Comprehensive analysis of metastasis-related genes reveals a gene signature predicting the survival of colon cancer

Running title: Metastasis genes related to colon cancer

Haotang Wei [1#], Jilin Li [2#], Minzhi Xie [2], Ronger Lei [3], Bangli Hu [2]

1 Department of Gastrointestinal Surgery, the Third Affiliated Hospital of Guangxi Medical University, Nanning 530031, China

2 Department of Gastroenterology, the First Affiliated Hospital of Guangxi Medical University, Nanning 530021, China

3 Department of Research, Affiliated Tumor Hospital of Guangxi Medical University, Nanning 530021, China.

#These authors contributed equally to this work.

Corresponding author: Bang-li Hu

Department of Research,

Affiliated Tumor Hospital of Guangxi Medical University,

71 Hedi Road, Nanning 530021, China

Tel: +86-0771- 5321865; Fax +86-0771- 5321865

E-mail: gxhubl@163.com

Abstract

Objective The mechanism underlying colon cancer metastasis remains unclear. This study aimed to elucidate the genes alteration during the metastasis of colon cancer and identify genes that are crucial to the metastasis and survival of colon cancer patients. **Methods** The dataset of primary and metastasis tissue of colon cancer, and dataset of high and low metastasis capability of colon cancer cells were selected as training cohort, and the overlapped differentially expressed genes (DEGs) were screened from the training cohort. The functional enrichment analysis for the overlapped DEGs was performed. The prognostic value of overlapped DEGs was analyzed in TCGA dataset, and a gene signature was developed using genes that are related to the overall survival (OS). The prognostic value of the gene signature was further confirmed in a validation cohort. **Results** 184 overlapped DEGs were screened from the training cohort. Functional enrichment analysis revealed the significant gene functions and pathways of the overlapped DEGs. Four hub genes (OXCT1, ACTN4, IL-8, ITGA3) were identified using protein-protein network analysis. Six genes (ALDH2, NEDD9, FLNA, LBR, TWF1, SRSF1) were closely related to the OS of colon cancer patients. A gene signature was developed used these six genes based on their risk score, and the validation cohort indicated that the prognostic value of this gene signature was high in the prediction of colon cancer patients. **Conclusions** Our study demonstrates a gene profiles related to the metastasis of colon cancer, and identify a six-gene signature that acts as an independent biomarker on the prognosis of colon cancer.

Introduction

Colorectal cancer (CRC) is one of the leading malignant cancers in the world, and the colon cancer accounts for a large part of CRC (Siegel et al. 2017a; Siegel et al. 2017b). During the last three decades, with the development of new therapies, such as bevacizumab and anti-epidermal growth factor receptor antibody (Castro et al. 2013; Knijn et al. 2010), great improvement in survival has been achieved for colon cancer patients with localized- and regional disease. However, cancer metastasis remains one of the main causes that lead to death of patient. Compared with the early stage of colon cancer, the prognosis of patients with distant metastasis remained poor (Siegel et al. 2017b). Accumulation of genetic mutations is recognized as one of the important causes that results in the pathogenesis and progression of colon cancer (Sameer 2013). Hence, exploring the mechanisms of cancer metastasis and searching suitable predictors are crucial to the diagnosis and treatment of colon cancer.

Previously, the TNM stage and pathological characteristics of colon cancer are commonly used to predict the prognosis and facilitate treatment for colon cancer patients. But there are some limitations of these methods (Marzouk & Schofield 2011). Recently, several novel biomarkers have been tested with the aim to improve the prediction of therapeutic response and prognosis of colon cancer patients (Demirkol et al. 2017; Hu et al. 2014; Xu et al. 2017), which provide a lot of help in the diagnosis and treatment of colon cancer, but the results were inconsistent and need to further study.

Thus far, metastasis is a major factor for the poor prognosis of colon cancer patients, while liver is the most common organ where colon cancer cells metastasize, but the molecular mechanism underlying distant metastasis remains unclear. Therefore, a comprehensive analysis for the molecular alteration and identify prognostic indicator is key for the

management of colon cancer patients with distant metastasis. In this study, by using the colon cancer data from gene expression omnibus (GEO) and The Cancer Genome Atlas (TCGA), we analyzed the data of colon cancer in primary tumor samples and liver metastasis samples to unveil the genes that key to the development colon cancer metastasis and the potential prognostic indicators.

Materials and methods

Patient datasets

The colon cancer tissue and cells microarray data (GSE40367 (Roessler et al. 2015) and GSE2509 (Provenzani et al. 2006)) was retrieval and download from the GEO (http://www.ncbi.nlm.nih.gov/geo/) database in the National Center for Biotechnology Information (NCBI) as the training cohort. The GSE40367 dataset includes seven colon adenocarcinoma (COAD) with liver metastasis species and eight COAD primary tumor species. The GSE2905 dataset includes two colon cancer cell lines (SW480: low metastasis capability and SW620: high metastasis capability). The prognostic value of genes was analyzed using the data of COAD from TCGA. To validate the results from training cohort, we used the GSE41258 (Sheffer et al. 2009) dataset that includes 390 species as validation cohort. Because the data were downloaded freely from GEO and TCGA database, the approval by an ethics committee of Guangxi Medical University was not needed.

Identification of overlapped DEGs

R language (version 3.4.2) and packages of Bioconductor were conducted to screen the differentially expressed genes (DEGs) between primary tumor tissue and liver metastasis

tissue in GSE40367, the DEGs between SW480 and SW620 in GSE2905 were also screened. Genes that fulfill the criteria of p value < 0.05 and |logFC| ≥1 were defined as the DEGs. Then the intersected DEGs of GSE40367 and GSE2905 were defined as overlapped DEGs. The probe level GSE data was converted into gene expression values in order to measure each gene before screening of DEGs. If one gene is corresponding to multiple probe sets, we used the average data of the multiple probe as the gene expression values (Qin et al. 2012). We also eliminated genes that >20% values of the total samples were missing as previous study did (Liew et al. 2011). After pre-processing the data, t-test methods were used to screen the DEGs using limma package.

Functional enrichment analysis of overlapped DEGs

Gene Ontology (GO) mainly includes three categories, namely, Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). To investigate the functional level of DEGs, DEGs were underwent GO analysis by using DAVID (https://david.ncifcrf.gov/). The significant GO categories were defined with the p < 0.05. Then the DAVID online tool was also used to perform the KEGG pathway analysis, which is to identify the pathways that genes were enrichment, and the significant pathways was defined as the p value <0.05.

Integration of PPI network and subnetwork analysis

Generally, the protein–protein interaction (PPI) network was used to identity key genes and gene modules that associated with the pathogenesis of diseases by constructing the interaction network from the gene level. In this study, the data of PPI network for overlapped DEGs was calculated from the Search Tool for the Retrieval of Interacting Genes (STRING)

database (http://www.string- db.org/). Then, Cytoscape software (version 3.5.2) was used to construct the PPI network for all the overlapped DEGs. After the establishment of the PPI network, we used the Molecular Complex Detection (MCODE) to perform module analysis and identify the hub genes of each module based on the score of each gene in the module.

Acquisition of a gene signature from the training cohort

The association of overlapped DEGs with the overall survival (OS) of colon cancer patients was analyzed in COAD dataset from TCGA database using Cox regression analysis in survival package of R. The gene with p value <0.05 was considered to be independent prognostic factor. In order to estimate the relative contribution of multiple hub genes for the survival prediction of the colon cancer patients, the hub genes was applied to develop a prognostic model using the risk scoring system. In brief, the risk score system was calculated based on a linear combination of the expression level multiplied by the regression coefficient derived from the multivariate cox regression model (β) with the following formula: Risk score = expression of $Gene_1$× $\beta_1 Gene_1$ + expression of $Gene_2$× $\beta_2 Gene_2$ + …expression of $Gene_n$× $\beta_n Gene_n$ [12]. Using the median risk score as the cutoff, patients were divided into high risk group and low risk group. Kaplan-Meier curves were used to estimate the survival for patients between the high risk group and low risk group. p <0.05 was defined as significantly different. The time-dependent receiver operating characteristic (ROC) curve analysis for the gene signature was preformed using the R package "survivalROC" (Heagerty et al. 2000). All statistical analyses were performed using R software (version 3.4.2) and Bioconductor.

Results

Overlapped genes of CRC cells and tissues

The DEGs of GES2905 and GSE40367 were screened based on the selection criteria after preprocessing raw data. 341 DEGs between colon cancer cells lines SW420 and SW680 cells were identified in GES2905 dataset, and 7339 DEGs between primary and metastasis tumor specimens were identified in GSE40367 dataset. By overlapped the DEGs from the two datasets, we obtained 184 overlapped genes that differentially expressed in both CRC cells and tissues. The result was displayed in Fig. 1.

GO and KEGG enrichment analysis

The gene functions of the 184 overlapped DEGs were then analyzed by GO and KEGG enrichment analysis. Using the DAVID online tools, we found that the most enriched GO terms of DEGs that related to BP was Signal transduction, and the MF was protein binding, and the CC was cytoplasm. The KEGG pathway analysis based on the GO results revealed that thyroid hormone synthesis was the most significant pathway of the overlapped DEGs. The results were shown in Fig. 2.

PPI network and Module screening analysis

Using the data from STRING database, a PPI network for the 184 DEGs consisting of 133 nodes and 138 edges was constructed by Cytoscape software. The overall PPI network was shown in Fig. 3A. Then the plug-ins MCODE identified four modules from the PPI network, with the OXCT1 (sore: 4), ACTN4 (sore: 3), IL-8 (sore: 3), ITGA3 (sore: 3) as the hub genes of each module. The top three modules were shown in Fig. 3B-D.

Prognostic value of overlapped DEGs

The prognostic value of 184 overlapped DEGs was analyzed using the COAD dataset of TCGA by the multiple Cox regression analysis after adjusted the data of age, gender and TNM stage, the results showed that only ALDH2, NEDD9, FLNA, LBR, TWF1, SRSF1 were independent genes that associated with the OS of colon cancer patients, with the beta value as -1.343, -0.051, 0.492, -0.020, -0.181 and -1.938, respectively. We developed a six-gene signature by calculating the risk score of each gene, and divided the patients into high risk group and low risk group based on the median of risk score (Fig. 4A), the survival status and genes expression level is shown in Fig. 4B-C. The survival analysis revealed that this six-gene signature with high risk group predicted poor OS of colon cancer patients compared with low risk group, as the p value <0.001 (Fig. 5A). Using survival ROC analysis, we found that the risk score of this six-gene signature has a moderate predict value for the 1-, 3-, 5-year OS of colon cancer patients, as the value of area under ROC curve (AUC) was 0.686, 634, 618, respectively (Fig. 5B).

Validation cohort confirm the prognostic value of the six-gene signature

The prognostic value of the six-gene signature for the OS of colon patients was further determined in the validation cohort (GSE41258 datasets, 390 colon patients, mean follow-up 65.3 months). Using the risk score model and cut-off value of the training cohort, we divided the validation cohort (390 patients) into high-risk group (n=195) and low-risk group (n=195), respectively. In agreement with results of the training cohort, the results of this six-gene signature in the validation cohort also indicated an obvious difference between high risk

group and low risk group with regard to the OS of colon cancer patients (p=0.005, log-rank test)

Discussion

Metastasis accounts for 90% of the mortalities of colon cancer patients and thus becomes the most lethal characteristic of colon cancer (Li et al. 2017). Colon cancer patients with localized- and regional -disease have a good prognosis (the 5-year survival rates up to 91.1%), but patients with distant metastasis has a much worse prognosis (the 5-year survival rates drop to 13.3%) (Siegel et al. 2017a). Furthermore, the failure of treatment are mostly caused by the ~~the~~ metastatic dissemination of primary tumors (Deliu et al. 2014; Stein & Schlag 2007). At a molecular level, the distinct metastasis colon cancer is molecularly and clinically distinct from the primary site of origin (Zarour et al. 2017). Thus, analyzing the molecular alteration of colon cancer with distinct metastasis is benefit for the identification of candidate targets for early diagnosis and treatment of advanced stage of colon cancer patients.

To date, some biomarkers have been identified to be the candidate targets for early diagnosis and treatment of colon cancer and rectum cancer, including genes, miRNA, lncRNA and the related signatures. Some gene signatures that related to the metastasis potential of each tumor have been described with promising results. Vellinga et al (Vellinga et al. 2017) designed a lymph angiogenic gene set and applied it to large datasets of CRC, and found that this lymph angiogenic gene set was related to the worse prognosis. Rokavec et al (Rokavec et al. 2017) reported a single gene, RBM47, was down-regulation during CRC progression may promote epithelial-mesenchymal transition and metastasis. Furthermore, proteomic studies used exosomes that from cancer cell lines and identified four candidates genes (MET, S100A8, S100A9, TNC) associated with CRC metastasis (Ji et al. 2013). Other

biomarkers, such as a four -miRNA signature (let-7i, miR-10b, miR-221 and miR-320a) (Hur et al. 2015), and a six- lncRNA signature were reported to be promising biomarkers for the metastasis and prognosis of CRC (Hu et al. 2014).

In this study, we identified 4 hub genes from the subnetwork of the PPI network that are OXCT1, ACTN4, IL-8 and ITGA3. OXCT1 is a key enzyme involved in the process of ketone body metabolism, and catalyzes the first and rate-determining step of ketolysis. OXCT1 can be converted into acetyl-CoA during the process of metabolism and takes part in the tricarboxylic acid cycle for the oxidation and the production of ATP (Zhang & Xie 2017). The role of OXCT1 has been defined in several cancers, including colorectal cancer, and the OXCT1 was overexpressed in the metastatic CRC cell line CC-M3 (Lee et al. 2016). ACTN4 is a non-muscle-type alpha-actinin, it plays an important role in regulating cytoskeleton organization and involves transcriptional regulation of gene expression. ACTN4 encodes a non-muscle, alpha-actinin isoform that concentrates in the cytoplasm, and participates in metastatic processes, and facilitates the motility, invasion and metastasis of cancer cells. In CRC, ACTN4 was reported to promote CRC cell line invasion by suppressing focal adhesion maturation (Fukumoto et al. 2015). ITGA3 belongs to the integrin family, joins a beta 1 subunit to form an intact integrin and interacts with several extracellular matrix proteins (Nagata et al. 2013). ITGA3 is found to be associated with lymphatic dissemination and local invasiveness in cancers. One study showed that ITGA3 was over-expressed in stages III versus I of CRC patients, and related to the OS and disease-free survival (Linhares et al. 2015). IL-8 is an important pro-inflammatory chemokine and plays a role in the recruitment of leukocytes to the sites of infection or tissue injury. A growing of evidence has suggested that paracrine signaling by tumor-derived IL-8 promotes the trafficking of neutrophils and myeloid-derived suppressor cells into the tumor microenvironment, which is associated with the dampen anti-tumor immune responses (David et al. 2016). Lambrechts et al

(~~Lambrechts et al.~~ 2015) observed that IL-8 plasma levels at baseline and subsequent increases in IL-8 were associated with worse progression-free survival of metastatic CRC patients. These studies confirmed our results that the hub genes of the PPI network were crucial to the metastasis of colon cancer, but further studies remains warrant to determine the underlying mechanism.

Similar to the previous studies, in this study, a signature constructed by six genes was showed to be good predictor for the OS of colon cancer, among these six genes, the role of ALDH2, NEDD9, SRSF1 and FLNA were reported to be associated with the CRC in several studies. ALDH2 is essential for the metabolism and detoxification of a wide range of endogenous and exogenous aldehyde substrates. It is the rate-limiting enzyme in the ethanol metabolism, oxidizing acetaldehyde to acetic acid both in the liver and other tissues (Chen et al. 2016). As a novel biological marker, ALDH2 displays an attractive prospect in the screening, diagnosis and evaluation of the prognosis of many diseases, and the genetic polymorphism of ALDH2 was significantly correlated with the susceptibility to CRC (Li et al. 2016). NEDD9 is a non-catalytic scaffolding protein, assembles complexes involving oncogenic kinases, and regulates the magnitude and duration of cell signaling cascades that controls multiple processes, which are important to the development and progression of tumors (Shagisultanova et al. 2015). Study has showed that downregulation of NEDD9 by apigenin can suppresses migration, invasion, and metastasis of CRC cells (Dai et al. 2016). With regard to the SRSF1, there was a study reported that phosphorylation of SRSF1 regulated alternative splicing of tumor-related Rac1b in CRC cells (Goncalves et al. 2014). SRSF1 is classified as exonic splicing enhancers, and recognizes degenerate purine-rich sequence motifs. SRSF1 can promote the recognition of both constitutive and alternative exons during the process of spliceosomal assembly (Sanford et al. 2009). FLNA is an actin-binding protein and expressing ubiquitously in the body. FLNA involves in many cell

signaling pathways and is important in the maintenance of cell shape and motility. The mutation of FLNA has been shown to associate with the neuronal migration, vascular function, connective tissue integrity, and skeletal development (Shelmerdine et al. 2017); FLNA showed low expression in CRC, and was high correlated with the incidence and development of CRC (Tian et al. 2015). LBR is a transmembrane protein of the inner nuclear membrane. LBR interacts with heterochromatin and B-type lamins through its nucleoplasmic amino-terminal domain, and is phosphorylated throughout the cell cycle (Duband-Goulet et al. 1998). TWF1 belongs to the ADF-H family, and is a conserved actin-binding protein. It regulates diverse morphological processes through sequestering ADP-actin monomers or capping filament barbed ends (Paavilainen et al. 2007). However, no study has yet reported the role of LBR and TWF1 in CRC, so their role need be further studied in colon cancers.

Compared with previous studies, this study screened the metastasis related genes by overlapping the DEGs from cancer tissues and cell lines. These tissues included primary colon cancer tissues and liver metastasis tissues, and the cell lines included primary (SW480) and metastatic (SW620) human isogenic colorectal cancer cell lines, thus, the overlapped DEGs could be more reliable in reflecting the genes altered in metastatic colon cancer. We also performed a comprehensive analysis for these overlapped DEGs, using GO and KEGG analysis, and identified the function of the genes and the pathways they involved, then the PPI network and subnetwork analysis revealed the gene-gene interaction and identified four hub genes that are crucial to the network, which provided an insight into the mechanism of colon cancer metastasis. Furthermore, we analyzed the prognostic value of the overlapped DEGs, and identified six genes that related to the OS of colon cancer patients. We finally developed a six-gene signature to test its prognostic value and validated by an independent dataset. Therefore, these results were more informative and provided ~~a~~ reliable prognostic factors pinpointing a subset of patients with poor prognosis.

However, some limitations need to be noted in this study. First, although metastasis-related genes of colon cancer were identified and the prognostic value of them were validated in our study, the results were calculated from microarray or RNA-sequencing technique datasets, thus, lack of functional validation of the target genes is one of the major limitations of this study. Therefore, a thorough functional experiment for these genes and corresponding downstream events to reveal novel diagnostic and therapeutic targets for colon cancer is necessary. Second, the development of colon cancer metastasis can be caused by many factors, such as KRAS, BRAF mutation, microsatellite instability, which has been proven to be closely related to the colon cancer, but due to the limited ~~of~~ datasets, we did not perform stratified analysis based on these factors, future studies should conduct this analysis to explore the difference under different conditions. Third, the mean time of the follow-up in validation cohort was 65.3 months, thus, a study including a longer follow-up time is warranted to validate our results in the future.

Conclusions

This study screens a gene profiles involving in the metastasis of colon cancer, and identifies four hub genes from the gene profiles. We also identify and validate a six-gene signature that can be served as an indicator of prognosis of colon cancer. Some genes that are not yet proved to be associated with colon cancer metastasis may represent new therapeutic targets.

Abbreviation:

CRC: colorectal cancer; DEGs: differentially expressed genes; GO: Gene Ontology; KEGG:

Kyoto Encyclopedia of Genes and Genomes; OS: overall survival; ROC: receiver operating characteristic curve; OXCT1: 3-oxoacid CoA-transferase 1; ACTN4: actinin alpha 4; IL-8: interleukin 8; ITGA3: integrin subunit alpha 3; ALDH2: aldehyde dehydrogenase 2; NEDD9: neural precursor cell expressed, developmentally down-regulated 9; FLNA: filamin A; LBR: lamin B receptor; TWF1: twinfilin actin binding protein 1; SRSF1: serine and arginine rich splicing factor 1; TNC: tenascin C; RBM47: RNA binding motif protein 47; MET: MET proto-oncogene, receptor tyrosine kinase

Reference

Castro E, Goh C, Olmos D, Saunders E, Leongamornlert D, Tymrakiewicz M, Mahmud N, Dadaev T, Govindasami K, Guy M, Sawyer E, Wilkinson R, Ardern-Jones A, Ellis S, Frost D, Peock S, Evans DG, Tischkowitz M, Cole T, Davidson R, Eccles D, Brewer C, Douglas F, Porteous ME, Donaldson A, Dorkins H, Izatt L, Cook J, Hodgson S, Kennedy MJ, Side LE, Eason J, Murray A, Antoniou AC, Easton DF, Kote-Jarai Z, and Eeles R. 2013. Germline BRCA mutations are associated with higher risk of nodal involvement, distant metastasis, and poor survival outcomes in prostate cancer. *J Clin Oncol* 31:1748-1757. 10.1200/JCO.2012.43.1882

Chen CH, Joshi AU, and Mochly-Rosen D. 2016. The Role of Mitochondrial Aldehyde Dehydrogenase 2 (ALDH2) in Neuropathology and Neurodegeneration. *Acta Neurol Taiwan* 25(4):111-123. 10196099/254111 [pii]

Dai J, Van Wie PG, Fai LY, Kim D, Wang L, Poyil P, Luo J, and Zhang Z. 2016. Downregulation of NEDD9 by apigenin suppresses migration, invasion, and metastasis of colorectal cancer cells. *Toxicol Appl Pharmacol* 311:106-112.

10.1016/j.taap.2016.09.016

David JM, Dominguez C, Hamilton DH, and Palena C. 2016. The IL-8/IL-8R Axis: A Double Agent in Tumor Immune Resistance. *Vaccines (Basel)* 4. 10.3390/vaccines4030022

Deliu IC, Georgescu EF, and Bezna MC. 2014. Analysis of prognostic factors in colorectal carcinoma. *Rev Med Chir Soc Med Nat Iasi* 118:808-816.

Demirkol S, Gomceli I, Isbilen M, Dayanc BE, Tez M, Bostanci EB, Turhan N, Akoglu M, Ozyerli E, Durdu S, Konu O, Nissan A, Gonen M, and Gure AO. 2017. A Combined ULBP2 and SEMA5A Expression Signature as a Prognostic and Predictive Biomarker for Colon Cancer. *J Cancer* 8:1113-1122. 10.7150/jca.17872

Duband-Goulet I, Courvalin JC, and Buendia B. 1998. LBR, a chromatin and lamin binding protein from the inner nuclear membrane, is proteolyzed at late stages of apoptosis. *J Cell Sci* 111 (Pt 10):1441-1451.

Fukumoto M, Kurisu S, Yamada T, and Takenawa T. 2015. alpha-Actinin-4 enhances colorectal cancer cell invasion by suppressing focal adhesion maturation. *PLoS One* 10:e0120616. 10.1371/journal.pone.0120616

Goncalves V, Henriques AF, Pereira JF, Neves Costa A, Moyer MP, Moita LF, Gama-Carvalho M, Matos P, and Jordan P. 2014. Phosphorylation of SRSF1 by SRPK1 regulates alternative splicing of tumor-related Rac1b in colorectal cells. *RNA* 20:474-482. 10.1261/rna.041376.113

Heagerty PJ, Lumley T, and Pepe MS. 2000. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56:337-344.

Hu Y, Chen HY, Yu CY, Xu J, Wang JL, Qian J, Zhang X, and Fang JY. 2014. A long

non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* 5:2230-2242. 1895 10.18632/oncotarget.1895

Hur K, Toiyama Y, Schetter AJ, Okugawa Y, Harris CC, Boland CR, and Goel A. 2015. Identification of a metastasis-specific MicroRNA signature in human colorectal cancer. *J Natl Cancer Inst* 107. 10.1093/jnci/dju492

Ji H, Greening DW, Barnes TW, Lim JW, Tauro BJ, Rai A, Xu R, Adda C, Mathivanan S, Zhao W, Xue Y, Xu T, Zhu HJ, and Simpson RJ. 2013. Proteome profiling of exosomes derived from human primary and metastatic colorectal cancer cells reveal differential expression of key metastatic factors and signal transduction components. *Proteomics* 13:1672-1686. 10.1002/pmic.201200562

Knijn N, Tol J, and Punt CJ. 2010. Current issues in the targeted therapy of advanced colorectal cancer. *Discov Med* 9:328-336.

Lambrechts D, Thienpont B, Thuillier V, Sagaert X, Moisse M, Peuteman G, Pericay C, Folprecht G, Zalcberg J, Zilocchi C, Margherini E, Chiron M, and Van Cutsem E. 2015. Evaluation of efficacy and safety markers in a phase II study of metastatic colorectal cancer treated with aflibercept in the first-line setting. *Br J Cancer* 113:1027-1034. 10.1038/bjc.2015.329

Lee CL, Huang CJ, Yang SH, Chang CC, Huang CC, Chien CC, and Yang RN. 2016. Discovery of genes from feces correlated with colorectal cancer progression. *Oncol Lett* 12:3378-3384. 10.3892/ol.2016.5069

Li H, Chen YX, Wen JG, and Zhou HH. 2017. Metastasis-associated in colon cancer 1: A promising biomarker for the metastasis and prognosis of colorectal cancer. *Oncol Lett* 14:3899-3908. 10.3892/ol.2017.6670

Li R, Zhao Z, Sun M, Luo J, and Xiao Y. 2016. ALDH2 gene polymorphism in different types of cancers and its clinical significance. *Life Sci* 147:59-66. 10.1016/j.lfs.2016.01.028

Liew AW, Law NF, and Yan H. 2011. Missing value imputation for gene expression data: computational techniques to recover missing data from available information. *Brief Bioinform* 12:498-513. 10.1093/bib/bbq080

Linhares MM, Affonso RJ, Jr., Viana Lde S, Silva SR, Denadai MV, de Toledo SR, and Matos D. 2015. Genetic and Immunohistochemical Expression of Integrins ITGAV, ITGA6, and ITGA3 As Prognostic Factor for Colorectal Cancer: Models for Global and Disease-Free Survival. *PLoS One* 10:e0144333. 10.1371/journal.pone.0144333

Marzouk O, and Schofield J. 2011. Review of histopathological and molecular prognostic features in colorectal cancer. *Cancers (Basel)* 3:2767-2810. 10.3390/cancers3022767

Nagata M, Noman AA, Suzuki K, Kurita H, Ohnishi M, Ohyama T, Kitamura N, Kobayashi T, Uematsu K, Takahashi K, Kodama N, Kawase T, Hoshina H, Ikeda N, Shingaki S, and Takagi R. 2013. ITGA3 and ITGB4 expression biomarkers estimate the risks of locoregional and hematogenous dissemination of oral squamous cell carcinoma. *BMC Cancer* 13:410. 10.1186/1471-2407-13-410

Paavilainen VO, Hellman M, Helfer E, Bovellan M, Annila A, Carlier MF, Permi P, and Lappalainen P. 2007. Structural basis and evolutionary origin of actin filament capping by twinfilin. *Proc Natl Acad Sci U S A* 104:3113-3118. 10.1073/pnas.0608725104

Provenzani A, Fronza R, Loreni F, Pascale A, Amadio M, and Quattrone A. 2006. Global alterations in mRNA polysomal recruitment in a cell model of colorectal cancer

progression to metastasis. *Carcinogenesis* 27:1323-1333. 10.1093/carcin/bgi377

Qin S, Kim J, Arafat D, and Gibson G. 2012. Effect of normalization on statistical and biological interpretation of gene expression profiles. *Front Genet* 3:160. 10.3389/fgene.2012.00160

Roessler S, Lin G, Forgues M, Budhu A, Hoover S, Simpson RM, Wu X, He P, Qin LX, Tang ZY, Ye QH, and Wang XW. 2015. Integrative genomic and transcriptomic characterization of matched primary and metastatic liver and colorectal carcinoma. *Int J Biol Sci* 11:88-98. 10.7150/ijbs.10583

Rokavec M, Kaller M, Horst D, and Hermeking H. 2017. Pan-cancer EMT-signature identifies RBM47 down-regulation during colorectal cancer progression. *Sci Rep* 7:4687. 10.1038/s41598-017-04234-2

Sameer AS. 2013. Colorectal cancer: molecular mutations and polymorphisms. *Front Oncol* 3:114. 10.3389/fonc.2013.00114

Sanford JR, Wang X, Mort M, Vanduyn N, Cooper DN, Mooney SD, Edenberg HJ, and Liu Y. 2009. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res* 19:381-394. 10.1101/gr.082503.108

Shagisultanova E, Gaponova AV, Gabbasov R, Nicolas E, and Golemis EA. 2015. Preclinical and clinical studies of the NEDD9 scaffold protein in cancer and other diseases. *Gene* 567:1-11. 10.1016/j.gene.2015.04.086

Sheffer M, Bacolod MD, Zuk O, Giardina SF, Pincas H, Barany F, Paty PB, Gerald WL, Notterman DA, and Domany E. 2009. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proc Natl Acad Sci U S A* 106:7131-7136. 10.1073/pnas.0902232106

Shelmerdine SC, Semple T, Wallis C, Aurora P, Moledina S, Ashworth MT, and Owens CM. 2017. Filamin A (FLNA) mutation-A newcomer to the childhood interstitial lung disease (ChILD) classification. *Pediatr Pulmonol* 52:1306-1315. 10.1002/ppul.23695

Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, and Jemal A. 2017a. Colorectal cancer statistics, 2017. *CA Cancer J Clin* 67:177-193. 10.3322/caac.21395

Siegel RL, Miller KD, and Jemal A. 2017b. Cancer Statistics, 2017. *CA Cancer J Clin* 67:7-30. 10.3322/caac.21387

Stein U, and Schlag PM. 2007. Clinical, biological, and molecular aspects of metastasis in colorectal cancer. *Recent Results Cancer Res* 176:61-80.

Tian ZQ, Shi JW, Wang XR, Li Z, and Wang GY. 2015. New cancer suppressor gene for colorectal adenocarcinoma: filamin A. *World J Gastroenterol* 21:2199-2205. 10.3748/wjg.v21.i7.2199

Vellinga TT, Kranenburg O, Frenkel N, Ubink I, Marvin D, Govaert K, van Schelven S, Hagendoorn J, and Borel Rinkes IH. 2017. Lymphangiogenic Gene Expression Is Associated With Lymph Node Recurrence and Poor Prognosis After Partial Hepatectomy for Colorectal Liver Metastasis. *Ann Surg*. 10.1097/SLA.0000000000002430

Xu G, Zhang M, Zhu H, and Xu J. 2017. A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. *Gene* 604:33-40. 10.1016/j.gene.2016.12.016

Zarour LR, Anand S, Billingsley KG, Bisson WH, Cercek A, Clarke MF, Coussens LM, Gast CE, Geltzeiler CB, Hansen L, Kelley KA, Lopez CD, Rana SR, Ruhl R, Tsikitis VL, Vaccaro GM, Wong MH, and Mayo SC. 2017. Colorectal Cancer Liver Metastasis:

Evolving Paradigms and Future Directions. *Cell Mol Gastroenterol Hepatol* 3:163-173. 10.1016/j.jcmgh.2017.01.006

Zhang S, and Xie C. 2017. The role of OXCT1 in the pathogenesis of cancer as a rate-limiting enzyme of ketone body metabolism. *Life Sci* 183:110-115. 10.1016/j.lfs.2017.07.003

Your manuscript has been improved; however, (1) many English errors or space problems have been identified. I have marked a few of them in filled yellow in your v3, which need to be corrected, and a careful proofreading is needed.

The use of the available GSE biosets to perform bioinformatics to search for biomarkers is of merit. However, in your future studies, you may try Partek Genomics software and Illumina Correlation Engine for biosets analysis, but also need biological or clinical verifications to impact the field.