

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

Haotang Wei^{1,*}, Jilin Li^{2,*}, Minzhi Xie², Ronger Lei³ and Bangli Hu²

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

² Department of Research, Affiliated Tumor Hospital of Guangxi Medical University, Nanning, China

³ Department of Gastroenterology, First Affiliated Hospital of Guangxi Medical University, Nanning, China

* These authors contributed equally to this work.

ABSTRACT

Objective: The mechanism underlying colon cancer metastasis remain unclear. This study aimed to elucidate the genes alteration during the metastasis of colon cancer and identify genes that 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 were analyzed in The Cancer Genome Atlas dataset, and a gene signature was developed using genes that related to the overall survival (OS). The prognostic value of the gene signature was further confirmed in a validation cohort.

Results: A total of 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 (3-oxoacid CoA-transferase 1, actinin alpha 4, interleukin 8, integrin subunit alpha 3) were identified using protein-protein network analysis. Six genes (aldehyde dehydrogenase 2, neural precursor cell expressed, developmentally down-regulated 9, filamin A, lamin B receptor, twinfilin actin binding protein 1, serine and arginine rich splicing factor 1) were closely related to the OS of colon cancer patients. A gene signature was developed using 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.

Submitted 14 March 2018

Accepted 10 July 2018

Published 21 August 2018

Corresponding author

Bangli Hu, hubangli@gxmu.edu.cn

Academic editor

Jie Liu

Additional Information and
Declarations can be found on
page 11

DOI 10.7717/peerj.5433

© Copyright

2018 Wei et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Gastroenterology and Hepatology, Oncology

Keywords Colon cancer, Metastasis, Genes, Signature, Prognosis

INTRODUCTION

Colorectal cancer (CRC) is one of the leading malignant cancers in the world, and colon cancer accounts for a large part of CRC (Siegel *et al.*, 2017; Siegel, Miller & Jemal, 2017). During the last three decades, with the development of new therapies, such as asbevacizumab and anti-epidermal growth factor receptor antibody (Castro *et al.*, 2013; Knijn, Tol & Punt, 2010), survival outcomes for colon cancer patients with localized and regional disease have improved greatly. However, cancer metastasis remains one of the main causes of death in these patients. One studied showed that the prognosis of patients with distant metastasis remained poor compared to the prognosis of patients with early-stage colon cancer (Siegel, Miller & Jemal, 2017). Accumulation of genetic mutations is recognized as one of the important causes for 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 tumor, node, metastasis (TNM) stage and pathological characteristics of colon cancer have been used to predict the prognosis and facilitate treatment for colon cancer patients. However, there are some limitations in using these methods to assess patients (Marzouk & Schofield, 2011). Recently, several novel biomarkers have been tested with the aim of improving the prediction of therapeutic response and prognosis of colon cancer patients (Demirkol *et al.*, 2017; Hu *et al.*, 2014; Xu *et al.*, 2017). These have aided in the diagnosis and treatment of colon cancer, but the results were inconsistent and need to be further studied.

Thus far, metastasis is a major factor for poor prognoses of colon cancer patients, and the liver is the most common organ affected. However, the molecular mechanism underlying distant metastasis remains unclear. Therefore, a comprehensive analysis of the molecular alterations to identify prognostic indicators is key for the management of colon cancer patients with distant metastasis. In this study, we used the colon cancer data from gene expression omnibus (GEO) and The Cancer Genome Atlas (TCGA) and analyzed the data of patients with colon cancer using primary tumor samples and liver metastasis samples to unveil the genes 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)) were retrieved and downloaded from the GEO (<http://www.ncbi.nlm.nih.gov/geo/>) database of National Center for Biotechnology Information as the training cohort. The GSE40367 dataset includes seven colon adenocarcinoma (COAD) with liver metastasis species and eight COAD primary tumor species. The GSE2509 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 confirm the results from training cohort, we used the GSE41258 (Sheffer *et al.*, 2009) dataset that includes 390 species

as a validation cohort. Because the data were download freely from GEO and TCGA database, approval of the ethics committee of Guangxi Medical University was not needed.

Identification of overlapped DEGs

R language software (version 3.4.2) and Bioconductor packages were applied to screen the differentially expressed genes (DEGs) between primary tumor tissue and liver metastasis tissue in [GSE40367](#). The DEGs between SW480 and SW620 in [GSE2509](#) were also screened. Genes that fulfilled the criteria of having a p -value < 0.05 and $|\log_{2}FC| \geq 1$ were defined as the DEGs. Then, the intersected DEGs of [GSE40367](#) and [GSE2509](#) were defined as overlapped DEGs. The probe level GSE data were converted into gene expression values to measure each gene before the screening of DEGs. If one gene corresponded to multiple probe sets, we used the average data of the multiple probes as the gene expression values ([Qin et al., 2012](#)). We also eliminated genes that had over $>20\%$ values of the total samples as a previous study did ([Liew, Law & Yan, 2011](#)). After pre-processing the data, t -test methods were used to screen the DEGs using a limma package.

Functional enrichment analysis of overlapped DEGs

Gene ontology (GO) includes three categories, namely, the biological process (BP), molecular function (MF), and the cellular component (CC). To investigate the functional level of DEGs, the genes underwent GO analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID, <https://david.ncifcrf.gov/>). The significant GO categories were defined as those with $p < 0.05$. Following this, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was conducted to identify significant pathways with gene enrichment using a DAVID online tool. We defined pathways as significant if the p -value was <0.05 .

Integration of PPI network and subnetwork analysis

Protein–protein interaction (PPI) networks was used to identify key genes and important gene modules which are involved in cancer development by assessing gene interaction levels. In this study, the data of PPI network for overlapped DEGs was acquired 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 for the establishment of a PPI network for all overlapped DEGs. After the establishment of a PPI network, module analysis was carried out by molecular complex detection to detect the gene modules of the PPI network. The hub of genes of each module was identified based on the score of each gene in the module.

Acquisition of a gene signature from the training cohort

The association of overlapping DEGs with the overall survival OS of colon cancer patients was analyzed in the COAD dataset from the TCGA database using a Cox regression analysis in the survival package of R. The gene with $p < 0.05$ was considered to be an independent prognostic factor. To estimate the relative contribution of multiple genes

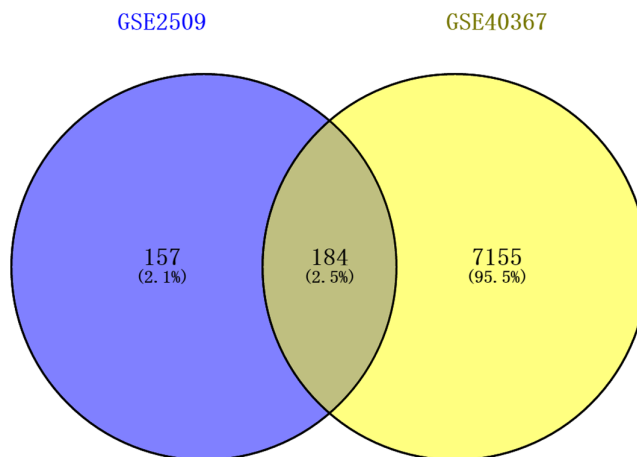


Figure 1 Venn diagrams showing intersected genes overlapping between time-series genes (TSGs) and differentially expressed genes (DEGs). [Full-size](#) DOI: 10.7717/peerj.5433/fig-1

for survival prediction of colon cancer patients, the hub genes were 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 (the β value) with the following formula as previously reported: Risk score = expression of Gene₁ \times β_1 Gene₁ + expression of Gene₂ \times β_2 Gene₂ + ... expression of Gene_n \times β_n Gene_n (Hur *et al.*, 2015). Using the median risk score as the cutoff, patients were divided into high-risk and low-risk groups. Kaplan–Meier curves were used to estimate the difference in survival for patients between the high-risk group and low-risk groups. $p < 0.05$ was defined as significantly different. Time-dependent receiver operating characteristic (ROC) curve analysis for the gene signature was performed using the R package “survivalROC” (Heagerty, Lumley & Pepe, 2000). All statistical analyses were performed using R software and Bioconductor.

RESULTS

Overlapped genes of CRC cells and tissues

The DEGs of GSE2509 and GSE40367 were screened on the basis of the selection criteria after preprocessing the raw data. A total of 341 DEGs between colon cancer cells lines SW420 and SW680 cells were identified in GSE2509 dataset, and 7339 DEGs between primary and metastasis tumor specimens were identified in the GSE40367 dataset. By overlapping the DEGs from the two datasets, we obtained 184 overlapped genes that differentially expressed in both CRC cells and tissues. The result is shown 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 related to BP were Signal transduction, and the MF was Protein

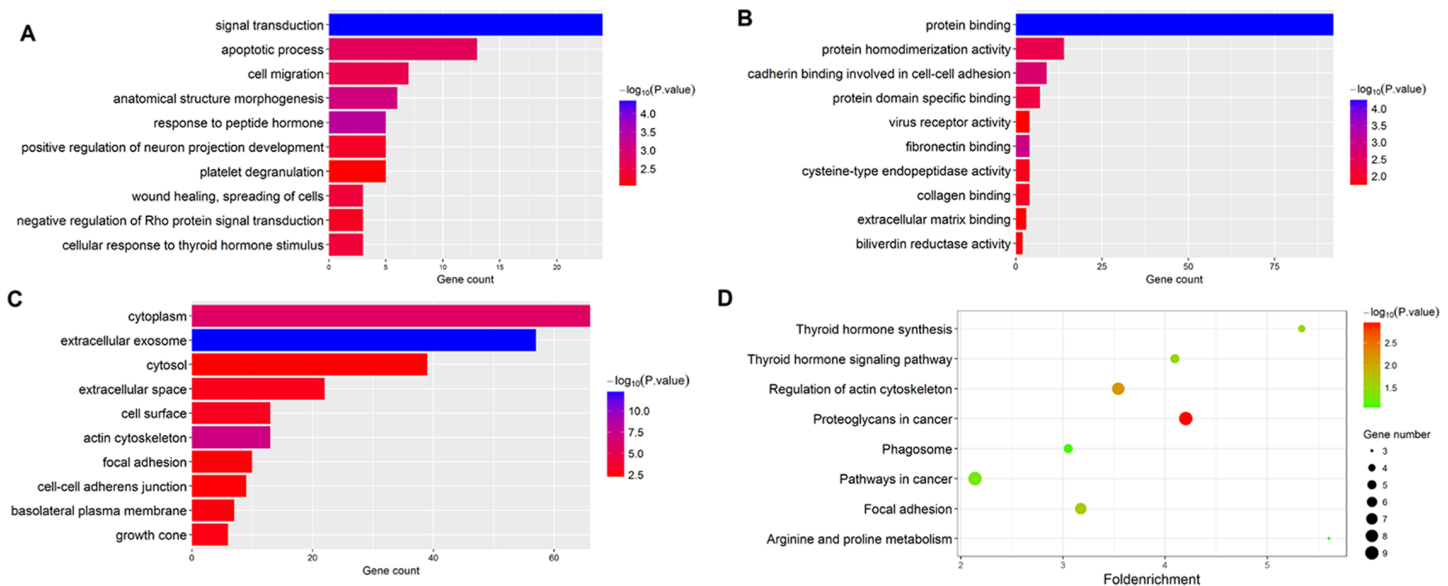


Figure 2 Functional enrichment analysis of overlapped DEGs. (A) GO enrichment analysis of overlapped DEGs in biological processes; (B) GO enrichment analysis of overlapped DEGs in molecular function; (C) GO enrichment analysis of overlapped DEGs in cellular component; (D) KEGG analysis of overlapped DEGs. [Full-size](#) DOI: 10.7717/peerj.5433/fig-2

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 are shown in Fig. 2.

PPI network and module screening analysis

Using data from the STRING database, a PPI network for the 184 DEGs consisting of 133 nodes and 138 edges was constructed using Cytoscape software. The overall PPI network is shown in Fig. 3A. Following this, the plug-in MCODE was used to detect the modules in the network, and four modules from the PPI network were identified, with the 3-oxoacid CoA-transferase 1 (*OXCT1*) (sore: 4), actinin alpha 4 (*ACTN4*) (sore: 3), interleukin 8 (*IL-8*) (sore: 3), integrin subunit alpha 3 (*ITGA3*) (sore: 3) as the hub genes of each module. The top three modules are shown in Figs. 3B–3D.

Prognostic value of overlapped DEGs

The prognostic value of the 184 overlapping DEGs was analyzed using the COAD dataset of TCGA by multiple Cox regression analysis after adjusting the data for age, sex, and TNM stage. The results showed that only aldehyde dehydrogenase 2 (*ALDH2*), neural precursor cell expressed, developmentally down-regulated 9 (*NEDD9*), filamin A (*FLNA*), lamin B receptor (*LBR*), twinfilin actin binding protein 1 (*TWF1*), and serine and arginine rich splicing factor 1 (*SRSF1*) were independent genes that associated with the OS of colon cancer patients, with β values of -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 a high-risk group and a low-risk group on the basis of the median risk score (Fig. 4A), the survival status, and the genes expression level shown in Figs. 4B–4C.

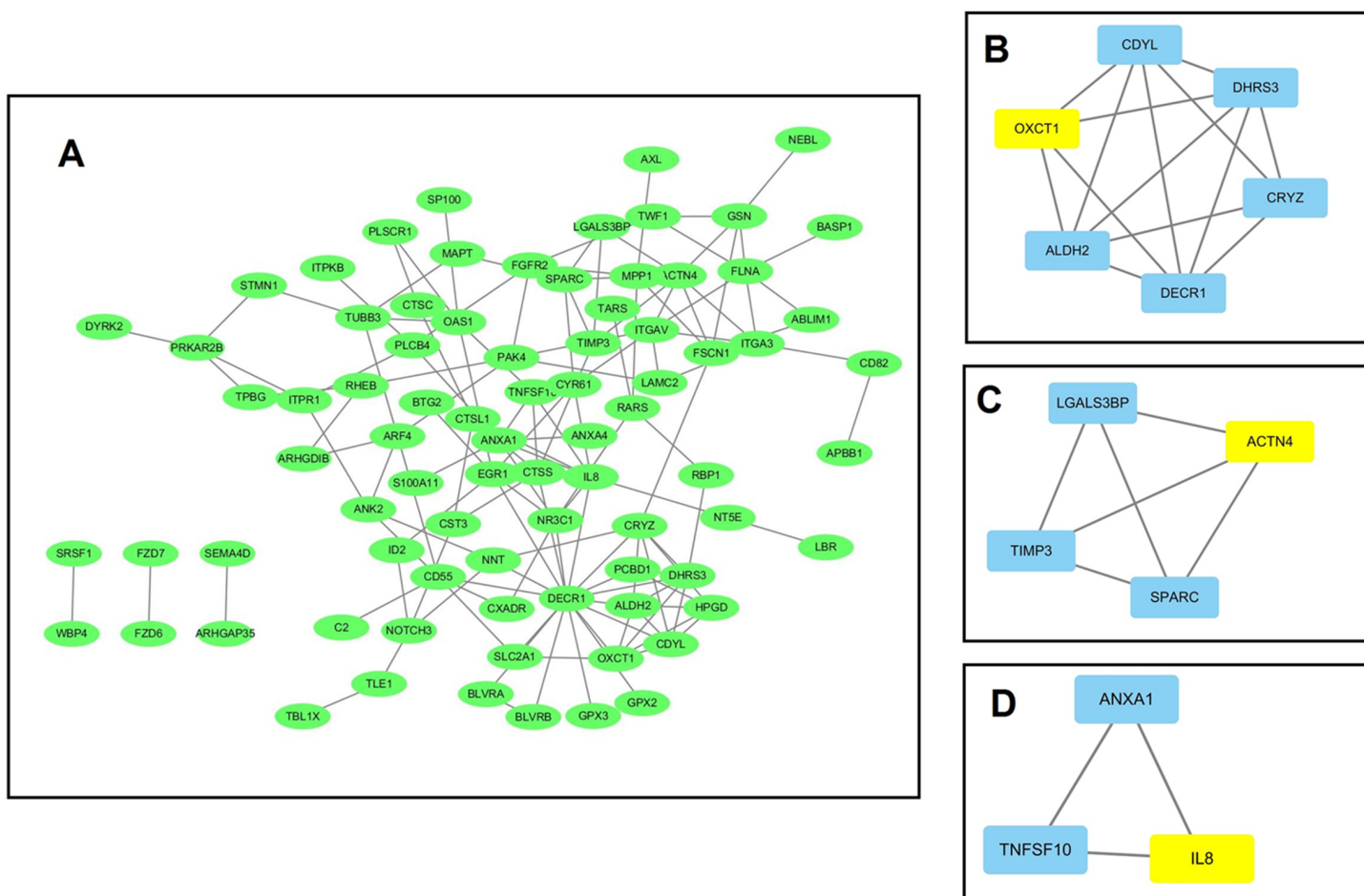


Figure 3 The protein–protein interaction (PPI) network analysis of the 326 overlapped DEGs. (A) PPI network for the overlapped DEGs; (B–D) three major submodules of the PPI network. [Full-size !\[\]\(1679558f37f6db0dd8360a2a7e913e90_img.jpg\) DOI: 10.7717/peerj.5433/fig-3](https://doi.org/10.7717/peerj.5433/fig-3)

The survival analysis revealed that this six-gene signature within the high-risk group predicted poor OS of colon cancer patients compared to that in the low-risk group, with $p < 0.001$ (Fig. 5A). Using survival ROC analysis, we found that the risk score of this six-gene signature could moderately predict the 1-, 3-, 5-year OS of colon cancer patients, as the value of the area under ROC curve was 0.686, 634, and 618, respectively (Fig. 5B).

Validation cohort confirms 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). By using the same risk score model and cutoff value derived from the training cohort, 390 patients of the validation cohort were classified into either a high-risk group ($n = 195$) or a low-risk group ($n = 195$). Consistent with results of the training cohort, the results of this six-gene signature indicated an obvious difference between the high-risk group and low-risk group with regard to the OS of colon cancer patients ($p = 0.005$, log-rank test).

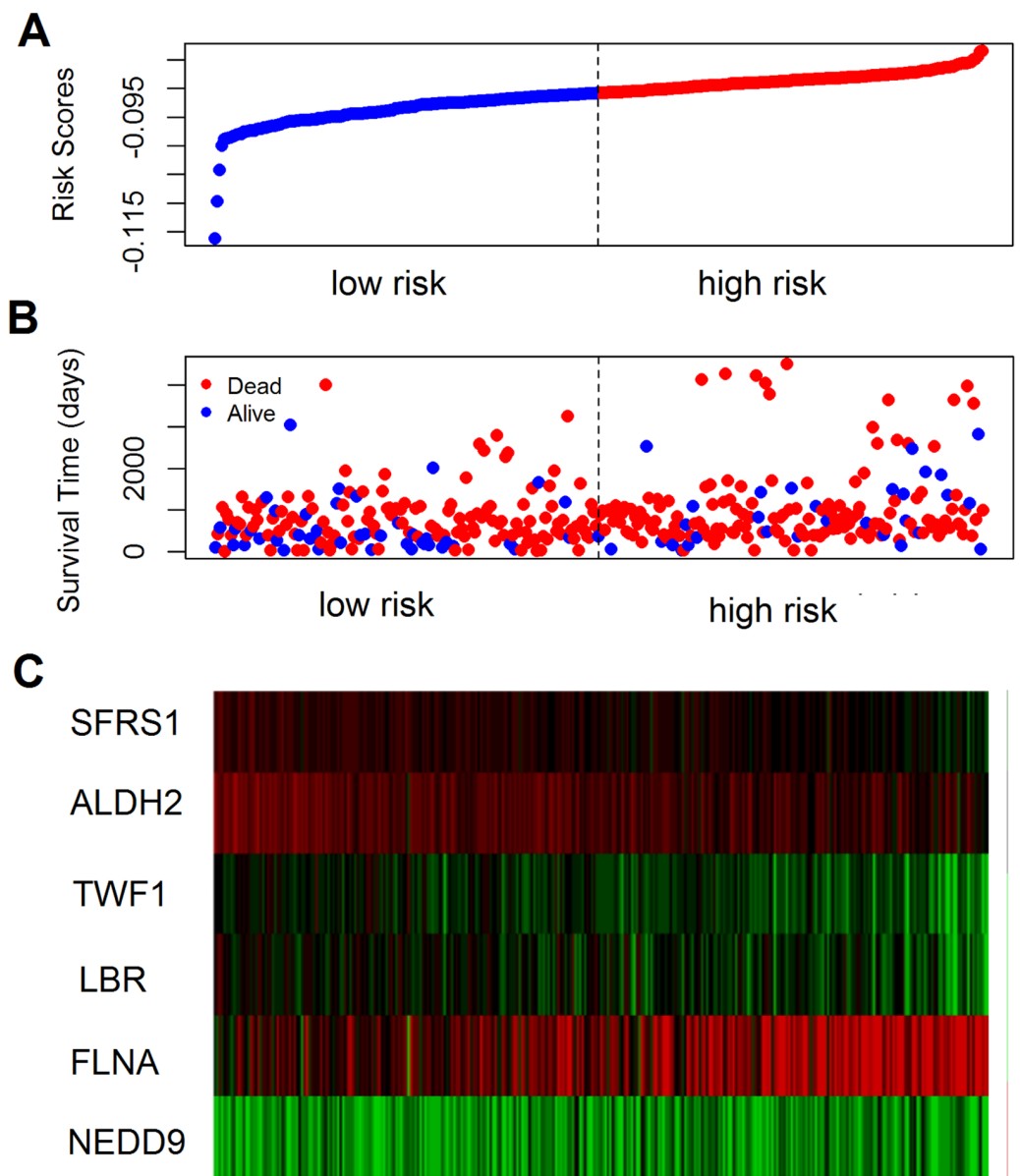


Figure 4 The prognostic performance of the six-gene signature of colon cancer. (A) Risk score curve of the six-gene signature; (B) patient survival status and time distribution by risk score; (C) heatmap showed the expression of six genes by risk score. [Full-size !\[\]\(fcc3264021d438d9732560e78099f674_img.jpg\) DOI: 10.7717/peerj.5433/fig-4](https://doi.org/10.7717/peerj.5433/fig-4)

DISCUSSION

Metastasis accounts for 90% of the mortalities of colon cancer patients and thus has become 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 rate is up to 91.1%), but patients with distant metastasis have a much worse prognosis (the 5-year survival rate has dropped to 13.3%) (*Siegel et al., 2017*). Furthermore, the failure of treatment is mostly caused by the metastatic dissemination of primary tumors (*Deliu, Georgescu & Bezna, 2014; Stein & Schlag, 2007*). At a molecular level, distinct

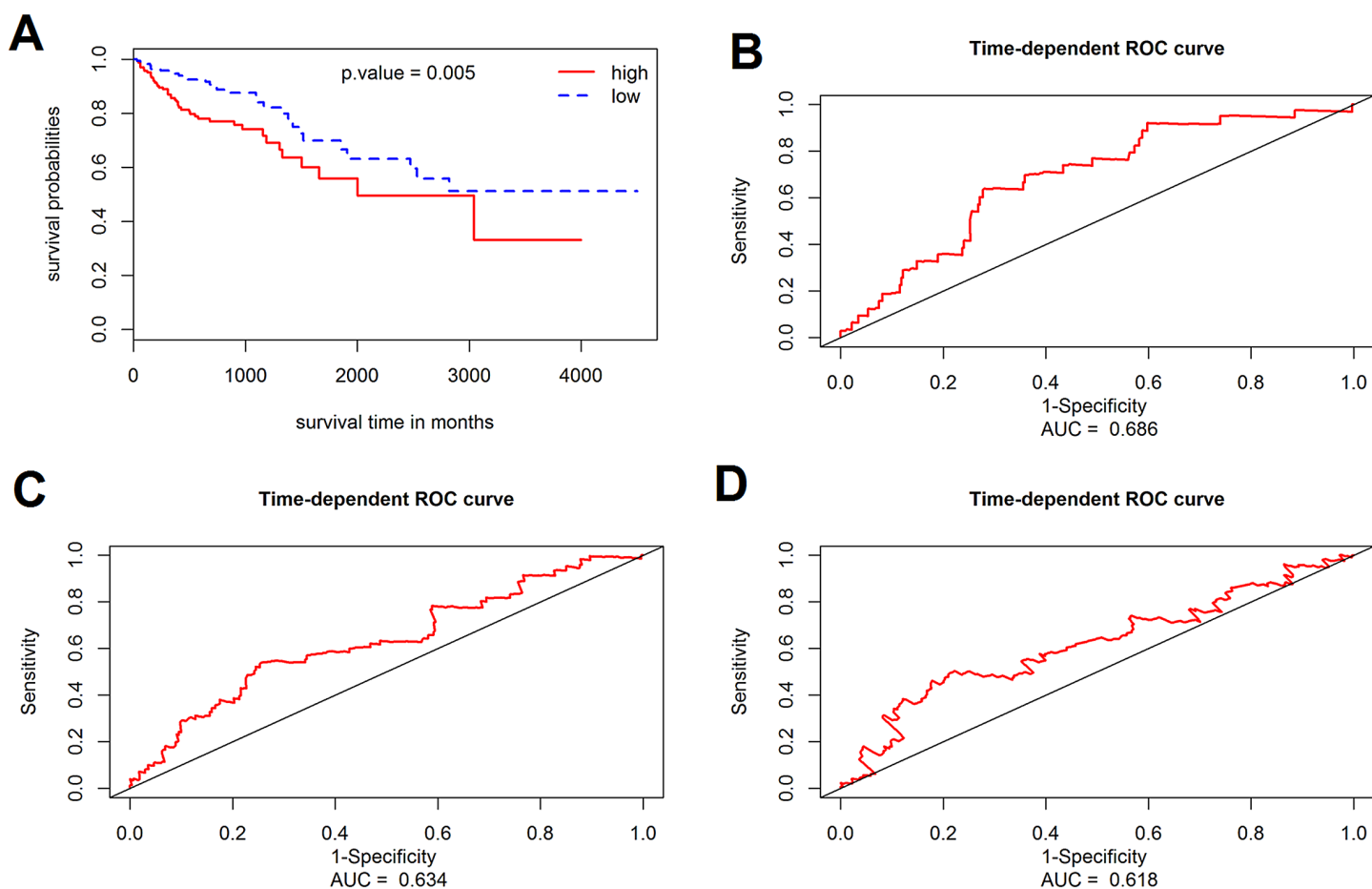


Figure 5 Survival analysis for the gene signature. (A) The Kaplan–Meier test of the risk score for the overall survival of colon cancer between high-risk and low-risk patients; (B–D) the prognostic value of the risk score showed by the time-dependent receiver operating characteristic (ROC) curve for predicting the 1, 3, 5 years overall survival, respectively. [Full-size !\[\]\(1663bb69f307a960345edb0e712f8c02_img.jpg\) DOI: 10.7717/peerj.5433/fig-5](https://doi.org/10.7717/peerj.5433/fig-5)

metastasis of 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 beneficial for the identification of candidate targets of early diagnosis and treatment of advanced stage colon cancer.

To date, some biomarkers have been identified as candidate targets for early diagnosis and treatment of colon cancer and rectal cancer, including genes, miRNA, lncRNA, and the related signatures. Some gene signatures related to the metastasis potential of each tumor have been described, with promising results. Vellinga *et al.* (2017) designed a lymphangiogenic gene set, applied it to large datasets of CRC, and found that this lymphangiogenic gene set was related to a worse prognosis. Rokavec *et al.* (2017) reported a single gene, RNA binding motif protein 47, down-regulated during CRC progression may promote epithelial–mesenchymal transition and metastasis. Furthermore, proteomic studies used exosomes that from cancer cell lines and identified four candidate genes MET proto-oncogene, receptor tyrosine kinase, S100A8, S100A9, tenascin C 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 four hub genes from the subnetwork of the PPI network: 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 adenosine triphosphate (ATP) ([Zhang & Xie, 2017](#)). The role of OXCT1 has been implicated in several cancers, included CRC, and OXCT1 was overexpressed in the metastatic CRC cell line CC-M3 ([Lee et al., 2016](#)). ACTN4 is a non-muscle-type alpha-actinin, which plays an important role in regulating cytoskeleton organization and is involved in transcriptional regulation of gene expression. ACTN4 encodes a nonmuscle, alpha-actinin isoform that is concentrated in the cytoplasm; 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 β_1 subunit to form an intact integrin and interacts with several extracellular matrix proteins ([Nagata et al., 2013](#)). ITGA3 was found to be associated with lymphatic dissemination and local invasiveness in cancers. One study showed that ITGA3 was overexpressed in stages III vs I of CRC patients and is related to 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. Growing evidence suggests 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 dampened anti-tumor immune responses ([David et al., 2016](#)). [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 in 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 are warranted to determine the underlying mechanism.

Similar to the previous studies, in this study, a signature constructed by six genes was shown to be a good predictor of the OS of patients with colon cancer. Among these six genes, the 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. ALDH2 is the rate-limiting enzyme in the ethanol metabolism, oxidizing acetaldehyde to acetic acid both in the liver and other tissues ([Chen, Joshi & Mochly-Rosen, 2016](#)). As a novel biological marker, ALDH2 is an attractive prospect in the screening, diagnosis, and evaluation of the prognosis of many diseases, and the genetic polymorphism of ALDH2 significantly correlated with the susceptibility to CRC ([Li et al., 2016](#)). NEDD9 is a non-catalytic scaffolding protein that assembles complexes involving oncogenic kinases and regulates the magnitude and duration of cell signaling cascades that control multiple processes important to the development

and progression of tumors (*Shagisultanova et al., 2015*). Studies have shown that downregulation of NEDD9 by apigenin can suppresses migration, invasion, and metastasis of CRC cells (*Dai et al., 2016*). With regard to SRSF1, one study reported that phosphorylation of SRSF1 regulated alternative splicing of tumor-related Rac1b in CRC cells (*Goncalves et al., 2014*). SRSF1 is classified as an exonic splicing enhancer and recognizes degenerate purine-rich sequence motifs. In addition, it 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 expressed ubiquitously in the body, involved in many cell-signaling pathways, and it is important in the maintenance of cell shape and motility. The mutation of FLNA has been shown to be associated with the neuronal migration, vascular function, connective tissue integrity, and skeletal development (*Shelmerdine et al., 2017*); FLNA showed low expression in CRC patients and was significantly correlated with the incidence and development of CRC (*Tian et al., 2015*). LBR, a transmembrane protein of the inner nuclear membrane, interacts with heterochromatin and B-type lamins through its nucleoplasmic amino-terminal domain and is phosphorylated throughout the cell cycle (*Duband-Goulet, Courvalin & Buendia, 1998*). TWF1 belongs to the ADF-H family and is a conserved actin-binding protein. TWF1 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 and therefore their role needs to be investigated further 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 CRC cell lines. Thus, the overlapped DEGs could more reliably reflect the gene alteration in metastatic colon cancer. Some limitations of this study need to be noted. First, although metastasis-related genes of colon cancer were identified and their prognostic value was validated in our study, the results were calculated from microarray or RNA-sequencing technique datasets and thus lack of functional validation of the target genes is one of the major limitations of this study. Therefore, thorough functional experiments 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, and microsatellite instability, which have been proven to be closely related to colon cancer, but because of the limitations of the datasets, we did not perform a stratified analysis on the basis of these factors. Future studies should be conducted to explore the difference under different conditions. Third, the mean time of follow-up in the validation cohort was 65.3 months; thus, a study including a longer follow-up time is warranted to validate our results in the future. Fourth, this study only used a GSE dataset to perform bioinformatics to search for biomarkers. Future studies should implement Partek Genomics software and Illumina Correlation Engine for bioset analyses.

CONCLUSIONS

In this study, we screened gene profiles involved in the metastasis of colon cancer and identified four hub genes from the profiles. We also identified and validated a six-gene signature that can serve as a prognostic indicator of colon cancer. Some genes that have not yet been proven to be associated with colon cancer metastasis may represent new therapeutic targets.

ABBREVIATIONS

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.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Scientific research and technology development project of Nanning (20173018-3; 20163120). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Scientific research and technology development project of Nanning; 20173018-3; 20163120.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Haotang Wei conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Jilin Li performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Minzhi Xie performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Ronger Lei conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Bangli Hu conceived and designed the experiments, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data can be downloaded from the GEO database: accession numbers: [GSE40367](#) and [GSE2509](#), and TCGA (colon cancer).

REFERENCES

- 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, Eccles R. 2013. Germline BRCA mutations are associated with higher risk of nodal involvement, distant metastasis, and poor survival outcomes in prostate cancer. *Journal of Clinical Oncology* **31**(14):1748–1757 DOI [10.1200/JCO.2012.43.1882](#).
- Chen CH, Joshi AU, Mochly-Rosen D. 2016. The role of mitochondrial aldehyde dehydrogenase 2 (ALDH2) in neuropathology and neurodegeneration. *Acta Neurol Taiwan* **25**(4):111–123.
- Dai J, Van Wie PG, Fai LY, Kim D, Wang L, Poyil P, Luo J, Zhang Z. 2016. Downregulation of NEDD9 by apigenin suppresses migration, invasion, and metastasis of colorectal cancer cells. *Toxicology and Applied Pharmacology* **311**:106–112 DOI [10.1016/j.taap.2016.09.016](#).
- David JM, Dominguez C, Hamilton DH, Palena C. 2016. The IL-8/IL-8R axis: a double agent in tumor immune resistance. *Vaccines* **4**(3):22 DOI [10.3390/vaccines4030022](#).
- Deliu IC, Georgescu EF, Bezna MC. 2014. Analysis of prognostic factors in colorectal carcinoma. *Revista medico-chirurgicala a Societatii de Medici si Naturalisti din Iasi* **118**(3):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, Gure AO. 2017. A combined ULBP2 and SEMA5A expression signature as a prognostic and predictive biomarker for colon cancer. *Journal of Cancer* **8**(7):1113–1122 DOI [10.7150/jca.17872](#).
- Duband-Goulet I, Courvalin JC, Buendia B. 1998. LBR, a chromatin and lamin binding protein from the inner nuclear membrane, is proteolyzed at late stages of apoptosis. *Journal of Cell Science* **111**(Pt 10):1441–1451.

- Fukumoto M, Kurisu S, Yamada T, Takenawa T. 2015. α -Actinin-4 enhances colorectal cancer cell invasion by suppressing focal adhesion maturation. *PLOS ONE* **10**(4):e0120616 DOI [10.1371/journal.pone.0120616](https://doi.org/10.1371/journal.pone.0120616).
- Goncalves V, Henriques AF, Pereira JF, Neves Costa A, Moyer MP, Moita LF, Gama-Carvalho M, Matos P, Jordan P. 2014. Phosphorylation of SRSF1 by SRPK1 regulates alternative splicing of tumor-related Rac1b in colorectal cells. *RNA* **20**(4):474–482 DOI [10.1261/rna.041376.113](https://doi.org/10.1261/rna.041376.113).
- Heagerty PJ, Lumley T, Pepe MS. 2000. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**(2):337–344 DOI [10.1111/j.0006-341x.2000.00337.x](https://doi.org/10.1111/j.0006-341x.2000.00337.x).
- Hu Y, Chen HY, Yu CY, Xu J, Wang JL, Qian J, Zhang X, Fang JY. 2014. A long non-coding RNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* **5**(8):2230–2242 DOI [10.18632/oncotarget.1895](https://doi.org/10.18632/oncotarget.1895).
- Hur K, Toiyama Y, Schetter AJ, Okugawa Y, Harris CC, Boland CR, Goel A. 2015. Identification of a metastasis-specific MicroRNA signature in human colorectal cancer. *Journal of the National Cancer Institute* **107**(3):dju492 DOI [10.1093/jnci/dju492](https://doi.org/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, 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**(10–11):1672–1686 DOI [10.1002/pmic.201200562](https://doi.org/10.1002/pmic.201200562).
- Knijn N, Tol J, Punt CJ. 2010. Current issues in the targeted therapy of advanced colorectal cancer. *Discovery Medicine* **9**(47):328–336.
- Lambrechts D, Thienpont B, Thuillier V, Sagaert X, Moisse M, Peuteman G, Pericay C, Folprecht G, Zalberg J, Zilocchi C, Margherini E, Chiron M, 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. *British Journal of Cancer* **113**(7):1027–1034 DOI [10.1038/bjc.2015.329](https://doi.org/10.1038/bjc.2015.329).
- Lee CL, Huang CJ, Yang SH, Chang CC, Huang CC, Chien CC, Yang RN. 2016. Discovery of genes from feces correlated with colorectal cancer progression. *Oncology Letters* **12**(5):3378–3384 DOI [10.3892/ol.2016.5069](https://doi.org/10.3892/ol.2016.5069).
- Li H, Chen YX, Wen JG, Zhou HH. 2017. Metastasis-associated in colon cancer 1: a promising biomarker for the metastasis and prognosis of colorectal cancer. *Oncology Letters* **14**(4):3899–3908 DOI [10.3892/ol.2017.6670](https://doi.org/10.3892/ol.2017.6670).
- Li R, Zhao Z, Sun M, Luo J, Xiao Y. 2016. ALDH2 gene polymorphism in different types of cancers and its clinical significance. *Life Sciences* **147**:59–66 DOI [10.1016/j.lfs.2016.01.028](https://doi.org/10.1016/j.lfs.2016.01.028).
- Liew AWC, Law NF, Yan H. 2011. Missing value imputation for gene expression data: computational techniques to recover missing data from available information. *Briefings in Bioinformatics* **12**(5):498–513 DOI [10.1093/bib/bbq080](https://doi.org/10.1093/bib/bbq080).
- Linhares MM, Affonso RJ Jr, Viana Lde S, Silva SR, Denadai MV, De Toledo SR, 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**(12):e0144333 DOI [10.1371/journal.pone.0144333](https://doi.org/10.1371/journal.pone.0144333).
- Marzouk O, Schofield J. 2011. Review of histopathological and molecular prognostic features in colorectal cancer. *Cancers* **3**(2):2767–2810 DOI [10.3390/cancers3022767](https://doi.org/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, Takagi R. 2013. ITGA3 and ITGB4 expression biomarkers estimate the risks of locoregional and

- hematogenous dissemination of oral squamous cell carcinoma. *BMC Cancer* **13**(1):410 DOI [10.1186/1471-2407-13-410](https://doi.org/10.1186/1471-2407-13-410).
- Paavilainen VO, Hellman M, Helfer E, Bovellan M, Annila A, Carlier MF, Permi P, Lappalainen P. 2007.** Structural basis and evolutionary origin of actin filament capping by twinfilin. *Proceedings of the National Academy of Sciences of the United States of America* **104**(9):3113–3118 DOI [10.1073/pnas.0608725104](https://doi.org/10.1073/pnas.0608725104).
- Provenzani A, Fronza R, Loreni F, Pascale A, Amadio M, Quattrone A. 2006.** Global alterations in mRNA polysomal recruitment in a cell model of colorectal cancer progression to metastasis. *Carcinogenesis* **27**(7):1323–1333 DOI [10.1093/carcin/bgi377](https://doi.org/10.1093/carcin/bgi377).
- Qin S, Kim J, Arafat D, Gibson G. 2012.** Effect of normalization on statistical and biological interpretation of gene expression profiles. *Frontiers in Genetics* **3**:160 DOI [10.3389/fgene.2012.00160](https://doi.org/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, Wang XW. 2015.** Integrative genomic and transcriptomic characterization of matched primary and metastatic liver and colorectal carcinoma. *International Journal of Biological Sciences* **11**(1):88–98 DOI [10.7150/ijbs.10583](https://doi.org/10.7150/ijbs.10583).
- Rokavec M, Kaller M, Horst D, Hermeking H. 2017.** Pan-cancer EMT-signature identifies RBM47 down-regulation during colorectal cancer progression. *Scientific Reports* **7**(1):4687 DOI [10.1038/s41598-017-04234-2](https://doi.org/10.1038/s41598-017-04234-2).
- Sameer AS. 2013.** Colorectal cancer: molecular mutations and polymorphisms. *Frontiers in Oncology* **3**:114 DOI [10.3389/fonc.2013.00114](https://doi.org/10.3389/fonc.2013.00114).
- Sanford JR, Wang X, Mort M, Vanduy N, Cooper DN, Mooney SD, Edenberg HJ, Liu Y. 2009.** Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Research* **19**(3):381–394 DOI [10.1101/gr.082503.108](https://doi.org/10.1101/gr.082503.108).
- Shagisultanova E, Gaponova AV, Gabbasov R, Nicolas E, Golemis EA. 2015.** Preclinical and clinical studies of the NEDD9 scaffold protein in cancer and other diseases. *Gene* **567**(1):1–11 DOI [10.1016/j.gene.2015.04.086](https://doi.org/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, Domany E. 2009.** Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America* **106**(17):7131–7136 DOI [10.1073/pnas.0902232106](https://doi.org/10.1073/pnas.0902232106).
- Shelmerdine SC, Semple T, Wallis C, Aurora P, Moledina S, Ashworth MT, Owens CM. 2017.** Filamin A (FLNA) mutation-A newcomer to the childhood interstitial lung disease (ChILD) classification. *Pediatric Pulmonology* **52**(10):1306–1315 DOI [10.1002/ppul.23695](https://doi.org/10.1002/ppul.23695).
- Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, Jemal A. 2017.** Colorectal cancer statistics, 2017. *CA: A Cancer Journal for Clinicians* **67**(3):177–193 DOI [10.3322/caac.21395](https://doi.org/10.3322/caac.21395).
- Siegel RL, Miller KD, Jemal A. 2017.** Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians* **67**(1):7–30 DOI [10.3322/caac.21387](https://doi.org/10.3322/caac.21387).
- Stein U, Schlag PM. 2007.** Clinical, biological, and molecular aspects of metastasis in colorectal cancer. *Recent Results in Cancer Research* **176**:61–80 DOI [10.1007/978-3-540-46091-6_7](https://doi.org/10.1007/978-3-540-46091-6_7).
- Tian ZQ, Shi JW, Wang XR, Li Z, Wang GY. 2015.** New cancer suppressor gene for colorectal adenocarcinoma: filamin A. *World Journal of Gastroenterology* **21**(7):2199–2205 DOI [10.3748/wjg.v21.i7.2199](https://doi.org/10.3748/wjg.v21.i7.2199).

- Vellinga TT, Kranenburg O, Frenkel N, Ubink I, Marvin D, Govaert K, Van Schelven S, Hagendoorn J, Borel Rinkes IH. 2017.** Lymphangiogenic gene expression is associated with lymph node recurrence and poor prognosis after partial hepatectomy for colorectal liver metastasis. *Annals of Surgery* **266**(5):765–771 DOI [10.1097/SLA.0000000000002430](https://doi.org/10.1097/SLA.0000000000002430).
- Xu G, Zhang M, Zhu H, Xu J. 2017.** A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. *Gene* **604**:33–40 DOI [10.1016/j.gene.2016.12.016](https://doi.org/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, Mayo SC. 2017.** Colorectal cancer liver metastasis: evolving paradigms and future directions. *Cellular and Molecular Gastroenterology and Hepatology* **3**(2):163–173 DOI [10.1016/j.jcmgh.2017.01.006](https://doi.org/10.1016/j.jcmgh.2017.01.006).
- Zhang S, Xie C. 2017.** The role of OXCT1 in the pathogenesis of cancer as a rate-limiting enzyme of ketone body metabolism. *Life Sciences* **183**:110–115 DOI [10.1016/j.lfs.2017.07.003](https://doi.org/10.1016/j.lfs.2017.07.003).