methylation biomarker for lung cancer (*Rauch et al., 2012*). In our study, the expression of *EVX2* was higher in CIMP-L subgroup consistent with its mechanism of epigenetic regulation. Our six-genes risk signature was an independent prognostic biomarker of GC, with patients in high-risk groups showing significantly worse prognosis than those in low-risk groups. Our results support the notion that gene risk signature might have more predictive power than traditional prognostic parameters. The prognostic performance of our signature was validated in the TCGA dataset, and the external datasets GSE13861 and GSE62254. Further, it compared favorably to two other gene-based signatures

(*Deng et al., 2018*; *Zhao et al., 2019*), based on ROC and AUC analyses predicting 1-, 3- and 5-year overall survival. In our study, distinct biological analyses in risk score stratified subgroups indicated that keratinization and keratinization-related processes may play an important role in GC progression. At present, there is no research about the role of keratinization in GC and it needs further studies. In addition, the neuroactive ligand-receptor interaction was shown to be the most significant pathway in stratified subgroups and revealed to be involved in apoptosis and cell proliferation (*Zan & Li, 2019*).

CONCLUSION

In summary, we first identified an accurate and comprehensive association between CIMP and clinical prognosis in GC, where high CIMP indicated better patient prognosis. We then developed and validated a six-genes prognostic signature related to CIMP that can predict the survival of patients with GC, where higher risk score indicated worse patient prognosis. This signature could be an effective tool in clinical practice as a supplement to traditional staging system to indicate progression and predict overall survival of GC. However, our study has some limitations. Firstly, it was based on a retrospective design, so the numbers of patients with the same clinical features in CIMP subgroups were not comparable. Also, the number of datasets used to validate the clinical prognostic signature is not large, and further validation by future prospective studies is desirable. Knowledge of the signature genes in GC development is currently scarce and further experiments are needed to verify their potential molecular mechanisms.

ACKNOWLEDGEMENTS

We thank Jie Shen and Liang Liu for comments on this article. We appreciate Ganxun Li and Furong Liu for assistance with the bioinformatics analysis.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 81702897 and 81572861). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 81702897 and 81572861.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Zhuo Zeng performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.

- Daxing Xie performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Jianping Gong conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The raw data are available from TCGA project TCGA-STAD and from the NCBI GEO GC database (accession numbers: GSE13861, GSE62254). The code used to analyze the data is available as a Supplemental File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.9624#supplemental-information.

REFERENCES

An C, Choi IS, Yao JC, Worah S, Xie K, Mansfield PF, Ajani JA, Rashid A, Hamilton SR, Wu TT. 2005. Prognostic significance of CpG island methylator phenotype and microsatellite instability in gastric carcinoma. *Clinical Cancer Research* **11(2 Pt 1)**:656–663.

Aryee MJ, Wu Z, Ladd-Acosta C, Herb B, Feinberg AP, Yegnasubramanian S, Irizarry RA. 2011. Accurate genome-scale percentage DNA methylation estimates from microarray data. *Biostatistics* **12(2)**:197–210 DOI 10.1093/biostatistics/kxq055.

Bang YJ, Kim YW, Yang HK, Chung HC, Park YK, Lee KH, Lee K-W, Kim YH, Noh S-I, Cho JY, Mok YJ, Kim YH, Ji J, Yeh T-S, Button P, Sirzén F, Noh SH. 2012. Adjuvant capecitabine and oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): a phase 3 open-label, randomised controlled trial. *Lancet* **379(9813)**:315–321 DOI 10.1016/S0140-6736(11)61873-4.

Ben Ayed-Guerfali D, Benhaj K, Khabir A, Abid M, Bayrouti MI, Sellami-Boudawara T, Gargouri A, Mokdad-Gargouri R. 2011. Hypermethylation of tumor-related genes in tunisian patients with gastric carcinoma: clinical and biological significance. *Journal of Surgical Oncology* **103(7)**:687–694 DOI 10.1002/jso.21875.

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **68(6)**:394–424 DOI 10.3322/caac.21492.

Cancer Genome Atlas Research. 2014. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513(7517)**:202–209 DOI 10.1038/nature13480.

Cheng X, Yin H, Fu J, Chen C, An J, Guan J, Duan R, Li H, Shen H. 2019. Aggregate analysis based on TCGA: TTN missense mutation correlates with favorable prognosis in lung squamous cell carcinoma. *Journal of Cancer Research and Clinical Oncology* **145(4)**:1027–1035 DOI 10.1007/s00432-019-02861-y.

Cho JY, Lim JY, Cheong JH, Park YY, Yoon SL, Kim SM, Kim S-B, Kim H, Hong SW, Park YN, Noh SH, Park ES, Chu I-S, Hong WK, Ajani JA, Lee J-S. 2011. Gene expression signature-based prognostic risk score in gastric cancer. *Clinical Cancer Research* **17(7)**:1850–1857 DOI 10.1158/1078-0432.CCR-10-2180.

Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G, Noushmehr H. 2016. TCGAbiolinks:

an R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Research* **44(8)**:e71 DOI 10.1093/nar/gkv1507.

Craig JM, Bickmore WA. 1994. The distribution of CpG islands in mammalian chromosomes. *Nature Genetics* **7(3)**:376–382 DOI 10.1038/ng0794-376.

Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, Ye XS, Do I-G, Liu S, Gong L, Fu J, Jin JG, Choi MG, Sohn TS, Lee JH, Bae JM, Kim ST, Park SH, Sohn I, Jung S-H, Tan P, Chen R, Hardwick J, Kang WK, Ayers M, Hongyue D, Reinhard C, Loboda A, Kim S, Aggarwal A. 2015. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nature Medicine* **21(5)**:449–456 DOI 10.1038/nm.3850.

Deng G, Qu J, Zhang Y, Che X, Cheng Y, Fan Y, Zhang S, Na D, Liu Y, Qu X. 2017. Gastric cancer-derived exosomes promote peritoneal metastasis by destroying the mesothelial barrier. *FEBS Letters* **591(14)**:2167–2179 DOI 10.1002/1873-3468.12722.

Deng X, Xiao Q, Liu F, Zheng C. 2018. A gene expression-based risk model reveals prognosis of gastric cancer. *PeerJ* **6(4)**:e4204 DOI 10.7717/peerj.4204.

Fujita Y, Yoshioka Y, Ochiya T. 2016. Extracellular vesicle transfer of cancer pathogenic components. *Cancer Science* **107(4)**:385–390 DOI 10.1111/cas.12896.

Gui J, Li H. 2005. Penalized cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics* **21(13)**:3001–3008 DOI 10.1093/bioinformatics/bti422.

Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, Flagg K, Hou J, Zhang H, Yi S, Jafari M, Lin D, Chung C, Caughey BA, Li G, Dhar D, Shi W, Zheng L, Hou R, Zhu J, Zhao L, Fu X, Zhang E, Zhang C, Zhu J-K, Karin M, Xu R-H, Zhang K. 2017. DNA methylation markers for diagnosis and prognosis of common cancers. *Proceedings of the National Academy of Sciences of the United States of America* **114(28)**:7414–7419 DOI 10.1073/pnas.1703577114.

He W, Zhang H, Han F, Chen X, Lin R, Wang W, Qiu H, Zhuang Z, Liao Q, Zhang W, Cai Q, Cui Y, Jiang W, Wang H, Ke Z. 2017. CD155T/TIGIT Signaling Regulates CD8[+] T-cell Metabolism and Promotes Tumor Progression in Human Gastric Cancer. *Cancer Research* **77(22)**:6375–6388 DOI 10.1158/0008-5472.Can-17-0381.

Hu Y, He C, Liu JP, Li NS, Peng C, Yang-Ou YB, Yang X-Y, Lu N-H, Zhu Y. 2018. Analysis of key genes and signaling pathways involved in Helicobacter pylori-associated gastric cancer based on The Cancer Genome Atlas database and RNA sequencing data. *Helicobacter* **23(5)**:e12530 DOI 10.1111/hel.12530.

Hu S, Yin X, Zhang G, Meng F. 2019. Identification of DNA methylation signature to predict prognosis in gastric adenocarcinoma. *Journal of Cellular Biochemistry* **120(7)**:11708–11715 DOI 10.1002/jcb.28450.

Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA. 2007. The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biology* **8(9)**:R183 DOI 10.1186/gb-2007-8-9-r183.

Hughes LA, Melotte V, De Schrijver J, De Maat M, Smit VT, Bovee JV, French PJ, Van den Brandt PA, Schouten LJ, De Meyer T, Van Criekinge W, Ahuja N, Herman JG, Weijenberg MP, Van Engeland M. 2013. The CpG island methylator phenotype: what's in a name? *Cancer Research* **73(19)**:5858–5868 DOI 10.1158/0008-5472.Can-12-4306.

Jia D, Lin W, Tang H, Cheng Y, Xu K, He Y, Geng W, Dai Q. 2019. Integrative analysis of DNA methylation and gene expression to identify key epigenetic genes in glioblastoma. *Aging* **11(15)**:5579–5592 DOI 10.18632/aging.102139.

Jung T, Castellana D, Klingbeil P, Hernández IC, Vitacolonna M, Orlicky DJ, Roffler SR, Brodt P, Zöller M. 2009. CD44v6 dependence of premetastatic niche preparation by exosomes. *Neoplasia* **11(10)**:1093–1105 DOI 10.1593/neo.09822.

Kang GH, Lee S, Kim WH, Lee HW, Kim JC, Rhyu MG, Ro JY. 2002. Epstein-barr virus-positive gastric carcinoma demonstrates frequent aberrant methylation of multiple genes and constitutes CpG island methylator phenotype-positive gastric carcinoma. *American Journal of Pathology* **160(3)**:787–794 DOI 10.1016/s0002-9440(10)64901-2.

Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, Liu XQ, Sher X, Jung H, Lee M, Lee S, Park SH, Park JO, Park YS, Lim HY, Lee H, Choi M, Talasaz A, Kang PS, Cheng J, Loboda A, Lee J, Kang WK. 2018. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nature Medicine* **24(9)**:1449–1458 DOI 10.1038/s41591-018-0101-z.

Li H, Adachi Y, Yamamoto H, Min Y, Ohashi H, Ii M, Arimura Y, Endo T, Lee C-T, Carbone DP, Imai K, Shinomura Y. 2011. Insulin-like growth factor-I receptor blockade reduces tumor angiogenesis and enhances the effects of bevacizumab for a human gastric cancer cell line. *MKN45 Cancer* **117(14)**:3135–3147 DOI 10.1002/cncr.25893.

Li X, Pasche B, Zhang W, Chen K. 2018a. Association of MUC16 mutation with tumor mutation load and outcomes in patients with gastric cancer. *JAMA Oncology* **4(12)**:1691–1698 DOI 10.1001/jamaoncol.2018.2805.

Li G, Xu W, Zhang L, Liu T, Jin G, Song J, Wu J, Wang Y, Chen W, Zhang C, Chen X, Ding Z, Zhu P, Zhang B. 2019. Development and validation of a CIMP-associated prognostic model for hepatocellular carcinoma. *EBioMedicine* **47**:128–141 DOI 10.1016/j.ebiom.2019.08.064.

Li Q, Zheng ZC, Ni CJ, Jin WX, Jin YX, Chen Y, Zhang XH, Chen ED, Cai YF. 2018b. Correlation of cystatin E/M with clinicopathological features and prognosis in triple-negative breast cancer. *Annals of Clinical and Laboratory Science* **48(1)**:40–44.

Liu X-S, Lin X-K, Mei Y, Ahmad S, Yan C-X, Jin H-L, Yu H, Chen C, Lin C-Z, Yu J-R. 2019. Regulatory T cells promote overexpression of Lgr5 on gastric cancer cells via TGF-beta1 and confer poor prognosis in gastric cancer. *Frontiers in Immunology* **10**:1741 DOI 10.3389/fimmu.2019.01741.

Liu H, Wu X, Wang S, Deng W, Zan L, Yu S. 2013. In vitro repolarized tumor macrophages inhibit gastric tumor growth. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics* **20(7)**:275–280 DOI 10.3727/096504013X13639794277563.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15(12)**:550 DOI 10.1186/s13059-014-0550-8.

Mathiak M, Warneke VS, Behrens H-M, Haag J, Böger C, Krüger S, Röcken C. 2017. Clinicopathologic characteristics of microsatellite instable gastric carcinomas revisited: urgent need for standardization. *Applied Immunohistochemistry & Molecular Morphology* **25(1)**:12–24 DOI 10.1097/pai.0000000000000264.

Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. 2018. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Research* **28(11)**:1747–1756 DOI 10.1101/gr.239244.118.

Miyatani K, Saito H, Murakami Y, Watanabe J, Kuroda H, Matsunaga T, Fukumoto Y, Osaki T, Nakayama Y, Umekita Y, Ikeguchi M. 2016. A high number of IgG4-positive cells in gastric

cancer tissue is associated with tumor progression and poor prognosis. *Virchows Archiv* **468(5)**:549–557 DOI 10.1007/s00428-016-1914-0.

Moarii M, Reyal F, Vert JP. 2015. Integrative DNA methylation and gene expression analysis to assess the universality of the CpG island methylator phenotype. *Human Genomics* **9(1)**:26 DOI 10.1186/s40246-015-0048-9.

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. 2015. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* **12(5)**:453–457 DOI 10.1038/nmeth.3337.

Oue N, Oshimo Y, Nakayama H, Ito R, Yoshida K, Matsusaki K, Yasui W. 2003. DNA methylation of multiple genes in gastric carcinoma: association with histological type and CpG island methylator phenotype. *Cancer Science* **94(10)**:901–905 DOI 10.1111/j.1349-7006.2003.tb01373.x.

Park S-Y, Kook MC, Kim YW, Cho N-Y, Jung N, Kwon H-J, Kim T-Y, Kang GH. 2010. CpG island hypermethylator phenotype in gastric carcinoma and its clinicopathological features. *Virchows Archiv* **457(4)**:415–422 DOI 10.1007/s00428-010-0962-0.

Peinado H, Alečković M, Lavotshkin S, Matei I, Costa-Silva B, Moreno-Bueno G, Hergueta-Redondo M, Williams C, García-Santos G, Ghajar CM, Nitadori-Hoshino A, Hoffman C, Badal K, Garcia BA, Callahan MK, Yuan J, Martins VR, Skog J, Kaplan RN, Brady MS, Wolchok JD, Chapman PB, Kang Y, Bromberg J, Lyden D. 2012. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nature Medicine* **18(6)**:883–891 DOI 10.1038/nm.2753.

Polom K, Marano L, Marrelli D, De Luca R, Roviello G, Savelli V, Tan P, Roviello F. 2018. Meta-analysis of microsatellite instability in relation to clinicopathological characteristics and overall survival in gastric cancer. *British Journal of Surgery* **105(3)**:159–167 DOI 10.1002/bjs.10663.

Powell AGMT, Soul S, Christian A, Lewis WG. 2018. Meta-analysis of the prognostic value of CpG island methylator phenotype in gastric cancer. *British Journal of Surgery* **105(2)**:e61–e68 DOI 10.1002/bjs.10742.

Rauch TA, Wang Z, Wu X, Kernstine KH, Riggs AD, Pfeifer GP. 2012. DNA methylation biomarkers for lung cancer. *Tumour Biology* **33(2)**:287–296 DOI 10.1007/s13277-011-0282-2.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43(7)**:e47 DOI 10.1093/nar/gkv007.

Sarrouilhe D, Mesnil M. 2019. Serotonin and human cancer: a critical view. *Biochimie* **161**:46–50 DOI 10.1016/j.biochi.2018.06.016.

Sato Y, Inokuchi M, Takagi Y, Kojima K. 2019. IGFBP1 is a predictive factor for haematogenous metastasis in patients with gastric cancer. *Anticancer Research* **39(6)**:2829–2837 DOI 10.21873/anticanres.13411.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13(11)**:2498–2504 DOI 10.1101/gr.1239303.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102(43)**:15545–15550 DOI 10.1073/pnas.0506580102.

Sun P, Zhu X, Shrubsole MJ, Ness RM, Hibler EA, Cai Q, Long J, Chen Z, Li G, Hou L, Smalley WE, Edwards TL, Giovannucci E, Zheng W, Dai Q. 2017. Genetic variation in SLC7A2 interacts with calcium and magnesium intakes in modulating the risk of colorectal polyps. *Journal of Nutritional Biochemistry* **47**:35–40 DOI 10.1016/j.jnutbio.2017.04.016.

Tan PY, Chang CW, Chng KR, Wansa KD, Sung WK, Cheung E. 2012. Integration of regulatory networks by NKX3-1 promotes androgen-dependent prostate cancer survival. *Molecular and Cellular Biology* **32(2)**:399–414 DOI 10.1128/MCB.05958-11.

Tang H, Deng M, Tang Y, Xie X, Guo J, Kong Y, Ye F, Su Q, Xie X. 2013. miR-200b and miR-200c as prognostic factors and mediators of gastric cancer cell progression. *Clinical Cancer Research* **19(20)**:5602–5612 DOI 10.1158/1078-0432.Ccr-13-1326.

Ueki T, Toyota M, Sohn T, Yeo CJ, Issa JP, Hruban RH, Goggins M. 2000. Hypermethylation of multiple genes in pancreatic adenocarcinoma. *Cancer Research* **60(7)**:1835–1839.

Vaissiere T, Sawan C, Herceg Z. 2008. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutation Research* **659(1–2)**:40–48 DOI 10.1016/j.mrrev.2008.02.004.

Warneke VS, Behrens HM, Hartmann JT, Held H, Becker T, Schwarz NT, Rocken C. 2011. Cohort study based on the seventh edition of the TNM classification for gastric cancer: proposal of a new staging system. *Journal of Clinical Oncology* **29(17)**:2364–2371 DOI 10.1200/JCO.2010.34.4358.

Wei L, Jin Z, Yang S, Xu Y, Zhu Y, Ji Y. 2018. TCGA-assembler 2: software pipeline for retrieval and processing of TCGA/CPTAC data. *Bioinformatics* **34(9)**:1615–1617 DOI 10.1093/bioinformatics/btx812.

Wilkerson MD, Hayes DN. 2010. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26(12)**:1572–1573 DOI 10.1093/bioinformatics/btq170.

Ye F, Tang H, Liu Q, Xie X, Wu M, Liu X, Chen X, Xie X. 2014. miR-200b as a prognostic factor in breast cancer targets multiple members of RAB family. *Journal of Translational Medicine* **12**:17 DOI 10.1186/1479-5876-12-17.

Yin S, Huang J, Li Z, Zhang J, Luo J, Lu C, Xu H, Xu H. 2017. The prognostic and clinicopathological significance of tumor-associated macrophages in patients with gastric cancer: a meta-analysis. *PLOS ONE* **12(1)**:e0170042 DOI 10.1371/journal.pone.0170042.

Yu G, Wang LG, Han Y, He QY. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* **16(5)**:284–287 DOI 10.1089/omi.2011.0118.

Zan XY, Li L. 2019. Construction of lncRNA-mediated ceRNA network to reveal clinically relevant lncRNA biomarkers in glioblastomas. *Oncology Letters* **17(5)**:4369–4374 DOI 10.3892/ol.2019.10114.

Zhao W, Chen R, Zhao M, Li L, Fan L, Che XM. 2015. High glucose promotes gastric cancer chemoresistance in vivo and in vitro. *Molecular Medicine Reports* **12(1)**:843–850 DOI 10.3892/mmr.2015.3522.

Zhao L, Jiang L, He L, Wei Q, Bi J, Wang Y, Yu L, He M, Zhao L, Wei M. 2019. Identification of a novel cell cycle-related gene signature predicting survival in patients with gastric cancer. *Journal of Cellular Physiology* **234(5)**:6350–6360 DOI 10.1002/jcp.27365.