Proteomics and liquid biopsy characterization of human EMT-related metastasis in colorectal cancer

Linhai Yan (yanlinhai000@163.com)
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital
https://orcid.org/0000-0001-5848-7913

Hai-Ming Ru
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Si-Si Mo
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Chun-Yin Wei
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Dai-Mou Li
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Di Zhang
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Hao Yuan
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Xian-Wei Mo
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Wei-Zhong Tang
Guangxi Cancer Hospital and Guangxi Medical University Affiliated Cancer Hospital

Research Article

Keywords: colorectal cancer metastasis, Proteomics; liquid biopsy, Epithelial-Mesenchymal Transition, COL6A1

Posted Date: April 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-464010/v1

Abstract

Introduction: Circulating tumor cells (CTCs) undergo epithelial-mesenchymal transition (EMT), and the heterogeneity of EMT possibly affects colorectal cancer metastasis (mCRC) evolution. The aim of this study was to identify novel metastasis-related proteins and EMT-related pathways.

Methods: The EMT status of the CTCs derived from CRC patients with liver metastasis was determined on the basis of surface markers. Comparative proteomic analysis was then performed on matched pairs of primary tumors, adjacent para-tumor and liver metastases tissues. An optimized proteomic workflow including data independent acquisition (DIA) and parallel reaction monitoring (PRM) was used to screen for novel EMT-related protein clusters.

Results: The proportion of the unstable epithelial/mesenchymal (E/M)-type CTCs correlated significantly with distant metastases. We screened 105 proteins related to EMT from 4,752 proteins identified in all samples, of which 40 proteins were differentially expressed across the different tissues. We identified a novel EMT-related protein cluster (e.g., GNG2, COL6A1, COL6A2, DCN, COL6A3, LAMB2, TNXB, CAVIN1) and found that the expression levels of the core EMT-related proteins KRAS and ERBB2 were altered during metastasis progression. The proteomics data indicated that KRAS, ERBB2, COL6A1 and CAVIN1 are promising EMT-related metastatic biomarkers.

Conclusions: The plasticity of EMT phenotypes in the CTCs are key to CRC metastasis, prognosis and treatment outcome. Therapies targeting this aggressive CTC subset and the related proteins may suppress metastatic evolution.

Introduction

Colorectal cancer (CRC) is the third most prevalent cancer worldwide, and the second leading cause of cancer-related mortality[1]. The major cause of death in CRC patients is metastasis, and the primary and most common site of metastatic CRC (mCRC) is the liver[2–4]. Early detection can improve clinical outcomes for CRC patients with liver metastases.

Circulating tumor cells (CTCs) are the direct cause of cancer metastasis, and specific types of CTCs give rise to metastatic foci under different conditions. In addition, the CTC load correlates with poor progression free survival in mCRC[5–7]. Epithelial-Mesenchymal Transition (EMT), a reversible biological program wherein epithelial cells gradually transform into the highly invasive stromal cells, is a key event in tumor metastasis[8]. CTCs enhance migration and aggressiveness through EMT, and interstitial CTCs are a biomarker of cancer progression[9–11]. Furthermore, the predominantly epithelial CTC subtypes (E and E/M) have stronger metastatic and proliferative abilities[12]. However, the prognostic impact of different CTC phenotypes is still ambiguous.

Proteomics-driven precision medicine (PDPM) relies on the detection of very low levels of protein biomarkers in the early stages of cancer through highly sensitive proteomics, i.e. Qualitative and

quantitative analysis of all proteins in a biological unit. The current challenges in proteomics technology are the analytical speed, proteome coverage depth and quality of data analysis. Data Independent Acquisition (DIA) system can simultaneously analyze the proteomes of multiple samples, and therefore obviate the above limitations and scan more data without losing low abundance proteins[13]. In addition, parallel reaction monitoring (PRM) mass spectrometry is a high throughput technology that can simultaneously verify dozens of proteins. The samples can be directly detected by mass spectrometry without the need for specific antibodies, which can improve the accuracy and success rate of verification[14]. Combining non-labeled target protein screening by DIA and further verification by PRM is a viable strategy for identifying novel disease-related biomarkers[15, 16].

In the present study, we found that the transient epithelial/mesenchymal (E/M)-type CTCs from mCRC patients have the strongest metastatic abilities. Subsequently, we used the comparative proteomic approach to identify novel biomarkers and EMT-related pathways in the matched pairs of primary tumor tissues, adjacent mucosal tissues and liver metastatic tissues from symptomatic early T staging ($T_2N_xM_1$) mCRC patients. The proteomics data was analyzed by DIA and PRM. Our findings indicate that different CTC subpopulations stratified on the basis of the EMT phenotype and related proteomics should be considered as targets for multimodal therapy.

Materials And Methods

Samples and patients collection

Circulating tumor cells(CTCs)from 100 patients with CRC were analyzed. 12 mCRC cases for which matched pairs of primary tumor tissues, adjacent mucosa tissues and liver metastases tissues were evaluated in the study (online supplementary methods).

Circulating tumor cells isolation and identification

70 histologically confirmed mCRC and 30 CRC cases (total = 100) that had matched pairs of tissue and peripheral blood samples for CTC analysis using a Canpatrol ® system were employed. For further details, refer to the online supplementary methods section.

Data independent acquisition (DIA) and Parallel Reaction Monitoring (PRM)

All proteome analyses were performed by a Q-Exactive HF mass spectrometer (Thermo, USA) equipped with a Nanospray Flex source (Thermo, USA). Samples were loaded and separated by a C18 column (50 cm × 75 μm) on an EASY-nLCTM 1200 system (Thermo, USA). (Online supplementary methods).

Results And Discussion

Clinicopathological characteristics of mCRC

Most CRC patient exhibit metastases at the initial diagnosis due to the occult nature of the process. To uncover underlying mechanisms involved in metastasis at early stage and develop a clinically practicable

therapy, we comprehensively investigated the CRC with metastasis across multiple dimensions (Fig. 1A). The CTC data of 70 mCRC patients, 30 CRC patients and 10 healthy controls, and the proteomic data of 12 CRC patients with liver metastasis ($T_2N_xM_1$) were collected. After screening for the CTC subtype with strongest correlation to distant metastasis, we selected 12 mCRC patients at early T stage (T_2) with distant metastasis, and analyzed the proteomics data of their primary tumor, normal intestinal epithelium and metastatic liver tissues. Finally, the CancerSEA and TIMER database were used to analyze the single cell function and immune invasion degree of these selected possible influencing factors and predicted proteins The primary and metastatic lesions of all patients who underwent CTC detection were assessed by HE staining, lymphocyte immunophenotyping and TUNEL staining, and the expression levels of MLH2, MSH6, PMS2 and MLH1 were also analyzed (Fig. 1B).

Unstable tumor and CTCs EMT state correlate with the clinical outcome of mCRC patients

Given that CTCs are the source of distant metastasis, we next assessed the correlation between the EMT-related molecular alterations in primary tumor cells and CTCs. To this end, we analyzed the EMT and microsatellite instability (MSI) status-related genes in a cohort of 70 mCRC patients with liver metastases and 30 CRC patients. The expression of CK7, CK20, vimentin, β-catenin, Ki67, S100, P53, MLH1, PMS2, MSH2 and MSH6 in the tumor tissues were evaluated by IHC. The Canpatrol® system was used to determine the expression levels of CK8, CK18, CK19, vimentin and Twist in the CTCs. As is evident from the expression of EMT proteins in CTCs (Fig. 1C) and the clinical characteristics of the patients (Fig. 1D), the CTC load was higher in patients with definite metastasis versus the non-metastatic patients, which is also consistent with previous studies[10]. However, the number of CTCs was not correlated to the extent of metastasis, lymphocyte infiltration, vascular invasion and nerve invasion in the primary tumor, and therefore cannot totally reflect the metastasis status. As shown in Fig. 1E, CK8 and CK18 showed the highest expression levels in both the CRC tissues and CTCs, with considerable heterogeneity. In addition, 60 (57.14%) tissue samples and 62 (82.85%) CTC samples expressed Twist. Forty-five patients expressed Twist in tumor tissues and CTCs, whereas both were negative in 9 patients. Vimentin was expressed in 50 (62.85%) tissues and 57 (77.14%) CTCs samples, and both were respectively positive and negative in 39 and 11 patients. Furthermore, 66 (94.28%) tissues were microsatellite stable (MSS), and 2 patients (2.85%) had respectively low-frequency (MSI-L) and high-frequency (MSI-H) microsatellite instability. In summary, the expression levels of EMT-related were overall consistent between tumor tissues and CTCs, and CTCs profiling may provide valuable insights into the EMT status of CRC tumors. In addition, since the MSI status does not affect EMT and the MSI-H patients account for only a small fraction of all mCRC patients, the MSI state need not be taken into account when analyzing EMT. Based on whether the core EMT molecules were expressed in both or neither tumor tissues and matched CTCs, the patients were stratified as stable and unstable E/M respectively. As shown in Fig. 1F, 20/70 mCRC patients (28.57%) had stable E/M and 50 patients (71.43%) had unstable E/M. In addition, 32/50 patients with unstable E/M (64%) showed inconsistent expression of one biomarker, 17 patients (34%) expressed two biomarkers differentially between tissues and CTCs, and one patient (2%) showed inconsistent expression of three biomarkers. Furthermore, the PFS of mCRC patients with unstable E/M was

significantly shorter than that of patients with stable E/M (12.42 months vs 14.07 months, p = 0.0427; Fig. 1G). Thus, unstable E/M status is prognostically relevant in mCRC.

Hybrid epithelial/mesenchymal phenotype (E/M) CTCs contributes to liver metastasis of mCRC

The immunophenotype of tumor cells change constantly during EMT, on the basis of which they are classified as epithelial (E), predominantly epithelial/mesenchymal mixed (E/m), predominantly mesenchymal/epithelial mixed phenotype (M/e), and mesenchymal (M)[12]. Based on our findings, we propose that cells undergoing EMT can be stratified into the stable $EpCAM^+CK8^+CK18^+CK19^+$ epithelial (E), stable $vimentin^+twist^+$ mesenchymal (M) and the transitioning unstable $EpCAM^+CK8^+CK18^+CK19^+vimentin^+twist^+$ epithelial/mesenchymal (E/M) phenotypes (Fig. 2A). As shown in Fig. 2B, E/M was detected in 63/70 (90%) mCRC patients and was the major CTC phenotype in this cohort, but was rarely seen among the CRC patients. In addition, 42 mCRC patients showed high frequency of the E/M phenotype (more than 3 biomarkers). Furthermore, the number of E/M CTCs correlated significantly with the size of liver metastases ($r = 0.4051$; $p = 0.0005$), whereas the E and M phenotypes did not show any correlation (Fig. 2C). In addition, the total CTC load was also significantly correlated with the size of metastatic nodules in patients with predominantly E/M phenotype ($r = 0.3624$, $p = 0.002$) compared to all phenotypes combined ($r = 0.3624$, $p = 0.002$; Fig. 2C). Furthermore, the number of E/M CTCs and carcinoembryonic antigen (CEA) levels were significantly correlated ($r = 0.5376$, $p < 0.0005$) compared to all phenotypes combined ($r = 0.4355$, $p = 0.0002$; Fig. 2D). In conclusion, the EMT status of the CTCs is a reliable indicator of metastasis in CRC, and the number of unstable E/M types is a measure of the extent of metastasis.

Mass spectrometric analysis about proteomics characterization of mCRC

The proteomes of paired primary tumor (Ca), normal para-tumor intestinal epithelium (P) and liver metastatic tissues (Liver) from 12 $T_2N_xM_1$ mCRC patients were analyzed by DIA and PRM methods. The patients were divided into groups of 3 (n = 4 each) (Supplementary Fig. 1A). All patients presented with superficial muscular, perineural and vascular invasion, which are indicative of the early stages of metastasis (Fig. 3A). Since T2 stage can progress to distant metastasis, the factors identified in these patients may be more accurate indicators of metastasis compared to those identified in T3 and T4 mCRC patients.

LC-MS/MS identified 44,815 peptides encompassing 4,752 protein groups with a false discovery rate (FDR) of 1%, and the P value of most proteins was less than 0.01 (Supplementary Fig. 1B). Principal component analyses and hierarchical clustering analyses revealed distinct protein expression patterns between the paired Ca, P and Liver samples (Fig. 3B and 3C). Interestingly, the top five highly-expressed proteins in the same tissue were similar among patients. Furthermore, using fold change ≥ 1.2 and p-value < 0.05 as the thresholds, we identified 780 differentially expressed proteins (DEPs) in P vs Liver, of which 459 were upregulated and 321 were downregulated in the former. Likewise, there were 771 DEPs in Ca samples relative to P, of which 435 and 336 proteins were respectively up- and downregulated. Finally, 337 DEPs were identified in Ca vs Liver samples, including 203 up-regulated proteins and 134 down-regulated proteins (Supplementary Fig. 1C and 1D). Overall, 1262 proteins were differentially expressed in

the Ca and Liver samples relative to P (Fig. 3D), and may be associated with metastasis. We classified these DEPs into the following four groups: (1) down-regulated in Ca vs P and up-regulated in P vs Liver, (2) up-regulated in both, (3) up-regulated in Ca vs P and down-regulated in P vs Liver, and (4) down-regulated in both. KEGG pathway analysis revealed that the top 10 enriched pathways in each group were mostly related to metabolism (Fig. 3E). In addition, the glycolysis/gluconeogenesis pathway was common to all groups, whereas the spliceosome pathway was most enriched among the DEPs between Ca and P samples (Fig. 3E). Similar trends were observed with the top 50 enriched pathways as well (Supplementary Fig. 1E). Gene ontology (GO) enrichment analysis of the DEPs further revealed that the most significant biological processes were creatine biosynthesis and fructose biosynthesis, cellular components were collagen type IV trimer and exosomes, and molecular functions were 11-beta-hydroxysteroid dehydrogenase [NAD (P)] activity and signal recognition particle binding involved in CRC with liver metastasis (Fig. 3F). As shown in Fig. 3G, proliferation associated protein PRG3 and anti-apoptosis related protein SERPINB9 in Ca vs Liver group, and cyclin-dependent kinases CDKS up-regulated in both primary tumor and liver metastasis tissues. The top 50 DEPs showed significant correlation in each group (Fig. 3H). Taken together, the identified DEPs are mostly involved in genetic information processing, metabolism, organismal systems, environmental information processing and human disease-related signaling pathways.

EMT-related metastasis core kinesin and structural helper proteins

Given the complex nature of metastasis, we performed network analysis of the proteomics data. Based on the above DIA analysis, 25 DEPs were identified in all comparison groups. ADI1, which is involved in cysteine and methionine metabolism, showed the most significant difference in expression levels between P and Liver groups. The largest fold change was exhibited by TPSB2, which regulates the influenza signaling pathway (Fig. 4A). We further screened 105 EMT-related proteins from a total of 4752, of which 40 were differentially expressed in each group (Fig. 4B). As shown in Fig. 4C and Supplementary Fig. 1F, the 40 DEPs common to all pairs were mainly enriched in PI3K-Akt signaling pathway, Wnt signaling pathway, Notch signaling pathway and TGF-beta signaling pathway, which are closely related to EMT (Fig. 4C). The differentially expressed EMT-related proteins were classified into four groups (Fig. 4D): (1) up-regulated in Ca and Liver vs P, (2) up-regulated in Ca vs Liver, (3) down-regulated in Ca and Liver vs P, and (4) down-regulated in Ca vs Liver. Integration of the proteomics data with the 4 major signaling pathways (Fig. 4E) showed that the Notch signaling pathway was activated and TGF-beta signaling pathway was suppressed in the primary and metastatic tumor tissues. PI3K-Akt signaling pathway is the enrichment of the most proteins. EMT-related proteins regulate DNA repair, angiogenesis, cell proliferation, apoptosis, glycolysis/gluconeogenesis, protein synthesis, cell cycle and proteolysis (Fig. 4E). Based on these findings, we hypothesize that the network of structural helper proteins synergize with the core kinesin during metastasis, and are potential early indicators of metastasis, as well as therapeutic targets.

Clusters of EMT inferred by an integrated proteomic analysis

Protein-protein interaction (PPI) network was constructed with the EMT-related DEPs, which revealed four functional clusters. Cluster (iii) was mainly involved in ECM-receptor interaction, focal adhesion, PI3K-Akt

signaling pathway and human papillomavirus infection signaling pathways (Fig. 5A and Supplementary Fig. 6A). Cluster (i) was the biggest and showed the strongest correlation with EMT, including proteins involved in Notch signaling pathway, HIF-1 signaling pathway, Wnt signaling pathway and protein processing in endoplasmic reticulum signaling pathway (Fig. 5B and Supplementary Fig. 6B). Furthermore, cluster (ii) consisted of proteins involved in Wnt signaling pathway, ribosome biogenesis in eukaryotes and adherens junction signaling pathway (Fig. 5C and Supplementary Fig. 6C), and cluster (iv) in glycolysis/gluconeogenesis, citrate cycle (TCA cycle), pyruvate metabolism and PPAR signaling pathway (Fig. 5D and Supplementary Fig. 6D). Interestingly, KRAS and ERBB2 were the core proteins in only clusters (iii) and (i). The main DEPs in cluster (iii) were GNG2, COL6A1, COL6A2, DCN, COL6A3, LAMB2, TNXB and CAVIN1 (Fig. 5E and 5F), and may be most significantly correlated with EMT and CRC metastasis.

Validation of EMT-related protein cluster by PRM

The differential expression levels of KRAS, ERBB2, GNG2, COL6A1, COL6A2, DCN, COL6A3, LAMB2, TNXB and CAVIN1 were analyzed by LC-MS/MS in the PRM mode (Supplementary Fig. 7A). All the quantified peptides for EMT-related proteins exhibited an excellent linear fit between the observed retention time and the iRT in the library (Supplementary Fig. 2A). In addition, the retention time and quality of inner-label iRT were stable, and the error of quality was small (Supplementary Fig. 2B and 2C). Thus, our method is highly reliable for peptide identification.

The DEPs in cluster (i), including GNG2, COL6A1, COL6A2, DCN, COL6A3, LAMB2, TNXB and CAVIN1, were down-regulated in the Ca and Liver samples compared to P tissues, which was also consistent with the DIA results (Fig. 6A). In addition, the expression levels of COL6A1, COL6A2, COL6A3 and DCN were higher than that of CAVIN1, GNG2, LAMB2 and TNXB (Fig. 6B), and all but CAVIN1 were up-regulated in the Liver vs Ca tissues. The peaks of the KRAS and ERBB2 peptides in PRM demonstrated the superior specificity and stability of our results (Fig. 6C and 6D), and similar observations were made with CAVIN1, COL6A1, COL6A2, COL6A3, DCN, GNG2, LAMB2 and TNXB (Supplementary Fig. 2D). KRAS and ERBB2 remained the core EMT-related proteins among the 40 DEPs (Fig. 7E and Supplementary Fig. 7). While KRAS was consistently up-regulated during CRC progression, ERBB2 was significantly over-expressed in the Ca but down-regulated in the Liver tissues (Fig. 6F and 6G).

Functional analysis of the core protein and EMT-related protein cluster

The infiltrating immune cells in the tumors with KRAS mutation were next assessed using the TIMER database. Surprisingly, KRAS mutation was associated with fewer macrophages ($P < 0.01$) and neutrophils ($P < 0.01$) in COAD, and less macrophages ($P < 0.01$) in READ (Fig. 7A). The distribution of the target proteins were further visualized using t-SNE analysis (Fig. 7B), which showed that CAVIN1, COL6A1 and COL6A3 were concentrated in the same cell cluster, indicating their involvement in similar malignant processes (Fig. 7B). After annotating CAVIN1, COL6A1 and COL6A3 as EMT-binding clusters, we analyzed the infiltration level of B cells, CD 8 + T cells, CD4 + T cells, macrophages, neutrophils and dendritic cells (DCs) in the CRC tumors (Fig. 7C), and detected significant correlation between the protein expression levels and the number of infiltrating CD4 + T cells, macrophages, neutrophils and DCs (Fig. 7C and

Supplementary Fig. 3A), GNG2, COL6A2, DCN, LAMB2, TNXB, CAVIN1 and COL6A1 have same tendency in infiltration level of B cells, CD 8 + T cells, CD4 + T cells, macrophages, neutrophils and dendritic cells (DCs) in the CRC tumors (Supplementary Fig. 3B). Our results suggest that infiltration of the above immune cells likely promote EMT in CRC. According to the CancerSEA database, the function of CAVIN1 (cor = 0.59, $p \leq 0.001$) and COL6A1 (cor = 0.58, $p \leq 0.01$) at the single cell level were significantly related to EMT (Fig. 7D and Supplementary Fig. 4). Since CAVIN1 and COL6A1 were also part of the EMT-binding clusters, they can be considered EMT-driving partner proteins (Fig. 7E). In conclusion, the protein network regulating metastasis may also affect the tumor immune micro-environment.

Discussion

Our findings suggest that EMT in CTCs is a transient process, and most CTCs in metastatic CRC are highly heterogeneous and co-express both epithelial and mesenchymal biomarkers. Therefore, instead of the conventional classification system, we broadly stratified the CTCs as stable or unstable based on their EMT phenotype to evaluate clinical prognosis. The plasticity of the epithelial and mesenchymal CTC phenotypes contributed to liver metastasis of CRC, and are therefore promising indicators of early metastatic events[17]. Previous studies show that CTCs undergoing EMT have superior migration, self-seeding and chemoresistance abilities[18–20]. Consistent with this, we found that the unstable hybrid E/M CTCs have enhanced ability to metastasize to the liver, which is consistent with the high degree of epithelial–mesenchymal plasticity observed in this phenotype[21].

The proteins secreted by tumor cells are quickly diluted in the bloodstream by nearly thousand-fold. Tumor markers (CEA, AFP, CA125, CA199 and CA153) do not accurately reflect the disease (Supplementary Fig. 5). Therefore, we used DIA and PRM MS-based workflow to detect biomarkers of early micro-metastases. The DIA data sets were qualitatively and quantitatively mined using the highly specific fragment ion maps in a spectral library[22], which ensures accuracy and high efficiency[23]. During liquid phase separation, all fragment ion maps of each target parent ion are recorded by PRM[24]. Its advantage lies in the use of ultra-high resolution orbit rap quality analyzer that can separate noise from the real signal[24].

Studies show the existence of rare populations with EMT potential within the tumor that are the source of metastasis[25], and the EMT-related metastatic potential may be present even during cancer initiation[26]. Since our data confirmed significant proteomic changes during CRC metastasis, we screened for the differentially expressed EMT-related proteins and constructed EMT-related protein clusters by analyzing the enriched pathways. The largest cluster consisted of core proteins including GNG2, COL6A1, COL6A2, DCN, COL6A3, LAMB2, TNXB and CAVIN1, which were down-regulated in the primary tumor and liver metastatic tissues. This cluster is primarily involved in regulating ECM-receptor interaction, focal adhesion, PI3K-Akt signaling pathway and human papillomavirus infection signaling pathway, and also correlated strongly with higher infiltration of CD4 + T cells, macrophages, neutrophils and dendritic cells. Lower immune cell infiltration and differential activation of specific tumor-intrinsic pathways contribute to immune escape[27], which is consistent with our findings as well. Since EMT-related clusters included

both EMT-binding proteins (CAVIN1, COL6A1 and COL6A3) and EMT-driving partner proteins (CAVIN1 and COL6A1), a heterogeneous population of cells with EMT potential may reside within the tumor that drive metastasis. In addition, EMT has also been associated with tumor immune escape via activation of key immune checkpoints[28]. In this regard, the EMT-driving partner proteins (CAVIN1 and COL6A1) are promising biomarkers and therapeutic targets for early metastasis.

In conclusion, liquid biopsies and proteomics data of both tumor tissues and blood samples indicate that EMT biomarkers are promising prognostic factors of mCRC. The CTCs are a highly heterogenous population, and frequently co-exhibit epithelial and mesenchymal features. The clinical significance of CTCs depends on the “stable” or “unstable” phenotype. The EMT-related clusters including receptor proteins (CAVIN1, COL6A1 and COL6A3), interacting proteins (CAVIN1 and COL6A1) and core proteins (KRAS and ERBB2) influence the distant liver metastatic cascade in CRC. Finally, KRAS, ERBB2, COL6A1 and CAVIN1 are early diagnostic biomarkers of liquid biopsy and therapeutic targets for CRC.

Abbreviations

CTCs: Circulating Tumor Cells; EMT: Epithelial-Mesenchymal Transition; mCRC: Metastasis Colorectal Cancer; DIA: Data Independent Acquisition; PRM: Parallel Reaction Monitoring; E/M-type: Epithelial/Mesenchymal Type; PRG3: Proliferation Associated Protein 3; SERPINB9: Serine Protease Inhibitor B9; CDKS: Cyclin-Dependent Kinases; GNG2: G Protein γ2 Subunit; COL6A1: Collagen type VI α1 Chain; LAMB2: Laminin Beta2; TNXB: Tenascin XB; CAVIN1: Caveolae-associated protein 1; KRAS: Kirsten Rat Sarcoma.

Declarations

Ethics statement

Sample collection and research were in accordance with regulations issued by the National Health Commission of China and the ethical standards formulated in the Helsinki Declaration.Written informed consent was obtained from all patients. The permission for retrospective study was obtained from the institutional review board of Guangxi Medical University Cancer Hospital.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

LH Y designed the study, led data analyses and wrote the manuscript. WZ T secured funding. HM R and SS M obtained clinical information, contributed to the design, experimental work, analysis and writing of the manuscript. DM L contributed to the analysis of the study and writing of the manuscript. D Z contributed to the experimental work. H Y obtained and documented clinical information. XW M recruited

patients, obtained blood samples and contributed to documentation of clinical information. All authors contributed to experimental design, critical discussion of the findings and to the final manuscript.

Funding

This work was supported by China Postdoctoral Science Foundation (No. 2019M653812XB). The National Science Foundation (81973533). Guangxi University High-level Innovation Team and the Project of Outstanding Scholars Program (2019AC03004) and Guangxi Science and Technology Project (AD19245197).

Acknowledgements

We are grateful to Changsee Guangxi Medical Technology Co., Ltd for the guidance of proteins analysis.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

References

1 Bray F , Ferlay J , Soerjomataram I , Siegel R L , Torre L A , Jemal A : Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.

2 Okholm C , Mollerup T K, Schultz N A, Strandby R B, Achiam M P: Synchronous and metachronous liver metastases in patients with colorectal cancer. Dan Med J 2018;65:A5524.

3 Tang J T, Chen H R, Wong C C, Liu D B, Li T, Wang X H, et al: DEAD-box helicase 27 promotes colorectal cancer growth and metastasis and predicts poor survival in CRC patients. Oncogene 2018;37:3006-3021.

4 Engstrand J , Nilsson H, Strömberg C, Jonas E, Freedman J: Colorectal cancer liver metastases – a population-based study on incidence, management and survival. BMC Cancer 2018;18:78.

5 Klein C A: Selection and adaptation during metastatic cancer progression. Nature 2013;501:365-72.

6 Au S H, Storey B D, Moore J C, Tang Q, Chen Y L, Javaid S, et al : Clusters of circulating tumor cells traverse capillary-sized vessels. Proc Natl Acad Sci U S A 2016;113:4947-52.

7 Hüsemann Y, Geigl J B, Schubert F, Musiani P, Meyer M, Burghart E, et al: Systemic spread is an early step in breast cancer. Cancer Cell 2008;13:58-68.

8 Bidard F C, Peeters D J, Fehm T, Nolé F, Gisbert-Criado R, Mavroudis D, et al: Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. Lancet Oncol 2014;15:406-14.

9 Yu M, Bardia A, Wittner B S, Stott S L, Smas M E, Ting D T, et al: Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science 2013;339:580-4.

10 Fischer K R, Durrans A, Lee S, Sheng J, Li F H, Wong S T C, et al: Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature 2015;527:472-6.

11 Zheng X F, Carstens J L, Kim J, Scheible M, Kaye J, Sugimoto H,et al: Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature 2015;527:525-530.

12 Li J M, Vranken J G Van, Vaites L P, Schweppe D K, Huttlin E L, Etienne C, et al: TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. Nat Methods 2020;17:399-404.

13 Latonen L, Afyounian E, Jylhä A,Nättinen J, Aapola U, Annala M,et al: Integrative proteomics in prostate cancer uncovers robustness against genomic and transcriptomic aberrations during disease progression. Nat Commun 2018;9:1176.

14 Piazza I, Kochanowski K, Cappelletti V, Fuhrer T, Noor E, Sauer U, et al: A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. Cell 2018;172:358-372.e23.

15 Musa Y R, Boller S, Puchalska M, Grosschedl R, Mittler G: Comprehensive Proteomic Investigation of Ebf1 Heterozygosity in Pro-B Lymphocytes Utilizing Data Independent Acquisition. J Proteome Res 2018;17:76-85.

16 Cortés-Hernández L E, Eslami-S Z, Alix-Panabières C: Circulating tumor cell as the functional aspect of liquid biopsy to understand the metastatic cascade in solid cancer. Mol Aspects Med 2020; 72:100816.

17 Liu X, Li J J, Cadilha B L, Markota A, Voigt C, Huang Z, et al: Epithelial-type systemic breast carcinoma cells with a restricted mesenchymal transition are a major source of metastasis. Sci Adv 2019; 5: eaav4275.

18 Papadaki M A, Stoupis G, Theodoropoulos P A, Mavroudis D, Georgoulias V, Agelaki S: Circulating Tumor Cells with Stemness and Epithelial-to-Mesenchymal Transition Features Are Chemoresistant and Predictive of Poor Outcome in Metastatic Breast Cancer. Mol Cancer Ther 2019; 18:437-447.

19 Wang W Y, Wan L, Wu S, Yang J, Zhou Y, Liu F, et al: Mesenchymal marker and LGR5 expression levels in circulating tumor cells correlate with colorectal cancer prognosis. Cell Oncol (Dordr)

2018;41:495-504..

20 Genna A, Vanwynsberghe A, Villard A V, Pottier C, Ancel J, Polette M , et al: EMT-Associated Heterogeneity in Circulating Tumor Cells: Sticky Friends on the Road to Metastasis. Cancers (Basel) 2020; 12:1632.

21 Rauniyar N, Peng Gang, Lam T T, Zhao H Y, Mor G, Williams K R : Data-Independent Acquisition and Parallel Reaction Monitoring Mass Spectrometry Identification of Serum Biomarkers for Ovarian Cancer. Biomark Insights 2017;12:1177271917710948.

22 Sanda M, Goldman R: Data Independent Analysis of IgG Glycoforms in Samples of Unfractionated Human Plasma. Anal Chem 2016;88:10118-10125.

23 Rauniyar N: Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. Int J Mol Sci 2015;16:28566-81.

24 Lawson D A, Kessenbrock K , Davis R T, Pervolarakis N, Werb Z: Tumour heterogeneity and metastasis at single-cell resolution. Nat Cell Biol 2018;20:1349-1360.

25 Gupta P B, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C: Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. Cell Stem Cell 2019;24:65-78.

26 Yang L, Li A T , Lei Q Y, Zhang Y: Tumor-intrinsic signaling pathways: key roles in the regulation of the immunosuppressive tumor microenvironment. J Hematol Oncol 2019;12:125.

27 Lou Y Y, Diao L X, Cuentas E R P, Denning W L, Chen L M, Fan Y H, et al: Epithelial-Mesenchymal Transition Is Associated with a Distinct Tumor Microenvironment Including Elevation of Inflammatory Signals and Multiple Immune Checkpoints in Lung Adenocarcinoma. Clin Cancer Res 2016;22:3630-42.

28 Mak M P, Tong P, Diao L X, Cardnell R L, Gibbons D L, William W N, et al: A Patient-Derived, Pan-Cancer EMT Signature Identifies Global Molecular Alterations and Immune Target Enrichment Following Epithelial-to-Mesenchymal Transition. Clin Cancer Res 2016;22:609-20.

Figures

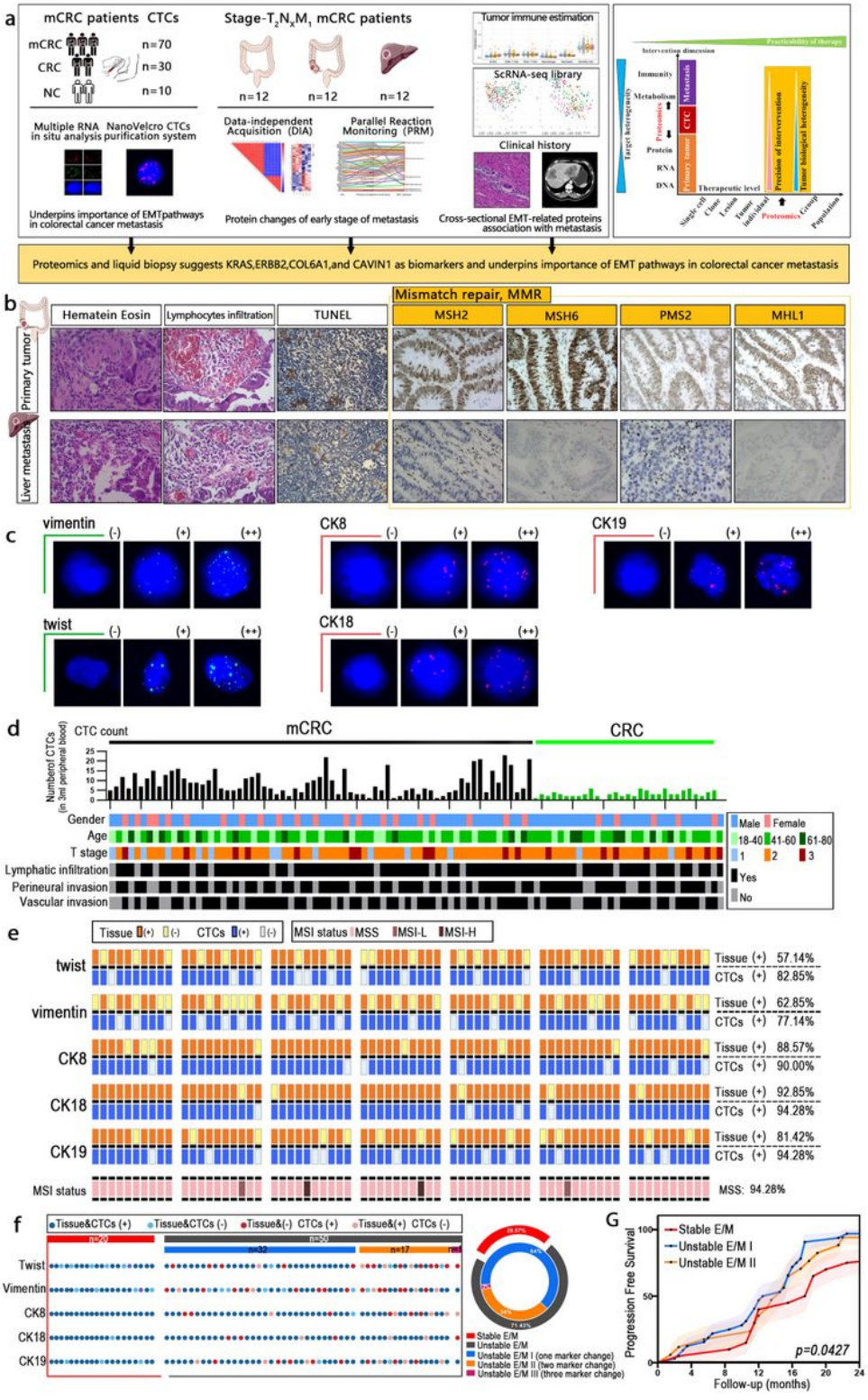

Figure 1

Unstable tumor and CTCs EMT state correlate with the clinical outcome of mCRC patients.(A)A scheme showing experiments and integrated analyses which were performed. (B) The morphological heterogeneity of typical samples under microscope. (C) CTCs were detected by Canpatrol® system (red: CK8, CK18, and CK19, green: vimentin and twist). (D) CTCs profile and associated clinicopathologic features of all the 70 mCRC patients and 30 CRC patients. (E) The landscape of EMT molecular

alterations detected in CTCs and the alterations detected in paired tissue samples is presented. (F) Shown are proportions of stability molecular expression CTCs in individual patients; red: stable, gray: instability. (G) Progress Free Survival (PFS) of mCRC patients was stratified according to the presence of stability molecular expression.

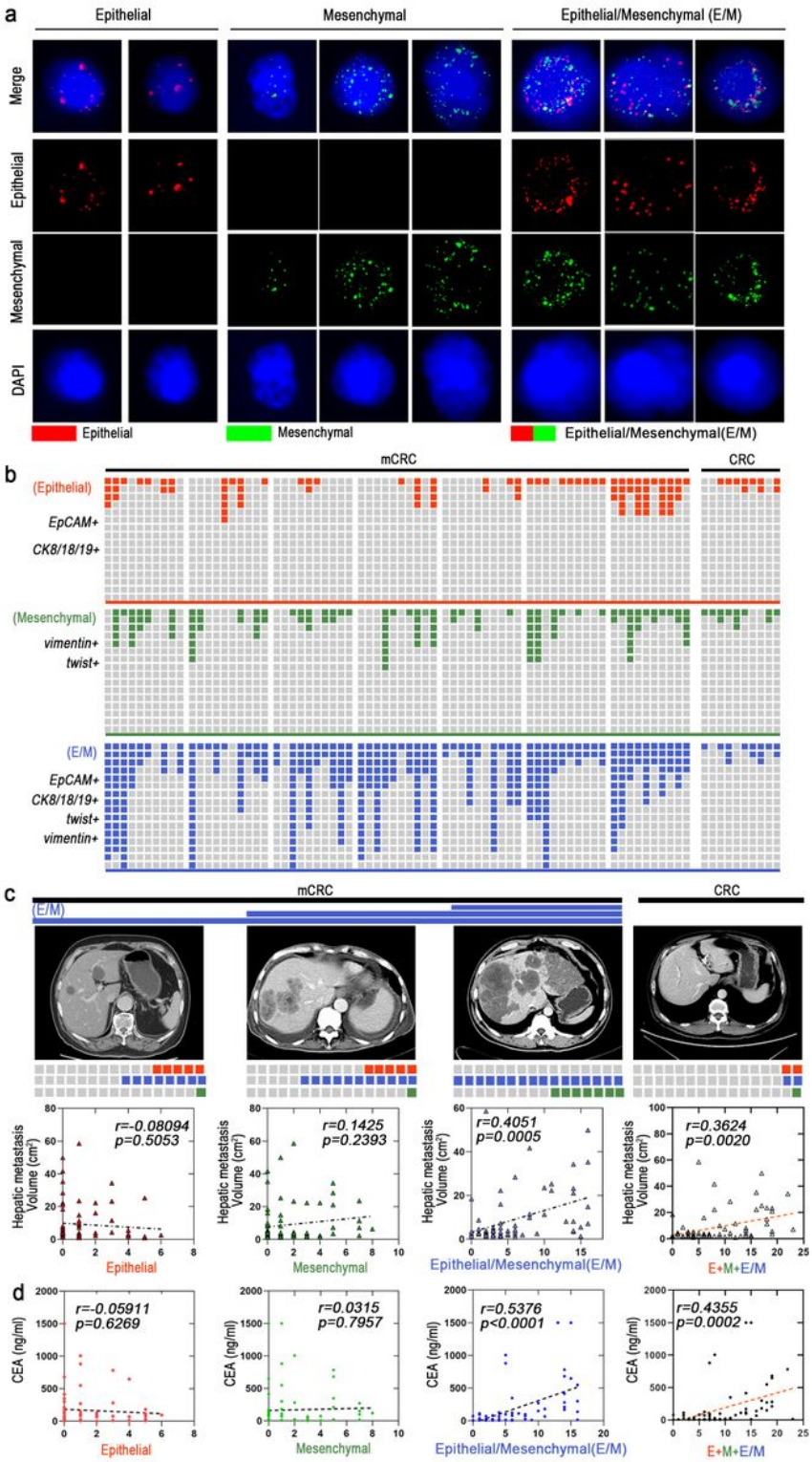

Figure 2

Hybrid epithelial/mesenchymal phenotype (E/M) CTCs contributes to liver metastasis of mCRC. (A) CTCs were detected by Canpatrol® system (red: EpCAM, CK8, CK18, and CK19, green: vimentin and twist). DAPI (4',6-diamidino-2-phenylindole). (B) Shown are proportions of E, M, and E/M phenotype CTCs in individual mCRC patients (n=70) and CRC patients (n=30). (C) E, M, and E/M phenotype CTCs to predict the liver metastatic tumor volume are depicted as a linear regression model. P value are indicated. (D) E, M, and E/M phenotype CTCs to predict the carcino-embryonic antigen (CEA) are depicted as a linear regression model. P value are indicated.

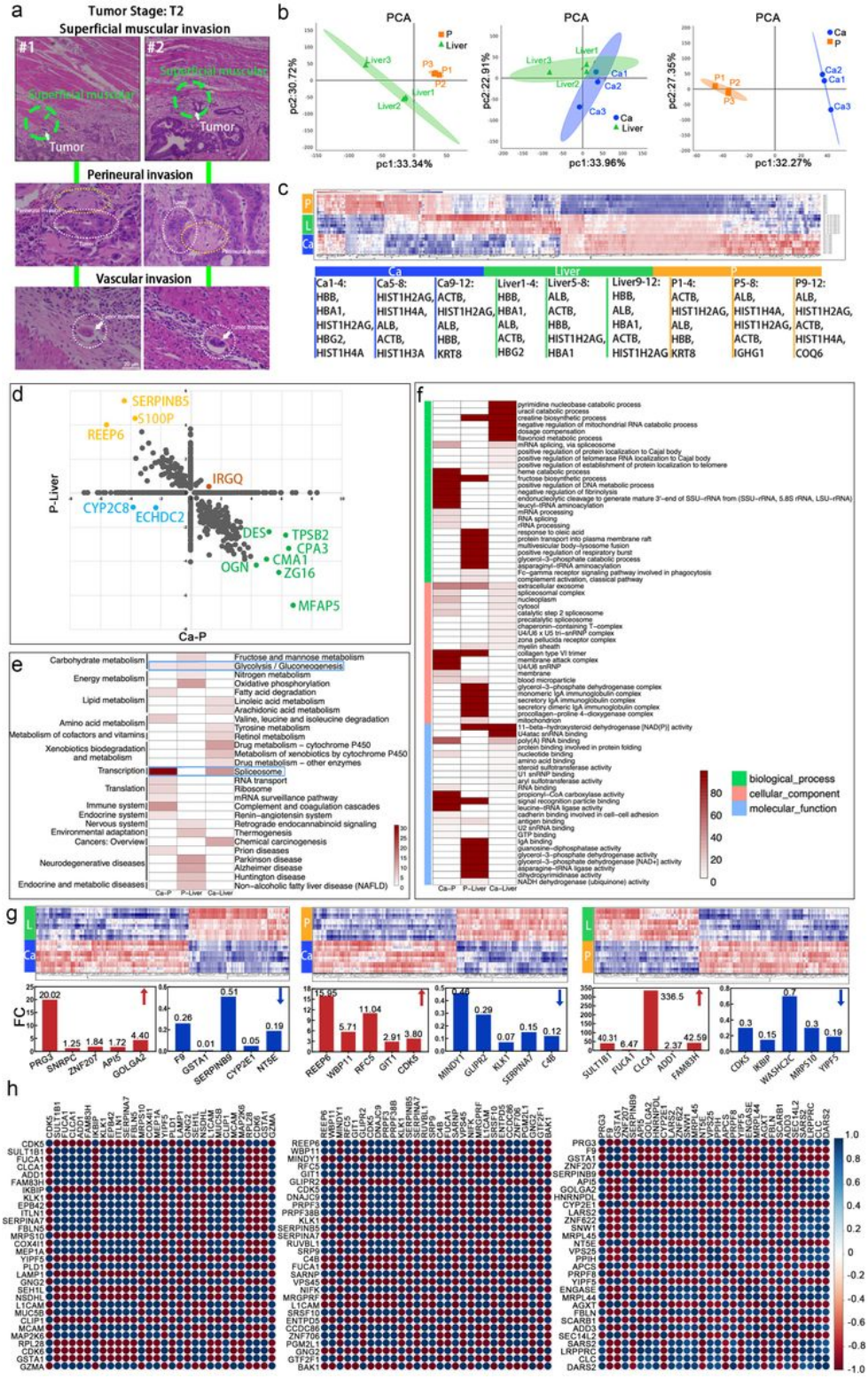

Figure 3

Mass spectrometric analysis about proteomics characterization of mCRC. (A) mCRC patients in T2 stage with perineural and vascular invasion. (B) Principal-component analysis of protein expression in each group. (C) Heat map of all protein expressions identified and quantified by mass spectrometry in the proteomic analysis. (D) Merge of Ca-P and P-Liver proteomic data divide proteins in four groups. (E) The heatmap shows the significance of top 10 cellular processes being enriched by proteins in Ca-P, P-Liver and Ca-Liver. (F) The heatmap shows the significance of top 30 GO term being enriched by proteins in Ca-P, P-Liver and Ca-Liver. (G) Proteomic analysis reveals distinct protein expression patterns in paired tumor (Ca), para-tumor (P) and liver metastasis tissues (Liver). Heat maps of the differentially expressed proteins show clearly distinctive patterns of protein expression between disease groups. (H) Correlation analysis of top 50 differentially expressed proteins in Ca-Liver, Ca-P and P-Liver.

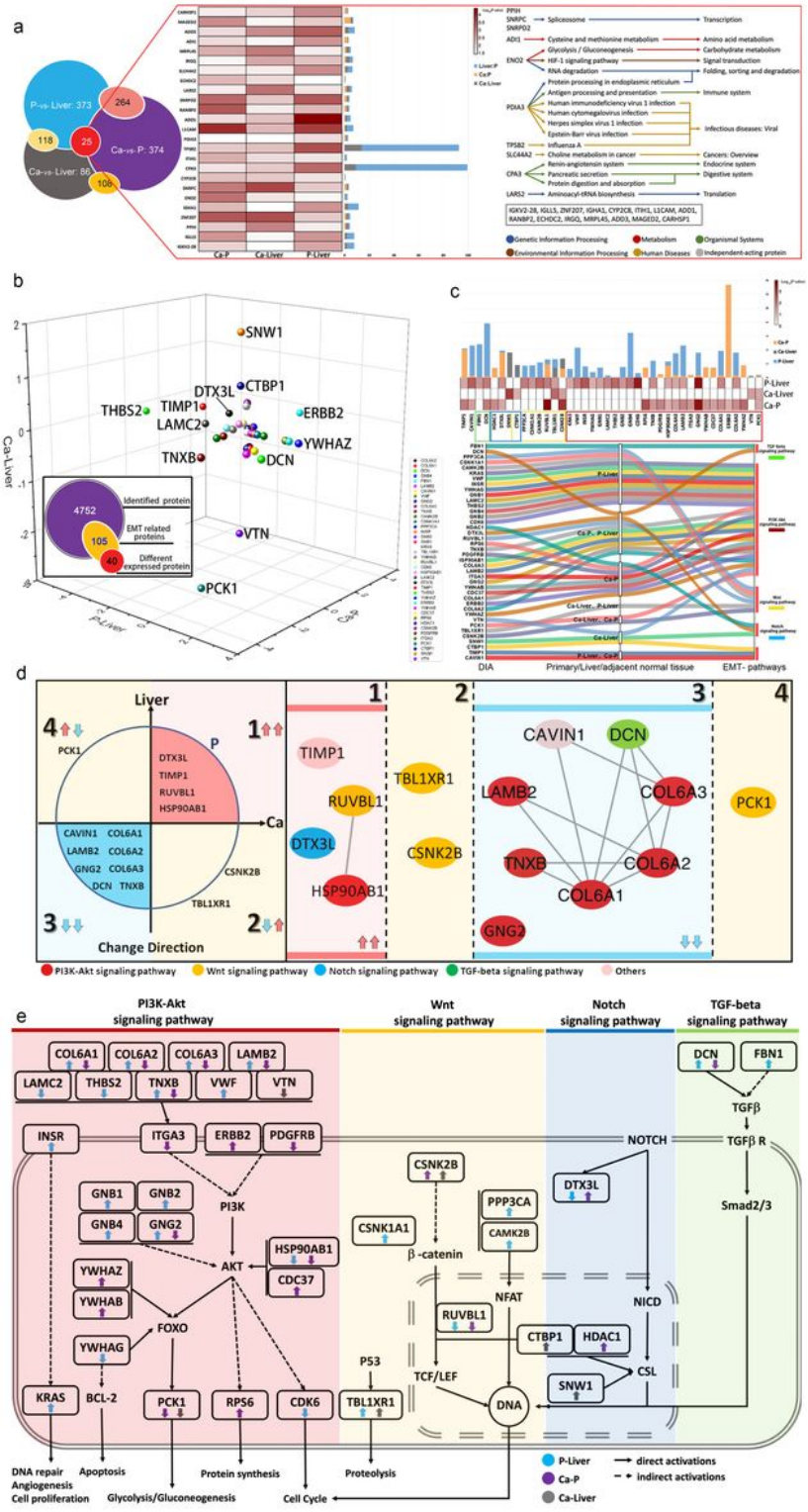

Figure 4

EMT-related metastasis core kinesin and structural helper proteins. (A) Relationship of differentially expressed proteins among Ca-Liver, Ca-P and P-Liver. Summary of 25 covariously expressed proteins in the three groups. (B) The change directions of 40 EMT-related differentially expressed proteins among Ca-Liver, Ca-P and P-Liver are shown in three-dimensional scatter plot. Log2 (FC) of protein levels in P-Liver (x axis), Ca-P (y axis) and Ca-Liver (z axis) are shown. (C) Integrated analysis of 40 EMT-related proteins.

(D) 40 EMT-related Proteins can be divided into four groups: (1) upregulated in both Ca-P and Liver-P; (2) upregulated in Ca-Liver; (3) downregulated in both Ca-P and Liver-P; (4) downregulated in Ca-Liver. (E) Overview of signaling pathways based on integrated proteogenomic analysis.

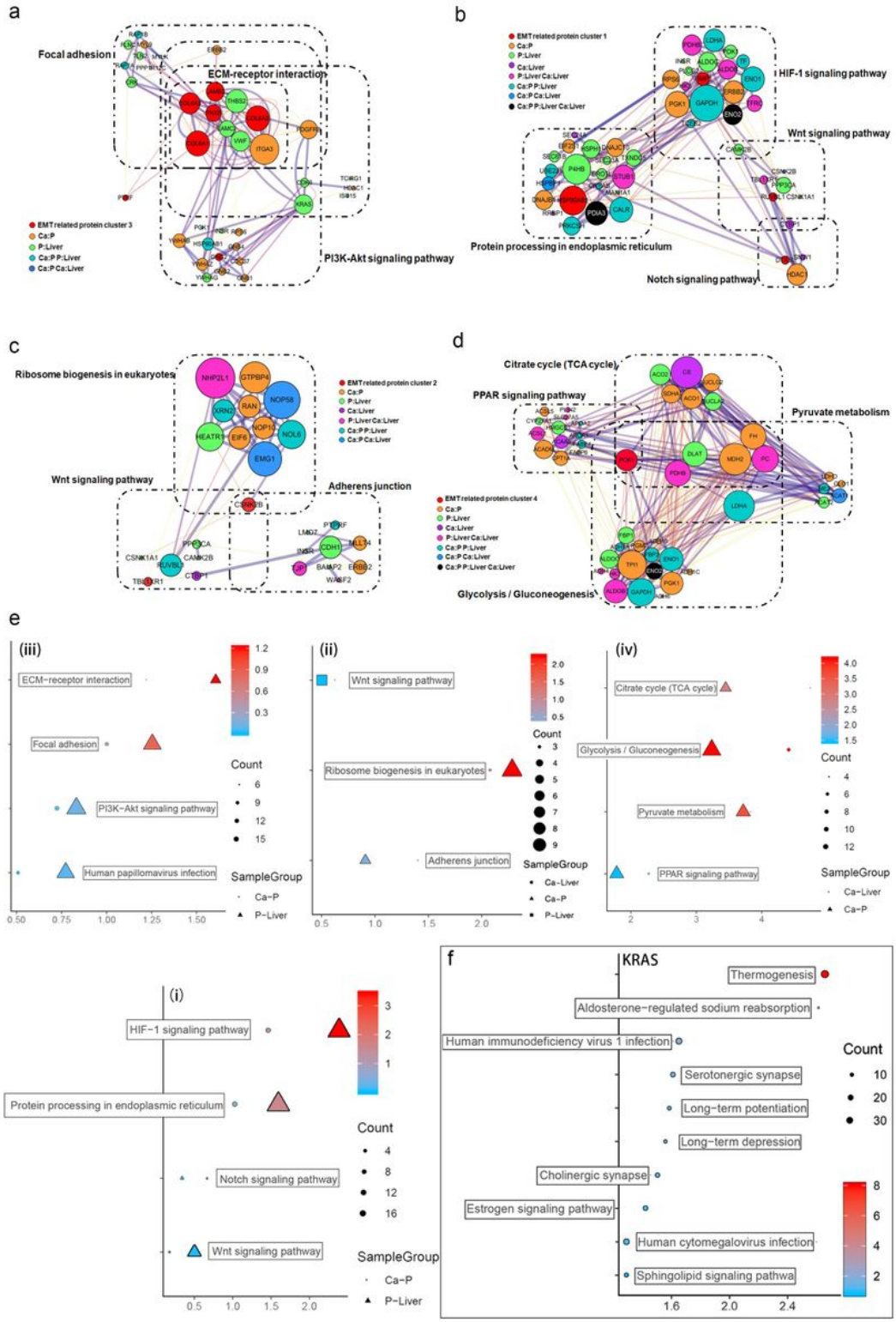

Figure 5

Clusters of EMT inferred by an integrated proteomic analysis. (A) All proteins in the pathway of differentially expressed proteins form the biggest EMT-related protein cluster. Interconnected network

modules identified in some clusters. (B) All proteins in the pathway of differentially expressed proteins form EMT-related protein cluster (1). (C) All proteins in the pathway of differentially expressed proteins form EMT-related protein cluster (2). (D) All proteins in the pathway of differentially expressed proteins (PCK1) form EMT-related protein cluster (2), which majorly involved in glycolysis/gluconeogenesis, citrate cycle (TCA cycle), pyruvate metabolism and PPAR signaling pathway. (E) Major KEGG enrichment bubble chart in four EMT-related protein cluster. (F) Major KEGG enrichment bubble chart of KRAS.

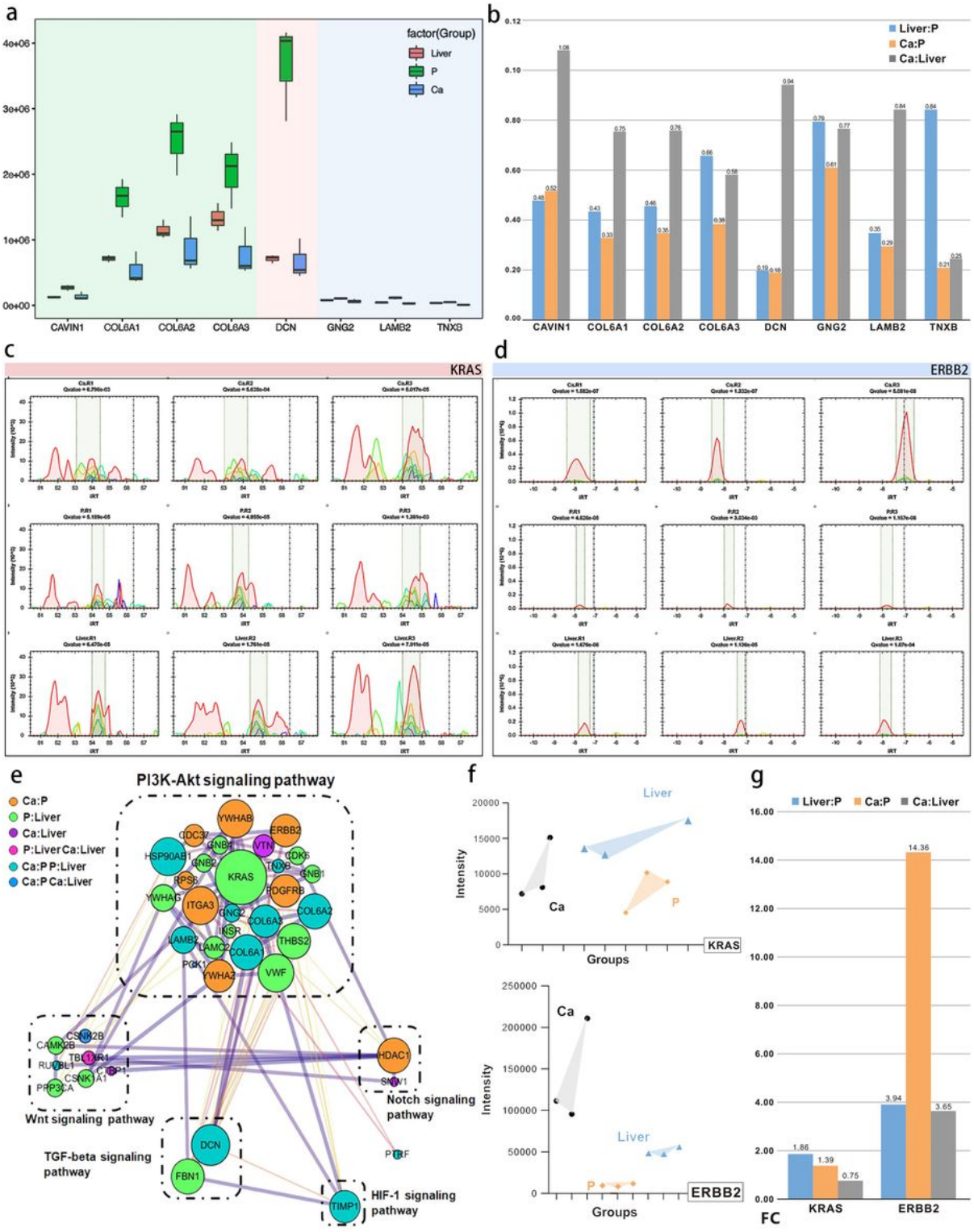

Figure 6

Validation of EMT-related protein cluster by PRM. (A) PRM analysis of EMT-related protein in the validation cohort. (B) The change of target protein expression. (C-D) Peak Diagram of the target peptide, including KRAS and ERBB2. (E) PPI analysis of 40 EMT-related proteins. (F) PRM analysis of core proteins in the same validation cohort. (G) The change of KRAS and ERBB2 expression.

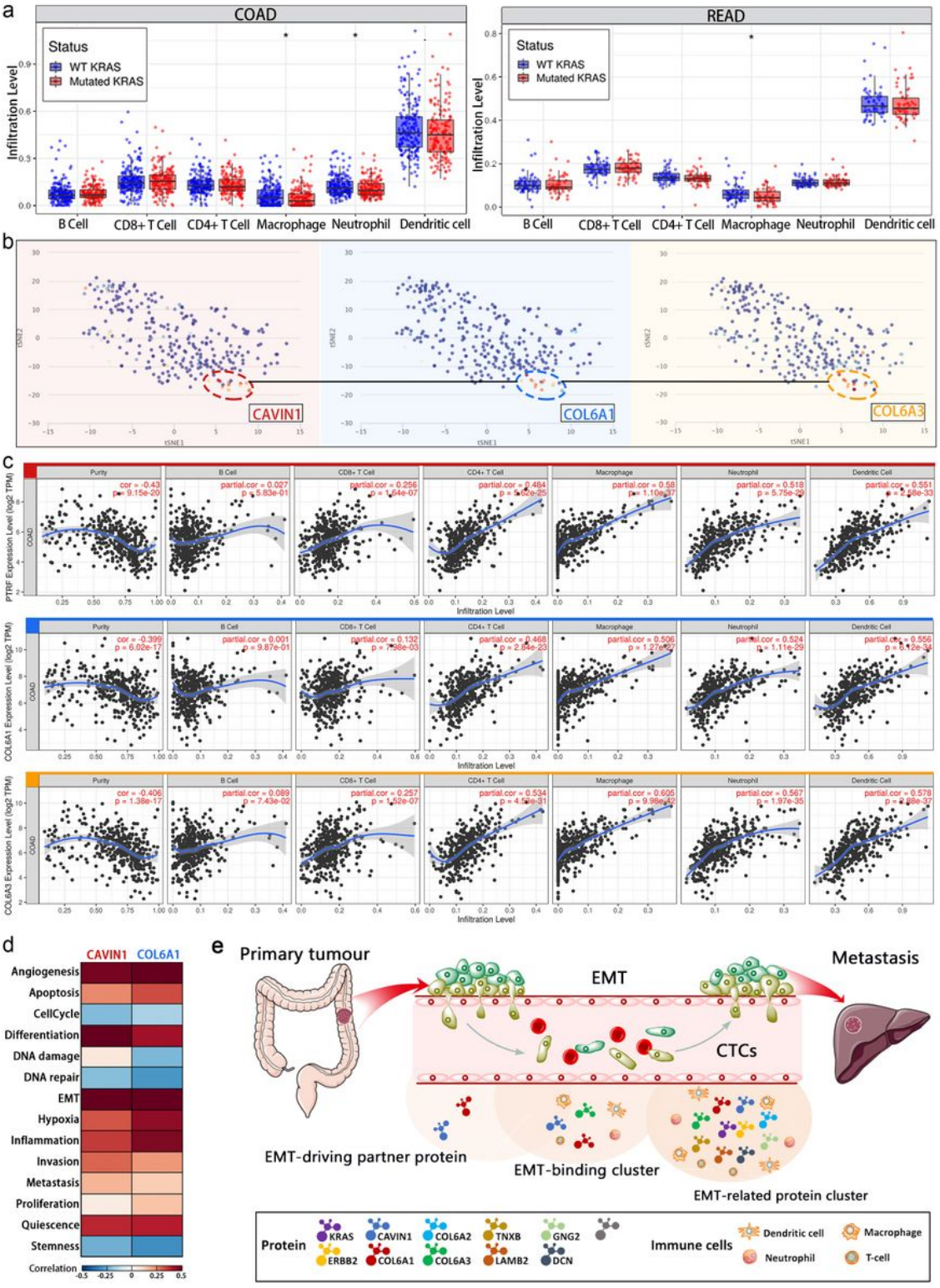

Figure 7

Functional analysis of the core protein and EMT-related protein cluster. (A) Relationship between KRAS mutation and immunologic infiltration abundance. (B) T-SNE describes the distribution of CRC cells. (C) Correlations between the expression of proteins in EMT-binding cluster and immune infiltration levels in COAD. (D) Correlations between EMT-driving partner proteins and functional states in CRC single-cell datasets by CancerSEA analysis. (E) Schematic representation of EMT during the metastatic cascade.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile.docx
- FigureS1.jpg
- FigureS2.jpg
- FigureS3.jpg
- FigureS4.jpg
- FigureS5.jpg
- FigureS6.jpg
- FigureS7.jpg
- SupplementaryTable1.docx
- SupplementaryTable2.docx
- SupplementaryTable3.docx
- SupplementaryTable4.docx
- supplementarymaterial1.docx
- supplementarymaterial2.docx
- supplementarymaterial3.docx