

Construction of a co-expression network and prediction of metastasis markers in colorectal cancer patients with liver metastasis

Lihong Lin^{1#}, Xiuxiu Zeng^{1#}, Shanyan Liang¹, Yunzhi Wang², Xiaoyu Dai¹, Yuechao Sun^{1,3*}, Zhou Wu^{1*}

¹Department of Anorectal Surgery, Hwa Mei Hospital, University of Chinese Academy of Sciences, Ningbo, China; ²School of Health Sciences, University of Sydney, Lidcombe, NSW, Australia; ³Ningbo Institute of Life and Health Industry, University of Chinese Academy of Sciences, Ningbo, China

Contributions: (I) Conception and design: Y Sun, Z Wu; (II) Administrative support: Y Sun, Z Wu; (III) Provision of study materials or patients: L Lin, X Zeng; (IV) Collection and assembly of data: S Liang, Y Wang; (V) Data analysis and interpretation: X Zeng, X Dai; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

*These authors share co-corresponding authorship.

Correspondence to: Zhou Wu. Department of Anorectal Surgery, Hwa Mei Hospital, University of Chinese Academy of Sciences, 41 Xibei Street, Ningbo 315000, China. Email: nbsdeyywz283641@126.com; Yuechao Sun. Ningbo Institute of Life and Health Industry, University of Chinese Academy of Sciences, 499 Huancheng North Road, Ningbo 315000, China. Email: sunyuechao@ucas.ac.cn.

Background: Colorectal cancer (CRC) is a common global malignancy associated with high invasiveness, high metastasis, and poor prognosis. CRC commonly metastasizes to the liver, where the treatment of metastasis is both difficult and an important topic in current CRC management.

Methods: Microarrays data of human CRC with liver metastasis (CRCLM) were downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database to identify potential key genes. Differentially expressed (DE) genes (DEGs) and DEmiRNAs of primary CRC tumor tissues and metastatic liver tissues were identified. Microenvironment Cell Populations (MCP)-counter was used to estimate the abundance of immune cells in the tumor micro-environment (TME), and weighted gene correlation network analysis (WGCNA) was used to construct the co-expression network analysis. Gene Ontology and Kyoto Encyclopaedia of Gene and Genome (KEGG) pathway enrichment analyses were conducted, and the protein-protein interaction (PPI) network for the DEGs were constructed and gene modules were screened.

Results: Thirty-five pairs of matched colorectal primary cancer and liver metastatic gene expression profiles were screened, and 610 DEGs (265 up-regulated and 345 down-regulated) and 284 DEmiRNAs were identified. The DEGs were mainly enriched in the complement and coagulation cascade pathways and renin secretion. Immune infiltrating cells including neutrophils, monocytic lineage, and cancer-associated fibroblasts (CAFs) differed significantly between primary tumor tissues and metastatic liver tissues. WGCN analysis obtained 12 modules and identified 62 genes with significant interactions which were mainly related to complement and coagulation cascade and the focal adhesion pathway. The best subset regression analysis and backward stepwise regression analysis were performed, and eight genes were determined, including *F10*, *FGG*, *KNG1*, *MBL2*, *PROC*, *SERPINA1*, *CAV1*, and *SPP1*. Further analysis showed four genes, including *FGG*, *KNG1*, *CAV1*, and *SPP1* were significantly associated with CRCLM.

Conclusions: Our study implies complement and coagulation cascade and the focal adhesion pathway play a significant role in the development and progression of CRCLM, and *FGG*, *KNG1*, *CAV1*, and *SPP1* may be metastatic markers for its early diagnosis.

Keywords: Colorectal liver metastasis; co-expression network; metastasis markers; complement system activation; tumor immune microenvironments

Submitted Sep 23, 2022. Accepted for publication Oct 18, 2022. doi: 10.21037/jgo-22-965 View this article at: https://dx.doi.org/10.21037/jgo-22-965

Introduction

Colorectal cancer (CRC) is one of the most common digestive tract malignancies worldwide, with the third highest incidence and second highest mortality (1). More than 2.2 million new cases of CRC are expected to occur globally by 2030, including more than 1.1 million deaths. Despite the continuous improvement of screening strategies and treatment methods, approximately 25% of CRC patients have already developed metastases at initial diagnosis and half of patients will develop metastatic disease (2,3). The 5-year survival rate of such advanced CRC patients is low, which seriously threatens human health (4). According to medical data, the liver is the most common target organ of metastasis from CRC (5,6).

Between 16% and 26% of patients have liver metastases at the time of diagnosis, and 18–25% will have liver metastases after primary radical resection (7). There has been a significant improvement in the survival of CRC including CRC with liver metastasis (CRCLM) in recent decades. Patients without any metastasis had a 5-year OS of 75.1% compared to 25.2%, 45.7%, and 12.7% respectively for patients with liver-only metastases, lung only metastases, and liver and lung metastases combined.

The key problems of liver metastasis in CRC currently are the lack of serum markers such as carcinoembryonic antigen (CEA), carbohydrate antigen (CA)199 which have high sensitivity and specificity in the diagnosis of CRCLM (8,9). Now the diagnosis of CRCLM still depends on imaging examination and focus biopsy, and the mechanism of CRCLM has not been fully elucidated (10). The treatment of liver metastasis is both difficult and an important topic in current CRC management. Several treatment options are available for patients with CRCLM, of which the surgical resection is the first choice. Other treatments include the chemotherapy, interventional therapy, chemoradiotherapy, targeted therapy and multidisciplinary comprehensive treatment (11-13).

The combination of systemic chemotherapy with resection has been shown to provide a 5-year survival of 50%, while only about 5% for patients treated with palliative intent (14,15). Therefore, there have been many attempts to explore the underlying mechanism of liver metastasis, and gene expression profiling has become

a popular strategy to identify genes involved in the progression and the prognosis of different cancers including CRC (16-19).

Weighted gene co-expression network analysis (WGCNA) is a systematic biology method that describes the correlation patterns between genes in sequencing samples (20) by constructing weighted gene co-expression networks based on gene expression and dividing these into co-expression modules. Gene expression in the same co-expression module is similar and implies the similar functions. WGCNA has been widely used to identify key genes or therapeutic targets in many diseases, including various cancers (21-23).

In this study we used gene expression analysis in the diagnosis of liver metastasis from the aspects of gene molecules and signalling pathways to analyze the interaction between gene proteins for the diagnosis of CRCLM. Our results can be used to help the classification, screening, diagnosis, treatment, and prognosis of CRC. Further the results of potential biomarkers detection can be used to predict or diagnose whether CRC patients may suffer from liver metastasis, providing certain treatment options for subsequent targeted therapy. We present the following article in accordance with the STREGA reporting checklist (available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-965/rc).

Methods

Data extraction and integration

Microarray data of human CRC were downloaded from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm. nih.gov/geo/query) containing 1,374 datasets. We screened the data of CRCLM, and finally filtered two databases, including GSE81558 (Affymetrix Human Gene Expression Array & Affymetrix Multispecies miRNA-3 Array, containing 19 primary CRC and paired liver metastases) and GSE35834 (Affymetrix Human Exon 1.0 ST Array & Affymetrix Multispecies miRNA-1 Array, containing 16 primary CRC and paired liver metastases). Among these 35 pairs of patients, 10 pairs were synchronous liver metastases, and 25 pairs were metachronous liver metastases. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

R package in SilicoMerging (24) was used to merge the two datasets and further removed batch effects (25) to obtain the final dataset, and a total of 35 cases of primary CRC and 35 paired liver metastases were included in the integrated analysis.

Identification of differentially expressed genes (DEGs)

To identify the features of CRCLM, the DEGs between primary CRC tissues and liver metastases tissues were analyzed. In addition, the DEGs between synchronous liver metastases and metachronous liver metastases were identified. Probes were converted into gene symbols, and the average expression value was used as the only value of the gene with multiple corresponding probes. Data were then transformed using a log2 transformation, and the DEGs were identified using limma package at a cut-off value $|log_2[fold$ change (FC)] | >1.2 and an adjusted P<0.05 (26). Kyoto Encyclopaedia of Gene and Genome (KEGG) pathway enrichment analysis of candidate DEGs was performed with the R-package clusterProfiler version 3.14.3. An adjusted P value <0.05 was considered statistically significant.

Construction of the competing endogenous RNA (ceRNA) network

The expression difference between the CRC and CRCLM was utilized to determine the DEmiRNAs and DEGs. The adjusted P value and the absolute log value of fold-change (log₂|FC|) were calculated using R software with limma package. The criteria of P value <0.05 and log₂|FC| >1.2 were adopted to select the DEmiRNAs. Targeted mRNAs of the collected DEmiRNAs were predicted using miRDB (http://www.mirdb.org/). The selected targeted mRNAs were merged with DEGs. The overlapped gene sets were analyzed with Venn Plot, and the pairs of miRNAs and mRNAs were subsequently constructed.

Estimation of the tumor microenvironment (TME)

The Microenvironment Cell Populations (MCP)-counter (27) was used to estimate the abundance of T cells, CD8⁺ T cells, cytotoxic lymphocytes, B lymphocytes, natural killer (NK) cells, monocytic lineage, myeloid dendritic cells, neutrophils, endothelial cells (ECs), and fibroblasts in the TME for each sample. MCP-counter scores were defined as the log2 average expression of the TME for each

population. The difference of the MCP-counter scores of the two groups was compared by Student's *t*-test. A P value <0.05 was considered statistically significant. A higher score means the more significant ratio of the corresponding parts in the TME. In the subsequent WGCNA, these scores of immune infiltrating cells were further combined into module correlation analysis for mining immune-related modules and genes.

WGCNA for the determination of key genes

The WGCNA was used for constructing co-expression network analysis and mining modules related to clinical information (28). We screened genes with the top 25% by analysis of variance (ANOVA) and formed 12 effective modules by hierarchical clustering of the topological overlap matrix. Metastasis related genes were selected from the significant modules associated with metastasis and high level of immune infiltrations by correlation analysis and KEGG pathway enrichment analysis.

Regression models for predicting liver metastases

We used the STRING database to construct the proteinprotein interaction (PPI) network of the above-mentioned genes and screened out hub genes with the most extensive associations with others. Cytoscape (Version 3.9.1) was used to map the network relationships of the above genes. The optimal subset regression (leaps, R) and the backward stepwise regression (MASS, R) were further used to select the core role genes in the process of CRCLM. Finally, the obtained Akaike Information Criterion (AIC) value reached the minimum value and a regression model for predicting liver metastasis in CRC was constructed.

Statistical analysis

To calculate the mean and standard deviation, the numerical data were analyzed by one-way ANOVA. A two-tailed Student's *t*-test was used to verify the significance of the differences between the groups. P value <0.05 was considered as statistical significance.

Results

Data pre-processing

To identify novel gene signatures in CRCLM patients, we

Journal of Gastrointestinal Oncology, Vol 13, No 5 October 2022



Figure 1 Flowchart of construction and analysis of co-expression networks. WGCNA, weighted gene correlation network analysis; MCP, Microenvironment Cell Populations; DEGs, differentially expressed genes; KEGG, Kyoto Encyclopaedia of Gene and Genome; PPI, protein-protein interaction; GSEA, gene set enrichment analysis.

analyzed gene expression profiles by microarray (available at https://cdn.amegroups.cn/static/public/10.21037 jgo-22-965-1.xlsx), and after the batch effects across platforms were removed and adjusted (Figure S1), a total of 17,541 probes were selected for further analysis.

Through the following analysis (*Figure 1*), we constructed the co-expression network and best-fitting logistic regression models, and screened the potential biomarkers associated with CRCLM.

Identification of DEGs between the primary tumor and liver metastases, and establishment of the DEGs-DEmiRNAs regulatory network

To identify the DEGs associated with CRCLM, we performed differential expression analysis on the primary

tumor tissues and metastatic liver tissues of the above 35 CRC patients. As shown in *Figure 2A* and vailable at https://cdn.amegroups.cn/static/public/10.21037 jgo-22-965-2.xlsx, a total of 610 DEGs were identified, of which 265 were up-regulated and 345 were downregulated. After KEGG functional enrichment analysis, the down-regulated DEGs showed significant enrichment in renin secretion, the cyclic adenosine monophosphate (cAMP) signalling pathway, and pathways in cancer (*Figure 2B*), while up-regulated DEGs were enriched in complement and coagulation cascades, chemical carcinogenesis, and retinol metabolism (*Figure 2C*).

We further analyzed the mRNA-miRNA interaction to determine whether some DEGs regulated the liver metastasis by miRNA, and based on the aforementioned cut-off criteria, a set of 284 DEmiRNAs were



Figure 2 Analysis of differential expression genes in CRCLM. (A) The volcano plot shows identification of the DEmRNAs in the primary tumor tissues and metastatic liver tissues. The red colour represents up-regulated genes, while the green colour indicates down-regulated genes. KEGG functional enrichment circle map of down-regulated (B) and up-regulated (C) genes. (D) Construction of the miRNA-mRNA regulatory network. CRCLM, colorectal cancer with liver metastasis; DE, differentially expressed; KEGG, Kyoto Encyclopaedia of Gene and Genome.

identified (available at https://cdn.amegroups.cn/static/ public/10.21037jgo-22-965-3.xlsx), among which only 11 DEmiRNAs have been reported in the miRDB database (available at https://cdn.amegroups.cn/static/ public/10.21037jgo-22-965-4.xlsx). The DEmiRNAsmRNA regulatory network comprised of eight DEmiRNAs, 17 down-regulated DEGs, and six up-regulated DEGs (*Figure 2D*).

As the DEGs were enriched in the complement and coagulation cascade pathways and renin secretion, it was necessary to further explore the relationship between the immune infiltration environment and CRC metastasis. However, no DEGs related to these two pathways were identified in the DEmiRNAs-DEGs regulatory network, suggesting genes associated with these two pathways regulate CRCLM but not through miRNAs.

Differential expression analysis of CRC with synchronous and metachronous liver metastases

As synchronous metastases of CRC are considered to hold worse prognostic value compared with metachronous metastases (29,30), we further explored the potential factors responsible for the differences in tumor behaviour. We identified 186 DEGs in the colorectal tissues of synchronous liver metastases versus metachronous liver metastases from CRC, among which 99 DEGs were up-regulated and 87 were down-regulated (*Figure 3A*)

2430



Figure 3 Differential expression analysis of different liver metastasis types in CRC patients. Heat map of DEGs between synchronous and metachronous liver metastases in primary colorectal tissues (A) and metastatic liver tissues (B) (the top 60 DEGs are shown). (C) Venn diagram shows the up-regulated DEGs and down-regulated DEGs in colorectal tissues and metastatic liver tissues of CRCLM. CRC, colorectal cancer; DEGs, differentially expressed genes; CRCLM, colorectal cancer with liver metastasis.



Figure 4 MCP-counter scores of the infiltrating immune cells in primary colorectal tissues (A) and metastatic liver tissues (B). MCP, Microenvironment Cell Populations; NK, natural killer.

and table available at https://cdn.amegroups.cn/static/ public/10.21037jgo-22-965-5.xlsx). Moreover, 58 upregulated DEGs and 44 down-regulated DEGs were identified in liver tissues between synchronous liver metastases and metachronous liver metastases of CRC (*Figure 3B* and Table S1).

A Venn diagram was then used to display the number of DEGs between the two tissues, and the overlapping DEGs were further analyzed (*Figure 3C*). Notably, *RPL22L1*, *SNCAIP*, *SYNPR*, and *KRT40* were highly expressed in metachronous liver metastases, and *CRISP1*, *PIWIL4*, *TRIM36*, *ATP12A*, and *ZDHHC11B* were highly expressed in synchronous liver metastases.

Immune infiltrating cells may be associated with CRCLM

Immune cell infiltrations were then characterized in colorectal tissues and liver tissues. We first calculated the abundance of immune cells using the MCP-counter algorithm, and the results showed cancer-associated fibroblasts (CAFs) were more abundant than any other cell types in the TME (*Figure 4*) (Table S2). Moreover, there was a significant difference in the level of CAFs between primary tumor tissues and metastatic liver tissues (P=0.04).

While neutrophils showed the most significant difference in MCP-counter score (P=0.0003) between the two tissues, other immune cells including monocytic lineage (P=0.02), ECs (P=0.05), and B lymphocytes (P=0.04) also differed significantly. This indicates there are

significant changes in the immune environment between primary and metastatic lesions, although whether there is a causal relationship or correlation with metastasis needs to be further explored.

Construction of weighted gene co-expression network and PPI network

WGCNA can build gene co-expression networks to mine network modules closely related with clinical traits, and was used to determine the interactions of the correlation between immune infiltrating cells and genes. A total of 12 modules were obtained by clustering the dissimilarity based on the consensus topological overlap, and the clustering dendrograms and module-trait relationship are shown in *Figure 5A* and online (available at https://cdn.amegroups. cn/static/public/10.21037jgo-22-965-6.xlsx).

As there were 441 genes identified in the yellow module and it had the highest association (0.6) with liver metastases in the heat map of the module-trait relationship, this module was selected for further analysis. The results showed 441 genes were significantly enriched in complement and coagulation cascades (Figure S2A) and, interestingly, only neutrophils among the immune infiltration cell types had a strong correlation (0.79) with the module. Based on the STRING database, 145 of the 441 genes had high interactions (combined score >0.4), including 41 genes involved in complement and coagulation cascades (*Figure 5B* and Figure S2B). Collectively, we speculated the



Figure 5 Construction of WGCNA network and PPI network. (A) A gene dendrogram containing 12 modules was obtained by clustering the dissimilarity based on consensus topological overlap. Each module contains a set of highly connected genes and is represented by different colours. The correlation between consensus modules and the clinical indicators (including the abundance of immune factors) of CRCLM was constructed. The intensity and direction of correlation are indicated by different colours (red, positive; blue, negative), the numbers in the table represent the correlation coefficients, and P values are printed in parentheses below. PPI network of genes from the yellow (B) and brown (C) modules. The size of the ellipse corresponds to the combined score among the gene members in the module, and the magenta circles (B) and yellow circles (C) represent the gene sets (KEGG enrichment analysis) of the most enriched pathways in their respective module networks. DC, dendritic cell; WGCNA, weighted gene correlation network analysis; PPI, protein-protein interaction; CRCLM, colorectal cancer with liver metastasis; ME, module eigengene; NK, natural killer; KEGG, Kyoto Encyclopaedia of Gene and Genome.

complement and coagulation cascade pathways might be the key pathways for the liver metastasis of CRC, and the differences may be caused by changes to the TME.

The brown module had a strong correlation with ECs (0.83) and CAFs (0.87), and KEGG analysis (Figure S2C) results demonstrated these genes were mainly involved in the focal adhesion and phosphoinositide 3-kinase (PI3K)-Akt signalling pathways. Further, 158 genes were identified in the brown module (*Figure 5C*), and 21 were involved in regulation of the focal adhesion pathway (Figure S2D).

In summary, we identified the yellow module and brown module as significantly related to CRCLM, and the 62 genes with significant interactions were mainly related to complement and coagulation cascade and focal adhesion pathways.

Construction of best-fitting logistic regression models and identification of potential metastasis markers

Best subset regression selects the best model from all

possible subsets according to some goodness-of-fit criteria (31,32). To find the best fit model for CRCLM from all possible subset models, we first performed optimal subset regression analysis of 41 hub genes associated with complement and coagulation cascades in the yellow module and 10 genes were identified (adjusted R^2 =0.85, Figure S3A). Following the same method, eight genes (adjusted R^2 =0.75) were screened from the set of 21 hub genes related to the focal adhesion pathway in the brown module (Figure S3B). The best subset regression analysis and backward stepwise regression analysis were performed on the above 18 genes, and finally, eight genes (adjusted R^2 =0.9, AIC =-240.56) were determined, including *F10*, *FGG*, *KNG1*, *MBL2*, *PROC*, *SERPINA1*, *CAV1*, and *SPP1* (Figure S3C,S3D).

The best-fitting logistic regression model was then established to calculate the probability of liver metastasis in CRC. When P value >0.5, CRC samples were deemed to be more prone to have liver metastases, while P value <0.5 indicated a low risk of developing liver metastases.

In addition, four genes including *FGG*, *KNG1*, *CAV1*, and *SPP1* were significantly associated with CRCLM (P<0.0001, Figure S3D), and we hypothesized they might be potential metastatic markers for its early diagnosis.

Discussion

CRC is characterized by an aggressive phenotype, poor clinical outcomes, and a high rate of mortality, especially when liver metastasis occurs (33,34). Accordingly, there is an urgent need to find potential molecular biomarkers for CRC that could help to enhance the diagnosis and treatment efficacy of CRCLM.

Recent studies have shown numerous molecular biomarkers in cancer (35-37). For example, KNG1 and FGG have been found to be correlated with the progression and prognosis of hepatocellular carcinoma (38,39), and SPP1 and CAV1 with overall survival in lung adenocarcinoma (40,41). Currently, there are no suitable serum markers with have high sensitivity and specificity in the diagnosis of CRCLM. Tumor markers such as CEA, CA19-9, CA242, CA72-4 in combination with biochemical markers such lactase dehydrogenase (LDH), gamma glutamyl transpeptidase (GGT) for liver function may improve the sensitivity and specificity for screening liver metastases in patients with CRC (42). Combining evaluation of SATB2 with CK20 and CDX2 to form a three-marker panel in liver biopsy tissues by immunohistochemistry could be used to detect the CRCLM (43). Our results revealed 62 genes

Lin et al. Liver metastasis-associated genes in human CRC

with significant interactions which were mainly related to complement and coagulation cascade and the focal adhesion pathway, and the *FGG*, *KNG1*, *CAV1*, and *SPP1* genes were significantly associated with liver metastasis in CRC. Taken together, these findings suggest these genes might be metastatic markers for the early diagnosis of CRCLM.

Gene expression can be influenced by different factors including miRNAs (44,45), and to explore the possible roles of miRNAs in the complement and coagulation cascade pathways and renin secretion, we further analyzed the mRNA-miRNA interaction in miRNA array. However, no DEGs related to these two pathways were identified in the DEmiRNAs-DEGs regulatory network, suggesting the genes of these two pathways may not regulate CRC metastasis through miRNAs.

Previous studies have revealed the relationships between immune cell infiltration, tumor purity, and biomarker gene expression to be quite valuable for developing appropriate immunotherapy (46,47). Studies have explored the correlations between tumor purity and marker genes expression to predict the clinical outcomes in different cancers (47,48). Liu et al. have assessed the immune cells in normal and cancerous human head and neck squamous cell carcinoma (HNSC) tissues using the CIBERSORT method (49) and found different cells, including neutrophil, B cells, CD4⁺ T cells, and CD8⁺ T cells were increased in the cancerous tissues compared with normal controls. Consistently, in our study, we also uncovered significant differences in neutrophils, monocytic lineage, and CAFs between primary tumor tissue and their liver metastases, which may help clinicians gain deeper insights into the TME landscape of CRCLM (Figure 6).

TME is a population of cells constituted with tumor cells and stromal cells recruited by tumor cells including T cells, macrophages, fibroblasts and so on (50). In CRCLM, the liver-infiltrating immune cells include neutrophils, myeloidderived suppressor cells (MDSCs), Monocyte, tumorassociated macrophages (TAMs), metastasis-associated macrophages (MAMs), NK cells and so on (51). Moreover, the infiltrating immune cells found in the primary tumor of CRC differs from those in liver metastases in terms of their spatial distribution. More immunosuppressive cells are present in the liver metastases, with the most pronounced differential distribution found for macrophages (52). The infiltrated immune cells in the TME are closely associated with tumorigenesis and metastasis, as well as the cellular response to therapy. CD3⁺ and CD8⁺ T cell density correlates with better prognosis and high regulatory Journal of Gastrointestinal Oncology, Vol 13, No 5 October 2022



Figure 6 The complement and focal adhesion pathway in the TME. MBL, mannan-binding lectin; MASP1, MBL-associated serine protease 1; MASP2, MBL-associated serine protease 2; MAC, membrane attack complex; FAC, focal adhesion complex; IACs, integrin adhesion complexes; NK, natural killer; TME, tumor micro-environment.

T cell (Treg cell) infiltrate correlates with shorter overall survival (53,54). Both the density and morphology of TAMs correlate with survival in patients with CRCLM (55,56). Lymphocyte infiltration and plasma cell infiltration are linked to prolonged patient survival in CRCLM (57).

The composition of the TME includes complex immune factors which play an important role in liver metastasis of CRC. The lectin pathway is one of three pathways of complement pathway and sees MBL and MASP2 form a complex, which cleaves C4 and C2 and generates C3 convertase. C3 convertase promotes the proliferative activation of complement and promotes the formation of C5 convertase on the cell surface, which binds to C6, C7, and C8, forming a C5b-8 complex, which polymerizes several C9 molecules, forming the membrane attack complex (MAC). Complement system activation and generation of anaphylatoxins orchestrate an inflammatory and immune responses. Anaphylatoxins activate macrophages, neutrophils, and monocytic lineages, resulting in the production of cytokines which increase vascular permeability and enhance neutrophil extravasation and chemotaxis (58-60). ECs and fibroblasts are also key players

in the TME and produce complement proteins, express regulators, and a lower level of complement receptors.

Force sensing between cells and cell matrix via integrin and their associated integrin adhesion complexes (IACs) is involved in cell migration, matrix remodelling, and mechanosensing (61). In addition, focal adhesion kinase (FAK), regulated by integrin signalling, plays an important role in promoting TME remodelling, and activated FAK can promote angiogenesis in ECs (62,63) and recruit immune cells such as fibroblasts (64). However, direct evidence on how genes including *FGG*, *KNG1*, *CAV1*, and *SPP1* regulate liver metastasis and immune cells infiltration in CRC is lacking, and the precise pathways and mechanisms requires further study.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (No. 42006093), the Medical Health Science and Technology project of Ningbo (No. 2021Y31), the Natural Science Foundation of Ningbo (No. 2021J323), the Research Initiation Project of the Ningbo

Lin et al. Liver metastasis-associated genes in human CRC

Institute of Life and Health Industry, the Chinese Academy of Sciences (Nos. 2021YJY1010 and 2022YJY0204), the Ningbo Medical Key Discipline (No. 2022-B11), and the Public Welfare Foundation of Ningbo (No. 2021S10).

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-965/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jgo.amegroups.com/article/view/10.21037/jgo-22-965/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
- Carlomagno C, De Stefano A, Rosanova M, et al. Multiple treatment lines and prognosis in metastatic colorectal cancer patients. Cancer Metastasis Rev 2019;38:307-13.
- Van Cutsem E, Oliveira J; ESMO Guidelines Working Group. Advanced colorectal cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol 2009;20 Suppl 4:61-3.
- 4. Li N, Lu B, Luo C, et al. Incidence, mortality, survival, risk factor and screening of colorectal cancer: A comparison

among China, Europe, and northern America. Cancer Lett 2021;522:255-68.

- Siegel RL, Miller KD, Fuchs HE, et al. Cancer Statistics, 2021. CA Cancer J Clin 2021;71:7-33.
- Engstrand J, Nilsson H, Strömberg C, et al. Colorectal cancer liver metastases - a population-based study on incidence, management and survival. BMC Cancer 2018;18:78.
- 7. Horn SR, Stoltzfus KC, Lehrer EJ, et al. Epidemiology of liver metastases. Cancer Epidemiol 2020;67:101760.
- Michl M, Taverna F, Kumbrink J, et al. Biomarker alterations associated with distinct patterns of metastatic spread in colorectal cancer. Virchows Arch 2021;478:695-705.
- Bai R, Shi Z, Li D, et al. Gene expression profile of human colorectal cancer identified NKTR as a biomarker for liver metastasis. Aging (Albany NY) 2022;14:6656-67.
- Zhou H, Liu Z, Wang Y, et al. Colorectal liver metastasis: molecular mechanism and interventional therapy. Signal Transduct Target Ther 2022;7:70.
- Zheng W, Wu F, Fu K, et al. Emerging Mechanisms and Treatment Progress on Liver Metastasis of Colorectal Cancer. Onco Targets Ther 2021;14:3013-36.
- 12. Alhumaid A, AlYousef Z, Bakhsh HA, et al. Emerging paradigms in the treatment of liver metastases in colorectal cancer. Crit Rev Oncol Hematol 2018;132:39-50.
- Van Cutsem E, Cervantes A, Adam R, et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. Ann Oncol 2016;27:1386-422.
- Nordlinger B, Sorbye H, Glimelius B, et al. Perioperative FOLFOX4 chemotherapy and surgery versus surgery alone for resectable liver metastases from colorectal cancer (EORTC 40983): long-term results of a randomised, controlled, phase 3 trial. Lancet Oncol 2013;14:1208-15.
- Xu YXX, Ye YB, Chen T, et al. Role of hepatic surgery in colorectal cancer multiple liver metastasis. Zhonghua Wai Ke Za Zhi 2021;59:816-20.
- Chakravarty D, Solit DB. Clinical cancer genomic profiling. Nat Rev Genet 2021;22:483-501.
- Moncada R, Barkley D, Wagner F, et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat Biotechnol 2020;38:333-42.
- Latha NR, Rajan A, Nadhan R, et al. Gene expression signatures: A tool for analysis of breast cancer prognosis and therapy. Crit Rev Oncol Hematol 2020;151:102964.
- 19. Mousavi M, Goodarzi MT, Kassaee SM, et al.

2436

Journal of Gastrointestinal Oncology, Vol 13, No 5 October 2022

Clinicopathological, Immunohistochemical, and PMS2 Gene Expression Profiling of Patients with Sporadic Colorectal Cancer. Arch Iran Med 2021;24:86-93.

- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.
- Zhang Y, Luo J, Liu Z, et al. Identification of hub genes in colorectal cancer based on weighted gene co-expression network analysis and clinical data from The Cancer Genome Atlas. Biosci Rep 2021;41:BSR20211280.
- 22. Zheng H, Liu H, Li H, et al. Weighted Gene Coexpression Network Analysis Identifies a Cancer-Associated Fibroblast Signature for Predicting Prognosis and Therapeutic Responses in Gastric Cancer. Front Mol Biosci 2021;8:744677.
- Chen M, Chen S, Yang D, et al. Weighted Gene Coexpression Network Analysis Identifies Crucial Genes Mediating Progression of Carotid Plaque. Front Physiol 2021;12:601952.
- Taminau J, Meganck S, Lazar C, et al. Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. BMC Bioinformatics 2012;13:335.
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8:118-27.
- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
- 27. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol 2016;17:218.
- Su R, Jin C, Zhou L, et al. Construction of a ceRNA network of hub genes affecting immune infiltration in ovarian cancer identified by WGCNA. BMC Cancer 2021;21:970.
- van Rooijen KL, Shi Q, Goey KKH, et al. Prognostic value of primary tumour resection in synchronous metastatic colorectal cancer: Individual patient data analysis of firstline randomised trials from the ARCAD database. Eur J Cancer 2018;91:99-106.
- Bakkers C, Lurvink RJ, Rijken A, et al. Treatment Strategies and Prognosis of Patients With Synchronous or Metachronous Colorectal Peritoneal Metastases: A Population-Based Study. Ann Surg Oncol 2021;28:9073-83.
- 31. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam

Med Community Health 2020;8:e000262.

- 32. Zhang Z. Variable selection with stepwise and best subset approaches. Ann Transl Med 2016;4:136.
- Liu W, Zhang W, Xu Y, et al. A Prognostic Scoring System to Predict Survival Outcome of Resectable Colorectal Liver Metastases in this Modern Era. Ann Surg Oncol 2021;28:7709-18.
- Nishioka Y, Kawaguchi Y, Kothari AN, et al. Prognostic and Therapeutic Implications of Tumor Biology, Including Gene Alterations, in Colorectal Liver Metastases. J Gastrointest Surg 2021;25:1591-600.
- Su CY, Huang GC, Chang YC, et al. Analyzing the Expression of Biomarkers in Prostate Cancer Cell Lines. In Vivo 2021;35:1545-8.
- 36. Nunes-Xavier CE, Zaldumbide L, Mosteiro L, et al. Protein Tyrosine Phosphatases in Neuroblastoma: Emerging Roles as Biomarkers and Therapeutic Targets. Front Cell Dev Biol 2021;9:811297.
- Lou J, Zhang L, Lv S, et al. Biomarkers for Hepatocellular Carcinoma. Biomark Cancer 2017;9:1-9.
- 38. Bai Q, Liu H, Guo H, et al. Identification of Hub Genes Associated With Development and Microenvironment of Hepatocellular Carcinoma by Weighted Gene Coexpression Network Analysis and Differential Gene Expression Analysis. Front Genet 2020;11:615308.
- Gu Y, Li J, Guo D, et al. Identification of 13 Key Genes Correlated With Progression and Prognosis in Hepatocellular Carcinoma by Weighted Gene Coexpression Network Analysis. Front Genet 2020;11:153.
- Luo X, Feng L, Xu W, et al. Weighted gene co-expression network analysis of hub genes in lung adenocarcinoma. Evol Bioinform Online 2021;17:11769343211009898.
- 41. Wang H, Zhang Z, Xu K, et al. Exploration of estrogen receptor-associated hub genes and potential molecular mechanisms in non-smoking females with lung adenocarcinoma using integrated bioinformatics analysis. Oncol Lett 2019;18:4605-12.
- Wu XZ, Ma F, Wang XL. Serological diagnostic factors for liver metastasis in patients with colorectal cancer. World J Gastroenterol 2010;16:4084-8.
- Zhang YJ, Chen JW, He XS, et al. SATB2 is a Promising Biomarker for Identifying a Colorectal Origin for Liver Metastatic Adenocarcinomas. EBioMedicine 2018;28:62-9.
- Ali Syeda Z, Langden SSS, Munkhzul C, et al. Regulatory Mechanism of MicroRNA Expression in Cancer. Int J Mol Sci 2020;21:1723.
- 45. Mollaei H, Safaralizadeh R, Rostami Z. MicroRNA replacement therapy in cancer. J Cell Physiol

Lin et al. Liver metastasis-associated genes in human CRC

2438

2019;234:12369-84.

- 46. Lawal B, Lin LC, Lee JC, et al. Multi-Omics Data Analysis of Gene Expressions and Alterations, Cancer-Associated Fibroblast and Immune Infiltrations, Reveals the Onco-Immune Prognostic Relevance of STAT3/CDK2/4/6 in Human Malignancies. Cancers (Basel) 2021;13:954.
- Gong Z, Zhang J, Guo W. Tumor purity as a prognosis and immunotherapy relevant feature in gastric cancer. Cancer Med 2020;9:9052-63.
- Deng Y, Song Z, Huang L, et al. Tumor purity as a prognosis and immunotherapy relevant feature in cervical cancer. Aging (Albany NY) 2021;13:24768-85.
- Liu G, Yuan C, Ma J, et al. Influence of Immune Microenvironment on Diagnosis and Prognosis of Head and Neck Squamous Cell Carcinoma. Front Oncol 2021;11:604784.
- Baghban R, Roshangar L, Jahanban-Esfahlan R, et al. Tumor microenvironment complexity and therapeutic implications at a glance. Cell Commun Signal 2020;18:59.
- 51. Zeng X, Ward SE, Zhou J, et al. Liver Immune Microenvironment and Metastasis from Colorectal Cancer-Pathogenesis and Therapeutic Perspectives. Cancers (Basel) 2021;13:2418.
- 52. He Y, Han Y, Fan AH, et al. Multi-perspective comparison of the immune microenvironment of primary colorectal cancer and liver metastases. J Transl Med 2022;20:454.
- Wang Y, Lin HC, Huang MY, et al. The Immunoscore system predicts prognosis after liver metastasectomy in colorectal cancer liver metastases. Cancer Immunol Immunother 2018;67:435-44.
- Katz SC, Bamboat ZM, Maker AV, et al. Regulatory T cell infiltration predicts outcome following resection of colorectal cancer liver metastases. Ann Surg Oncol 2013;20:946-55.
- 55. Donadon M, Torzilli G, Cortese N, et al. Macrophage morphology correlates with single-cell diversity and

Cite this article as: Lin L, Zeng X, Liang S, Wang Y, Dai X, Sun Y, Wu Z. Construction of a co-expression network and prediction of metastasis markers in colorectal cancer patients with liver metastasis. J Gastrointest Oncol 2022;13(5):2426-2438. doi: 10.21037/jgo-22-965 prognosis in colorectal liver metastasis. J Exp Med 2020;217:e20191847.

- Cavnar MJ, Turcotte S, Katz SC, et al. Tumor-Associated Macrophage Infiltration in Colorectal Cancer Liver Metastases is Associated With Better Outcome. Ann Surg Oncol 2017;24:1835-42.
- 57. Kurebayashi Y, Kubota N, Sakamoto M. Immune microenvironment of hepatocellular carcinoma, intrahepatic cholangiocarcinoma and liver metastasis of colorectal adenocarcinoma: Relationship with histopathological and molecular classifications. Hepatol Res 2021;51:5-18.
- 58. Holers VM. Complement and its receptors: new insights into human disease. Annu Rev Immunol 2014;32:433-59.
- 59. Afshar-Kharghan V. The role of the complement system in cancer. J Clin Invest 2017;127:780-9.
- 60. Lu P, Ma Y, Wei S, et al. The dual role of complement in cancers, from destroying tumors to promoting tumor development. Cytokine 2021;143:155522.
- 61. Humphries JD, Chastney MR, Askari JA, et al. Signal transduction via integrin adhesion complexes. Curr Opin Cell Biol 2019;56:14-21.
- 62. Wu HJ, Hao M, Yeo SK, et al. FAK signaling in cancerassociated fibroblasts promotes breast cancer cell migration and metastasis by exosomal miRNAs-mediated intercellular communication. Oncogene 2020;39:2539-49.
- 63. Chen XL, Nam JO, Jean C, et al. VEGF-induced vascular permeability is mediated by FAK. Dev Cell 2012;22:146-57.
- 64. Wang X, Zhou Q, Yu Z, et al. Cancer-associated fibroblast-derived Lumican promotes gastric cancer progression via the integrin β1-FAK signaling pathway. Int J Cancer 2017;141:998-1010.

(English Language Editor: B. Draper)

Table S1 DEGs identified in the liver tissues of CRC patients by comparing the synchronous and metachronous liver metastases

Тад	logFC	AveExpr	t	P value	adj.P.Val	В
PAGE2	-1.0588	8.834422	-2.24183	0.03189	0.999863	-3.51181
GRPR	-0./3281	10.40107	-2.68668	0.011006	0.999863	-2.78859
SI C5A1	-0.69059	13.64516	-2.12049	0.0412	0.999863	-3.6792
PCDHGA10	-0.60875	12.88078	-2.36057	0.024008	0.999863	-3.31814
MAGEC1	-0.51665	10.84899	-3.66535	0.00082	0.999863	-1.00408
CAPN6	-0.50413	12.20878	-2.4412	0.019892	0.999863	-3.19115
SYNPR	-0.48609	10.98673	-2.19649	0.034842	0.999863	-3.56779
FZD3	-0.47134	12.80018	-2.26848	0.029641	0.999863	-3.45974
GDPD1	-0.46835	12.23495	-2.74397	0.009548	0.999863	-2.69144
SNCAIP	-0.4631	12.71017	-2.44112	0.019896	0.999863	-3.19129
CEL	-0.46104	12 24604	-2.09391	0.02089	0.999863	-3.71751
GLB1L3	-0.45568	11.07772	-2.89959	0.006445	0.999863	-2.42204
SHISA9	-0.45109	13.05117	-3.49356	0.001323	0.999863	-1.33282
TUBB8	-0.45079	11.062	-2.34694	0.024776	0.999863	-3.33933
KRT40	-0.45029	11.63853	-2.58329	0.01417	0.999863	-2.96093
ANXA13	-0.44575	13.21112	-2.89741	0.006481	0.999863	-2.42586
KLHL32	-0.42718	11.55337	-2.37344	0.023303	0.999863	-3.29805
PPBP	-0.42602	12.54828	-2.11922	0.041314	0.999863	-3.68103
MAGEC2	-0.41789	11.1698	-2.05843	0.047131	0.999863	-3.7681
	-0.4166	9.241159	-2.32105	0.026295	0.999863	-3.37937
KRT75	-0.38235	11.93962	-2.21422	0.003756	0.999863	-2.05086
OR1A2	-0.36087	11.34791	-2.11318	0.041862	0.999863	-3.68976
OR10A5	-0.34537	11.14506	-2.53951	0.015748	0.999863	-3.03271
C20orf96	-0.34393	12.32322	-2.24036	0.031584	0.999863	-3.50223
ANGPTL5	-0.33751	7.421051	-2.17226	0.036767	0.999863	-3.6036
WFDC11	-0.33615	10.45289	-2.20951	0.033844	0.999863	-3.54842
MUC7	-0.33369	10.82261	-2.10059	0.043025	0.999863	-3.70791
EREG	-0.33121	13.95892	-2.41999	0.020908	0.999863	-3.22483
ZNF214	-0.32956	11.58271	-2.10793	0.042344	0.999863	-3.69734
USP29	-0.31581	12.00768 10.05404	-2.23059 -2 17741	0.031924	0.000863	-3.5094 -3.59601
TCEAL6	-0.31393	12.38811	-3.44433	0.001515	0.999863	-1.4261
AADACL2	-0.30081	10.27	-2.14081	0.039406	0.999863	-3.64967
CFTR	-0.29948	14.04541	-2.3174	0.026516	0.999863	-3.38499
WFDC9	-0.29731	10.84212	-2.81509	0.007988	0.999863	-2.56929
OR10G8	-0.29592	12.18128	-2.04379	0.048636	0.999863	-3.7888
ASB14	-0.28425	10.65682	-2.64965	0.012055	0.999863	-2.85077
IL36G	-0.27303	11.93714	-2.49854	0.017368	0.999863	-3.09921
DSC2	-0.26774	13.8035	-2.43185	0.020334	0.999863	-3.20602
GISFIL TMEM14EP	-0.26605	11.07386	-2.08572	0.039652	0.999863	-3.72924
ZSCAN5A	0.263856	12.0826	2.251756	0.030783	0.999863	-3.48506
ISL1	0.267669	11.28736	2.289225	0.028278	0.999863	-3.42817
CDCA5	0.268206	13.12138	2.419712	0.020921	0.999863	-3.22526
РХК	0.27467	13.07765	2.088919	0.044129	0.999863	-3.72466
RPTN	0.27978	11.00448	2.412981	0.021254	0.999863	-3.23591
GFPT2	0.281686	12.63883	2.248426	0.031015	0.999863	-3.49008
TEX13A	0.284459	11.87127	2.6339	0.012529	0.999863	-2.87706
NKAIN3	0.288958	10.93199	2.043366	0.048679	0.999863	-3.78939
EPHX3 FAM167A	0.290299	12.66914	2.640093	0.012341	0.999863	-2.86673
ZNF75D	0.295719	12.36248	2.137618	0.039683	0.999863	-3.65433
IQCF2	0.308313	10.60545	2.62379	0.012842	0.999863	-2.89389
CLEC9A	0.312115	10.75696	2.16988	0.036961	0.999863	-3.60711
ZNF655	0.312992	13.31962	2.621054	0.012928	0.999863	-2.89844
POU4F3	0.316591	11.97574	2.379243	0.022992	0.999863	-3.28898
SNORA77	0.327117	12.8573	3.170393	0.003179	0.999863	-1.93606
TRIM36	0.33054	12.42079	2.089804	0.044045	0.999863	-3.72339
SH2D1B	0.330661	12.60858	2.083214	0.044678	0.999863	-3.73282
ADGRE1	0.336423	11.84148	2.407019	0.021552	0.999863	-3.24532
PIWIL4	0.345199	13.45551	2.664765	0.011616	0.999863	-2.82545
KBTBD8	0.345277	11.20389	2.087047	0.044309	0.999863	-3.72734
ARNTL	0.348979	13.09691	4.406286	9.66E-05	0.789844	0.452101
BNIP3	0.352699	13.88717	2.19846	0.034896	0.999863	-3.56792
CRISP1	0.353202	9.863335	2.493978	0.017558	0.999863	-3.10658
	0.358893	9.159359	2.099194	0.043156	0.999863	-3.70992
GAD1	0.365212	11 90011	2 27235	0.029382	0.999863	-3 45387
PPIL6	0.366826	11.92791	2.426618	0.020585	0.999863	-3.21432
CDC14C	0.367293	10.77132	2.733105	0.00981	0.999863	-2.70995
HUNK	0.381564	12.76172	2.68875	0.01095	0.999863	-2.7851
PLEKHG1	0.3935	12.76051	2.629646	0.01266	0.999863	-2.88415
RNASE3	0.395957	11.72401	2.258932	0.030288	0.999863	-3.47421
SDR42E1	0.417584	12.83105	3.271672	0.002424	0.999863	-1.74943
CYP26A1	0.428998	11.8507	2.2282	0.032458	0.999863	-3.52049
SLAIN1	0.44549	12.2229	2.379446	0.022981	0.999863	-3.28866
CD226	0.465618	11 60287	2 279453	0.000135	0.769644	-3 44307
DPH3P1	0.487039	9.846845	2.097148	0.043348	0.999863	-3.71286
CXorf58	0.501813	9.168746	2.226285	0.032598	0.999863	-3.52336
TAS2R31	0.504158	11.09448	2.5893	0.013966	0.999863	-2.95103
KIAA0319	0.506881	12.32017	2.635829	0.01247	0.999863	-2.87385
RPGRIP1	0.524015	10.06052	2.571256	0.014589	0.999863	-2.98074
SPAG16	0.531744	12.71616	2.487472	0.017831	0.999863	-3.11706
MIR99AHG	0.570395	9.453328	2.132419	0.040138	0.999863	-3.66189
ADGRG2	0.579713	11.264	2.115231	0.041675	0.999863	-3.6868
	U.588606	8.623054	2.078193	0.045166	0.999863	-3./3999
ZDHHC11B	0.633251	12.8559	2.568465	0.014688	0.999863	-2.98533
SPINK7	0.672686	8.644482	2.068891	0.046082	0.999863	-3.75325
CNTLN	0.680737	11.48618	2.691111	0.010886	0.999863	-2.78112
FUT9	0.699768	8.441847	4.407687	9.62E-05	0.789844	0.454887
IL13RA2	0.719021	10.39282	2.628168	0.012838	0.999863	-2.89482
ASPA	0.749501	11.36381	2.245045	0.031252	0.999863	-3.49517
OR5H15	0.750191	8.460118	2.521068	0.016459	0.999863	-3.06273
TAS2R46	0.853411	9.728956	2.104017	0.043163	0.999863	-3.70747
ILSKA	0.907942	11.46483	2.058268	0.047148	0.999863	-3.76833
C19orf18	0 92955	10 87/52	2 51527	0.016685	0 000863	_3 07108

DEGs, differentially expressed genes; FC, fold change; CRC, colorectal cancer.

© Journal of Gastrointestinal Oncology. All rights reserved.

https://dx.doi.org/10.21037/jgo-22-965



Figure S1 Comparison of GEO data before and after merging. (A) Data distribution plot before removing batch effects. (B) Data distribution plot after removing batch effects. (C) Intersection of genes between the two datasets. The first column represents the number of genes in the two GEO datasets, the second column represents the unique elements of the one set, and the third column represents the element values corresponding to the intersection. GEO, Gene Expression Omnibus.

Table S2 The score of immune cells using the MCP-counter algorithm and the t-test results between the primary tumor tissues and metastatic liver tissues in CRC patients

GSE –	T cells		CD8 ⁺ T cells		Cytotoxic lymphocytes		B lymphocytes		NK cells		Monocytic lineage		Myeloid dendritic cells		Neutrophils		Endothelial cells		CAFs	
	Colon	Liver	Colon	Liver	Colon	Liver	Colon	Liver	Colon	Liver	Colon	Liver	Colon	Liver	Colon	Liver	Colon	Liver	Colon	Liver
GSM875949	9492.66	8637.84	3693.20	4080.81	4803.36	4569.50	8065.97	6431.44	4789.10	6077.04	11749.03	8513.35	5118.98	4768.10	8079.20	8102.92	9690.91	8212.87	16875.99	14451.03
GSM875961	7765.78	8587.02	3091.83	2810.58	4573.58	4569.23	6112.37	6595.73	5169.89	6055.82	9303.02	9765.86	4363.34	5346.13	7849.02	8832.61	8431.24	8693.21	16190.20	14005.01
GSM875951	8342.42	9365.74	3406.83	2841.26	6646.76	8171.18	5959.64	6576.60	6605.05	5533.79	8333.74	15971.09	4468.91	6397.20	8838.85	9625.56	9329.17	9821.49	16215.43	15960.41
GSM876001	7366.96	9877.54	3465.13	3719.79	4243.84	5136.33	5747.41	6626.74	4348.40	5782.70	9079.73	13262.40	4260.90	5033.66	8018.96	9481.70	9001.80	8815.40	15406.68	16142.36
GSM875953	7607.24	8130.31	3301.49	3663.54	4065.95	4773.36	5677.82	6254.73	4636.21	5239.75	7951.48	10675.24	4519.26	5145.72	8091.63	8116.45	8762.41	8917.77	15295.46	15175.63
GSM875962	8536.45	8498.39	3075.47	2878.08	3928.01	4016.84	7606.82	6088.89	5106.47	4388.43	9586.65	14443.02	4900.99	5275.43	7899.25	9465.82	8747.16	9108.08	12474.54	15436.15
GSM875956	8456.24	8226.21	3473.31	3146.04	4253.95	4764.65	6268.97	6407.26	4325.25	5471.09	10445.89	12042.05	5164.56	4926.88	8515.18	8820.36	8776.67	9025.85	15630.20	15120.82
GSM876003	8414.17	8473.72	3090.81	2530.35	4738.99	4848.27	7367.74	6403.97	5423.01	5554.09	7817.05	12503.13	5180.39	5521.63	8535.21	8259.33	9149.07	8888.64	14552.67	15912.07
GSM875959	8002.73	7993.15	3256.49	2229.67	4470.12	4751.31	6459.34	6014.25	5386.25	5556.68	8456.24	10480.94	5213.45	5097.59	8369.02	8518.32	8926.32	8801.55	14499.52	15618.29
GSM875990	7641.90	7431.30	3231.94	4517.52	3916.58	4290.41	6019.57	5489.87	4923.79	4780.07	7510.15	9808.47	4329.74	4470.32	8641.31	9078.01	8319.57	8746.40	13873.95	15043.24
GSM875980	8568.10	8065.26	2611.15	1552.62	4347.07	4366.81	6638.04	6021.03	5738.92	4242.70	11146.44	10636.59	5758.87	4461.42	8250.64	9743.16	8187.24	8013.54	11252.59	15326.19
GSM875955	8245.78	8116.54	3636.95	3382.29	4841.37	4547.25	6420.76	6301.37	4541.80	5201.98	12104.25	11160.45	4503.52	4601.75	9390.85	9317.31	10073.56	9082.54	17229.45	14451.41
GSM876008	8064.57	8155.04	3739.22	3161.38	4930.35	4418.99	6091.91	5309.12	4891.42	4964.04	8483.12	6762.10	4837.52	3781.93	7962.51	8263.94	8304.09	7704.38	13370.22	11899.34
GSM876006	13072.24	7620.47	6160.03	1664.10	6821.51	4247.19	13634.70	5977.20	5810.47	4019.43	12065.16	11638.45	6002.90	4722.13	8325.46	7877.79	8899.59	7758.81	16479.91	14015.57
GSM875979	7723.34	8137.29	2999.78	3047.85	4325.09	4448.68	6240.83	6759.17	5103.29	5180.10	10389.76	10801.57	4405.83	4803.52	9083.64	9963.15	9379.77	9128.80	16782.19	15750.07
GSM875992	8275.59	8706.32	2548.76	3259.56	4014.67	3818.27	5869.99	6327.07	4471.36	4306.95	8785.85	11946.87	4962.05	4998.67	7883.79	8890.78	9123.02	9305.13	16138.06	14636.56
GSM2157173	8516.17	7922.34	2773.90	3070.54	4256.06	4288.03	6966.02	6221.15	6553.47	6319.96	9157.22	9492.13	4566.81	4952.81	8413.29	8302.66	9086.56	9307.16	15743.41	16110.13
GSM2157174	8281.11	7498.82	2577.09	2738.72	5212.38	4321.42	6476.33	6432.52	6506.76	5871.11	12129.17	9997.23	4775.13	4162.19	8102.29	7694.13	9465.35	8027.84	17023.89	13666.21
GSM2157175	8487.52	10967.74	3257.85	2308.97	4401.72	4406.07	7252.60	7711.57	4626.37	6697.50	9244.95	9951.46	5022.10	4610.88	8813.66	9711.49	8704.65	8905.59	15712.25	14125.90
GSM2157176	8223.18	8540.23	3197.00	3023.01	3654.61	4583.65	7325.38	6591.76	4362.47	4856.30	8233.08	10659.28	5002.81	5272.41	8160.38	9690.20	8769.97	8551.72	15356.06	15055.01
GSM2157177	8884.69	8892.60	2507.68	3601.08	3876.30	4944.24	6040.71	5994.38	4702.91	5271.00	10675.24	12619.35	5976.50	6816.07	8241.53	10265.68	9804.74	9246.40	16950.17	15507.74
GSM2157178	10991.48	8018.98	5249.73	2800.52	5905.21	4445.38	11055.48	5989.94	5416.73	6724.73	13802.44	9272.11	6755.67	4156.01	8527.72	8497.02	10338.15	7905.14	16726.85	13555.75
GSM2157179	8500.72	9610.88	2778.66	3617.24	4630.72	4965.35	8490.84	6919.02	4968.59	5341.41	10698.25	11924.25	5148.86	6699.79	9097.93	11243.99	9953.01	9355.29	16859.53	15379.03
GSM2157180	7354.14	7530.65	2757.74	3268.31	3494.55	4346.73	6193.02	6502.41	5090.98	4952.86	10787.48	12204.62	4198.66	4402.44	8300.30	8478.66	9162.11	8164.59	15572.12	13107.06
GSM2157181	8197.95	12564.21	3753.20	4584.18	4017.06	7951.54	6969.19	8410.65	4790.76	5000.45	8026.88	16381.18	4138.76	6823.90	7780.35	9699.68	7949.48	11448.54	12571.91	17916.91
GSM2157182	8690.36	8119.06	3055.33	4183.91	4369.93	3969.18	6399.34	6304.24	4738.28	4839.08	10220.74	11512.75	5201.41	5061.80	8331.84	8557.64	9083.77	9050.01	15397.24	15265.02
GSM2157183	9226.14	7643.32	3110.48	3249.29	5501.47	4338.43	7718.10	5555.22	5038.44	4727.62	12878.90	9608.54	5386.96	4268.01	8630.73	8163.59	9910.63	7959.13	16618.08	13365.89
GSM2157184	8075.91	8116.72	2751.08	2714.00	4369.51	3610.87	6942.60	6635.36	4915.46	4812.66	7451.59	14061.41	3692.61	4814.31	8469.20	8215.77	7916.90	9783.16	12669.70	15900.01
GSM2157185	8258.06	9437.04	3365.29	2467.75	4389.32	4861.81	6424.50	6570.24	5722.80	5992.31	12742.84	12754.64	4944.56	5274.91	9781.99	10673.85	9425.77	9246.81	16169.27	15121.73
GSM2157186	8564.86	8314.38	3478.43	2752.03	5908.45	4624.65	6016.48	6116.94	4109.01	4581.03	13431.93	13797.67	6774.20	6684.28	8363.69	7999.98	9205.22	9066.95	17043.81	15093.77
GSM2157187	9103.11	8342.29	3065.79	3082.90	4656.51	3994.08	7268.40	5980.68	5033.10	4500.43	13452.63	12057.29	5855.82	4787.83	8950.02	9450.47	10926.09	8144.19	17588.11	14432.97
GSM2157188	9757.40	8541.03	2688.33	3797.89	5682.56	5283.51	7605.03	6847.68	5378.31	4580.92	15952.05	12791.41	6374.61	4331.88	9045.12	8638.39	10251.78	8029.59	18112.33	13732.39
GSM2157189	8437.55	8977.80	2909.86	3499.34	4411.56	5470.58	7019.65	8308.94	4446.68	6293.94	9232.69	12301.71	4703.26	5361.04	8135.88	9795.80	8743.37	9551.55	15331.90	16313.41
GSM2157190	9590.90	9174.88	2112.16	2529.55	3901.10	4001.06	7211.63	7261.42	6308.11	5815.04	8599.62	9241.47	5463.15	4577.56	8755.19	8389.76	9814.61	9535.48	15954.32	16371.47
GSM2157191	8423.35	8550.50	5687.09	2694.99	6122.66	4652.70	5934.31	7366.82	4802.78	5394.63	12617.23	9167.44	5623.90	3944.26	8543.76	8282.37	9528.44	8380.10	17654.70	15687.66
t Stat	0.038496296 1.053208564		-0.173136213 1.7820		60816	-1.203729721		-2.062928077		0.192223258		-3.761478906		1.662788425		1.718682233				
P (one-tailed test)	0.484	475857 0.14983653		83653	0.431785498		0.041836379		0.118504062		0.023411769		0.424355422		0.000318927		0.052773957		0.047383376	
t (one-tailed test)	1.6909	90924255 1.690924255		24255	1.690924255 1.690924		24255	1.690924255		1.690924255		1.690924255		1.690924255		1.690924255		1.690924255		
P (two-tailed test)	0.96951714 0.299673061		573061	0.8635	0.863570996 0.083672759		72759	0.237008124 0.046823		23539	9 0.848710844		0.000637855		0.105547913		0.094766752			
t (two-tailed test)	2.0322	2.032244509 2.032244509		244509	2.032244509 2.032244509		44509	2.032244509 2.032244509			44509	2.032244509		2.032244509		2.032244509		2.032244509		

CRC, colorectal cancer; MCP, Microenvironment Cell Populations; GSE, gene set enrichment; CAFs, cancer-associated fibroblasts; NK, natural killer.



Figure S2 Functional enrichment analysis of genes in candidate modules. Top 10 enrichment results of pathway analysis by KEGG analysis for co-expressed genes in yellow (A) and brown (C) modules. Top eight enrichment results of genes by KEGG analysis in the PPI network of yellow (B) and brown (D) modules. The shade of colour of the inner ring represents the P value. KEGG, Kyoto Encyclopaedia of Gene and Genome; PPI, protein-protein interaction; PI3K, phosphoinositide 3-kinase; ECM, extracellular matrix.



Figure S3 The best-fitting regression model for the analysis of CRCLM. The optimal subset regression of hub genes in the yellow module (A), brown module (B), and further filtering the two (C). The left of the forest plot is the adjusted R^2 value corresponding to the combination of different gene subsets, and the colour of the grid corresponds to the value. (D) The result of the backward stepwise regression. *, P<0.05; ***, P<0.001. CRCLM, colorectal cancer with liver metastasis.