Int J Clin Exp Med 2016;9(12):23198-23205
www.ijcem.com /ISSN:1940-5901/IJCEM0034116

*Original Article*

# Identification of potential biomarkers in pathogenesis and liver metastasis of colorectal cancer: a bioinformatics study

Xi Ma, Huiqun Sun, Jingnan Fu, Xing Gao, Xiangyu Li

*Department of Anorectal Surgery, The Affiliated Hospital of Logistics University of Chinese People's Armed Police Forces, Tianjin, People's Republic of China*

Received June 20, 2016; Accepted August 15, 2016; Epub December 15, 2016; Published December 30, 2016

**Abstract:** Colorectal cancer (CRC) is the third leading cancer-related death cause worldwide and liver is one of its common metastasis sites. The purpose of this study was to explore the mechanisms of its development and identify some potential therapeutic targets. Gene expression dataset GSE41258 was downloaded from the Gene Expression Omnibus database. Differentially expressed genes (DEGs) in primary CRC samples compared with normal mucosae and DEGs in liver metastatic CRC samples compared with the primary CRC samples were obtained through limma package implemented in *R*. GO and pathway analysis of DEGs were conducted in the Database for Annotation, Visualization and Integrated Discovery. Protein-protein interaction (PPI) and miRNA-Gene regulation network were constructed based on the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) and the TargetScan database respectively and visualized via the Cytoscape software. A total of 1048 DEGs, which contained 365 up- and 683 down-regulated ones, and 364 DEGs, which contained 150 up- and 214 down-regulated ones, were obtained in primary and liver metastatic CRC samples respectively. GO and pathway analysis indicated that processes that related to cell activity were enriched in DEGs of primary CRC samples, while for DEGs of liver metastatic CRC samples, substance activity or binding were found to significantly enriched. Besides, the overlaps between those two lists of DEGs were mainly involved in the GO or pathways related to cell activity, substance binding and immune system. What's more, some proved and novel biomarkers, such as *MEIS1, TCF21, hsa-miR-142-5p,* were screened out in our study, which could improve our understanding about the progression of CRC. Through bioinformatics, we could obtain some potential biomarkers and processes involved in the incidence and development of CRC, which would be helpful in its studies and treatment.

**Keywords:** Colorectal cancer, gene expression omnibus, STRING, targetScan, biomarkers

## Introduction

Colorectal cancer (CRC) is one of the most common cancers with high morbidity and mortality. In western countries, a large part of populations suffer from it and 10 percent of those individuals will develop into malignant tumor [1]. Like other solid tumors, distant metastatic is one of its main features contribute its poor prognosis and conjunction with lungs, liver are the most commonly involved sites [2]. Some therapeutics has been proposed for liver-metastasis CRC, such as surgical resection, chemotherapy, radiotherapy and etc, but it is still defined as an incurable disease in the present moment [3]. So it becomes particularly important for its early diagnostic and the development of new therapeutics.

In the past few decades, lots of molecular studies were carried out and some factors have been shown to contribute to the presence of liver metastasis of CRC and its poor prognosis. In the study of Nakayama et al, real time PCR was conducted and copy number variant of PRL-3 was identified as a representation of poor prognosis of CRC and might be associated with its liver metastasis [4]. Through microRNA (miRNA) microarray, Perilli *et al* screened out some differentially expressed miRNAs in liver metastatic CRC samples compared with the normal ones, which could be helpful for its diagnosis and treatment [5].

For different CRC patients with liver metastasis, the same therapeutics might result in distinct outcome. So it is important for the selection of

different therapeutics or drugs for different patients, i.e. personalized medical. The first we should do is to identify the molecular targets which could significantly affect the prognosis of the therapeutics. Until now, some factors have been obtained that could play important roles in the treatment of CRC. In the study of JA et al, the expression of *AURKA, PTGS2* and *MMP9* were identified as prognostic marker in liver metastatic CRC samples and could be helpful in the selection of patients who would be benefit from adjuvant systemic therapy [6]. Through the analysis of aCGH microarray, Bruin et al inferred that the copy number variation of *ERC1* and *ERC2* could be used as indication of prognosis of hepatic resection in liver metastatic CRC patients.

The rapid growth of high-throughput technologies, such as gene microarray and next generation sequencing (NGS), make it possible to profile the expression of thousands of genes simultaneously and many studies screened out some potential therapeutic targets for lots of diseases (particular tumors) based on them. In this study, gene expression analysis about CRC with and without liver metastasis, which based on gene microarray was carried out. Some core genes that might play important roles in the progression of CRC were identified and the involved biological process and pathways were obtained. Besides, the interactions among the candidate CRC related genes, as well as the miRNA-Gene regulation pairs were identified, which would be helpful in the early diagnosis and treatment of CRC.

## Material and methods

### *Gene expression dataset*

The microarray dataset GSE41258 was downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/), which deposited by Sheffer *et al*, [7] contained gene expression data from colon cancer patients with or without liver or lung metastasis and their corresponding normal mucosae at Memorial Sloan-Kettering Cancer Center. The Affymetrix Human Genome U133A Array platform was used for the detection of gene expression profile.

### *Microarray data preprocessing*

The raw CEL microarray data were imported into R and affy [8] package was used for the background correction and robust multi array normalization. Affymetrix probe IDs were converted to official gene symbol and the expression value was summarized for genes corresponding to multi probes.

### *Identification of differentially expressed genes*

Here, limma [9] package was used for the identification of differentially expressed genes (DEGs). Student's t-test was conducted on the expression value between control and case samples. Besides, Bonferroni & Hochberg (BH) method was used for the correction of *P* value. Genes which satisfied the criteria of adjusted *P* value less than 0.05 and fold change bigger than 2 or less than 0.5, i.e. |log2 (fold change)|>1, were screened out.

### *Functional and pathway enrichment analysis*

To identify the involved functions and pathways of DEGs, we uploaded them to the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/) [10] and all of the human genes were used as the background gene list. Here, Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with *P* value less than 0.05 were screened out as the involved functions and pathways of DEGs.

### *Construction of protein-protein interaction network*

In most of the human phenotypes, genes tend to function together, rather than individual. So it is important to explore the interactions among genes in diseases. In this study, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) [11] database, which manipulates the interactions among genes or proteins from multiple sources, such as high-throughput experiments, bioinformatics prediction and etc, was used for the identification of interactions among DEGs. Combined score (a representation of reliability of interactions) bigger than 0.4 was used as the threshold for the selection of interactions and protein-protein interaction (PPI) network was constructed via Cytoscape [12] software.

### *Construction of miRNA-Gene regulation network*

As a kind of small non-coding RNA, miRNA plays important roles in the regulation of gene

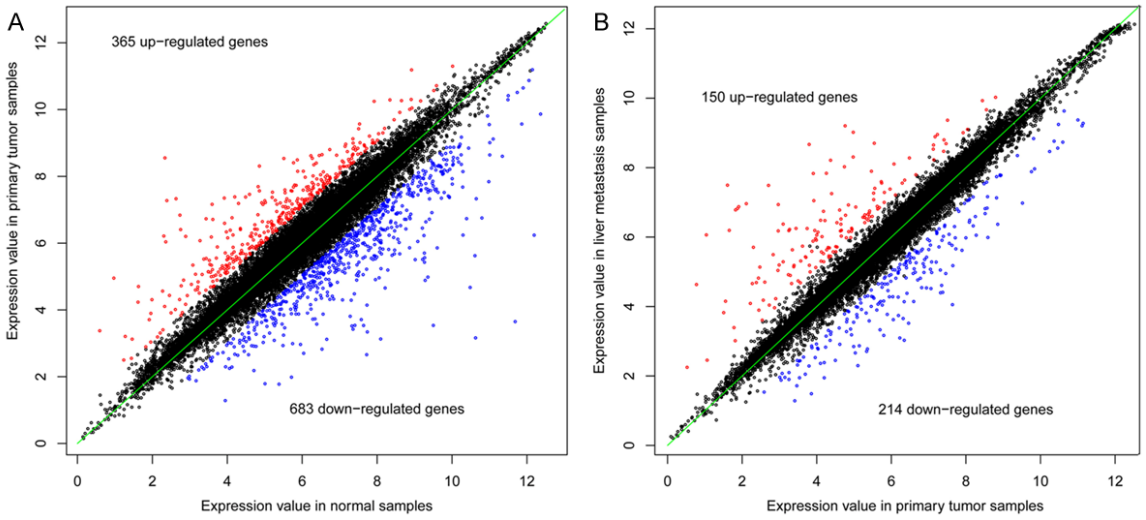


**Figure 1.** DEGs in primary and liver metastatic CRC samples. A: DEGs in primary CRC samples compared with the corresponding normal mucosae. B: DEGs in liver metastatic CRC samples compared with primary CRC ones.

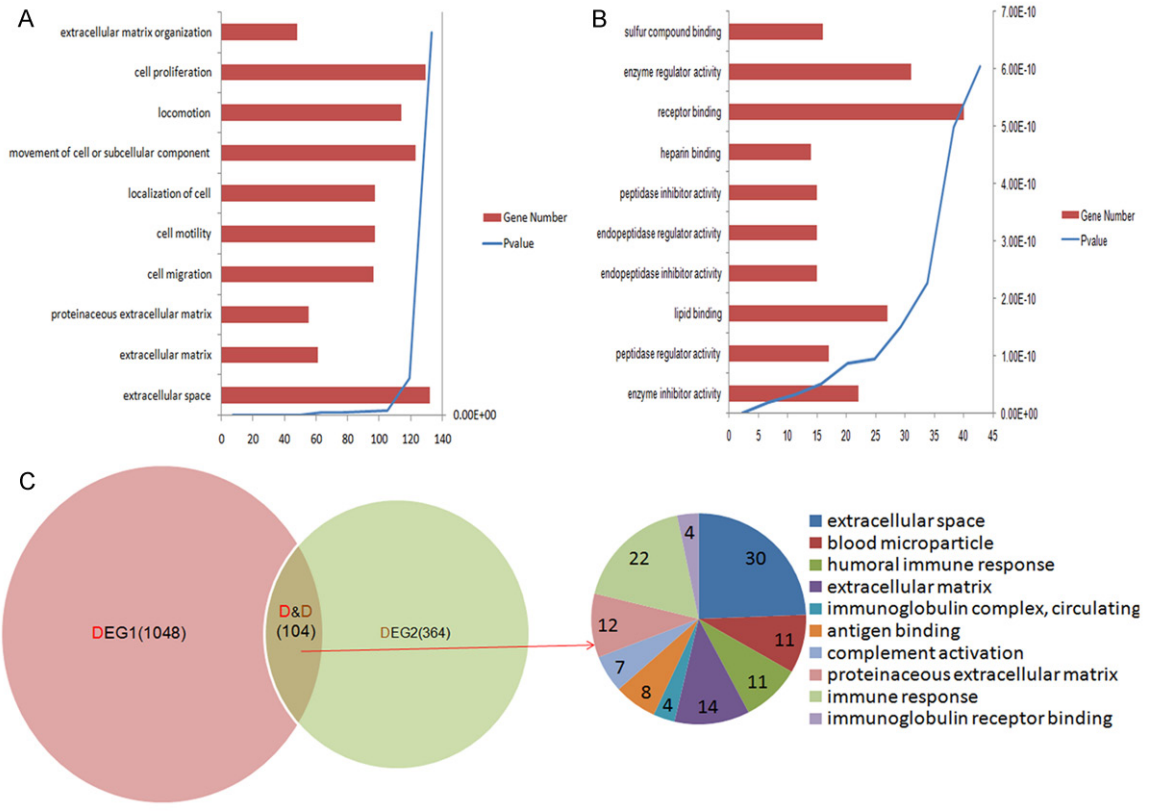


**Figure 2.** Top 10 most significantly enriched GO terms. A: Top 10 most significantly enriched GO terms of DEGs in primary CRC samples. B: Top 10 most significantly enriched GO terms of DEGs in liver metastatic CRC samples. C: Top 10 most significantly enriched GO terms of the overlaps between the two lists of DEGs.

expression through repressing translation or degrading protein at post-translation stage. In this study, Target Scan [13] database was used to identify the potential miRNAs that might regulate DEGs and miRNA-Gene regulation network was built based on Cytoscape software.

**Table 1.** Enriched KEGG pathways of the overlapped genes

| Pathway Name | Gene Number | *P* value |
|---|---|---|
| PPAR signaling pathway | 5 | $2.77\times10^{-5}$ |
| Pancreatic secretion | 5 | $1.35\times10^{-4}$ |
| Fat digestion and absorption | 3 | $1.24\times10^{-3}$ |
| Rheumatoid arthritis | 4 | $1.36\times10^{-3}$ |
| Proximal tubule bicarbonate reclamation | 6 | $6.27\times10^{-3}$ |
| Amoebiasis | 3 | $1.90\times10^{-2}$ |
| TNF signaling pathway | 3 | $1.95\times10^{-2}$ |
| Vascular smooth muscle contraction | 3 | $2.50\times10^{-2}$ |
| Hedgehog signaling pathway | 5 | $2.88\times10^{-2}$ |
| Basal cell carcinoma | 4 | $3.31\times10^{-2}$ |
| Wnt signaling pathway | 3 | $3.56\times10^{-2}$ |
| Creatine pathway | 5 | $3.59\times10^{-2}$ |
| Synthesis and degradation of ketone bodies | 3 | $4.59\times10^{-2}$ |
| Glycolysis/Gluconeogenesis | 4 | $4.62\times10^{-2}$ |
| Complement and coagulation cascades | 4 | $5.00\times10^{-2}$ |

KEGG, Kyoto Encyclopedia of Genes and Genomes.

## Results

*Differentially expressed genes*

In this study, 1048 DEGs (365 up- and 683 down-regulated) in primary CRC samples compared with normal mucosae and 364 DEGs (150 up- and 214 down-regulated) in CRC samples with liver metastasis compared with primary CRC samples were obtained. The numbers of up- and down-regulated genes in the two lists of DEGs were shown in **Figure 1A** and **1B**.

*Enriched GO terms and KEGG pathways*

Here, a total of 126 and 85 enriched GO terms were identified for DEGs in primary and metastatic CRC samples. Besides, we obtained 56 enriched GO terms for the overlaps between the two lists of DEGs. The top 10 most significantly enriched GO terms according to *P* value were shown in **Figure 2**. Besides, **Table 1** demonstrated the 15 enriched KEGG pathways of the 104 overlapped DEGs.

*Protein-protein network*

Based on the STRING database, a total of 764 interactions were obtained among the 104 overlapped DEGs and the PPI network was shown in **Figure 3**. In the PPI network, some core genes (genes with more direct interactions in the network) were identified, such as *CAV1*, *SPP1* and etc, which might be potential therapeutic targets in the progression of CRC.

*miRNA-Gene regulation network*

A total of 258 miRNAs which possible regulate the 104 overlapped DEGs were got and 496 miRNA-Gene regulation pairs were identified among those miRNAs and genes (**Figure 4**). What's more, some genes were found to have more direct neighbors in both PPI and miRNA-Gene regulation networks. **Table 2** showed the genes which have at least 5 direct neighbors in both PPI and miRNA-Gene networks. As an example, the expression values of *MEIS1* in primary, metastatic CRC samples and their corresponding mucosa were shown in **Figure 5**.

## Discussion

CRC is a common malignant tumor with high mortality and morbidity and liver is one of the most frequent metastatic sites. In this study, through the analysis of gene expression datasets in primary, metastatic CRC samples and their corresponding mucosa, we identified the candidate genes which might contribute to the progression of CRC. Function and pathway enrichment analysis indicated processes related to cell activity, immune system and etc were enriched in those candidate genes. What's more, the analysis of PPI and miRNA-Gene regulation networks got some potential molecular targets which might play important roles in the pathogenesis and metastasis of CRC.

Since its appearance, gene microarray has been applied in many studies to explore the mechanisms and potential biomarkers of many diseases, including CRC. Through the combined analysis of published microarray studies, Zhang et al screened out 8 candidate genes which including *RPN2*, *HMGB1*, *AARS*, *IGFBP3*, *STAT1*, *HYOU1*, *NQO1* and *PEA15* were significantly associated with the pathogenesis and development of CRC [14]. In the study of Akram *et al*, via the microarray study followed by RT-PCR, *CCR1* and *CCRL2* were identified play important roles in CRC

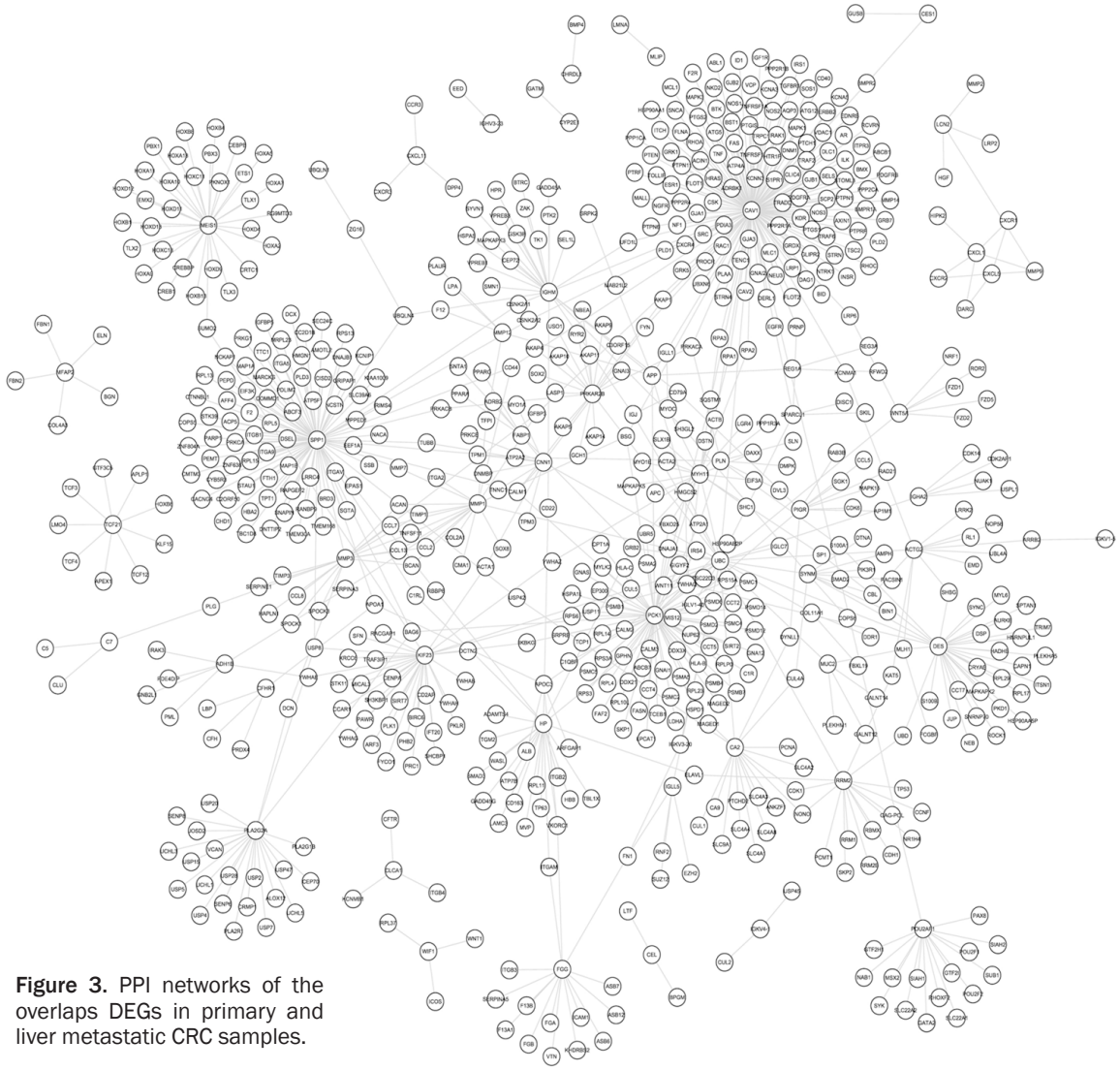

**Figure 3.** PPI networks of the overlaps DEGs in primary and liver metastatic CRC samples.

liver metastasis [15]. In this study, reanalysis of published microarray about the incidence and metastasis of CRC, several potential therapeutic targets were identified. *MEIS1* is one of the biomarkers we obtained and encodes a homebox protein which has been proved to be involved in neoplasia. Here, it directly interacts with 32 other genes in PPI network and targeted by 64 miRNAs in miRNA-Gene regulation network. Besides, **Figure 5** showed that the expression values of *MEIS1* were significantly decreased with the progression of CRC. So it might be an important cancer suppressor gene in CRC and play important roles in the CRC liver metastasis. Consistent with our study, Crist *et al* also verified its decreased expression values in CRC samples compared with the corresponding normal ones [16]. Besides, through methylation-specific PCR, Dihal *et al* also identified the methylation sites of *MEIS1* in BRAF mutated colon tumors which could result in its down regulation [17]. Together with those two studies, our study further validated the roles of *MEIS1* in the progression of CRC.

Enrichment analysis obtained some enriched functions of DEGs in primary and liver metastatic CRC samples, as well as the overlaps. Some differences were identified, e.g. DEGs in primary CRC samples were mainly enriched for biological processes that related to cell activity, while DEGs in liver metastatic ones mainly involved in the substance activity or factor binding, such as sulfur compound binding, receptor binding, enzyme inhibitor binding and etc. This might explain the differences between the incidence and metastasis of

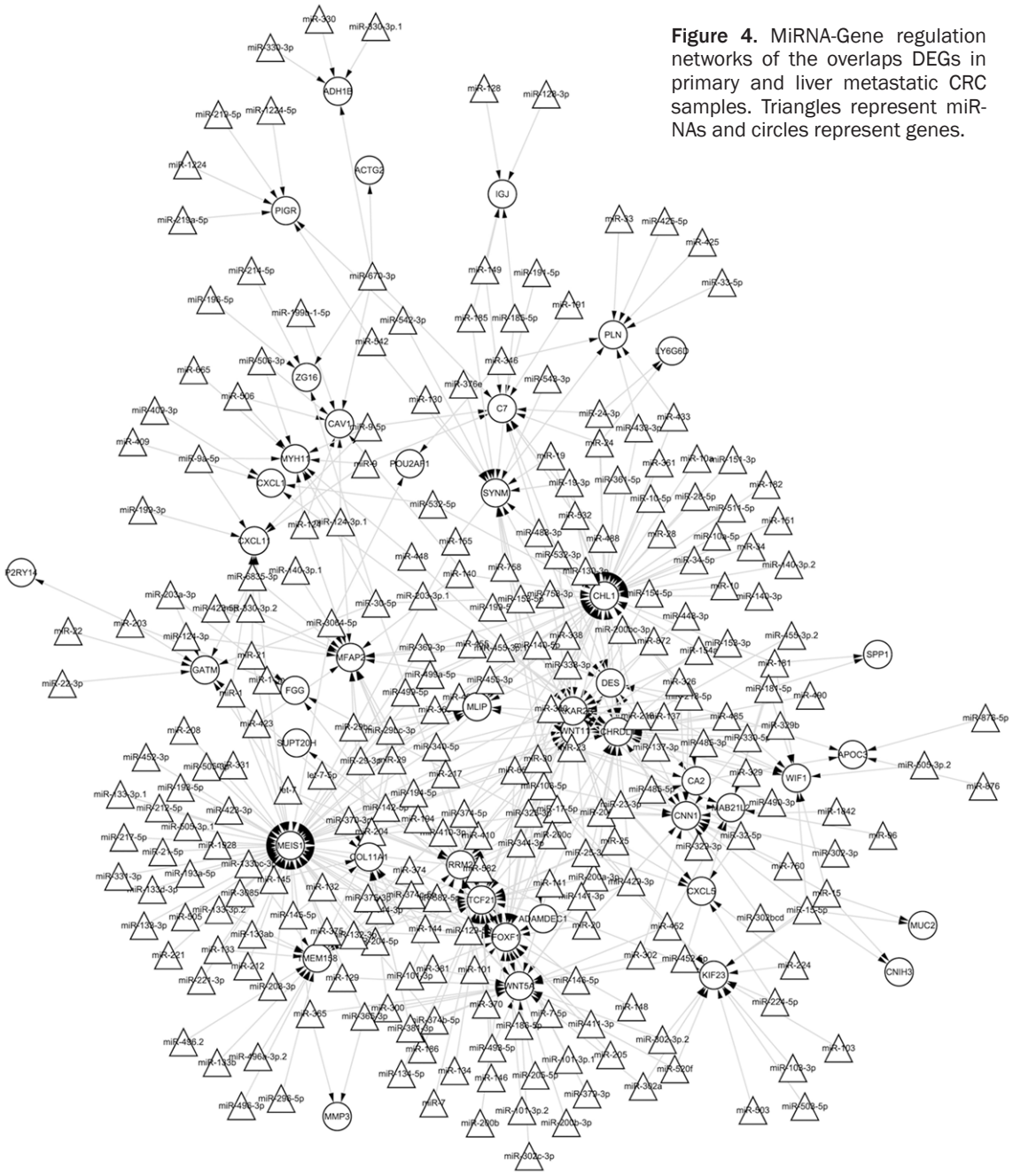

**Figure 4.** MiRNA-Gene regulation networks of the overlaps DEGs in primary and liver metastatic CRC samples. Triangles represent miRNAs and circles represent genes.

CRC to some extent. For the overlapped genes, we identified GO terms and pathways not only associated with cell and substance activity, but also immune and blood system, e.g. blood microparticle, humoral immune response, etc. This further illustrated the importance of immune and vascular in the progression of cancer including CRC, which have been fully proved in the previous study [18-20].

MiRNA is a kind of small non-coding RNA which could regulate gene expression in post-transcription through base-pairing with complementary sequences within mRNA. Here, we identified the miRNAs might regulate overlapped genes through TargetScan database, which manipulates the validated and predicted miRNA-Gene regulation pairs. Among which, *hsa-miR-142-5p* were found to regulate 6 overlapped genes. Some studies have focused on its roles in the development of multi cancers, including CRC. In the study of Chanda *et al*, *hsa-miR-142-5p* was considered to influence the development of CRC by

**Table 2.** Genes with degree bigger than 5 in both PPI and miRNA-Gene network

| Gene Name | Degree_miRNA_Gene | Degree_PPI |
|---|---|---|
| MEIS1 | 64 | 32 |
| TCF21 | 28 | 9 |
| PRKAR2B | 25 | 19 |
| WNT5A | 21 | 7 |
| MFAP2 | 18 | 5 |
| CNN1 | 16 | 12 |
| SYNM | 15 | 12 |
| KIF23 | 14 | 33 |
| RRM2 | 12 | 16 |
| MYH11 | 10 | 21 |
| DES | 9 | 31 |
| CAV1 | 7 | 143 |
| PIGR | 6 | 13 |

PPI, Protein-protein interaction. Note: Degree is the number of direct neighbors of genes in PPI and miRNA-Gene regulation network.

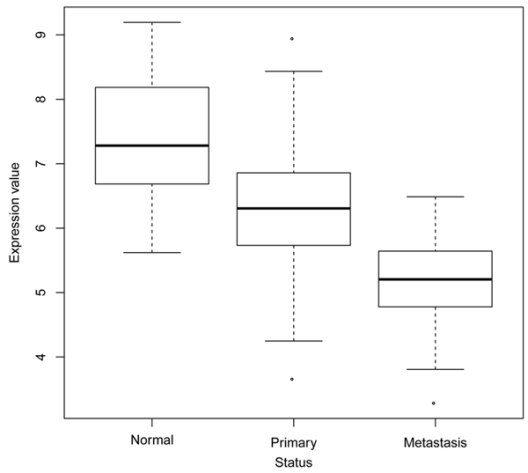


**Figure 5.** Expression values of MEIS1 in normal mucosae, primary and liver metastatic CRC samples.

targeting TGFβ signaling [21]. Besides, in our study it was identified to regulate *MEIS1*, which was proved to inhibit CRC progression in previous and our studies. In addition, *TCF21* is one of the targets of *hsa-miR-142-5p*, which has more direct neighbors in both PPI and miRNA-Gene regulation networks and it has also been identified as an important tumor suppressor gene through inhibiting cell proliferation, migration and invasion in CRC [22]. So *hsa-miR-142-5p* might act as a carcinogen in CRC and more studies would be preferred to illustrate its roles until the clinical application.

In conclusion, this study identified some potential biomarkers in the progression of CRC through the analysis of gene expression profiles. Besides, through GO and pathway enrichment analysis, we obtained some differences of processes take place in the incidence and liver metastasis of CRC, which could improve our understanding and facilitate its early diagnosis and treatment.

## Acknowledgements

We would like to thank all the members of our research group for their enthusiastic participation in this study.

## Disclosure of conflict of interest

None.

**Address correspondence to:** Dr. Xi Ma, Department of Anorectal Surgery, The Affiliated Hospital of Logistics University of Chinese People's Armed Police Forces, 220 Chenglin Road, Hedong District, Tianjin 300162, People's Republic of China. Tel: +8615510993958; E-mail: jinma8@sina.com

## References

[1] Kinzler KW and Vogelstein B. Lessons from hereditary colorectal cancer. Cell 1996; 87: 159-170.

[2] Vatandoust S, Price TJ and Karapetis CS. Colorectal cancer: Metastases to a single organ. World J Gastroenterol 2015; 21: 11767-11776.

[3] Costi R, Leonardi F, Zanoni D, Violi V and Roncoroni L. Palliative care and end-stage colorectal cancer management: the surgeon meets the oncologist. World J Gastroenterol 2014; 20: 7602-7621.

[4] Nakayama N, Yamashita K, Tanaka T, Kawamata H, Ooki A, Sato T, Nakamura T and Watanabe M. Genomic gain of the PRL-3 gene may represent poor prognosis of primary colorectal cancer, and associate with liver metastasis. Clin Exp Metastasis 2016; 33: 3-13.

[5] Perilli L, Pizzini S, Bisognin A, Mandruzzato S, Biasiolo M, Facciolli A, Rossi E, Esposito G, Rugge M, Pilati P, Mocellin S, Nitti D, Bortoluzzi S and Zanovello P. Human miRNome profiling in colorectal cancer and liver metastasis development. Genom Data 2014; 2: 184-188.

[6] Goos JA, Coupe VM, van de Wiel MA, Diosdado B, Delis-Van Diemen PM, Hiemstra AC, de Cuba EM, Belien JA, Menke-van der Houven van Oordt CW, Geldof AA, Meijer GA, Hoekstra OS, Fijneman RJ; DeCoDe PET Group. A prog-

nostic classifier for patients with colorectal cancer liver metastasis, based on AURKA, PTGS2 and MMP9. Oncotarget 2016; 7: 2123-34.
[7] Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F, Paty PB, Gerald WL, Notterman DA and Domany E. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proc Natl Acad Sci U S A 2009; 106: 7131-7136.
[8] Gautier L, Cope L, Bolstad BM and Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004; 20: 307-315.
[9] Diboun I, Wernisch L, Orengo CA and Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma. BMC Genomics 2006; 7: 252.
[10] Huang da W, Sherman BT and Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37: 1-13.
[11] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ and von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43: D447-452.
[12] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-2504.
[13] Shin C, Nam JW, Farh KK, Chiang HR, Shkumatava A and Bartel DP. Expanding the microRNA targeting code: functional sites with centered pairing. Mol Cell 2010; 38: 789-802.
[14] Nguyen MN, Choi TG, Nguyen DT, Kim JH, Jo YH, Shahid M, Akter S, Aryal SN, Yoo JY, Ahn YJ, Cho KM, Lee JS, Choe W, Kang I, Ha J and Kim SS. CRC-113 gene expression signature for predicting prognosis in patients with colorectal cancer. Oncotarget 2015; 6: 31674-31692.
[15] Akram IG, Georges R, Hielscher T, Adwan H and Berger MR. The chemokines CCR1 and CCRL2 have a role in colorectal cancer liver metastasis. Tumour Biol 2016; 37: 2461-71.
[16] Crist RC, Roth JJ, Waldman SA and Buchberg AM. A conserved tissue-specific homeodomain-less isoform of MEIS1 is downregulated in colorectal cancer. PLoS One 2011; 6: e23665.
[17] Dihal AA, Boot A, van Roon EH, Schrumpf M, Farina-Sarasqueta A, Fiocco M, Zeestraten EC, Kuppen PJ, Morreau H, van Wezel T and Boer JM. The homeobox gene MEIS1 is methylated in BRAF (p.V600E) mutated colon tumors. PLoS One 2013; 8: e79898.
[18] Quigley DA and Kristensen V. Predicting prognosis and therapeutic response from interactions between lymphocytes and tumor cells. Mol Oncol 2015; 9: 2054-62.
[19] Kaser SA, Mattiello D and Maurer CA. Distant Metastasis in Colorectal Cancer is a Risk Factor for Anastomotic Leakage. Ann Surg Oncol 2016; 23: 888-93.
[20] Maurel J and Postigo A. Prognostic and Predictive Biomarkers in Colorectal Cancer. From the Preclinical Setting to Clinical Practice. Curr Cancer Drug Targets 2015; 15: 703-715.
[21] Chanda S, Nandi S and Chawla-Sarkar M. Rotavirus induced miR-142-5p elicits proviral milieu by targeting Non-canonical TGFbeta signalling and apoptosis in cells. Cell Microbiol 2016; 18: 733-47.
[22] Dai Y, Duan H, Duan C, Zhou R, He Y, Tu Q and Shen L. Down-regulation of TCF21 by hypermethylation induces cell proliferation, migration and invasion in colorectal cancer. Biochem Biophys Res Commun 2016; 469: 430-6.