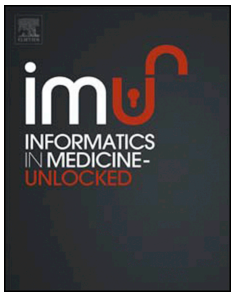


Revealing novel biomarkers involved in development and progression of gastric cancer by comprehensive bioinformatics analysis

Homa Akhavan, Sina Ramezani, Zinat Shams, Saied Hosseini-Asl

PII: S2352-9148(21)00120-9

DOI: https://doi.org/10.1016/j.imu.2021.100630

Reference: IMU 100630

To appear in: *Informatics in Medicine Unlocked*

Received Date: 19 April 2021

Revised Date: 27 May 2021

Accepted Date: 31 May 2021

Please cite this article as: Akhavan H, Ramezani S, Shams Z, Hosseini-Asl S, Revealing novel biomarkers involved in development and progression of gastric cancer by comprehensive bioinformatics analysis, *Informatics in Medicine Unlocked* (2021), doi: https://doi.org/10.1016/j.imu.2021.100630.

Revealing Novel Biomarkers involved in Development and Progression of Gastric Cancer by Comprehensive Bioinformatics Analysis

Homa Akhavan[1], Sina Ramezani[2], Zinat Shams[3], Saied Hosseini-Asl[1*]

[1]Department of Genomics Research, Digestive Disease Research Center, Ardabil University of Medical Sciences, Ardabil, Iran

[2] Department of Biology, University of Guilan, Rasht, Iran

[3]Department of biological science, kharazmi university, Tehran, Iran

*Corresponding author: Associate Professor Saied Hosseini-Asl, Department of Genomics Research, Digestive Disease Research Center, Ardabil University of Medical Sciences, Ardabil, Iran

Email: saied.hosseiniasl@arums.ac.ir

Tel: +98 45 33712634; Fax: +98 45 33261449

Abstract: Gastric cancer (GC) is the third cause of cancer mortality in the world but the molecular mechanisms underlying the pathogenesis of GC remain little known. This study aimed to provide novel insights into GC tumorigenesis and identify potential key genes for the clinical management of patients through comprehensive bioinformatics analysis. mRNA (GSE26942, GSE66229, and GSE54129) and miRNA (GSE26595) microarray datasets were downloaded and Differentially expressed genes (DEGs) and differentially expressed miRNAs (DEmiRs) were obtained using R software. The FunRich database was applied to analyze the function and pathways enrichment of DEGs. Protein-protein interaction (PPI) network was assessed using STRING and visualized by Cytoscape software. Then, the value of key genes were validated. There were 516 DEGs that overlapped in three expression profile datasets and predicted targets of DEmiRs. DEGs were mainly enriched in biological processes related to apoptosis and regulation of nucleobase, nucleoside, nucleotide, and nucleic acid metabolism. Pathway analysis illustrated that DEGs were enriched in P53 signaling pathway, pathways in cancer, PI3K-AKT signaling pathway, small cell lung cancer, MicroRNAs in cancer, and apoptosis. We identified 5 genes (*CEMIP*, *CLDN1*, *SERPINE1*, *PMEPA1*, and *LIFR*) that were common amongst all three datasets and predicted targets of DEmiRs, with a good performance in predicting overall survivals. Furthermore, we constructed miRNAs–mRNAs network, which revealed miRNAs and genes involved in the development and progression of GC, including hsa-miR-421, hsa-miR-193a-3p, hsa-miR-576-5p, hsa-miR-1246, *CTC1*, *RGMB*, *E2F6*, *IGF1*, *JARID2*, and *PHKA1*. The findings of this study improved the understanding of molecular mechanisms of GC and the roles of identified DEmiRs in GC through interactions with DEGs may provide potential targets for GC diagnosis and treatment.

Key words: Gastric cancer; GC; differentially expressed genes; Bioinformatics analysis; differentially expressed gene; microRNA

1. Introduction:

Being one of the major malignancies worldwide gastric cancer (GC) is the fifth (5.7%) most commonly diagnosed cancer and third (8.2%) leading cause of cancer-related mortality[1]. In 2020, GC was diagnosed among over one million people and caused approximately 769,000 deaths[2]. Although significant progresses have been achieved in recent years, the current standard treatments for patients with advanced GC stages remain unsatisfactory. The 5-year survival rate of advanced GC is 20–40% after radical gastrectomy combined with chemotherapy[3, 4]. Due to little symptoms in early stage of GC, the diagnosis is often delayed and some patients still suffer from unusual patterns of systemic recurrence[5]. Currently, the most frequent serum based tumor biomarkers for early detection of GC, including pepsinogen and α-fetoprotein (AFP), the carbohydrate antigens (CA), CA19-9, CA24-2, CA72-4, CA50, and CA125, as well as carcinoembryonic antigen (CEA), but sensitivity and specificity of these biomarkers are poor. Therefore, it is critical to identify molecular mechanisms and biomarkers that can be used for diagnosing GC and predicting recurrence.

Recent advances in high-throughput microarray technologies could be applied to shed new light on the molecular mechanisms of human diseases. The major public databases such as Gene Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA) are powerful public data repositories used to find and analyze differentially expressed genes (DEGs) corresponding to the

carcinogenesis and progression of various cancers [6-9]. Gene expression profiling combined with bioinformatics analysis has been employed to identify DEGs and signaling pathways that are associated with tumorigenesis and the tumor grade in human GC[7]. Yong et al utilized GEO, Oncomine, and other databases to examine the expression of PPP2CA in colorectal cancer and suggested that PPP2CA has an oncogene role and could be used as a prognostic biomarker in the progression of colorectal cancer[10]. *Wang et al.* confirmed *SERPINH1* as a core gene involved in the regulation of GC development and promotes migration, cell cycle, and proliferation of GC cells via bioinformatics analysis and in vitro experiments[11]. *Liu et al.* used GEO, protein–protein interaction networks and other databases for bioinformatics analysis to examine DEGs related to invasion and metastases of GCs. After that, GC tissues were analyzed for validating bioinformatics results that high levels of *BGN* expression were associated with GC clinicopathological characteristics, including microvascular tumor thrombus, lymph node metastases, and vessel invasion [12].

MicroRNAs (miRNAs) are a large group of small non-coding RNAs of ~22 nucleotides which act as crucial regulators of gene expression through binding to the 3' untranslated region (3'UTR) of target mRNAs resulting in post- transcriptional inhibition of gene expression[13, 14]. Differentially expressed miRNAs (DEmiRs) have been reported to be correlated with the onset and progression of multiple tumor types, such as GC[15].For example, A study on GC revealed up-regulation of miR-106a and a member of miR-17 in GC[16].

Although a number of studies on DEGs and DEmiRs have been conducted and some of their functions in molecular functions, biological processes, and different pathways have been reported, there are still questions about how the DEGs and microRNAs interact through molecular pathways. Therefore, analyzing DEGs and DEmiRs illuminating the interactions

network among them is vital for understanding the molecular mechanisms of GC's causes and pathogenesis, which leads to further investigations for predictive and curative purposes.

In the present study, we identified differentially expressed genes and microRNAs by analyzing three GC mRNA microarray datasets and one microRNA dataset. The purpose of this study was to distinguish key genes and miRNAs in GC using bioinformatics analysis to identify new potential diagnostic, therapeutic molecular markers of GC.

2. Materials and methods

2.1. Microarray data

Three gene expression datasets (GSE26942, GSE66229, and GSE54129) and one miRNA expression dataset (GSE26595) were downloaded from the GEO database[17, 18]. The GSE26942 dataset was comprised of 205 GC and 12 normal gastric tissue sample mRNA expression datasets; GSE66229 included 300 GC tissue and 100 non-cancer tissues sample mRNA expression datasets; GSE54129 contained 111 GC and 21 normal gastric tissue sample mRNA expression datasets, and GSE26595 was comprised of 60 GC tissues and 8 non-cancer tissues.

2.2. Identification of DE-microRNAs and DEGs

All data were processed using R software (version 3.5.1, https://www.r-project.org/), and the LIMMA package (Linear Models for Microarray Data) was applied to identify DEmiRs and DEGs between GC tissue samples and control samples. To detect the DEmiRs, the p-value <0.01 and logFC > 1 cutoff criterion were obtained in the screening. For DEGs, the threshold was P-value < 0.05 and logFC> 0.01.

2.3. Prediction of targets for differentially expressed miRNAs

MultiMiR package (http://multimir.ucdenver.edu/) was used to predict targets of miRNAs[19]. MultiMiR package was used to predict targets of miRNAs by miRTarBase, TarBase, and miRcode, and only the target genes predicted in all three databases were selected for the following analysis. Venn diagram was applied to obtain a miRNA-target relationship, which were matched with DEGs acquired by microarray analysis to identify the interaction between DEmiRs and DEGs.

2.4. Functional enrichment analyses

FunRich (http://funrich.org/ faq) , which is an analysis tool applied to predict molecular function, biological processes, cellular components, and pathways of the selected target genes[20]. Statistical cutoff of enrichment analyses in FunRich software was set to <0.05 as usual and default quantity in researches.

2.5. Protein–protein interaction network generation and module analysis

The protein–protein interaction (PPI) network was constructed using the Search Tool for the Retrieval of Interacting Genes and Proteins (STRING) database (https://string-db.org)[21], followed by visualization using Cytoscape software[22]. The Molecular Complex Detection (MCODE) plug-in was applied to screen modules of hub genes from the PPI network with degree cutoff = 2, max. Depth = 100, k-core = 2, and node score cutoff = 0.2[23].

2.6. Survival analysis of key genes

Gene Expression Profiling Interactive Analysis (GEPIA) is a web server specialized for analyzing the RNA- seq data, which was utilized to compare mRNA expression between 211 GC samples and 408 normal samples based on data from TCGA database (https://portal.gdc.cancer.gov/). GEPIA (http://gepia.cancer-pku.cn) and Kaplan-Meier plotter (KM plotter) tool (http://kmplot.com/analysis/index.php?p=background) were applied to validate the role of the key genes in the progression of GC as well as the transcriptional levels in normal gastric and GC samples to predict the prognostic value of the key genes in GC patients[24, 25].

2.7. Construction of predicted miRNAs–mRNAs network

Target genes of DEmiRs were predicted using MultiMiR package. We used CytoHubba package to construct the regulatory network of predicted miRNAs-mRNAs, then The top 10 mRNA and miRNAs with the highest degree were discovered.

3. Results

3.1. Identification of DEGs and DEmiRs

A total of 15 DEmiRs were identified, among which 9 miRNAs were significantly up-regulated, and 6 miRNAs were significantly down-regulated (Table1) in GC tissues compared to non- cancerous gastric tissues. After searching by miRTarBase, TarBase, and miRcode databases, a total of 1716 target genes were predicted for DEmiRs. Overlap targets of DEmiRs with selected DEGs were identified, and VennDiagram was constructed for showing these overlap genes (Figure1), Details are in Supplementary. Out of 516 commonly DEGs among at least one of three datasets(GSE26942, GSE66229, and GSE54129) and predicted targets of DEmiRs, 5 genes including 4 up-regulated(*CEMIP*, *CLDN1*, *SERPINE1*, and *PMEPA1*) and a down-regulated gene (*LIFR*) were common amongst all three datasets and predicted targets of

158 DEmiRs. We determined that hsa-miR-421 and hsa-miR-193a-3p were the main DEmiRs which
159 target these 5 genes. Moreover, among three datasets, 57common DEGs which were not
160 predicted as targets of DEmiRs were obtained. Volcano Plots are presented in Supplementary
161 Figure S1
162
163
164
165
166
167 **Table 1.**significantly up-regulated and downregulated miRNAs in GC tissues compared with
168 normal gastric tissues.

	miRNAs	logFC	P.Value	adj.P.Val
Highly expressed miRNAs	hsa-miR-425	1.110298	3.97E-06	1.88E-06
	hsa-miR-10a	1.242239	1.15E-05	0.000129
	hsa-miR-98	1.273005	4.58E-05	0.001665
	hsa-let-7d*	1.369433	3.43E-05	0.0006
	hsa-miR-421	1.397872	4.72E-06	0.008625
	hsa-miR-576-5p	1.458824	8.12E-06	0.000185
	hsa-miR-1246	1.747138	9.95E-09	2.22E-05
	hsa-miR-196b	2.033751	2.13E-06	0.001918
	hsa-miR-135b	2.344572	3.24E-07	8.66E-05
Low expressed miRNAs	hsa-miR-204	-2.47754	9.82E-14	4.03E-11

hsa-miR-363	-1.77033	6.83E-06	0.000175
hsa-miR-29c*	-1.61882	1.76E-07	1.56E-05
hsa-miR-193a-3p	-1.57946	3.57E-06	0.000116
hsa-miR-20b	-1.4071	5.51E-05	0.000868
hsa-miR-193b	-1.39806	6.12E-05	0.000929

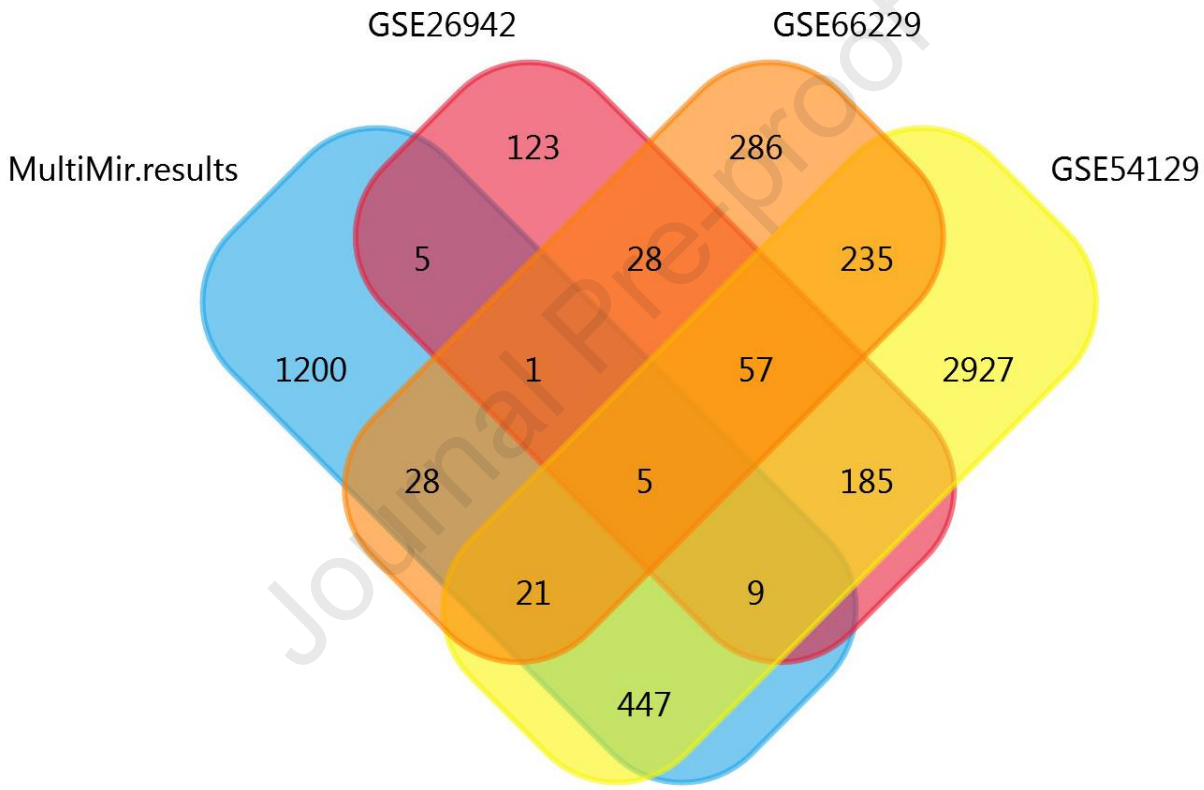
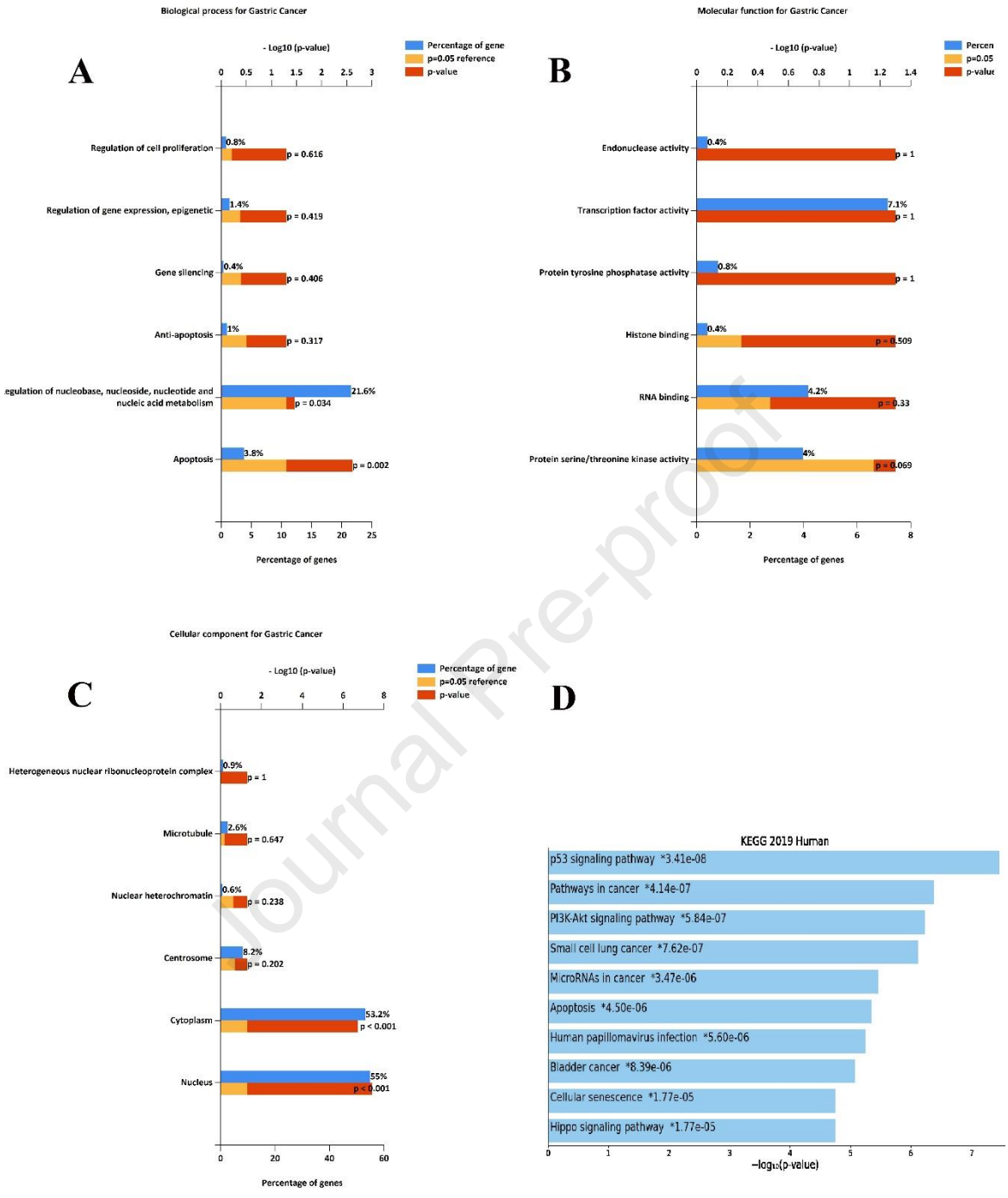


Figure1. Venn diagram analysis showing overlap of dysregulated genes among three datasets (GSE26942, GSE66229, and GSE54129) and predicted targets of miRNAs (multiMir.results). Different colors meant different datasets.

3.2. Functional and pathway enrichment analysis for identified target genes

To further understand the function and mechanism, DEGs are imported into the online enrichment analysis tool, FunRich to identify GO analysis and KEGG pathway in GC. In the biological process (BP) term of GO analysis, the results indicated that genes were significantly enriched in apoptosis and regulation of nucleobase, nucleoside, nucleotide, and nucleic acid metabolism (figure2A). Regarding cellular component (CC) term, the DEGs were mainly involved in nucleus and cytoplasm (figure2B). In addition, cell component analysis showed that DEGs were enriched in protein serine/threonine kinase activity (figure2C). KEGG pathway analysis demonstrated that the DEGs were primarily related to P53 signaling pathway, pathways in cancer, PI3K-AKT signaling pathway, small cell lung cancer, MicroRNAs in cancer, and apoptosis (figure2D).



189

190 Figure2. GO And KEGG Analyses of overlapping DEGs in GC and predicted targets for
191 DEmiRs: (A) biological process, (B) molecular function, (C) cellular component, and (D) KEGG
192 pathway analysis.

3.3. PPI network of DEGs and module analysis

The STRING database was adopted to determine the PPI pairs among the 516 overlapping DEGs, and the PPI network was constructed and visualized by the Cytoscape software. Cytoscape MCODE was applied to screen modules within the PPI network. A significant module was obtained from the MCODE plug-in in Cytoscape, which contained 28 nodes and 162 edges (figure3). The Biological pathway analysis revealed that these genes were significantly enriched in Trail signaling, Insulin pathway, class 1 PI3K signaling, Arf6 signaling events, and EGF receptor signaling.

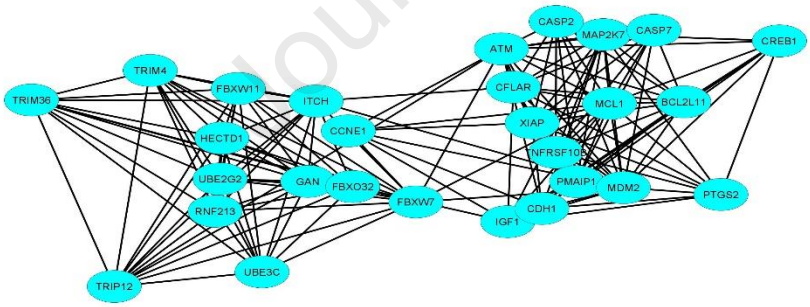


Figure3. Protein–protein interaction (PPI) network. The PPI network consisted of 28 nodes and 162 edges were constructed by overlapping DEGs.

3.4.The five key genes validation

The prognostic values of *CEMIP*, *CLDN1*, *SERPINE1*, *PMEPA1*, and *LIFR* were obtained from KM plotter (figure4). The curves showed that overexpression of the five genes are significantly related to decreased overall survival times of GC patients. Moreover, Gene expression validations were performed using GEPIA (figure5). Results indicated that mRNA expression levels of *CEMIP*, *CLDN1*, *SERPINE1*, and *PMEPA1* were significantly upregulated in GC tissues compared to those in non- cancerous gastric tissues, while *LIFR* downregulated in GC samples compared to normal samples. For expression information of key genes see Supplementary Figure S2

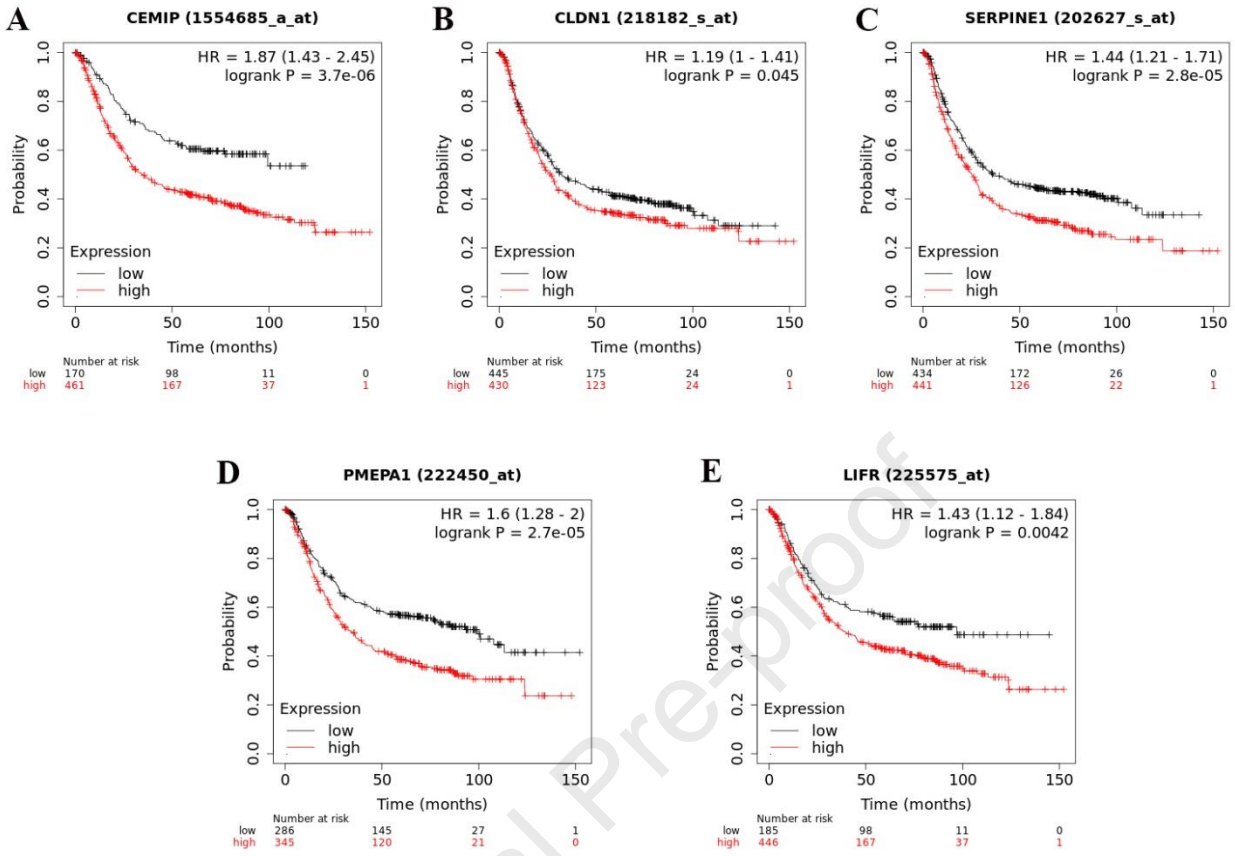


Figure4. Kaplan-Meier overall survival analyses of patients with GC based on expression of the five key genes. (A) CEMIP, (B) CLDN1, (C) SERPINE1, (D) PMEPA1, (E) LIFR.

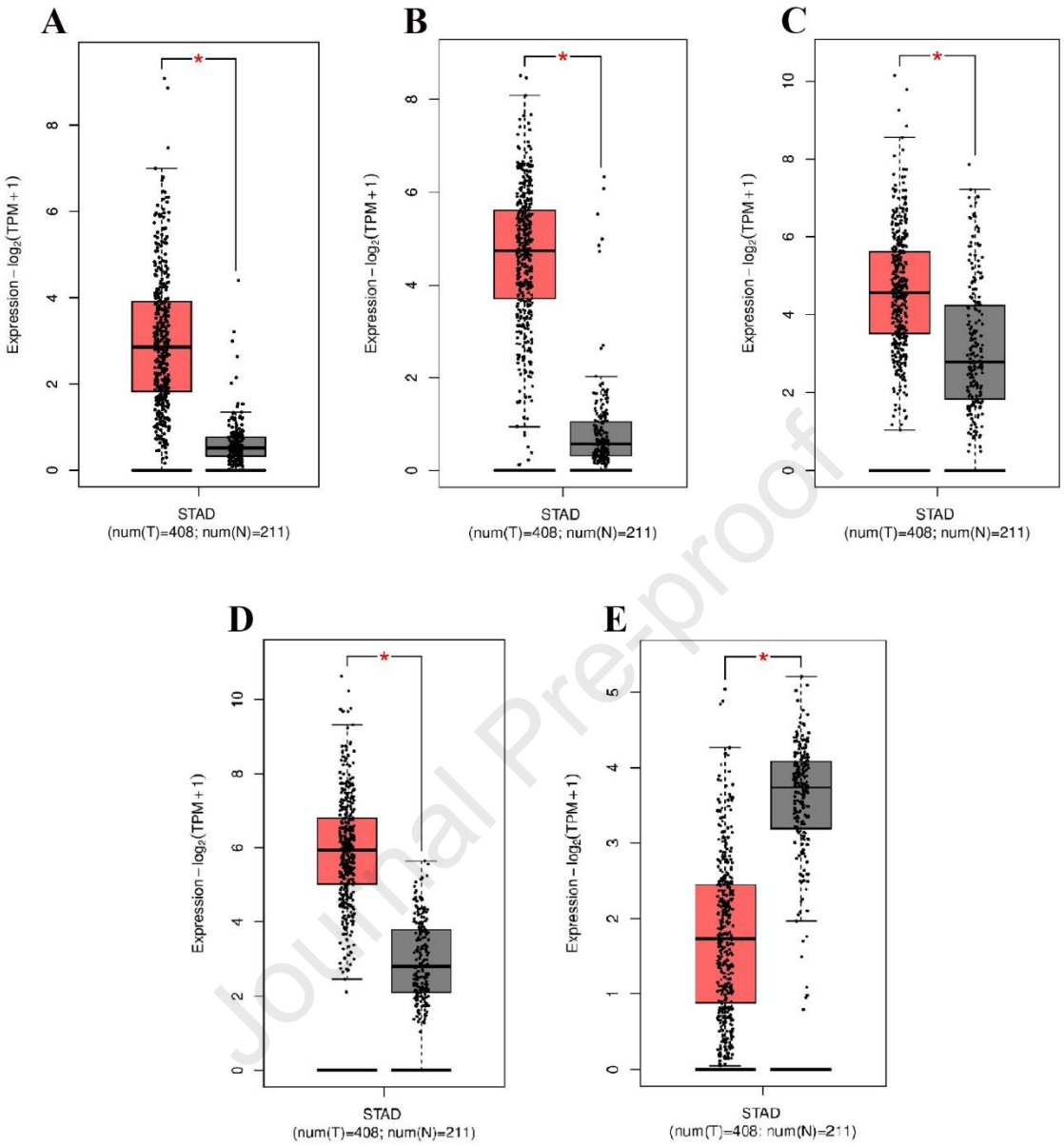


Figure5. Validation of the mRNA expression levels CEMIP (A), CLDN1(B), SERPINE1(C), PMEPA1(D) , and LIFR (E).in GC and gastric brain tissues using GEPIA. These five box plots are based on 408 GC samples (marked in red) and 211 normal samples (marked in gray). *P<0.05 was considered statistically significant.

3.5. Construction of predicted miRNAs–mRNAs network

Based on the predicted miRNA-mRNA relationship of 12 DEmiRs, DEmiR-mRNA regulatory network was obtained. The DEmiR-mRNA regulation network was shown in Supplementary. The top 4 miRNAs with higher degrees included hsa-miR-421 (up-regulated, degree = 236), hsa-miR-193a-3p (down-regulated, degree = 125), hsa-miR-576-5p (up-regulated, degree = 101), and hsa-miR-1246 (up-regulated, degree = 44). The top 6 genes with higher degrees included *CTC1*, *RGMB*, *E2F6*, *IGF1*, *JARID2*, and *PHKA1*.

4. Discussion

In this study, mRNA and miRNA expression profiles were integrated to evaluate changes of genes (DEGs) and miRNA (DEmir) expression in GC. A total of 15 DEmiRs (9 up- and 6 down-regulated miRNAs) and 516 DEGs were found by analyzing four gene expression profiles containing a combined 676 GC tissue samples and 141 normal gastric tissue samples. The results of functional enrichment analyses of the DEGs revealed that the genes enriched in a number of biological processes, such as apoptosis and regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism. It has been revealed that de novo nucleotide synthesis, which is essential for cancer cell proliferation, is directly regulated by tumor suppressors and oncogenes [26-29]. KEGG pathway analysis demonstrated that the DEGs were involved in P53 signaling pathway, PI3K-AKT signaling pathway, small cell lung cancer, MicroRNAs in cancer, and apoptosis. P53 is a tumor suppressor gene and serves as a cellular stress sentinel for DNA damage and other cellular stresses[30]. TP53 mutations increase with the progression of GC from normal gastric mucosa[31, 32]. Phosphoinositide 3-kinase (PI3K)/protein kinase B (AKT) signaling pathway is one of the key signaling pathways in the formation and progression of many cancers[33]. Researchers have demonstrated the promoting effects of the PI3K/Akt/mTOR pathway in cell growth, metastasis, resistance to chemotherapy metabolism, and survival[34].

More interestingly, overexpression of PIK3CA could enhance the metastasis of gastric carcinoma through aberrant activation of PI3K1Akt signaling[35]. Similarly, targeted blockade of this pathway may inhibit gastric cancer growth and metastasis through downregulating the expression of *MMP-2* and *Ki-67*[34]. *CEMIP*, *CLDN1*, *SERPINE1*, *PMEPA1*, and *LIFR* were common genes among DEGs of threes datasets and predicted targets of DEmiRs and are regulated by hsa-miR-421 and hsa-miR-193a-3p. Subsequently, survival analysis of the relationship between the expression of the five genes and postoperative survival of patients indicated that these genes were significantly correlated with the overall survival of patients with GC.

Cell migration inducing protein currently (CEMIP) is a Wnt-related protein, enriched in lung tumor-derived exosomes, breast, and exosomes brain metastatic, promotes BrM by generating a pro-metastatic environment[36]. Overexpression of CEMIP is related to uncontrolled proliferation and invasion of the tumor with distant metastasis, dedifferentiation, and lower survival of cancer patients. Up-regulation of CEMIP also protect the malignant tumor from the strict microenvironment in low glucose and hypoxia[37]. Over-expression of CEMIP has been reported in various cancer cells, such as gastric, lung, cervix, kidney, and colorectal cancer[38, 39]. CLDN1 is one of the integral membrane proteins essential for the maintenance of normal epithelium, particularly barrier formation, signal transduction, and cell polarity[40]. Down-regulation of CLDN1 could lead to the destruction of tight junctions and loss of cell-to-cell adhesion correlated with the development of the neoplastic phenotype in epithelial cells[41, 42]. *Singh et al* reported that CLDN1 protected colon cancer cells from anoikis, a form of apoptosis happening when cells detach from the extracellular matrix (ECM)[43]. Anoikis is a crucial mechanism in the maintenance of tissue development and homeostasis. CLDN1 has dual role as

oncogene and tumor suppressor, as well as it is a negative and positive prognostic factor in various cancers including gastric, colon, lung, breast, and ovarian [44-49]. Some investigations on colon and ovarian cancer have reported a role of CLDN1 on metastatic processes through activation of metalloproteinases, increasing migration, and reducing apoptosis. The elevated expression of CLDN1 in gastric cancer is associated with metastasis, tumor invasion, poor outcome, lymph node metastasis, and TNM stage [42, 50, 51]. SERPINE1 is a key regulator of the uPA system through inhibiting urokinase plasminogen activator (uPA) and principal inhibitor of tissue plasminogen activator (tPA)[52]. SERPINE1 plays a crucial role in different types of tumors not only as an oncogene but also serve as a new prognostic factor in certain cancers, including bladder cancer, oesophageal cancer, human melanoma, cell lung cancer, oral squamous cell carcinoma, and head and neck cancer[53-59]. It has also been indicated that down-regulation of SERPINE1 has a tumor-suppressive role in the phenotype of glioma tumor cells by activating p53 signaling pathway and inhibited the nasopharyngeal carcinoma migration and cell invasion in vitro[60, 61]. Upregulation of SERPINE1 has been shown in GC tissues compared with normal tissues, and overexpression of SERPINE1 is significantly associated with poor prognosis and unfavorable clinical features in patients with GC[62]. PMEPA1 is a type Ib transmembrane protein and involves in the transforming growth factor beta (TGF-β) signaling pathway. The TGF- β is a crucial regulator of homeostasis and suppresses tumor progression at the early stage of tumorigenesis[63]. TMEPAI protein was reported to regulate differentiation of epithelial tissues and cell proliferation, suggesting its function in the development of malignant tumors. Beside, a significant upregulation of PMEPA1 has been identified in malignant tissues of GC patients, and its higher expression was associated with poor prognosis[64, 65]. Leukemia inhibitory factor (LIF) is a type of cytokine which involves in various diseases, including cancer,

carcinogens, differentiation and Regulates cell proliferation[66]. LIF and LIFR expression are correlated with tumor differentiation, tumor stage, lymphovascular invasion, pTNM stage, lymph node, and metastasis in GC cells[67]. It has been identified that hsa-miR-421, which targets CREBZF, could play an important role in the development of GC and knock-down of this miRNA leads to an increased expression of CREBZF expression in GC[68]. An investigation Human Endothelial Cells revealed that SERPINE1 is a target gene of miR-421[69]. Dysregulation of miR-193a family in numerous malignancies has been reported and increasing evidence has been shown their pivotal roles in cancer pathways [70-72]. Several studies previously revealed that miR-193a-3p is a neoplasm suppressor in different cancers, including thyroid cancer, breast cancer, lung cancer, hepatocellular cancer, and colorectal cancer [73-78]. Furthermore, studies indicated that the expression levels of miR-193a-5p was significantly decreased in GC compared to adjacent normal tissue [79, 80].

The findings of the miRNAs–mRNAs network revealed a high degree of hsa-miR-421, hsa-miR-193a-3p, hsa-miR-576-5p, and hsa-miR-1246, as well as *CTC1*, *RGMB*, *E2F6*, *IGF1*, *JARID2*, and *PHKA1* were genes with the highest degree of connectivity, indicating that these miRNAs and mRNAs might play key roles in the development of GC. It has been suggested that hsa-miR-1246, which is upregulated in a human gastric cancer cell line, may play important roles in the progression of GC, and exosomal miR1246 in serum could serve as a biomarker for the early diagnosis of GC[81, 82]. The E2F family of transcription factors regulate the expression of genes in various cellular processes such as, control of cell cycle, DNA damage response, differentiation, and apoptosis [83, 84]. The expression of E2F6, a member of E2F family, was significantly correlated with favorable overall survival of male patients and could be applied as novel prognostic markers to improve the survival rate and prognostic accuracy in GC[85, 86].

Insulin-like growth factors (IGFs) can stimulate differentiation and cellular proliferation and have pathogenic roles in cancer [87-89]. Specifically, *Li et al* reported a Significant increased levels of serum IGF1 in GC patients[90]. Nevertheless, investigations on the regulatory mechanism and prognostic value of hsa-miR-576-5p, *CTC1*, *RGMB*, *JARID2*, and *PHKA1* in GC have seldom been reported. The present study has the following limitations that should be noticed in future studies. Firstly, lack of experimental and clinical validation. Secondly, considering that we utilized available online tools with default options in several steps of the project, investigation of the expression level of identified key genes and miRNAs of GC in different contexts such as gender, age, tumor stage, and smoking habit was not applicable. Moreover, DEG limitation were logFC>1 & LogFC<1 & PValue<0.01, DEM limitation were logFC>1 & LogFC<1 & PValue<0.0001.

5. Conclusion:

In the present study, we identified several genes and miRNAs that closely associated with GC occurrence and development, including *CEMIP*, *CLDN1*, *SERPINE1*, *PMEPA1*, *LIFR*, hsa-miR-193a-3p, and hsa-miR-421. Moreover, further studies are required to assess the effects of hsa-miR-576-5p, *CTC1*, *RGMB*, *JARID2*, and *PHKA1* on incidence of GC and improve the reliability and reproducibility of our results. The results provide important information about the critical roles of these genes in GC initiation and progression, which could be used for the diagnosis and treatment of GC patients..

Conflict of interest:

The authors declare that they have no conflict of interest to disclose.

References:

1. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* 2018. **68**(6): p. 394-424.
2. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* 2021. **71**(3): p. 209-249.
3. Mokadem, I., et al., *Recurrence after preoperative chemotherapy and surgery for gastric adenocarcinoma: a multicenter study.* 2019. **22**(6): p. 1263-1273.
4. Nashimoto, A., et al., *Gastric cancer treated in 2002 in Japan: 2009 annual report of the JGCA nationwide registry.* 2013. **16**(1): p. 1-27.
5. Okines, A., et al., *Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* 2010. **21**: p. v50-v54.
6. Li, T., et al., *Identification of hub genes with prognostic values in gastric cancer by bioinformatics analysis.* 2018. **16**(1): p. 1-12.
7. Wang, K., et al., *Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer.* 2014. **46**(6): p. 573-582.
8. Jiang, P. and X.S.J.N.g. Liu, *Big data mining yields novel insights on cancer.* 2015. **47**(2): p. 103-104.
9. Akbani, R., et al., *A pan-cancer proteomic perspective on The Cancer Genome Atlas.* 2014. **5**(1): p. 1-15.
10. Yong, L., Z. YuFeng, and B.J.I.J.o.S. Guang, *Association between PPP2CA expression and colorectal cancer prognosis tumor marker prognostic study.* 2018. **59**: p. 80-89.
11. Wang, F., et al., *Identifying the hub gene in gastric cancer by bioinformatics analysis and in vitro experiments.* 2020. **19**(11): p. 1326-1337.
12. Liu, Z., et al., *Identification and Analysis of Key Genes Driving Gastric Cancer Through Bioinformatics.* 2021. **25**(1): p. 1-11.
13. Nilsen, T.W.J.T.i.G., *Mechanisms of microRNA-mediated gene regulation in animal cells.* 2007. **23**(5): p. 243-249.
14. Bartel, D.P.J.c., *MicroRNAs: genomics, biogenesis, mechanism, and function.* 2004. **116**(2): p. 281-297.
15. Hwang, J., et al., *MicroRNA expression profiles in gastric carcinogenesis.* 2018. **8**(1): p. 1-8.
16. Zhu, M., et al., *Similarly up-regulated microRNA-106a in matched formalin-fixed paraffin-embedded and fresh frozen samples and the dynamic changes during gastric carcinogenesis and development.* 2014. **210**(12): p. 909-915.
17. Edgar, R., M. Domrachev, and A.E.J.N.a.r. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* 2002. **30**(1): p. 207-210.
18. Barrett, T., et al., *NCBI GEO: archive for high-throughput functional genomic data.* 2009. **37**(suppl_1): p. D885-D890.

19. Ru, Y., et al., *The multiMiR R package and database: integration of microRNA–target interactions along with their disease and drug associations.* 2014. **42**(17): p. e133-e133.
20. Pathan, M., et al., *FunRich: An open access standalone functional enrichment and interaction network analysis tool.* 2015. **15**(15): p. 2597-2601.
21. Szklarczyk, D., et al., *STRING v10: protein–protein interaction networks, integrated over the tree of life.* 2015. **43**(D1): p. D447-D452.
22. Doncheva, N.T., et al., *Cytoscape StringApp: network analysis and visualization of proteomics data.* 2018. **18**(2): p. 623-632.
23. Chin, C.-H., et al., *cytoHubba: identifying hub objects and sub-networks from complex interactome.* 2014. **8**(4): p. 1-7.
24. Nagy, Á., et al., *Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets.* 2018. **8**(1): p. 1-9.
25. Tang, Z., et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses.* 2017. **45**(W1): p. W98-W102.
26. Graves, L.M., et al., *Regulation of carbamoyl phosphate synthetase by MAP kinase.* 2000. **403**(6767): p. 328-332.
27. Ben-Sahra, I., et al., *Stimulation of de novo pyrimidine synthesis by growth signaling through mTOR and S6K1.* 2013. **339**(6125): p. 1323-1328.
28. Saha, A., et al., *Akt phosphorylation and regulation of transketolase is a nodal point for amino acid control of purine synthesis.* 2014. **55**(2): p. 264-276.
29. Qian, X., et al., *Conversion of PRPS hexamer to monomer by AMPK-mediated phosphorylation inhibits nucleotide synthesis in response to energy stress.* 2018. **8**(1): p. 94-107.
30. Stegh, A.H.J.E.o.o.t.t., *Targeting the p53 signaling pathway in cancer therapy–the promises, challenges and perils.* 2012. **16**(1): p. 67-83.
31. Shiao, Y.-H., et al., *Genetic and immunohistochemical analyses of p53 independently predict regional metastasis of gastric cancers.* 2000. **9**(6): p. 631-633.
32. Gleeson, C., et al., *Comparison of p53 and DNA content abnormalities in adenocarcinoma of the oesophagus and gastric cardia.* 1998. **77**(2): p. 277-286.
33. Vara, J.Á.F., et al., *PI3K/Akt signalling pathway and cancer.* 2004. **30**(2): p. 193-204.
34. Willems, L., et al., *PI3K and mTOR signaling pathways in cancer: new data on targeted therapies.* 2012. **14**(2): p. 129-138.
35. Liu, J.-F., et al., *Up-regulation of PIK3CA promotes metastasis in gastric carcinoma.* 2010. **16**(39): p. 4986.
36. Rodrigues, G., et al., *Tumour exosomal CEMIP protein promotes cancer cell colonization in brain metastasis.* 2019. **21**(11): p. 1403-1412.
37. Li, L., et al., *Central role of CEMIP in tumorigenesis and its potential as therapeutic target.* 2017. **8**(12): p. 2238.
38. Birkenkamp-Demtroder, K., et al., *Repression of KIAA1199 attenuates Wnt-signalling and decreases the proliferation of colon cancer cells.* 2011. **105**(4): p. 552-561.
39. Terashima, M., et al., *KIAA1199 interacts with glycogen phosphorylase kinase β-subunit (PHKB) to promote glycogen breakdown and cancer cell survival.* 2014. **5**(16): p. 7040.
40. Tsukita, S. and M.J.C.o.i.c.b. Furuse, *Claudin-based barrier in simple and stratified cellular sheets.* 2002. **14**(5): p. 531-536.
41. Tobioka, H., et al., *Occludin expression decreases with the progression of human endometrial carcinoma.* 2004. **35**(2): p. 159-164.
42. Jung, H., et al., *The expression of claudin-1, claudin-2, claudin-3, and claudin-4 in gastric cancer tissue.* 2011. **167**(2): p. e185-e191.

43. Singh, A.B., A. Sharma, and P.J.C. Dhawan, *Claudin-1 expression confers resistance to anoikis in colon cancer cells in a Src-dependent manner.* 2012. **33**(12): p. 2538-2547.
44. Myal, Y., et al., *Claudin 1 in breast tumorigenesis: revelation of a possible novel "claudin high" subset of breast cancers.* 2010. **2010**.
45. Blanchard, A.A., et al., *Claudin 1 expression in basal-like breast cancer is related to patient age.* 2013. **13**(1): p. 1-13.
46. Chang, T.L., et al., *Claudin-1 has tumor suppressive activity and is a direct target of RUNX3 in gastric epithelial cells.* 2010. **138**(1): p. 255-265. e3.
47. Qin, W., et al., *MicroRNA-155 is a novel suppressor of ovarian cancer-initiating cells that targets CLDN1.* 2013. **587**(9): p. 1434-1439.
48. Chao, Y.-C., et al., *Claudin-1 is a metastasis suppressor and correlates with clinical outcome in lung adenocarcinoma.* 2009. **179**(2): p. 123-133.
49. Dhawan, P., et al., *Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer.* 2005. **115**(7): p. 1765-1776.
50. Wu, Y.-L., et al., *Expression transformation of claudin-1 in the process of gastric adenocarcinoma invasion.* 2008. **14**(31): p. 4943.
51. Huang, J., et al., *The expression of claudin 1 correlates with β-catenin and is a prognostic factor of poor outcome in gastric cancer.* 2014. **44**(4): p. 1293-1301.
52. Dellas, C., D.J.J.T. Loskutoff, and haemostasis, *Historical analysis of PAI-1 from its discovery to its potential role in cell motility and disease.* 2005. **93**(04): p. 631-640.
53. Pavón, M.A., et al., *uPA/uPAR and SERPINE1 in head and neck cancer: role in tumor resistance, metastasis, prognosis and therapy.* 2016. **7**(35): p. 57351.
54. Dhanda, J., et al., *SERPINE1 and SMA expression at the invasive front predict extracapsular spread and survival in oral squamous cell carcinoma.* 2014. **111**(11): p. 2114-2121.
55. Klein, R.M., et al., *SERPINE 1 expression discriminates site-specific metastasis in human melanoma.* 2012. **21**(7): p. 551-554.
56. Klimczak-Bitner, A.A., et al., *Expression of MMP9, SERPINE1 and miR-134 as prognostic factors in esophageal cancer.* 2016. **12**(5): p. 4133-4138.
57. Zhang, G., et al., *Validation and clinicopathologic associations of a urine-based bladder cancer biomarker signature.* 2014. **9**(1): p. 1-10.
58. Li, X., et al., *Overexpression of CEP72 promotes bladder urothelial carcinoma cell aggressiveness via epigenetic CREB-mediated induction of SERPINE1.* 2019. **189**(6): p. 1284-1297.
59. Lin, X., et al., *PAI-1/PIAS3/Stat3/miR-34a forms a positive feedback loop to promote EMT-mediated metastasis through Stat3 signaling in Non-small cell lung cancer.* 2017. **493**(4): p. 1464-1470.
60. Wu, D.M., et al., *MircoRNA-1275 promotes proliferation, invasion and migration of glioma cells via SERPINE1.* 2018. **22**(10): p. 4963-4974.
61. Sang, Y., et al., *TEL2 suppresses metastasis by down-regulating SERPINE1 in nasopharyngeal carcinoma.* 2015. **6**(30): p. 29240.
62. Liao, P., et al., *Genome-scale analysis identifies SERPINE1 and SPARC as diagnostic and prognostic biomarkers in gastric cancer.* 2018. **11**: p. 6969.
63. Watanabe, Y., et al., *TMEPAI, a transmembrane TGF-β-inducible protein, sequesters Smad proteins from active participation in TGF-β signaling.* 2010. **37**(1): p. 123-134.
64. Karbyshev, M.S., et al., *Development of novel monoclonal antibodies for evaluation of transmembrane prostate androgen-induced Protein 1 (TMEPAI) expression patterns in gastric cancer.* 2018. **24**(2): p. 427-438.
65. Zhang, L., et al., *PMEPA1 induces EMT via a non-canonical TGF-β signalling in colorectal cancer.* 2019. **23**(5): p. 3603-3615.

66. Pinho, V., et al., *Leukemia inhibitory factor: Recent advances and implications in biotechnology.* 2020. **52**: p. 25-33.
67. Bian, S.B., et al., *Leukemia inhibitory factor promotes gastric cancer cell proliferation, migration, and invasion via the LIFR–Hippo–YAP pathway.* 2020.
68. Kim, Y.J., et al., *Potential miRNA-target interactions for the screening of gastric carcinoma development in gastric adenoma/dysplasia.* 2018. **15**(6): p. 610.
69. Marchand, A., et al., *miR-421 and miR-30c inhibit SERPINE 1 gene expression in human endothelial cells.* 2012. **7**(8): p. e44532.
70. Shirafkan, N., et al., *microRNA-193a-5p inhibits migration of human HT-29 colon cancer cells via suppression of metastasis pathway.* 2019. **120**(5): p. 8775-8783.
71. Xie, Z.C., et al., *A meta-analysis and bioinformatics exploration of the diagnostic value and molecular mechanism of miR-193a-5p in lung cancer.* 2018. **16**(4): p. 4114-4128.
72. Wang, S., Y. Diao, and B.J.N. Zhu, *MiR-193a-5p suppresses cell proliferation and induces cell apoptosis by regulating HOXA7 in human ovarian cancer.* 2020. **506**(3): p. 169-162.
73. Liu, Y., et al., *Down-regulation of MiR-193a-3p dictates deterioration of HCC: a clinical real-time qRT-PCR study.* 2015. **21**: p. 2352.
74. Deng, W., et al., *Quantitative proteomic analysis of the metastasis-inhibitory mechanism of miR-193a-3p in non-small cell lung cancer.* 2015. **35**(5): p. 1677-1688.
75. Yu, M., et al., *PTP 1B markedly promotes breast cancer progression and is regulated by miR-193a-3p.* 2019. **286**(6): p. 1136-1153.
76. Takahashi, H., et al., *microRNA-193a-3p is specifically down-regulated and acts as a tumor suppressor in BRAF-mutated colorectal cancer.* 2017. **17**(1): p. 1-14.
77. Mamoori, A., et al., *Clinical and biological significance of miR-193a-3p targeted KRAS in colorectal cancer pathogenesis.* 2018. **71**: p. 145-156.
78. Santarpia, L., et al., *A miRNA signature associated with human metastatic medullary thyroid carcinoma.* 2013. **20**(6): p. 809-23.
79. Chou, N.-H., et al., *MiR-193a-5p and-3p play a distinct role in gastric cancer: miR-193a-3p suppresses gastric cancer cell growth by targeting ETS1 and CCND1.* 2018. **38**(6): p. 3309-3318.
80. Huang, Y., et al., *LINC00152 down-regulated miR-193a-3p to enhance MCL1 expression and promote gastric cancer cells proliferation.* 2018. **38**(3).
81. Shi, Y., et al., *Exosomal miR-1246 in serum as a potential biomarker for early diagnosis of gastric cancer.* 2020. **25**(1): p. 89-99.
82. Yin, C., et al., *Differential expression profile analysis of cisplatin-regulated miRNAs in a human gastric cancer cell line.* 2019. **20**(2): p. 1966-1976.
83. Trimarchi, J.M. and J.A.J.N.r.M.c.b. Lees, *Sibling rivalry in the E2F family.* 2002. **3**(1): p. 11-20.
84. Chen, H.-Z., S.-Y. Tsai, and G.J.N.R.C. Leone, *Emerging roles of E2Fs in cancer: an exit from cell cycle control.* 2009. **9**(11): p. 785-797.
85. Manicum, T., et al., *Prognostic values of E2F mRNA expression in human gastric cancer.* 2018. **38**(6).
86. Liu, X., C.J.M.T.-M. Hu, and C. Development, *Novel Potential Therapeutic Target for E2F1 and Prognostic Factors of E2F1/2/3/5/7/8 in Human Gastric Cancer.* 2020. **18**: p. 824-838.
87. Sara, V.R. and K.J.P.R. Hall, *Insulin-like growth factors and their binding proteins.* 1990. **70**(3): p. 591-614.
88. Tian, Z., et al., *IGF2R expression is associated with the chemotherapy response and prognosis of patients with advanced NSCLC.* 2014. **34**(5): p. 1578-1588.
89. Chen, J., et al., *Estrogen receptor beta-mediated proliferative inhibition and apoptosis in human breast cancer by calycosin and formononetin.* 2013. **32**(6): p. 1790-1797.

90. Li, S., X. Lei, and J. Zhang, *Insulin-like growth factor 1 promotes growth of gastric cancer by inhibiting foxo1 nuclear retention (Retraction of Vol 36, Pg 4519, 2015)*. 2017, SAGE PUBLICATIONS LTD 1 OLIVERS YARD, 55 CITY ROAD, LONDON EC1Y 1SP, ENGLAND.

Supplementary figures

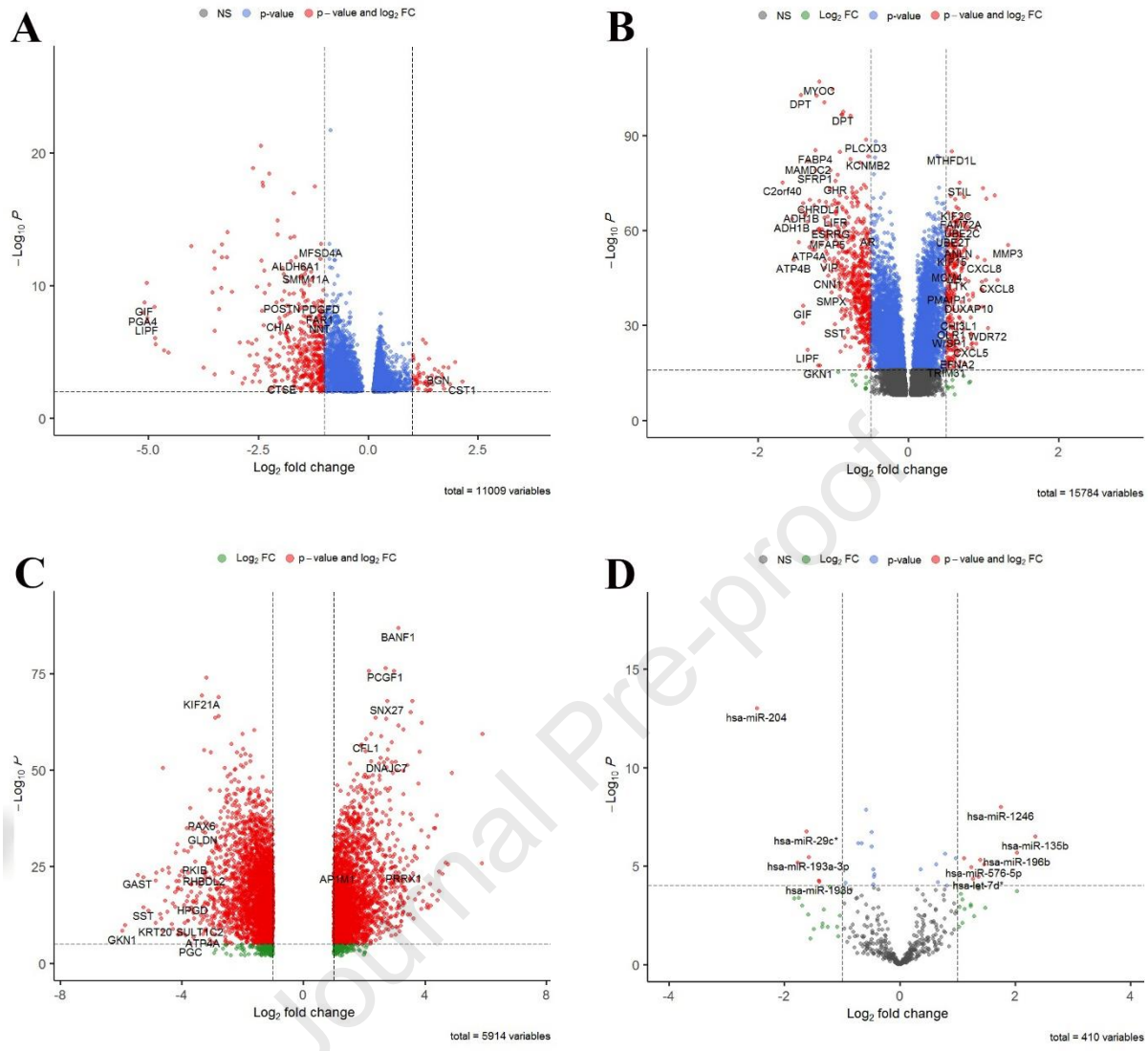
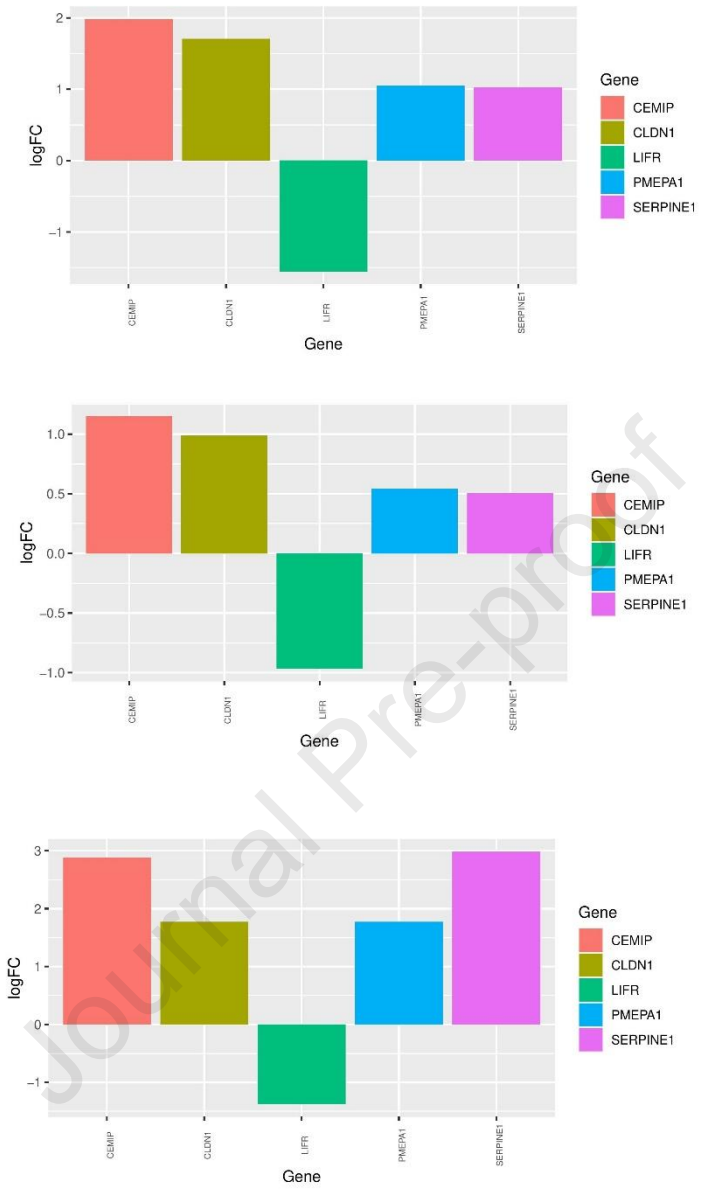


Figure S1. Volcano plot of DEGs related to GSE26942 (A), GSE66229 (B), GSE54129(C), and DEmiRs GSE26595 (D).



543

544 **Figure S2.** Expression information of key genes related to GSE26942 (A), GSE66229 (B), and

545 GSE54129(C)

546

Acknowledgments

This manuscript has not been published previously nor is it under consideration for publication elsewhere. The authors are grateful to those who have participated in this research work.

Conflict of interest:

The authors declare that they have no conflict of interest to disclose.