Int J Clin Exp Med 2018;11(7):7111-7117
www.ijcem.com /ISSN:1940-5901/IJCEM0061896

Original Article

Identification of potential biomarkers for risk analysis of colorectal cancer using a combined bioinformatics analysis

Xiaoyan Zhang[1,2*], Zebin Hu[3*], Zhao Liang[4], Qiang Wu[5], Bing Xiu[5], Ping Li[5], Dong Li[5], Mingmin Chen[2], Hengjun Gao[2,6]

[1]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China; [2]National Engineering Center for Biochip at Shanghai, Shanghai 201203, China; [3]Institue for In Vitro Diagnostic Reagents Control, The National Institutes for Food and Drug Control (NIFDC), China; [4]Tianjin First Center Hospital, Tianjin, China; [5]Tongji Hospital Affiliated to Tongji University, Shanghai, China; [6]Tongji Institute of Digestive Disease and Department of Gastroenterology, Tongji University, Shanghai, China. [*]Equal contributors.

Received April 20, 2017; Accepted January 8, 2018; Epub July 15, 2018; Published July 30, 2018

Abstract: Colorectal cancer (CRC) is among the three most commonly diagnosed cancers in the world. To screen molecular biomarkers for the diagnosis of CRC, three gene expression databases (GSE74602, GSE31279 and GSE10950) and one microRNA (miRNA) microarray database (GSE41655) were downloaded from Gene Expression Omnibus (GEO) and analyzed in this study. The different expressed genes (DEGs) and different expressed miRNAs (DEMs) between tumor and normal colorectal samples were calculated by GEO2R. DAVID online analysis was performed to investigate the gene ontology (GO) and KEGG pathway enrichment of up-regulated and down-regulated DEGs respectively. The protein-protein interactions between DEGs were calculated by STRING. The hub genes were subsequently calculated by MCODE. Finally, the prognostic role of hub genes was evaluated by SurvExpress online tool. Through this calculation, 303 DEGs were screened out, 113 among which were up-regulated genes and the rest 189 were down-regulated genes. 10 hub genes (IL8, CXCL1, CXCL10, CCL20, CCL21, GRM8, CXCL12, SST, PYY and LPAR1) were identified by MCODE. Moreover, 113 DEMs, including 56 up-regulated miRNA and 57 down regulated miRNA, were screened out. CXCL12 was positively correlated with a down regulated miRNA has-let-7f by miRecords prediction. Finally, 10 hub genes could classify CRC patients into low- and high-risk groups in four databases. In conclusion, we screened out 10 hub genes that of great significance for the prediction of CRC prognosis.

Keywords: Colorectal cancer, DEGs, PPI, MCODE, SurvExpress

Introduction

Colorectal cancer (CRC) is among the three most commonly diagnosed cancers in the world [1]. Owing to early detection and advanced therapy, the mortality and incidence of CRC was decreased in recent years [2]. As early detection of CRC by colonoscopy was not only expensive but also taking a great risk, biomarkers were developed to screen CRC in early stage.

Several biomarkers have been proposed for the prediction of CRC risk and treatment response to improve the outcome of patients. For example, CEA, cytokeratin 20, and epidermal growth factor receptor (EGFR) transcripts from patients with CRC can indicate disease stage and prognosis [3]. However, the RT-PCR based detection of these biomarkers in circulation tumor cells was still too expensive. It is emerging to find more sensitive and specific molecular biomarkers for the diagnosis of CRC.

Microarray expression data for CRC was widely used for investigation of carcinogenic processing mechanism [4]. Three gene expression data and one miRNA microarray data from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) were downloaded and analyzed in this study. Our aim of this study was to screen out genes for the diagnosis of CRC.

Materials and methods

Gene expression data

Three gene expression datasets (GSE74602, GSE31279 and GSE10950) and one miRNA ex-

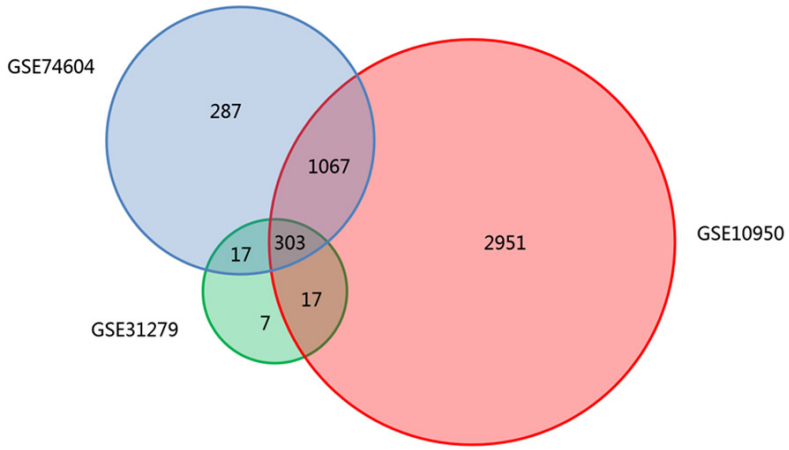

Figure 1. Identification of different expressed genes in GSE74604, GSE 10950 and GSE31279. |logFC|>1 and adj. *P*<0.05 were considered to be the threshold values.

pression profile (GSE41655) were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). The array date of GSE74602 contained 30 paired tumor and normal colorectal samples. GSE10950 included 24 paired normal and tumor colorectal specimens. GSE31279 consisted of 35 paired tumor and normal colorectal tissue specimens. The miRNA expression profile array datasets GSE41655 included 33 tumor and 15 normal colorectal mucosas.

Data processing

The different expressed genes (DEGs) and different expressed miRNAs (DEMs) between tumor and normal colorectal samples were calculated by GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r). The adjusted *P* (adj. *P*) values corrected the occurrence of false positive results by Benjamini and Hochberg false discovery rate method. |logFC|>1 and adj. *P*<0.05 were considered to be the threshold values.

Functional and pathway enrichment analysis

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment of DEGs were analyzed by DAVID online tools (https://david.ncifcrf.gov/). *P*<0.05 was considered significantly.

Protein-protein interaction network construction and modules selection

The protein-protein interaction (PPI) between DEGs were calculated by STRING online tools (http://www.string-db.org/) and visualized by Cytoscape. The hub genes were further calculated by MCODE with degree cutoff=2, node score cutoff=0.2, k-core=2 and max. depth=100 [5].

Survival analysis of hub genes

The prognostic role of hub genes screened out by MCODE was evaluated by SurvExpress online tool (http://bioinformatica.mty.itesm.mx/) [6]. All hub genes were incorporated to define low and high risk by using prognostic index (PI), the liner component of the Cox model. $PI=\beta_1 x_1+\beta_2 x_2+.....+\beta_p x_p$ (x_i is the expression value and the β_i is obtained from the Cox fitting). The classical Cox model (all genes were included in a unique model) was used to evaluate the prognostic role of hub genes. The high and low risk groups were splited by median PI values (the higher PI values for higher risk).

Results

Identification of DEGs

By GEO2R analysis, 1676, 323 and 4347 DEGs were screened out from GSE74604, GSE31279 and GSE10950 respectively. There were 303 common DEGs in total, 113 of which were up-regulated genes while the other 189 were down-regulated genes (**Figure 1**).

Gene ontology and pathway enrichment analysis

GO and KEGG pathway analysis of DEGs revealed that the up-regulated genes were enriched in chemokine activity, biosynthesis of amino acids and TNF signaling pathway while the down-regulated genes were enriched in muscle contraction and nitrogen metabolism (**Table 1**).

PPI network screening the hub genes in CRC

To identify the hub genes in all selected DEGs, PPI network analysis was performed using STRING with combined score≥0.7 and visualized by Cytoscape (**Figure 2A**). Functional gene

Table 1. Top 10 gene ontology and pathway enrichment of DEGs by DAVID

Term	Function	*P* Value
Up-regulated genes		
GO:0004252	Serine-type endopeptidase activity	2.59E-05
GO:0005615	Extracellular space	3.81E-06
GO:0005576	Extracellular region	1.15E-03
GO:0005578	Proteinaceous extracellular matrix	1.59E-03
GO:0031528	Microvillus membrane	2.73E-03
GO:0008009	Chemokine activity	3.32E-03
GO:0070062	Extracellular exosome	3.40E-03
GO:0030574	Collagen catabolic process	4.66E-03
GO:0006508	Proteolysis	7.21E-03
GO:0051301	Cell division	8.10E-03
hsa01230	Biosynthesis of amino acids	1.79E-02
hsa04668	TNF signaling pathway	4.51E-02
hsa04114	Oocyte meiosis	4.83E-02
Down-regulated genes		
GO:0006936	Muscle contraction	1.31E-07
GO:0015701	Bicarbonate transport	2.69E-07
GO:0030018	Z disc	3.44E-07
GO:0008307	Structural constituent of muscle	3.58E-06
GO:0008201	Heparin binding	2.10E-04
GO:0006730	One-carbon metabolic process	2.26E-04
GO:0070062	Extracellular exosome	2.80E-04
GO:0004089	Carbonate dehydratase activity	3.31E-04
GO:0003779	Actin binding	5.01E-04
GO:0008285	Negative regulation of cell proliferation	7.22E-04
hsa00910	Nitrogen metabolism	8.71E-04
hsa04976	Bile secretion	1.09E-03
hsa04964	Proximal tubule bicarbonate reclamation	2.16E-03
hsa04972	Pancreatic secretion	4.08E-03
hsa00830	Retinol metabolism	6.32E-03
hsa04971	Gastric acid secretion	9.49E-03
hsa04960	Aldosterone-regulated sodium reabsorption	9.78E-03
hsa04270	Vascular smooth muscle contraction	1.14E-02
hsa04530	Tight junction	1.99E-02
hsa00982	Drug metabolism-cytochrome P450	4.24E-02

modules were identified by MCODE and subsequently one significant gene module, including 10 nodes and 45 edges, were screened out (**Figure 2B**). The 10 hub genes included IL8, CXCL1, CXCL10, CCL20, CCL21, GRM8 (seed), CXCL12, SST, PYY and LPAR1. The hub genes were enriched in chemokine signaling pathway. CXCL1, CXCL10, CCL20, IL8 and GRM8 were up-regulated genes, and the other five genes (CXCL12, CCL21, SST, PYY and LPAR1) were down-regulated genes in CRC patients.

miRNA-DEGs pairs

According to GEO2R analysis, 113 DEMs, including 56 up-regulated miRNA and 57 down regulated miRNA, were screened out from miRNA expression profile (GSE41655). The potential targets of DEMs were predicted by miRecords (http://c1.accurascience.com/miRecords/). In compared with DEGs, up regulated expressed gene SERPINB5 was a target of has-miR-1 and has-miR-21. The down-regulated DEG MEF2D was a target of has-miR-17 and has-miR-20a [7] while EPB41L3 was a target of has-miR-223 [8]. Moreover, As predicted by miRDB (http://mirdb.org/miRDB/), one of the hub genes CXCL12 was a potential target of a down-regulated miRNA has-miR-765.

Virtual validation of hub genes

All hub genes were incorporated to differentiate low- and high- risk CRC patients in multi databases by SurvExpress online tool. The results showed that CRC patients were able to be classified into low- and high-risk groups significantly in four database by hub genes (GSE41258, 244 samples, *P*=4.227e-07, **Figure 3A**; SurvExpress 808 samples, *P*=5.341e-09, **Figure 3B**; GSE14333, 260 samples, *P*=3.054e-03, **Figure 3C**; TCGA, 467 samples, *P*=0.01458, **Figure 3D**).

Discussion

Although the survival time of primary CRC patients could be prolonged by regional resection, the survival rate of CRC patients decreased once metastasis occurred. Therefore, underlying the mechanism of CRC metastasis would be of great importance for new therapy developing of metastasis CRC. In this study, 10 hub genes (CXCL1, IL8, CXCL10, CXCL12, CCL20, CCL21, GRM8, SST, PYY and LPAR1) were screened out, six of which were chemokines (CXCL1, IL8, CXCL10, CXCL12, CCL20 and CCL21). Chemok-

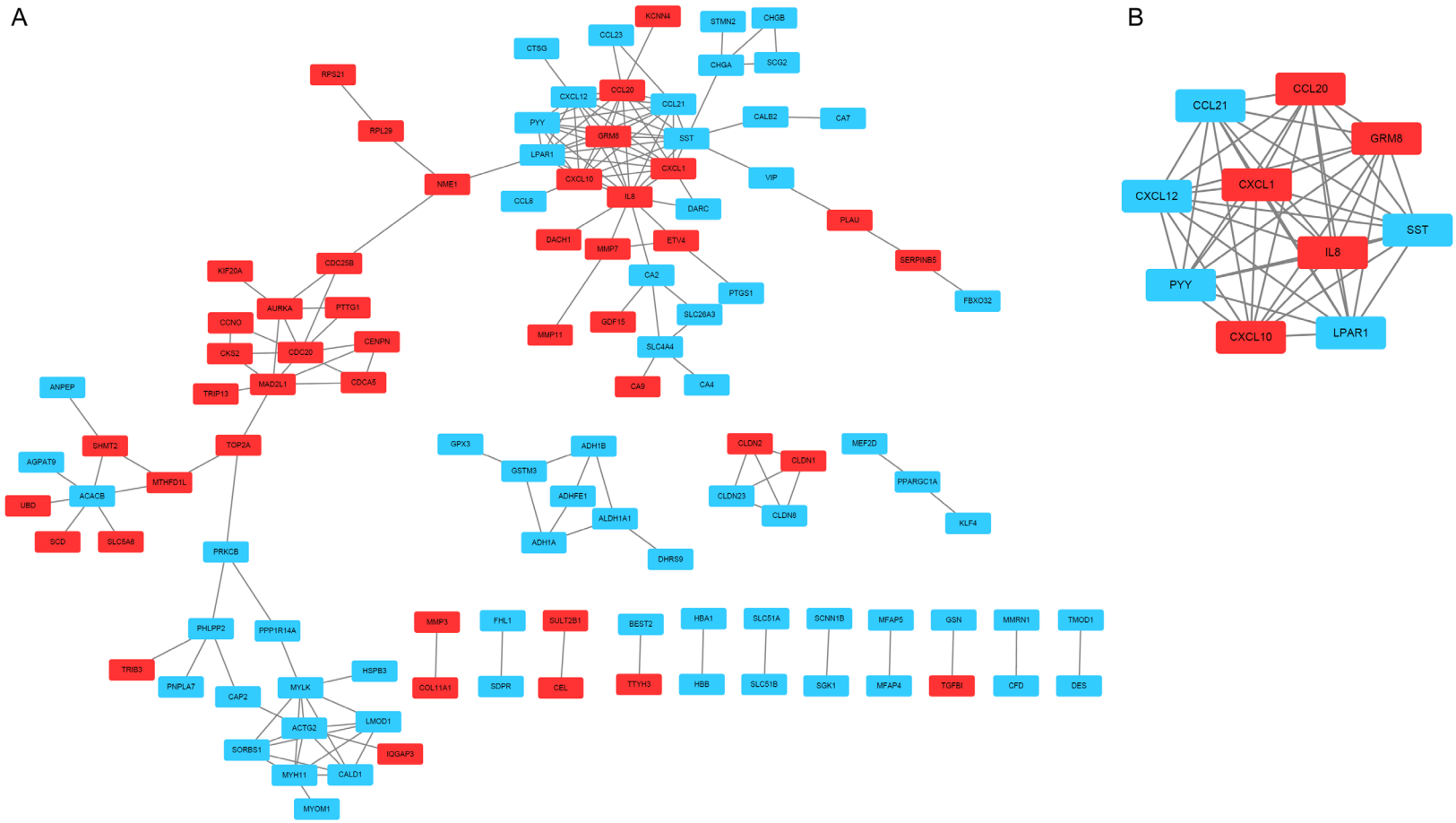

Figure 2. Protein-protein interaction network of DEGs analyzed by STRING. Combined score≥0.7 was considered threshold (A); A significant module in PPI network analyzed by MCODE (B). The blue boxes represent down regulated genes, and the red boxes represent up regulated genes.

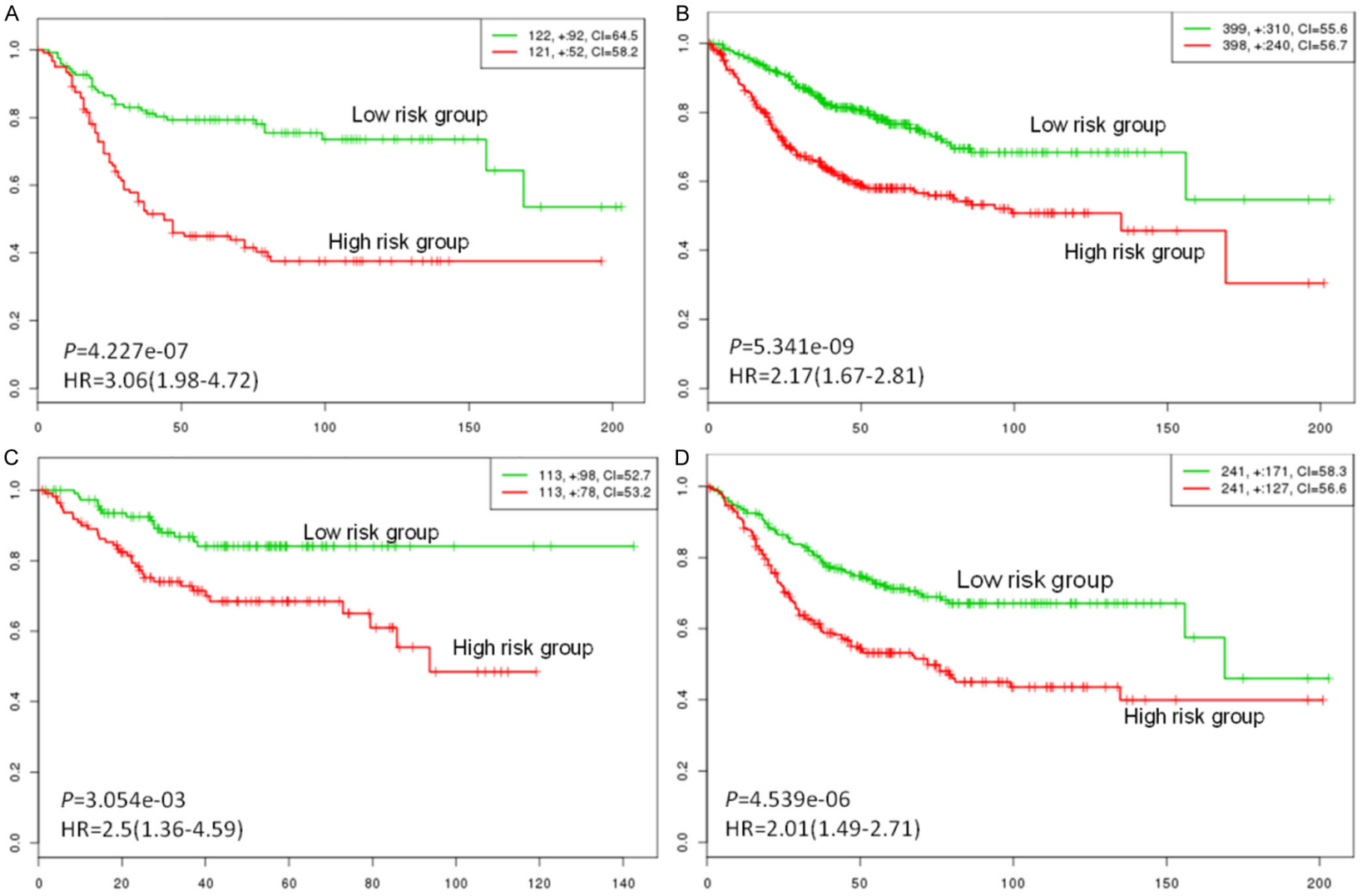

Figure 3. Kaplan-Meier curves for different risk CRC groups divided by hub genes in four databases. A. GSE41258, 244 samples; B. SurvExpress 808 samples; C. GSE14333, 260 samples; D. TCGA, 467 samples.

ines are a family of small peptides that binding to G-protein-coupled receptors, which is one of the key factors that are responsible for cancer metastasis in some types of cancer including CRC. The role of the six chemokines screened out in this study was proved to be vital in CRC. CXCL12 was one component of the representative chemokine signaling pathway CXCL12/CXCR4. Although the function of CXCL12 in CRC was still debatable, its vital role is undoubted [9-12]. CXCL1 and IL8 which were all bind to CXCR2 were proved to take an important role in angiogenesis in CRC. Consistently with our results, up regulated CXCL1, CCL20 and down regulated CCL21 in CRC specimens was validated in some previous works [13, 14]. CXCL10 was reported to be a prognostic factor in CRC [15, 16].

The other four genes involved in the hub genes were also proved to be participated in tumorigenic progression of CRC. The down-regulation of PYY was related to the metastasis of CRC which was in consistent with the observation in our study [17]. Somatostatin (SST), a regulatory inhibitory peptide, played a vital role in inhibition of gastrointestinal motility, epithelial proliferation and hormone release. Moreover, SST inhibited cell proliferation and induced apoptosis through somatostatin receptor signaling, and thus involved in the CRC tumorigenesis [18, 19]. Metabotropic glutamate receptor 8 (GRM8) was considered as a surface marker of colon cancer cells and mutation of GRM8 was related to human cancers [20].

Accumulating evidence suggested that miRNA played a vital role in CRC tumorigenisis via their target genes [21]. MiRNA has-let-7 was down-regulated in various cancers including CRC [21], and acted as a tumor suppressor by suppressing the expression of KRAS. In this study, we found CXCL12 as the potential target gene of has-let-7f, suggesting has-let-7f might function through chemokines activity.

In conclusion, our study was performed to screen out 10 hub genes for CRC risk prediction and validated in four databases. Further work that validating the prognosis prediction function of the 10 hub genes in large CRC samples was needed to be done.

Disclosure of conflict of interest

None.

Address correspondence to: Hengjun Gao, National Engineering Center for Biochip at Shanghai, Shanghai 201203, China; Tongji Institute of Diges-tive Disease and Department of Gastroenterology, Tongji University, Shanghai, China. E-mail: hengjun_gao@shbiochip.com

References

[1] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2017. CA Cancer J Clin 2017; 67: 7-30.

[2] Mocellin S, Baretta Z, Roqué I Figuls M, Sola I, Martin-Richard M, Hallum S and Bonfill Cosp X. Second-line systemic therapy for metastatic colorectal cancer. Cochrane Database Syst Rev 2017; 1: CD006875.

[3] Tsouma A, Aggeli C, Lembessis P, Zografos GN, Korkolis DP, Pectasides D, Skondra M, Pissimissis N, Tzonou A and Koutsilieris M. Multiplex RT-PCR-based detections of CEA, CK20 and EGFR in colorectal cancer patients. World J Gastroenterol 2010; 16: 5965-5974.

[4] Chang J, Huang L, Cao Q and Liu F. Identification of colorectal cancer-restricted microRNAs and their target genes based on high-throughput sequencing data. Onco Targets Ther 2016; 9: 1787-1794.

[5] Bader GD and Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003; 4: 2.

[6] Aguirre-Gamboa R, Gomez-Rueda H, Martinez-Ledesma E, Martinez-Torteya A, Chacolla-Huaringa R, Rodriguez-Barrientos A, Tamez-Pena JG and Trevino V. SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. PLoS One 2013; 8: e74250.

[7] Beveridge NJ, Tooney PA, Carroll AP, Tran N and Cairns MJ. Down-regulation of miR-17 family expression in response to retinoic acid induced neuronal differentiation. Cell Signal 2009; 21: 1837-1845.

[8] Li X, Zhang Y, Zhang H, Liu X, Gong T, Li M, Sun L, Ji G, Shi Y, Han Z, Han S, Nie Y, Chen X, Zhao Q, Ding J, Wu K and Daiming F. miRNA-223 promotes gastric cancer invasion and metastasis by targeting tumor suppressor EPB41L3. Mol Cancer Res 2011; 9: 824-833.

[9] Stanisavljevic L, Assmus J, Storli KE, Leh SM, Dahl O and Myklebust MP. CXCR4, CXCL12 and the relative CXCL12-CXCR4 expression as prognostic factors in colon cancer. Tumour Biol 2016; 37: 7441-7452.

[10] Izumi D, Ishimoto T, Miyake K, Sugihara H, Eto K, Sawayama H, Yasuda T, Kiyozumi Y, Kaida T, Kurashige J, Imamura Y, Hiyoshi Y, Iwatsuki M, Iwagami S, Baba Y, Sakamoto Y, Miyamoto Y, Yoshida N, Watanabe M, Takamori H, Araki N,

Tan P and Baba H. CXCL12/CXCR4 activation by cancer-associated fibroblasts promotes integrin beta1 clustering and invasiveness in gastric cancer. Int J Cancer 2016; 138: 1207-1219.

[11] Song ZY, Gao ZH, Chu JH, Han XZ and Qu XJ. Downregulation of the CXCR4/CXCL12 axis blocks the activation of the Wnt/beta-catenin pathway in human colon cancer cells. Biomed Pharmacother 2015; 71: 46-52.

[12] Sakai N, Yoshidome H, Shida T, Kimura F, Shimizu H, Ohtsuka M, Takeuchi D, Sakakibara M and Miyazaki M. CXCR4/CXCL12 expression profile is associated with tumor microenvironment and clinical outcome of liver metastases of colorectal cancer. Clin Exp Metastasis 2012; 29: 101-110.

[13] Rubie C, Frick VO, Wagner M, Schuld J, Graber S, Brittner B, Bohle RM and Schilling MK. ELR+CXC chemokine expression in benign and malignant colorectal conditions. BMC Cancer 2008; 8: 178.

[14] Mumtaz M, Wagsater D, Lofgren S, Hugander A, Zar N and Dimberg J. Decreased expression of the chemokine CCL21 in human colorectal adenocarcinomas. Oncol Rep 2009; 21: 153-158.

[15] Agesen TH, Sveen A, Merok MA, Lind GE, Nesbakken A, Skotheim RI and Lothe RA. ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. Gut 2012; 61: 1560-1567.

[16] Jiang Z, Xu Y and Cai S. CXCL10 expression and prognostic significance in stage II and III colorectal cancer. Mol Biol Rep 2010; 37: 3029-3036.

[17] Adrian TE, Ballantyne GH, Zucker KA, Zdon MJ, Tierney R and Modlin IM. Lack of peptide YY immunoreactivity in adenomatous colonic polyps: evidence in favor of an adenoma-carcinoma sequence. J Surg Res 1988; 44: 561-565.

[18] Brenner H, Kloor M and Pox CP. Colorectal cancer. Lancet 2014; 383: 1490-1502.

[19] Liu Y, Chew MH, Tham CK, Tang CL, Ong SY and Zhao Y. Methylation of serum SST gene is an independent prognostic marker in colorectal cancer. Am J Cancer Res 2016; 6: 2098-2108.

[20] Sewda K, Coppola D, Enkemann S, Yue B, Kim J, Lopez AS, Wojtkowiak JW, Stark VE, Morse B, Shibata D, Vignesh S and Morse DL. Cell-surface markers for colon adenoma and adenocarcinoma. Oncotarget 2016; 7: 17773-17789.

[21] Ghanbari R, Mosakhani N, Sarhadi VK, Armengol G, Nouraee N, Mohammadkhani A, Khorrami S, Arefian E, Paryan M, Malekzadeh R and Knuutila S. Simultaneous underexpression of let-7a-5p and let-7f-5p microRNAs in plasma and stool samples from early stage colorectal carcinoma. Biomark Cancer 2015; 7: 39-48.