

Article

Development and Experimental Validation of a Novel Prognostic Signature for Gastric Cancer

Chengcheng Liu ^{1,2,3,†}, Yuying Huo ^{1,†} , Yansong Zhang ⁴, Fumei Yin ⁵ , Taoyu Chen ⁶, Zhenyi Wang ^{2,7} , Juntao Gao ^{2,7,*}, Peng Jin ⁸, Xiangyu Li ^{1,*} , Minglei Shi ^{2,3} and Michael Q. Zhang ^{2,3,7,9}

¹ School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China

² MOE Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic & Systems Biology, BNRist, Tsinghua University, Beijing 100084, China

³ School of Medicine, Tsinghua University, Beijing 100084, China

⁴ School of Life Sciences, Peking University, Beijing 100871, China

⁵ Medical School of Chinese PLA, Beijing 100853, China

⁶ Department of Bioinformatics, School of Basic Medicine, Peking University Health Center, Beijing 100191, China

⁷ Department of Automation, Tsinghua University, Beijing 100084, China

⁸ Senior Department of Gastroenterology, First Medical Center of Chinese PLA General Hospital, Beijing 100036, China

⁹ Department of Biological Sciences, Center for Systems Biology, University of Texas at Dallas, Richardson, TX 75080, USA

* Correspondence: jtgao@mail.tsinghua.edu.cn (J.G.); lixiangyu@bjtu.edu.cn (X.L.)

† These authors contributed equally to this work.

Simple Summary: Gastric cancer (GC) accounts for a considerable amount of morbidity and mortality worldwide. This study developed and experimentally validated a prognostic risk gene signature (PRGS). This is a stable and robust signature for assessing the prognosis of gastric cancer. We performed multiple analyses through consensus clustering and binary classification to assess the robustness of the PRGS in other independent datasets. Additionally, this PRGS exhibited a superior accuracy compared to most traditional clinical markers, including molecular features, and other published signatures. Besides, we also detected the tumor purity, immune cell infiltration, and oncogenic mutation status of high- and low-PRGS groups.

Abstract: Background: Gastric cancer is a malignant tumor with high morbidity and mortality. Therefore, the accurate recognition of prognostic molecular markers is the key to improving treatment efficacy and prognosis. Methods: In this study, we developed a stable and robust signature through a series of processes using machine-learning approaches. This PRGS was further experimentally validated in clinical samples and a gastric cancer cell line. Results: The PRGS is an independent risk factor for overall survival that performs reliably and has a robust utility. Notably, PRGS proteins promote cancer cell proliferation by regulating the cell cycle. Besides, the high-risk group displayed a lower tumor purity, higher immune cell infiltration, and lower oncogenic mutation than the low-PRGS group. Conclusions: This PRGS could be a powerful and robust tool to improve clinical outcomes for individual gastric cancer patients.

Keywords: gastric cancer; prognostic signature; tumor microenvironment; oncogenic mutation; clinical outcomes; machine learning



Citation: Liu, C.; Huo, Y.; Zhang, Y.; Yin, F.; Chen, T.; Wang, Z.; Gao, J.; Jin, P.; Li, X.; Shi, M.; et al. Development and Experimental Validation of a Novel Prognostic Signature for Gastric Cancer. *Cancers* **2023**, *15*, 1610. <https://doi.org/10.3390/cancers15051610>

Academic Editor: Alain P. Gobert

Received: 11 November 2022

Revised: 18 January 2023

Accepted: 16 February 2023

Published: 5 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gastric cancer (GC) is a leading cause of cancer morbidity and mortality worldwide [1]. According to GLOBOCAN 2020, there were about 1,089,103 new cases of gastric cancer patients (5.6% of the total cancer burden) [1]. Disease progression and a lack of effective treatment cause most of the mortality [2]. Therefore, preventing “high-risk” GC is the

key to improving clinical outcomes. The tumor, node, metastasis (TNM) classification and the American Joint Committee on Cancer (AJCC) classification [3] are commonly used methods to assess the risk and treatment demand for patients in the clinical setting. Nevertheless, due to the limitation of the current grading system, it cannot provide the best clinical treatment for patients. For example, in the clinic, the decision of adjuvant chemotherapy (ACT) is mainly dependent on the clinical–pathological stage rather than molecular biological characteristics [4]. This approach is insufficient and may result in latent overtreatment or undertreatment. Hence, in the era of individualized treatment, it is imperative to identify effective biomarkers to optimize the prognosis of GC.

The ideal biomarker should have a consistent expression within and between tumor tissues to perform stably among all patients. Hence, a multigene signature may be an effective approach for addressing this heterogeneity. To date, there are three categories of clinically important GC markers, CEA, CA19-9, and CA72-4, with positive rates of 21.1%, 27.8%, and 30.0%, respectively [5]. However, they are present in a limited number of patients with GC, and the sensitivity and specificity of these biomarkers are not sufficient. With the development of bioinformatics technology, many prognostic biomarkers have been published [6–8]. Unfortunately, most of the identified biomarkers failed in the validation.

GC is a heterogeneous malignant disease. Histologically, the human gastric mucosa can be divided into three zones, i.e., the cardiac zone, the fundus/corpus zone, and the pyloric zone [9,10]. These zones differ vastly in their histology, regeneration rates, and profiles [11]. In addition, there are different morbidity and mortality rates among GC in these three zones. It is believed that GC patients with the GC localized in the cardiac zone have the worst prognosis. As the patient's age progresses, the location of the gastric cancer moves upward, and the incidence of it occurring in the cardiac zone increases [12]. The composition of cell types within them is also discrepant. For instance, the quantities of gastric parietal cells are considerably different in these three zones. Namely, parietal cells account for 25%, 50–100%, and 0–1% of all cells in the cardiac zone, the fundus/corpus zone, and the pyloric zone, respectively [12]. Of note, most researchers who investigate gastric cancer biomarkers have ignored the differences between these three zones and analyzed them as if all gastric cancers are of one single area. Hence, it is essential to construct gene co-expression networks and select prognostic-related genes separately for GC data in these three zones.

In this work, we attempted to computationally develop and experimentally validate a prognostic risk gene signature with 1226 GC patients from three independent public datasets (Tables S1–S4), a gastric cancer cell line, and several clinical samples to assess the prognosis, tumor growth, and molecular characterization of GC. A multi-step procedure of machine-learning approaches was performed to develop and cross-validate the prognostic risk gene signature (PRGS) model based on the co-expression networks of GC in the cardiac zone, the fundus/corpus zone, and the pyloric zone. This PRGS may help optimize precision treatment and further improve the clinical outcomes of GC patients.

2. Materials and Methods

2.1. Data Acquisition and Processing

In this study, we used three independent public datasets, including 1226 GC patients obtained from the UCSC Toil Recompute Compendium of The Cancer Genome Atlas TARGET and Genotype Tissue Expression project datasets (TCGA target GTEX, primary_site = stomach) [13] and the Gene Expression Omnibus (GEO) (Tables S1–S4). These datasets (TCGA target GTEX, GSE66229, and GSE15459), which encompass complete overall survival (OS) information, were used. Among these, we converted the RNA-seq raw read count from the TCGA target GTEX database to transcripts per kilobase million (TPM) and then log₂-transformed. And the data has been removed batch effects among these patients [14]. We retrieved GSE15459 and GSE66229 from the Affymetrix[®] GPL570 platform (Human Genome U133 Plus 2.0 Array). We selected the most highly expressed

probe for each gene to ensure reliable results in consensus clustering and we reserved all probes of each gene to ensure accurate results in binary classification.

The ATAC-seq and somatic variation data were obtained from the database of the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/>, accessed on 1 February 2022).

2.2. Human Tissue Specimens

In this study, the collection of gastric tissues was approved by the Department of Gastroenterology, Seventh Medical Center of Chinese PLA General Hospital, Beijing, China, on 1 June 2022. Overall, one advanced GC sample, an early GC sample, and a normal gastric sample were collected. All patients provided written informed consent, and the ethics committee of the PLA General Hospital also approved our research.

2.3. Differentially Expressed Analysis and Weighted Correlation Network Analysis

The DEGs between cancer and normal samples were detected by edgeR [15]. The genes with an absolute \log_2 (fold change) ≥ 1.5 were considered to be significant differentially expressed genes (DEGs) between tumor and normal tissues. The volcano plots of upregulated or downregulated genes were generated by the ggplot2 R 4.0 package. The Venn diagram was plotted by the VennDiagram R 4.0 package.

The co-expression gene networks were constructed by the WGCNA package [16]. To recognize modules of significantly correlated clusters in the cardiac zone, the fundus/corpus zone, and the pyloric zone, the module that displayed the highest correlation was selected.

2.4. Construction of the Co-Expression Network

The brown, purple, and salmon modules of the result of WGCNA were used for the edges, signifying the correlations in the cardiac zone, the fundus/corpus zone, and the pyloric zone, respectively. The filter criterion of a weight value was set to greater than 0.02.

A total of 25,350, 1819, and 706 edges and 308, 69, and 54 nodes correlated with these zones were separately obtained and processed in Cytoscape 3.8.1 [17]. To construct the network and select the hub genes, the Cytoscape software was used with the CytoHubba method [18].

2.5. PRGS Signature Generation

LASSO is a regularization and dimensionality reduction technique combined with Cox models, which can be applied in biomarker screening [19]. To identify the hub genes, the top 25 genes of the co-expression network from the cardiac zone, the fundus/corpus zone, and the pyloric zone by Cytoscape were inputted into the LASSO–Cox regression [19]. To ensure the stability of the gene and the model, this procedure was repeated 1000 times. Then, the regression coefficients of each gene were calculated (Table S5). Ultimately, genes with positive regression coefficients were selected. Four genes (APOB, VCAN, ABCA6, and CTSF) whose p -values in the Kaplan–Meier analysis were also < 0.05 were identified to generate the PRGS model. The PRGS risk score of GC patients was calculated by the formula: $\text{PRGS score} = \sum_{k=1}^4 \binom{4}{1} \beta_k \times \text{RNA}_k$.

In this formula, k means the genes in the PRGS—we used $k = 1, 2, 3, 4$ to index the genes in the PRGS—the β value was the multivariate Cox regression coefficient; and RNA_k means the expression level of gene k of each patient.

2.6. Chromatin Accessibility Analysis

The peak regions on chromosomes were shown by the R package chromosome locator. Using the R-packaged ChIPseeker, the alignment can be mapped to peaks in the TSS region to build a signature matrix. Peak position annotation and motif enrichment analyses were performed using HOMER (V4.10). In the range of ± 1 kb around the TSS, the peak overlapping the gene initiator was considered as the peak of gene regulation.

2.7. Consensus Clustering

This process was accomplished through the ConsensusClusterPlus package [20]. We subsampled 80% of the samples, and then used the k-means algorithm to divide each subsample into k (k = 9) according to the Euclidean distance. This procedure was repeated 1000 times. Finally, using the PAC with the smallest k value, the optimal cluster (k = 2) was derived.

2.8. Binary Classification

- (1) We pre-processed the GSE66229 data and selected the top 5 genes from the three generated co-expression networks of the three zones' expression spectrum matrices, then divided the data into normal and tumor samples. We also generated the GSE66229 data and generated PRGS, CEA [21], and GCscore [6] expression spectrum matrices.
- (2) We performed 5-fold cross validation using the logistic regression classifier (LR) and the random forest classifier (RF) on the GSE66229 and computed the ROC curve for each fold. We then calculated the means of every fitting curve to generate the plot. We then divided the data into five subsets based on the sample tags "Tumor/Normal", "Stage 1/Normal", "Stage 2/Normal", "Stage 3/Normal", and "Stage 4/Normal". Details of the data are shown in Table S4.

The parameters were set as follows: RF, max_depth = 5, n_estimators = 5, random_state = 123; LR, solver = "liblinear", penalty = "l2", C = 1.0. The rest were set as default. We trained and validated each fold and calculated the FPR, TPR, and AUC. We fitted each result into the ROC curve using np.interp and took the means of all fitting curves to craft the figure.

2.9. Haematoxylin–Eosin (HE) and Immunohistochemistry (IHC)

For HE staining, a 4% paraformaldehyde solution was used to fix GC tissues. After 24 h, the GC tissues were dewaxed in xylene, dehydrated in ethanol, and subjected to hematoxylin staining (5 min), then dehydrated in eosin solution (10 s), dehydrated in graded alcohol, removed in xylene, and sealed with neutral glue. They were observed and photographed with a microscope (Olympus, Tokyo, Japan).

For IHC, 5% bovine serum was used to incubate sections. Then, they were mixed with primary antibodies (anti-APOB, 1:100; anti-VCAN, 1:100; anti-ABCA6, 1:200; and anti-CTSF, 1:200, Abcam, UK) and a secondary antibody (1:800, Abcam, Cambridge, UK). Then, they were stained with a DAB kit and photographed with an optical microscope (Olympus, Japan).

2.10. Cell Culture, Transfection, and Immunostaining

GES-1 gastric cancer cells (a gift from Prof. Jun Qin) were cultured with 5% CO₂ at 37 °C in Dulbecco's modified Eagle's medium (DMEM) (Thermo Fisher, Waltham, MA, USA, C11995500BT) with 20% FBS and 100 mg/mL penicillin/streptomycin.

For the siRNA treatment, Lipofectamine™ 3000 Transfection Reagent (Thermo Fisher, L3000001) was used to transfect cells using a standard procedure.

For immunostaining, poly-L-lysine-coated coverslips were used to culture cells for 72 h to perform the siRNA transfection; 4% paraformaldehyde was used to fix the cells for 10 min, and then the cells were washed with PBSTr buffer (PBS plus 0.1% Triton X-100, Sigma-Aldrich, St. Louis, MO, USA T8787). The cells were incubated with the anti-phosphorylated histone 3 (pH3) antibody overnight at 4 °C. The cells were incubated with a secondary antibody at room temperature and washed with PBSTr buffer. Then, 0.01 mg/mL DAPI and the Vectashield Antifade Mounting Medium (Vector Laboratories, San Francisco, CA, USA H-1200) were used to incubate and mount the cells, respectively.

2.11. Flow Cytometric Analysis

For flow cytometric analysis of the cell cycle with propidium iodide (PI) staining, a standard procedure was used [22].

2.12. Quantitative Real-Time PCR

The Eastep Super Total RNA Extraction Kit (Promega, Madison, WI, USA LS1040) was used to extract the total RNA of siRNA-treated GES-1 cells. The Eastep RT Master Mix Kit (Promega, LS2050) was used to synthesize the cDNA. An Applied Biosystems 7500 real-time PCR system (Thermo Fisher) was used to perform real-time PCRs with the PowerUp SYBR Green Master Mix (Thermo Fisher, A25776). The comparative CT method and Graphpad Prism 8 (GraphPad Software, La Jolla, CA, USA) were used to analyze the data. All experiments were repeated three times.

2.13. Cell Infiltration Estimation

ESTIMATE [23] and CIBERSORT [24] were used to evaluate immune infiltrates. The immune scores, stromal scores, and tumor purity were calculated by the ESTIMATE algorithm. CIBERSORT [24] was used to analyze the levels of infiltrating immune and stromal cells.

2.14. Tumor Mutation Status Analysis

Maftools was used to calculate significantly mutated genes ($p < 0.05$) between the low- and high-PRGS groups [25]. A one-sided z-test and two-sided Chi-square test were used to calculate the statistical test for the proportion of mutations ($p < 0.05$).

2.15. Functional Enrichment Analysis

A functional enrichment analysis was performed on DEGs and peak-related genes. The possible peak-related genes of GO/KEGG enrichment were used to analyze the ClusterProfiler package in R [26].

Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database were used to annotate the tumor-related pathways. The gene set variation (GSVA) method [27] was used to enrich the pathways. We used the enrichment score of the GSVA to obtain the expression pathway of the PRGS.

3. Results

3.1. Construction of the Gene Modules

The workflow of this work is shown in Figure 1.

First, we identified the differentially expressed genes (DEGs) in the TCGA target GTEx with the edgeR package independently to screen DEGs in normal samples and GC samples from the cardiac zone, the fundus/corpus zone, and the pyloric zone. For GC in the cardiac zone, 1498 upregulated and 883 downregulated DEGs were identified. A total of 1661 upregulated and 826 downregulated DEGs were found between normal and GC samples from the fundus/corpus zone. There were 1542 upregulated and 889 downregulated DEGs for GC in the pyloric zone compared with normal samples (Table S5). The distributions of DEGs are shown by volcano plots (Figure 2A). There were 1553 DEGs shared among the GC samples from the three zones (Figure 2B). We adopted GO and KEGG enrichment analysis methods to investigate the annotation of the DEGs. The DEGs of GC samples from the cardiac zone were mainly enriched in sensory perception (GO:0050907), the collagen-containing extracellular matrix (GO:0062023), the sarcomere (GO:0030017), and olfactory transduction. Sensory perception (GO:0050907), the ion channel complex (GO:0034702), olfactory receptor activity (GO:0004984), and olfactory transduction were detected in GC samples from the fundus/corpus zone. In addition, the GC samples from the pyloric zone were highly associated with digestion (GO:0007586), the collagen-containing extracellular matrix (GO:0062023), receptor–ligand activity (GO:0048018), and neuroactive ligand–receptor interactions (Figure S1A,B).

To further identify the gene modules related to GC in the cardiac zone, the fundus/corpus zone, and the pyloric zone, the WGCNA method [16] was applied. We assured a scale-free network (soft threshold = 3) with a high scale independence and a low mean connectivity (near 0) (Figure S1C). DEGs in the GC samples from the cardiac zone, the

fundus/corpus zone, and the pyloric zone were respectively divided into 22 modules by a cluster analysis (Figure 2C). The brown module related to GC in the pyloric zone was the most significant ($cor = 0.76, p = 2 \times 10^{-112}$). For GC in the cardiac zone and the fundus/corpus zone, the purple and salmon modules were chosen according to the correlation ($cor = 0.22, 0.33; p = 3 \times 10^{-8}, 1 \times 10^{-16}$).

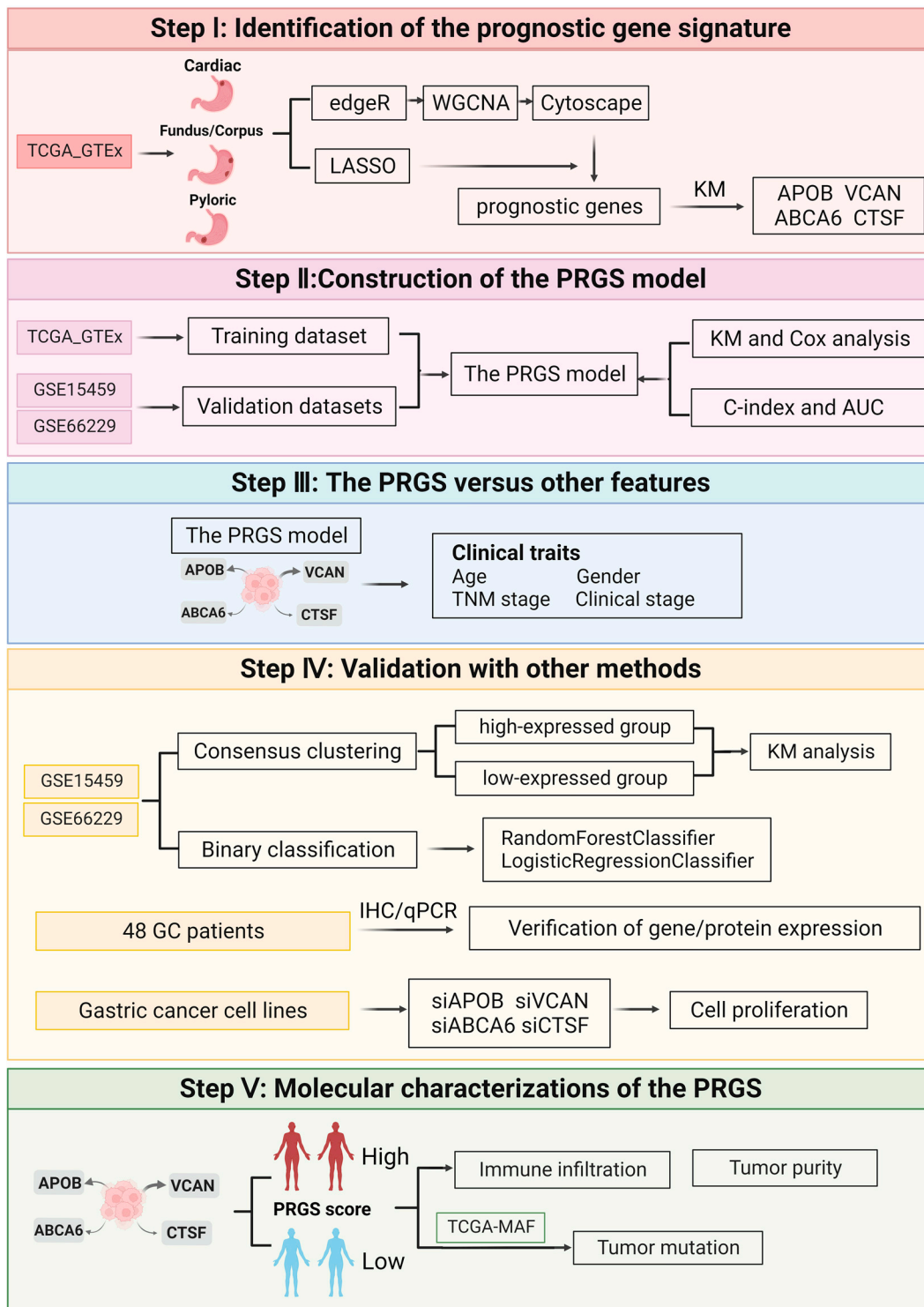


Figure 1. The overall design of this study.

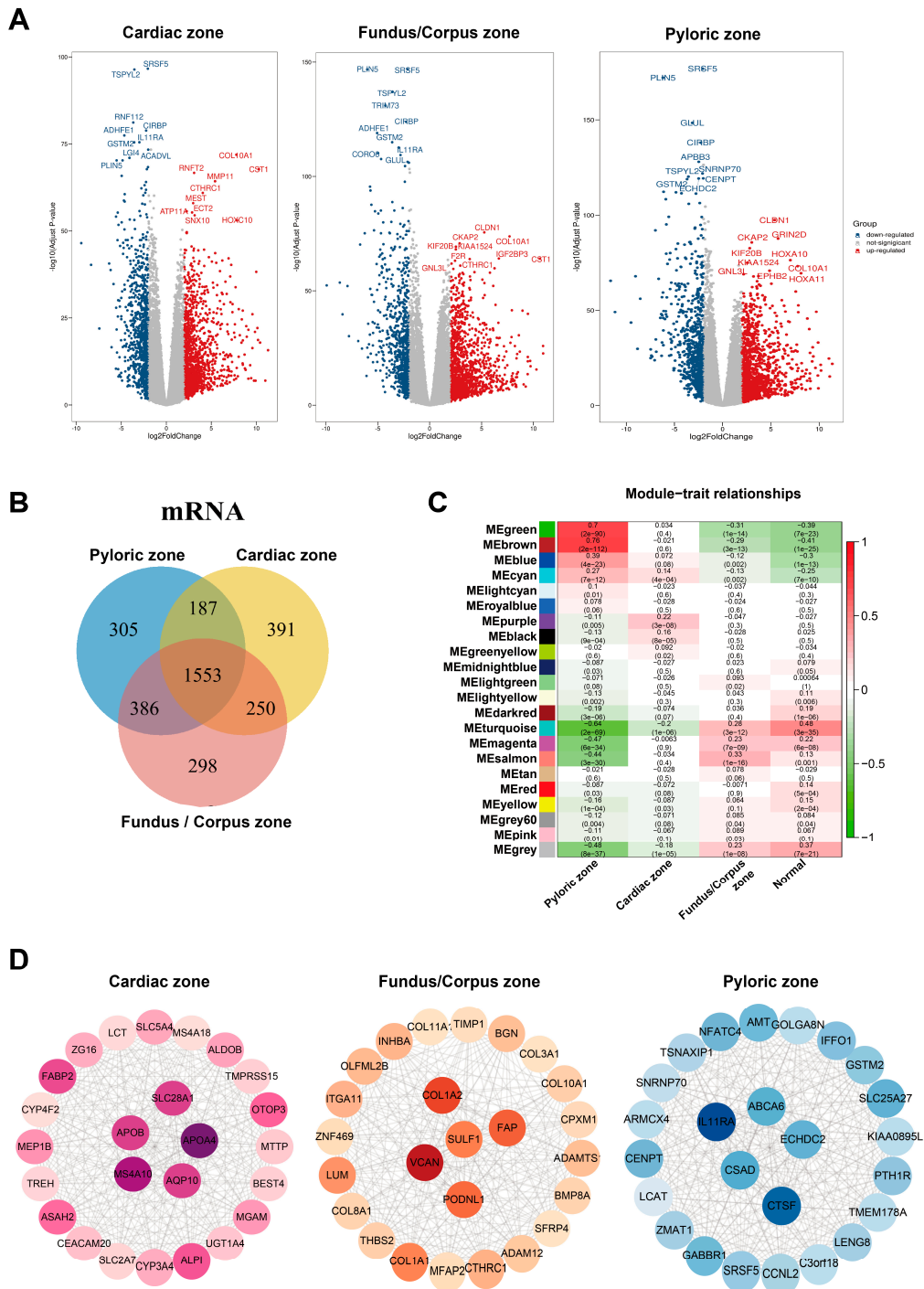


Figure 2. Identified co-expression networks of the cardiac zone, the fundus/corpus zone, and the pyloric zone. (A) Volcano plots of DEGs in the cardiac zone, the fundus/corpus zone, and the pyloric zone. Red and blue spots represent significant up- and down-regulated RNAs, respectively. The remark for the gene symbols represents the significant up- and down-regulated RNAs. (B) Venn diagram of the DEGs in the cardiac zone, the fundus/corpus zone, and the pyloric zone. Different colors indicate GC in different zones. The numbers represent the common DEGs among GC in different zones. (C) Correlation analysis between module eigengenes and the three zones. (D) Co-expression networks of the cardiac zone, the fundus/corpus zone, and the pyloric zone.

We selected the brown, purple, and salmon modules for the edges, representing the correlations in GC in the cardiac zone, the fundus/corpus zone, and the pyloric zone,

respectively, by the WGCNA algorithm [16]. The Cytoscape software [16] was used to visualize the gene co-expression networks [17], and Cytohubba was used to select hub genes [18] (Figure 2D and Table S6). We also performed a GO and KEGG enrichment analysis of the purple (correlated with the cardiac zone), salmon (correlated with the pyloric zone), and brown (correlated with the fundus/corpus zone) modules (Figure S2A–F).

We performed a survival analysis of the genes from the co-expression networks of the three zones on GC patients in the TCGA target GTEx and GSE66229 datasets (GSE15459 did not have information on the three zones). The results indicated that the top five genes (APOA4, MS4A10, SLC28A1, AQP10, and APOB) from the cardiac zone co-expression network correlated most significantly with the outcomes of GC patients in the cardiac zone compared to GC patients of the other two zones (Figure S3A,B). The same was true for patients with GC in the fundus/corpus zone. The top five genes (VCAN, COL1A2, FAP, PODNL1, and SULF1) from the fundus/corpus zone co-expression network also displayed the vastest correlation with the fundus/corpus zone GC patients compared to GC patients of the other two zones (Figure S3C,D). Meanwhile, we performed binary classification using the logistic regression classifier (LR) and random forest classifier (RF) with the hub genes (APOA4, MS4A10, SLC28A1, AQP10, and APOB) of GC in the cardiac zone as the feature genes to predict whether patients have GC, and we achieved the highest AUC in patients with GC of the cardiac zone compared to the other two zones in the GSE66229 datasets (Figure S3E). The same was true for the hub genes (VCAN, COL1A2, FAP, PODNL1, and SULF1) of GC patients in the fundus/corpus zone (Figure S3F). These results further indicated that discriminating among different zones of GC is of great importance.

3.2. Construction and Cross-Validation of the PRGS Model in Gastric Cancer Cohorts

To further determine the prognostic genes related to the three zones of GC, we continued to generate predictive genes using the TCGA target GTEx and cross-validated these genes with two independent datasets (GSE66229 and GSE15459). Based on the expression profiles of 25 genes correlated with GC in the cardiac zone, the fundus/corpus zone, and the pyloric zone, a LASSO–Cox regression analysis [19] generated the predictive genes. This process was repeated 1000 times with the glmnet R package to ensure the stability of the gene [28]. In the LASSO regression, the partial likelihood of deviance reached the minimum value to obtain the optimal λ (Figure S4A). Six, four, and three genes from GC in the cardiac zone, the fundus/corpus zone, and the pyloric zone, respectively, with positive LASSO coefficients were subjected to the log-rank test and the Kaplan–Meier curve (Figure 3A), which identified a final set of four genes. There were four genes with the maximum coefficient of GC in the three zones, including APOB (p -value = 0.028, FDR = 0.0042) correlated with GC in the cardiac zone, VCAN (p -value = 9.2×10^{-25} , FDR = 0.0004) correlated with GC in the fundus/corpus zone, and ABCA6 (p -value = 0.0028, FDR = 0.0042) as well as CTSF (p -value = 0.023, FDR = 0.023) correlated with GC in the pyloric zone. For each of these four genes, the increased expression level was vastly associated with a worse OS for gastric cancer patients (Figure 3A). Besides, these four genes (APOB, VCAN, ABCA6, and CTSF) are included in the DEG lists in the GC data of respective zones in dataset GSE66229.

We next depicted the PRGS expression level at a single-cell resolution. We analyzed the scRNA-seq data derived from 29 gastric cancer and 11 normal gastric tissues [29]. After the quantity control and removal batch effect, we obtained a total of 200,954 cells majorly comprising lymphoid cells (CD8A and KLRD1 positive), plasma (TNFRSF17 positive), epithelial cells (CDH1 positive), macrophages (CD163 positive), fibroblasts (FN1 and LUM positive), B cells (MS4A1 positive), mast cells (KIT positive), and pericytes (NOTCH3 positive) (Figures S4B–E and S5A). As shown in Figure S5B, VCAN was expressed in macrophages and fibroblasts; ABCA6 and CTSF were mainly expressed in fibroblasts; and APOB was expressed in epithelial cells (Figure S5B). We also observed that GC samples in different Lauran classifications (intestinal, diffuse, and mixed types) exhibited higher PRGS scores than normal samples (Figure S6A–D). We computed global PRGS scores for

all cell types and found that fibroblasts had the highest PRGS scores compared to other cell types (Figure S6D).

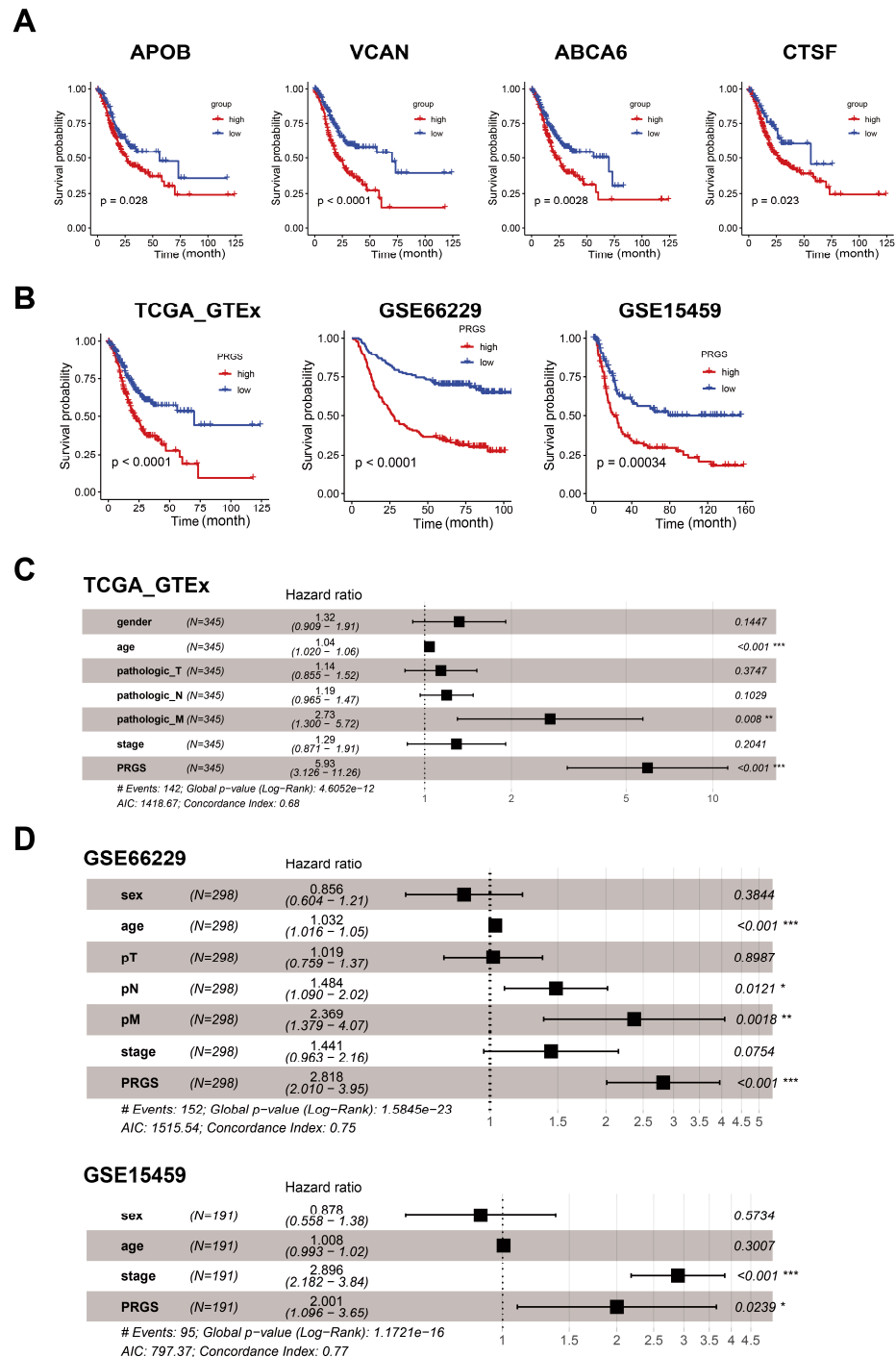


Figure 3. Cox and survival analysis of the PRGS model. (A) Kaplan–Meier curves of overall survival (OS) according to the PRGS in TCGA target GTEx. (B) Kaplan–Meier curves of OS according to the PRGS in TCGA target GTEx (log-rank test: $p = 7.75 \times 10^{-5}$); GSE66229 (log-rank test: $p = 2.11 \times 10^{-11}$); and GSE15459 (log-rank test: $p = 0.00034$). (C) Multivariate Cox regression of PRGS regarding OS in TCGA target GTEx ($n = 345$). (D) Multivariate Cox regression of PRGS regarding OS in GSE66229 ($n = 298$) and GSE15459 ($n = 191$). Statistic tests: two-sided Wald test. Data are presented as hazard ratio (HR) \pm 95% confidence interval (CI). * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

We also identified the chromatin accessibility of these four genes; thus, we analyzed the relative enriched proportions of coding regions, intergenic regions, introns, exons, and upstream and downstream regions (Figure S7A–C) with ATAC-seq data from TCGA. The peak annotation demonstrated that the peaks of these genes were more likely to be located in promoter regions (Figure S7D). We also performed motif enrichment and calculated potential regulatory TFs within the 200 bp range of gene loci based on genes from the PRGS model with HOMER (Figure S8A). A KEGG analysis of the peaks was also performed (Figure S8B–E).

The risk score for each patient was then calculated using the expression matrix of these four genes (APOB, VCAN, ABCA6, and CTSF) weighted by their regression coefficients in the Cox model (Table S7). All patients were divided into high- and low-PRGS groups by the survminer package [30]. As can be seen from Figure 3B, the overall survival (OS) was significantly lower in the high-PRGS group relative to the low-PRGS group in the TCGA target GTEx training dataset and the two validation datasets (all with $p < 0.05$) (Figure 3B and Table S8). In Figure S9A, we see the distribution of the PRGS scores for the patient group, as well as the relationship between the PRGS and survival time. The cut-off for the high- and low-PRGS groups was 50%.

We also evaluated the PRGS in Lauren and WHO histotypes. The results shown in Figure S9B–D also indicate that the overall survival (OS) of all classifications was significantly lower in the high-PRGS group relative to the low-PRGS group in the TCGA target GTEx training dataset and the two validation datasets (all with $p < 0.05$) (The GSE66229 and GSE15459 only provide the information of Lauren classification). Then, we assessed the OS status in different TNM stage groups. As shown in Figure S10A,B,D,E, the high-PRGS group was tightly correlated with a worse OS status in late stages in TCGA target GTEx (T3, T4, N2, and N3) and in GSE66229 (T2, T3, N1, N2, and N3). The high-PRGS group was vastly associated with a worse OS status in the M0 stage and most of the patients in these datasets were in the M0 stage (Figure S10C,F and Table S4).

We noticed that patients with a high PRGS were distributed in all different stages and OS statuses (Figure S11), indicating that our PRGS model is prognosis-specific and capable of assessing the risk of GC patients regardless of stage and status, though clinically, it would be potentially used at later stages for precision.

To investigate whether the prognostic value of the PRGS based on the four genes was an independent risk factor, a multivariate Cox regression analysis was performed; the results indicated that the PRGS was notably associated with OS compared with other clinical characteristics (age, sex, clinical stages, and T, N, and M stages), thus validating that the PRGS is robust in independently predicting the GC prognosis (Figure 3C). Similarly, the PRGS was still an independent risk factor for OS in the validation datasets (GSE66229 and GSE15459) (Figure 3D). We further explored the performance of the PRGS with other characters and found that the performance of the PRGS was better than that of other factors, including gender (whether male or female), age, pathological stages (T1~4, N0~3, M0~M1), and clinical stages (stage I~IV) in TCGA target GTEx, GSE66229, and GSE15459 (Figure S12A–C).

The discrimination of the PRGS was measured by an ROC analysis, with 1-, 3-, and 5-year AUCs of 0.601, 0.663, and 0.717 in TCGA target GTEx; 0.607, 0.709, and 0.707 in GSE66229; and 0.657, 0.674, and 0.711 in GSE15459 (Figure 4A and Table S9). The C-index (95% confidence interval) of the PRGS was the highest compared with other factors (gender; age; T, N, and M; clinical stage) in the TCGA target GTEx cohorts (Figure S13).

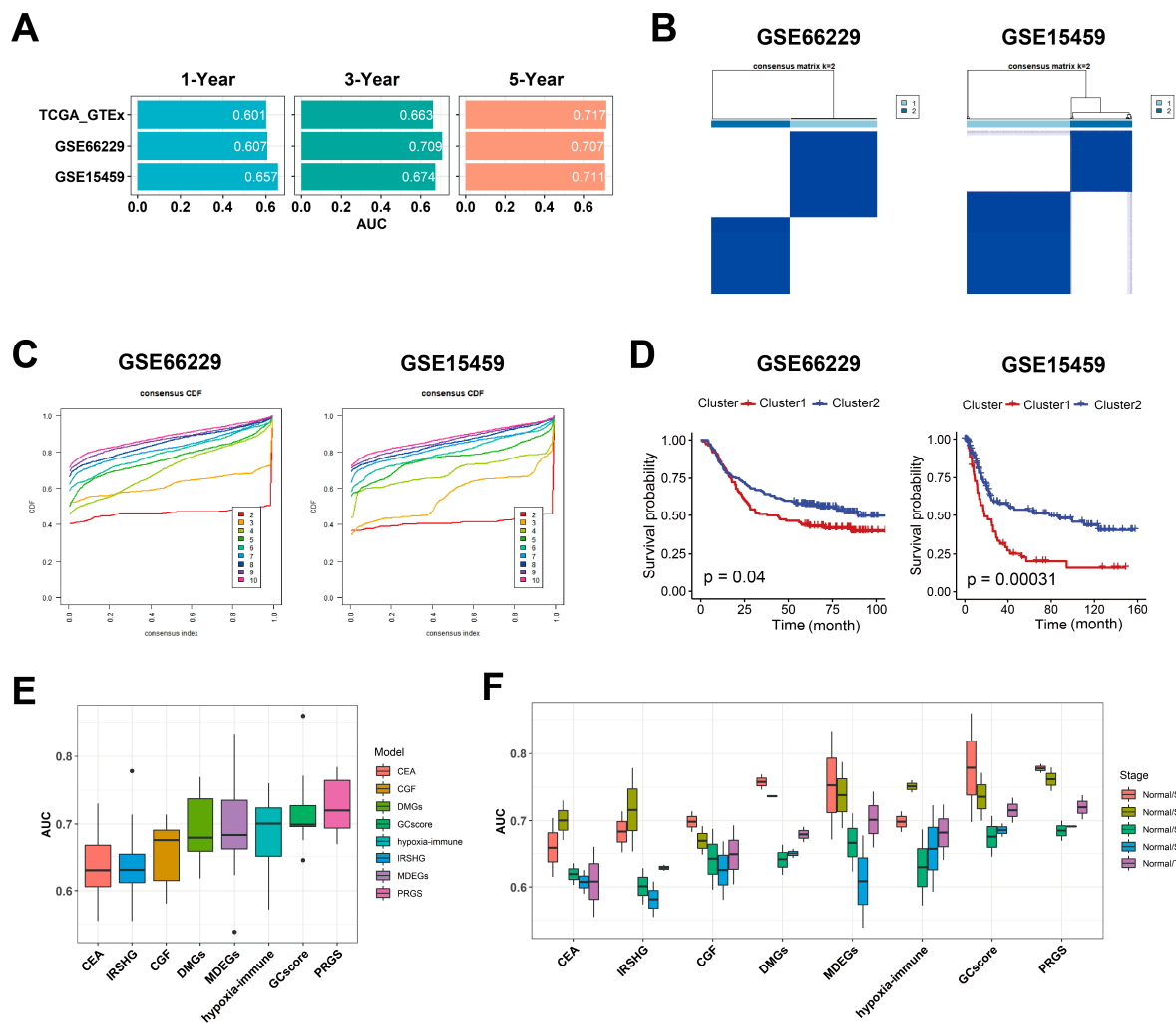


Figure 4. Evaluation of the PRGS signature. (A) A time-dependent ROC analysis for predicting OS at 1, 3, and 5 years in TCGA target GTE_x, GSE66229, and GSE15459. (B) The consensus score matrix of all samples in GSE66229 and GSE15459 when $k = 2$. A higher consensus score between two clusters indicates they are more likely to be grouped into the same cluster in different iterations. (C) The CDF curves of the consensus matrix for each k . (D) Kaplan–Meier analysis showed that patients in Cluster 1 exhibited a worse OS in both the GSE66229 ($p = 0.04$) and GSE15459 ($p = 0.00031$) cohorts. (E) Box plot of AUC of two classifiers (LR, logistic regression; RF, random forest) with eight gene signatures as features in GSE66229. (F) Box plot of AUC of CEA, GCscore, and PRGS in clinical stages I to IV. CEA [21] is three common clinical GC biomarkers, including CEACAM1, CEACAM5, and CEACAM6; GCscore [6] includes FANCA, DUSP3, HIST1H3B, CLNS1A, and FANCC; PRGS includes APOB, VCAN, ABCA6, and CTSF; IRSHG includes seven immune-related signatures (TGFB1, NOX4, F2R, TLR7, CIITA, RBP5, and KIR3DL3) [21]; CGF includes two cadherin gene signatures, CDH2 and CDH6 [31]; DMGs include six metastasis-related gene signatures (TMEM132, PCOLCE, UPK1B, PM20D1, FLJ35024, and SLITRK2) [32]; MDEGs include eight methylation-based gene signatures (TREM2, RAI14, NRP1, YAP1, MATN3, PCSK5, INHBA, and MICAL2) [33]; and hypoxia-immune-based gene signature includes CXCR6, PPP1R14A, and TAGLN [34].

3.3. PRGS Are Significantly Related to Clinical Outcomes

To further validate the performance of our PRGS model, we conducted multiple analyses to evaluate the robustness of these four prognostic genes in the aforementioned independent datasets (GSE66229 and GSE15459). First, through consensus clustering based on the four prognostic genes, we applied a consensus cluster analysis to all GC samples, resulting in the division of the samples into k clusters ($k = 2-9$). The cumulative distribution

function (CDF) curves of the consensus score matrix and proportion of ambiguous clustering (PAC) statistic indicated that the optimal number was obtained when $k = 2$ [35]. We classified GC patients in GSE66229 and GSE15459 into Cluster 1 and Cluster 2 (Figure 4B,C and Tables S10 and S11), respectively. The Kaplan–Meier curve indicated significant OS differences between the two clusters via the log-rank test, and the OS of the patients in Cluster 1 was significantly worse than that of Cluster 2 (Figure 4D). As shown in Figure S14A,E, the overall expression levels of the PRGS in Cluster 1 were higher than in Cluster 2 for both datasets. When $k = 3$ and 4, the Kaplan–Meier curve indicated that patients in Cluster 1 with the lower PRGS expression level had a vastly better OS than that of the other clusters (Figure S14B–D,F–H). Thus, for these four genes, the increased expression level was associated with a worse survival rate for gastric cancer patients. The observed consistency suggests that the expression levels of these four genes are vastly correlated with the OS of patients.

There are many prognostic gene expression signatures that have been developed based on bioinformatics methods. To compare the performance of the PRGS with other signatures, we selected five cancer stem cell-related feature genes in GCscore risk models (FANCA, DUSP3, HIST1H3B, CLNS1A, and FANCC) [6], three common clinical GC biomarkers (CEACAM1, CEACAM5, and CEACAM6) [21], seven immune-related signatures (TGFB1, NOX4, F2R, TLR7, CIITA, RBP5, and KIR3DL3) [31], two cadherin gene signatures (CDH2 and CDH6) [32], six metastasis-related gene signatures (TMEM132, PCOLCE, UPK1B, PM20D1, FLJ35024, and SLITRK2) [33], eight methylation-based gene signatures (TREM2, RAI14, NRP1, YAP1, MATN3, PCSK5, INHBA, and MICAL2) [34], and three hypoxia-immune-based gene signatures (CXCR6, PPP1R14A, and TAGLN) [36] as features to classify gastric cancer patients with different machine-learning (ML) classifiers. We utilized the logistic regression (LR) classifier and random forest (RF) classifier ML models to predict whether patients had GC in GSE66229 (GSE15459 did not have normal samples). We performed binary classification with them at the same time. The results indicated that the PRGS had a low sensitivity to different classifiers while it had the highest and most robust AUC (Figures 4E and S14A,B). Additionally, when performing classification tasks on the patients in different clinical stages (stages I, II, III, and IV), the PRGS had the highest and most robust accuracy (Figures 4F, S15A,B and S16A). Hence, the PRGS signature had optimized effects in classifying GC samples from control samples, which could serve as a potential feature for examining patients' prognoses. Together, we believe that the expression matrix based on these four genes (APOB, VCAN, ABCA6, and CTSF) as features for screening GC samples could properly assist classifiers in distinguishing cancerous samples from normal samples, acquiring satisfactory precision from test datasets and predicting whether patients have gastric cancer in each stage accurately.

3.4. Experimental Validation of the PRGS in the Clinical Samples and Cell Lines

To examine the protein expression levels of the four genes (APOB, VCAN, ABCA6, and CTSF) in gastric cancer, we performed immunohistochemistry (IHC) on the pathological section of the samples from patients of different stages, including normal, paracancerous tissue, early gastric cancer (EGC), and advanced gastric cancer (AGC) (Figure S16B–D). Compared with the samples from the normal tissues, the expression levels of the four proteins (APOB, VCAN, ABCA6, and CTSF) were much higher in EGC (Figure 5A–D). These four genes showed strong expression in almost all of the AGC samples and showed weak expressions in a portion of the paracancerous areas (Figure 5E–H). These results indicated a higher expression level for the four genes (APOB, VCAN, ABCA6, and CTSF) in GC compared to normal and paracancerous lesions. Furthermore, these proteins have already reached high expression levels in EGC samples compared to the matching normal samples.

To investigate how these four highly-expressed genes influence tumorigenesis, we further performed anti-phosphorylated histone 3 (pH3) immunostaining and flow cytometry-based cell cycle assays (Figure 6A–D). Consistent with the results in human cancer tissues, all four genes were highly expressed in GES-1 and BGC803 gastric cancer cells (Figure 6E,F).

Knocking down any of these four genes independently led to reduced phosphorylated histone 3, a well-established biomarker [37] of cell proliferation (Figure 6A,B). Consistently, the cell cycles of GC cells shifted from the S/G2/M to the G1 state upon the knockdown of APOB, VCAN, ABCA6, and CTSF, as demonstrated in flow cytometric analyses of the cell cycle with propidium iodide staining (Figure 6C,D). Note that the knockdown efficacy of the siRNAs used in our experiments were validated by quantitative real-time PCR (Figure 6E,F). These data unraveled that APOB, VCAN, ABCA6, and CTSF play important roles in the incidence of GC.

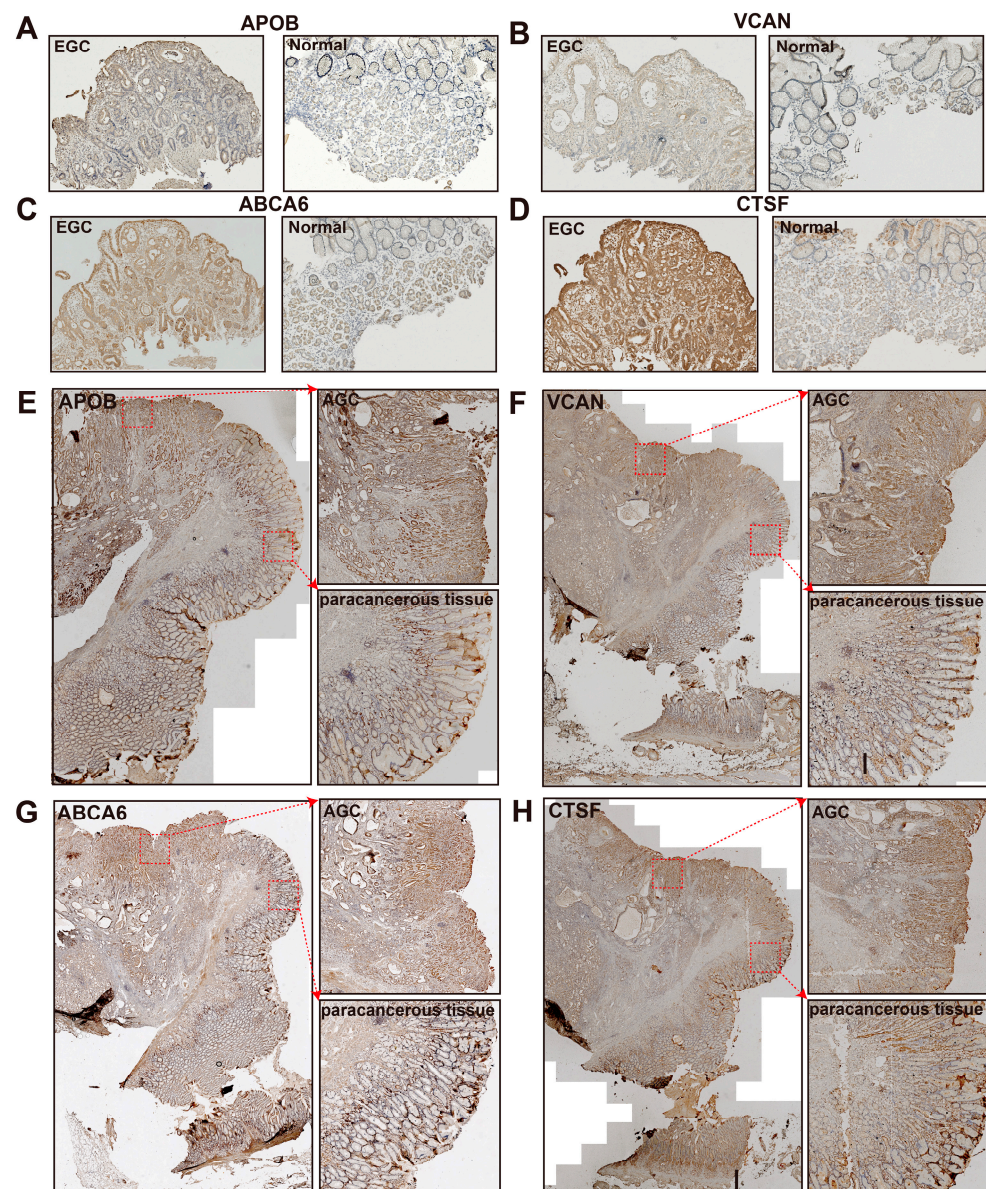


Figure 5. The results of immunohistochemistry (IHC). (A–D) IHC staining of APOB (A), VCAN (B), ABCA6 (C), and CTSF (D) in an early gastric cancer sample (EGC, left) and a normal sample (right). (E–H) IHC staining of APOB (E), VCAN (F), ABCA6 (G), and CTSF (H) in advanced gastric cancer (AGC). The pathological section of AGC shows the overall site (left), tumor region (AGC, upper right), and paracancerous tissue region (lower left).

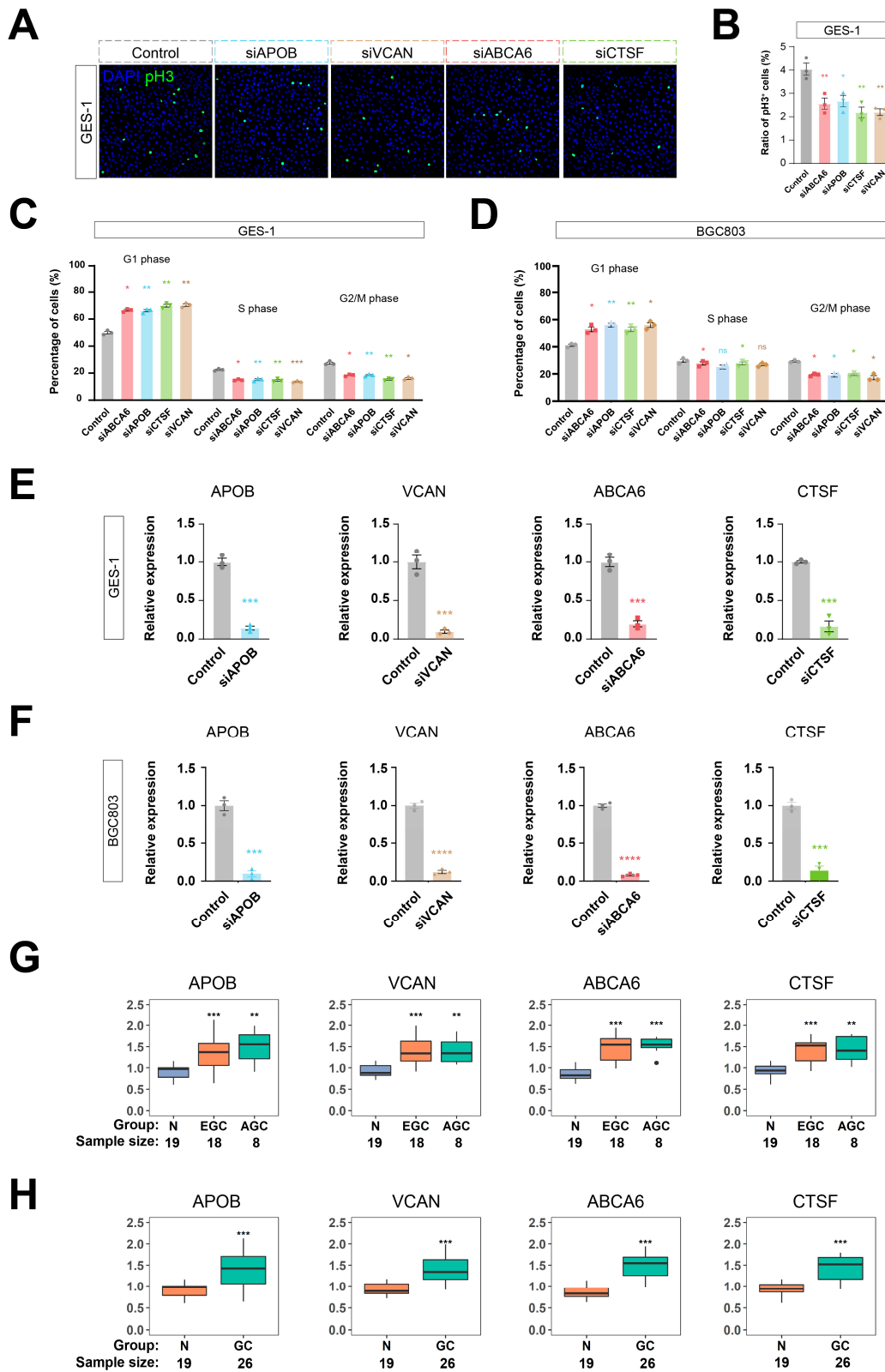


Figure 6. Knockdown of PRGS inhibited proliferation of gastric cancer cells. (A–D) Knockdown of any one of APOB, VCAN, ABCA6, or CTSF by siRNA inhibited cell proliferation of GES-1 and BGC803 gastric cancer cells. Shown are the anti-phosphorylated histone 3 (pH3) immunostaining (A,B) and flow cytometry-based cell cycle assays (C,D) of the APOB, VCAN, ABCA6, or CTSF siRNA

treatment. For (A), scale bar is 10 μm . For (B–F), $n = 3$. One-way ANOVA and Tukey's multiple comparison tests were used. In all figures, standard errors of the mean were represented. NS is not significant. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$, **** $p < 0.0001$. (G) Box plot of PRGS expression level in normal, early gastric cancer, and advanced gastric cancer patients. (H) Box plot of PRGS expression level in normal and gastric cancer patients. N means normal gastric tissues; EGC means early gastric cancer tissues; AGC means advanced gastric cancer tissues; and GC means gastric cancer tissues.

We detected the gene expression of the PRGS using an RT-PCR assay in a clinical cohort that included 19 normal gastric samples, 18 EGC patients, and 8 AGC patients by conducting qRT-PCR experiments (Figure 6G,H). The result indicated that the expression level in GC patients was significantly higher than in normal samples. These assays supported that our PRGS model was robust.

3.5. The Immune Cell Infiltration between the High- and Low-PRGS Patients

Immune-infiltrating cells in the tumor microenvironment (TME) can modulate the tumor phenotype. We assessed tumor purity as well as stromal and immune scores using the ESTIMATE algorithm in TCGA target GTEx samples [23]. ESTIMATE generates three types of scores, namely stromal scores indicating the presence of stroma, immune scores representing the infiltration of immune cells, and an ESTIMATE score, which infers tumor purity. Samples with a low tumor purity show high ESTIMATE scores [23]. The results demonstrated that stromal and immune cells significantly increased along with malignancy progression (from stages I to IV). The ESTIMATE score also increased from stage I to stage IV, whereas tumor purity decreased in higher grades (Figure 7A and Table S12) in accordance with previous studies [38], which illustrated that lower tumor purity correlates with severer malignancy.

As expected, a high abundance of stromal cells and immune cells and a low tumor purity were shown in the high-PRGS group (Figure 7A). Hence, the high-PRGS group was positively correlated with malignancy. To further chart the underlying immune cells, we implemented the CIBERSORT algorithm to infer the differential abundance between the high- and low-PRGS patients [24]. The high-PRGS group had a very high level of naïve B cells ($p = 0.00014$), monocytes ($p = 0.0312$), M2 macrophages ($p = 1.6 \times 10^{-5}$), and resting mast cells ($p = 5.6 \times 10^{-5}$) (Figure 7B and Table S13), whereas T follicular helper cells ($p = 3.9 \times 10^{-7}$), Treg ($p = 2.1 \times 10^{-6}$), resting NK cells ($p = 0.00716$), and activated mast cells ($p = 0.03602$) exhibited a consistent negative correlation with the low-PRGS group. Then, we determined whether the infiltrating immune cells mentioned above could be associated with patient survival. In agreement with previous studies indicating the promoting roles of M2 macrophage cells in tumor progression [39] and naïve B cells when they differentiate into Breg cells in the tumor microenvironment, participating in tumor metastasis [40], we discovered that the naïve B cells and M2 macrophage cells appeared to be associated with poor survival (Figure 7C,D), consistent with the PRGS groups. In addition, the regulatory T cells (Treg), T follicular helper (Tfh) cells, and M0 macrophage cells showed a negative correlation with patient survival (Figure 7E,F), which accords with the low-PRGS group, where there were more Treg and Tfh cells than in the high-PRGS group. Together, our data indicate that the high-PRGS group was characterized by a TME with a high immune cell infiltration and a low tumor purity.

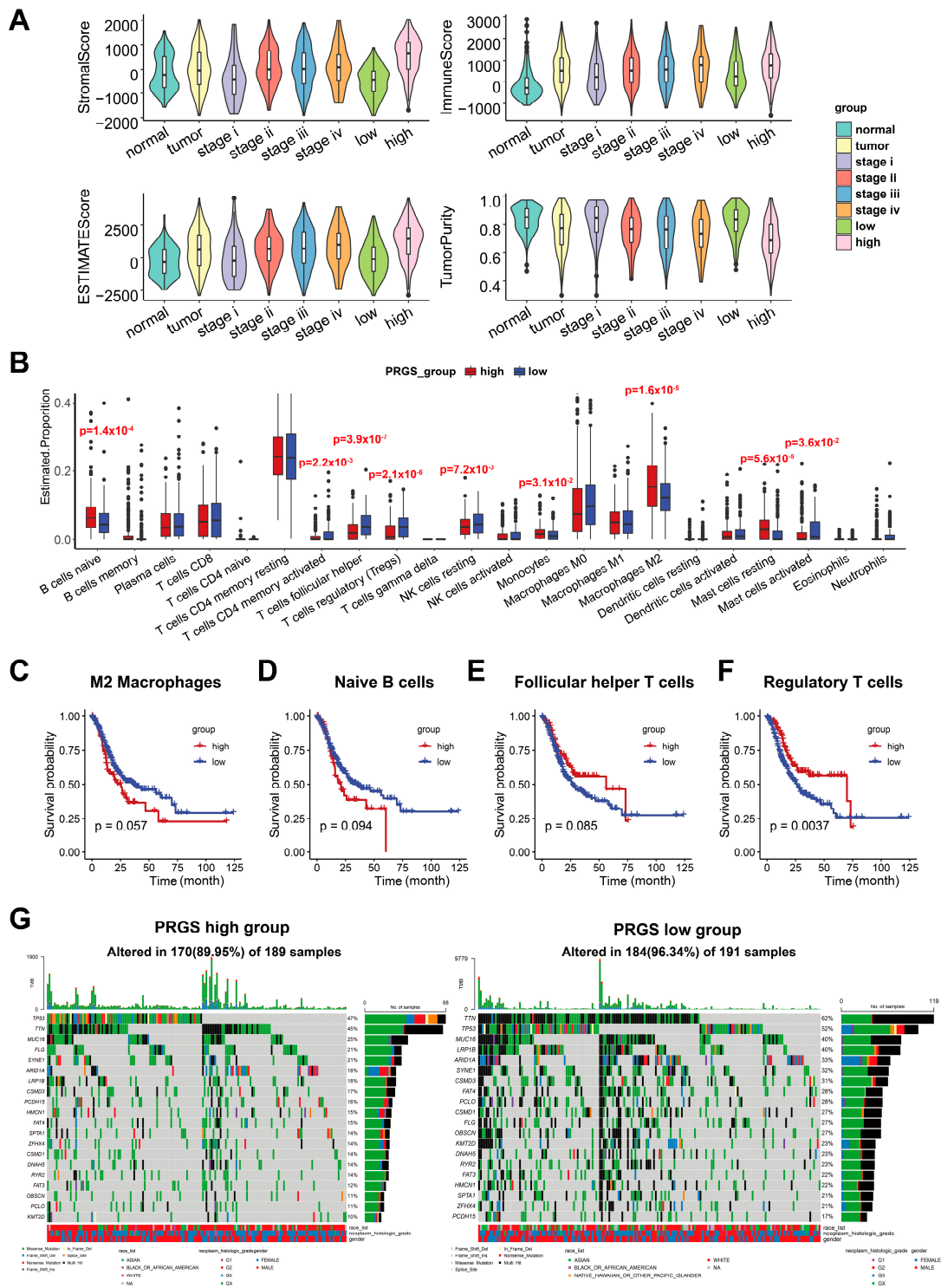


Figure 7. The landscape of immune infiltrations and tumor mutation status in the low- and high-PRGS groups. **(A)** Distinct distribution of stromal score (upper left), immune score (upper right), ESTIMATE score (lower left), and tumor purity (lower right) among GC patients in TCGA target GTEx. **(B)** The immune-infiltrating cells in the TME were determined based on the CIBERSORT algorithm in the TCGA target GTEx. **(C–F)** Kaplan–Meier curves of OS according to the M2 macrophages **(C)**, naïve B cells **(D)**, follicular helper T cells **(E)**, and regulatory T cells **(F)** in TCGA target GTEx. **(G)** OncoPlot of significantly mutated genes in high- (left) and low- (right) PRGS groups. The mutation types with their frequencies are presented.

3.6. Mutation Status in GC Patients in the High- and Low-PRGS Groups

To investigate PRGS-related mechanisms in GC, we also analyzed somatic mutations. When comparing the mutant frequency between samples of the low- and high-PRGS groups, we detected more mutations in the low-PRGS group than the high-PRGS group, indicating that more mutations led to a lower PRGS and GC risk (Figure 7G). Past studies have revealed sophisticated correlations between mutations and tumor prognoses, e.g., TP53 mutations have significant associations with poor outcomes in kidney renal clear cell carcinoma, head and neck squamous cell carcinoma, and acute myeloid leukemia, as well as improved outcomes in ovarian serous cystadenocarcinoma [41]. Besides, IDH1 and MUC16 mutations are associated with an improved prognosis in gastric cancer [41–45]. All of the top 20 frequent mutations were significantly enriched in cases with patients in both groups (Figure 7G). As a reference, we utilized the GCscore [6] as feature genes (FANCA, DUSP3, HIST1H3B, CLNS1A, and FANCC) to classify GC patients from the TCGA dataset into high- and low-risk groups (Figure S17A). The top 20 genes with the highest mutation frequencies also had relatively higher mutation frequencies in the low-risk group than in the high-risk group (Figure S17B). This result exhibited a similar tendency as our PRGS. We also found that MUC16 ($p = 0.051$), CSMD1 ($p = 0.054$), and FAT4 ($p = 0.033$) mutants had better outcomes than wild type (WT), while TP53 and TTN, genes with the highest mutation frequencies, did not show differences in the outcomes between their mutants and the WT (Figure S17C–G). Frequent mutations in TNN, TP53, and MUC16 were significantly enriched in the high- and low-PRGS groups, which were within expectations according to previous reports [46].

Additionally, APOB portrayed a mutation percentage of 8% and 18% in patients of the high- and low-PRGS groups, respectively. The mutation percentage for VCAN was 6% and 10% in the high- and low-PRGS groups; for ABCA6, it was 2% and 3%; and for CTSF, it was 1% and 3% (Figure S18A). Previous work [47] has reported APOB and VCAN as mutation driver genes.

Moreover, we observed significant co-occurrences among mutations of these genes. Among the top 20 genes with the highest mutation probabilities in gastric cancer, the majority of these gene mutations were co-occurring. The gene mutations in the low-PRGS group displayed more significant co-occurrences than those in the high-PRGS group (Figure S18B). There were also gene pairs that were mutually exclusive. For example, previous studies have reported that ARID1A mutant tumors display p53 pathway activation, and that ARID1A directly regulates TP53 target genes [48,49]; our analysis also displayed that in the high-PRGS group, TP53 and ARID1A mutations were significantly mutually exclusive. On the other hand, within the low-PRGS group, KMT2D and FAT4 were also significantly mutually exclusive along with the previous two. These results also indicate a possible pair of mutually exclusive mutations in patients within the low-PRGS group.

Somatic mutations were further explored on the basis of oncogenic signaling pathways. We summarized the gene expression levels and mutations of common oncogenic signaling pathways from patients of the high-PRGS group and low-PRGS group. For instance, disturbing the hippo pathway promoted GC proliferation and metastasis [50], and dysregulation of the MAPK pathway promoted cell metabolism, proliferation, apoptosis, and migration [51]. Our results also indicated that both the hippo and MAPK pathways had high mutation percentages for the samples from the high-PRGS and low-PRGS groups (Figure S18C). We discovered that for the most investigated cancer pathways, the patients from the high-PRGS group had higher pathway enrichment scores than the patients in the low-PRGS group (Figure S18D). Yet, the percentage of patients with mutations was lower in the high-PRGS group than in the low-PRGS group, which might have arisen from other molecular mechanisms.

4. Discussion

To date, there have rarely been effective prognostic genes identified for detecting and predicting gastric cancer prognoses. There is a noticeable difference among GC in

the cardiac zone, the fundus/corpus zone, and the pyloric zone. However, researchers might neglect some significant genes when considering mixed data. Hence, we constructed co-expression networks of GC in these three zones separately to select prognostic genes and build the PRGS. Different treatment options mean that patients need better-individualized evaluations when implementing clinical decisions. These can be used as reliable biomarkers for the diagnosis of “high-risk” GC patients.

In this study, WGCNA and Cytoscape were applied to identify the gene co-expression networks of GC in the cardiac zone, the fundus/corpus zone, and the pyloric zone. With the expression profiles of these genes in TCGA target GTEx and two independent datasets, the LASSO-Cox regression model was applied to develop a prognostic gene signature. The prognostic analysis demonstrated that this PRGS was a deleterious indicator of the OS. Besides, the PRGS demonstrated a high accuracy and consistent performance for TCGA target GTEx. We also performed cross-validation using two independent GEO datasets (GSE66229 and GSE15459), which indicated great potential for the clinical application of the PRGS.

The common tools for evaluating clinical outcomes and making therapeutic schedules include T, N, M, and the clinical stage [4]. Remarkably, our signature worked independently of these factors and had a vastly superior performance in predicting prognoses. Besides, we took the PRGS as features to classify GC and normal samples with machine-learning classifiers. All of the classification accuracy based on the PRGS signature as features were highest with independent datasets (GSE66229) and for several clinical stages (stages I, II, III, and IV).

There are a number of prognostic gene signatures based on characteristic genes, such as stem cell-related gene signatures [6] and DNA methylation-related gene signatures [34]. These gene signatures overlook the differences among tumor locations. Although several of these signatures have been developed, few have been implemented in clinical experiments and even fewer have undergone rigorous validation. We compared the PRGS with these signatures to classify GC patients, and the PRGS had a better performance than the other signatures. To further confirm the clinical significance of the PRGS, we conducted a validation assay using qRT-PCR on 45 frozen gastric cancer tissues. The results further supported the validity of the PRGS as a clinical marker. Consequently, our signature has the potential to be a useful clinical tool for the prognosis determination of GC.

To further test the clinical explanation of the PRGS, the experimental validation was based on the IHC results from different clinical stages of independent GC patients, validating our prior findings and assessing their feasibility. This indicated that the protein expression level of the PRGS signature was significantly higher in GC patients. This conclusion was further experimentally validated in GES-1 and BGC803 gastric cancer cells, as the knockdown of these four genes led to cell growth inhibition by regulating the cell cycle. Hence, the PRGS signature could serve as a promising surrogate for assessing the prognosis of GC in clinical settings.

Pathologists generally determine tumor purity by visual evaluation, which is affected by the sensitivity of histopathology, interobserver bias, and variability in accuracy [52]. Our results show that the high-PRGS-group patients had a lower tumor purity and higher levels of immune and stromal cell infiltration compared with patients in the low-PRGS group. Previous studies have reported that tumor cells can dominate the microenvironment [52], which has given rise to the hypothesis that malignant GC recruits abundant surrounding cells and subjugates them to compose a protective shield. Therefore, a lower tumor purity and correlated cellular heterogeneity may contribute to a worse prognosis for GC. In our study, we found that the PRGS was positively correlated with several infiltration cells, such as M2 macrophages and naïve B cells, which are positively correlated with the prognosis of GC patients; furthermore, the PRGS was negatively correlated with follicular helper T cells, M0 macrophages, and regulatory T cells, which are negatively correlated with the prognosis of GC patients.

We found that GC patients in the low-PRGS group had a higher rate of oncogenic mutations. It should be noted that the association of gene mutations with cancer outcomes is sophisticated. Take MUC16 and IDH1 as examples: past GC studies have illustrated that MUC16 mutations could activate the p53 pathway and DNA repair pathway, which are all tumor suppressor pathways [42,43]; thus, improved outcomes of GC might be expected from mutated MUC16 [34,36]. Mutations in IDH1, as a tumor suppressor in human glioma cells through the negative regulation of Wnt/ β -catenin signaling, improves survival conditions [44,45]. In addition, mutations in the top 20 genes had a high frequency of co-mutations. However, further investigations are needed for a deep understanding of the mechanisms of mutations in GC.

5. Conclusions

In a word, based on a series of bioinformatics, machine-learning-based algorithms, and experimental validation, we developed a powerful and robust signature for assessing the prognosis of GC patients. This PRGS model may be a promising tool for screening and monitoring individual GC patients.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers15051610/s1>, Figure S1. (A) Bubble plot of GO enrichment for GC in the cardiac zone, the fundus/corpus zone, and the pyloric zone. Green represents biological processes; red represents cellular components; blue represents molecular functions. (B) Identify differentially expressed genes (DEGs) and KEGG enrichment of DEGs for GC in the cardiac zone, the fundus/corpus zone, and the pyloric zone. The color of the bubble represents the significance of p_{adjust} . (C) Analysis of network topology for different soft-threshold power. The left panel shows the impact of soft threshold power on the scale-free topology fit index; the right panel displays the impact of soft-threshold power on the mean connectivity. Figure S2. GO and KEGG enrichment of genes in the purple, brown and salmon modules. (A–C) GO enrichment of the purple module (correlated with the cardiac zone) (A), salmon module (correlated with fundus/corpus zone) (B) and brown module (correlated with the pyloric zone) (C). (D–F) KEGG enrichment of the purple module (correlated with the cardiac zone) (D), salmon module (correlated with the cardiac zone) (E) and brown module (correlated with the pyloric zone) (F). Figure S3. Survival analysis and binary classification of the genes from the co-expression networks (A) Kaplan-Meier curves of OS according to the top5 genes (APOA4, MS4A10, SLC28A1, AQP10, APOB) from the cardiac zone co-expression network separate of GC in the cardiac zone (left, $p = 0.074$), the fundus/corpus zone (middle, $p = 0.22$), and the pyloric zone (right, $p = 0.73$) in TCGA target GTEEx. (B) Kaplan-Meier curves of OS according to the top5 genes (APOA4, MS4A10, SLC28A1, AQP10, APOB) from the cardiac zone co-expression network separate of GC in the cardiac zone (left, $p = 0.059$), the fundus/corpus zone (middle, $p = 0.15$), and the pyloric zone (right, $p = 0.33$) in GSE66229. (C) Kaplan-Meier curves of OS according to the top5 genes (VCAN, COL1A2, FAP, PODNL1, SULF1) from the fundus/corpus co-expression network separate of GC in the cardiac zone (left, $p = 0.16$), the fundus/corpus zone (middle, $p = 5.32 \times 10^{-6}$), and the pyloric zone (right, $p = 0.73$) in TCGA target GTEEx. (D) Kaplan-Meier curves of OS according to the top5 genes (VCAN, COL1A2, FAP, PODNL1, SULF1) from the fundus/corpus co-expression network separate of GC in the cardiac zone (left, $p = 0.83$), the fundus/corpus zone (middle, $p = 0.0047$), and the pyloric zone (right, $p = 0.045$) in GSE66229. (E,F) BoxPlot of AUC of two classifiers (LR, Logistic Regression; RF, Random Forest) with the top5 genes (APOA4, MS4A10, SLC28A1, AQP10, APOB) and the top5 genes (VCAN, COL1A2, FAP, PODNL1, SULF1) from the cardiac zone co-expression network (E) and the fundus/corpus co-expression network (F) separately as features in TCGA target GTEEx and GSE66229. The cardiac zone means GC in the cardiac zone; the Fundus/Corpus zone means GC in the fundus/corpus zone; the Pyloric zone means GC in the pyloric zone. Figure S4. (A) In the TCGA target GTEEx cohort (383 GC samples with clinical information), the optimal λ was determined when the partial likelihood deviance reached the minimum value and further generated Lasso coefficients of the most useful prognostic genes. Data are presented as mean \pm 95% confidence interval [CI]. (B) Cell-type markers. The expression levels of cell-type markers across cell types are shown. Cell-type marker genes were identified in Wilcoxon rank-sum test (FDR < 0.01, and fold change > 1.5) and only the top genes are shown in the figure. (C) The UMAP plot of 200,954 cells to visualize cell-type clusters. (D) The UMAP plot of both samples. (E) The UMAP plot of both cells from normal and

tumor tissues (blue means tumor tissues; red means normal tissues). Figure S5. (A) The UMAP plot of cell types marked by specific marker genes. The epithelial cell is marked by CDH1, and including pit mucous cell (Pit, MUC5AC and TFF1 positive), chief cell (Chief, LIPF and PGA3 positive), and intestinal cell (Intestinal, REG4 and TFF3 positive). Fibroblast is FN1, LUM, DCN positive; Pericytes is RGS5 and NOTCH3 positive; Endothelial cell is PLVAP and ACKR positive; T cell is CD8A positive; T regulatory cell (Treg) is IL2RA positive; NK cell is KLRD1 positive; B cell is MS4A1 positive; Plasma is TNFRSF 17 positive; mast cell is KIT positive; Macrophage is CD163 positive; Dendritic cell is PLD4 positive. (B) The UMAP plot of PRGS expression level. Figure S6. (A–C) The UMAP plot of all samples (A), normal samples (B), tumor samples (C) according to PRGS score. (D) BoxPlot of PRGS in different cell types. Normal, normal samples; tumor, tumor samples; Diffuse, diffuse-type gastric cancer; Intestinal, intestinal-type gastric cancer; Mixed, mixed-type gastric cancer. Figure S7. (A) Genome characterization and enrichment analysis. (A) Distribution map of transcription factors and TSS. (B) location distribution of Peaks on the genome. (C) relative proportions of gene coding regions, intergenic regions, introns, exons, and upstream and downstream regions. (D) Browser showing chromatin accessibility status of PRGS signature (APOB, VCAN, ABCA6, CTSF). The grey bars highlighted peaks of PRGS signature in promoter regions. All of the genes showed duplication of two samples. Figure S8. (A) Motif enrichment analysis of common genes around open chromatin regions. Fra1 and Foxa2 were found to be significantly enriched with APOB; KLF6 and Erra were found to be significantly enriched with VCAN; Fos and GABPA were found to be significantly enriched with ABCA6; BATF and Sp5 were found to be significantly enriched with CTSF (B–D) GO enrichment of the GC with ATAC-seq data. (E) KEGG enrichment of the GC with ATAC-seq data. Figure S9. (A) Establishment of PRGS model dividing patients into high- and low- groups in the TCGA target GTEX (left), GSE66229 (middle) and GSE15459 (right). The cutoff value on the top (red: high expression; blue: low expression); survival statues on the bottom. (B–D) Kaplan–Meier curves of overall survival (OS) according to the PRGS of the different Lauren/WHO histotypes patients in TCGA target GTEX (B), GSE15459 (C), and GSE66229 (D). Intestinal means Intestinal-type gastric cancer patients; Diffuse means diffuse-type gastric cancer patients; Adenocarcinoma means gastric adenocarcinoma patients, and accurate histotypes is unknown; Tubular adenocarcinoma means tubular adenocarcinoma patients. Figure S10. (A) Survival analysis of the different TNM stages patients. (A–F) Kaplan–Meier curves of overall survival (OS) according to the PRGS of the different TNM stages patients in TCGA target GTEX (A–C), and GSE66229 (D–F). Figure S11. The integrated sankey diagram portrays the underlying correlations across the PRGS group, stage and OS status in the TCGA target GTEX cohort (345 samples with information on the clinical stage). Figure S12. Multivariate Cox analysis evaluating independently predictive ability of PRGS model and other clinical characteristics for OS. (A) Multivariate Cox regression of PRGS regarding OS in TCGA target GTEX (345 samples with information on the clinical stage). (B) Multivariate Cox regression of PRGS regarding OS in GSE66229 (n = 298). (C) Multivariate Cox regression of PRGS regarding OS in GSE15459 (n = 191). Statistic tests: two-sided Wald test. Data are presented as hazard ratio (HR) ± 95% confidence interval [CI]. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. Figure S13. The performance of PRGS was compared with other clinical and molecular variables in predicting prognosis. Statistic tests: two-sided z-score test. Figure S14. (A) Heatmap showed that the expression level of cluster1 is higher than cluster2 in GSE66229. (B) The consensus score matrix of all samples in GSE66229 when $k = 3$ or 4. (C) Kaplan–Meier analysis showed that patients in Cluster 1 exhibited better OS in the GSE66229 cohort when $k = 3$ (left, $p = 0.048$) or $k = 4$ (right, $p = 0.0071$). (D) Heatmap showed that the expression level of cluster1 is lower than others in GSE66229 when $k = 3$ (left) or $k = 4$ (right). (E) Heatmap showed that the expression level of cluster1 is higher than cluster2 in GSE15459. (F) The consensus score matrix of all samples in GSE15459 when $k = 3$ or 4. (G) Kaplan–Meier analysis showed that patients in Cluster 1 exhibited better OS in the GSE15459 cohort when $k = 3$ (left, $p = 0.00064$) or $k = 4$ (right, $p = 0.00064$). (H) Heatmap showed that the expression level of cluster1 is lower than others in GSE15459 when $k = 3$ (left) or $k = 4$ (right). C1 means cluster1, C2 means cluster2, C3 means cluster3, C4 means cluster4. Figure S15. Evaluation of the PRGS signature with machine learning classifiers in GSE66229. (A,B) ROC curve of binary classification with logistic regression classifier (A) and random forest classifier (B) for all and different stage(stage I–IV) GC patients with PRGS signature as features. “Tumor” means GC patients; “Normal” means gastric tissues with a non-tumor state. Figure S16. (A) BoxPlot of AUC in two classifiers (LR, Logistic Regression; RF, Random Forest), respectively. (B–D) HE staining of advanced gastric cancer (AGC) (B), normal tissue (C) and early gastric cancer (D). Figure S17. Tumor mutation status of GCscore

(FANCA, DUSP3, HIST1H3B, CLNS1A, FANCC) [6] and Kaplan–Meier curves of high frequency mutant genes. (A) Kaplan–Meier curves of OS according to the GCscore [6] in TCGA. (B) OncoPlot of significantly mutated genes in high- (left) and low- (right) GCscore groups. The mutation types with their frequencies were presented. (C–G) Kaplan–Meier curves of OS between the oncogenic mutation and wild type (WT) of MUC16 (C), CSMD1 (D), FAT4 (E), TP53 (F) and TTN (G). Figure S18. Tumor mutation status of high- and low- PRGS groups. (A) The mutation percentages of PRGS signature (APOB, VCAN, ABCA6, CTSF) in high-(left) and low-(right) PRGS groups. (B) Interaction effect of genes mutating differentially in patients in the high-(left) and the low-(right) PRGS groups. (C) The mutation percentages of nine common oncogenes signaling pathways in high-(left) and low-(right) PRGS groups. (D) BoxPlot of nine common oncogenes signaling pathways expression profile in the high-(red) and low-(blue) PRGS groups. Table S1. Details of baseline information in TCGA target GTEX. Table S2. Details of baseline information in GSE15459. Table S3. Details of baseline information in GSE66229. Table S4. Datasets in use. Table S5. The remaining genes of each step. Table S6. Top25 genes selected by Cytoscape. Table S7. LASSO coefficients. Table S8. TCGA-STAD database with PRGS groups. Table S9. The evaluation metrics of PRGS in training and testing cohorts. Table S10. GSE66229 database with consensus clustering. Table S11. GSE15459 database with consensus clustering. Table S12. The stromal, immune cell infiltration and purity in TCGA target GTEX samples by ESTIMATE. Table S13. The immune cell infiltration in all TCGA target GTEX samples by CIBERSORT.

Author Contributions: Conception and design: M.Q.Z., C.L. and X.L.; experimental operation: Y.Z.; collection and assembly of data: C.L., X.L. and Y.H.; provision of materials and patient information: P.J. and F.Y.; manuscript writing and revision: M.Q.Z., C.L., X.L., M.S., Y.H., Y.Z., P.J., F.Y., T.C., Z.W. and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Key Research and Development Program of China (2017YFA0505503), Fundamental Research Funds for the Central Universities (2019RC045), Beijing Municipal Natural Science Foundation (Z200021), National Key Research and Development Program of China (2021YFE0201100, 2022YFA1103401), the National Natural Science Foundation of China (81890991), CAS Interdisciplinary Innovation Team (JCTD-2020-04).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: In this study, the gastric tissues used were approved by the Department of Gastroenterology, Seventh Medical Center of Chinese PLA General Hospital, Beijing, China, on 1 June 2022. Overall, one advanced GC sample, an early GC sample, and a normal gastric sample were collected. All patients provided written informed consent, and the ethics committee of PLA General Hospital also approved our research.

Data Availability Statement: The datasets that support the findings of this study are available from the Cancer Genome Atlas TARGET and Genotype Tissue Expression project datasets (TCGA target GTEX, primary_site = stomach) and the NCBI Gene Expression Omnibus (GEO) under accession numbers GSE66229 and GSE15459.

Acknowledgments: We thank Jianming Zeng (University of Macau), all the members of his bioinformatics team, and Biotrainee for generously sharing their experience and codes, as well as the use of the biostudio's high-performance computing cluster (<https://biostudio.cloud>, accessed on 1 January 2022) at Biotrainee and The Shanghai HS Biotech Co., Ltd. for conducting the research reported in this paper. We thank Jun Qin and Xianju Li, Beijing Proteome Research Center, for giving us the gastric cancer cell line GSE-1. We thank Zijie Shen from the College of Agriculture and Biotechnology, Zhejiang University; Haotian Wang from Kunming Institute of Zoology, CAS; and Wenbo Guo, Department of Automation, Tsinghua University for data processing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)]
2. Fong, C.; Johnston, E.; Starling, N. Neoadjuvant and Adjuvant Therapy Approaches to Gastric Cancer. *Curr. Treat. Options Oncol.* **2022**, *23*, 1247–1268. [[CrossRef](#)] [[PubMed](#)]
3. Wittekind, C. The development of the TNM classification of gastric cancer. *Pathol. Int.* **2015**, *65*, 399–403. [[CrossRef](#)] [[PubMed](#)]

4. Egner, J.R. AJCC cancer staging manual. *JAMA* **2010**, *304*, 1726–1727. [[CrossRef](#)]
5. Shimada, H.; Noie, T.; Ohashi, M.; Oba, K.; Takahashi, Y. Clinical significance of serum tumor markers for gastric cancer: A systematic review of literature by the Task Force of the Japanese Gastric Cancer Association. *Gastric Cancer* **2014**, *17*, 26–33. [[CrossRef](#)]
6. Wei, C.; Chen, M.; Deng, W.; Bie, L.; Ma, Y.; Zhang, C.; Liu, K.; Shen, W.; Wang, S.; Yang, C.; et al. Characterization of gastric cancer stem-like molecular features, immune and pharmacogenomic landscapes. *Briefings Bioinform.* **2022**, *23*, bbab386. [[CrossRef](#)]
7. Cui, K.; Yao, S.; Liu, B.; Sun, S.; Gong, L.; Li, Q.; Fei, B.; Huang, Z. A novel high-risk subpopulation identified by CTSL and ZBTB7B in gastric cancer. *Br. J. Cancer* **2022**, *127*, 1450–1460. [[CrossRef](#)] [[PubMed](#)]
8. Cheong, J.H.; Wang, S.C.; Park, S.; Porembka, M.R.; Christie, A.L.; Kim, H.; Kim, H.S.; Zhu, H.; Hyung, W.J.; Noh, S.H.; et al. Development and validation of a prognostic and predictive 32-gene signature for gastric cancer. *Nat. Commun.* **2022**, *13*, 774. [[CrossRef](#)]
9. Helander, H.F. The cells of the gastric mucosa. *Int. Rev. Cytol.* **1981**, *70*, 217–289. [[PubMed](#)]
10. Hoffmann, W. Regeneration of the gastric mucosa and its glands from stem cells. *Curr. Med. Chem.* **2008**, *15*, 3133–3144. [[CrossRef](#)] [[PubMed](#)]
11. Hoffmann, W. Current Status on Stem Cells and Cancers of the Gastric Epithelium. *Int. J. Mol. Sci.* **2015**, *16*, 19153–19169. [[CrossRef](#)] [[PubMed](#)]
12. Lai, S. A human mode of intestinal type gastric carcinoma. *Med. Hypotheses* **2019**, *123*, 27–29. [[CrossRef](#)] [[PubMed](#)]
13. Vivian, J.; Rao, A.A.; Nothhaft, F.A.; Ketchum, C.; Armstrong, J.; Novak, A.; Pfeil, J.; Narkizian, J.; Deran, A.D.; Musselman-Brown, A.; et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat. Biotechnol.* **2017**, *35*, 314–316. [[CrossRef](#)] [[PubMed](#)]
14. Chen, H.M.; MacDonald, J.A. Network analysis of TCGA and GTEx gene expression datasets for identification of trait-associated biomarkers in human cancer. *STAR Protoc.* **2022**, *3*, 101168. [[CrossRef](#)] [[PubMed](#)]
15. Law, C.W.; Alhamdoosh, M.; Su, S.; Dong, X.; Tian, L.; Smyth, G.K.; Ritchie, M.E. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* **2016**, *5*. [[CrossRef](#)]
16. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)]
17. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
18. Chin, C.H.; Chen, S.H.; Wu, H.H.; Ho, C.-W.; Ko, M.-T.; Lin, C.-Y. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **2014**, *8*, S11. [[CrossRef](#)] [[PubMed](#)]
19. Tibshirani, R. The lasso method for variable selection in the Cox model. *Stat. Med.* **1997**, *16*, 385–395. [[CrossRef](#)]
20. Wilkerson, M.D.; Hayes, D.N. ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **2010**, *26*, 1572–1573. [[CrossRef](#)]
21. Beauchemin, N.; Arabzadeh, A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev.* **2013**, *32*, 643–671. [[CrossRef](#)]
22. Shen, Y.; Vignali, P.; Wang, R. Rapid profiling cell cycle by flow cytometry using concurrent staining of DNA and mitotic markers. *Bio-Protocol* **2017**, *7*, e2517. [[CrossRef](#)]
23. Yoshihara, K.; Shahmoradgoli, M.; Martínez, E.; Vegesna, R.; Kim, H.; Torres-Garcia, W.; Trevino, V.; Shen, H.; Laird, P.W.; Levine, D.A.; et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **2013**, *4*, 1–11. [[CrossRef](#)]
24. Newman, A.M.; Steen, C.B.; Liu, C.L.; Gentles, A.J.; Chaudhuri, A.A.; Scherer, F.; Khodadoust, M.S.; Esfahani, M.S.; Luca, B.A.; Steiner, D.; et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **2019**, *37*, 773–782. [[CrossRef](#)]
25. Mayakonda, A.; Lin, D.-C.; Assenov, Y.; Plass, C.; Koeffler, H.P. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **2018**, *28*, 1747–1756. [[CrossRef](#)] [[PubMed](#)]
26. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **2012**, *16*, 284–287. [[CrossRef](#)] [[PubMed](#)]
27. Hänzelmann, S.; Castelo, R.; Guinney, J. GSEA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinform.* **2013**, *14*, 7. [[CrossRef](#)]
28. Friedman, J.; Hastie, T.; Tibshirani, R. glmnet: Lasso and elastic-net regularized generalized linear models. *R Package Version* **2009**, *1*, 1–24.
29. Kumar, V.; Ramnarayanan, K.; Sundar, R.; Padmanabhan, N.; Srivastava, S.; Koiwa, M.; Yasuda, T.; Koh, V.; Huang, K.K.; Tay, S.T.; et al. Single-Cell Atlas of Lineage States, Tumor Microenvironment, and Subtype-Specific Expression Programs in Gastric Cancer. *Cancer Discov.* **2022**, *12*, 670–691. [[CrossRef](#)]
30. Kassambara, A.; Kosinski, M.; Biecek, P.; Fabian, S. Package ‘Survminer’. Drawing Survival Curves Using ‘ggplot2’ (R Package Version 03 1). 2017. Available online: <https://cran.r-project.org/web/packages/survminer/survminer.pdf> (accessed on 11 November 2022).
31. Geng, H.; Dong, Z.; Zhang, L.; Yang, C.; Li, T.; Lin, Y.; Ke, S.; Xia, X.; Zhang, Z.; Zhao, G.; et al. An immune signature for risk stratification and therapeutic prediction in helicobacter pylori-infected gastric cancer. *Cancers* **2022**, *14*, 3276. [[CrossRef](#)] [[PubMed](#)]

32. Luo, S.; Lin, R.; Liao, X.; Li, D.; Qin, Y. Identification and verification of the molecular mechanisms and prognostic values of the cadherin gene family in gastric cancer. *Sci. Rep.* **2021**, *11*, 23674. [[CrossRef](#)]
33. Hu, B.-L.; Xie, M.-Z.; Li, K.-Z.; Li, J.-L.; Gui, Y.-C.; Xu, J.-W. Genome-wide analysis to identify a novel distant metastasis-related gene signature predicting survival in patients with gastric cancer. *Biomed. Pharmacother.* **2019**, *117*, 109159. [[CrossRef](#)]
34. Li, Y.; Sun, R.; Zhang, Y.; Yuan, Y.; Miao, Y. A methylation-based mRNA signature predicts survival in patients with gastric cancer. *Cancer Cell Int.* **2020**, *20*, 284. [[CrossRef](#)] [[PubMed](#)]
35. Şenbabaoğlu, Y.; Michailidis, G.; Li, J.Z. Critical limitations of consensus clustering in class discovery. *Sci. Rep.* **2014**, *4*, 6207. [[CrossRef](#)] [[PubMed](#)]
36. Liu, Y.; Wu, J.; Huang, W.; Weng, S.; Wang, B.; Chen, Y.; Wang, H. Development and validation of a hypoxia-immune-based microenvironment gene signature for risk stratification in gastric cancer. *J. Transl. Med.* **2020**, *18*, 201. [[CrossRef](#)] [[PubMed](#)]
37. Chen, J.; Ou, Y.; Luo, R.; Wang, J.; Wang, D.; Guan, J.; Li, Y.; Xia, P.; Chen, P.R.; Liu, Y. SAR1B senses leucine levels to regulate mTORC1 signalling. *Nature* **2021**, *596*, 281–284. [[CrossRef](#)]
38. Zhang, C.; Cheng, W.; Ren, X.; Wang, Z.; Liu, X.; Li, G.; Han, S.; Jiang, T.; Wu, A. Tumor Purity as an Underlying Key Factor in Glioma Tumor Purity in Glioma. *Clin. Cancer Res.* **2017**, *23*, 6279–6291. [[CrossRef](#)]
39. Boutilier, A.J.; Elswa, S.F. Macrophage polarization states in the tumor microenvironment. *Int. J. Mol. Sci.* **2021**, *22*, 6995. [[CrossRef](#)]
40. Sarvaria, A.; Madrigal, A.; Saudemont, A. B cell regulation in cancer and anti-tumor immunity. *Cell. Mol. Immunol.* **2017**, *14*, 662–674. [[CrossRef](#)]
41. Kandoth, C.; McLellan, M.D.; Vandin, F.; Ye, K.; Niu, B.; Lu, C.; Xie, M.; Zhang, Q.; McMichael, J.F.; Wyczalkowski, M.A.; et al. Mutational landscape and significance across 12 major cancer types. *Nature* **2013**, *502*, 333–339. [[CrossRef](#)] [[PubMed](#)]
42. Li, X.; Pasche, B.; Zhang, W.; Chen, K. Association of *MUC16* Mutation With Tumor Mutation Load and Outcomes in Patients With Gastric Cancer. *JAMA Oncol.* **2018**, *4*, 1691–1698. [[CrossRef](#)]
43. Huang, Y.J.; Cao, Z.F.; Wang, J.; Yang, J.; Wei, Y.-H.; Tang, Y.-C.; Cheng, Y.-X.; Zhou, J.; Zhang, Z.X. Why *MUC16* mutations lead to a better prognosis: A study based on The Cancer Genome Atlas gastric cancer cohort. *World J. Clin. Cases* **2021**, *9*, 4143. [[CrossRef](#)] [[PubMed](#)]
44. Myung, J.K.; Cho, H.J.; Park, C.K.; Kim, S.K.; Phi, J.H.; Park, S.H. *IDH1* mutation of gliomas with long-term survival analysis. *Oncol. Rep.* **2012**, *28*, 1639–1644. [[CrossRef](#)] [[PubMed](#)]
45. Yao, Q.; Cai, G.; Yu, Q.; Shen, J.; Gu, Z.; Chen, J.; Shi, W.; Shi, J. *IDH1* mutation diminishes aggressive phenotype in glioma stem cells. *Int. J. Oncol.* **2018**, *52*, 270–278. [[CrossRef](#)]
46. Zhu, Z.; Fu, H.; Wang, S.; Yu, X.; You, Q.; Shi, M.; Dai, C.; Wang, G.; Cha, W.; Wang, W. Whole-exome sequencing identifies prognostic mutational signatures in gastric cancer. *Ann. Transl. Med.* **2020**, *8*, 1484. [[CrossRef](#)]
47. Wang, K.; Yuen, S.T.; Xu, J.; Lee, S.P.; Yan, H.H.N.; Shi, S.T.; Siu, H.C.; Deng, S.; Chu, K.M.; Law, S.; et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **2014**, *46*, 573–582. [[CrossRef](#)]
48. Reske, J.J.; Wilson, M.R.; Holladay, J.; Siwicki, R.A.; Skalski, H.; Harkins, S.; Adams, M.; Risinger, J.I.; Hostetter, G.; Lin, K.; et al. Co-existing *TP53* and *ARID1A* mutations promote aggressive endometrial tumorigenesis. *PLoS Genet.* **2021**, *17*, e1009986. [[CrossRef](#)]
49. Zang, Z.J.; Cutcutache, I.; Poon, S.L.; Zhang, S.L.; McPherson, J.R.; Tao, J.; Rajasegaran, V.; Heng, H.L.; Deng, N.; Gan, A.; et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat. Genet.* **2012**, *44*, 570–574. [[CrossRef](#)] [[PubMed](#)]
50. Qiao, Y.; Li, T.; Zheng, S.; Wang, H. The Hippo pathway as a drug target in gastric cancer. *Cancer Lett.* **2018**, *420*, 14–25. [[CrossRef](#)]
51. Magnelli, L.; Schiavone, N.; Staderini, F.; Biagioni, A.; Papucci, L. MAP kinases pathways in gastric cancer. *Int. J. Mol. Sci.* **2020**, *21*, 2893. [[CrossRef](#)] [[PubMed](#)]
52. Cohen, D.A.; Dabbs, D.J.; Cooper, K.L.; Amin, M.; Jones, T.E.; Jones, M.W.; Chivukula, M.; Trucco, G.A.; Bhargava, R. Interobserver agreement among pathologists for semiquantitative hormone receptor scoring in breast carcinoma. *Am. J. Clin. Pathol.* **2012**, *138*, 796–802. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.