Research Paper

Identification and validation of key genes in gastric cancer: insights from *in silico* analysis, clinical samples, and functional assays

Xiaofeng Pei[1,*], Yuanling Luo[1,*], Huanwen Zeng[1,*], Muhammad Jamil[2], Xiaodong Liu[3], Bo Jiang[4]

[1]Department of Oncology, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai 519000, China
[2]PARC Arid Zone Research Center, Dera Ismail Khan 29050, Pakistan
[3]Department of Pharmacy, The 922 Hospital of Joint Logistics Support Force, PLA, Hengyang 421002, China
[4]Department of Emergency, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai 519000, China
[*]Equal contribution

Correspondence to: Muhammad Jamil, Xiaodong Liu, Bo Jiang; **email:** jamilmatrah@parc.gov.pk; 3178075938@qq.com, https://orcid.org/0009-0003-2527-2467; 13631220892@163.com, https://orcid.org/0009-0001-3698-3913

ABSTRACT

Introduction: The underlying mechanisms of gastric cancer (GC) remain unknown. Therefore, in this study, we employed a comprehensive approach, combining computational and experimental methods, to identify potential key genes and unveil the underlying pathogenesis and prognosis of GC.
Methods: Gene expression profiles from GEO databases (GSE118916, GSE79973, and GSE29272) were analyzed to identify DEGs between GC and normal tissues. A PPI network was constructed using STRING and Cytoscape, followed by hub gene identification with CytoHubba. Investigations included expression and promoter methylation analysis, survival modeling, mutational and miRNA analysis, gene enrichment, drug prediction, and *in vitro* assays for cellular behaviors.
Results: A total of 83 DEGs were identified in the three datasets, comprising 41 up-regulated genes and 42 down-regulated genes. Utilizing the degree and MCC methods, we identified four hub genes that were hypomethylated and up-regulated: COL1A1, COL1A2, COL3A1, and FN1. Subsequent validation of their expression and promoter methylation on clinical GC samples through targeted bisulfite sequencing and RT-qPCR analysis further confirmed the hypomethylation and overexpression of these genes in local GC patients. Furthermore, it was observed that these hub genes regulate tumor proliferation and metastasis in *in vivo* and exhibited mutations in GC patients.
Conclusion: We found four potential diagnostic and prognostic biomarkers, including COL1A1, COL1A2, COL3A1, and FN1 that may be involved in the occurrence and progression of GC.

INTRODUCTION

Gastric cancer (GC) stands as the sixth most frequently diagnosed cancer globally, with the second-highest mortality rate among malignant tumors [1]. While the 5-year overall survival rate for early-stage GC patients can reach 95% [2], it remains around 50% for those in the advanced stage, even with comprehensive treatment approaches involving surgery [2, 3]. The low survival rate of GC is primarily attributed to tumor recurrence and metastasis [4]. As a result, it becomes crucial to delve into the potential molecular mechanisms that drive the malignant biological behavior of GC cells. Moreover, the discovery of efficient early diagnostic techniques and dependable molecular markers for recurrence monitoring and prognosis evaluation holds significant importance.

Despite notable progress in comprehending the molecular intricacies of GC and the emergence of targeted therapeutic options, the effectiveness of existing targeted therapies remains limited for certain patients [5, 6]. Therefore, further research aimed at uncovering novel and more effective targeted approaches is essential to improve patient outcomes and overcome these challenges.

In recent years, the utilization of microarray and RNA-sequencing technology has emerged as a powerful and efficient tool in the quest for promising biomarkers to aid in cancer diagnosis, treatment, and prognosis [7, 8]. These technologies have led to the accumulation of a vast amount of data, accessible through public database platforms like Gene Expression Omnibus (GEO) [9] and The Cancer Genome Atlas (TCGA) [10, 11]. By leveraging the wealth of information in GEO and TCGA, scientists can uncover novel molecular signatures and candidate biomarkers that may have diagnostic, prognostic, or therapeutic implications in the fight against cancer [12]. The integration of GEO data with other experimental approaches enables a deeper understanding of cancer biology and aids in the early detection and treatment of cancer. By performing experimental validation, researchers can verify the expression patterns, molecular interactions, and functional roles of the identified biomarkers, ultimately strengthening the confidence in their potential clinical utility.

Various investigations have been undertaken to analyze the abnormal gene expression patterns associated with GC. Despite advanced research, these studies yielded inconsistent results [13–16]. Therefore, to address the challenges posed by diverse technological platforms and small sample sizes, integrated bioinformatics approaches have been embraced in cancer studies, yielding a wealth of valuable biological insights. In pursuit of gaining profound insights into the influence of Differentially Expressed Genes (DEGs) on the molecular pathogenesis of GC, our study aimed to explore novel signature genes associated with GC. For this purpose, we adopted a multi-level validation approach to rigorously examine and confirm the relevance and significance of these signature genes in the context of GC.

MATERIALS AND METHODS

Methodology

The overall methodology employed in this study is depicted in Figure 1.

Collection of clinical specimens

We acquired paired fresh cancer tissue specimens along with control samples from 39 patients who underwent surgical resection of GC at the District Headquarter
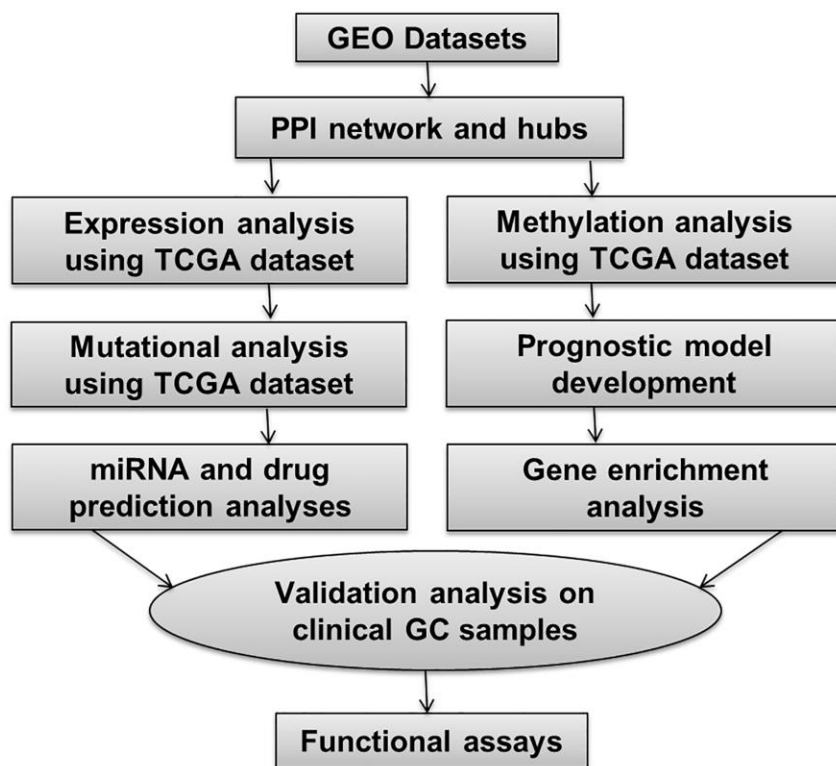

Figure 1. This figure illustrates the overall methodology utilized in the present study.

Hospital (DHQ), Teaching Hospital, Dera Ismail Khan, Khyber Pakhtunkhwa (KPK) between 2019 and 2023. None of the patients had received any chemotherapy or radiation therapy prior to the surgery. The collected tissue samples were promptly frozen in liquid nitrogen and stored at −80°C until DNA and RNA isolation. The study received ethical approval in accordance with the Helsinki Declaration, and informed written consent was obtained from all participants.

Microarray data acquisition, DEGs, and hub genes identification

We retrieved three datasets of microarray, namely GSE118916, GSE79973, and GSE29272, from the GEO database at http://www.ncbi.nlm.nih.gov/geo/. The selection criteria for appropriate GC datasets were as follows: studies involving pharmacological manipulation, interfering molecules like miRNAs, siRNAs, or gene therapies, knockdown cultures, or artificially induced mutations were excluded. Only studies with a minimum of ten control and experimental samples, exclusively conducted in Homo sapiens, and providing clear descriptions of protocols or samples were chosen. Additionally, datasets with raw data availability, excluding those with treated data only, and studies conducted on platforms belonging to Affymetrix, Illumina, or Agilent manufacturers were included. Samples from metastasized tissues were also excluded. A total of 16 microarray datasets were reviewed, and GSE118916, GSE79973, and GSE29272 were selected based on sample size adequacy for further analysis.

For our study, we specifically selected paired GC tissues and their corresponding adjacent tissues. In cases where multiple probes were associated with a particular gene, we calculated the average expression level to represent its final expression. The initial microarray data from each series underwent processing using the R software package (version 3.6.1; http://www.R-project.org/). Following the transformation to a log2 scale, we set the cutoff criteria for identifying DEGs as |Log2 fold change (FC)| > 1 and adjusted $P < 0.01$. To visualize the common DEGs across the three datasets, we generated a Venn diagram utilizing Venny (version 2.1; https://bioinfogp.cnb.csic.es/tools/venny/index.html). Subsequently, we selected these overlapping DEGs for further investigation of hub genes.

In order to identify hub genes from the overlapping DEGs, we first constructed a protein-protein interaction (PPI) network using the STRING database (https://string-db.org/) [17]. During PPI construction all the active interactions sources were

utilized including textmining, experiments, database, co-expression, neighborhood, genefusion, and co-occurrence with 1st shell. The minimum required interaction score was set to 0.4 (medium confidence) and the protein nodes having no interaction with other proteins were removed from the network. This allowed us to explore the interactions and associations between the identified genes. Next, we employed the CytoHubba application within the Cytoscape software to analyze the PPI network and pinpoint the hub genes based on the degree method [18].

TCGA-datasets-based expression validation analysis

In this study, we harnessed the power of three crucial databases, including UALCAN (https://ualcan.path.uab.edu/cgi-bin/ualcan-res.pl) [19], OncoDB (https://oncodb.org/) [20], and GEPIA (http://gepia.cancer-pku.cn/) [21] to validate the expression of hub genes on the GC TCGA expression datasets. UALCAN offers a user-friendly interface, providing researchers with valuable insights into gene expression patterns and conducting in-depth analyses using TCGA data. OncoDB serves as a comprehensive repository of oncogenomic information, shedding light on cancer-related gene expression and molecular alterations. GEPIA, on the other hand, is a web-based tool enabling researchers to explore gene expression data and conduct interactive analyses across diverse cancer types.

Proteomic expression analysis

In this study, the UALCAN database was used to conduct proteomic expression analysis of the hub genes across GC and normal tissues. By utilizing this database, we gained essential information on the protein expression patterns of these hub genes, enhancing our understanding of their potential roles in GC and normal tissue biology.

Promoter methylation analysis

In our research, we utilized the promoter methylation features available in the UALCAN [19] and OncoDB [20] databases. These databases provide essential information on the epigenetic modifications of genes, particularly focusing on promoter methylation, which can regulate gene expression. Leveraging the data from these resources, we conducted a comprehensive promoter methylation analysis of the hub genes identified across GC and normal tissue samples. By examining the methylation patterns of these hub genes, we aimed to gain valuable insights into potential regulatory mechanisms that could influence gene expression in the context of GC.

Mutational and co-expressed gene analyses

In the current study, we utilized the mutational analysis feature of the cBioPortal database (https://www.cbioportal.org/) [22], a powerful resource for exploring genomic alterations in various cancers. cBioPortal provides comprehensive genomic and clinical data from numerous cancer studies, enabling researchers to assess the mutation landscape of specific genes of interest. We conducted an extensive mutational analysis of the hub genes across GC samples using cBioPortal with default settings. By examining the mutational status of these hub genes, we aimed to uncover potential genetic alterations that could influence their functions and contribute to the development and progression of GC. In addition to this, we also used this database with default settings to identify mutually co-expressed genes with hub genes in GC.

Survival analysis and the construction of a prognostic model

In this study, we employed two important methodologies to explore the prognostic implications of the hub genes. Firstly, we utilized the GEPIA [21], a powerful online tool for conducting survival analysis, to assess the association between the expression levels of the hub genes and patient outcomes. GEPIA allowed us to investigate the impact of these genes on overall survival and disease-free survival in specific cancer types. Secondly, to construct a robust prognostic model, we implemented the Cox regression method [23] via R. This approach enabled us to develop a predictive model that could accurately stratify patients based on their risk of adverse clinical outcomes.

Enrichment and miRNA prediction analyses

In our study, we utilized three important databases, namely DAVID, [24] miRDB, and ENCORI [25], to gain further insights into the functional implications of the hub genes identified. DAVID (Database for Annotation, Visualization, and Integrated Discovery) offers a comprehensive platform for gene functional annotation and enrichment analysis. We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the hub genes using the DAVID database. We utilized miRDB and ENCORI database to conduct miRNA prediction analysis of the hub genes.

Genomic DNA and RNA isolation

Total cell DNA from tissue samples was extracted using an organic method [26], while total RNA was extracted using TRIZol method [27]. We employed the NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) to assess the concentration and purity of the extracted DNA and RNA, ensuring that the A260/A280 ratio fell within the range of 1.8 to 2.0.

Library preparation for targeted bisulfite sequencing analysis

In brief, total DNA (1 µg) was fragmented into approximately 200–300 bp fragments using a Covarias sonication system (Covarias, Woburn, MA, USA). Following purification, the DNA fragments underwent repair and phosphorylation of blunt ends using a mixture of T4 DNA polymerase, Klenow Fragment, and T4 polynucleotide kinase. The repaired fragments were then 3′ adenylated using Klenow Fragment (3′–5′ exo-) and ligated with adapters containing 5′-methylcytosine instead of 5′-cytosine and index sequences using T4 DNA Ligase. The constructed libraries were quantified using a Qubit fluorometer with the Quant-iT dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA) and sent to Beijing Genomic Institute (BGI), China for targeted bisulfite sequencing. Following sequencing, the methylation data was normalized into beta values.

Real time quantitative PCR (RT-qPCR)

The RNA extracted was transcribed into cDNA using the Prime-Script RT reagent kit (TaKaRa, Dalian, China). RT-qPCR analysis was performed on an ABI 7500 Real-Time PCR System (Applied Biosystems, USA) using the SYBR Premix Ex Taq™ II kit (TaKaRa). The expression levels were normalized to β-actin. All experiments were independently conducted in triplicate. The $2^{(-\Delta\Delta Ct)}$ method was employed to assess the relative expression of each hub gene [28]. This method quantifies gene expression changes by comparing the cycle threshold (Ct) values of a target gene between control and experimental groups. It normalizes to reference genes, yielding a fold change value ($2^{(-\Delta\Delta Ct)}$), indicating whether the gene is up-regulated (>1), down-regulated (<1), or unchanged (=1).

Receiver operating characteristic (ROC) curve generation

Based on the RT-qPCR and targeted bisulfite-seq expression and methylation data, ROC curves of identified DEGs expression were generated using the SRPLOT web source (https://bioinformatics.com.cn/srplot).

Drug prediction analysis

In our study, we harnessed the drug prediction feature of the DrugBank (http://www.drugbank.ca) database [29], a comprehensive resource that provides valuable

information on drug-target interactions and drug-related data. Leveraging this feature, we aimed to identify potential drugs that could target the hub genes identified in our study. By exploring the vast database of drug-target interactions, we sought to uncover drugs that may have regulatory effects on the expression of the hub genes.

Cell culture and transfection

AGS cell line (obtained from the American Type Culture Collection, ATCC, Manassas, VA, USA) were maintained in a 37°C incubator with 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) from Hyclone (Logan, UT, USA) supplemented with 10% fetal bovine serum (FBS) obtained from Gibco (Waltham, MA, USA). Following this, we carried out gene knockdown experiments targeting COL1A1, COL1A2, COL3A1, and FN1 genes. These knockdowns were achieved by transfecting the cell with two siRNA constructs specific to each gene-namely, si-COL1A1, si-COL1A2, si-COL3A1, and si-FN1. The transfection was facilitated using Lipofectamine 3000 from Invitrogen (Waltham, MA, USA). The cells were subsequently cultured for an additional 48 hours following transfection.

Following are the sequences of the utilized siRNAs:

si-COL1A1-1: 5′-TTGGTGTTGTGCGATGACGTG-3′; si-COL1A1-2: 5′-GTACGTCCGGTTGTATGTA-3′; si-COL1A2-1: 5′-GGACCCGTTGGCAAAGATG-3′; si-COL1A2-2: 5′-CACCAGGAGGACCAGGAG-3′; si-COL1A3-1: 5′-CUAUGCGGAUAGAGAUGUCTT-3′; si-COL1A3-2: 5′-GAGGAAACAGAGGTGAAAGA GG-3′; si-FN1-1 sense: 5′-CCAUUUCACCUU CAGACAATT-3′; si-FN1-1 anti-sense: 5′-UUGU CUGGGUGAAAUGGTT-3′; si-FN2-1 sense: 5′-GCAAGCAGCAACAAUUUTT-3′; si-FN2-1 anti-sense: 5′-AAAUUGGCUUGCUGAUUGCTT-3′.

RNA extraction and RT-qPCR

Total RNA from the cell lines was extracted using TRIZol method [30] and RT-qPCR analysis was performed according to the instructions as discussed above.

Cell counting kit-8 (CCK-8) assays

After the transfection process, AGS cells were plated in 96-well plates at a concentration of 1×10^5 cells/mL and allowed to proliferate for 48 hours. To assess cell viability, we employed a CCK-8 kit (provided by Meilunbio, China), following the manufacturer's instructions. Absorbance measurements at 450 nm were taken using a Bio-Rad model 550 microplate reader.

Colony-forming assays

The cells were distributed into 6-well plates, with each well receiving 500 cells, and were then cultured for 48 hours. Subsequently, the cells were exposed to the correct doses of ATO (2 μM for AGS cells). Following one-week incubation, the cells were immobilized using 4% paraformaldehyde sourced from Thermo Fisher Scientific (Waltham, MA, USA). Afterward, they were subjected to staining with 2% crystal violet from Thermo Fisher Scientific (USA). Colonies that were clearly visible and consisted of at least 50 cells per clone were enumerated under a microscope.

Statistics analysis

DEGs were identified using a *t*-test [31]. While for GO and KEGG enrichment analysis, we used Fisher's Exact test for computing difference [32]. Correlational analyses were carried out using the Pearson method. For comparisons, a student *t*-test was adopted in the current study. All the analyses were carried out in R version 3.6.3 software.

Availability of data and materials

The data supporting the findings of the article are available in the GEO database at https://www.ncbi.nlm.nih.gov/geo/.

RESULTS

Microarray data acquisition, DEGs, and hub genes identification

We obtained three gene expression profiles (GSE118916, GSE79973, and GSE29272) from the GEO database, comprising a total of 318 samples, including 159 GC and 159 matched adjacent control tissues. From GSE118916, a total of 1295 DEGs were identified, consisting of 651 up-regulated and 644 down-regulated genes. In GSE79973, a total of 376 DEGs were screened, including 132 up-regulated and 244 down-regulated genes. Similarly, GSE29272 yielded 330 DEGs, comprising a total of 165 up-regulated and 165 down-regulated genes. Volcano plots depicting the DEGs in each dataset were illustrated in Figure 2A–2C. Among these datasets, 83 genes (41 up-regulated and 43 down-regulated) were common, and were chosen for further analysis (Figure 2D).

To investigate the interactions between the 83 DEGs, we utilized the STRING database to construct a PPI network. The resulting PPI network was generated using Cytoscape (Figure 3A) and comprised 83 nodes with 253 interactions (Figure 3A). Subsequently, the

PPI network was analyzed using the cytoHubba application in Cytoscape to identify hub genes (Figure 3B). This analysis involved two algorithms, degree and MCC, provided by cytoHubba. Based on these two algorithms, up-regulated COL1A1 (Collagen, type I, alpha 1), COL1A2 (Collagen, type I, alpha 2), COL3A1 (Collagen, type III, alpha 1), and FN1 (Fibronectin) in GC samples were identified as the hub genes (Figure 3C). These genes exhibited significant connectivity within the network, suggesting their potential importance in the regulatory network of the DEGs.
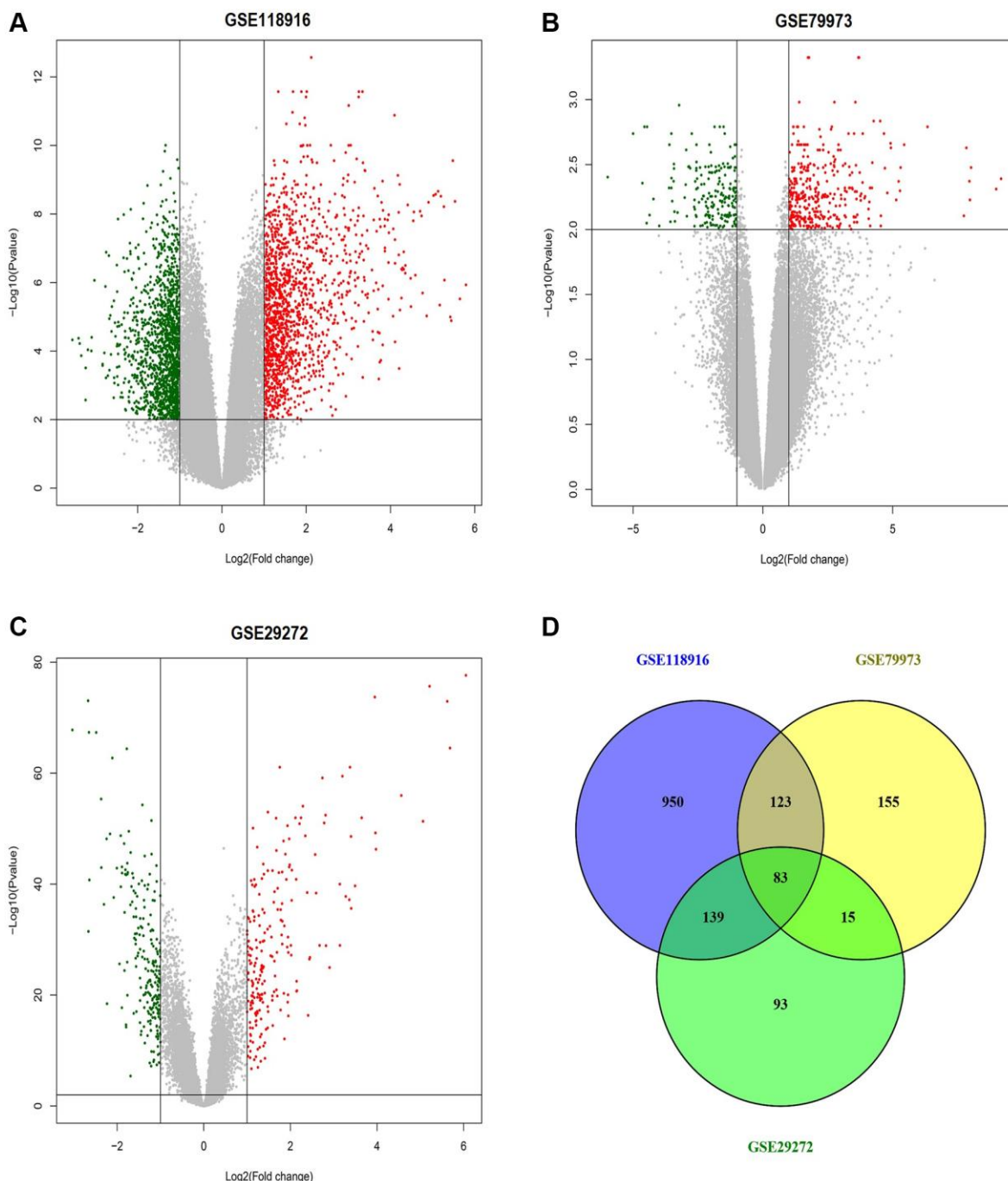
Figure 2. This figure depicts the process of identifying differentially expressed genes (DEGs) across the GSE118916, GSE79973, and GSE29272 datasets related to gastric cancer (GC). (**A**) Volcano plot of differentially expressed genes (DEGs) in the GSE118916 dataset. (**B**) Volcano plot of DEGs in the GSE79973 dataset. (**C**) Volcano plot of DEGs in the GSE29272 dataset. (**D**) Venn diagram showing the overlap of DEGs among the three datasets (GSE118916, GSE79973, and GSE29272). Red dots represent up-regulated genes, and green dots represent down-regulated genes. The numbers in Venn diagram represent the count of unique and overlapping genes among the datasets. *P*-value < 0.05.

Expression validation based on TCGA datasets

To confirm the mRNA expression levels of COL1A1, COL1A2, COL3A1, and FN1 in GC samples compared to controls from the TCGA database, we utilized UALCAN, OncoDB, and GEPIA for data integration and visualization. These hub genes (COL1A1, COL1A2, COL3A1, and FN1) exhibited significant overexpression ($p < 0.05$) in GC samples relative to controls (Figure 4A–4C), which was consistent with the findings from the GEO datasets. Furthermore, the expression of COL1A1, COL1A2, COL3A1, and FN1 varied across different stages of GC (Figure 4D). These results provide additional evidence supporting the up-regulation of these hub genes in GC.

Proteomic expression analysis of COL1A1, COL1A2, COL3A1, and FN1

In our study, we conducted proteomic expression analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC samples compared to controls using the UALCAN database. The results revealed that the protein levels of these genes were significantly higher in GC samples compared to controls (Figure 5). The findings were consistent with the mRNA expression data, further

validating the up-regulation of COL1A1, COL1A2, COL3A1, and FN1 in GC relative to control samples.

Promoter methylation levels of COL1A1, COL1A2, COL3A1, and FN1

In our study, we conducted promoter methylation analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC samples compared to controls using the UALCAN and OncoDB databases. The results revealed that these genes exhibited hypomethylation in GC samples relative to controls (Figure 6). This finding suggests that the promoter regions of COL1A1, COL1A2, COL3A1, and FN1 undergo reduced methylation levels in GC, which could potentially contribute to their up-regulation of these genes in GC (Figure 6).

Mutational and co-express gene analysis of COL1A1, COL1A2, COL3A1, and FN1

We conducted a mutational analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC samples using the cBioPortal database. The results revealed that these genes were mutated in approximately 15.79% of the analyzed GC samples (Figure 7A). Among the mutated GC samples, missense mutations were particularly
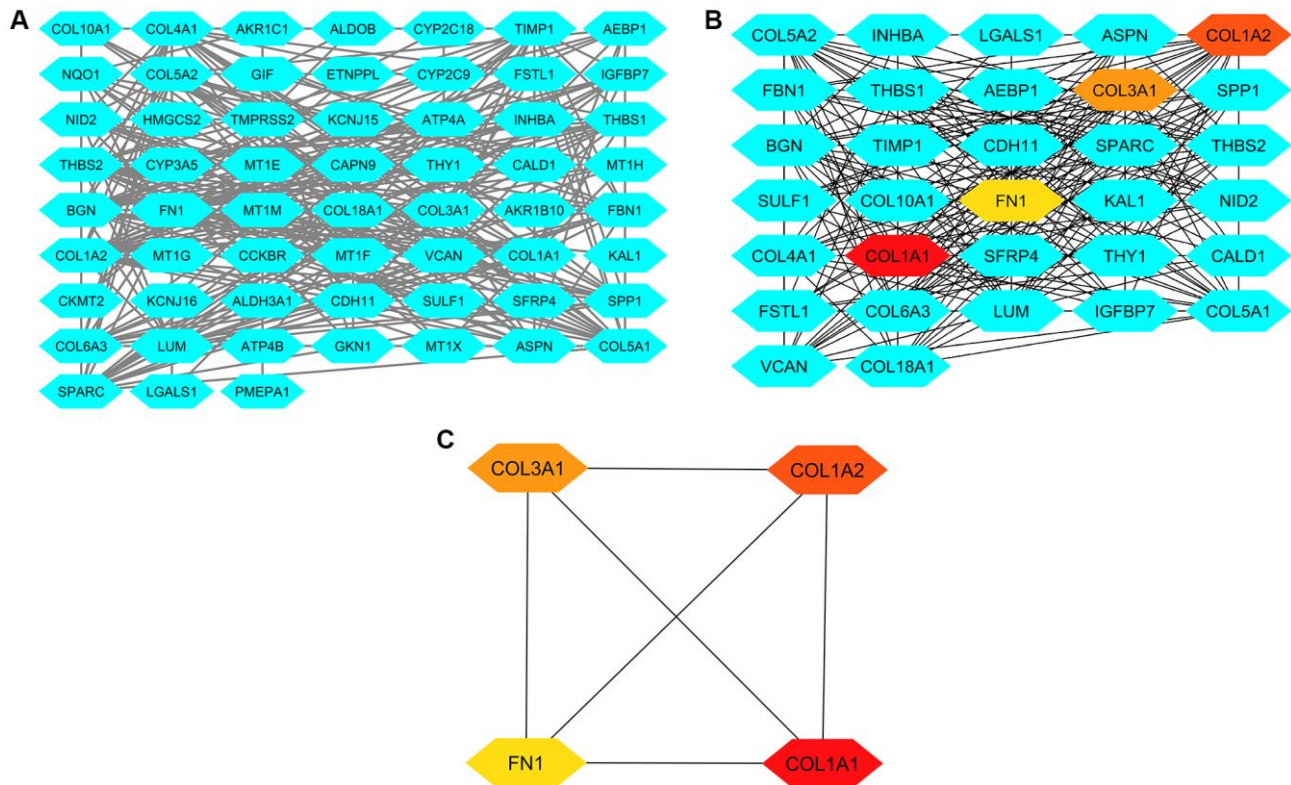
Figure 3. This figure illustrates the process of constructing protein-protein interaction (PPI) networks, analyzing them, and identifying hub genes. (**A**) Panel A presents the PPI network formed by the 83 common DEGs from GSE118916, GSE79973, and GSE29272. (**B**) Panel B displays the PPI network of these DEGs highlighting hub genes identified through degree and MCC methods. (**C**) Panel C showcases a refined PPI network focusing solely on the four identified hub genes.

prevalent, with C>T substitution mutation being the most common type of mutation observed (Figure 7B). These findings suggest that these hub genes undergo genetic alterations in a subset of GC cases, and the prevalence of missense mutations, particularly C>T substitutions, underscores their potential significance in the molecular landscape of GC.

Additionally, in our study, we observed that COL1A1, COL1A2, COL3A1, and FN1, being the hub genes, exhibited co-expression patterns with other genes in GC samples. Notably, the co-expression analysis revealed that MTM and COL5A1 were among the genes showing significant co-expression with the hub genes (Figure 7C). This finding suggests that MTM and COL5A1 might be functionally linked to the hub genes and potentially involved in shared biological processes or pathways related to GC development and progression.

Survival analysis and the construction of a prognostic model

We conducted survival analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC patients using the GEPIA

database. The results revealed a significant association between higher expression levels of these genes and worse overall survival (OS) of the GC patients. This finding suggests that increased expression of COL1A1, COL1A2, COL3A1, and FN1 may be indicative of poorer prognosis in GC (Figure 8A).

To develop the prognostic model based on COL1A1, COL1A2, COL3A1, and FN1 genes, we utilized the TCGA_STAD dataset as the training dataset, and the GSE84437, GSE84433, GSE84426, GSE28541, GSE26901, GSE26899, GSE26253, GSE183136, and GSE13861 datasets served as validation datasets. The construction of our prognostic model involved a stepwise Cox regression approach, which integrated hazard ratio, c-index, and risk score parameters. By evaluating the predictive performance of our prognostic model using the c-index, we confirmed its effectiveness and robustness in assessing the prognosis of patients with GC (Figure 8B, 8C). The incorporation of multiple datasets for validation strengthens the reliability of our prognostic model and supports its potential clinical utility in predicting patient outcomes in GC.
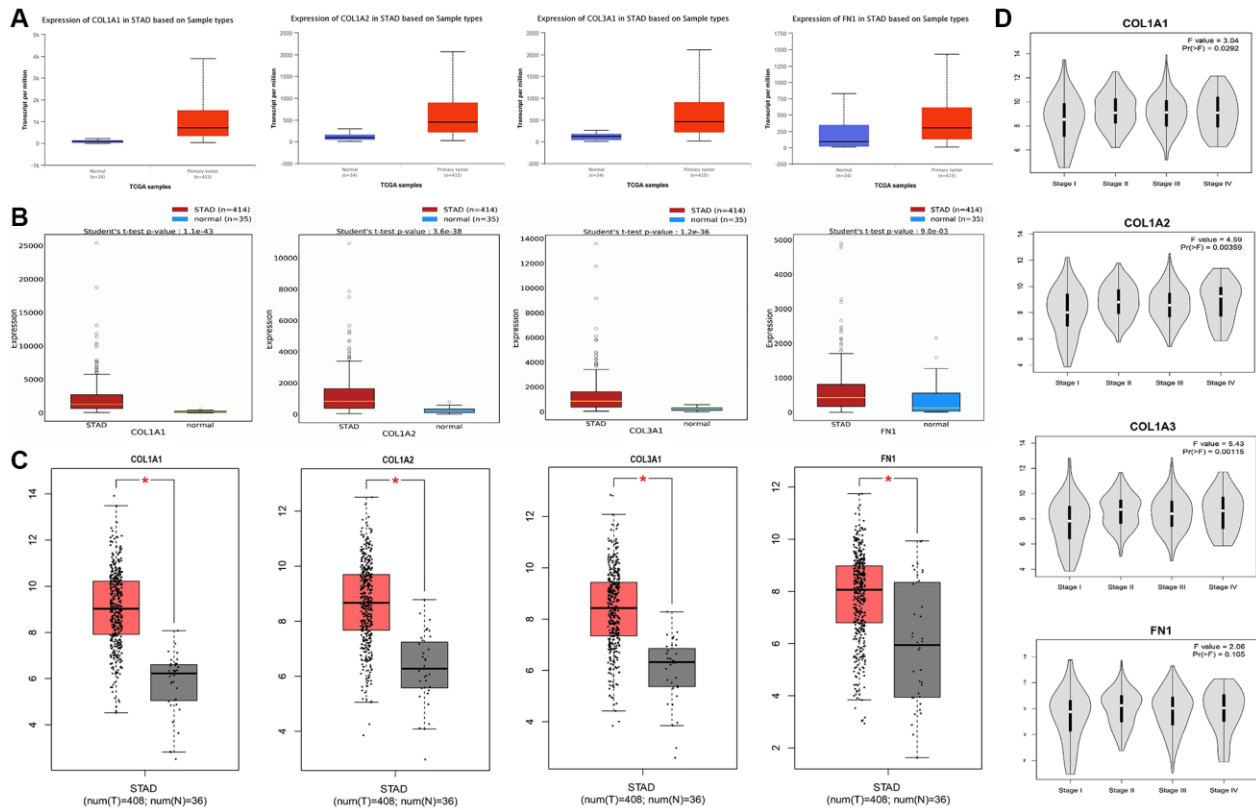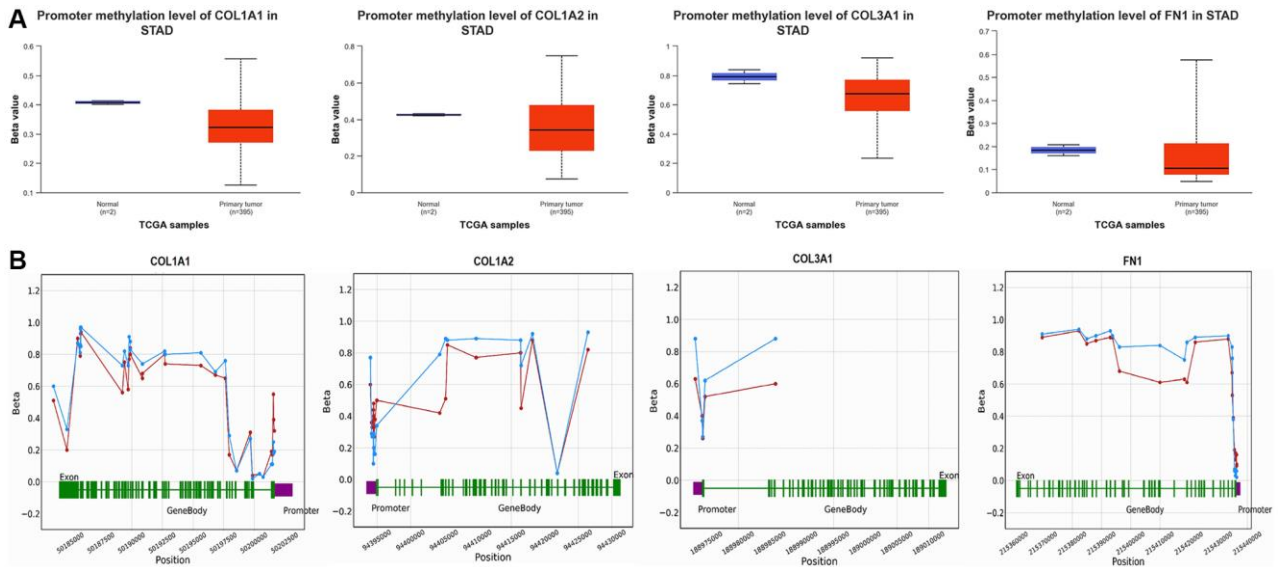

Figure 4. mRNA expression analysis of COL1A1, COL1A2, COL3A1, and FN1 using additional TCGA datasets of gastric cancer (GC). (**A**) Expression analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC and normal samples via UALCAN database. (**B**) Expression analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC and normal samples via OncoDB database. (**C**) Expression analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC and normal samples via GEO GEPIA. (**D**) Expression analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC samples belonging to different cancer stages. *P*-value < 0.05.
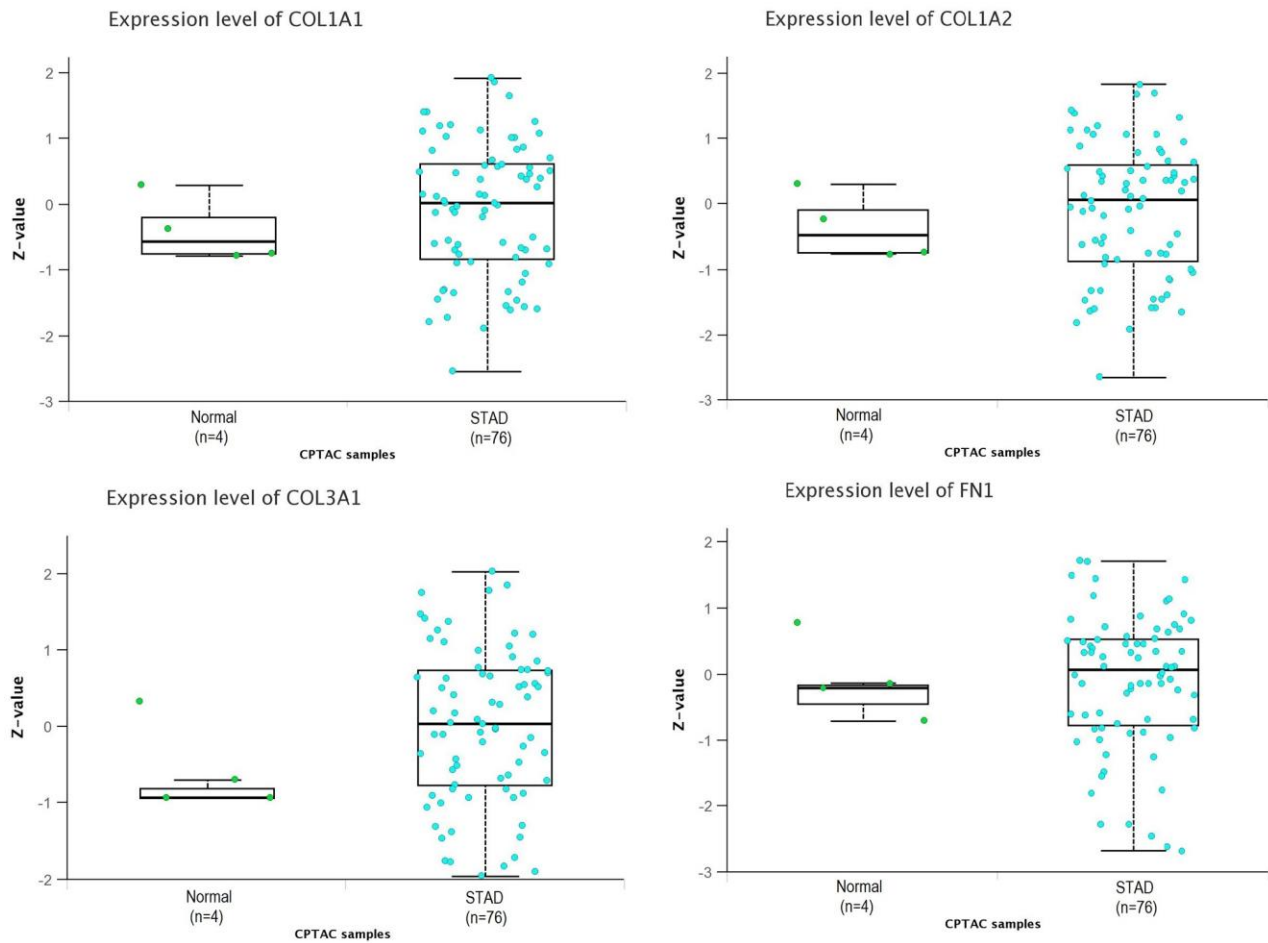
Figure 5. Proteomic expression analysis of COL1A1, COL1A2, COL3A1, and FN1 using additional database. This figure presents the proteomic expression analysis of COL1A1, COL1A2, COL3A1, and FN1 in gastric cancer (GC) and normal samples via UALCAN database. *P*-value < 0.05.


Figure 6. Promoter methylation and survival analyses of COL1A1, COL1A2, COL3A1, and FN1. (**A**) Promoter methylation analysis of COL1A1, COL1A2, COL3A1, and FN1 in gastric cancer (GC) and normal samples via UALCAN. (**B**) Promoter methylation analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC and normal samples via OncoDB. *P*-value < 0.05.

Enrichment and miRNA prediction analyses

In our study, we conducted enrichment analysis of COL1A1, COL1A2, COL3A1, and FN1 to gain insights into their functional roles. The analysis revealed that these hub genes are involved in a wide range of diverse GO terms and KEGG) pathways (Figure 9A–9D). These findings indicate that COL1A1,

COL1A2, COL3A1, and FN1 may play crucial roles in various BP, CC, and MF.

Furthermore, in our study, we utilized miRDB and ENCORI to predict the regulatory miRNAs targeting COL1A1, COL1A2, COL3A1, and FN1. The analysis via both databases revealed a total of 40 miRNAs that potentially target these hub genes (Figure 9E).
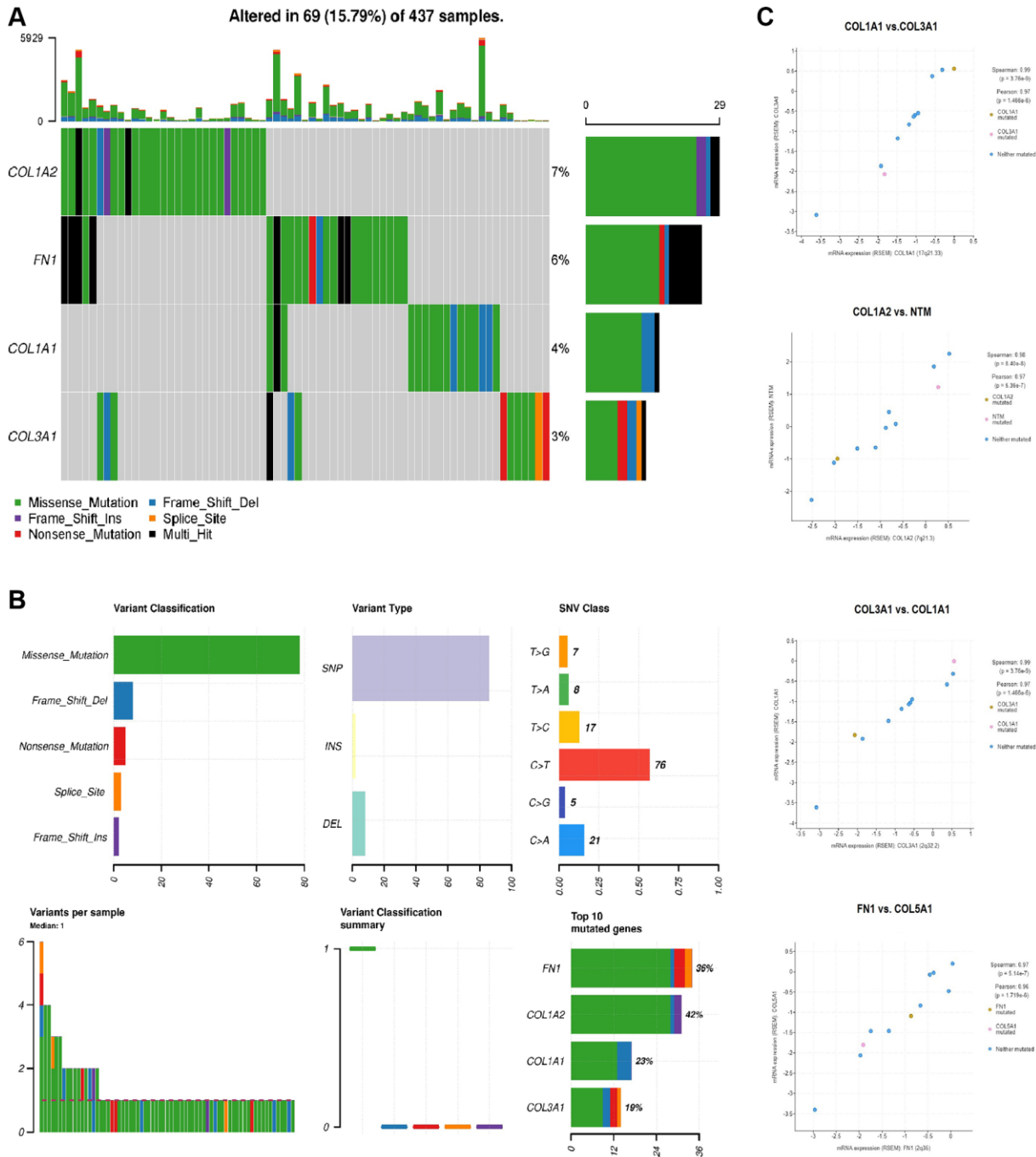

Figure 7. Mutational and co-express gene analysis of COL1A1, COL1A2, COL3A1, and FN1. (**A**) Detail of the mutational frequencies of COL1A1, COL1A2, COL3A1, and FN1 gens in gastric cancer (GC) samples. (**B**) Detailed summary of the mutations found in COL1A1, COL1A2, COL3A1, and FN1 genes across GC samples. (**C**) Significant co-expressed genes along with overexpressed COL1A1, COL1A2, COL3A1, and FN1 genes in GC samples. *P*-value < 0.05.

Remarkably, hsa-miR-29b-3p was found to target all hub genes simultaneously (Figure 9F). This observation suggests that hsa-miR-29b-3p may play a crucial role in the post-transcriptional regulation of COL1A1, COL1A2, COL3A1, and FN1, potentially modulating their expression levels.

Validation of COL1A1, COL1A2, COL3A1, and FN1 gene expression in clinical GC samples via RT-qPCR

To validate the results obtained from the GEO expression dataset, cDNA from both GC and control
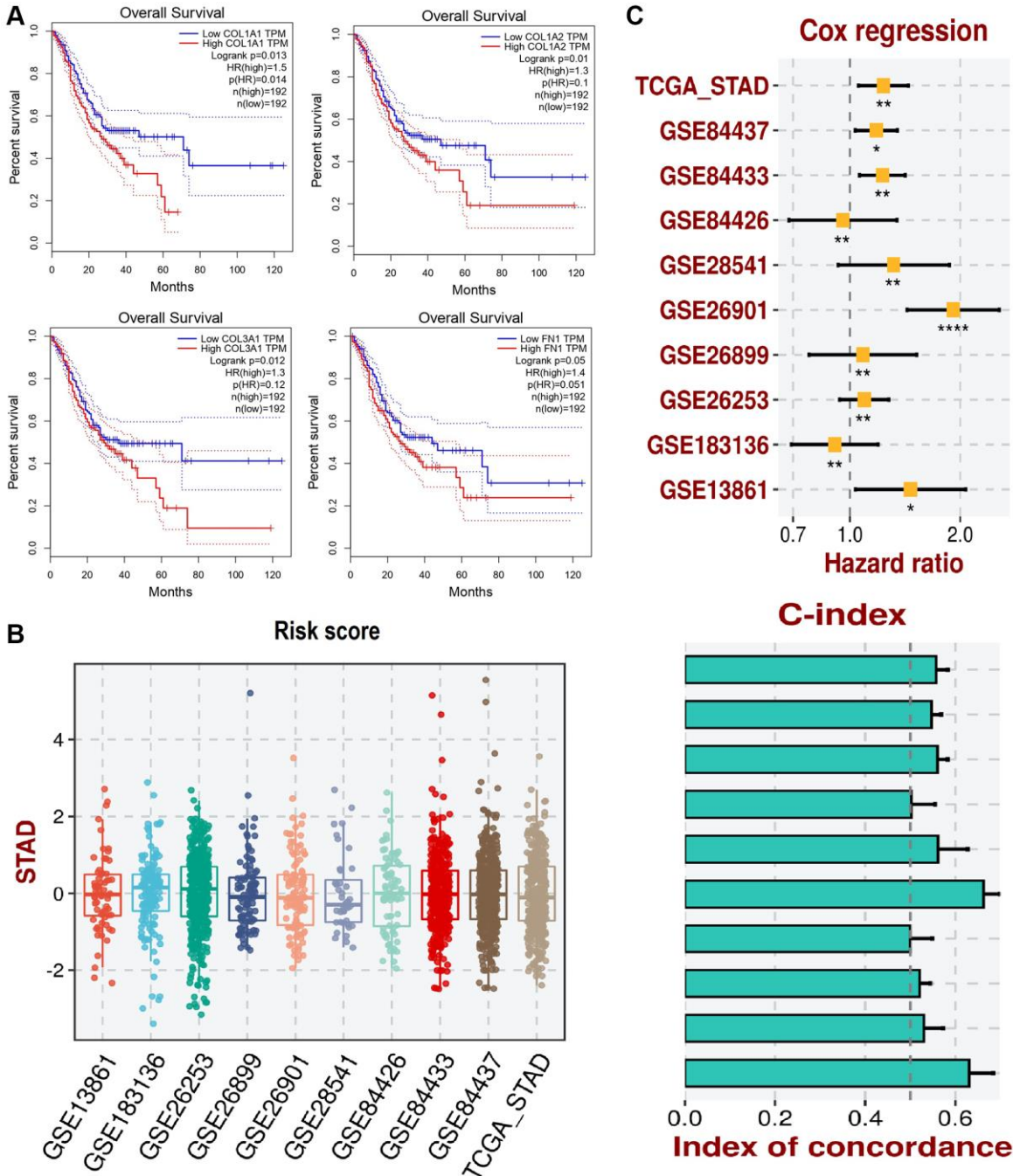
Figure 8. This figure illustrates the survival analysis and development of a prognostic model using the gene expression data of COL1A1, COL1A2, COL3A1, and FN1. (**A**) The survival analysis of these genes in gastric cancer (GC) patients is conducted via GEPIA. (**B**) Box plots representing the risk scores of patients in various GEO datasets and the TCGA_STAD dataset. (**C**) Cox regression analysis forest plot showing the hazard ratios of COL1A1, COL1A2, COL3A1, and FN gene expression levels for overall survival in STAD across different datasets and concordance index (C-index) bar plot for the predictive performance of the COL1A1, COL1A2, COL3A1, and FN gene expression models. The yellow squares represent the hazard ratio (HR) for each dataset with 95% confidence intervals. *P*-value < 0.05.

tissue samples was utilized for RT-qPCR analysis of COL1A1, COL1A2, COL3A1, and FN1. The results, as depicted in Figure 10A, demonstrated a significant increase in the expression levels of COL1A1, COL1A2, COL3A1, and FN1 in the GC sample group ($n = 39$) compared to the control group ($n = 39$, p-value < 0.05). Additionally, the ROC curves for COL1A1 (AUC: 1.0, p-value < 0.05), COL1A2 (AUC: 1.0, p-value < 0.05), COL3A1 (AUC: 1.0, p-value < 0.05), and FN1 (AUC: 1.0, p-value < 0.05) based on the expression levels exhibited significant diagnostic potential, sensitivity, and specificity (Figure 10B).

Targeted bisulfite-seq analysis to analyze promoter methylation levels of COL1A1, COL1A2, COL3A1, and FN1 in clinical GC samples

To assess the extent of promoter methylation in the hub genes COL1A1, COL1A2, COL3A1, and FN1 within clinical GC samples, we enrolled a total of 39 individuals diagnosed with GC, along with 39 healthy individuals from the Pakistani population. In both the GC and control groups, a high rate of bisulfite conversion (C to T) exceeding 99.1% was observed, and there were no notable differences in the read mapping rate between the two groups. Following stringent quality control measures, all 39 samples from the GC group and 39 samples from the control group were deemed suitable for subsequent analysis. Our analysis revealed a significant pattern of hypomethylation across all candidate genes (COL1A1, COL1A2, COL3A1, and FN1) in GC samples compared to the control group (Figure 11A). Furthermore, the ROC curves were generated for COL1A1 (AUC: 1.0, p-value $<$

0.05), COL1A2 (AUC: 1.0, p-value < 0.05), COL3A1 (AUC: 1.0, p-value < 0.05), and FN1 (AUC: 1.0, p-value < 0.05) based on their methylation levels (Figure 11B). These ROC curves demonstrated significant diagnostic potential with high AUC values of 1.0, indicating excellent discriminatory power. Additionally, the ROC curves exhibited remarkable sensitivity and specificity in distinguishing between GC and controls (Figure 11B).

Drug prediction analysis

The management of GC often involves medical treatment as the primary approach. Therefore, the careful selection of suitable candidate drugs becomes essential. In this current investigation, we utilized the DrugBank database to explore potential therapeutic drugs for GC, focusing on the identified hub genes (COL1A1, COL1A2, COL3A1, and FN1) as potential targets for treatment. Notably, our investigation yielded two important drugs deemed suitable for the treatment of GC with respect to identified hub genes, namely Acetaminophen and Cytarabine (Table 1).

Functional verification of the *in vitro* and *in vivo* roles of COL1A1, COL1A2, COL3A1, and FN1 in GC

The COL1A1, COL1A2, COL3A1, and FN1 genes work synergistically to regulate processes such as cell migration, invasion, and tissue remodeling. Therefore, the simultaneous silencing of COL1A1, COL1A2, COL3A1, and was carried out in AGS cells using siRNA to analyze their functional synergetic impact on
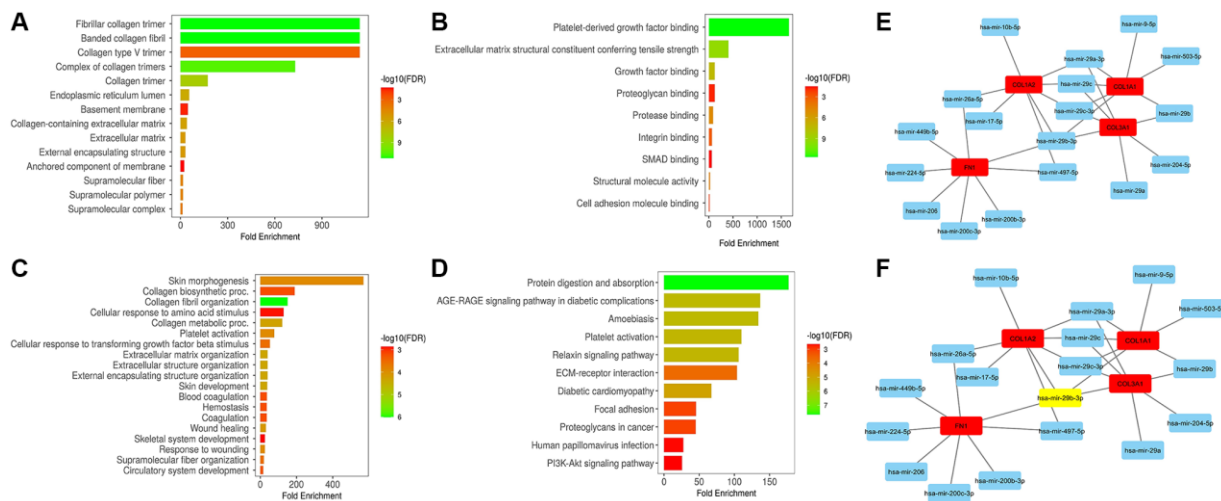
Figure 9. This figure showcases the gene enrichment and miRNA prediction analyses of COL1A1, COL1A2, COL3A1, and FN1. (**A**) Displays the associated cellular component (CC) terms. (**B**) Illustrates the associated molecular function (MF) terms. (**C**) Presents the associated biological process (BP) terms. (**D**) Shows the associated Kyoto Encyclopedia of Genes and Genomes (KEGG) terms. (**E**) Exhibits a protein-protein interaction (PPI) network of COL1A1, COL1A2, COL3A1, and FN1 along with their associated 40 miRNAs. (**F**) Demonstrates another PPI network of these genes and 40 miRNAs, highlighting the most significant miRNA (has-miR-29b-3p) in the network. *P*-value < 0.05.

the different parameters. The silencing efficiency was checked with the help of RT-qPCR. As shown in Figure 12A, reduced expression of COL1A1, COL1A2, COL3A1, and FN1 was observed in transfected AGS cell as compared to control AGS cells (Figure 12A).

Further assessments, via CCK-8 assays and colony-forming assays, indicated that the knockdown of COL1A1, COL1A2, COL3A1, and FN1 led to a reduction in cellular proliferation when compared to the control AGS cells (Figure 12B–12D).
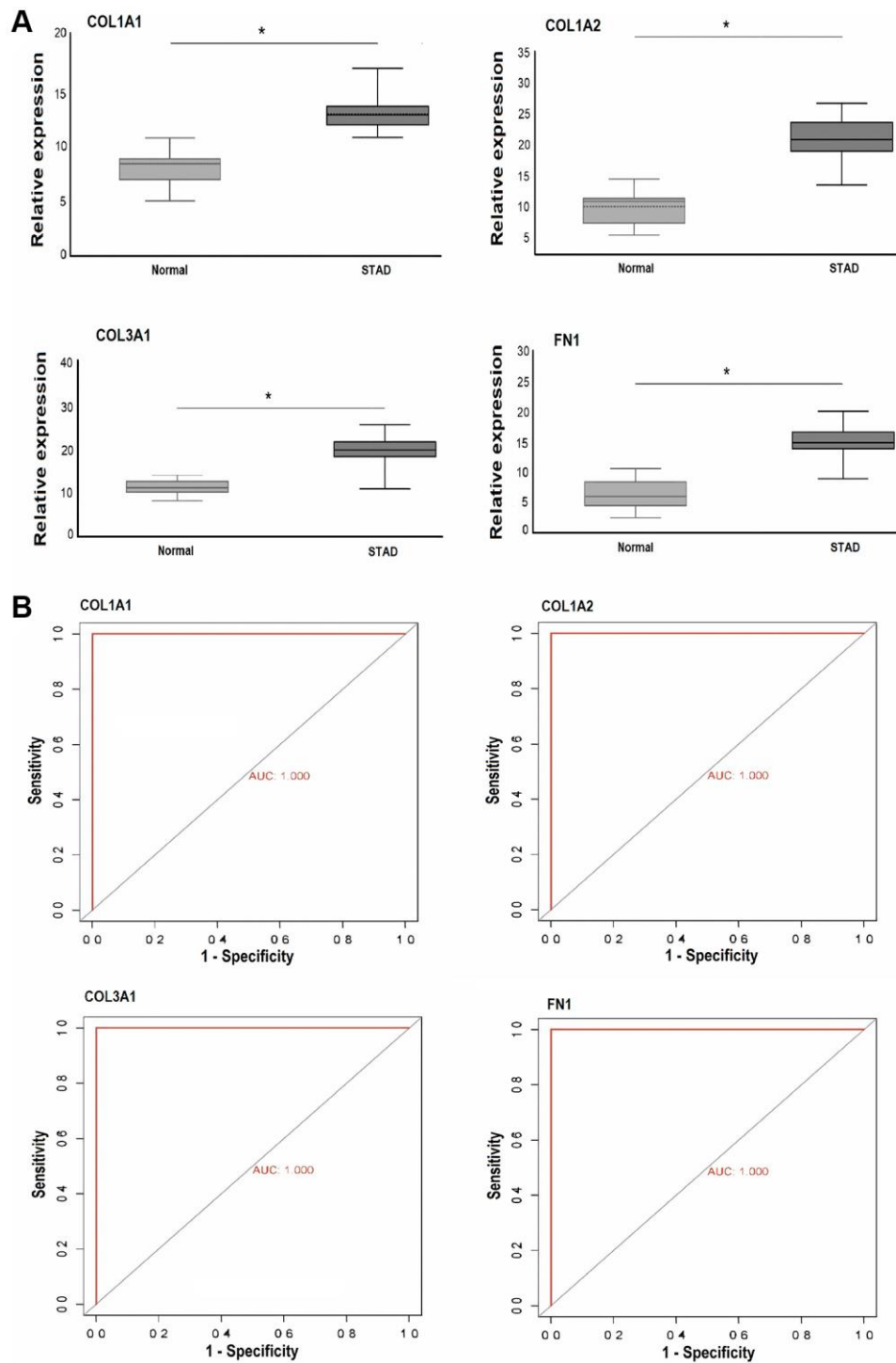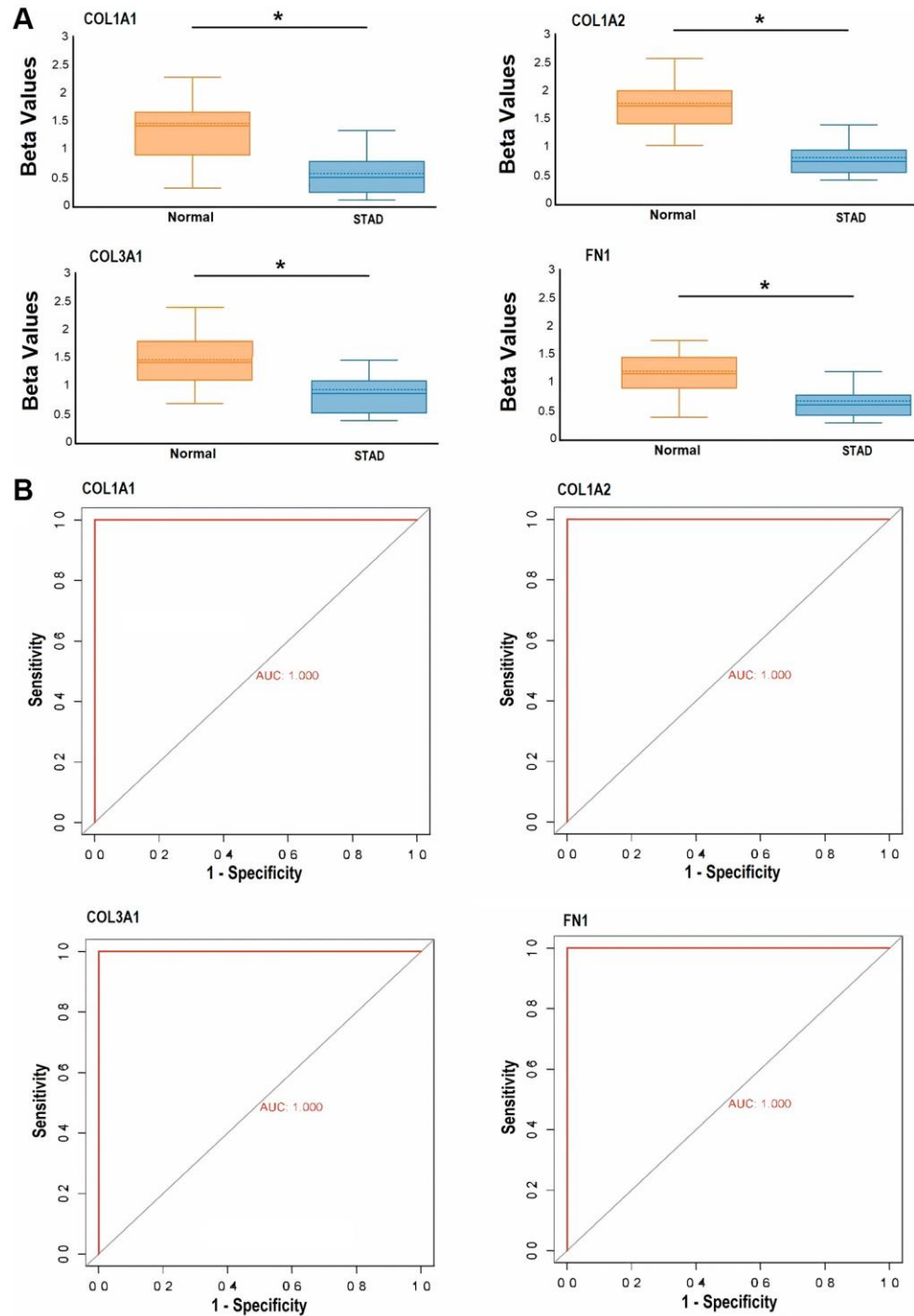

Figure 10. This figure depicts the relative expression and receiver operating characteristic (ROC) curve analysis of COL1A1, COL1A2, COL3A1, and FN1 in Pakistani gastric cancer (GC) patients and normal controls. (**A**) Presents the relative expression analysis of these genes in Pakistani GC patients and control samples via RT-qPCR. (**B**) Shows the ROC curves based on RT-qPCR expression of COL1A1, COL1A2, COL3A1, and FN1. A significance level of $P < 0.05$ was utilized as the selection criteria. P-value < 0.05.

DISCUSSION

In this study, we initially integrated three microarray expression profiles obtained from the GEO database, leading to the identification of 83 DEGs between GC and normal gastric tissues, with 41 up-regulated and 42 down-regulated genes. Subsequently, utilizing the degree and MCC methods, we designated COL1A1, COL1A2, COL3A1, and FN1 as hub genes, which exhibited significant up-regulation in GC. Furthermore,

Figure 11. Targeted bisulfite sequencing-based methylation level exploration and receiver operating characteristic (ROC) curve analysis of the hub genes, including COL1A1, COL1A2, COL3A1, and FN1 in Pakistani gastric cancer (GC) patients and normal controls. (**A**) Beta value-based methylation analysis of COL1A1, COL1A2, COL3A1, and FN1 in Pakistani GC patients and control samples, and (**B**) targeted bisulfite sequencing-based ROC curves of the COL1A1, COL1A2, COL3A1, and FN1 methylation level. *P*-value < 0.05.

Table 1. DrugBank-based DEGs-associated drugs.

Sr. no	Hub gene	Drug name	Effect	Reference	Group
1	COL1A1	Acetaminophen Cytarabine	Decrease expression of COL1A1 mRNA	A20418 A20508	Approved
2	COL1A2	Acetaminophen Cytarabine	Decrease expression of COL1A2 mRNA	A20418 A20508	Approved
3	COL3A1	Acetaminophen Cytarabine	Decrease expression of COL3A1 mRNA	A20418 A20508	Approved
4	FN1	Acetaminophen Cytarabine	Decrease expression of FN1 mRNA	A20418 A20508	Approved

we validated the expression of these hub genes on additional GC datasets from TCGA and clinical samples collected from Pakistani GC patients. The expression validation analysis further confirmed the significant up-regulation of COL1A1, COL1A2, COL3A1, and FN1 in GC patients compared to controls.

COL1A1, encoding the alpha-1 chain of collagen type I, plays a critical role in the extracellular matrix (ECM) and is essential for maintaining tissue integrity and strength [33]. This protein is known to be involved in cell adhesion, migration, and proliferation, making it a key player in various biological processes [34]. Dysregulation of COL1A1 has been implicated in tumorigenesis and cancer progression in multiple malignancies. Research has demonstrated the significance of COL1A1 in different cancers. For instance, in breast cancer, up-regulated

COL1A1 has been associated with tumor growth, invasion, and metastasis, promoting a pro-tumorigenic microenvironment [35]. Similarly, in pancreatic cancer, higher expression of COL1A1 has been found to enhance tumor cell proliferation and migration [36]. In hepatocellular carcinoma, overexpressed COL1A1 contributes to tumor progression and metastasis by modulating the tumor microenvironment [37]. In lung cancer, COL1A1 higher expression has been linked to tumor invasiveness and poor patient prognosis [38].

COL1A2, which codes for the alpha-2 chain of collagen type I, plays a critical role as a fundamental building block in the extracellular matrix (ECM), ensuring the integrity and strength of various tissues [39]. Like COL1A1, COL1A2 is involved in various cellular processes, including cell adhesion, migration, and proliferation, making it an important player in cancer
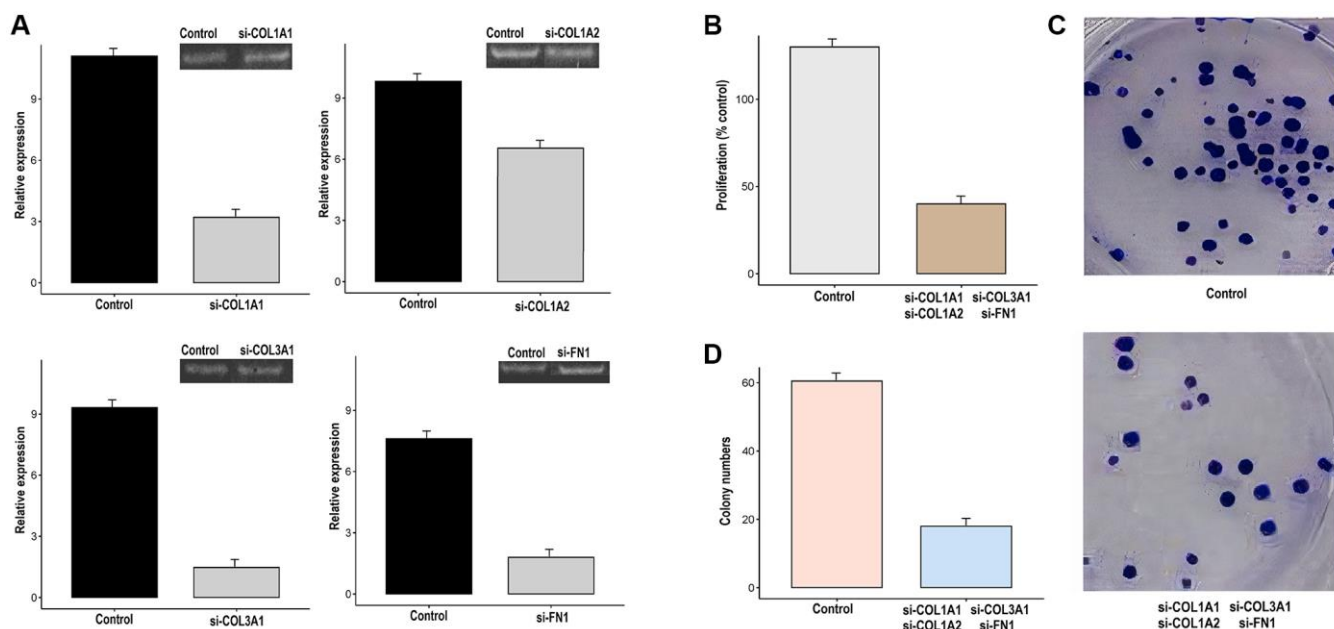
Figure 12. Knockdown of COL1A1, COL1A2, COL3A1, and FN1 impairs the growth and metastatic potential of gastric cancer (GC) cells (AGS). (**A**) The transfection efficiency of si-COL1A1, si-COL1A2, si-COL3A1, and si-FN1 was checked with the help of RT-qPCR, (**B**) AGS control and transfected cells were analyzed proliferation, (**C**, **D**) colony formation.

biology [40]. Emerging research has shed light on the role of COL1A2 in different cancer types. For example, in breast cancer, up-regulation of COL1A2 has been associated with increased tumor invasiveness and metastasis [41]. In GC, COL1A2 has been identified as a potential biomarker for tumor progression and prognosis [42]. Moreover, in lung cancer, COL1A2 expression has been linked to tumor growth and metastasis [43]. In ovarian cancer, COL1A2 has been associated with tumor cell proliferation and migration [44]. In colorectal cancer, COL1A2 has been found to play a role in tumor invasion and metastasis [45]. Taken together, the evidence highlights the importance of COL1A2 in various cancers, with its dysregulation contributing to tumor aggressiveness and metastasis.

COL3A1, encoding the alpha-1 chain of collagen type III, is an essential component of the extracellular matrix (ECM) that provides structural support and elasticity to tissues [46]. Research has highlighted the role of COL3A1 in various cancer types. Such as, in colorectal cancer, elevated COL3A1 expression has been associated with tumor growth, progression, and metastasis, indicating its potential as a prognostic marker [47, 48]. Similarly, in ovarian cancer, overexpressed COL3A1 has been found to promote tumor cell migration and invasion, contributing to disease aggressiveness [49]. Moreover, in hepatocellular carcinoma, higher expression of COL3A1 has been implicated in tumor growth and angiogenesis, affecting patient prognosis [50]. These studies underscore the significance of COL3A1 in cancer biology and highlight its potential as a therapeutic target and diagnostic marker.

FN1 encodes for a significant glycoprotein responsible for cell adhesion, migration, and tissue remodeling, playing a crucial role in these processes [51]. FN1 interactions with various components of the extra-cellular matrix (ECM) are essential for facilitating cell-matrix interactions and maintaining tissue organization [52]. Studies have provided valuable insights into the diverse role of FN1 in different cancer types. In breast cancer, elevated FN1 expression has been linked to tumor invasiveness and metastasis, contributing to poor patient outcomes [53]. Similarly, in pancreatic cancer, FN1 has been associated with tumor progression and resistance to therapy, indicating its potential as a therapeutic target [54]. Moreover, in GC, FN1 over-expression has been correlated with tumor aggressiveness and lymph node metastasis, suggesting its significance as a prognostic biomarker [55].

During present study, we observed hypomethylation of COL1A1, COL1A2, COL3A1, and FN1 promoter regions in GC. Previous studies have also reported dysregulation of COL1A1, COL1A2, COL3A1, and

FN1 promoter methylation in various cancers. Hypermethylation of the promoters has been linked to the down-regulation of these genes in breast, gastric, and colorectal cancers, contributing to tumor growth and invasion [56–59]. Conversely, hypomethylation of these genes has been observed in ovarian and lung cancers, leading to their overexpression and association with aggressive tumor phenotypes [60, 61].

Survival analysis of COL1A1, COL1A2, COL3A1, and FN1 in GC patients revealed the relevance of these genes with poor OS. Earlier studies have also reported the association of COL1A1, COL1A2, COL3A1, and FN1 expression with OS in different other cancer patients. For example, in breast cancer, high expression of COL1A1 has been associated with worse OS and distant metastasis [40]. In lung cancer, increased COL1A2 expression has been correlated with poorer OS and advanced tumor stage [62]. Similarly, up-regulation of COL3A1 and FN1 has been linked to unfavorable OS in stomach and ovarian cancers, respectively [63, 64].

Additionally, the results of this study emphasized that hsa-miR-29b-3p is a shared regulator of COL1A1, COL1A2, COL3A1, and FN1 expression. Dysregulation of this miRNA may play a role in the abnormal expression of COL1A1, COL1A2, COL3A1, and FN1, potentially contributing to the observed alterations in their expression levels. The hsa-miR-29b-3p has been implicated in tumorigenesis and cancer progression in different cancer types. For instance, in breast cancer, overexpressed hsa-miR-29c-3p has been shown to inhibit tumor cell migration and invasion by targeting specific genes involved in metastasis [65]. In GC, up-regulated hsa-miR-29c-3p has been reported to suppress tumor growth and induce apoptosis, indicating its tumor-suppressive function [66]. Conversely, in colorectal cancer, elevated hsa-miR-29c-3p has been found to promote cancer cell proliferation and invasiveness, suggesting an oncogenic role in this context [67].

The study presents valuable insights into GC through comprehensive analysis of gene expression profiles. Despite its strengths, including multi-database integration and identification of key hub genes, several limitations warrant consideration. The relatively small sample size and heterogeneity across datasets may limit result generalization. Additionally, biological variability and confounding factors were not fully addressed, potentially impacting result accuracy. Moreover, validation methods primarily focused on *in silico* and *in vitro* analyses, lacking extensive clinical validation. Lastly, the study's single-omics approach overlooks other molecular layers crucial for a holistic understanding of GC. Addressing these limitations in future research could enhance the study's clinical relevance and translational impact.

CONCLUSION

This extensive investigation, combining experimental analyses and computational approaches, enabled the identification of DEGs linked to GC. Among these DEGs, four promising biomarkers (COL1A1, COL1A2, COL3A1, and FN1) were discovered, demonstrating potential diagnostic and prognostic implications in GC. Additionally, these genes hold promise as therapeutic targets for the treatment of GC, presenting new opportunities for targeted interventions.

Abbreviations

GC: Gastric Cancer; GEO: Gene Expression Omnibus; DEGs: Differentially Expressed Genes; PPI: Protein-Protein Interaction; OS: Overall Survival; TCGA: The Cancer Genome Atlas; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.

AUTHOR CONTRIBUTIONS

Pei X and Luo Y conceived and designed the research. Zeng H, Jamil M, and Liu X analyzed and interpreted the data and wrote original draft of the manuscript which was approved by Bo J.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

ETHICAL STATEMENT AND CONSENT

The ethics committee of the Pakistan Agriculture Research Center (PARC) approved (Ref # 2987) the study in compliance with the Helsinki Declaration. Additionally, informed written consent was obtained from all participants prior to sample collection.

FUNDING

REFERENCES

1. Shin WS, Xie F, Chen B, Yu P, Yu J, To KF, Kang W. Updated Epidemiology of Gastric Cancer in Asia: Decreased Incidence but Still a Big Challenge. Cancers (Basel). 2023; 15:2639.
   https://doi.org/10.3390/cancers15092639
   PMID:37174105

2. Zhong N, Yu Y, Chen J, Shao Y, Peng Z, Li J. Clinicopathological characteristics, survival outcome and prognostic factors of very young gastric cancer. Clin Exp Med. 2023; 23:437–45.
   https://doi.org/10.1007/s10238-022-00822-3
   PMID:35451668

3. Liu L, Lin J, Zhao J, Yan P. Analysis of clinicopathologic characteristics and prognosis of gastric cancer in patients <40 years. Medicine (Baltimore). 2023; 102:e34635.
   https://doi.org/10.1097/MD.0000000000034635
   PMID:37653814

4. Chen Y, Sun Z, Wan L, Chen H, Xi T, Jiang Y. Tumor Microenvironment Characterization for Assessment of Recurrence and Survival Outcome in Gastric Cancer to Predict Chemotherapy and Immunotherapy Response. Front Immunol. 2022; 13:890922.
   https://doi.org/10.3389/fimmu.2022.890922
   PMID:35572498

5. Ansari KK, Wagh V, Saifi AI, Saifi I, Chaurasia S. Advancements in Understanding Gastric Cancer: A Comprehensive Review. Cureus. 2023; 15:e46046.
   https://doi.org/10.7759/cureus.46046
   PMID:37900456

6. Angom RS, Nakka NMR, Bhattacharya S. Advances in Glioblastoma Therapy: An Update on Current Approaches. Brain Sci. 2023; 13:1536.
   https://doi.org/10.3390/brainsci13111536
   PMID:38002496

7. Thakur S, Ghosh S. Chapter 18 - Recent advances in transcriptomic biomarker detection for cancer. In: Ali MA, Lee J, (eds.). Transcriptome Profiling. Academic Press. 2023; 453–78.
   https://doi.org/10.1016/B978-0-323-91810-7.00007-8

8. Das S, Dey MK, Devireddy R, Gartia MR. Biomarkers in Cancer Detection, Diagnosis, and Prognosis. Sensors (Basel). 2023; 24:37.
   https://doi.org/10.3390/s24010037
   PMID:38202898

9. Clough E, Barrett T. The Gene Expression Omnibus Database. Methods Mol Biol. 2016; 1418:93–110.
   https://doi.org/10.1007/978-1-4939-3578-9_5
   PMID:27008011

10. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015; 19:A68–77.
    https://doi.org/10.5114/wo.2014.47136
    PMID:25691825

11. Usman M, Hameed Y. GNB1, a novel diagnostic and prognostic potential biomarker of head and neck and liver hepatocellular carcinoma. J Cancer Res Ther. 2023.
    https://doi.org/10.4103/jcrt.jcrt_1901_20

12. Hu H, Umair M, Khan SA, Sani AI, Iqbal S, Khalid F, Sultan R, Abdel-Maksoud MA, Mubarak A, Dawoud

TM, Malik A, Saleh IA, Al Amri AA, et al. CDCA8, a mitosis-related gene, as a prospective pan-cancer biomarker: implications for survival prognosis and oncogenic immunology. Am J Transl Res. 2024; 16:432–45.
https://doi.org/10.62347/WSEF7878
PMID:38463578

13. Rajgopal S, Fredrick SJ, Parvathi VD. CircRNAs: Insights into Gastric Cancer. Gastrointest Tumors. 2021; 8:159–68.
https://doi.org/10.1159/000517303
PMID:34722469

14. 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.
https://doi.org/10.3389/fgene.2018.00265
PMID:30065754

15. Luu Truong Thanh H, Hoang TM, Hoang Van H. Identification of Hub Genes and Potential Pathogenesis in Gastric Cancer Based on Integrated Gene Expression Profile Analysis. Asian Pac J Cancer Prev. 2024; 25:885–92.
https://doi.org/10.31557/apjcp.2024.25.3.885
PMID:38546071

16. Zhao T, Chen Z, Liu W, Ju H, Li F. Identification of Hub Genes Associated with Gastric Cancer via Bioinformatics Analysis and Validation Studies. Int J Gen Med. 2023; 16:4835–48.
https://doi.org/10.2147/IJGM.S432284
PMID:37908756

17. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019; 47:D607–13.
https://doi.org/10.1093/nar/gky1131
PMID:30476243

18. Killcoyne S, Carter GW, Smith J, Boyle J. Cytoscape: a community-based framework for network modeling. Methods Mol Biol. 2009; 563:219–39.
https://doi.org/10.1007/978-1-60761-175-2_12
PMID:19597788

19. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BVS, Varambally S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017; 19:649–58.
https://doi.org/10.1016/j.neo.2017.05.002
PMID:28732212

20. Tang G, Cho M, Wang X. OncoDB: an interactive online database for analysis of gene expression and viral infection in cancer. Nucleic Acids Res. 2022; 50:D1334–9.
https://doi.org/10.1093/nar/gkab970
PMID:34718715

21. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. 2017; 45:W98–102.
https://doi.org/10.1093/nar/gkx247
PMID:28407145

22. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013; 6:pl1.
https://doi.org/10.1126/scisignal.2004088
PMID:23550210

23. Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. BMC Med Res Methodol. 2014; 14:116.
https://doi.org/10.1186/1471-2288-14-116
PMID:25323009

24. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022; 50:W216–21.
https://doi.org/10.1093/nar/gkac194
PMID:35325185

25. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020; 48:D127–31.
https://doi.org/10.1093/nar/gkz757
PMID:31504780

26. Gupta N. DNA Extraction and Polymerase Chain Reaction. J Cytol. 2019; 36:116–7.
https://doi.org/10.4103/JOC.JOC_110_18
PMID:30992648

27. Rio DC, Ares M Jr, Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc. 2010; 2010:pdb.prot5439.
https://doi.org/10.1101/pdb.prot5439
PMID:20516177

28. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25:402–8.
https://doi.org/10.1006/meth.2001.1262
PMID:11846609

29. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S,

Tzur D, Gautam B, Hassanali M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res. 2008; 36:D901–6.
https://doi.org/10.1093/nar/gkm958
PMID:18048412

30. Xu W, Li H, Hameed Y, Abdel-Maksoud MA, Almutairi SM, Mubarak A, Aufy M, Alturaiki W, Alshalani AJ, Mahmoud AM, Li C. Elucidating the clinical and immunological value of m6A regulator-mediated methylation modification patterns in adrenocortical carcinoma. Oncol Res. 2023; 31:819–31.
https://doi.org/10.32604/or.2023.029414
PMID:37547754

31. Kim TK. T test as a parametric statistic. Korean J Anesthesiol. 2015; 68:540–6.
https://doi.org/10.4097/kjae.2015.68.6.540
PMID:26634076

32. Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restor Dent Endod. 2017; 42:152–5.
https://doi.org/10.5395/rde.2017.42.2.152
PMID:28503482

33. Iacobescu GL, Iacobescu L, Popa MIG, Covache-Busuioc RA, Corlatescu AD, Cirstoiu C. Genomic Determinants of Knee Joint Biomechanics: An Exploration into the Molecular Basis of Locomotor Function, a Narrative Review. Curr Issues Mol Biol. 2024; 46:1237–58.
https://doi.org/10.3390/cimb46020079
PMID:38392197

34. Ouyang Z, Dong L, Yao F, Wang K, Chen Y, Li S, Zhou R, Zhao Y, Hu W. Cartilage-Related Collagens in Osteoarthritis and Rheumatoid Arthritis: From Pathogenesis to Therapeutics. Int J Mol Sci. 2023; 24:9841.
https://doi.org/10.3390/ijms24129841
PMID:37372989

35. Liu J, Shen JX, Wu HT, Li XL, Wen XF, Du CW, Zhang GJ. Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target. Discov Med. 2018; 25:211–23.
PMID:29906404

36. Chakravarthy D, Muñoz AR, Su A, Hwang RF, Keppler BR, Chan DE, Halff G, Ghosh R, Kumar AP. Palmatine suppresses glutamine-mediated interaction between pancreatic cancer and stellate cells through simultaneous inhibition of survivin and COL1A1. Cancer Lett. 2018; 419:103–15.
https://doi.org/10.1016/j.canlet.2018.01.057
PMID:29414301

37. Shields MA, Dangi-Garimella S, Redig AJ, Munshi HG. Biochemical role of the collagen-rich tumour microenvironment in pancreatic cancer progression. Biochem J. 2012; 441:541–52.
https://doi.org/10.1042/BJ20111240
PMID:22187935

38. Li X, Sun X, Kan C, Chen B, Qu N, Hou N, Liu Y, Han F. COL1A1: A novel oncogenic gene and therapeutic target in malignancies. Pathol Res Pract. 2022; 236:154013.
https://doi.org/10.1016/j.prp.2022.154013
PMID:35816922

39. Mienaltowski MJ, Gonzales NL, Beall JM, Pechanec MY. Basic Structure, Physiology, and Biochemistry of Connective Tissues and Extracellular Matrix Collagens. Adv Exp Med Biol. 2021; 1348:5–43.
https://doi.org/10.1007/978-3-030-80614-9_2
PMID:34807414

40. Yin W, Zhu H, Tan J, Xin Z, Zhou Q, Cao Y, Wu Z, Wang L, Zhao M, Jiang X, Ren C, Tang G. Identification of collagen genes related to immune infiltration and epithelial-mesenchymal transition in glioma. Cancer Cell Int. 2021; 21:276.
https://doi.org/10.1186/s12935-021-01982-0
PMID:34034744

41. Shi Y, Zheng C, Jin Y, Bao B, Wang D, Hou K, Feng J, Tang S, Qu X, Liu Y, Che X, Teng Y. Reduced Expression of METTL3 Promotes Metastasis of Triple-Negative Breast Cancer by m6A Methylation-Mediated COL3A1 Up-Regulation. Front Oncol. 2020; 10:1126.
https://doi.org/10.3389/fonc.2020.01126
PMID:32766145

42. Li D, Yin Y, He M, Wang J. Identification of Potential Biomarkers Associated with Prognosis in Gastric Cancer via Bioinformatics Analysis. Med Sci Monit. 2021; 27:e929104.
https://doi.org/10.12659/MSM.929104
PMID:33582701

43. Ji J, Zhao L, Budhu A, Forgues M, Jia HL, Qin LX, Ye QH, Yu J, Shi X, Tang ZY, Wang XW. Let-7g targets collagen type I alpha2 and inhibits cell migration in hepatocellular carcinoma. J Hepatol. 2010; 52:690–7.
https://doi.org/10.1016/j.jhep.2009.12.025
PMID:20338660

44. Li S, Li H, Xu Y, Lv X. Identification of candidate biomarkers for epithelial ovarian cancer metastasis using microarray data. Oncol Lett. 2017; 14:3967–74.
https://doi.org/10.3892/ol.2017.6707
PMID:28943904

45. Yu Y, Liu D, Liu Z, Li S, Ge Y, Sun W, Liu B. The inhibitory effects of COL1A2 on colorectal cancer cell proliferation, migration, and invasion. J Cancer. 2018; 9:2953–62.
https://doi.org/10.7150/jca.25542

PMID:30123364

46. Doherty EL, Aw WY, Warren EC, Hockenberry M, Whitworth CP, Krohn G, Howell S, Diekman BO, Legant WR, Nia HT, Hickey AJ, Polacheck WJ. Patient-derived extracellular matrix demonstrates role of COL3A1 in blood vessel mechanics. Acta Biomater. 2023; 166:346–59.
https://doi.org/10.1016/j.actbio.2023.05.015
PMID:37187299

47. Torres S, Bartolomé RA, Mendes M, Barderas R, Fernandez-Aceñero MJ, Peláez-García A, Peña C, Lopez-Lucendo M, Villar-Vázquez R, de Herreros AG, Bonilla F, Casal JI. Proteome profiling of cancer-associated fibroblasts identifies novel proinflammatory signatures and prognostic markers for colorectal cancer. Clin Cancer Res. 2013; 19:6006–19.
https://doi.org/10.1158/1078-0432.CCR-13-1130
PMID:24025712

48. Shahrajabian MH, Sun W. Mechanism of Action of Collagen and Epidermal Growth Factor: A Review on Theory and Research Methods. Mini Rev Med Chem. 2024; 24:453–77.
https://doi.org/10.2174/1389557523666230816090054
PMID:37587815

49. Zhou J, Yang Y, Zhang H, Luan S, Xiao X, Li X, Fang P, Shang Q, Chen L, Zeng X, Yuan Y. Overexpressed COL3A1 has prognostic value in human esophageal squamous cell carcinoma and promotes the aggressiveness of esophageal squamous cell carcinoma by activating the NF-κB pathway. Biochem Biophys Res Commun. 2022; 613:193–200.
https://doi.org/10.1016/j.bbrc.2022.05.029
PMID:35598375

50. Chai ZT, Zhu XD, Ao JY, Wang WQ, Gao DM, Kong J, Zhang N, Zhang YY, Ye BG, Ma DN, Cai H, Sun HC. microRNA-26a suppresses recruitment of macrophages by down-regulating macrophage colony-stimulating factor expression through the PI3K/Akt pathway in hepatocellular carcinoma. J Hematol Oncol. 2015; 8:56.
https://doi.org/10.1186/s13045-015-0150-4
PMID:26021873

51. Kulus J, Kulus M, Kranc W, Jopek K, Zdun M, Józkowiak M, Jaśkowski JM, Piotrowska-Kempisty H, Bukowska D, Antosik P, Mozdziak P, Kempisty B. Transcriptomic Profile of New Gene Markers Encoding Proteins Responsible for Structure of Porcine Ovarian Granulosa Cells. Biology (Basel). 2021; 10:1214.
https://doi.org/10.3390/biology10111214
PMID:34827207

52. Horwacik I. The Extracellular Matrix and Neuroblastoma Cell Communication-A Complex Interplay and Its Therapeutic Implications. Cells. 2022; 11:3172.
https://doi.org/10.3390/cells11193172
PMID:36231134

53. Saatci O, Kaymak A, Raza U, Ersan PG, Akbulut O, Banister CE, Sikirzhytski V, Tokat UM, Aykut G, Ansari SA, Dogan HT, Dogan M, Jandaghi P, et al. Targeting lysyl oxidase (LOX) overcomes chemotherapy resistance in triple negative breast cancer. Nat Commun. 2020; 11:2416.
https://doi.org/10.1038/s41467-020-16199-4
PMID:32415208

54. Byers LA, Diao L, Wang J, Saintigny P, Girard L, Peyton M, Shen L, Fan Y, Giri U, Tumula PK, Nilsson MB, Gudikote J, Tran H, et al. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. Clin Cancer Res. 2013; 19:279–90.
https://doi.org/10.1158/1078-0432.CCR-12-1558
PMID:23091115

55. Xu B, Bai Z, Yin J, Zhang Z. Global transcriptomic analysis identifies SERPINE1 as a prognostic biomarker associated with epithelial-to-mesenchymal transition in gastric cancer. PeerJ. 2019; 7:e7091.
https://doi.org/10.7717/peerj.7091
PMID:31218131

56. Wang R, Fu L, Li J, Zhao D, Zhao Y, Yin L. Microarray Analysis for Differentially Expressed Genes Between Stromal and Epithelial Cells in Development and Metastasis of Invasive Breast Cancer. J Comput Biol. 2020; 27:1631–43.
https://doi.org/10.1089/cmb.2019.0154
PMID:32429691

57. Zhang L, Lu Q, Chang C. Epigenetics in Health and Disease. Adv Exp Med Biol. 2020; 1253:3–55.
https://doi.org/10.1007/978-981-15-3449-2_1
PMID:32445090

58. Ullah L, Hameed Y, Ejaz S, Raashid A, Iqbal J, Ullah I, Ejaz SA. Detection of novel infiltrating ductal carcinoma-associated BREast CAncer gene 2 mutations which alter the deoxyribonucleic acid-binding ability of BReast CAncer gene 2 protein. J Cancer Res Ther. 2020; 16:1402–7.
https://doi.org/10.4103/jcrt.JCRT_861_19
PMID:33342804

59. Ahmad M, Khan M, Asif R, Sial N, Abid U, Shamim T, Hameed Z, Iqbal MJ, Sarfraz U, Saeed H. Expression characteristics and significant diagnostic and prognostic values of ANLN in human cancers. Int J Gen Med. 2022; 1957–72.

60. Jayaraman H, Anandhapadman A, Ghone NV. In Vitro

and In Vivo Comparative Analysis of Differentially Expressed Genes and Signaling Pathways in Breast Cancer Cells on Interaction with Mesenchymal Stem Cells. Appl Biochem Biotechnol. 2023; 195:401–31.
https://doi.org/10.1007/s12010-022-04119-9
PMID:36087230

61. Wieder R. Fibroblasts as Turned Agents in Cancer Progression. Cancers (Basel). 2023; 15:2014.
https://doi.org/10.3390/cancers15072014
PMID:37046676

62. Li T, Gao X, Han L, Yu J, Li H. Identification of hub genes with prognostic values in gastric cancer by bioinformatics analysis. World J Surg Oncol. 2018; 16:114.
https://doi.org/10.1186/s12957-018-1409-3
PMID:29921304

63. Hao S, Lv J, Yang Q, Wang A, Li Z, Guo Y, Zhang G. Identification of Key Genes and Circular RNAs in Human Gastric Cancer. Med Sci Monit. 2019; 25:2488–504.
https://doi.org/10.12659/MSM.915382
PMID:30948703

64. Gu Z, Cui X, Wang X. Construction of Prognostic Prediction Model for Stomach Adenocarcinoma Based on the TCGA Database. Res Sq. 2020.
https://doi.org/10.21203/rs.3.rs-114928/v1

65. Chen J, Lou W, Ding B, Wang X. Overexpressed pseudogenes, DUXAP8 and DUXAP9, promote growth of renal cell carcinoma and serve as unfavorable prognostic biomarkers. Aging (Albany NY). 2019; 11:5666–88.
https://doi.org/10.18632/aging.102152
PMID:31409759

66. Zhang S, Xiang X, Liu L, Yang H, Cen D, Tang G. Bioinformatics Analysis of Hub Genes and Potential Therapeutic Agents Associated with Gastric Cancer. Cancer Manag Res. 2021; 13:8929–51.
https://doi.org/10.2147/CMAR.S341485
PMID:34876855

67. Zhang S, Jin J, Tian X, Wu L. hsa-miR-29c-3p regulates biological function of colorectal cancer by targeting SPARC. Oncotarget. 2017; 8:104508–24.
https://doi.org/10.18632/oncotarget.22356
PMID:29262657