Identification of novel therapeutic target genes in colon cancer by integrative analysis

Meidan Fang[1,3], Xin Jin[2], Zhenxiang Pan[3], Xianquan An[3], Qianchuang Sun[3], Zhengyang Lu[1], Shuang Wang[3], Tongjun Liu[1*]

[1]Department of General Surgery, Second Hospital of Jilin University, 218 Ziqiang Street, Changchun 130041, China

[2]Department of Oncology and Hematology, Second Hospital of Jilin University, 218 Ziqiang Street, Changchun 130041, China

[3]Department of Anesthesiology, Second Hospital of Jilin University, 218 Ziqiang Street, Changchun 130041, China

*** Correspondence:** Tongjun Liu, Emails: tongjunliu@163.com

Abstract

Colon cancer is one of the most common tumors and have a has a high fatality rate in United States. Both genetic and epigenetic factors contribute to colon cancer. Nowadays, gene expression profiles with microarrays of colon cancer are emerging in GEO database and TCGA database. Though some studies have revealed promising colorectal carcinoma-associated genes or gene signatures, the results of these data may have discrepancies due to sample size or heterogeneity. Robust rank aggregation (RRA) is an unbiased bioinformatics method to integrate individual microarray profiles and can reveal a relatively accurate gene expression signatures based on individual microarray profiles. In the present studies, we performed RRA to integrate microarray profiles of colon cancer in GEO database. We found 105 differentially expressed genes between colon cancer tissue and normal adjacent mucosa. The results of GO function and KEGG pathways enrichment analysis showed that IL-17 signaling and TNF signaling pathway are the most upreguluated pathways during colonic carcinogenesis. PPI network

analysis of top 20 differentially expressed genes indicated that upregulated genes MS4A12, CLCA4, GUCA2B and downregulated genes GDF15, CXCL8 and MMPs played a significant role in colonic carcinogenesis. Our study would pave a way to explore therapeutic target genes and genetic mechanism of colon cancer.

Colorectal cancer is the third leading morbidity of all cases of cancer in United States in both sexes and lists top three in cases of malignant tumor death[1, 2]. To date, surgery, radiotherapy and chemotherapy are the predominant modalities to treat colon cancer. However, some patients presented in advanced stage disease at the time of initial diagnosis and were not sensitive to radiotherapy and chemotherapy[3-7]. Moreover, high rate of relapse and metastasis challenging the therapy of colon cancer. Colon cancer results from both genetic and epigenetic factors. Nowadays, gene expression analysis by microarray has been popularized to explore the abnormal alterations of genes in colon cancer. Gene expression profiles with microarrays of colon cancer are emerging in GEO database and TCGA database and some studies have revealed promising colorectal carcinoma-associated genes or gene signatures. However, there are discrepancies in these profiles and some gene expression signatures have shown thousands of abnormal alterations of genes due to a small sample size. Thus, the obtained results from individual microarray assays correspondingly lack credibility. In order to overcome these limitations,

we performed a robust rank aggregation (RRA) approach to integrate individual microarray profiles of colon cancer in GEO database[8]. RRA is an unbiased integrated bioinformatics method to obtain a relatively accurate gene expression signatures based on individual microarray profiles. Our study would help to unravel novel therapeutic target genes in colon cancer and prioritize the putative targets for genome research of the development of colon cancer.

Materials and Methods

GEO database was searched for human colon cancer gene expression profiling studies by microarrays which data had been uploaded prior to February 10st,2019. We screened the studies which analyzed the comparison between human colon primary cancer tissues and paired normal adjacent mucosa. The studies profiled cell lines or preselected candidate genes only were excluded. Nonhuman studies were also excluded. In the targeted datasets, gene names were normalized by gene symbols. First, we use GEO2R, an online tool designed by GEO database, to perform the gene expression comparisons in individual datasets. All genes expression fold changes and adjust p value by Bonferroni correction were listed in a new table for nest analysis. A gene with absolute value of $\log_2$ FC >1 and adjust p value <0.05 was considered differentially expressed gene.

Datasets construction and statistical analysis

The extracted genes were ranked based on fold changes (FC) value. The absolute value of $\log_2 FC$ >1 was considered as upregulated and downregulated gene. The upregulated and downregulated gene data were analyzed separately. We conducted the RRA approach using an R package RobustRankAggreg to rank the genes consistently. The RRA approach normalized upregulated and downregulated matrixes on the robust rank aggregation algorithm. Acquired p values of each gene from each dataset were averaged and exported adjust p values by Bonferroni correction. The adjust p values of genes in the normalized matrixes less than 0.05 were considered statistically significant and listed to a new table with a rank of maximal possibility. The results were visualized as a heatmap.

Enrichment analysis

Enrichment analyses for KEGG pathways and Gene Ontology terms were carried out with R package Colorspace, Stringi and Bioclite.

Analysis of protein–protein interaction (PPI) network

To determine the function of the proteins that differentially expressed genes encoded, PPI network of these genes were conducted by an online tool (String, https://string-db.org/) and visualized by Cytoscape software. The PPI network identified for the differentially expressed genes was screened at a genome-wide scale, with both end nodes having

these genes. The network construction using methods based on genomic context and structure information.

Results

Selection of microarray datasets

We selected four microarray datasets retrieved in GEO database according to our methods, including GSE74604, GSE10950, GSE41328, GSE44861. The 4 datasets provided the gene expression profiles on both human colon primary cancer tissues and paired normal adjacent mucosa. The details of 4 selected datasets are shown in Table 1. Thus, there are total 120 colon cancer tissues and 119 normal adjacent colon tissues included in the integrative analysis.

Table1 Characteristic of included microarray datasets

GSE ID	Platform	Cancer	Normal	Sample
GSE10950	GPL6104	24	24	Tissue
GSE41328	GPL570	10	10	Tissue
GSE44861	GPL3921	56	55	Tissue
GSE74604	GPL6104	30	30	Tissue

Identification of differentially expressed genes for the 4 datasets

The gene names in 4 matrixes are all named by gene symbols. The details of analysis of differentially expressed genes in 4 datasets are shown in Table 2. There are great disparities in the analysis results of 4 datasets. After RRA analysis, there are 54 upregulated and 51 downregulated differentially expressed genes depending on adjust p value and FC level (Table 3). The expression data of top 20 upregulated and downregulated genes are visualized by heatmap (Fig. 1).

Table2 Analysis of differentially expressed genes in 4 datasets

GSE ID	Total	downregulated	upregulated
GSE10950	4256	2108	2148
GSE41328	1090	500	590
GSE44861	345	199	146
GSE74604	1523	850	673

Table3 Total differentially expressed genes by RRA analysis

Down gene	adjPvalue	logFC	Up gene	adjPvalue	logFC

AQP8	4.57E-08	4.72720943	CDH3	3.58E-08	-3.67986752
GUCA2B	1.53E-07	4.421364355	MMP1	4.57E-08	-3.748842023
MS4A12	6.49E-07	4.353688153	CEMIP	8.75E-08	-4.130847765
CA4	1.14E-06	3.880860245	MMP7	1.06E-07	-3.691066865
GUCA2A	1.15E-06	4.30961205	REG1A	9.19E-07	-3.436978545
CLCA4	2.45E-06	4.194705725	COL11A1	7.57E-06	-3.289600543
ABCA8	1.55E-05	3.027552125	CXCL8	9.17E-06	-3.50585515
VIP	0.000156227	3.25807052	DPEP1	1.17E-05	-3.27317428
MT1M	0.000204194	3.239553328	REG1B	2.13E-05	-2.880870738
ZG16	0.000247505	3.37732275	NFE2L3	4.84E-05	-2.439834425
CHGA	0.00033234	3.323546673	MMP3	5.25E-05	-3.232301755
CHP2	0.000392983	3.230416283	SLCO4A1	5.92E-05	-2.43565447
CA2	0.000451822	3.023311263	NEBL	7.44E-05	-2.46451018
C7	0.000524849	3.100203123	TRIB3	0.000121178	-2.554282583
HSD17B2	0.000586543	2.628305953	FOXQ1	0.000143249	-3.6556851
SRPX	0.000774801	2.072468748	GDF15	0.000198381	-2.178515845
KRT24	0.001147198	2.62911698	CLDN1	0.000208512	-3.767214238
STMN2	0.001147198	2.09422074	SERPINB5	0.000269605	-2.286883628
CD177	0.00178474	2.750098913	RNF43	0.000435733	-2.120888255
AKR1B10	0.001978663	2.87357106	VSNL1	0.000456856	-2.327361893
SLC26A3	0.002003198	2.464498488	CXCL1	0.000501395	-2.771952085
SCNN1B	0.002864129	2.702405648	TESC	0.000574214	-2.55776469
CEACAM7	0.003141314	2.626836815	KRT6B	0.001235405	-2.090268745

SLC26A2	0.003370948	2.452674105	ASCL2	0.001524251	-2.899765445
SCGN	0.003459233	2.341081023	MMP11	0.002003198	-2.46003824
ANPEP	0.005298331	2.818059513	KRT23	0.002148091	-3.088601435
LRRC19	0.006673337	1.755399385	TCN1	0.002538314	-2.58469684
SPIB	0.007930513	2.05418281	CLDN2	0.003370948	-2.568057325
TMEM100	0.008022327	1.9262158	PHLDA1	0.00713921	-2.347639063
MYOT	0.01164761	2.244524825	CEL	0.007659795	-1.91199531
CCDC69	0.013315938	1.771146475	CKAP2	0.008114936	-1.768637828
PYY	0.013587683	2.143855983	GTF2IRD1	0.009085758	-1.819794498
ITM2A	0.015157376	1.632515515	MMP10	0.009923106	-2.245710883
TUBAL3	0.015158313	2.127066783	PUS7	0.011526124	-1.732081048
DPT	0.018355172	2.442392265	ENC1	0.013725105	-1.828605248
LGALS2	0.020688916	2.442078413	CXCL2	0.013914152	-2.08948386
ADTRP	0.021643267	2.11327697	LCN2	0.014003073	-1.71664808
TNFRSF17	0.02263026	2.088488285	LRP8	0.017512367	-2.089614165
CWH43	0.022831641	2.133663915	CCL20	0.017512367	-1.958783235
CLEC3B	0.028842817	2.29524777	IL11	0.022231499	-1.697978553
CHGB	0.029769525	1.67734968	SOX9	0.024279186	-2.141384858
GCG	0.030901285	2.149127873	COL10A1	0.026922705	-2.518080643
NXPE4	0.032312537	2.426322073	COL1A1	0.027871751	-1.765750748
SCARA5	0.032952304	2.355340375	LY6G6D	0.029836154	-1.91370104
BCHE	0.036142253	2.321360205	SLC7A5	0.029836154	-1.984627793
DHRS11	0.036277938	2.017980565	SLC12A2	0.03051606	-1.74024574

MT1E	0.039994177	2.378484588	INHBA	0.030901285	-1.64504031
PLAC8	0.0410497	2.335502728	GGH	0.041554743	-1.527724548
MT1H	0.041239039	1.737616973	FABP6	0.045148419	-2.119505288
TSPAN1	0.042304596	1.417806193	MYC	0.047204188	-1.810236035
ADAMTS1	0.043160539	1.41711367	PPAT	0.049690431	-1.559291
CFD	0.043584777	2.52026848			
BEST2	0.044890493	2.134921558			
FABP1	0.044890493	1.629451835			

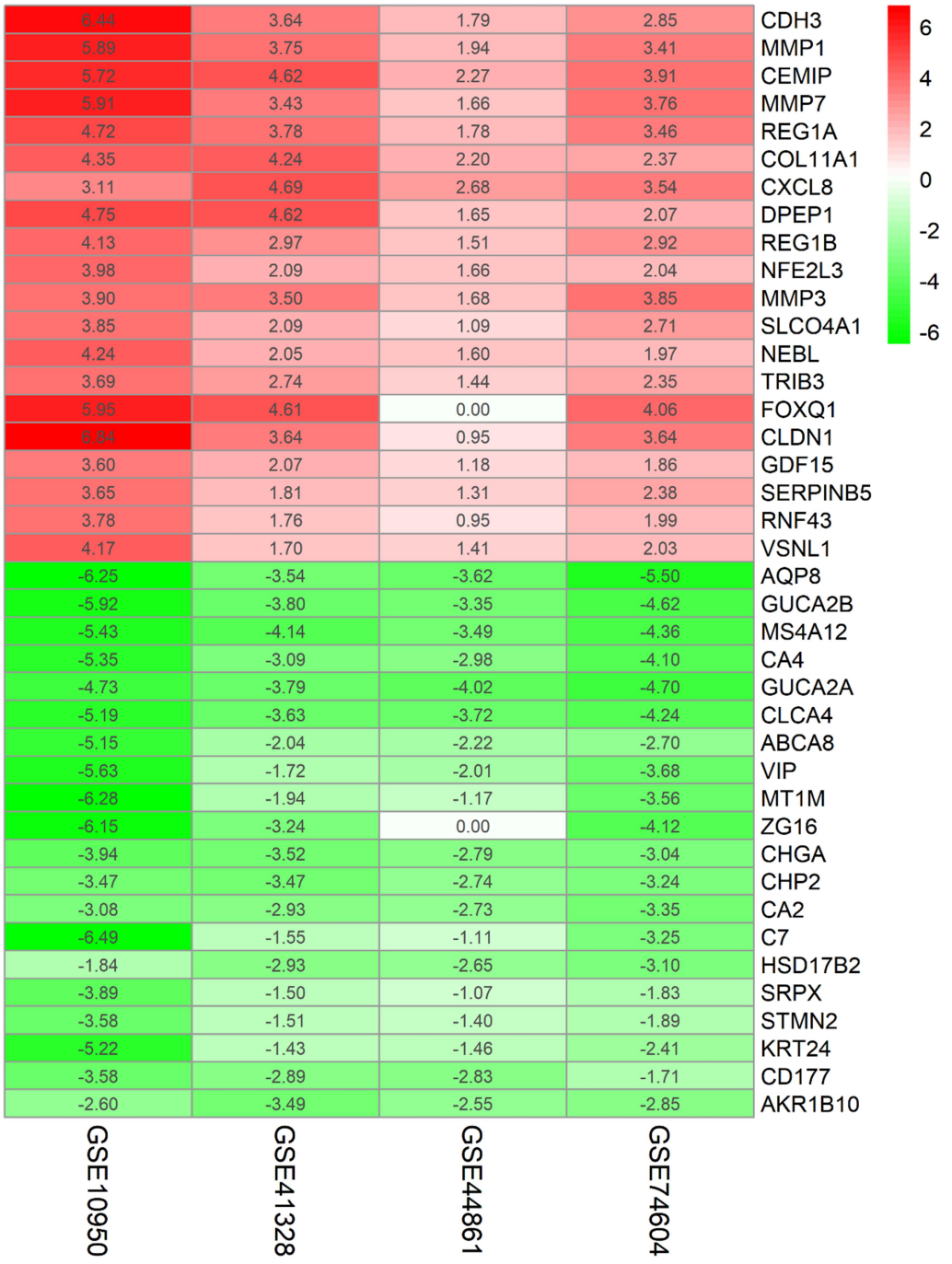

Figure 1 Top 20 differentially expressed genes in 4 datasets

Functional annotation and KEGG pathway for differentially expressed

genes

To understand the roles of these genes further, we performed GO function enrichment and KEGG pathway enrichment of all differentially expressed genes. The results are shown in Figure 2 and Figure 3. In downregulated genes, enriched GO terms mainly include hormone related activities, transmembrane transporter activities and energy metabolism. Enriched KEGG pathways in downregulated genes cover mineral absorption, nitrogen metabolism, proximal tubule bicarbonate reclamation, pancreatic secretion. In upregulated genes, enriched GO terms mainly include the cytokine, chemokine and growth factor activities and their receptor binding, metallopeptidase and metalloendopeptidase activities, serine involved activities, and extracellular matrix structural constituent conferring tensile strength. Enriched KEGG pathways in upregulated genes include IL-17 signaling pathway, cytokine-cytokine receptor interaction, bladder cancer, TNF signaling pathway, protein digestion and absorption, chemokine signaling pathway.

PPI network analysis of differentially expressed genes

The PPI network for the total and top 20 differentially expressed genes with significant interaction relation are shown in Figure 4 and

Figure 5, respectively. The downregulated and upregulated genes are marked by blue and yellow color orderly.

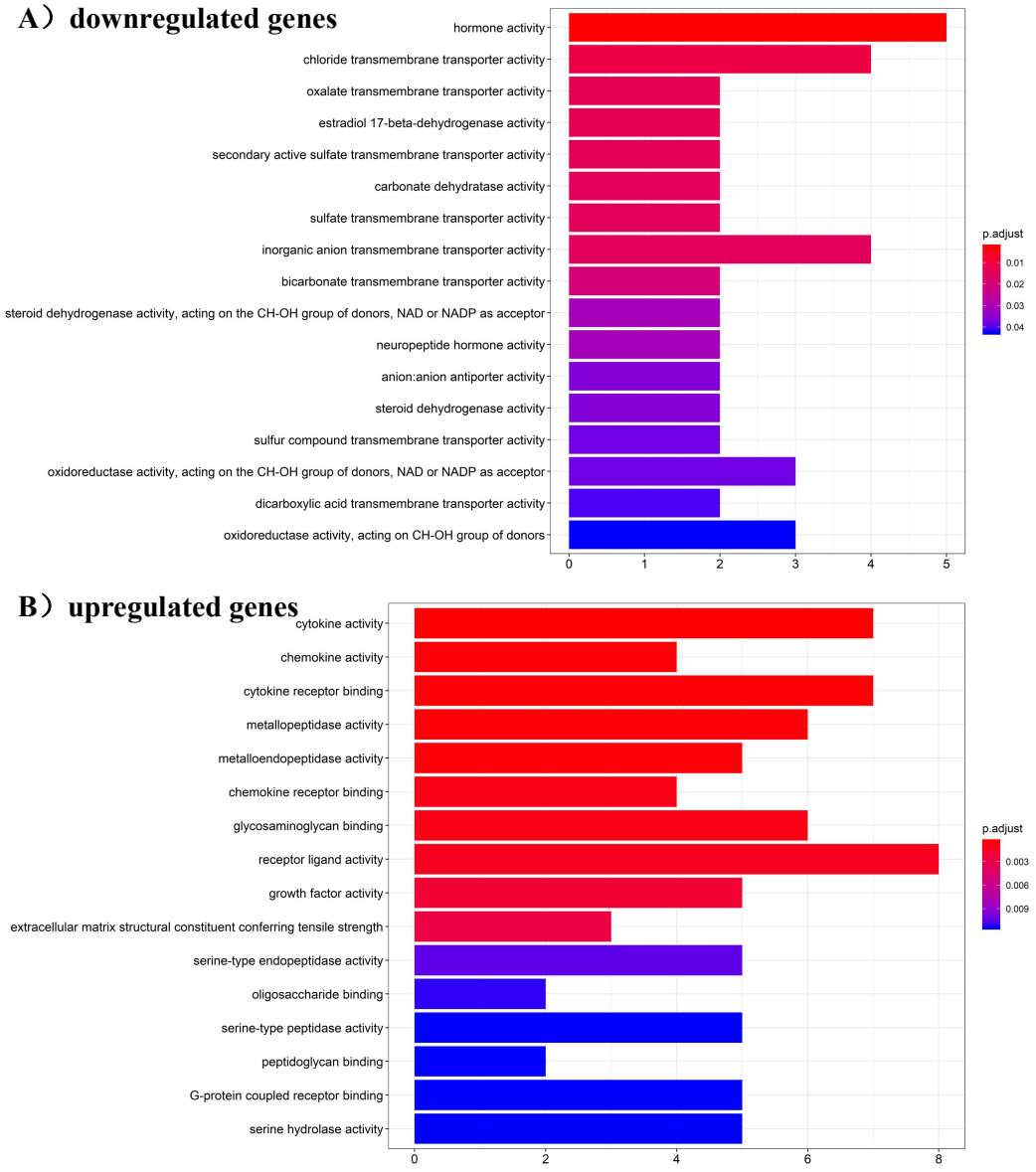

Figure 2 GO function enrichment of differentially expressed genes

A）downregulated genes

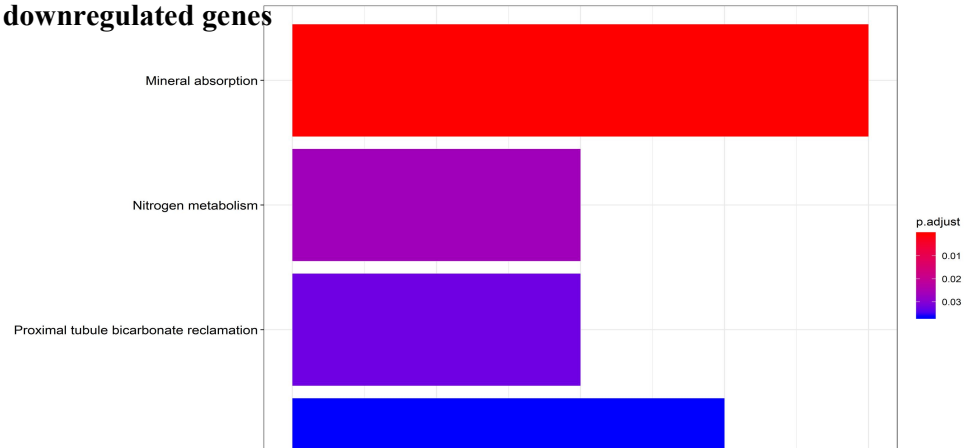

Figure 4 PPI network of total differentially expressed genes

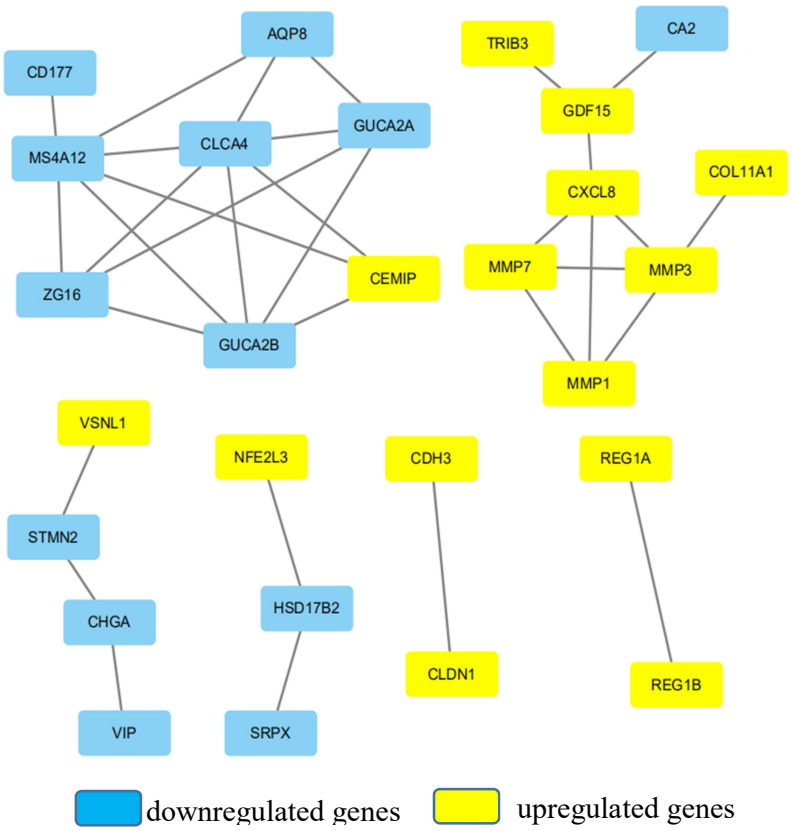

Figure 5 PPI network of top 20 differentially expressed genes

Discussion

There are huge discrepancies between the results from analysis of single microarray profile. Generally, the differentially expressed genes from a large samples dataset are more persuasive. Thus, integrative analysis of microarray profiles is a easy and significant approach to enlarge samples. In our results, the number of differentially expressed genes are more less than individual analysis, and it make easier to further analysis. 51 up-regulated genes and 54 down-regulated are extracted from the integrative analysis. In the heatmap, expression of top 20 differentially genes are shown in all datasets. Except FOXQ1 and ZG16 in GSE44861, expression of the other genes are consistent in the 4 datasets. Therefore our approach and result of integrative analysis is credible. These differentially expressed genes involved in various biological processes and pathways according to the GO function and KEGG pathways enrichment analysis. Totally, these genes indicate a hyperactivity of proliferation and hypoactivity of reabsorption of nutrients. In the results, we found that IL-17 signaling and TNF signaling pathway are the most upreguluated pathways during colonic

carcinogenesis. In many tumors, IL-17 and TNF-α are co-expressed by T helper 17 (TH17) cells. TNF-α can enhance mRNA and protein of PD-L1 in colon cancer cell lines, and IL-17 also increases the expression level of PD-L1 in colon cancer cell lines[9]. PD-L1 plays a significant role in tumor immune escape. To know these genes better, we conduct PPI network analysis. The results in top 20 differentially expressed genes indicated that upregulated genes MS4A12, CLCA4, GUCA2B and downregulated genes GDF15, CXCL8 and MMPs played a significant role in colonic carcinogenesis.

In colonic carcinogenesis, MS4A12, as a colon-specific gene, participates in proliferation and chemotaxis mediated by epidermal growth factor (EGF)[10, 11] and regulate the differentiation[10, 12]. Decreased expression of MS4A12 inhibits differentiation and indicates a poor survival in colon cancer[12, 13]. CLCA4 is a member of the calcium sensitive chloride conductance family of proteins. CLCA4 can inhibit cell proliferation, migration, and invasion in many cancer through suppressing PI3K/AKT signaling[14-17]. There is a lack of evidence of the effect of CLCA4 on colon cancer. According to studies of CLCA4, we can infer that decreased expression of CLCA4 is a promoting factor in colonic carcinogenesis. GUCA2B encodes uroguanylin and regulates proliferation, metabolism and barrier function in colon via binding and activating GUCY2C [18]. A lower expression GUCA2B may also promote

colon cancer. Overexpression GDF15 was proved to promote EMT and metastasis in colorectal cancer via TGFβ/Smad2,3 signaling[19]. Increased expression of CXCL8 can enhance cell proliferation, migration and invasion of colon cancer through PI3K/Akt/NF-κB signaling[20]. MMPs (MMP1, MMP3, MMP7) can also promote colonic carcinogenesis[21-31]. These results demonstrated the crucial function of the differentially expressed genes in colonic carcinogenesis.

In summary, our integrative analysis provided a useful method to mine the data from GEO database and revealed a well understanding of the molecular mechanism in colonic carcinogenesis. The pathways and genes we shown may be potential therapeutic targets and paved a way for further studies of colon cancer. However, the molecular mechanisms of cancer are complicated and there are some limits in our method. Therefore further studies remain necessary to reveal the molecular mechanism of colonic carcinogenesis.

Conflict of interest

All authors declared that there were no conflicts of interest.

References

[1] Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017. CA Cancer J Clin. 2017. 67(3):

177-193.

[2] Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin. 2017. 67(1): 7-30.

[3] Cascinu S, Poli D, Zaniboni A, et al. The prognostic impact of primary tumour location in patients with stage II and stage III colon cancer receiving adjuvant therapy. A GISCAD analysis from three large randomised trials. Eur J Cancer. 2019. 111: 1-7.

[4] Bregni G, Rebuzzi SE, Sobrero A. The Optimal Duration of Adjuvant Therapy for Stage III Colon Cancer: the European Perspective. Curr Treat Options Oncol. 2019. 20(1): 8.

[5] Leijssen L, Dinaux AM, Amri R, Kunitake H, Bordeianou LG, Berger DL. The Impact of a Multivisceral Resection and Adjuvant Therapy in Locally Advanced Colon Cancer. J Gastrointest Surg. 2019. 23(2): 357-366.

[6] Venook AP. Advances in Adjuvant Therapy for Colon Cancer: P value or Practical Value. J Clin Oncol. 2018. 36(15): 1461-1462.

[7] Ilson DH. Adjuvant therapy in colon cancer: less is more. Lancet Oncol. 2018. 19(4): 442-443.

[8] Võsa U, Kolde R, Vilo J, Metspalu A, Annilo T. Comprehensive meta-analysis of microRNA expression using a robust rank aggregation approach. Methods Mol Biol. 2014. 1182: 361-73.

[9] Wang X, Yang L, Huang F, et al. Inflammatory cytokines IL-17 and TNF-α up-regulate PD-L1 expression in human prostate and colon cancer cells. Immunol Lett. 2017. 184: 7-14.

[10] Koslowski M, Türeci O, Huber C, Sahin U. Selective activation of tumor growth-promoting Ca2+ channel MS4A12 in colon cancer by caudal type homeobox transcription factor CDX2. Mol Cancer. 2009. 8: 77.

[11] Koslowski M, Sahin U, Dhaene K, Huber C, Türeci O. MS4A12 is a colon-selective store-operated calcium channel promoting malignant cell processes. Cancer Res. 2008. 68(9): 3458-66.

[12] He L, Deng HY, C Wang X. Decreased expression of MS4A12 inhibits differentiation and predicts early stage survival in colon cancer. Neoplasma. 2017. 64(1): 65-73.

[13] Dalerba P, Kalisky T, Sahoo D, et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol. 2011. 29(12): 1120-7.

[14] Liu Z, Chen M, Xie LK, et al. CLCA4 inhibits cell proliferation and invasion of hepatocellular carcinoma by suppressing epithelial-mesenchymal transition via PI3K/AKT signaling. Aging (Albany NY). 2018. 10(10): 2570-2584.

[15] Yu Y, Walia V, Elble RC. Loss of CLCA4 promotes epithelial-to-mesenchymal transition in breast cancer cells. PLoS One. 2013. 8(12): e83943.

[16] Bundela S, Sharma A, Bisen PS. Potential therapeutic targets for oral cancer: ADM, TP53, EGFR, LYN, CTLA4, SKIL, CTGF, CD70. PLoS One. 2014. 9(7): e102610.

[17] Hou T, Zhou L, Wang L, Kazobinka G, Zhang X, Chen Z. CLCA4 inhibits bladder cancer cell proliferation, migration, and invasion by suppressing the PI3K/AKT pathway. Oncotarget. 2017. 8(54): 93001-93013.

[18] Pattison AM, Merlino DJ, Blomain ES, Waldman SA. Guanylyl cyclase C signaling axis and colon cancer prevention. World J Gastroenterol. 2016. 22(36): 8070-7.

[19] Li C, Wang J, Kong J, et al. GDF15 promotes EMT and metastasis in colorectal cancer. Oncotarget. 2016. 7(1): 860-72.

[20] Shen T, Yang Z, Cheng X, et al. CXCL8 induces epithelial-mesenchymal transition in colon cancer cells via the PI3K/Akt/NF-κB signaling pathway. Oncol Rep. 2017. 37(4): 2095-2100.

[21] Felton J, Hu S, Raufman JP. Targeting M3 Muscarinic Receptors for Colon Cancer Therapy. Curr Mol Pharmacol. 2018. 11(3): 184-190.

[22] Said AH, Hu S, Abutaleb A, et al. Interacting post-muscarinic receptor signaling pathways potentiate matrix metalloproteinase-1 expression and invasion of human colon cancer cells. Biochem J. 2017. 474(5): 647-665.

[23] Davudian S, Shajari N, Kazemi T, et al. BACH1 silencing by siRNA inhibits migration of HT-29 colon cancer cells through reduction of metastasis-related genes. Biomed Pharmacother. 2016. 84: 191-198.

[24] Pietruszewska W, Bojanowska-Poźniak K, Kobos J. Matrix metalloproteinases MMP1, MMP2, MMP9 and their tissue inhibitors TIMP1, TIMP2, TIMP3 in head and neck cancer: an immunohistochemical study. Otolaryngol Pol. 2016. 70(3): 32-43.

[25] Okusha Y, Eguchi T, Sogawa C, et al. The intranuclear PEX domain of MMP involves proliferation, migration, and metastasis of aggressive adenocarcinoma cells. J Cell Biochem. 2018. 119(9): 7363-7376.

[26] Křivohlavá R, Grobárová V, Neuhöferová E, Fišerová A, Benson V. Interaction of colon cancer cells with glycoconjugates triggers complex changes in gene expression, glucose transporters and cell invasion. Mol Med Rep. 2018. 17(4): 5508-5517.

[27] Jagadish N, Parashar D, Gupta N, et al. A-kinase anchor protein 4 (AKAP4) a promising therapeutic target of colorectal cancer. J Exp Clin Cancer Res. 2015. 34: 142.

[28] Izadi F. Differential Connectivity in Colorectal Cancer Gene Expression Network. Iran Biomed J. 2019. 23(1): 34-46.

[29] Trumpi K, Frenkel N, Peters T, et al. Macrophages induce "budding" in aggressive human colon cancer subtypes by protease-mediated disruption of tight junctions. Oncotarget. 2018. 9(28): 19490-19507.

[30] Ertem FU, Zhang W, Chang K, et al. Oncogenic targets Mmp7, S100a9, Nppb and Aldh1a3 from transcriptome profiling of FAP and Pirc adenomas are downregulated in response to tumor suppression by Clotam. Int J Cancer. 2017. 140(2): 460-468.

[31] Klupp F, Neumann L, Kahlert C, et al. Serum MMP7, MMP10 and MMP12 level as negative prognostic markers in colon cancer patients. BMC Cancer. 2016. 16: 494.