Dual Regulatory Mechanisms of Expression and Mutation Involving Metabolism-Related Genes FDFT1 and UQCR5 during CLM

Yu-Shui Ma,1,2,11 Zhi-Jun Wu,1,11 Hong-Wei Zhang,4,11 Bo Cai,5,11 Tao Huang,6 Hui-Deng Long,2 Hong Xu,7 Yong-Zhong Zhao,8 Yu-Zhen Yin,2 Shao-Bo Xue,2 Liu Li,2 Cheng-Lin Liu,9 Ru-Ting Xie,2 Lin-Lin Tian,2 Ji-Bin Liu,10 Xu-Ming Wu,5 and Da Fu2

1Department of Radiotherapy, Nantong Tumor Hospital, Nantong 226631, China; 2Central Laboratory for Medical Research, Shanghai Tenth People’s Hospital, Tongji University School of Medicine, Shanghai 200072, China; 3Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development, College of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China; 4Analytical Chemistry Platforms, Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai 201210, China; 5Department of Chronic Diseases, Nantong Center for Disease Control and Prevention, Nantong 226007, China; 6Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200025, China; 7Department of Gastroenterology and Hepatology, Hangzhou Red Cross Hospital, Hangzhou 310003, China; 8Department of Medical Genetics, Southern Medical University, Guangzhou 510515, China; 9School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; 10Cancer Institute, Nantong Tumor Hospital, Nantong 226631, China

Colorectal cancer (CRC) is the third most common cancer worldwide, and liver metastasis presents a major cause of CRC-associated death. Extensive genomic analysis has provided valuable insight into the pathogenesis and progression of CRC; however, a comprehensive proteogenomic characterization of CRC liver metastasis (CLM) has yet to be reported. Here, we analyzed the proteomes of 44 paired normal colorectal tissues and CRC tissues with or without liver metastasis, as well as analyzed genomics of CRC characterized previously by The Cancer Genome Atlas (TCGA) to conduct integrated proteogenomic analyses. We identified a total of 2,170 significantly deregulated proteins associated with CLM, 14.88% of which were involved in metabolic pathways. The mutated peptide number was found to have potential prognosis value, and somatic variants revealed two metabolism-related genes UQCR5 and FDFT1 that frequently mutated only in the liver metastatic cohort and displayed dysregulated protein abundance with biological function and clinical significance in CLM. Proteogenomic characterization and integrative and comparative genomic analysis provides functional context and prognostic value to annotate genomic abnormalities and affords a new paradigm for understanding human colon and rectal cancer liver metastasis.

INTRODUCTION

Colorectal cancer (CRC) is a significant contributor of cancer morbidity and mortality.1–3 Almost half of CRC patients die within 5 years of diagnosis due to the development of recurrent disease and metastasis.4–8 Therefore, it is important to illuminate the molecular basis of CRC liver metastasis (CLM) in hopes of developing new effective treatment modalities. The Cancer Genome Atlas (TCGA) has characterized the genomic features of many types of human cancers, including CRC,9–11 and the Clinical Proteomic Tumor Analysis Consortium has also performed CRC-integrated proteomic analyses.12–15 However, the primary genetic basis of CLM has not been fully elucidated. Understanding the genetic and proteogenomic differences between primary colon cancer and associated metastases to the liver is essential for discovering metastasis-specific molecular biomarkers and for devising a better therapeutic approach for this disease.

RESULTS AND DISCUSSION

To identify protein variants associated with CLM, we collected tumor samples and paired normal colorectal tissues (PNs) from 44 CRC tumor samples with (MT, n = 23) or without liver metastasis (NM, n = 21) in comparison with PNs (Table S1) to perform liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based shotgun proteomics profiling (Figure S1). A total of 12,177,738 spectra were used in the Andromeda engine search, and 63,720 unique peptides were identified in an assembly of 5,758 protein groups with a protein-level false discovery rate (FDR) of 1.0%. Ingenuity pathway analysis with all 5,758 identified proteins showed that about 55% of the
proteins were from the cytoplasm, 28% were from the nucleus, 7% were from the plasma membrane, and 2% were from the extracellular space, whereas 8% of proteins remained unclassified (Figure S2A). The random predicted cellular distribution of the proteins supports the quality of the sample preparation. A scatterplot of protein abundance between CRC tissues and PNs showed that there was a great variation between the NM or MT tumors and PN (Figures S2B and S2C). However, the protein expression between the MT and NM group was positively correlated ($R^2 = 0.86$) (Figure S2D). These results suggest that metastatic and non-metastatic CRCs share similar protein profiles and that there are common molecular alterations at each stage of tumor development.\textsuperscript{16–18}

Among the 5,758 proteins, a total of 1,679 proteins were significantly altered between CRC tissue and normal colorectal tissues (Figure 1A, left bars). There were 2,170 proteins with significant differences in MT tissues when compared with PNs or NM tissues (Figures 1A, middle and right bars, and 1B). To explore the functions of proteins that are dysregulated in CLM, we used DAVID (Database for Annotation, Visualization and Integrated Discovery) analysis software to classify the gene ontology of the 2,170 significantly altered proteins in MT tissues\textsuperscript{19} (Figures S3A–S3C). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of differentially expressed proteins\textsuperscript{20} revealed that metabolic pathways were altered in the MT samples, with 323 significantly differentially expressed proteins (14.88%) involved in metabolic pathways (Figure S3D), which suggests that metabolism-related pathways may play important roles in the liver metastasis of CRC.\textsuperscript{21–24}

We performed unsupervised hierarchical clustering of the expression data from RNA-sequencing data to evaluate the genetic diversity between the primary and metastatic CRCs, based on the assumption that genetically similar CRCs and their matched metastatic foci would be closely related.\textsuperscript{25–27} As expected, the results showed that the MT and LM tissues had closely related expression profiles, indicating clonal and genetic similarity for these pairs. However, the NM samples were more distantly related, suggesting distinct genetic relationships between patients with and without metastasis (Figure S4A). A total of 2,299 genes from the RNA-sequencing were significantly changed in CLM (Figure S4B). KEGG pathway classification enrichment analysis indicated that 915 genes (39.8%) were enriched in metabolism pathways (Figure 1C), which was similar to the findings

---

**Figure 1. Differently Expressed Protein and mRNA-Protein Correlation Analysis**

(A) Numbers of differently expressed proteins ($\geq 2$-fold difference; $p < 0.05$ with a FDR q value $< 0.05$) between the 44 CRC and paired PN samples, and the 23 MT and 21 NM samples. (B) Significantly changed proteins among three groups from LC-MS/MS data. (C) KEGG pathway analysis of the 2,170 differentially expressed proteins. (D) Significantly changed genes among three groups from RNA-sequencing data. (E) 444 significantly changed genes showed significant mRNA-protein correlation, with a mean Spearman’s correlation coefficient of 0.71. Among these, 57 genes were enriched in metabolism pathways.
of the protein analysis (Figure S3D). These results provide additional confirmation that metabolism pathways may play a role in the carcinogenesis and development of CRC.28

For further confirmation of these results, we analyzed RNA-sequencing data from TCGA portal. A total of 6,585 significantly changed genes in CRC tissues were identified (Figure 1D). When compared with the 2,170 significantly changed CLM-associated proteins identified by LC-MS/MS, 444 significantly changed genes showed significant positive mRNA-protein correlation (Figure 1E, left two panels). To determine whether the concordance between protein and mRNA variation is related to the biological function of the gene product, we performed KEGG enrichment analysis, which indicated that among the 444 significantly deregulated genes or proteins, 57 are enriched in metabolic pathways (p = 2.38E−7) (Figure 1E, right panel). These findings further verify the role of metabolic pathway genes in CLM.

To examine the clinical significance of our findings, we evaluated the clinical characters data of 374 CRC patients from TCGA database for the 444 significantly deregulated genes or proteins (Table S2). Our results showed that 17 dysregulated genes were significantly associated with overall survival (OS) and progression-free survival (PFS) of CRC patients. Among them, high expression of four upregulated genes, BGN, EHD2, UQCR5, or COLIA2, was associated with poor survival of CRC patients.29–32 Furthermore, high expression of FDFT1 significantly prolongs the OS and PFS of CRC patients relative to low expression of FDFT1 (Table S3).

A fundamental goal of proteogenomics is to identify protein-coding alterations that are expressed at the protein level.33–35 However, standard database search approaches cannot identify variant peptides from MS/MS data.36–38 Therefore, we created a customized mutation database to search for single amino acid variants (SAAVs) in CRC. A SAAV library was prepared using 113,844 mutated sites in CRC tissues from cBioport, and 16,581 mutated proteins were identified, which constitutes 82.08% of 20,201 proteins in the human protein library (Figure 2A). We determined the total numbers of mutated and non-mutated peptides and tumor-specific mutant peptides (Figure 2B) and found that mutated peptide numbers in MT samples were significantly increased (Figure 2C), which indicates that the mutated peptide number has potential predictive value for CLM.

Among those, 140 SAAVs in 131 proteins occurred only in NM patients (Figure 2D; Table S4), and 223 proteins in 18 MT patients
had 256 SAAVs, of which 110 SAAVs in 100 proteins occurred in both NM and MT samples (Figure 2D; Table S5), and 203 SAAVs in 184 proteins only occurred in MT samples (Figure 2D; Table S6). Similarly to single nucleotide variants (SNVs), some mutations reported previously, such as those in TP53, antigen-presenting cells (APCs), vascular endothelial growth factor (VEGF), and SMAD4 were found. Among these, two mRNA-protein positively changed and metabolism-related proteins, FDFT1, a 47-kDa membrane-associated enzyme located at a branch point in the mevalonate pathway involved in the replication stage of the hepatitis virus C (HCV) life cycle and paclitaxel sensitivity in hypopharynx cancer cell, and UQCR5, a component of the ubiquinol-cytochrome c reductase complex, amplifying in primary breast cancer core biopsy samples and overexpressing in gastric cancer, were found to be an occurrence of somatic alterations, which was validated with Sanger sequencing, in the 18.2% (8/44) and 9.1% (4/44) of CRC, and are specific to CLM (Figure 3A). Mutation in FDFT1 was predicted to lose the function; however, mutation in UQCR5 was predicted to gain the function, and increased copy number enhanced UQCR5 gene expression and lead to increased protein expression when scale-invariant feature transform (SIFT) and PolyPhen2 were used for predicted function of FDFT1 and UQCR5 mutations (Figure 3B).

Moreover, FDFT1 knockdown (Figures S5A and S5B) or UQCR5 overexpression (Figures S5C and S5D) led to a significant increase of migrated distance (Figure 4A) and an obvious increase in matrigel invasion (Figure 4B). FDFT1 was found to downregulate in CRC tissues when compared with normal colorectal tissues (Figure 4C). Determination of UQCR5 gene expression identified overexpression (Figure 4D) and negatively correlated with FDFT1 expression in normal colorectal tissues and CRC tissues (Figure 4E).
Patients with lower FDFT1 expression were found to have shorter median OS (Figures S6A and S7A) and PFS (Figures S6B and S7B), and high UQCR5 expression was an important risk factor for OS (Figure S6A and S7A) and PFS (Figures S6B and S7B). Moreover, concomitant low expression of FDFT1 and high expression of UQCR5 correlated with a shorter median OS (Figures S6A and S7A) and PFS (Figures S6B and S7B) in CRC patients. These studies provide direct evidence for contribution of metabolism-related genes.

To the best of our knowledge, this is the first comprehensive study to use proteogenomic profiling of primary CRCs from patients with or without liver metastasis to define the dominant events of metastatic lesions in terms of their expression and mutation. Our comprehensive integrative analysis of 44 colorectal tumors and normal pairs provides a number of insights into the biology of CLM and identifies potential therapeutic targets. Moreover, our characterization of the annotated metastatic CRC proteome clarifies the power of integrating genomics (SNVs) and proteomics (SAAVs). This approach provides new insights into the roles of these protein alterations in CLM, which can be broadly extended to understand the roles of protein mutation in other cancers.

**MATERIALS AND METHODS**

Details of sample preparation and analysis are described in the Supplemental Materials and Methods. The study was examined and approved by the Ethics Committee of the Shanghai Tenth People's Hospital, Tongji University School of Medicine. This study was registered with ClinicalTrials.gov: NCT02917707). Forty-four paired fresh primary tumors from the colorectal and PNs were used for proteogenomic analysis. The reliability of the exome analysis and somatic variant identification strategies was assessed using PCR and Sanger sequencing. Nine specimens from three CRC patients with metastasis and three specimens from three CRC patients without liver metastasis were obtained for RNA-sequencing analysis. Human CRC cell line SW480 was used for scratch-wound and in vitro invasion assays. All calculations were performed with SPSS 20.0 software.

Whole-exome sequencing data from this study are available for download through NCBI Sequence Read Archive: PRJNA358865.
All RNA-seq data have been deposited in the Gene Expression Omnibus: GSE92914. All of the MS proteomics data have been deposited to iProX (https://www.iprox.org/index): IPX00083203 and IPX00083210.

SUPPLEMENTAL INFORMATION
Supplemental Information can be found online at https://doi.org/10.1016/j.omto.2019.04.008.

AUTHOR CONTRIBUTIONS
Y.-S.M., J.-B.L., X.-M.W., and D.F. designed and supervised the research and experiments. Y.-S.M., Z.-J.W., B.C., H.-W.Z., T.H., H.-D.L., H.X., Y.-Z.Z., Y.-Z.Y., R.-T.X., L.L.T., and D.F. analyzed the genetic data. Y.-S.M., Z.-J.W., B.C., H.-W.Z., T.H., S.-B.X., L.L., C.-L.L., J.-B.L., X.-M.W., and D.F. performed the proteogenomic experiments and analyzed the data. Y.-S.M., Z.-J.W., B.C., H.-W.Z., T.H., S.-B.X., L.L., R.-T.X., J.-B.L., and X.-M.W. performed tissue acquisition and clinical data collection. Y.-S.M., Z.-J.W., B.C., H.-W.Z., S.-B.X., L.L., R.-T.X., J.B.L., and D.F. performed functional validation experiments. All authors contributed to the final version of the manuscript.

CONFLICTS OF INTEREST
The authors declare no competing interests.

ACKNOWLEDGMENTS
We would like to thank Dr. Wei Zhang for data analysis and critical discussion of the manuscript. This study was supported partly by grants from the National Natural Science Foundation of China (81772932, 81472202, 81201535, 81302065, 81671716, 81301993, 81702243, 81372175, and 81472209), The Fundamental Research Funds for the Central Universities (22120170212 and 22120170117), the Shanghai Natural Science Foundation (12ZR1436000 and 16ZR1428900), the Shanghai Municipal Commission of Health and Family Planning (201540228 and 201440398), the Construction of Clinical Medical Center for Tumor Biological Samples in Nantong (HS2016004), and the Jiangsu 333 Program (BRA2017205).

REFERENCES
1. Laisse, P. (2019). The forkhead-box family of transcription factors: key molecular players in colorectal cancer pathogenesis. Mol. Cancer 18, 5.
2. Cristobal, I., Sanz-Alvarez, M., Torreon, B., Santos, A., Luque, M., Rojo, F., and Garcia-Foncillas, J. (2018). Potential therapeutic impact of miR-145 deregulation in colorectal cancer. Mol. Ther. 26, 1399-1400.
3. Wang, X., Sun, D., Tai, J., Chen, S., Hong, S., and Wang, L. (2019). ZNF280A promotes proliferation and tumorigenesis via inactivating the hippo-signaling pathway in colorectal cancer. Mol. Ther. Oncolytics 2, 204-213.
4. Li, X.N., Wang, Z., Ye, C.X., Zhao, B.C., Li, Z.L., and Yang, Y. (2018). RNA sequencing reveals the expression profiles of circRNA and indicates that circDDX17 acts as a tumor suppressor in colorectal cancer. J. Exp. Clin. Cancer Res. 37, 325.
5. Zhu, Y., Wang, C., Becker, S.A., Hurst, K., Nogueira, L.M., Findlay, V.J., and Camp, E.R. (2018). miR-145 antagonizes SNAI1-mediated stemness and radiation resistance in colorectal cancer. Mol. Ther. 26, 744-754.
6. Barrolla-Villava, M., Cervantes, A., and Bardelli, A. (2018). Preclinical models for precision oncology. Biochim Biophys Acta Rev Cancer 1870, 239-246.
7. O’Leary, M.P., Warner, S.G., Kim, S.I., Chaurasiya, S., Lu, J., Choi, A.H., Park, A.K., Woo, Y., Fong, Y., and Chen, N.G. (2018). A novel oncolytic chimeric orthopoxvirus encoding luciferase enables real-time view of colorectal cancer cell infection. Mol. Ther. Oncolytics 9, 13–21.
8. Bleau, A.M., Redrado, M., Nistal-Villan, E., Villalba, M., Exposito, F., Redin, E., de Abearsturi, A.L., Larzabal, L., Freire, J., Gomez-Roman, I., and Calvo, A. (2018). miR-146a targets c-met and abolishes colorectal cancer liver metastasis. Cancer Lett. 414, 257–267.
9. Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature 487, 330-337.
10. Xu, H., Wang, C., Song, H., Xu, Y., and Ji, G. (2019). RNA-Seq profiling of circular RNAs in human colorectal Cancer liver metastasis and the potential biomarkers. Mol. Cancer 18, 8.
11. Toiyama, Y., Okugawa, Y., Fleshman, J., Richard Boland, C., and Goel, A. (2018). MicroRNAs as potential liquid biopsy biomarkers in colorectal cancer: A systematic review. Biochim Biophys Acta Rev Cancer 1870, 274–282.
12. Kryeziu, K., Bruun, I., Guren, T.K., Sven, A., and Lothe, R.A. (2019). Combination therapies with HSPO9 inhibitors against colorectal cancer. Biochim Biophys Acta Rev Cancer 1871, 240–247.
13. Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., et al. NCI CPTAC (2014). Proteogenomic characterization of human colon and rectal cancer. Nature 513, 382–387.
14. Herrera, M., Llorens, C., Rodriguez, M., Herrera, A., Ramos, R., Gil, B., Candia, A., Larriba, M.J., Garre, P., Earl, J., et al. (2018). Differential distribution and enrichment of non-coding RNAs in exosomes from normal and Cancer-associated fibroblasts in colorectal cancer. Mol. Cancer 17, 114.
15. Li, J., Song, P., Jiang, T., Dai, D., Wang, H., Sun, J., Zhu, L., Xu, W., Feng, L., Shi, V.Y., et al. (2018). Heat shock factor 1 epigenetically stimulates glutaminase-1-dependent miTORS activation to promote colorectal carcinogenesis. Mol. Ther. 26, 1828–1839.
16. Bhardwaj, D., Patel, H., Ganguly, R., Shaih, A., Shukla, Y., Sharma, D., and Singh, P. (2018). Novel insights into adult and cancer stem cell biology. Stem Cells Dev. 27, 1527–1539.
17. Ribatti, D. (2018). An historical note on the cell theory. Exp. Cell Res. 364, 1–4.
18. Shastri, A., Choudhury, G., Teixeira, M., Gordon-Mitchell, S, Ramachandaram, N, Bernard, L., Bhattacharyya, S., Lopez, R., Pradhan, K., Giricz, O., et al. (2018). Antisense STAT3 inhibitor decreases viability of myelodysplastic and leukemic stem cells. J. Clin. Invest. 128, 5479–5488.
19. Ma, Y.S., Yu, F., Zhong, X.M., Lu, G.X., Cong, X.L., Xue, S.B., Xie, W.T., Hou, L.K., Pang, L.J., Wu, W., et al. (2018). miR-30 family reduction maintains self-renewal and promotes tumorigenesis in NSCLC-initiating cells by targeting oncogene T4451. Mol. Ther. 26, 2751–2756.
20. Zhang, Y.J., Ma, Y.S., Xia, Q., Yu, F., Lv, Z.W., Jia, C.Y., Jiang, X.Z., Zhang, L., Shao, Y.C., Xie, W.T., et al. (2018). MicroRNA mRNA integrated analysis based on a case of well-differentiated thyroid cancer with both metastasis and metastatic recurrence. Oncol. Rep. 40, 3803–3811.
21. Xing, F., Wang, S., and Zhou, J. (2018). The expression of microRNA-598 inhibits ovarian cancer cell proliferation and metastasis by targeting URI. Mol. Ther. Oncolytics 2, 9–15.
22. Jin, F., Yang, R., Wei, Y., Wang, D., Zhu, Y., Wang, X., Lu, Y., Wang, Y., Zen, K., and Li, L. (2019). HIF-1α-induced miR-23a-27a-24 cluster promotes colorectal cancer progression via reprogramming metabolism. Cancer Lett. 440-441, 211–222.
23. Pang, R., Law, W.L., Chu, A.C., Poon, J.T., Lam, C.S., Chow, A.K., Ng, L., Cheung, L.W., Lan, X.R., Lan, H.Y., et al. (2010). A subpopulation of CD26+ cancer stem cells with metastatic capacity in human colorectal cancer. Cell Stem Cell 6, 603-615.
24. Yue, B., Liu, C., Sun, H., Liu, M., Song, C., Cui, R., Qiu, S., and Zhong, M. (2018). A positive feed-forward loop between IncRNA-CYTOR and Wnt/β-catenin signaling promotes metastasis of colon cancer. Mol. Ther. 26, 1287–1298.
25. Lu, H.M., Yi, W.W., Ma, Y.S., Wu, W., Yu, F., Fan, H.W., Lv, Z.W., Yang, H.Q., Chang, Z.Y., Zhang, C., et al. (2018). Prognostic implications of decreased
