Identification of Potential Biomarkers and Therapy Targets involved in Gastric Cancer Using Bioinformatics Analysis

YangYang Teng

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Na Shan

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

GuangRong Lu

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

LeYi Ni

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

ZeJun Gao

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Yao Shen

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Sheng Xu

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

QinJian Wang

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

QingJi Ying

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Jing Zhang

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

HeJie Yu

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

ZhanXiong Xue

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Chao Xing

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

ZhenZhai Cai (✉ czz77@sina.com)

Wenzhou Medical College Second Affiliated Hospital: The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Research

Keywords: gastric cancer, Bioinformatics, WGCNA, GEO, TCGA, biomarker

Posted Date: August 31st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-837643/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Gastric cancer remains one of the five major malignant tumors in the world, posing a great threat to public life and health. As gene sequencing technology develops, it is urgent to find out specific molecular markers for cancer therapy. In this study, datasets of GSE13911, GSE30727, GSE63089 and GSE118916 were investigated by bioinformatics analysis, and differentially expressed genes (DEGs) between GC tissues and normal tissues were screened for potential cancer therapeutic targets.

Furthermore, the GSE63089 dataset was analyzed by Weighted Gene Co-expression Network Analysis (WGCNA), and the highly related genes were clustered. Then, the hub genes were searched using co-expression network and Molecular Complex Detection (MCODE) plug-in from Cytoscape software. Finally, ASPM, COL11A1 and CDC20 were obtained by intersection of hub genes and DEGs. The expressions of ASPM, COL11A1 and CDC20 gene in gastric cancer tissues and normal tissues from TCGA database were detected. For these genes, the least absolute shrink and selection operator (LASSO) Cox expression analysis was used to establish the prognostic risk model. COL11A1 and CDC20 genes were identified as candidate prognostic risk markers for GC. Analysis using qRT-PCR has shown that COL11A1 and CDC20 were significant differentially expressed between gastric cancer tissues and normal gastric tissues ($P < 0.01$). In conclusion, our study identifies specific DEGs involved in ECM process and metabolism by cytochrome P450 process, and these DEGs may be potential targets for GC therapy. The model constructed by COL11A1 and CDC20 genes can predict the prognosis risk of GC patients. Taken together, these findings provide reference for further analyses of key alterations during GC progression.

Introduction

Gastric cancer (GC) is one of the five major malignant tumors worldwide, with poor survival and high recurrence [1]. The 5-year survival rate of patients being diagnosed at an advanced stage of GC is below 20% [2]. Biomarkers are identified as indicators of normal biologic and pathological processes. Development of biomarkers play an important role in clinical diagnosis and treatment of GC. Previous study has suggested that monoclonal antibodies against HER2 may be used to treat GC with positive HER2 [3]. Trastuzumab, one of the monoclonal antibodies against HER2 protein, has been approved for the treatment of advanced GC patients [4]. In addition, approximately 40% of GC patients overexpress programmed death ligand 1 (PD-L1), indicating that PD-L1 can be designated as a potential biomarker for GC [5]. As a monoclonal antibody against PD-1, pembrolizumab has a certain curative effect on patients with PD-1 positive [6]. Therefore, monitoring the clinical prognosis of high-risk GC patients and discovering biomarkers and therapeutic targets of GC with clinical value are the current research hotspots.

In recent years, bioinformatics analysis of different gene expression databases is a hot topic in cancer research. These tools can identify potential biomarkers and therapeutic targets for certain disease, predict their cellular component, molecular function and pathway annotation, and construct related prognostic risk models. To improve the accuracy of the data in this study, we firstly screened four common mRNA

datasets and obtained 128 potential GC biomarkers composed of differentially expressed genes (DEGs) between GC tissues and adjacent non-cancer tissues. In addition, we analyzed one of the mRNA datasets using Weighted correlation network analysis (WGCNA) to construct the protein-protein interaction (PPI) network, followed by Molecular Complex Detection (MCODE) plug-in to obtain the hub genes. The application of powerful bioinformatics allowed us to identify novel GC biomarkers from co-existing mRNA in DEGs and hub genes. A prognostic risk model composed of COL11A1 and CDC20 gene was finally constructed using The Cancer Genome Atlas (TCGA) database. Finally, the core genes were verified using qRT-PCR. Quantitative RT-PCR results have shown that COL11A1 and CDC20 were significantly different between gastric cancer tissues and normal gastric tissues.

Results

Identification of DEGs

To obtain the DEGs, the following two methods were used to screen gene expression profiles of four GC GEO datasets (Fig. 1C). According to the cut-off thresholds, 1732, 200, 438 and 1397 DEGs were extracted from GSE13911, GSE30727, GSE63089 and GSE118916 datasets, respectively. Totally 45 DEGs were found to be co-expressed in four datasets (Fig. 1A). The results of GO enrichment analysis of these 45 genes were shown in Fig. 1B. Additionally, the RRA method was applied to integrate gene expression profiles of GC in four datasets, and 41 highly expressed genes and 83 low expressed genes were found based on the cut-off thresholds. The Fig. 1D presented the top 20 high expression genes and low expression genes, while the results of GO enrichment analysis of 20 high expression genes and 20 low expression genes were shown in Fig. 1E and 1F. The final DEGs were obtained through the above two methods (Table 2), followed by GO enrichment analysis and Reaction pathway enrichment analysis on these genes (Fig. 1G and 1H).

Table 2. DEGs.

Venn and DEGs tables1 selected by RRA; MCODE App in Cytoscape software was used to process genes obtained by WGCNA cluster analysis so as to obtain hub genes.

Method	Gene MALAT1ATP4BFAPESSRRGKCNJ16CLDN1FUT9KCNE2AQP4FBP2
Venn	COL12A1COL6A3FN1CPA2SULF1SPARCCAPN9LIPFHPGDPLAUSSSTLIFR GCPSCAULT1B1TPX2PIK3C2GCOL1A2UGT2B15GKN1TMED6 MMP12TNFRSF17SLC2A12KRT20PLA2G7C6orf58CLDN4FMO5OLR1 MMP7PLA2G2ALAMC2CXCL10SLC28A2FAPCLDN1SULF1MMP12COL12A1,
RAA	CLDN4SPP1FN1CLDN7PLA2G2ACOL6A3PLAU MMP7PLA2G7BUB1TPX2 ASPMCOL5A2CST1SPARCTIMP1CXCL9COL1A2IFITM1FNDC1LIFSERPINH1 KIF20ACOL11A1ANLN SERPINB5CHI3L1SFRP4CDC20MAD2L1THBS2ASPN UBE2TPI15TOP2ATRIM29ATP4BKNE2ESSRRGSSSTTMED6AQP4GIFGKN1 PSCA CPA2HPGDATP4AGCKCNJ16CAPN9KRT20LIPFSLC26A7FBP2ADH1C CHGAATFF2PIK3C2GC6ORF58GKN2AZGP1SLC28A2CKMT2CHIAFCGBP UGT2B15AADACALDOB PBLDCXCL17CYP2C18LIFRPGCREG1ASLC2A12 AKR7A3CKBFGD4VSIG2DUSP19RNASE1FAM3BPTPRZ1ANXA10CA2VSIG1 KCNJ13FMO5CYP3A5SOSTDC1MAP7D2CLIC6SULT1B1CAPN13MALADH7 HOMER2KLK11PDIA2NRG4SELENBP1ALDH1A1CYP2C8PDILTGHRLTRIM50 SLC26A9FUT9GCNT2ALDH3A1SH3BGRL2AKR1C1CWH43PRKACBSTX12 SIDT2RASSF6SYTL2

Enriched reaction pathway is mainly composed of Gastric acid secretion, Retinal metadata, Drug metabolism-cytochrome P450, Metabolism of xenobiology by cytochrome P450, Chemical carcinogenesis, and ECM receptor interaction etc. In GO biological process, DEGs were mainly related to digestion, extractive matrix (ECM) organization, extractive matrix organization, and xenobiological metallic process. In GO cellular component, these DEGs were enriched in collagen-containing ECM, endoplasmic reticulum lumen, and basolateral plasma membrane. In the GO molecular function, DGEs were related to ECM structural constituent, ECM structural constituent conferring tensile strength and oxidoreductase activity, acting on CH-OH group of donors. These results suggested that these DEGs may be the key genes of GC.

WGCNA and co-expression analysis

The genes in GSE63089 dataset were selected for WGCNA cluster analysis. A total of five color modules were obtained, of which the genes in blue, brown, turquoise and yellow modules were used for the subsequent analysis, while the genes in the gray module were excluded (Fig. 2). The co-expression network of these four modules was constructed by STRING and visualized using Cytoscape (Fig. 3). After that, these DEGs were screened and extracted by MCODE using Cytoscape software (Table 1). The obtained hub genes intersected with DEGs, which indicated that ASPM, COL11A1 and CDC20 genes may be the key genes of GC, and then the following analysis was carried out in TCGA dataset.

Table 1
Details of four data sets in GEO.

GEO	Platform	Normal	Tumor
GSE13911	GPL570	31	38
GSE30727	GPL5188	30	30
GSE63089	GLP5175	45	45
GSE118916	GPL15207	15	15

Expression of key genes in GC dataset and construction of prognosis risk model

The gene expression of ASPM, COL11A1 and CDC20 genes in 381 tumor samples and 32 normal gastric tissue samples were compared, which showed a significant difference in the expression of these genes between GC tissues and normal tissues (Fig. 3A). To predict the impact of these genes on clinical outcomes, the LASSO Cox regression algorithm was applied. COL11A1 and CDC20 genes were selected to construct prognosis model using prognostic RS formula as follows: prognosis model = $3.03995515383844 \times 10^{-5} * \text{COL11A1} - 1.53094355024689 \times 10^{-5} * \text{CDC20}$. The gene symbol in the model represented the gene count of the gene in the sample (Fig. 3B and 3C). To explore the correlations of COL11A1 and CDC20 genes with the prognosis of GC, the patients from TCGA database were divided into low-risk group and high-risk group based on the median RS, suggesting a higher mortality rate in high-risk group (Fig. 3D).

Correlation of RS in clinicopathological features of GC

After screening out COL11A1 and CDC20 genes by prognosis model, we systematically studied the relationship between these genes and the pathological characteristics of GC patients in TCGA dataset, including classification, stage, risk score, T status, M status, N status, age, gender, fustat (Fig. 4A). ROC curve showed that the RS could predict the 5-year survival rate of GC patients (Fig. 4B). The results of Cox regression analysis indicated that the RS could be used as an independent prognostic indicator for GC patients, while univariate and multivariate Cox regression analysis results demonstrated that overall survival (OS) was associated with risk score, gender, age, stage, T status, M status, and N status (Fig. 4C and 4D).

Verification of Core Genes

To verify the authenticity of the 2 core genes, we obtained gastric cancer tissues and normal gastric tissues from 15 patients with gastric cancer for analysis using qRT-PCR. The qRT-PCR results shown in Fig. 5A and Fig. 5B demonstrated that COL11A1 and CDC20 were significantly differentially expressed between gastric tissue and normal gastric tissue ($P < 0.01$).

Discussion

GC, also known as stomach cancer, seriously affects public health and life safety because most of patients admitted to the hospitals are advanced GC [1]. At present, the common treatment for GC includes surgery, chemotherapy, radiotherapy, targeted and immunotherapy, among which surgery is the most effective way to treat GC, but the recurrence rate is high [7]. Although cancer mortality has slowly decreased since 1991, it remains a challenge to explore potential therapeutic targets of GC due to increased drug resistance and heavy economic burden [8]. Therefore, we integrated four sets of microarray datasets from GEO databases and identified a total of 128 DEGs by means of two bioinformatics analysis methods, which mainly involved in ECM process and metabolism by cytochrome P450.

ECM is composed of different macromolecular substances secreted by cells into extracellular matrix, including collagen, proteoglycans/glycosaminoglycans and fibronectins, which forms a three-dimensional network around the cell [9, 10]. ECM also undertakes a variety of biological functions, such as cellular differentiation, cell growth, cell survival, cell migration, homeostasis and morphogenesis [10]. Tumor-related microenvironment is mainly composed of ECM, while the imbalance of ECM can lead to the development of tumor cells [11]. For example, the imbalance of ECM can prevent the behavior regulation of stromal cells, and promote tumor-related angiogenesis and inflammatory reaction, allowing a favorable microenvironment for tumor growth [11]. Therefore, the abnormal ECM dynamics is one of the common clinical causes of cancer [12].

Moreover, cytochrome P450s (P450s) consists of various hemoproteins and belongs to the largest and most functionally versatile superfamily. The functional range of P450 activity is of great significance in both microorganisms and humans [13]. Many studies have found that different members of the P450s family play important roles in activating different carcinogenic compounds [14–16]. P450s can not only promote the metabolism of carcinogens by regulating polycyclic aromatic hydrocarbon metabolism [17–20], but also affect the tumor metastasis and growth-related inflammatory process [21]. Hence, P450s plays an essential role in tumorigenesis, cancer metastasis, chemoprevention and chemotherapy [16]. Most pre-existing GC prognostic models have limitations since there is only one gene or mRNA involved in the model [22]. The clinical outcome of GC is found to be associated with multiple genes. In the current study, overlapped DEGs including ASPM, COL11A1 and CDC20 genes were noticed by using WGCNA and co-expression analysis. These DEGs not only controlled the number of genes in the prognosis model, but also improved the specificity and accuracy of the model. To further explore the relationship of ASPM, COL11A1 and CDC20 with GC, data from TCGA database was used to draw the heatmap of their expression in tumor tissues and normal tissues. The results showed a significant high expression of ASPM, COL11A1 and CDC20 in tumor tissues. After that, LASSO Cox regression analysis was used to screen COL11A1 and CDC20 genes, and constructed a prognostic risk model. According to RS, cancer patients were divided into high-risk group and low-risk group. The results of survival curve demonstrated that the prognosis of these two groups was remarkably different. The results of ROC curve showed that the RS predicted the 5-year survival rate of GC patients, while the multivariate Cox regression analysis

reflected that RS could be used as an independent prognostic indicator. Several studies have shown that COL11A1 gene can promote the cell proliferation, migration and invasion of GC cells [23]. Researchers from a previous study have also found an overexpressed COL11A1 gene in a study of recurrent non-small cell lung cancer, which can promote cell proliferation, migration, invasion and drug resistance [24]. Additionally, COL11A1 gene may also contribute greatly to the occurrence and development of ovarian cancer, breast cancer, pancreatic cancer and esophageal squamous cell carcinoma [25–28]. Moreover, the carcinogenic effect of CDC20 gene can cause ubiquitination and degradation of securin, leading to chromosome segregation and blockage of cell cycle, which means CDC20 may be a promising target for anti-tumor therapy of GC [29, 30]. Another study has reported that the phosphorylation of Mcl-1 can induce CDC20 gene to initiate cell apoptosis [31]. In addition, 2 core genes (COL11A1 and CDC20) were found to be significantly differentially expressed between gastric cancer tissue and normal gastric tissue using qRT-PCR. Therefore, from the clinical and bioinformatics point of view, COL11A1 and CDC20 genes can be used as prognostic risk markers of GC.

Conclusion

This study identifies specific DEGs involved in ECM process and metabolism by cytochrome P450 process, and these DEGs may be potential targets for GC therapy. The model constructed by COL11A1 and CDC20 genes can predict the prognosis risk of GC patients. Taken together, these findings provide reference for further analyses of key alterations during GC progression.

Materials And Methods

The source of gastric cancer related gene expression datasets

Microarray datasets related to GC were downloaded from the publicly Gene Expression Omnibus (GEO) databases (www.ncbi.nlm.nih.gov/geo): GSE13911, GSE30727, GSE63089 and GSE118916. All sample tissues were collected from GC patients in each dataset with sample size greater than or equal to 30 (Table 1). The above four mRNA datasets were used to find co-expressed genes between GC tissues and adjacent tissues. In addition, the RNA-seq transcriptome data of 381 tumor tissues and 32 normal tissues, and the relevant clinical information of 443 tumor patients from stomach adenocarcinoma (STAD) cohort were recorded in database of TCGA (<https://cancergenome.nih.gov/>).

Dataset preprocessing

After downloading the series files of GSE13911, GSE30727, GSE63089 and GSE118916 from GEO database, R software (version 4.0.0, <http://www.r-project.org/>) was applied to annotate the genes. Then, the ‘normalizeBetweenArrays’ function in limma package was carried out to normalize the samples of each files [32], and the results were saved in TXT format. The Venn diagrams were generated through the VennDiagram package, followed by apply the Robust Rank Aggrg (RRA) package to process the files for

the second time. Under the null hypothesis of uncorrelated input, the RRA algorithm detects genes that are ranked consistently better than expected under null hypothesis of uncorrelated inputs and assigns a significance score for each gene [33]. The cutoff criterion of DEGs were adjusted as $P < 0.05$ and $|\log \text{fold change (FC)}| \geq 1$. After obtaining the GC gene expression file from TCGA database, we used R software to annotate the gene again. After screening the related genes, the next step analysis was carried out.

Differentially Expressed Genes (DEGs) enrichment analysis

The clusterProfiler package [34] was used for Gene Ontology (GO; <http://geneontology.org/>) and Kyoto Encyclopedia of Genes and Genomes (KEGG; <https://www.genome.jp/kegg/>) pathway enrichment analyses of DEGs. False Discovery Rate (FDR) and adjusted $P < 0.05$ were considered as statistically significant differences.

Weighted correlation network analysis (WGCNA)

WGCNA in R software package is a comprehensive collection of R functions, which was used to establish co-expression network of highly related gene clusters, allowing the discovery of candidate biomarkers for GC [35]. The GSE63089 dataset was selected for identification of co-expression module because of the large sample size. The 'hclust' function in the fastcluster package was chose to cluster samples and remove abnormal samples (Fig. 1). Soft threshing power β was calculated by clicking 'westhold' R function (Fig. 2). The relevant parameters of the identification module were power = 7, min module size = 30, and merge cut height = 0.6.

PPI network construction and MCODE analysis

The Search Tool for the Retrieval of Interacting Genes (STRING, <https://string-db.org/>) database was used to determine the relationship between DEGs and form an interaction network [36]. The validated interactions with cut-off scores of 0.9 were considered as significant. The integration of PPI network was visualized by Cytoscape software v3.7.2[37]. The clusters of the co-expression networks were obtained using MCODE plug-in with node score cutoff = 0.2 and k-core = 2 [38]. The hub genes in sub-clusters were selected as candidate genes for subsequent GC analysis.

Least absolute shrinkage and selection operator (LASSO) analysis

To study the expression of DEGs in predicting prognosis, the R software package glmnet was used for LASSO Cox regression analysis. LASSO regression can reduce the weighted mean value of case mean square prediction error by identifying the regression coefficient of genes [39]. According to the expression of the selected genes in LASSO analysis and the correlation coefficient of Cox proportional hazard analysis, the risk score (RS) of death was calculated for patients with GC.

Statistical Analysis

The wilcox.test and rank sum test were applied to obtain the ASPM, COL11A1 and CDC20 gene expressions in 381 tumor tissues and 32 normal gastric tissues from TCGA dataset. A chi-square test

was used to compare the expression level of clinical data in patients with gastric cancer. Kaplan Meier curve analysis was performed according to the calculated RS to assess its relationship with overall survival. The intermediate value was defined as the cut-off. Univariate and multivariate Cox regression analysis were conducted to explore the relationship between RS and clinical prognostic features. ROC curve (AUC) was also used to verify the accuracy of the prognostic risk model. All incomplete or missing samples were removed from the analysis.

Verification of Core Genes

A total of 15 pairs of specimens (gastric cancer tissues and adjacent normal samples) were obtained from gastric cancer patients who received surgical treatment between December 2020 and March 2021 at The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University. All patients provided written informed consent prior to collection of specimens. Total RNA was extracted from tissues using TRIzol reagent. Total RNA was reverse transcribed into cDNA using the PrimeScript™ RT Master Mix (Takara, Wenzhou, China). Quantitative PCR primers were designed using Primer Premier 5.0 and Beacon Designer 7.8. Real-time PCR was used for statistical analysis of gene expression differences. Human GAPDH was used as the internal reference gene. The data were analyzed using the comparative cycle threshold (CT, $2^{-\Delta\Delta CT}$) method. Quantitative RT-PCR program settings are as follows: predenaturation at 95°C for 5 min, denaturation at 95°C for 30 sec, annealing/extension at 60°C for 30 sec, and repeat 40 cycles; other parameters are performed according to the instrument default settings.

Declarations

Acknowledgements

Not applicable.

Author Contributions

YangYang Teng, Na Shan and GuangRong Lu contributed to the design of the study. LeYi Ni, ZeJun Gao, Yao Shen, Sheng Xu, QinJian Wang, QingJi Ying, Jing Zhang performed statistical analysis and experiments. HeJie Yu, ZhanXiong Xue, Chao Xing and ZhenZhai Cai contributed to the supervision and writing of the study. All authors revised and approved the final version of the manuscript.

Funding

This work was supported by Zhejiang Provincial Natural Science Foundation of China (Grant No. LY17H200002) and Wen Zhou Science and Technology Bureau (Grant No. Y20170050).

Availability of data and material

The data used in the study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69(1):7–34.
2. Sun F, Sun H, Mo X, Tang J, Liao Y, Wang S, Su Y, Ma H. Increased survival rates in gastric cancer, with a narrowing gender gap and widening socioeconomic status gap: A period analysis from 1984 to 2013. *J Gastroenterol Hepatol.* 2018;33(4):837–46.
3. Gravalos C, Jimeno A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. *Ann Oncol.* 2008;19(9):1523–9.
4. Gomez-Martín C, Lopez-Rios F, Aparicio J, Barriuso J, García-Carbonero R, Pazo R, Rivera F, Salgado M, Salud A, Vázquez-Sequeiros E, Lordick F. A critical review of HER2-positive gastric cancer evaluation and treatment: from trastuzumab, and beyond. *Cancer Lett.* 2014;351(1):30–40.
5. Raufi AG, Klempner SJ. Immunotherapy for advanced gastric and esophageal cancer: preclinical rationale and ongoing clinical investigations. *J Gastrointest Oncol.* 2015;6(5):561–9.
6. Fife BT, Pauken KE. The role of the PD-1 pathway in autoimmunity and peripheral tolerance. *Ann N Y Acad Sci.* 2011;1217:45–59.
7. Johnston FM, Beckman M. Updates on Management of Gastric Cancer. *Curr Oncol Rep.* 2019;21(8):67.
8. Coutzac C, Pernot S, Chaput N, Zaanani A: **Immunotherapy in advanced gastric cancer, is it the future?** *Crit Rev Oncol Hematol* 2019, **133**:25–32.
9. Geckil H, Xu F, Zhang X, Moon S, Demirci U. Engineering hydrogels as extracellular matrix mimics. *Nanomedicine (Lond).* 2010;5(3):469–84.
10. Theocharis AD, Skandalis SS, Gialeli C, Karamanos NK. Extracellular matrix structure. *Adv Drug Deliv Rev.* 2016;97:4–27.
11. Lu P, Weaver VM, Werb Z. The extracellular matrix: a dynamic niche in cancer progression. *J Cell Biol.* 2012;196(4):395–406.
12. Barker HE, Chang J, Cox TR, Lang G, Bird D, Nicolau M, Evans HR, Gartland A, Erler JT. LOXL2-mediated matrix remodeling in metastasis and mammary gland involution. *Cancer Res.* 2011;71(5):1561–72.

13. Danielson PB. The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr Drug Metab.* 2002;3(6):561–97.
14. Badawi AF, Cavalieri EL, Rogan EG: **Role of human cytochrome P450 1A1, 1A2, 1B1, and 3A4 in the 2-, 4-, and 16alpha-hydroxylation of 17beta-estradiol.** *Metabolism* 2001, **50**(9):1001–1003.
15. Hecht SS, Stepanov I, Carmella SG. Exposure and Metabolic Activation Biomarkers of Carcinogenic Tobacco-Specific Nitrosamines. *Acc Chem Res.* 2016;49(1):106–14.
16. Hodges RE, Minich DM. Modulation of Metabolic Detoxification Pathways Using Foods and Food-Derived Components: A Scientific Review with Clinical Application. *J Nutr Metab.* 2015;2015:760689.
17. White IN, Razvi N, Gibbs AH, Davies AM, Manno M, Zaccaro C, De Matteis F, Pähler A, Dekant W. Neoantigen formation and clastogenic action of HCFC-123 and perchloroethylene in human MCL-5 cells. *Toxicol Lett.* 2001;124(1–3):129–38.
18. Moorthy B, Chu C, Carlin DJ. Polycyclic aromatic hydrocarbons: from metabolism to lung cancer. *Toxicol Sci.* 2015;145(1):5–15.
19. Rao PS, Kumar S. Polycyclic aromatic hydrocarbons and cytochrome P450 in HIV pathogenesis. *Front Microbiol.* 2015;6:550.
20. Agundez JA. Cytochrome P450 gene polymorphism and cancer. *Curr Drug Metab.* 2004;5(3):211–24.
21. Johnson AL, Edson KZ, Totah RA, Rettie AE. Cytochrome P450 ω -Hydroxylases in Inflammation and Cancer. *Adv Pharmacol.* 2015;74:223–62.
22. Suzuki F, Fujiwara Y, Hamura R, Haruki K, Sakamoto T, Shiba H, Yanaga K. Combination of Distance from Superior Mesenteric Artery and Serum CA19-9 as a Novel Prediction of Local Recurrence in Patients With Pancreatic Cancer Following Resection. *Anticancer Res.* 2019;39(3):1469–78.
23. Li A, Li J, Lin J, Zhuo W, Si J. COL11A1 is overexpressed in gastric cancer tissues and regulates proliferation, migration and invasion of HGC-27 gastric cancer cells in vitro. *Oncol Rep.* 2017;37(1):333–40.
24. Shen L, Yang M, Lin Q, Zhang Z, Zhu B, Miao C. COL11A1 is overexpressed in recurrent non-small cell lung cancer and promotes cell proliferation, migration, invasion and drug resistance. *Oncol Rep.* 2016;36(2):877–85.
25. Wu YH, Chang TH, Huang YF, Huang HD, Chou CY. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. *Oncogene.* 2014;33(26):3432–40.
26. Gu SQ, Luo JH, Yao WX. The regulation of miR-139-5p on the biological characteristics of breast cancer cells by targeting COL11A1. *Math Biosci Eng.* 2019;17(2):1428–41.
27. Kleinert R, Prenzel K, Stoecklein N, Alakus H, Bollschweiler E, Hölscher A, Warnecke-Eberz U. Gene Expression of Col11A1 Is a Marker Not only for Pancreas Carcinoma But also for Adenocarcinoma of the Papilla of Vater, Discriminating Between Carcinoma and Chronic Pancreatitis. *Anticancer Res.* 2015;35(11):6153–8.
28. Zhang B, Zhang C, Yang X, Chen Y, Zhang H, Liu J, Wu Q. Cytoplasmic collagen XI α 1 as a prognostic biomarker in esophageal squamous cell carcinoma. *Cancer Biol Ther.* 2018;19(5):364–72.

29. Wang L, Zhang J, Wan L, Zhou X, Wang Z, Wei W. Targeting Cdc20 as a novel cancer therapeutic strategy. *Pharmacol Ther.* 2015;151:141–51.
30. Kidokoro T, Tanikawa C, Furukawa Y, Katagiri T, Nakamura Y, Matsuda K. CDC20, a potential cancer therapeutic target, is negatively regulated by p53. *Oncogene.* 2008;27(11):1562–71.
31. Harley ME, Allan LA, Sanderson HS, Clarke PR. Phosphorylation of Mcl-1 by CDK1-cyclin B1 initiates its Cdc20-dependent destruction during mitotic arrest. *Embo j.* 2010;29(14):2407–20.
32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
33. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics.* 2012;28(4):573–80.
34. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* 2012;16(5):284–7.
35. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
36. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017;45(D1):D362-d368.
37. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
38. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;4:2.
39. Zhao Y, Simon R. Development and validation of predictive indices for a continuous outcome using gene expression profiles. *Cancer Inform.* 2010;9:105–14.

Figures

heatmap represents an independent gene chip, and each row represents the log2 (fold change) of a gene compared with para-carcinoma tissues in different microarrays. The heatmap shows the top 20 genes screened by Robust Rank Aggreg (RRA). E and F. GO enrichment analysis of high expression genes (E) and low expression genes (F). G and H. Gene ontology (GO) enrichment analysis (G) and Reactome pathway enrichment analysis (H) of final DEGs screened by two methods.

Figure 2

WGCNA analysis of GSE63089 dataset. The blue, brown, yellow and turquoise color indicate the different co-expression network modules for genes and the removal genes are shown in the gray module.

Figure 3

Data analysis of ASPM, COL11A1 and CDC20 genes in TCGA database. A. The expression levels of ASPM, COL11A1 and CDC20 genes in different samples, the higher the absolute value of gene expression, the darker the color, red: up-regulation, green: down-regulation. In the bar chart, blue represents normal tissue samples and red represents tumor samples. B and C. The coefficient calculated by multivariate Cox regression using LASSO. D. The survival of patients in high-risk group and low-risk group using Kaplan Meier method. * $P < 0.1$, ** $P < 0.05$, and *** $P < 0.01$.

Figure 4

Clinical data analysis of COL11A1 and CDC20 genes. A. The relationship between COL11A1 and CDC20 gene expression heatmap and clinicopathological characteristics. The higher the absolute value of gene expression, the darker the color, Red: up-regulated, Green: down-regulated. The bar chart represents the clustering results of different experimental groups. The information represented by different colors is shown on the right side of the picture. B. The ROC curve shows the accuracy of the risk score in predicting 5-year survival. C. The relationship between survival rate and clinicopathological features by Univariate Cox regression analysis, including risk score. D. The relationship between survival rate and clinicopathological features using Multivariate Cox regression analysis, including risk score. * $P < 0.1$, ** $P < 0.05$, and *** $P < 0.01$.

Figure 5

Quantitative RT-PCR verification of core genes. A. The expression of CDC20 gene in gastric cancer tissues(N) was significantly higher than that in normal gastric tissues(T) in groups 1, 7, 8, 9, 10, 11, 12, 13, 14 and 15 ($P<0.01$). B. In groups 1, 7, 8, 10, 11, 12, 13 and 15, the expression of COL11A1 gene in gastric cancer tissue(T) was significantly higher than that in normal gastric tissue(N) ($P<0.01$).