Identification of Significant Genes With Poor Colorectal Cancer Prognosis In Via Bioinformatical Analysis

Shaodong Li

Jilin University First Hospital

Ruizhi Dong

Jilin University First Hospital

Bin Liang

Jilin University First Hospital

Zhenhua Kang (✉ 1269232708@qq.com)

Jilin University First Hospital <https://orcid.org/0000-0001-6045-9663>

Research Article

Keywords: Bioinformatical analysis, Microarray, colorectal cancer, Differentially expressed gene

Posted Date: July 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-645864/v1>

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

[Read Full License](#)

Abstract

Purpose: Identification of significant genes with poor colorectal cancer prognosis in via bioinformatical analysis.

Method: Gene expression profiles of GSE74602, GSE110223, GSE113513 and GSE 141174 were available from GEO database. There are 65 CRC tissues and 65 normal tissues in the four profile datasets. Differentially expressed genes (DEGs) between CRC tissues and normal tissues were picked out by GEO2R tool and Venn diagram software. Next, we made use of the Database for Annotation, Visualization and Integrated Discovery (DAVID) to analyze Kyoto Encyclopedia of Gene and Genome (KEGG) pathway and gene ontology (GO). Then protein-protein interaction (PPI) of these DEGs with Search Tool for the Retrieval of Interacting Genes (STRING).

Results: There were total of 171 consistently expressed genes in the four datasets, including 148 up-regulated and 23 down-regulated genes. up-regulated DEGs were particularly enriched in oxidation-reduction process, in extracellular exosome, in zinc ion binding, in Metabolic pathways, Mineral absorption; and down-regulated DEGs in positive regulation of cell proliferation, in cytosol, in One carbon pool by folate. Furthermore, for the analysis of overall survival among those genes, Kaplan–Meier analysis was implemented and 30 of 88 genes had a significantly worse prognosis. For validation in Gene Expression Profiling Interactive Analysis (GEPIA), 13 of 30 genes were discovered highly expressed in CRC tissues compared to normal tissues. Furthermore, MYC & FGFR3 markedly enriched in the Bladder cancer pathway.

Conclusion: We have identified two significant up-regulated DEGs with poor prognosis in CRC , which could be potential therapeutic targets for CRC patients.

Introduction

Colorectal cancer(CRC) is a sort of common malignant tumor of digestive tract its morbidity and mortality rank third and fourth respectively in the world, and in China, the morbidity and mortality of CRC is also very high, and there is a tendency to increase, which seriously threatens the physical and mental health of human beings^[1–2]. Therefore, more reliable prognostic biomarkers should be explored as a target for improving the treatment effect and better understanding the underlying mechanism^[3–6].

Bioinformatics and Gene chip which were used for more than ten years can quickly detect differentially expressed genes and was proved to be a reliable technique that could make many slice data be produced and stored in public databases. Therefore, a large number of valuable clues could be explored for new research on the base of these data. Furthermore, many bioinformatical studies on CRC have been produced in recent years ^[7], which proved that the ntegrated bioinformatical methods could help us to further study and better exploring the underlying mechanisms.

In this study, first we choosed GSE74602,GSE110223 and GSE113513 and GSE 141174 from Gene Expression Omnibus (GEO). Second, we applied for GEO2R online tool and Venn diagram software to obtain the commonly differentially expressed genes (DEGs) in the three datasets above. Third, the Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to analyze these DEGs including molecular function (MF), cellular component (CC), biological process (BP) and Kyoto Encyclopedia of Gene and Genome (KEGG) pathways. Fourth, we established protein-protein interaction (PPI) network for additional analysis of the DEGs which would identify some core genes. Moreover, these core DEGs were imported into the Kaplan Meier plotter online database for the significant prognostic information ($P < 0.05$).

Method

Microarray data information

NCBI-GEO is regarded as a free public database of microarray/gene profile and we obtained the gene expression profile of GSE74602,GSE110223 and GSE113513 and GSE 141174 in colorectal cancer and normal tissue. Microarray data of GSE74602 and GSE 141174 were all on account of GPL6104 Platforms (Illumina humanRef-8 v2.0 expression beadchip) and GSE110223 was all on account of GPL96 Platforms [HG-U133A] Affymetrix Human Genome U133A Array and GSE113513 was all on account of GPL15207 Platforms [PrimeView] Affymetrix Human Gene Expression Array which included 30 CRC tissues and 30 normal tissues, 8 CRC tissues and 8 normal tissues, 13 CRC tissues and 13 normal tissues, 14 CRC tissues and 14 normal tissues.

Data processing of DEGs

DEGs between colorectal cancer and normal specimen were identified via GEO2R online tools^[8] with $|\log FC| > 1$ and adjust P value < 0.05 . Then, the raw data in TXT format were checked in Venn software online to detect the commonly DEGs among the three datasets. The DEGs with $\log FC < 0$ was considered as down-regulated genes, while the DEGs with $\log FC > 0$ was considered as an up-regulated gene.

Gene ontology and pathway enrichment analysis

Gene ontology analysis (GO) is a commonly used approach for defining genes and its RNA or protein product to identify unique biological properties of high-throughput transcriptome or genome data^[9]. KEGG is a collection of databases dealing with genomes, diseases, biological pathways, drugs, and chemical materials^[10]. DAVID which is an online bioinformatic tool is designed to identify a large number of genes or proteins function^[11]. We could use DAVID to visualize the DEGs enrichment of BP, MF, CC and pathways ($P < 0.05$).

PPI network

PPI information can be evaluated by an online tool, STRING (Search Tool for the Retrieval of Interacting Genes) [12]. Then, We could use STRING to visualize the potential correlation between these DEGs.(PPI enrichment p-value < 0.05)

Survival analysis and RNA sequencing expression of core genes

Kaplan Meier-plotter are a commonly used website tool for assessing the effect of a great number of genes on survival based on EGA, TCGA database and GEO (Affymetrix microarrays only) [13]. The log rank P value and hazard ratio(HR) with 95% confidence intervals were computed and showed on the plot. To validate these DEGs, we applied the GEPIA website to analyze the data of RNA sequencing expression on the basis of thousands of samples from the GTEx projects and TCGA [14].

Results

Identification of DEGs in colorectal cancer .There were 65 CRC tissues and normal tissues in our present study. Via GEO2R online tools, we extracted1643,12693497and2180 DEGs from GSE74602GSE110223GSE113513 and GSE 141174, respectively. Then, we used Venn diagram software to identify the commonly DEGs in the four datasets. Results showed that a total of 171commonly DEGs were detected, including 148 up-regulated genes(logFC > 0)and 23 down-regulated genes (logFC < 0) in the CRC tissues (Table 1 & Fig. 1).

Table 1

All 171 commonly differentially expressed genes (DEGs) were detected from four profile datasets in the CRC tissues compared to normal tissues

DEGs	Genes Name
Up-regulated	PLCE1 PTPRR HSD17B2 EPB41L4B ZG16 DHRS11 HSD11B2 PAPSS2 TRPM6 CES2 PYY VWA5A MAOA SLCO2A1 PDZD3 ZZEF1 GNA11 HHLA2 FGFR3 AHCYL2 CDHR5 PBLD PTPRH KLF4 MT1G AQP8 SGK1 AOC1 C1orf115 ACADVL LGALS4 ADH1C SGK2 SLC17A4 BMP2 UGT2A3 LRRC19 SST SCNN1B CLIC5 EPB41L3 PPP1R14D CWH43 MEP1A ACAA2 MALL GPD1L CA12 GUCA2A TP53I3 SLC22A5 MT1E CA4 APPL2 NAAA NAT2 KRT20 SLC26A2 TST ADTRP MT1F PLPP1 GCNT3 MS4A12 CEACAM1 HIST1H1C ABCA5 CDA SLC4A4 SI IQGAP2 SQRDL PHLPP2 PRDX6 AKR1B10 TUBAL3 DEFB1 ARL14 PIGZ CKB SLC26A3 CHGA CAPN9 DSC2 ABCB1 VILL PDE9A C4orf19 SMPD1 EDN3 SPON1 ITPKA PCK1 SMPDL3A ITM2C CASP7 SCGB2A1 TMPRSS2 ACADS SEMA6A DHRS9 PNLIPRP2 CA7 ENTPD5 MMP28 GUCA2B MGLL CHP2 CLCA4 UGDH MYO1A SLC36A1 NXPE4 FABP1 ETFDH MXI1 SLC1A1 FAM46A GDPD3 NR5A2 RBM47 FMO5 CA2 BCAS1 STAP2 ANPEP CEACAM7 ALPI ARHGAP44 TSPAN1 GCG CHST5 P2RX4 UGP2 CLDN8 EGLN3 MT1H VIPR1 PLAC8 GPA33 CA1 RETSAT SULT1B1 FXYD3 SELENBP1 GRAMD3 PARM1 CDHR2
Down-regulated	TRIB3 CDC25B PHLDA1 PRDX4 TCFL5 MYC NAP1L1 MMP11 LRP8 CEMIP CTPS1 COL4A1 SHMT2 PLAU GTF3A MTHFD2 SCD IFITM3 SLC7A5 FBL PPAT TESC TGFB1

DEGs gene ontology and KEGG pathway analysis in colorectal cancer

All 171 DEGs were analyzed by DAVID software and the results of GO analysis indicated that 1) The main genes enriched in biological processes (BP), up-regulated DEGs were particularly enriched in oxidation-

reduction process, bicarbonate transport, digestion, ion transmembrane transport and one-carbon metabolic process, and down-regulated DEGs in positive regulation of cell proliferation, cellular response to insulin stimulus and extracellular matrix organization; 2) The main genes enriched in cell component (CC),up-regulated DEGs were enriched in extracellular exosome, integral component of membrane, plasma membrane, extracellular spacea and integral component of plasma membrane, and down-regulated DEGs in cytosol, extracellular region and extracellular matrix; 3) The main genes enriched in molecular function (MF),up-regulated DEGs were enriched in zinc ion binding, oxidoreductase activity, transporter activity and hormone activity(Table 2).

Table2 Gene ontology analysis of differentially expressed genes in colorectal cancer

Expression	Category	Term	Count	%	p-Value	FDR
Up-regulated	GOTERM_BP_DIRECT	GO:0015701-bicarbonate transport	8	5.405405405	6.14E-08	4.60E-05
	GOTERM_BP_DIRECT	GO:0006730-one-carbon metabolic process	6	4.054054054	4.06E-06	0.001541992
	GOTERM_BP_DIRECT	GO:0007586-digestion	7	4.72972973	1.23E-05	0.00310071
	GOTERM_BP_DIRECT	GO:0030855-epithelial cell differentiation	6	4.054054054	2.66E-04	0.043717881
	GOTERM_BP_DIRECT	GO:0071276-cellular response to cadmium ion	4	2.702702703	3.32E-04	0.043717881
	GOTERM_BP_DIRECT	GO:0051114-oxidation-reduction process	15	10.13513514	3.46E-04	0.043717881
	GOTERM_BP_DIRECT	GO:0045926-negative regulation of growth	4	2.702702703	4.68E-04	0.044384436
	GOTERM_BP_DIRECT	GO:0071294-cellular response to zinc ion	4	2.702702703	4.68E-04	0.044384436
	GOTERM_BP_DIRECT	GO:0014533-regulation of intracellular pH	4	2.702702703	0.003115381	0.262730424
	GOTERM_BP_DIRECT	GO:0006811-ion transport	6	4.054054054	0.003857989	0.274673062
	GOTERM_BP_DIRECT	GO:0001558-regulation of cell growth	5	3.378378378	0.004154037	0.274673062
	GOTERM_BP_DIRECT	GO:0006814-sodium ion transport	5	3.378378378	0.004342625	0.274673062
	GOTERM_BP_DIRECT	GO:1902476-chloride transmembrane transport	5	3.378378378	0.007071591	0.409565793
	GOTERM_BP_DIRECT	GO:0034220-ion transmembrane transport	7	4.72972973	0.00754573	0.409565793
	GOTERM_BP_DIRECT	GO:0031539-fatty acid beta-oxidation using acyl-CoA dehydrogenase	3	2.027027027	0.009281157	0.469626556
	GOTERM_BP_DIRECT	GO:0010765-positive regulation of sodium ion transport	3	2.027027027	0.01031802	0.489461094
	GOTERM_BP_DIRECT	GO:0006011-UDP-glucose metabolic process	2	1.351351351	0.016251212	0.649193167
	GOTERM_BP_DIRECT	GO:0090675-intermicrovillar adhesion	2	1.351351351	0.016251212	0.649193167
	GOTERM_BP_DIRECT	GO:2001225-regulation of chloride transport	2	1.351351351	0.016251212	0.649193167
	GOTERM_BP_DIRECT	GO:1903385-regulation of homophilic cell adhesion	2	1.351351351	0.024278227	0.837598848
Down-regulated	GOTERM_BP_DIRECT	GO:0006730-one-carbon metabolic process	2	1.351351351	0.024278227	0.837598848
	GOTERM_BP_DIRECT	GO:0006065-UDP-glucuronate biosynthetic process	2	1.351351351	0.024278227	0.837598848
	GOTERM_BP_DIRECT	GO:0006805-anaerobic metabolic process	4	2.702702703	0.025925471	0.85540533
	GOTERM_BP_DIRECT	GO:0010922-positive regulation of phosphatase activity	2	1.351351351	0.032240219	0.978813062
	GOTERM_BP_DIRECT	GO:0032849-positive regulation of cellular pH reduction	2	1.351351351	0.032240219	0.978813062
	GOTERM_BP_DIRECT	GO:0007588-excretion	3	2.027027027	0.036538097	0.994757536
	GOTERM_BP_DIRECT	GO:0006821-chloride transport	3	2.027027027	0.042123688	0.994757536
	GOTERM_BP_DIRECT	GO:0007010-cytoskeleton organization	5	3.378378378	0.042936221	0.994757536
	GOTERM_BP_DIRECT	GO:0006685-sphingomyelin catabolic process	2	1.351351351	0.047971221	0.994757536
	GOTERM_BP_DIRECT	GO:0050428-3'-phosphoadenosine 5'-phosphosulfate biosynthetic process	2	1.351351351	0.047971221	0.994757536
	GOTERM_BP_DIRECT	GO:0071280-cellular response to copper ion	2	1.351351351	0.047971221	0.994757536
	GOTERM_CC_DIRECT	GO:0070062-extracellular exosome	54	36.48648649	2.56E-10	3.63E-08
	GOTERM_CC_DIRECT	GO:0016324-apical plasma membrane	14	9.459459459	5.46E-07	3.88E-05
	GOTERM_CC_DIRECT	GO:0031526-brush border membrane	6	4.054054054	5.04E-05	0.002384104
	GOTERM_CC_DIRECT	GO:0005615-extracellular space	23	15.54054054	8.90E-04	0.031609785
	GOTERM_CC_DIRECT	GO:0005903-brush border	5	3.378378378	0.001293643	0.036739466
	GOTERM_CC_DIRECT	GO:0005887-integral component of plasma membrane	15	15.54054054	0.001679058	0.039737704
	GOTERM_CC_DIRECT	GO:0005886-plasma membrane	47	31.75675676	0.006156794	0.124894957
	GOTERM_CC_DIRECT	GO:0045177-apical part of cell	4	2.702702703	0.021730409	0.331949844
	GOTERM_CC_DIRECT	GO:0005785-lysosomal membrane	7	4.72972973	0.021816788	0.331949844
	GOTERM_CC_DIRECT	GO:0044214-spanning component of plasma membrane	2	1.351351351	0.023519469	0.331949844
	GOTERM_CC_DIRECT	GO:0048471-perinuclear region of cytoplasm	11	7.432432432	0.025714425	0.331949844
	GOTERM_CC_DIRECT	GO:0030658-transport vesicle membrane	3	2.027027027	0.036211022	0.41350654
	GOTERM_CC_DIRECT	GO:0016021-integral component of membrane	52	35.13513514	0.037856233	0.41350654
Down-regulated	GOTERM_MF_DIRECT	GO:0004048-carbonate dehydratase activity	5	3.378378378	4.02E-06	0.001269138
	GOTERM_MF_DIRECT	GO:0048038-quinone binding	4	2.702702703	1.82E-04	0.019001371
	GOTERM_MF_DIRECT	GO:0008270-zinc ion binding	23	15.54054054	1.99E-04	0.019001371
	GOTERM_MF_DIRECT	GO:0016491-oxidoreductase activity	9	6.081081081	2.41E-04	0.019001371
	GOTERM_MF_DIRECT	GO:0005254-chloride channel activity	5	3.378378378	9.85E-04	0.062255297
	GOTERM_MF_DIRECT	GO:0017080-sodium channel regulator activity	4	2.702702703	0.00228976	0.117392741
	GOTERM_MF_DIRECT	GO:0051287-NAD binding	4	2.702702703	0.003658155	0.165139552
	GOTERM_MF_DIRECT	GO:0005179-hormone activity	5	3.378378378	0.007391889	0.291979628
	GOTERM_MF_DIRECT	GO:0004745-retinol dehydrogenase activity	3	2.027027027	0.009316935	0.32712793
	GOTERM_MF_DIRECT	GO:0005200-structural constituent of cytoskeleton	5	3.378378378	0.012667668	0.4002983
	GOTERM_MF_DIRECT	GO:0050660-flavin adenine dinucleotide binding	4	2.702702703	0.015493251	0.44507885
	GOTERM_MF_DIRECT	GO:0005215-transporter activity	6	4.054054054	0.025223419	0.664216701
	GOTERM_MF_DIRECT	GO:0009055-electron carrier activity	4	2.702702703	0.037523049	0.912098738
	GOTERM_MF_DIRECT	GO:0004064-arylesterase activity	2	1.351351351	0.048064632	0.980466331
	GOTERM_MF_DIRECT	GO:0008131-primary amine oxidase activity	2	1.351351351	0.048064632	0.980466331
	GOTERM_BP_DIRECT	GO:0032869-cellular response to insulin stimulus	3	13.04347826	0.004517664	0.941659992
	GOTERM_BP_DIRECT	GO:0071711-basement membrane organization	2	8.695652174	0.010435411	0.941659992
	GOTERM_BP_DIRECT	GO:0046653-tetrahydrofolate metabolic process	2	8.695652174	0.013027975	0.941659992
	GOTERM_BP_DIRECT	GO:0033628-regulation of cell adhesion mediated by integrin	2	8.695652174	0.014321825	0.941659992
	GOTERM_BP_DIRECT	GO:0008284-positive regulation of cell proliferation	4	17.39130435	0.022099566	1
	GOTERM_BP_DIRECT	GO:0046655-folic acid metabolic process	2	8.695652174	0.023333544	1
	GOTERM_BP_DIRECT	GO:0030198-extracellular matrix organization	3	13.04347826	0.026863138	1
	GOTERM_BP_DIRECT	GO:0006730-one-carbon metabolic process	2	8.695652174	0.038599552	1
	GOTERM_BP_DIRECT	GO:0006865-amino acid transport	2	8.695652174	0.044892384	1
	GOTERM_CC_DIRECT	GO:0005829-cytosol	9	39.13043478	0.033945931	1
	GOTERM_CC_DIRECT	GO:0005576-extracellular region	6	26.08695652	0.039561198	1
	GOTERM_CC_DIRECT	GO:0031012-extracellular matrix	3	13.04347826	0.049060058	1

KEGG analysis results were shown in Table 3 which demonstrated that up-regulated DEGs were particularly enriched in Metabolic pathways, Mineral absorption, Nitrogen metabolism and Pancreatic secretion while down-regulated DEGs in One carbon pool by folate.

Table 3 KEGG pathway analysis of differentially expressed genes in colorectal cancer

Expression	Pathway ID and Name	Count	%	p-Value	FDR
Up-regulated	hsa00910:Nitrogen metabolism	5	3.378378378	3.78E-05	0.004682299
	hsa01100:Metabolic pathways	30	20.27027027	7.26E-05	0.004682299
	hsa04978:Mineral absorption	6	4.054054054	1.51E-04	0.006512746
	hsa04964:Proximal tubule bicarbonate reclamation	4	2.702702703	0.002350437	0.075801581
	hsa00920:Sulfur metabolism	3	2.027027027	0.004672819	0.12055873
	hsa00040:Pentose and glucuronate interconversions	4	2.702702703	0.006654704	0.143076129
	hsa00071:Fatty acid degradation	4	2.702702703	0.012982141	0.23924232
	hsa00983:Drug metabolism - other enzymes	4	2.702702703	0.016600662	0.267685676
	hsa04972:Pancreatic secretion	5	3.378378378	0.023626408	0.338645174
	hsa00561:Glycerolipid metabolism	4	2.702702703	0.030529984	0.393836791
	hsa00830:Retinol metabolism	4	2.702702703	0.039221226	0.443770023
	hsa00982:Drug metabolism - cytochrome P450	4	2.702702703	0.045641408	0.443770023
	hsa04976:Bile secretion	4	2.702702703	0.047323285	0.443770023
	hsa00052:Galactose metabolism	3	2.027027027	0.048161088	0.443770023
Down-regulated	hsa00670:One carbon pool by folate	2	8.695652174	0.031543112	1

Protein–protein interaction network (PPI)

A total of 171 DEGs were imported into the DEGs PPI network complex which included 170 nodes and 204 edges, including 23 down-regulated and 143 up-regulated genes (Fig.2a). Then we applied STRING analysis and results showed that 87 central nodes were identified among the 160 nodes (Fig.2b).

Analysis of core genes by the Kaplan Meier plotter and GEPIA

Kaplan Meier plotter (<http://kmplot.com/analysis>) was utilized to identify 88 core genes survival data. It was found that 30 genes had a significantly worse survival while 58 had no significant ($P < 0.05$, Table 4 & Fig. 3). Then, GEPIA was used to dig up the 30 gene expression level between cancerous and normal people. Results reported that 13 of 30 genes reflected high expressed in CRC samples contrasted to normal samples ($P < 0.05$, Table 5 & Fig. 4).

Table 4 The prognostic information of the 88 key candidate genes

Category	Genes
Genes with significantly worse survival (P < 0.05)	MYC CEMIP MTHFD2 SCD TGFBI PLCE1 HSD17B2 ZG16 TRPM6 CES2 AHCYL2 FGFR3 BMP2 SLC22A5 KRT20 TST CDA SI PHLPP2 AKR1B10 TMPRSS2 DHRS9 MGLL MYO1A SLC36A1 NXPE4 SLC1A1 ANPEP CEACAM7 ALPI
Genes without significantly worse survival (P > 0.05)	TRIB3 CDC25B CTPS1 COL4A1 SHMT2 PLAU SLC7A5 FBL PYY MAOA PDZD3 GNA11 CDHR5 KLF4 AQP8 SGK1 ACADVL LGALS4 ADH1C SGK2 UGT2A3 SST SCN1B MEP1A ACAA2 CA12 GUCA2A CA4 NAAA MS4A12 SLC4A4 SQRDL DEFB1 SLC26A3 CHGA ABCB1 PDE9A EDN3 ITPKA PCK1 CASP7 ACADS PNLIPRP2 CA7 ENTPD5 GUCA2B CLCA4 UGDH FABP1 ETFDH MXI1 CA2 GCG UGP2 VIPR1 CA1 RETSAT CDHR2

Table 5 Validation of 30 genes via GEPIA

Category	Genes
Genes with high expressed in COAD (P < 0.05)	CEMIP CDA FGFR3 KRT20 MTHFD2 MYC MYO1A PLCE1 SCD TGFBI TMPRSS2 TST ZG16
Genes without high expressed in COAD (P > 0.05)	HSD17B2 TRPM6 CES2 AHCYL2 BMP2 SLC22A5 SI PHLPP2 AKR1B10 DHRS9 MGLL SLC36A1 NXPE4 SLC1A1 ANPEP CEACAM7 ALPI

Re-analysis of 13 selected genes via KEGG pathway enrichment

To understand the possible pathway of these 13 selected DEGs, KEGG pathway enrichment was re-analyzed via DAVID (P<0.05). Results showed that two genes(MYC,FGFR3) markedly enriched in the Bladder cancer pathway (P = 0.041, Table 6 & Fig. 5).

Table 6 Re-analysis of 13 selected genes via KEGG pathway enrichment

Pathway ID and Name	Count	%	p-Value	FDR	Genes
hsa05219:Bladder cancer	2	15.3846153846153	0.0410001159215224	1	MYC, FGFR3

To identify more useful prognostic biomarkers in CRC cancer, this study used bioinformatical methods on the basis of four profile datasets (GSE74602, GSE110223, GSE113513 and GSE141174). Sixty-five CRC specimens and normal specimens were enrolled in the present research. Via GEO2R and Venn software, we revealed a total of 171 commonly changed DEGs ($|\log FC| > 2$ and adjusted P value < 0.05) including 148 up-regulated ($\log FC > 0$) and 23 down-regulated DEGs ($\log FC < 0$). Then, Gene Ontology and Pathway Enrichment Analysis using DAVID methods showed that up-regulated DEGs were particularly enriched in oxidation-reduction process, in extracellular exosome, in zinc ion binding, in Metabolic pathways, Mineral absorption; and down-regulated DEGs in positive regulation of cell proliferation, in cytosol, in One carbon pool by folate. Next, DEGs PPI network complex of 170 nodes and 204 edges was constructed via the STRING online database. Furthermore, for the analysis of overall survival among those genes, Kaplan–Meier analysis was implemented and 30 of 88 genes had a significantly worse prognosis. For validation in Gene Expression Profiling Interactive Analysis (GEPIA), 13 of 30 genes were discovered highly expressed in CRC tissues compared to normal tissues. Furthermore, two genes (MYC, FGFR3) markedly enriched in the Bladder cancer pathway. In conclusion, we have identified two significant up-regulated DEGs with poor prognosis in CRC on the basis of integrated bioinformatical methods, which could be potential therapeutic targets for CRC patients.

MYC (oncogene) is highly expressed in colorectal cancer tissues, and is widely activated in colorectal cancer. It is involved in the regulation of growth, invasion and metastasis of colorectal cancer, and has been used as a research target for anti-tumor therapy for many years. Guo^[15] found that polyamine biosynthesis is often disordered in colorectal cancer, and there is a close relationship between polyamine metabolism pathway and oncogenic signaling pathway in the process of tumor development. Inhibition of SMS and Myc simultaneously has synergistic effects, and combined inhibition of SMS and Myc expression may provide a new therapeutic target and therapeutic strategy for the treatment of colorectal cancer. Bian^[16] shows that MYC family oncogene and DNA damage repair protein PARP 1 play an important role in the development and development of small cell lung cancer. MYC family genes are amplified or highly expressed in some small cell lung cancer cell lines and patient tissues. Combined inhibition of PARP and BET can effectively inhibit the proliferation and survival of MYC family gene-dependent small cell lung cancer tumor tissue. Fibroblast growth factor receptor (FGFR) is a receptor protein binding to Fibroblast growth factor family member protein ligand (FGF). A member of the tyrosine kinase family, it is involved in cell proliferation, stem cell differentiation, embryo development, migration, survival, angiogenesis and organogenesis through multiple signaling pathways. Studies have shown that the activation of FGFR signaling pathway is related to the occurrence and development of a variety of cancers. FGFR2 can also be phosphorylated, but the way and sequence of its phosphorylation and activation remain unclear. The phosphorylation of FGFR3 and FGFR4 is different from that of FGFR1.

Zhao^[17] have shown that effective components from the rhizome of *R. sinicum* inhibit the proliferation of colorectal cancer cells by targeting FGFR1 and down-regulating the expressions of p-JAK2, p-STAT3 and p-MEK1/2, and inhibit the growth of human colorectal cancer xenograft tumors with high FGFR1 expression. Fibroblast growth factor receptor 3 (FGFR3). Wang^[18] have shown that miR-99a-SP has target

gene binding sites with downstream FGFR3. MiR-99a-sp mimics significantly down-regulated FGFR3 after overloading; MiR-99a-SP can negatively regulate FGFR3, and the proliferation, metastasis and invasion of cancer cells are all decreased after targeted knockdown of FGFR3. Numerous studies have proved that MYC and FGFR genes were related to various types of cancer's progression. In this study, MYC and FGFR genes were also enriched in bladder cancer, and the mechanism of action of the two genes was also studied. MYC and FGFR genes, as important indicators for detection, could be used as effective targets for diagnosis and treatment of digestive tract tumors.

Conclusions

Taken above, our bioinformatics analysis study identified MYC and FGFR genes between CRC tissues and normal tissues on the base of four different microarray datasets. Results showed that MYC and FGFR genes could play key roles in the progression of CRC. However, these predictions should be verified by a series of experiments in the future. Anyway, these data may provide some useful information and direction into the potential bio-markers and biological mechanisms of CRC and other cancer.

Declarations

Funding: No funding was received to assist with the preparation of this manuscript.

Conflicts of interest: The authors have no relevant financial or non-financial interests to disclose.

Availability of data and material: All data generated or analysed during this study are included in this published article [and its supplementary information files].

Code availability: Not applicable.

Authors' contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Shaodong Li], [Ruizhi Dong], [Bin Liang] and [Zhenhua Kang]. The first draft of the manuscript was written by [Shaodong Li] proofreading and reviewing was finished by [Zhenhua Kang] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics approval : Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

References

1. BRAY F, FERLAY J, SOERJOMATARAM I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: a cancer journal for

- clinicians, 2018, 68(6):394-424.
2. CHEN W, ZHENG R, BAADE P D, et al. Cancer statistics in China, 2015 [J].CA: a cancer journal for clinicians, 2016, 66(2): 115-32.
 3. Szasz AM, Lanczky A, Nagy A, et al. Crossvalidation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget*. 2016;7(31):49322–33.
 4. Guilford P J , Holyoake A J . Markers for detection of gastric cancer[J]. 2007.
 5. Shiotani A , Iishi H , Uedo N , et al. Histologic and serum risk markers for noncardia early gastric cancer.[J]. *International Journal of Cancer*, 2005, 115(3):463-469.
 6. Duffy M J , Lamerz R , Haglund C , et al. Tumor markers in colorectal cancer, gastric cancer and gastrointestinal stromal cancers: European group on tumor markers 2014 guidelines update[J]. *International Journal of Cancer*, 2014, 134(11):2513-2522.
 7. Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, et al. Stromal contribution to the colorectal cancer transcriptome. *Nat Genet*. 2015;47(4):312–9.
 8. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23(14):1846–7.
 9. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25–9.
 10. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
 11. Huang d W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
 12. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(Database issue):D447–52.
 13. Szasz AM, Lanczky A, Nagy A, Forster S, Hark K, Green JE, et al. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget*. 2016;7(31):49322–33.
 14. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):w98–w102.
 15. Guo Y , Ye Q , Deng P , et al. Spermine synthase and MYC cooperate to maintain colorectal cancer cell survival by repressing Bim expression[J]. *Nature Communications*.
 16. Bian X , Wang X , Zhang Q , et al. The MYC Paralog-PARP1 Axis as a Potential Therapeutic Target in MYC Paralog-Activated Small Cell Lung Cancer[J]. *Frontiers in Oncology*, 2020, 10.
 17. Zhao R , Bu Y C , Wei L , et al. Acetylshikonin suppressed growth of colorectal tumour tissue and cells by inhibiting the intracellular kinase, T-lymphokine-activated killer cell-originated protein kinase. [J]. *British Journal of Pharmacology*, 2020, 177.

18. X Wang, Yin Z , Zhao Y , et al. Identifying potential prognostic biomarkers in head and neck cancer based on the analysis of microRNA expression profiles in TCGA database[J]. Molecular Medicine Reports, 2020, 21(3).

Figures

Figure 1

Authentication of 171 common DEGs in the four datasets (GSE74602, GSE110223, GSE113513 and GSE 141174) through Venn diagrams software(available online: <http://bioinformatics.psb.ugent.be/webtools/Venn/>). Different color meant different datasets. 148 DEGs were up-regulated in the four datasets ($\log FC > 0$); 23 DEGs were down-regulated in the four datasets ($\log FC < 0$).

Figure 2

Common DEGs PPI network constructed by STRING online database.aThere were a total of 171 DEGs in the DEGs PPI network complex. The nodes meant proteins; the edges meant the interaction of proteins.b Potential central correlation between these DEGs via STRING.

Figure 3

The prognostic information of the 88 core genes. Kaplan meier plotter online tools were used to identify the prognositc information of the 88 core genes and 30 of 88 genes had a significantly worse survival rate ($P < 0.05$)

Figure 4

Significantly expressed 13 genes in CRC patients compared to healthy people. To further identify the genes' expression level between CRC and normal people, 13 genes which were related with poor prognosis were analyzed by GEPIA website. 13 of 30 genes had significant expression level in CRC specimen compared to normal specimen (*P < 0.05). Red color means tumor tissues and grey color means normal tissues

Figure 5

Re-analysis of 13 selected genes by KEGGpathway enrichment. Two high expressed genes in CRC tissues with poor prognosis were re-analyzed by KEGG pathway enrichment.