

Article

Bioinformatics Prediction and Machine Learning on Gene Expression Data Identifies Novel Gene Candidates in Gastric Cancer

Medi Kori ¹  and Esra Gov ^{2,*} ¹ Department of Bioengineering, Marmara University, Istanbul 34854, Turkey² Department of Bioengineering, Adana Alparslan Türkeş Science and Technology University, Adana 01250, Turkey

* Correspondence: egov@atu.edu.tr

Abstract: Gastric cancer (GC) is one of the five most common cancers in the world and unfortunately has a high mortality rate. To date, the pathogenesis and disease genes of GC are unclear, so the need for new diagnostic and prognostic strategies for GC is undeniable. Despite particular findings in this regard, a holistic approach encompassing molecular data from different biological levels for GC has been lacking. To translate Big Data into system-level biomarkers, in this study, we integrated three different GC gene expression data with three different biological networks for the first time and captured biologically significant (i.e., reporter) transcripts, hub proteins, transcription factors, and receptor molecules of GC. We analyzed the revealed biomolecules with independent RNA-seq data for their diagnostic and prognostic capabilities. While this holistic approach uncovered biomolecules already associated with GC, it also revealed novel system biomarker candidates for GC. Classification performances of novel candidate biomarkers with machine learning approaches were investigated. With this study, AES, CEBPZ, GRK6, HPGDS, SKIL, and SP3 were identified for the first time as diagnostic and/or prognostic biomarker candidates for GC. Consequently, we have provided valuable data for further experimental and clinical efforts that may be useful for the diagnosis and/or prognosis of GC.

Keywords: gastric cancer; disease genes; diagnostic genes; prognostic genes; multi-omics; systems biology



Citation: Kori, M.; Gov, E. Bioinformatics Prediction and Machine Learning on Gene Expression Data Identifies Novel Gene Candidates in Gastric Cancer. *Genes* **2022**, *13*, 2233. <https://doi.org/10.3390/genes13122233>

Academic Editors: Kenta Nakai and Tun-Wen Pai

Received: 31 October 2022

Accepted: 25 November 2022

Published: 28 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 one of the leading causes of cancer deaths worldwide with a high prevalence. According to recent reports, GC is responsible for one in 13 deaths worldwide and was the fifth most common cancer worldwide in 2020 [1]. *Helicobacter pylori* infection is the main risk factor for GC, but other factors such as genetic and environmental factors also play a role [2]. Because GC is a heterogeneous disease, it is an attractive model for studying carcinogenesis and tumorigenesis. The exact mechanisms underlying the development of GC are still unknown despite significant progress in understanding the molecular causes of GC. Malignant transformation of gastric mucosa during the multistep process of GC pathogenesis is caused by a variety of genetic and molecular abnormalities that occur in GC [3]. Since one of the major causes of treatment failure in GC is drug resistance, a deeper knowledge of novel gene candidates is crucial to better understand the molecular mechanism of pathogenesis, which could improve patient survival [4].

Integrating multi-omics data can reveal the entire physical and functional architecture of cellular signaling and regulatory pathways. Moreover, it has been reported that a systems medicine approach involving the integration of gene expression data with multi-omics data reveals important and crucial genes in a pathological state [5]. Today, systems medicine is being used in several studies to identify significant disease genes, for example, in papillary thyroid cancer [6], acute myeloid leukemia [7], abdominal aortic aneurysm [8], ovarian

cancer stem cells [9], rheumatoid arthritis [10], colorectal cancer [11], and three different ovarian diseases [12].

The rapid development of cancer genomics has led to extensive information on crucial genes in malignancies. Although many bioinformatics web servers and tools have been developed to identify disease genes [13,14], there are still no clinically validated treatments and their efficacy is controversial for the vast majority of cancer genes. Recently, key genes [15], prognostic genes [16], and prognostic and diagnostic genes for GC [17] have been reported. In another study conducted by our research group, we analyzed several microarray datasets of GC and identified a prognostic differential co-expressed gene module and offered drug candidates by repurposing analysis [18].

Although several studies have been performed to identify disease genes in GC, the pathogenesis of GC is still unclear and new and efficient biomarker candidates are needed. In parallel with the advances in high-throughput technologies and the increasing number of omics data, the increasing use of systems biology approaches to understand diseases at the systems level and provide biomarker candidates is becoming increasingly important. In our study, we used an integrative multi-omics approach that differs from previous GC studies. In the present study, we adopted a systems biology approach by integrating gene expression data with comprehensive human biological networks to identify molecular signatures that allow us to identify important biomolecules which can be considered as biomarkers associated with GC. Accordingly, a meta-analysis of GC-associated transcriptomic datasets was performed and common differentially expressed genes (DEGs) were identified among the datasets. Gene set overrepresentation analysis was performed for the common DEGs. The DEGs were integrated into various human biological networks, including protein–protein interactions (PPI), transcriptional regulation, and protein–receptor interaction networks to identify hub proteins, reporter transcription factors (TFs), and reporter receptors. The diagnostic and prognostic value of reporter biomolecules was assessed using an independent cohort study. Finally, the novel prognostic and diagnostic biomarker candidates for GC were revealed and their classification power was evaluated using machine learning approaches (Figure 1). Consequently, we believe that the novel biomarker candidates presented here will be a crucial resource for understanding the pathogenesis of GC and can be considered as powerful diagnostic and prognostic biomarkers for further experimental and clinical studies for GC.

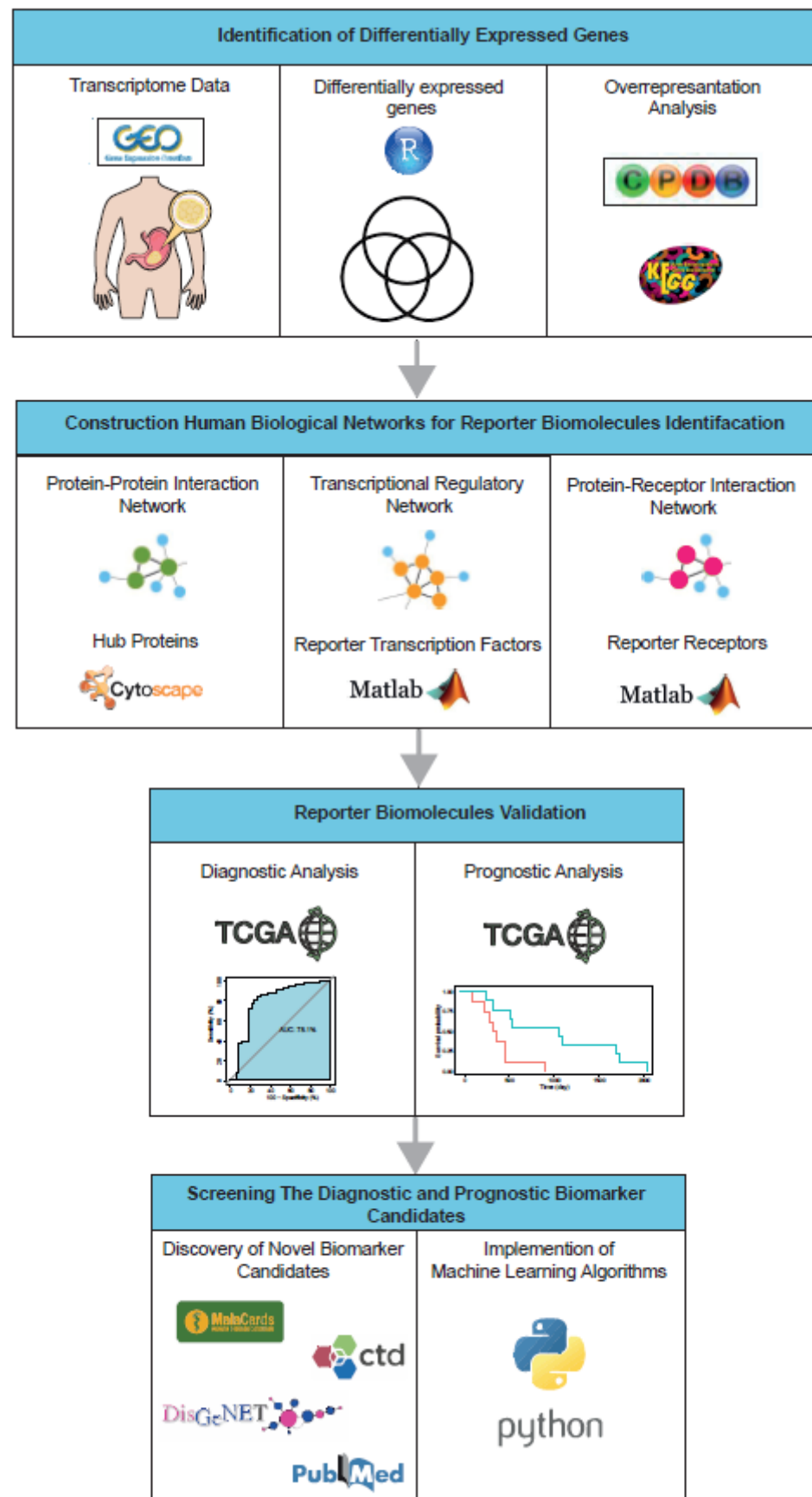


Figure 1. The computational flow employed in the study.

2. Materials and Methods

2.1. Gene Expression Datasets of Gastric Carcinoma

Three microarray datasets, including GSE19826 [19], GSE54129 (unpublished), and GSE79973 [20] from independent studies, were obtained from the Gene Expression Omnibus (NCBI-GEO) database [21] to meta-analyze transcriptome profiles in GC. Datasets were selected based on the following criteria: (i) samples consisted of two different phenotypes (i.e., cancerous vs. normal); (ii) each phenotype included at least 3 samples; (iii) the microarray platforms used were from the same platform. A total of 179 samples were analyzed, including 133 GC and 46 normal gastric tissue samples.

An independent adenocarcinoma dataset of the stomach (STAD) from The Cancer Genome Atlas (TCGA) [22], comprising 375 GC and 32 normal stomach tissue samples, was used as a validation dataset, for preclinical validation purposes (i.e., diagnostic and prognostic analyses), and for implementing machine learning approaches.

2.2. Identification of Differentially Expressed Genes

In this study, a well-established statistical analysis procedure [23,24] was used to identify DEGs. Briefly, the raw data (stored in CEL files) of each dataset were normalized by calculating the Robust Multi-Array Average (RMA) expression measure [25] implemented in the Affy package [26] of the R/Bioconductor platform (version 4.0.2) [27]. DEGs were identified from normalized expression values using the Linear Models for Microarray Data package (LIMMA) [28]. The Benjamini–Hochberg method was used to control for the false discovery rate (FDR). The adjusted p -value < 0.05 was used as the cutoff value to determine the statistical significance of the DEGs. To determine the regulatory patterns of DEGs, the fold-change thresholds were used as 2-fold change. Each data set was analyzed independently, and the results were comparatively analyzed to identify common signatures from these independent studies, and common DEGs were used in further analyses.

2.3. Gene Set Overrepresentation Analyses

Overrepresentation analyses were performed using ConsensusPathDB [29] to determine the functional annotations (i.e., biological pathways) of the DEGs. In the analyses, KEGG [30] and Reactome [31] were employed as the data sources for pathways. The p -values were determined using Fisher's exact test, and a false discovery rate was applied to control the p -values. An adjusted p -value < 0.01 was considered statistically significant.

2.4. Reconstruction of Protein–Protein Interaction Network and Identification of Hub Proteins

Physical PPIs among DEGs were extracted from the BioGRID database (MV-Physical-4.2.191) [32], which contains 51,745 physical and experimentally detected PPIs among 10,177 human proteins. The PPI sub-network was reconstructed for common DEGs with their first neighbors and visualized using Cytoscape (v3.5.0) [33]. To determine hub proteins (i.e., central proteins), topological analyses were performed using the Cytohubba plugin [34]. The dual metric approach that considers degree and betweenness centrality metrics (i.e., degree as a local metric and betweenness centrality as a global metric) was simultaneously used to identify hub proteins. The 10 proteins with the highest degree and betweenness centrality values in the PPI subnetwork were determined as hub proteins.

2.5. Identification of Reporter Transcription Factors and Receptors

The reporter molecules were identified using the reporter features algorithm [35], which was previously adapted for potential TFs and receptors [36]. Briefly, reporters were identified by integrating common DEGs gene expression data with relevant human biological networks (i.e., TF–target gene interactions and receptor–protein interactions). TF–target gene interaction information was obtained from the TRRUST (transcriptional regulatory relationships unraveled by sentence-based text-mining) database [37]. The proteins with receptor activity (GO: 0004872) were extracted from DAVID [38], PANTHER [39], and GeneCodis [40] databases, and the physical interactions of these receptors were extracted

from the human PPI network [32]. The reporter features algorithm [35] was implemented in MATLAB (R2016). The p -values of the calculated reporter molecules were controlled with FDR, and the reporter molecules with adjusted p -value < 0.001 were considered significant. Reporter TFs and receptors functions were analyzed using the PANTHER classification system [39].

2.6. Pre-Clinical Diagnostic Validation of Reporter Biomolecules

To evaluate the diagnostic performance of reporter biomolecules (i.e., hub proteins, TFs, and receptors), a receiver-operating characteristic curve (ROC) was used that utilized the parameters of sensitivity and specificity to predict diagnostic ability. To determine the overall diagnostic accuracy of diagnostic performance, the area under the roc curve (AUC) was calculated. A reporter biomolecule with an AUC value $\geq 70\%$ was considered statistically significant [41] and accepted as a diagnostic biomolecule.

2.7. Pre-Clinical Prognostic Validation of Reporter Biomolecules

To evaluate the prognostic performance of the reporter biomolecules (i.e., hub proteins, TFs, and receptors), we obtained clinical information from STAD samples of TCGA [22] and used it in the prognostic performance analyses. To determine the prognostic performance of each biomolecule, survival analyses were performed by dividing subjects into two groups (high and low risk) according to their prognostic index (PI), which is the linear component of the Cox model. The differences in gene expression values between the risk groups were represented by box plots. Survival signatures of reporter biomolecules were evaluated by Kaplan–Meier plots. The hazard ratio ($HR = (O1/E1)/(O2/E2)$) was calculated using the ratio between the relative mortality rate in group 1 and the relative mortality rate in group 2, where O and E are the observed and expected number of deaths, respectively. Reporter biomolecules with a log-rank p -value < 0.05 were considered statistically significant and accepted as prognostic biomolecules.

2.8. Screening the Association of Diagnostic and Prognostic Reporter Biomolecules with Gastric Cancer

Following the two preclinical validation analyses, an extensive search was performed to determine whether the diagnostic or/and prognostic reporter biomolecules found in the study had been previously associated with GC. The following databases and electronic search services were used throughout the association screening process: Malacards: The Human Disease Database [42], DisGeNET: a comprehensive platform for integrating information on genes and variants associated with human diseases [43], Comparative Toxicogenomics Database (CTD) [44], PubMed, Science-Direct, Scopus, and Web of Science. The reporter biomolecules, which have diagnostic or/and prognostic capabilities and were not associated with GC according to previous studies, were considered as novel biomarker candidates in this study.

2.9. Investigation of Classification Performances of Novel Candidate Biomarkers with Machine Learning Approaches

To better interpret new biomarker candidates, we applied several classification methods, a well-known and useful machine learning technique in biomarker discovery, to identify novel biomarker candidates. We implemented different classification algorithms, including K-Neighbors, MLP, Decision Tree, Random Forest, Gradient Boosting, CatBoost, LGBM, and XGB using the Python programming language [45]. The performance of these techniques was estimated based on the predictive accuracy of the classifiers.

3. Results

3.1. The Transcriptomic Signatures of Gastric Cancer: Identification of Differentially Expressed Genes

The individual statistical analyses of three gene expression datasets (GSE19826, GSE54129, and GSE79973) led to the identification of DEGs. The number of DEGs in each dataset showed a wide range from 791 to 4358 genes, and the highest number of DEGs was

identified in GSE54129. In all three datasets, no significant tendency toward a particular regulatory pattern (up- or down-regulation) was detected in the culminated DEGs; in other words, the difference between up- and down-regulation did not exceed 5%. Nevertheless, in both datasets (i.e., GSE19826 and GSE79973), up-regulated DEGs predominated (51.4% and 52.6%, respectively) compared with down-regulated DEGs. On the other hand, DEGs in dataset GSE54129 showed a stronger pattern of down-regulation compared with up-regulation (Figure 2A).

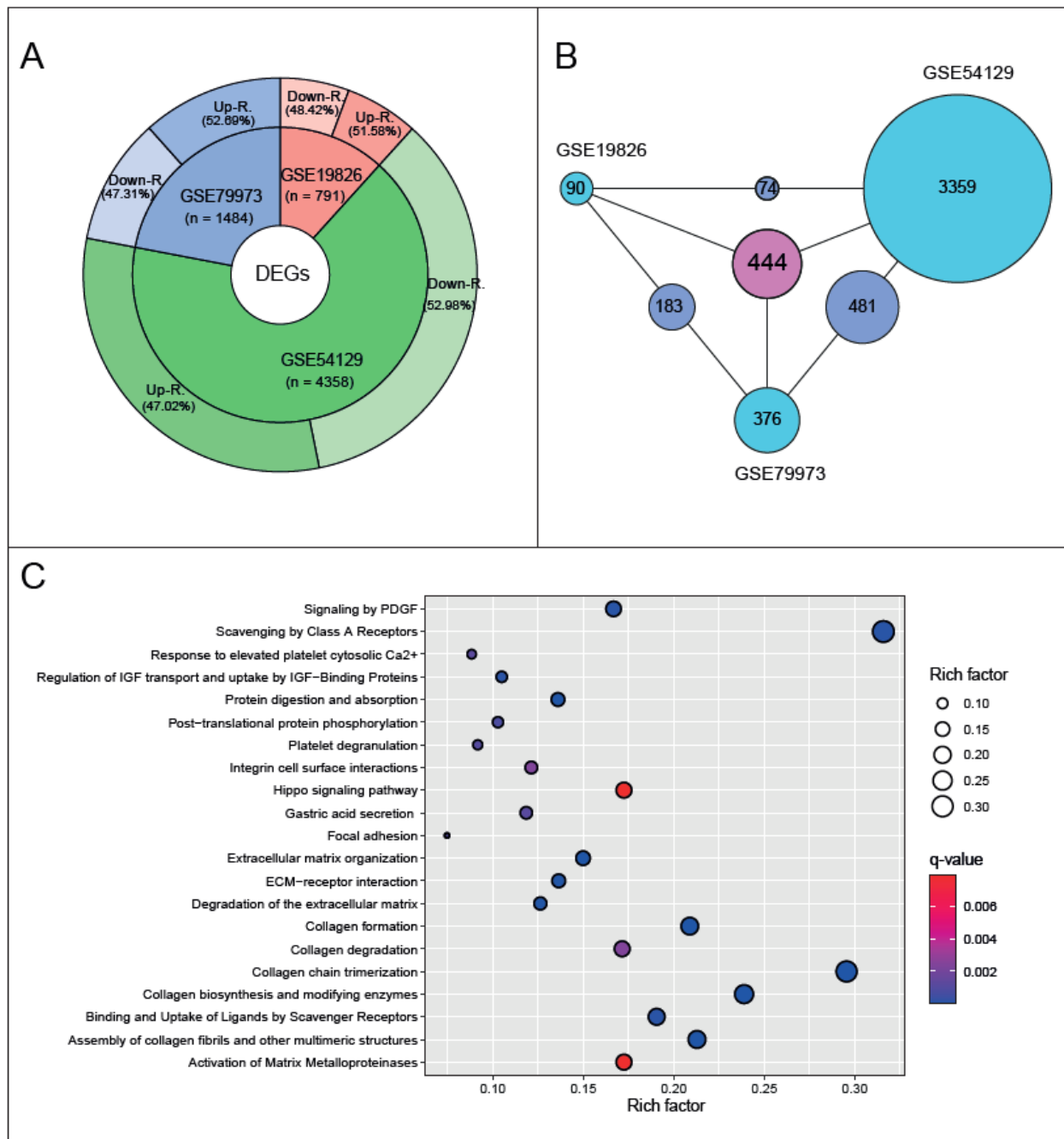


Figure 2. Meta-analysis of the three transcriptome datasets associated with gastric cancer. (A) Pie donut diagram shows the distribution of differentially expressed genes (DEGs) of the three transcriptome datasets. (B) The Venn diagram shows the DEGs common to the datasets. (C) The gene set overrepresentation analysis of the common DEGs.

Excluding the regulatory patterns of DEGs, the comparative analysis of the resulting DEGs showed that a total of 444 DEGs were common in the three GEO datasets (Figure 2B). To ensure consistency of the analysis, further analyses were performed using these common DEGs.

Overrepresentation analyses indicated that the common DEGs were significantly associated with cancer-associated molecular pathways, such as Hippo signaling, focal adhesion extracellular matrix (ECM) receptor interaction. Several processes that were associated with collagen synthesis or degradation were highlighted as pathways for common DEGs. Moreover, gastric acid segregation and protein digestion and the absorption pathway, which were highly associated with each other, come into prominence in overrepresentation analysis (Figure 2C).

3.2. The Proteomic Signatures of Gastric Cancer: Identification of Hub Proteins

To identify hub proteins, a PPI subnetwork was reconstructed around proteins encoded by the common DEGs of GC. The reconstructed network consisted of 974 proteins (i.e., 444 common proteins and their physically interacting first neighbors) and 1025 links (i.e., physical PPIs between these proteins).

The PPI network showed a scale-free topology and indicated the presence of hub proteins. Hub proteins, which play a central role in modular organization and information flow within the network, were identified by topological analysis that included degree and betweenness centrality metrics. The 10 proteins with the highest degree and betweenness centrality values were combined together and determined as hub proteins. As a result, a total of 15 hub proteins were determined, namely ACTN1, AGR2, BAG2, BMPR1A, DTL, FLNA, FN1, LGALS1, MECOM, MUC1, NEDD4L, PDGFRB, PDLIM7, TP53, and TRIM29 (Figure 3A).

3.3. The Regulatory Signatures of Gastric Cancer: Identification of Reporter Transcription Factors

The regulatory elements (i.e., TFs) controlling key transcriptional changes in GC genes were identified by integrating common DEGs with the transcriptional regulatory network using the reporter features algorithm. Accordingly, 20 TFs emerged with a significance level of p -value < 0.001 and were identified as reporter regulatory elements in the transcriptional control of genes in GC (Figure 3B). These reporter TFs included two rel homology transcription factors (NFKB1 and RELA), two zinc finger transcription factors (SP1 and SP3), and two DNA-binding transcription factors (CEBPZ and GZF1).

3.4. The Signaling Signatures of Gastric Cancer: Identification of Reporter Receptors

Reporter receptors of GC were determined in a similar manner that was used to determine reporter TFs. To identify reporter receptors, we integrated common DEGs with the receptor–protein interaction network by using the reporter features algorithm. According to the results, 23 proteins were identified as reporter receptors with a significance level of p -value < 0.001 (Figure 3C). Among the 23 reporter receptors, seven proteins belonged to the metalloprotease family (ECE1, MMP1, MMP14, MMP2, MMP3, MMP8, and MMP9), three reporter receptors belonged to the transmembrane signaling receptors (GRIK1, GRIK3, and TIE1), and three reporter receptors belonged to the G protein-coupled receptors (DRD1, GRM7, and GRM8).

3.5. Diagnostic and Prognostic Power of Reporter Biomolecules of Gastric Cancer

To pre-clinically validate the diagnostic and prognostic capabilities of the discovered reporter biomolecules, independent expression data from TCGA (TCGA-STAD) were used. The diagnostic property of each module was evaluated using ROC curves, and a reporter biomolecule with an AUC value $\geq 70\%$ was considered statistically significant and accepted as diagnostic. Subsequently, five hub proteins (33.3% of total hubs), 12 reporter TFs (60% of total reporter TFs), and 14 reporter receptors (60.8% of total reporter receptors) were found to be diagnostic reporter biomolecules (Figure 4A). Among the diagnostic biomolecules, a

hub protein, DTL, with an AUC score of 95%, a reporter TF, HOXC8, with an AUC score of 93.1%, and a reporter receptor, BUB1, with an AUC score of 94.1% were the most important diagnostic reporter biomolecules when statistical significance was considered (Figure 4B).

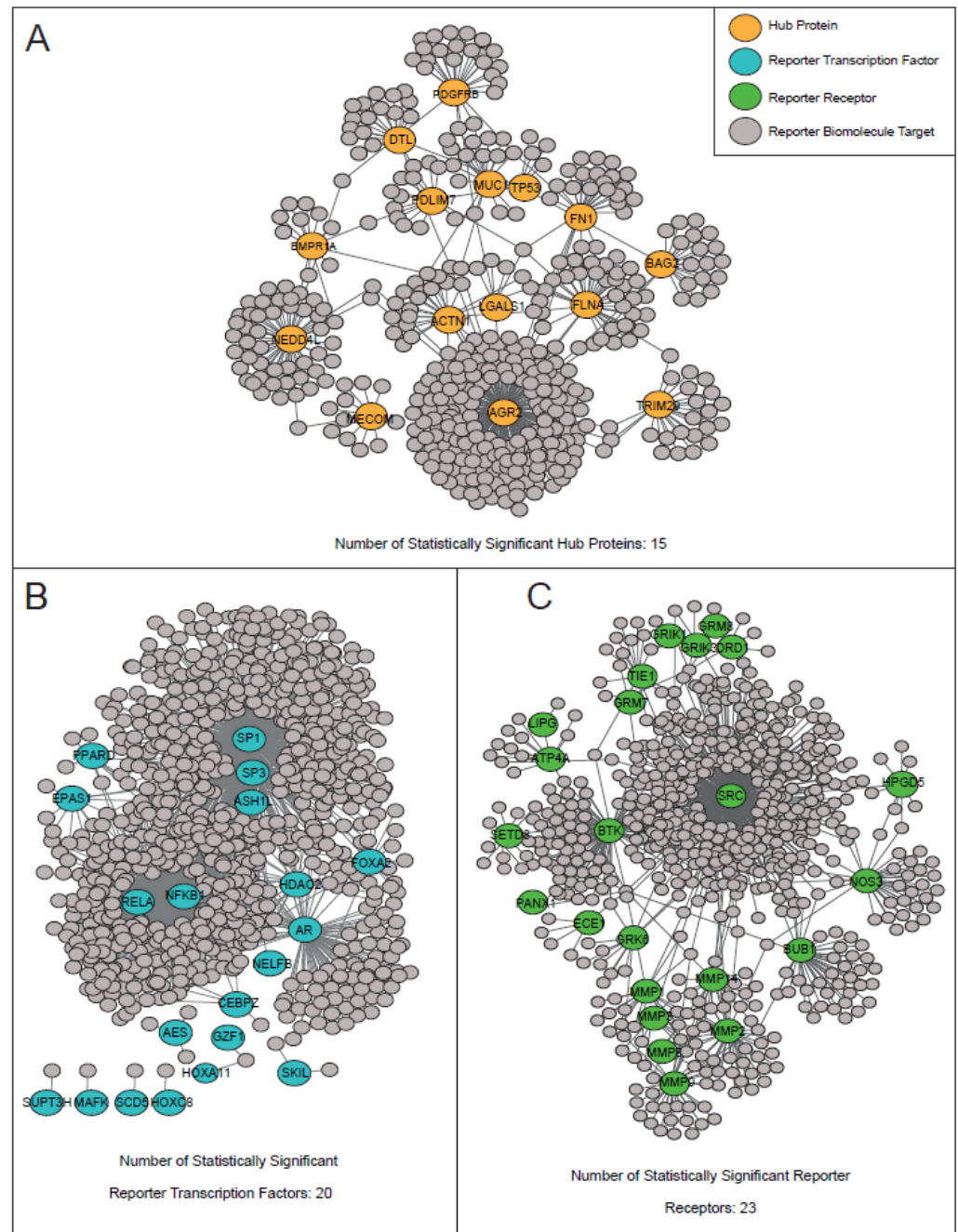


Figure 3. The reconstructed human biological networks. (A) The reconstructed protein–protein interaction (PPI) network. The revealed significant hub proteins according to employed topological parameters were shown in orange. (B) The reconstructed transcriptional regulatory interaction network. The statistically significant (p -value < 0.001) reporter transcription factors (TFs) were shown in blue. (C) The reconstructed protein–receptor interaction network interaction network. The statistically significant (p -value < 0.001) reporter receptors were shown in green.

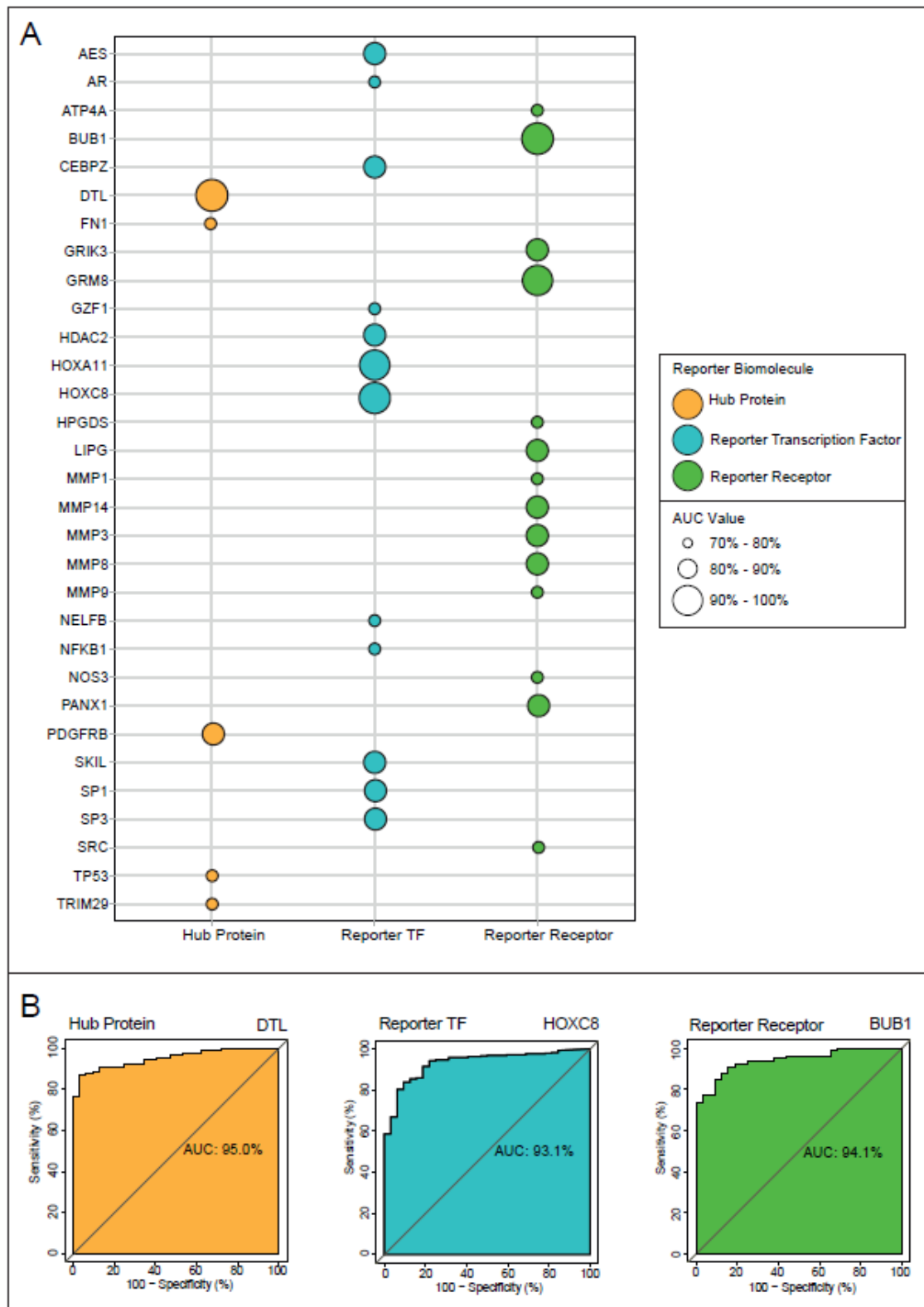


Figure 4. The diagnostic performance analyses of the reporter biomolecules. **(A)** The bubble plot representing the AUC values of the reporter biomolecules. Only the AUC values that were considered significant in the study were shown (AUC > 70%). Hub proteins are shown in orange, reporter transcription factors (TFs) in blue, and reporter receptors in green. **(B)** The major reporter biomolecules: a hub protein, a TF, and a reporter receptor according to their AUC values.

Patient information on overall survival was extracted from data from TCGA-STAD and used for prognostic performance analysis. Prognostic performance of reporter biomolecules was assessed using Kaplan–Meier survival charts based on risk groups and days of survival. The log-rank p -value and hazard ratios were considered to determine whether reporter biomolecules had a high impact on overall patient survival. As a result, a total of three hub proteins (i.e., PDGFRB, TP53, and TRIM29), four reporter TFs (i.e., AR, HOXA11, NELFB, and SKIL), and one reporter receptor (GRK6) had a high impact on patients' overall survival (log-rank p -value < 0.05). In addition, the differences in the expression levels of genes (encoding hub proteins, reporter TFs, or reporter receptors) between the risk groups showed that up-regulation of the expression of PDGFRB, AR, and SKIL was associated with a higher risk of GC, while down-regulation of the expression of TP53, TRIM29, HOXA11, NELFB, and GRK6 was associated with a higher risk of GC (Figure 5). With the exception of GRK6, all prognostic reporter biomolecules also showed high diagnostic performance. Thus, seven reporter biomolecules, namely AR, HOXA11, NELFB, PDGFRB, SKIL, TP53, and TRIM29 showed both statistically significant diagnostic and prognostic properties for GC. The reporter biomolecules that exhibited diagnostic or/and prognostic properties were considered as biomarker candidates for GC.

3.6. The Association of Diagnostic and Prognostic Reporter Biomolecules with Gastric Cancer

A total of 32 candidate biomarkers were identified by bioinformatics analysis. To determine whether the candidate biomarkers found were associated with GC from previous studies or were discovered for the first time with our study, we primarily examined GC-associated biomarkers and genes from three different publicly available databases. As a result, we obtained 1224 different GC-related biomarkers/genes from data repositories [42–44], including the 13 candidate biomarkers we proposed in this study. For the remaining 19 candidate biomarkers, we manually reviewed electronic search services and found that 13 candidate biomarkers had been previously associated with GC [46–58]. Consequently, we concluded that, to our knowledge, six diagnostic or/and prognostic reporter biomolecules, including AES, CEBPZ, GRK6, HPGDS, SKIL, and SP3, are proposed here for the first time as GC biomarker candidates (Table 1).

3.7. Classification Powers of Novel Candidate Biomarkers

A machine learning technique, classification, can be used to evaluate the potential of biomarkers identified by various statistical tests. Because an effective potential biomarker should be able to distinguish the diseased cohort from controls, we used several classification algorithms to determine the potential of our novel biomarkers. We evaluated the novel biomarker candidates based on the predictive accuracy of the classifier and found that the accuracy of the eight different classification methods ranged from 92.6% to 89.4% (Figure 6A), suggesting that the novel GC biomarkers we have provided here can efficiently discriminate the diseased samples from the controls. In addition, we used clinical data from our validation dataset (i.e., STAD-TCGA) [22] to test whether our candidates were informative in classifying alive and dead specimens. The accuracy of the classification results showed that our proposed novel biomarkers were not as successful in evaluating live and dead specimens compared with diseased and control specimens (accuracy ranged from 64.6% to 47.7%) (Figure 6B).

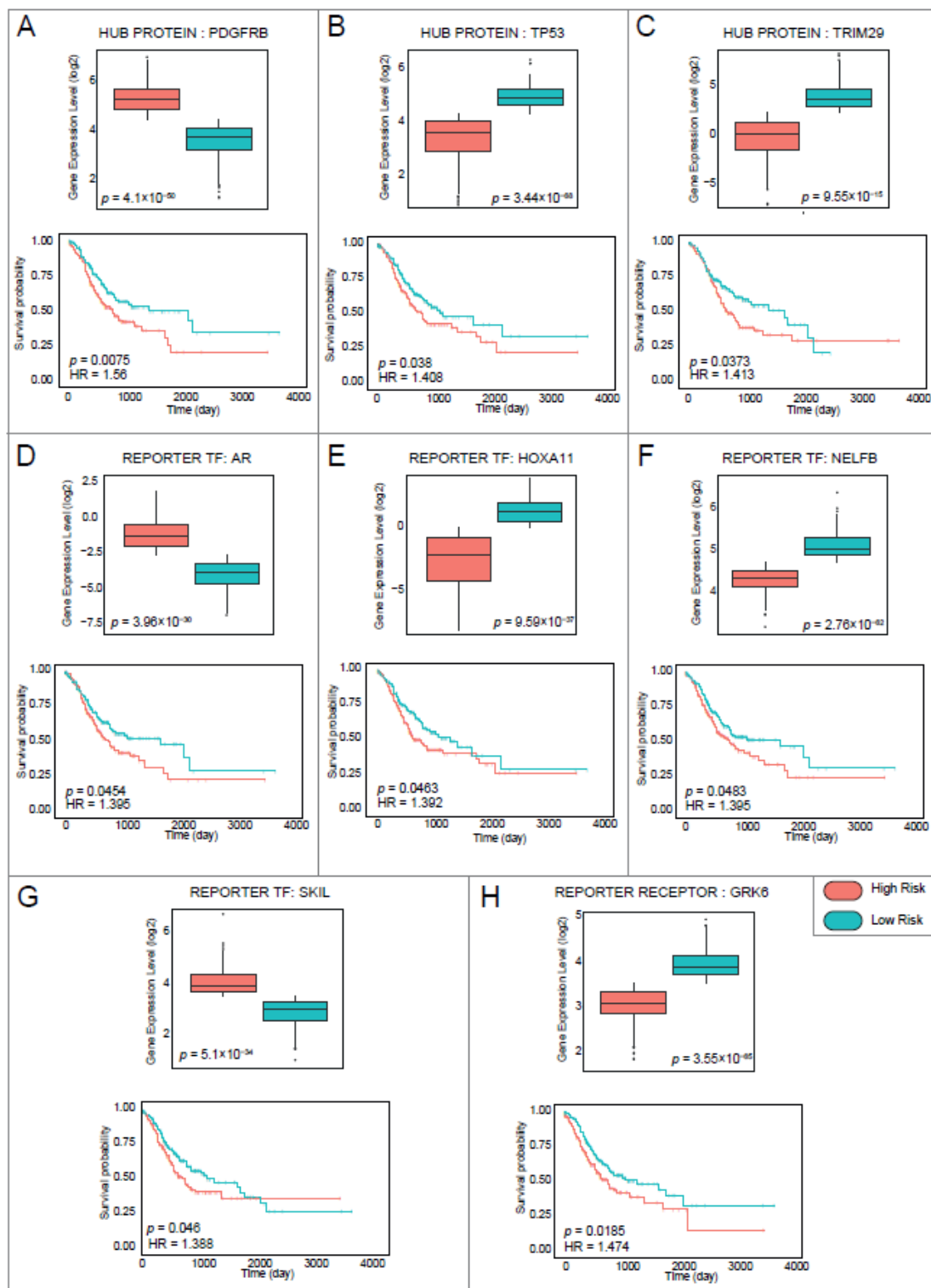
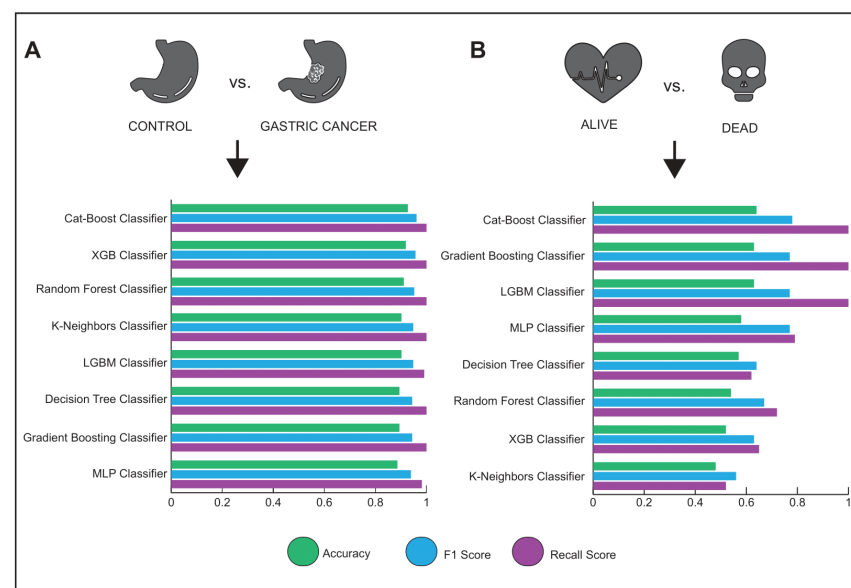


Figure 5. Analysis of prognostic performance of reporter biomolecules. Box plots showing expression levels of reporter biomolecules between low and high-risk groups with p -values. Kaplan–Meier plots estimating survival of patients with gastric cancer showing p -value and hazard ratio for each curve. (A) Hub protein: PDGFRB. (B) Hub protein: TP53. (C) Hub protein: TRIM29. (D) Reporter transcription factor (TF): AR. (E) Reporter TF: HOXA11. (F) Reporter TF: NELFB. (G) Reporter TF: SKIL. (H) Reporter receptor: GRK6. The high-risk group was shown in red, while the low-risk group was shown in blue.

Table 1. The association of diagnostic and prognostic candidate biomarkers with gastric cancer.

Type	Name	Diagnostic?	Prognostic?	Association with Gastric Cancer
Reporter Transcription Factor	DTL	+	−	[46]
	FN1	+	−	[42,43]
	PDGFRB	+	+	[42,43]
	TP53	+	+	[42–44]
	TRIM29	+	+	[47]
	AES	+	−	Novel
	AR	+	+	[42,43]
	CEBPZ	+	−	Novel
	GZF1	+	−	[48]
	HDAC2	+	−	[49]
	HOXA11	+	+	[50]
	HOXC8	+	−	[51]
	NELFB	+	+	[52]
	NFKB1	+	−	[42,43]
	SKIL	+	+	Novel
	SP1	+	−	[53]
	SP3	+	−	Novel
	ATP4A	+	−	[42]
	BUB1	+	−	[42]
	GRIK3	+	−	[54]
GRK6	−	+	Novel	
GRM8	+	−	[55]	
HPGDS	+	−	Novel	
LIPG	+	−	[56]	
MMP1	+	−	[42,43]	
MMP14	+	−	[42,43]	
MMP3	+	−	[42,43]	
MMP8	+	−	[57]	
MMP9	+	−	[42,43]	
NOS3	+	−	[42–44]	
PANX1	+	−	[58]	
SRC	+	−	[42,43]	

**Figure 6.** Machine learning analysis for novel diagnostic and/or prognostic biomarker candidates. (A) The accuracy, F1, and recall score plot of eight different classification algorithms for discriminating between diseased samples and controls. (B) The accuracy, F1, and recall score plot of eight different classification algorithms for discriminating between alive and dead samples.

4. Discussion

It is estimated that GC will be responsible for 770,000 deaths and 1.1 million new cancer cases worldwide in 2020. Worse, it is predicted that by 2040, GC cases will result in approximately 1.3 million deaths and approximately 1.8 million people will be diagnosed with the disease [59]. Although many GC studies have accumulated in the scientific community to date, recent cancer statistics estimate the global burden of GC and clearly demonstrate the need for new diagnostic and prognostic strategies for GC. Despite these GC-based studies, the intertwined structure of the cell has not been considered, which requires the integration of biological data (i.e., expression data) with human biological networks. To translate Big Data into system-level biomarkers, in this study, we integrated GC expression data with three different biological networks for the first time and captured transcripts, hub proteins, TFs, and receptor molecules of GC. In addition, to determine a reporter biomolecule as a “biomarker,” we assessed its diagnostic and prognostic performance in an independent cohort.

Based on individual analysis of three gene expression datasets, we found that hundreds of genes were differentially expressed in each dataset. However, to increase the reliability and robustness of the results, we combine information from multiple microarray datasets and focus only on common 444 DEGs. Analysis of the overrepresentation of common DEGs revealed significant biological pathways. Interestingly, four pathways related to collagen synthesis or degradation were found to be significant. It was known that the restructuring of the collagen components of the tumor microenvironment had a remarkable impact on cancer development and progression. For GC, it was reported that the collagen components in the tumor microenvironment rearrange quantitatively and qualitatively, and there was a significant correlation between the prognosis of GC and collagen. The study even concluded that collagen width can be used as a prognostic indicator for GC [60].

Reconstruction and topological analysis of the PPI network around the proteins encoded by the 444 common DEGs led to the identification of hub proteins that play a central role in the flow of information within the network. A total of 15 hub proteins appeared as reporter signal mediators in GC. Among them, PDGFRB, TP53, and TRIM29 have shown both high diagnostic and prognostic capacity, while DTL and FN1 have shown only high diagnostic capacity. These diagnostic and/or prognostic biomarker candidates have already been associated with GC (Table 1), so these results further strengthen our confidence in our observations.

In this study, the reporter features algorithm was adapted to identify reporter TFs and receptors. Transcriptional expression of the common transcripts of GC was controlled by 20 TFs, whereas 23 receptors played a central role in signal transduction. Of these reporters, 12 TFs and 15 reporter receptors showed significant diagnostic and/or prognostic results when cross-validation analysis was performed with independent RNA-seq data. Accordingly, eight TFs and 13 reporter receptors have already been shown to be associated with GC (Table 1). However, to the best of our knowledge, AES, CEBPZ, GRK6, HPGDS, SKIL, and SP3 have not yet been associated with GC and were considered as novel biomarker candidates in this study. It was found that novel biomarker candidates efficiently discriminate the diseased samples from the controls compared to the performance of discriminating in the live and dead specimens. Since, it may suggest that biomarker candidates have the diagnostic capability of GC.

AES (also known as TLE5) is a transcriptional modulator and a transcriptional co-repressor that represses associated proteins of the Groucho/TLE family. AES plays an active role in the formation and development of organs or cells such as heart, pituitary, ear, and blood cells [61]. AES has been associated with several types of cancer. Deficiency of AES has been shown to lead to invasion and metastasis of prostate cancer [62]. Similarly, it suppresses colon cancer invasion and metastasis by inhibiting the Notch signaling pathway [63].

A TF, CEBPZ, acts as an activator or suppressor depending on the cell state, and its expression is associated with cellular stress, cell cycle arrest, or programmed cell death [64].

CEBPZ expression and methylation have been remarkably correlated with acute myeloid leukemia [65]. In a recent study, its overexpression was also found in squamous cell carcinoma of the esophagus [66].

GRK6 belongs to the family of G protein-coupled receptor kinases. It is overexpressed in immune cells and is closely associated with inflammation-related processes [67]. Up-regulation of GRK6 has been associated with colorectal cancer and is considered a potential biomarker for predicting poor survival in colorectal cancer patients [68]. In contrast, downregulation of GRK6 has been suggested as a potential biomarker for predicting overall survival in patients with lung adenocarcinoma [69]. In addition, up- or down-regulation of GRK6 expression has been observed in patients with hepatocellular carcinoma [70] and medulloblastoma [71], respectively.

HPGDS belongs to the family of transferases and catalyzes the production of a prostaglandin, prostaglandin D2, which is considered an essential lipid regulator that plays a remarkable role in the immune system, for example, in the inflammatory response [72]. According to a genome-wide association study, HPGDS was significantly associated with germ cell tumors in the testis [73]. Moreover, in a recent study, HPGDS was considered a prognostic biomarker for lung adenocarcinomas, and it was suggested that HPGDS may provide clues to the aggressiveness of the disease [74].

SKIL (also known as SnoN) is a transcriptional co-repressor that negatively regulates TGF- β signaling. SKIL has been associated with many cancers. For example, upregulation or amplification of SKIL has been associated with breast cancer [75], squamous cell carcinoma of the esophagus [76], prostate cancer, squamous cell carcinoma of the head and neck, and non-small cell lung cancer [77]. In addition, SKIL has been associated with leukemia [78], ovarian cancer [79], and squamous cell carcinoma of the lung [80].

SP3, which has a highly conserved DNA-binding domain, can either promote or repress the transcriptional activity of the corresponding target genes that play a role in the cell cycle, differentiation, or carcinogenesis [81]. High expression of SP3 was observed in hepatocellular carcinoma tissues compared with control tissues [82]. In another study, SP3 is described as a driving force for cancer metastasis in sarcomas [83].

In summary, we present here for the first time the molecular codes of GC at the different system levels (i.e., hub proteins, receptor TFs, and receptors) based on an integrative multi-omics approach and machine learning algorithms. The bioinformatics and machine learning approach determined previously identified biomolecules associated with GC as well as novel diagnostic and/or prognostic biomarker candidates such as AES, CEBPZ, GRK6, HPGDS, SKIL, and SP3. We believe that these results will provide insights into the underlying mechanisms of GC progression as well as some powerful novel biomarker candidates for GC. Despite the tremendous significance of the results of this study, further efforts are needed to experimentally and clinically validate the insights gained here.

Author Contributions: Conceptualization, M.K. and E.G.; Methodology, M.K. and E.G.; Formal Analysis, M.K.; Writing—Original Draft Preparation, M.K.; Writing—Review and Editing, M.K. and E.G.; Supervision, E.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The datasets analyzed during the current study are available in The Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>, accessed on 14 April 2022) and The Genome Cancer Atlas (<https://portal.gdc.cancer.gov/>) (accessed on 23 May 2022). Protein interactome data are available in Biological General Repository for Interaction Datasets (<https://thebiogrid.org>, accessed on 25 July 2022).

Acknowledgments: We thank Betul Comertpay who provided help in machine learning analyses.

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.J.C. 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. Lyons, K.; Le, L.C.; Pham, Y.T.H.; Borron, C.; Park, J.Y.; Tran, C.T.; Tran, T.V.; Tran, H.T.-T.; Vu, K.T.; Do, C.D.; et al. Gastric cancer: Epidemiology, biology, and prevention: A mini review. *Eur. J. Cancer Prev.* **2019**, *28*, 397–412. [[CrossRef](#)] [[PubMed](#)]
3. Ho, S.W.T.; Tan, P. Dissection of gastric cancer heterogeneity for precision oncology. *Cancer Sci.* **2019**, *110*, 3405–3414. [[CrossRef](#)]
4. Biagioni, A.; Skalamera, I.; Peri, S.; Schiavone, N.; Cianchi, F.; Giommoni, E.; Magnelli, L.; Papucci, L. Update on gastric cancer treatments and gene therapies. *Cancer Metastasis Rev.* **2019**, *38*, 537–548. [[CrossRef](#)]
5. Correa, R.; Alonso-Pupo, N.; Rodríguez, E.W.H. Multi-omics data integration approaches for precision oncology. *Mol. Omics.* **2022**, *18*, 469–479. [[CrossRef](#)] [[PubMed](#)]
6. Gulfidan, G.; Soyulu, M.; Demirel, D.; Erdonmez, H.B.C.; Beklen, H.; Sarica, P.O.; Arga, K.Y.; Turanli, B. Systems biomarkers for papillary thyroid cancer prognosis and treatment through multi-omics networks. *Arch. Biochem. Biophys.* **2022**, *715*, 109085. [[CrossRef](#)] [[PubMed](#)]
7. Kelesoglu, N.; Kori, M.; Turanli, B.; Arga, K.Y.; Yilmaz, B.K.; Duru, O.A. Acute Myeloid Leukemia: New Multiomics Molecular Signatures and Implications for Systems Medicine Diagnostics and Therapeutics Innovation. *OMICS J. Integr. Biol.* **2022**, *26*, 392–403. [[CrossRef](#)] [[PubMed](#)]
8. Kori, M.; Cig, D.; Arga, K.Y.; Kasavi, C. Multiomics Data Integration Identifies New Molecular Signatures for Abdominal Aortic Aneurysm and Aortic Occlusive Disease: Implications for Early Diagnosis, Prognosis, and Therapeutic Targets. *OMICS J. Integr. Biol.* **2022**, *26*, 290–304. [[CrossRef](#)]
9. Gov, E. Co-expressed functional module-related genes in ovarian cancer stem cells represent novel prognostic biomarkers in ovarian cancer. *Syst. Biol. Reprod. Med.* **2020**, *66*, 255–266. [[CrossRef](#)]
10. Comertpay, B.; Gov, E. Identification of key biomolecules in rheumatoid arthritis through the reconstruction of comprehensive disease-specific biological networks. *Autoimmunity* **2020**, *53*, 156–166. [[CrossRef](#)]
11. Rahman, M.R.; Islam, T.; Gov, E.; Turanli, B.; Gulfidan, G.; Shahjaman, M.; Banu, N.A.; Haque, M.; Arga, K.Y.; Moni, M.A. Identification of Prognostic Biomarker Signatures and Candidate Drugs in Colorectal Cancer: Insights from Systems. *Biol. Anal. Med.* **2019**, *55*, 20. [[CrossRef](#)] [[PubMed](#)]
12. Kori, M.; Gov, E.; Arga, K.Y. Molecular signatures of ovarian diseases: Insights from network medicine perspective. *Syst. Biol. Reprod. Med.* **2016**, *62*, 266–282. [[CrossRef](#)] [[PubMed](#)]
13. Liu, C.J.; Hu, F.F.; Xia, M.X.; Han, L.; Zhang, Q.; Guo, A.Y. GSCALite: A web server for gene set cancer analysis. *Bioinformatics* **2018**, *34*, 3771–3772. [[CrossRef](#)] [[PubMed](#)]
14. Sondka, Z.; Bamford, S.; Cole, C.G.; Ward, S.A.; Dunham, I.; Forbes, S.A. The COSMIC Cancer Gene Census: Describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **2018**, *18*, 696–705. [[CrossRef](#)] [[PubMed](#)]
15. Liu, X.; Wu, J.; Zhang, D.; Bing, Z.; Tian, J.; Ni, M.; Zhang, X.; Meng, Z.; Liu, S. Identification of potential key genes associated with the pathogenesis and prognosis of gastric cancer based on integrated bioinformatics analysis. *Front. Genet.* **2018**, *9*, 265. [[CrossRef](#)]
16. Hou, J.Y.; Wang, Y.G.; Ma, S.J.; Yang, B.Y.; Li, Q.P. Identification of a prognostic 5-gene expression signature for gastric cancer. *J. Cancer Res. Clin. Oncol.* **2017**, *143*, 619–629. [[CrossRef](#)]
17. Chen, X.; Yang, Y.; Liu, J.; Li, B.; Xu, Y.; Li, C.; Xu, Q.; Liu, G.; Chen, Y.; Ying, J.; et al. Ndr4 hypermethylation is a potential biomarker for diagnosis and prognosis of gastric cancer in chinese population. *Oncotarget* **2017**, *8*, 8105–8119. [[CrossRef](#)]
18. Demirtas, T.Y.; Rahman, M.R.; Yurtsever, M.C.; Gov, E. Forecasting Gastric Cancer Diagnosis, Prognosis, and Drug Repurposing with Novel Gene Expression Signatures. *OMICS J. Integr. Biol.* **2022**, *26*, 64–74. [[CrossRef](#)]
19. Wang, Q.; Wen, Y.G.; Li, D.P.; Xia, J.; Zhou, C.-Z.; Wang, D.; Yan, D.-W.; Tang, H.-M.; Peng, Z.H. Upregulated INHBA expression is associated with poor survival in gastric cancer. *Med. Oncol.* **2012**, *29*, 77–83. [[CrossRef](#)]
20. Jin, Y.; He, J.; Du, J.; Zhang, R.X.; Yao, H.B.; Shao, Q.S. Overexpression of HS6ST2 is associated with poor prognosis in patients with gastric cancer. *Oncol. Lett.* **2017**, *14*, 6191–6197. [[CrossRef](#)]
21. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [[CrossRef](#)] [[PubMed](#)]
22. Tomczak, K.; Czerwińska, P.; Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Contemp. Oncol.* **2015**, *2015*, 68–77. [[CrossRef](#)]
23. Kori, M.; Gov, E.; Arga, K.Y. Novel Genomic Biomarker Candidates for Cervical Cancer as Identified by Differential Co-Expression Network Analysis. *OMICS J. Integr. Biol.* **2019**, *23*, 261–273. [[CrossRef](#)] [[PubMed](#)]
24. Kori, M.; Arga, K.Y.; Mardinoglu, A.; Turanli, B. Repositioning of Anti-Inflammatory Drugs for the Treatment of Cervical Cancer Sub-Types. *Front. Pharmacol.* **2022**, *13*, 884548. [[CrossRef](#)] [[PubMed](#)]
25. Bolstad, B.M.; Irizarry, R.A.; Astrand, M.; Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. *Bioinformatics* **2003**, *19*, 185–193. [[CrossRef](#)]
26. Gautier, L.; Cope, L.; Bolstad, B.M.; Irizarry, R.A. Affy—Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **2004**, *20*, 307–315. [[CrossRef](#)] [[PubMed](#)]

27. Huber, W.; Carey, V.J.; Gentleman, R.; Anders, S.; Carlson, M.; Carvalho, B.S.; Bravo, H.C.; Davis, S.; Gatto, L.; Girke, T.; et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **2015**, *12*, 115–121. [[CrossRef](#)]
28. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)]
29. Kamburov, A.; Stelzl, U.; Lehrach, H.; Herwig, R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* **2013**, *41*, D793–D800. [[CrossRef](#)] [[PubMed](#)]
30. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [[CrossRef](#)] [[PubMed](#)]
31. Joshi-Tope, G.; Gillespie, M.; Vastrik, I.; Schmidt, E.; de Bono, B.; Jassal, B.; Gopinath, G.R.; Wu, G.R.; Matthews, L.; Lewis, S.; et al. Reactome: A knowledgebase of biological pathways. *Nucleic Acids Res.* **2005**, *33*, D428–D432. [[CrossRef](#)] [[PubMed](#)]
32. Oughtred, R.; Stark, C.; Breitkreutz, B.J.; Rust, J.; Boucher, L.; Chang, C.; Tyers, M. The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **2019**, *47*, D529–D541. [[CrossRef](#)] [[PubMed](#)]
33. Smoot, M.E.; Ono, K.; Ruscheinski, J.; Wang, P.L.; Ideker, T. Cytoscape 2.8: New Features for Data Integration and Network Visualization. *Bioinformatics* **2011**, *27*, 431–432. [[CrossRef](#)] [[PubMed](#)]
34. 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* (Suppl. S4), S11. [[CrossRef](#)] [[PubMed](#)]
35. Patil, K.R.; Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2685–2689. [[CrossRef](#)] [[PubMed](#)]
36. Kori, M.; Arga, K.Y. Potential biomarkers and therapeutic targets in cervical cancer: Insights from the meta-analysis of transcriptomics data within network biomedicine perspective. *PLoS ONE* **2018**, *18*, e0200717. [[CrossRef](#)]
37. Han, H.; Cho, J.W.; Lee, S.; Yun, A.; Kim, H.; Bae, D.; Yang, S.; Kim, C.Y.; Lee, M.; Kim, E.; et al. TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* **2018**, *46*, D380–D386. [[CrossRef](#)]
38. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)]
39. Mi, H.; Ebert, D.; Muruganujan, A.; Mills, C.; Albou, L.-P.; Mushayamaha, T.; Thomas, P.D. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **2021**, *49*, D394–D403. [[CrossRef](#)]
40. Tabas-Madrid, D.; Nogales-Cadenas, R.; Pascual-Montano, A. GeneCodis3: A non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Res.* **2012**, *40*, 478–483. [[CrossRef](#)] [[PubMed](#)]
41. Mandrekar, J.N. Receiver operating characteristic curve in diagnostic test assessment. *J. Thorac. Oncol.* **2010**, *5*, 1315–1316. [[CrossRef](#)] [[PubMed](#)]
42. Espe, S. Malacards: The Human Disease Database. *J. Med. Libr. Assoc. JMLA* **2018**, *106*, 140–141. [[CrossRef](#)]
43. Piñero, J.; Bravo, À.; Queralt-Rosinach, N.; Gutiérrez-Sacristán, A.; Deu-Pons, J.; Centeno, E.; García, J.G.; Sanz, F.; Furlong, L.I. DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **2017**, *45*, D833–D839. [[CrossRef](#)] [[PubMed](#)]
44. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; Wieggers, J.; Wieggers, T.C.; Mattingly, C.J. Comparative Toxicogenomics Database (CTD): Update 2021. *Nucleic Acids Res.* **2021**, *49*, D1138–D1143. [[CrossRef](#)] [[PubMed](#)]
45. Schapire, R. *Machine Learning Algorithms for Classification*; Princeton University: Princeton, NJ, USA, 2015; p. 10.
46. Kobayashi, H.; Komatsu, S.; Ichikawa, D.; Kawaguchi, T.; Hirajima, S.; Miyamae, M.; Okajima, W.; Ohashi, T.; Kosuga, T.; Konishi, H.; et al. Overexpression of denticleless E3 ubiquitin protein ligase homolog (DTL) is related to poor outcome in gastric carcinoma. *Oncotarget* **2015**, *6*, 36615–36624. [[CrossRef](#)]
47. Farhadi, J.; Mehrzad, J.; Mehrad-Majd, H.; Motavalizadehkakhky, A. Clinical significance of TRIM29 expression in patients with gastric cancer. *Gastroenterol. Hepatol. Bed Bench* **2022**, *15*, 131–138.
48. Wang, X.; Liu, Y.; Niu, Z.; Fu, R.; Jia, Y.; Zhang, L.; Shao, D.; Du, H.; Hu, Y.; Xing, X.; et al. Prognostic value of a 25-gene assay in patients with gastric cancer after curative resection. *Sci. Rep.* **2017**, *8*, 7515. [[CrossRef](#)] [[PubMed](#)]
49. Song, J.; Noh, J.H.; Lee, J.H.; Eun, J.W.; Ahn, Y.M.; Kim, S.Y.; Lee, S.H.; Park, W.S.; Yoo, N.J.; Lee, J.Y.; et al. Increased expression of histone deacetylase 2 is found in human gastric cancer. *APMIS* **2005**, *113*, 264–268. [[CrossRef](#)]
50. Ignatavicius, P.; Dauksa, A.; Zilinskas, J.; Kazokaite, M.; Riauka, R.; Barauskas, G. DNA Methylation of HOXA11 Gene as Prognostic Molecular Marker in Human Gastric Adenocarcinoma. *Diagnostics* **2022**, *12*, 1686. [[CrossRef](#)]
51. Gu, H.; Zhong, Y.; Liu, J.; Shen, Q.; Wei, R.; Zhu, H.; Zhang, X.; Xia, X.; Yao, M.; Ni, M. The Role of miR-4256/HOXC8 Signaling Axis in the Gastric Cancer Progression: Evidence From lncRNA-miRNA-mRNA Network Analysis. *Front. Oncol.* **2022**, *11*, 793678. [[CrossRef](#)]
52. McChesney, P.A.; Aiyar, S.E.; Lee, O.J.; Zaika, A.; Moskaluk, C.; Li, R.; El-Rifai, W. Cofactor of BRCA1: A novel transcription factor regulator in upper gastrointestinal adenocarcinomas. *Cancer Res.* **2006**, *66*, 1346–1353. [[CrossRef](#)] [[PubMed](#)]
53. Shi, S.; Zhang, Z.G. Role of Sp1 expression in gastric cancer: A meta-analysis and bioinformatics analysis. *Oncol. Lett.* **2019**, *18*, 4126–4135. [[CrossRef](#)] [[PubMed](#)]
54. Gong, B.; Li, Y.; Cheng, Z.; Wang, P.; Luo, L.; Huang, H.; Duan, S.; Liu, F. GRIK3: A novel oncogenic protein related to tumor TNM stage, lymph node metastasis, and poor prognosis of GC. *Tumor Biol.* **2017**, *39*, 1010428317704364. [[CrossRef](#)] [[PubMed](#)]

55. Verma, R.; Sharma, P.C. Next generation sequencing-based emerging trends in molecular biology of gastric cancer. *Am. J. Cancer Res.* **2018**, *8*, 207–225. [[PubMed](#)]
56. Dong, X.; Wang, G.; Zhang, G.; Ni, Z.; Suo, J.; Cui, J.; Cui, A.; Yang, Q.; Xu, Y.; Li, F. The endothelial lipase protein is promising urinary biomarker for diagnosis of gastric cancer. *Diagn. Pathol.* **2013**, *8*, 45. [[CrossRef](#)]
57. Laitinen, A.; Hagström, J.; Mustonen, H.; Kokkola, A.; Tervahartiala, T.; Sorsa, T.; Böckelman, C.; Haglund, C. Serum MMP-8 and TIMP-1 as prognostic biomarkers in gastric cancer. *Tumor Biol.* **2018**, *40*, 1010428318799266. [[CrossRef](#)]
58. Ying, W.; Zheng, K.; Wu, Y.; Wang, O. Pannexin 1 Mediates Gastric Cancer Cell Epithelial-Mesenchymal Transition via Aquaporin 5. *Biol. Pharm. Bull.* **2021**, *44*, 1111–1119. [[CrossRef](#)]
59. Morgan, E.; Arnold, M.; Camargo, M.C.; Gini, A.; Kunzmann, A.T.; Matsuda, T.; Meheus, F.; Verhoeven, R.H.A.; Vignat, J.; Laversanne, M.; et al. The current and future incidence and mortality of gastric cancer in 185 countries, 2020–2040: A population-based modelling study. *Eclinical Med.* **2022**, *47*, 101404. [[CrossRef](#)]
60. Zhou, Z.H.; Ji, C.D.; Xiao, H.L.; Zhao, H.B.; Cui, Y.H.; Bian, X.W. Reorganized Collagen in the Tumor Microenvironment of Gastric Cancer and Its Association with Prognosis. *J. Cancer* **2017**, *8*, 1466–1476. [[CrossRef](#)]
61. Beagle, B.; Johnson, G.V. AES/GRG5: More than just a dominant-negative TLE/GRG family member. *Dev. Dyn.* **2010**, *239*, 2795–2805. [[CrossRef](#)] [[PubMed](#)]
62. Okada, Y.; Sonoshita, M.; Kakizaki, F.; Aoyama, N.; Itatani, Y.; Uegaki, M.; Sakamoto, H.; Kobayashi, T.; Inoue, T.; Kamba, T.; et al. Amino-terminal enhancer of split gene AES encodes a tumor and metastasis suppressor of prostate cancer. *Cancer Sci.* **2017**, *108*, 744–752. [[CrossRef](#)]
63. Kakizaki, F.; Sonoshita, M.; Miyoshi, H.; Itatani, Y.; Ito, S.; Kawada, K.; Taketo, M.M. Expression of metastasis suppressor gene AES driven by a Yin Yang (YY) element in a CpG island promoter and transcription factor YY2. *Cancer Sci.* **2016**, *107*, 1622–1631. [[CrossRef](#)] [[PubMed](#)]
64. Ramji, D.P.; Foka, P. CCAAT/enhancer-binding proteins: Structure, function and regulation. *Biochem. J.* **2002**, *365*, 561–575. [[CrossRef](#)] [[PubMed](#)]
65. Yao, D.M.; Qian, J.; Lin, J.; Wang, Y.-L.; Chen, Q.; Qian, Z.; Li, Y.; Wang, C.-Z.; Yang, J. Aberrant methylation of CCAAT/enhancer binding protein zeta promoter in acute myeloid leukemia. *Leuk. Res.* **2011**, *35*, 957–960. [[CrossRef](#)] [[PubMed](#)]
66. Huang, Y.; Lin, L.; Shen, Z.; Li, Y.; Cao, H.; Peng, L.; Qiu, Y.; Cheng, X.; Meng, M.; Lu, D.; et al. CEBPG promotes esophageal squamous cell carcinoma progression by enhancing PI3K-AKT signaling. *Am. J. Cancer Res.* **2020**, *10*, 3328–3344. [[PubMed](#)]
67. Stegen, M.; Engler, A.; Ochsenfarth, C.; Manthey, I.; Peters, J.; Siffert, W.; Frey, U.H. Characterization of the G protein-coupled receptor kinase 6 promoter reveals a functional CREB binding site. *PLoS ONE* **2021**, *16*, e0247087. [[CrossRef](#)]
68. Tao, R.; Li, Q.; Gao, X.; Ma, L. Overexpression of GRK6 associates with the progression and prognosis of colorectal carcinoma. *Oncol. Lett.* **2018**, *15*, 5879–5886. [[CrossRef](#)]
69. Yao, S.; Zhong, L.; Liu, J.; Feng, J.; Bian, T.; Zhang, Q.; Chen, J.; Lv, X.; Chen, J.; Liu, Y. Prognostic value of decreased GRK6 expression in lung adenocarcinoma. *J. Cancer Res. Clin. Oncol.* **2016**, *142*, 2541–2549. [[CrossRef](#)]
70. Li, Y.P. GRK6 expression in patients with hepatocellular carcinoma. *Asian Pac. J. Trop. Med.* **2013**, *6*, 220–223. [[CrossRef](#)]
71. Yuan, L.; Zhang, H.; Liu, J.; Rubin, J.B.; Cho, Y.J.; Shu, H.K.; Schniederjan, M.; MacDonald, T.J. Growth factor receptor-Src-mediated suppression of GRK6 dysregulates CXCR4 signaling and promotes medulloblastoma migration. *Mol. Cancer* **2013**, *12*, 18. [[CrossRef](#)]
72. Seo, M.J.; Oh, D.K. Prostaglandin synthases: Molecular characterization and involvement in prostaglandin biosynthesis. *Prog. Lipid Res.* **2017**, *66*, 50–68. [[CrossRef](#)]
73. Chung, C.C.; Kanetsky, P.A.; Wang, Z.; Hildebrandt, M.A.; Koster, R.; Skotheim, R.I.; Kratz, C.P.; Turnbull, C.; Cortessis, V.K.; Bakken, A.C.; et al. Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat. Genet.* **2013**, *45*, 680–685. [[CrossRef](#)] [[PubMed](#)]
74. Mao, H.; Luo, T.; Li, Q.; Xu, L.; Xie, Y. HPGDS is a novel prognostic marker associated with lipid metabolism and aggressiveness in lung adenocarcinoma. *Front. Oncol.* **2022**, *12*, 5788.
75. Deheuninck, J.; Luo, K. Ski and SnoN potent negative regulators of TGF-beta signaling. *Cell Res.* **2009**, *19*, 47–57. [[CrossRef](#)]
76. Akagi, I.; Miyashita, M.; Makino, H.; Nomura, T.; Hagiwara, N.; Takahashi, K.; Tajiri, T. SnoN overexpression is predictive of poor survival in patients with esophageal squamous cell carcinoma. *Ann. Surg. Oncol.* **2008**, *15*, 2965–2975. [[CrossRef](#)]
77. Hagerstrand, D.; Tong, A.; Schumacher, S.E.; Ilic, N.; Shen, R.R.; Cheung, H.W.; Hahn, W.C. Systematic Interrogation of 3q26 Identifies TLOC1 and SKIL as Cancer Drivers. *Cancer Discov.* **2013**, *3*, 1044–1057. [[CrossRef](#)]
78. Raffoul, F.; Campla, C.; Nanjundan, M. SnoN/SkiL, a TGFbeta signaling mediator: A participant in autophagy induced by arsenic trioxide. *Autophagy* **2010**, *6*, 955–957. [[CrossRef](#)]
79. Smith, D.M.; Patel, S.; Raffoul, F.; Haller, E.; Mills, G.B.; Nanjundan, M. Arsenic trioxide induces a beclin-1-independent autophagic pathway via modulation of SnoN/SkiL expression in ovarian carcinoma cells. *Cell Death Differ.* **2010**, *17*, 1867–1881. [[CrossRef](#)]
80. Lazarus, K.A.; Hadi, F.; Zambon, E.; Bach, K.; Santolla, M.F.; Watson, J.K.; Khaled, W.T. BCL11A interacts with SOX2 to control the expression of epigenetic regulators in lung squamous carcinoma. *Nat. Commun.* **2018**, *9*, 3327. [[CrossRef](#)]
81. Li, L.; Davie, J.R. The role of Sp1 and Sp3 in normal and cancer cell biology. *Ann. Anat.* **2010**, *192*, 275–283. [[CrossRef](#)]

82. Huang, Z.; Huang, L.; Shen, S.; Li, J.; Lu, H.; Mo, W.; Feng, Z. Sp1 cooperates with Sp3 to upregulate MALAT1 expression in human hepatocellular carcinoma. *Oncol. Rep.* **2015**, *34*, 2403–2412. [[CrossRef](#)] [[PubMed](#)]
83. Kajita, Y.; Kato, T., Jr.; Tamaki, S.; Furu, M.; Takahashi, R.; Nagayama, S.; Toguchida, J. The transcription factor Sp3 regulates the expression of a metastasis-related marker of sarcoma, actin filament-associated protein 1-like 1 (AFAP1L1). *PLoS ONE* **2013**, *8*, e49709. [[CrossRef](#)] [[PubMed](#)]