Molecular Therapy
Oncolytics
Original Article

AMERICAN SOCIETY of GENE & CELL THERAPY

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

Yu-Shui Ma,[1,2,3,11] 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 Fu[2]

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

Received 3 January 2019; accepted 30 April 2019;
https://doi.org/10.1016/j.omto.2019.04.008.

[11]These authors contributed equally to this work.

Correspondence: Ji-Bin Liu, Cancer Institute, Nantong Tumor Hospital, Nantong 226631, China.
E-mail: tians2008@163.com
Correspondence: Xu-Ming Wu, Department of Chronic Diseases, Nantong Center for Disease Control and Prevention, Nantong 226007, China.
E-mail: ntzlyyywk@163.com
Correspondence: Da Fu, Central Laboratory for Medical Research, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 200072, China.
E-mail: fu800da900@126.com

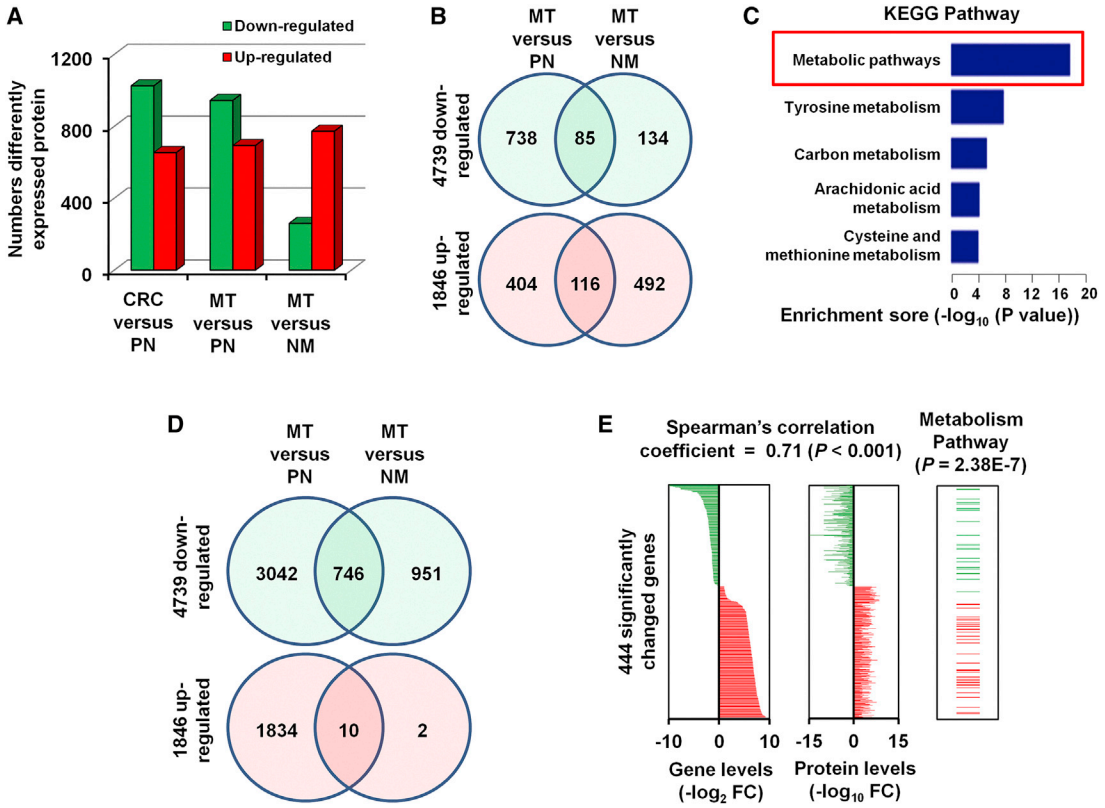

Figure 1. Differently Expressed Protein and mRNA-Protein Correlation Analysis

(A) Numbers of differently expressed proteins (≥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.

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.[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[19] (Figures S3A–S3C). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of differentially expressed proteins[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.[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.[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

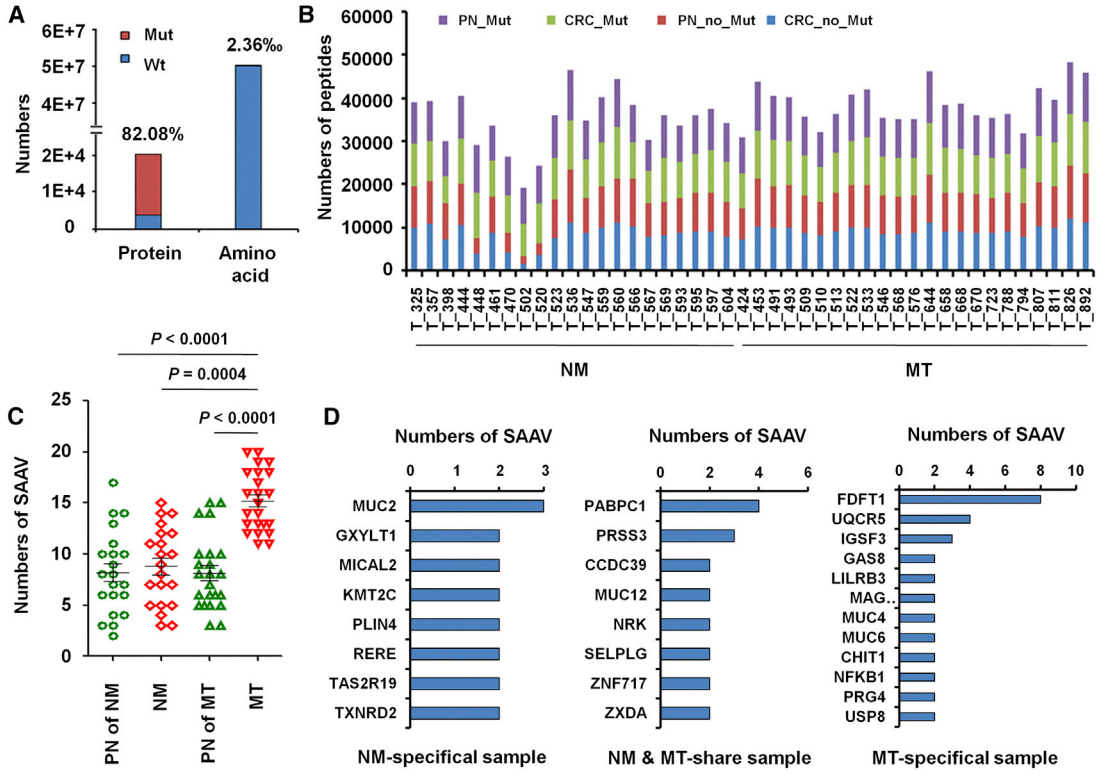

Figure 2. Numbers of SAAVs in Paired PN, NM, or MT Samples

(A) The proportion of mutated proteins and amino acids in CRC samples were calculated by comparing LC-MS/MS data for the standard protein library and SAAV library. (B) The mutated and non-mutated peptides numbers of 44 paired CRC tissues. (C) Numbers of SAAVs in 21 NM, 23 MT, and their PNs. (D) Numbers of NM-specific, MT-specific, and NM- and MT-shared SAAVs. The mutated peptides were identified by comparing LC-MS/MS data for the standard protein library and SAAV library.

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 *COL1A2*, 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

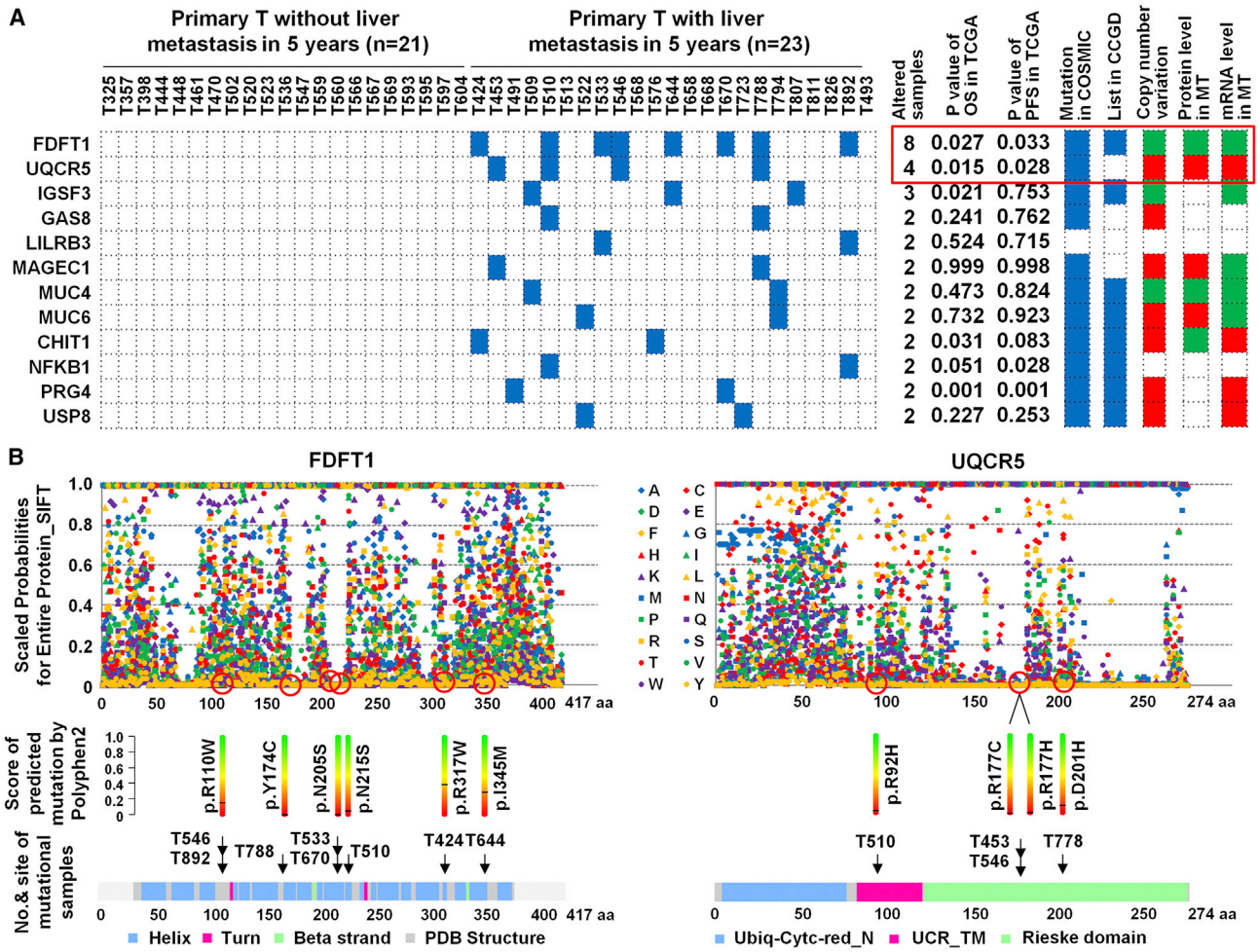

Figure 3. The Frequency and Distribution of Mutational Proteins Specifically Altered in MT for 44 CRC

(A) Heatmap comparing the frequency and distribution of 12 mutational proteins in 44 CRC (including 23 MT and 21 NM). Red, amplification or high expression; green, deletion or low expression. (B) Scaled probabilities for entire protein score of predicted mutation using SIFT and PolyPhen2 for FDFT1 and UQCR5 mutations.

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.[39–43] 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[44] and paclitaxel sensitivity in hypopharynx cancer cell,[45] and UQCR5, a component of the ubiquinol-cytochrome *c* reductase complex, amplifying in primary breast cancer core biopsy samples[46–48] and overexpressing in gastric cancer,[49] 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).

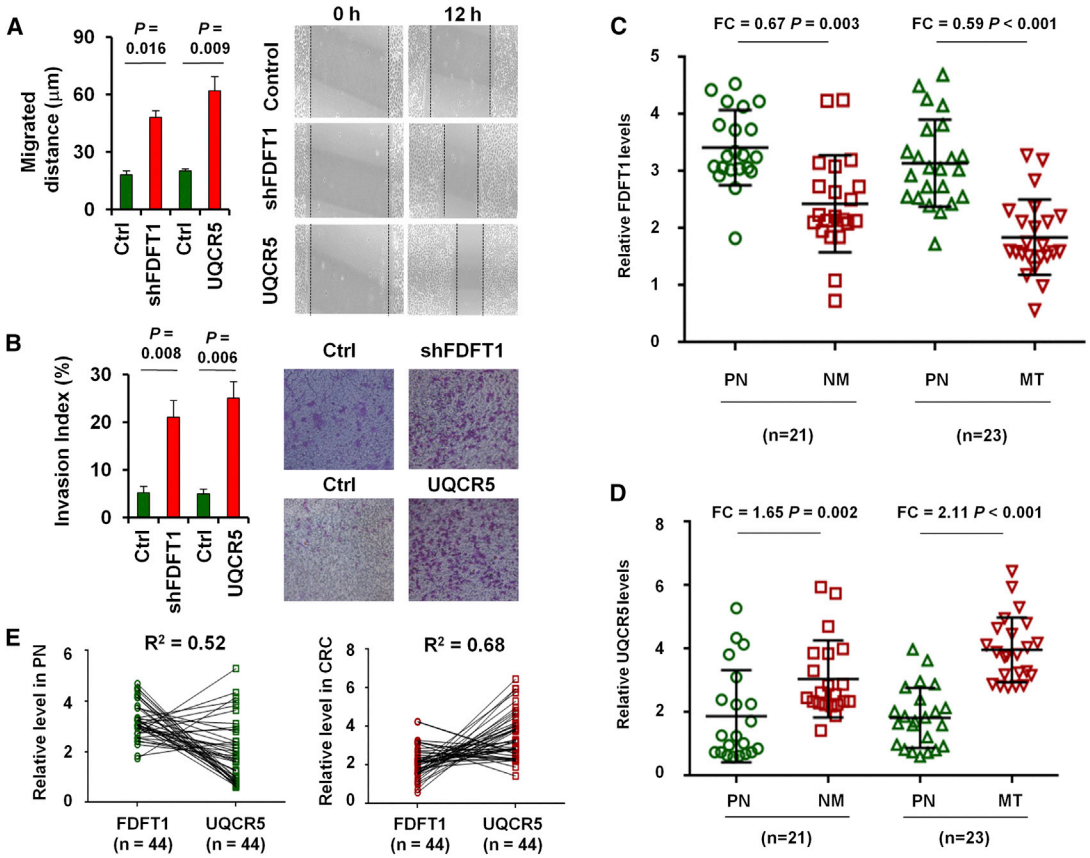

Figure 4. The Biological Function and Clinical Significance of FDFT1 and UQCR5 in CRC

Wound-healing assay (A) and migration abilities (B) of the parental and shFDFT1 or UQCR5 overexpressed SW480 cells. LC-MS/MS to quantify FDFT1 (C) and UQCR5 (D) levels in 44 paired CRC (including 23 MT and 21 NM) and normal colorectal tissues. (E) The expression correlation between FDFT1 and UQCR5 expression levels in 44 adjacently normal colorectal tissues or CRC tissues.

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-sequencing 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. Laissue, P. (2019). The forkhead-box family of transcription factors: key molecular players in colorectal cancer pathogenesis. Mol. Cancer *18*, 5.
2. Cristóbal, I., Sanz-Alvarez, M., Torrejón, B., Santos, A., Luque, M., Rojo, F., and García-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 tumorigenicity via inactivating the hippo-signaling pathway in colorectal cancer. Mol. Ther. Oncolytics *12*, 204–213.
4. Li, X.N., Wang, Z.J., 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. Ibarrola-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 Aberasturi, A.L., Larzabal, L., Freire, J., Gomez-Roman, J., 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, J., Guren, T.K., Sveen, A., and Lothe, R.A. (2019). Combination therapies with HSP90 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., Rodríguez, 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., Shin, V.Y., et al. (2018). Heat shock factor 1 epigenetically stimulates glutaminase-1-dependent mTOR activation to promote colorectal carcinogenesis. Mol. Ther. *26*, 1828–1839.
16. Bhartiya, D., Patel, H., Ganguly, R., Shaikh, 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., Choudhary, G., Teixeira, M., Gordon-Mitchell, S., Ramachandra, 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 TM4SF1. Mol. Ther. *26*, 2751–2765.
20. Zhang, Y.J., Ma, Y.S., Xia, Q., Yu, F., Lv, Z.W., Jia, C.Y., Jiang, X.X., 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 *12*, 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 lncRNA-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

microRNA-101-3p expression in patients with non-small cell lung cancer. Oncol. Lett. *16*, 7048–7056.

26. Yu, F., Liu, J.B., Wu, Z.J., Xie, W.T., Zhong, X.J., Hou, L.K., Wu, W., Lu, H.M., Jiang, X.H., Jiang, J.J., et al. (2018). Tumor suppressive microRNA-124a inhibits stemness and enhances gefitinib sensitivity of non-small cell lung cancer cells by targeting ubiquitin-specific protease 14. Cancer Lett. *427*, 74–84.

27. Ma, Y.S., Wu, Z.J., Bai, R.Z., Dong, H., Xie, B.X., Wu, X.H., Hang, X.S., Liu, A.N., Jiang, X.H., Wang, G.R., et al. (2018). DRR1 promotes glioblastoma cell invasion and epithelial-mesenchymal transition via regulating AKT activation. Cancer Lett. *423*, 86–94.

28. Bhullar, K.S., Lagarón, N.O., McGowan, E.M., Parmar, I., Jha, A., Hubbard, B.P., and Rupasinghe, H.P.V. (2018). Kinase-targeted cancer therapies: progress, challenges and future directions. Mol. Cancer *17*, 48.

29. Kwon, J.J., Factora, T.D., Dey, S., and Kota, J. (2018). A systematic review of miR-29 in cancer. Mol. Ther. Oncolytics *12*, 173–194.

30. Ma, Y.S., Huang, T., Zhong, X.M., Zhang, H.W., Cong, X.L., Xu, H., Lu, G.X., Yu, F., Xue, S.B., Lv, Z.W., and Fu, D. (2018). Proteogenomic characterization and comprehensive integrative genomic analysis of human colorectal cancer liver metastasis. Mol. Cancer *17*, 139.

31. Nimmakayala, R.K., Batra, S.K., and Ponnusamy, M.P. (2019). Unraveling the journey of cancer stem cells from origin to metastasis. Biochim Biophys Acta Rev Cancer *1871*, 50–63.

32. Ylä-Herttuala, S. (2017). The pharmacology of gene therapy. Mol. Ther. *25*, 1731–1732.

33. Bach, D.H., Park, H.J., and Lee, S.K. (2017). The dual role of bone morphogenetic proteins in cancer. Mol. Ther. Oncolytics *8*, 1–13.

34. Hou, Z., Guo, K., Sun, X., Hu, F., Chen, Q., Luo, X., Wang, G., Hu, J., and Sun, L. (2018). TRIB2 functions as novel oncogene in colorectal cancer by blocking cellular senescence through AP4/p21 signaling. Mol. Cancer *17*, 172.

35. Krishnamurthy, N., and Kurzrock, R. (2018). Targeting the Wnt/beta-catenin pathway in cancer: Update on effectors and inhibitors. Cancer Treat. Rev. *62*, 50–60.

36. Xu, S., Yue, Y., Zhang, S., Zhou, C., Cheng, X., Xie, X., Wang, X., and Lu, W. (2018). STON2 negatively modulates stem-like properties in ovarian cancer cells via DNMT1/MUC1 pathway. J. Exp. Clin. Cancer Res. *37*, 305.

37. van Duijnhoven, F.J.B., Jenab, M., Hveem, K., Siersema, P.D., Fedirko, V., Duell, E.J., Kampman, E., Halfweeg, A., van Kranen, H.J., van den Ouweland, J.M.W., et al. (2018). Circulating concentrations of vitamin D in relation to pancreatic cancer risk in European populations. Int. J. Cancer *142*, 1189–1201.

38. Li, H.P., Peng, C.C., Wu, C.C., Chen, C.H., Shih, M.J., Huang, M.Y., Lai, Y.R., Chen, Y.L., Chen, T.W., Tang, P., et al. (2018). Inactivation of the tight junction gene CLDN11 by aberrant hypermethylation modulates tubulins polymerization and promotes cell migration in nasopharyngeal carcinoma. J. Exp. Clin. Cancer Res. *37*, 102.

39. Chen, X., Guan, H., Liu, X.D., Xie, D.F., Wang, Y., Ma, T., Huang, B., and Zhou, P.K. (2018). p53 positively regulates the expression of cancer stem cell marker CD133 in HCT116 colon cancer cells. Oncol. Lett. *16*, 431–438.

40. Bressy, C., Hastie, E., and Grdzelishvili, V.Z. (2017). Combining oncolytic virotherapy with p53 tumor suppressor gene therapy. Mol. Ther. Oncolytics *5*, 20–40.

41. Liu, Y., Lv, J., Liu, J., Liang, X., Jin, X., Xie, J., Zhang, L., Chen, D., Fiskesund, R., Tang, K., et al. (2018). STAT3/p53 pathway activation disrupts IFN-β-induced dormancy in tumor-repopulating cells. J. Clin. Invest. *128*, 1057–1073.

42. Tzchori, I., Falah, M., Shteynberg, D., Levin Ashkenazi, D., Loberman, Z., Perry, L., and Flugelman, M.Y. (2018). Improved patency of ePTFE grafts as a hemodialysis access site by seeding autologous endothelial cells expressing Fibulin-5 and VEGF. Mol. Ther. *26*, 1660–1668.

43. Jiang, T., Ye, L., Han, Z., Liu, Y., Yang, Y., Peng, Z., and Fan, J. (2017). miR-19b-3p promotes colon cancer proliferation and oxaliplatin-based chemoresistance by targeting SMAD4: validation by bioinformatics and experimental analyses. J. Exp. Clin. Cancer Res. *36*, 131.

44. Park, E.M., Nguyen, L.N., Lim, Y.S., and Hwang, S.B. (2014). Farnesyl-diphosphate farnesyltransferase 1 regulates hepatitis C virus propagation. FEBS Lett. *588*, 1813–1820.

45. Xu, C.Z., Shi, R.J., Chen, D., Sun, Y.Y., Wu, Q.W., Wang, T., and Wang, P.H. (2013). Potential biomarkers for paclitaxel sensitivity in hypopharynx cancer cell. Int. J. Clin. Exp. Pathol. *6*, 2745–2756.

46. Natrajan, R., Mackay, A., Wilkerson, P.M., Lambros, M.B., Wetterskog, D., Arnedos, M., Shiu, K.K., Geyer, F.C., Langerød, A., Kreike, B., et al. (2012). Functional characterization of the 19q12 amplicon in grade III breast cancers. Breast Cancer Res. *14*, R53.

47. Sato, T., Chang, H.C., Bayeva, M., Shapiro, J.S., Ramos-Alonso, L., Kouzu, H., Jiang, X., Liu, T., Yar, S., Sawicki, K.T., et al. (2018). mRNA-binding protein tristetraprolin is essential for cardiac response to iron deficiency by regulating mitochondrial function. Proc. Natl. Acad. Sci. USA *115*, E6291–E6300.

48. Kim, H.C., Chang, J., Lee, H.S., and Kwon, H.J. (2017). Mitochondrial UQCRB as a new molecular prognostic biomarker of human colorectal cancer. Exp. Mol. Med. *49*, e391.

49. Jun, K.H., Kim, S.Y., Yoon, J.H., Song, J.H., and Park, W.S. (2012). Amplification of the UQCRFS1 gene in gastric cancers. J. Gastric Cancer *12*, 73–80.

OMTO, Volume *14*

Supplemental Information

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

Yu-Shui Ma, Zhi-Jun Wu, Hong-Wei Zhang, Bo Cai, Tao Huang, Hui-Deng Long, Hong Xu, Yong-Zhong Zhao, Yu-Zhen Yin, Shao-Bo Xue, Liu Li, Cheng-Lin Liu, Ru-Ting Xie, Lin-Lin Tian, Ji-Bin Liu, Xu-Ming Wu, and Da Fu

Supplementary Figures and Legends

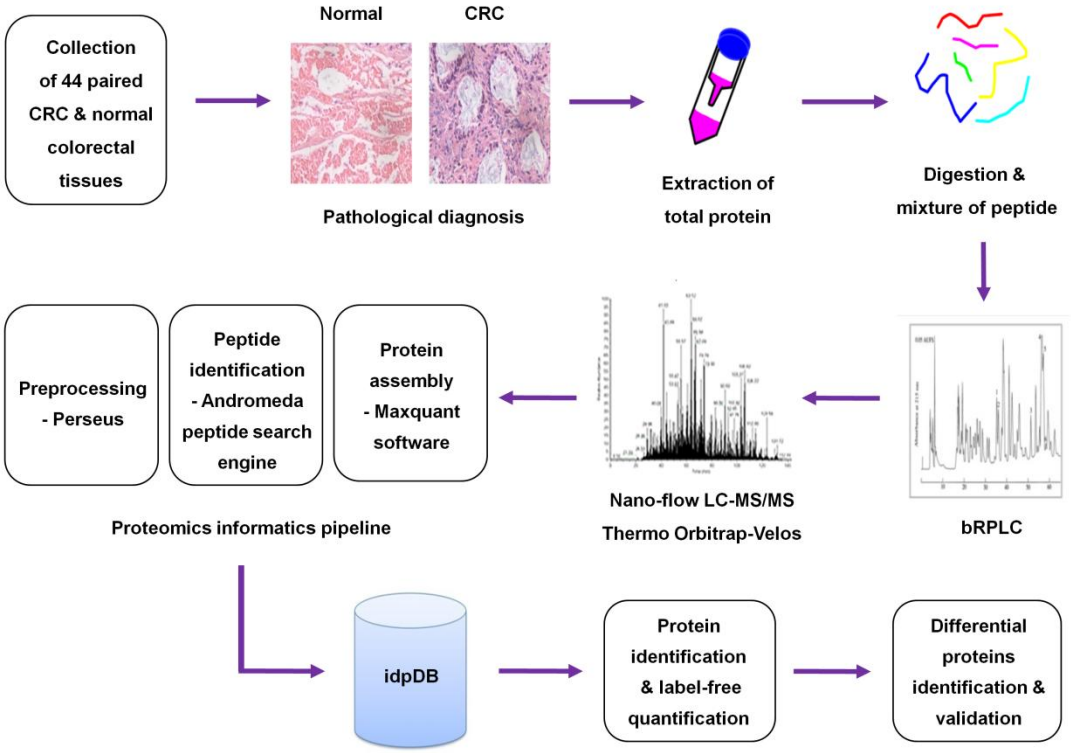

Figure S1. Mass-spectrometry-based proteomics workflow. Protein was extracted from 44 CRC and paired PN tissues and was used to generate tryptic digests. The resulting tryptic peptides were fractionated using off-line bRPLC Collected fractions were pooled and used with a Thermo Orbitrap-Velos MS instrument. Raw data were processed by Perseus software and then used for database and spectral library evaluation using the Andromeda peptide search engine. Identified peptides were assembled using Maxquant software. bRPLC, basic reverse-phase (high-pressure) liquid chromatography; CRC, colorectal cancer; PN paracarcinoma normal tissue.

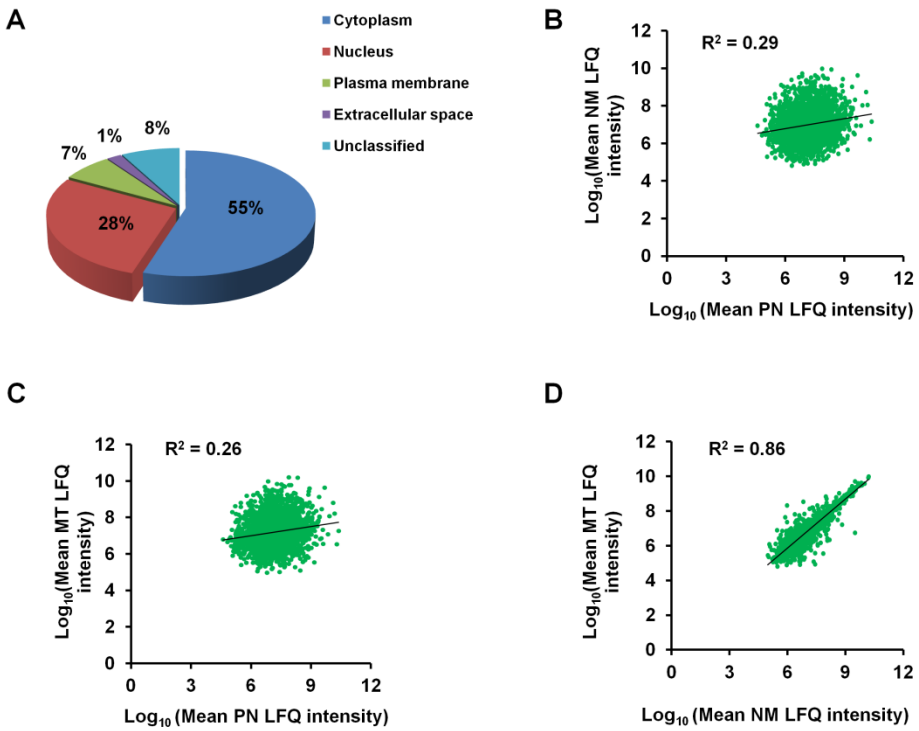

Figure S2. Peptide and protein identification of 44 paired CRC samples. A, The cellular distribution is shown for 5758 proteins identified from 44 CRC patients by mass spectrometry, with a false discovery rate of 1.0 %. Correlation analysis of protein expression between 21 NM and paired PN samples (B), 23 MT and paired PN samples (C), or 23 MT and 21 NM samples (D).

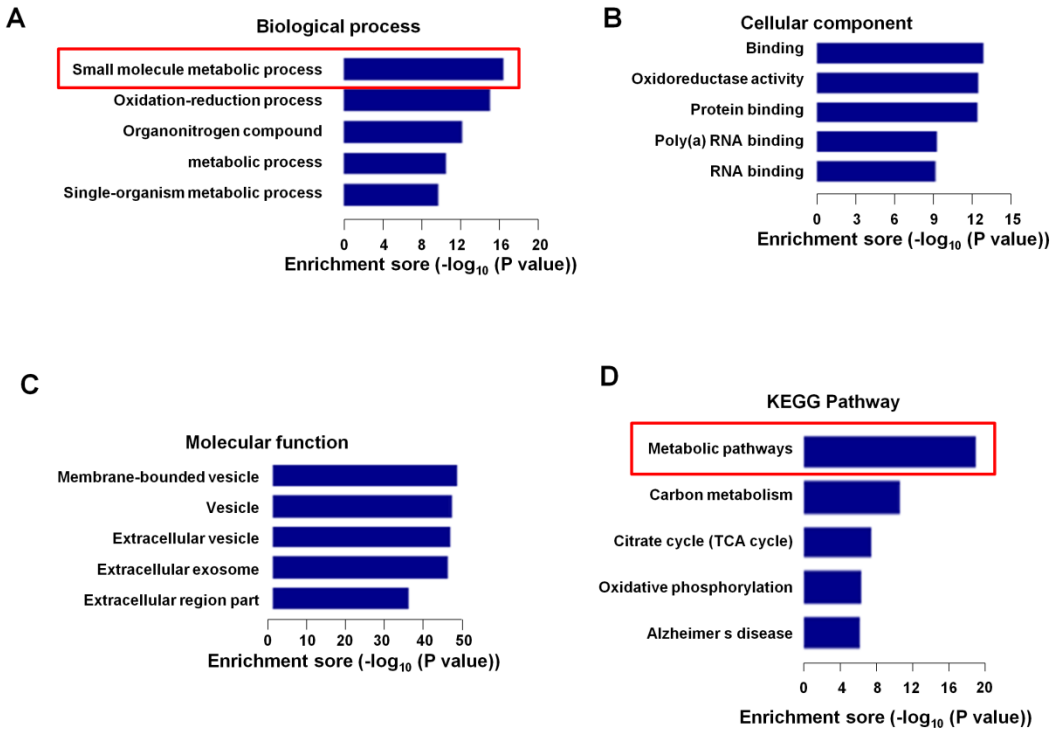

Figure S3. GO and KEGG pathway analysis of differently expressed protein. Gene Ontology using STRING online analysis software classified the 2170 differentially expressed proteins in CLM according to their biological process (A), cellular component (B) and molecular function (C). D, KEGG pathway analysis of the 2170 differentially expressed proteins.

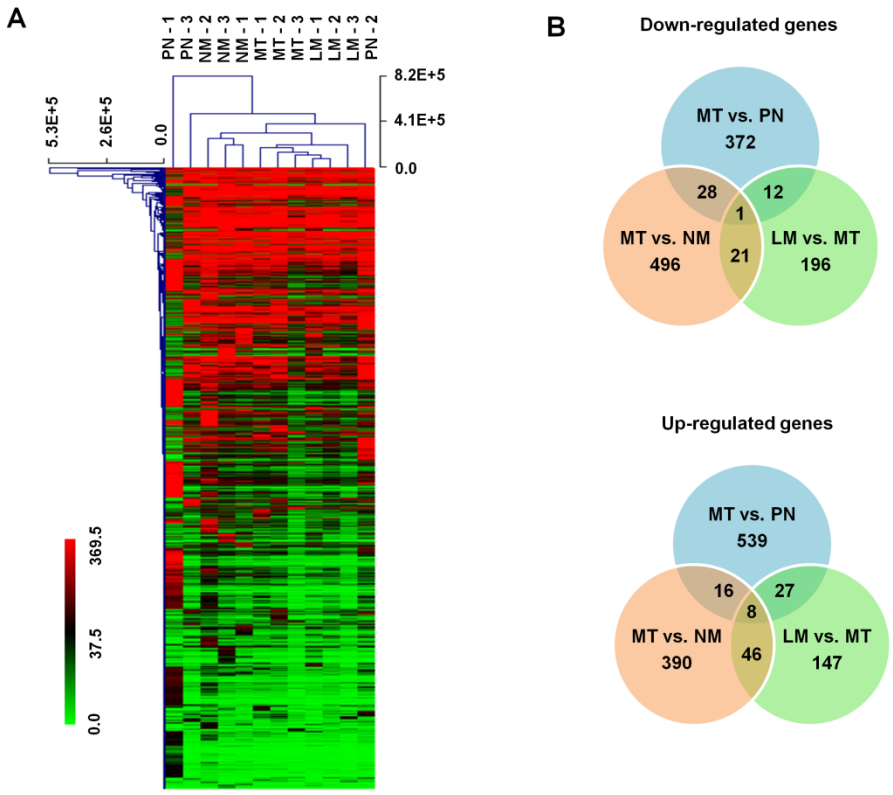

Figure S4. RNA sequencing and identification of differently expressed mRNAs. A, Hierarchical clustering for RNA-sequencing data of 12 samples. B, Significantly changed genes among three groups from RNA-sequencing data of 12 samples.

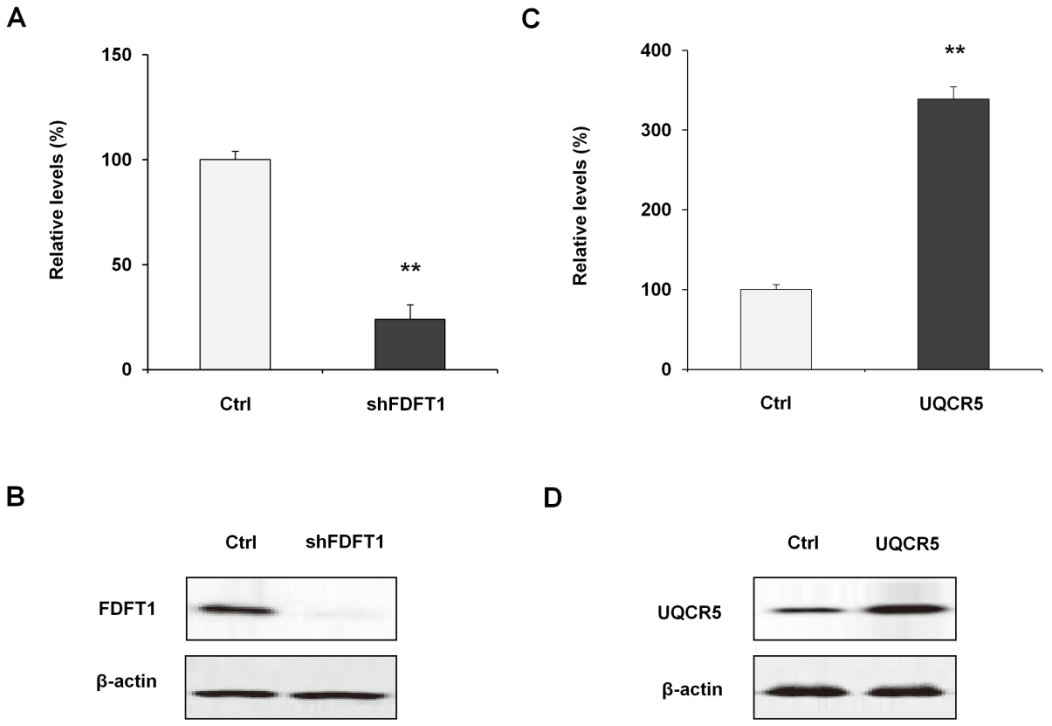

Figure S5. qRT-PCR and Western blot to quantify FDFT1 and UQCR5 levels in SW480 cells. A, qRT-PCR measurement of the levels of FDFT1 mRNA in SW480 cells treated with negative control or shFDFT1. A, Western blot measurement of the levels of FDFT1 protein in SW480 cells treated with negative control or shFDFT1. C, qRT-PCR measurement of the levels of UQCR5 mRNA in SW480 cells treated with negative control or UQCR5 overexpression plasmid. D, Western blot measurement of the levels of UQCR5 protein in SW480 cells treated with negative control or UQCR5 overexpression plasmid. Data shown are the means ± SD of three independent experiments. Statistical analyses were performed with one-way ANOVA (** = $P < 0.01$).

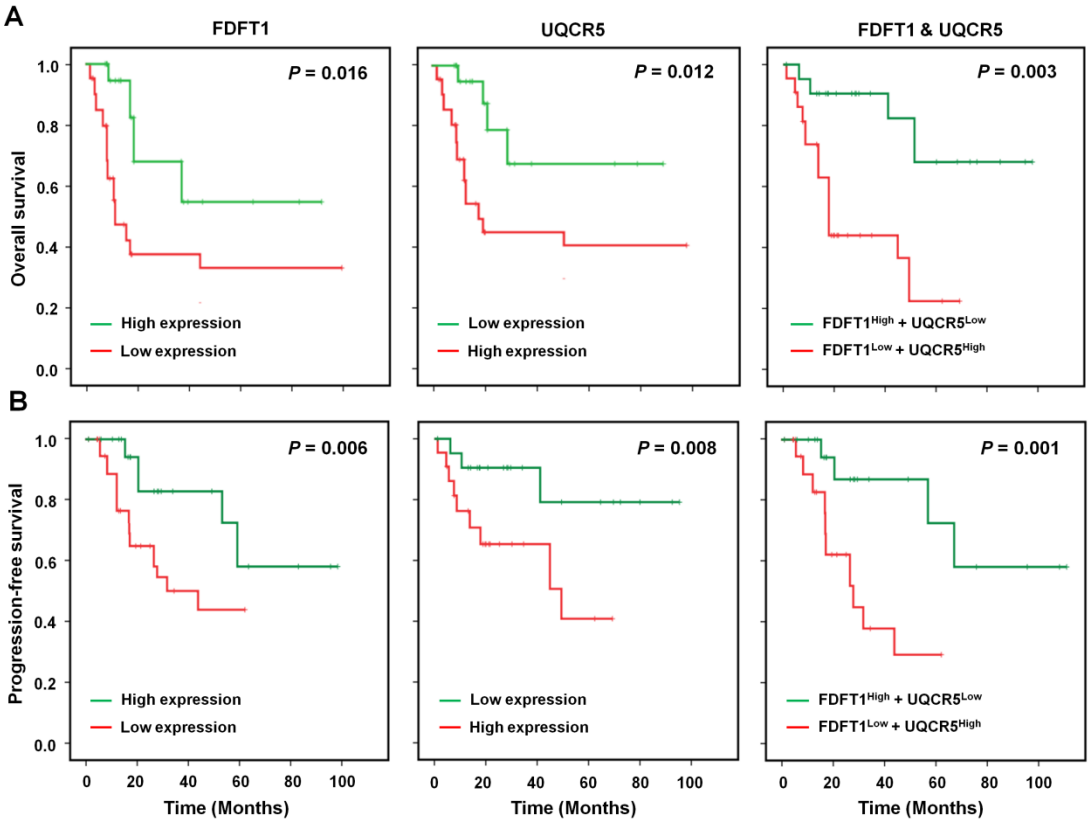

Figure S6. Clinical significance of FDFT1 and UQCR5 in 44 CRC. Kaplan-Meier survival analysis to evaluate the prognostic value of FDFT1 or/and UQCR5 expression for OS (A) and PFS (B) of 44 CRC patients.

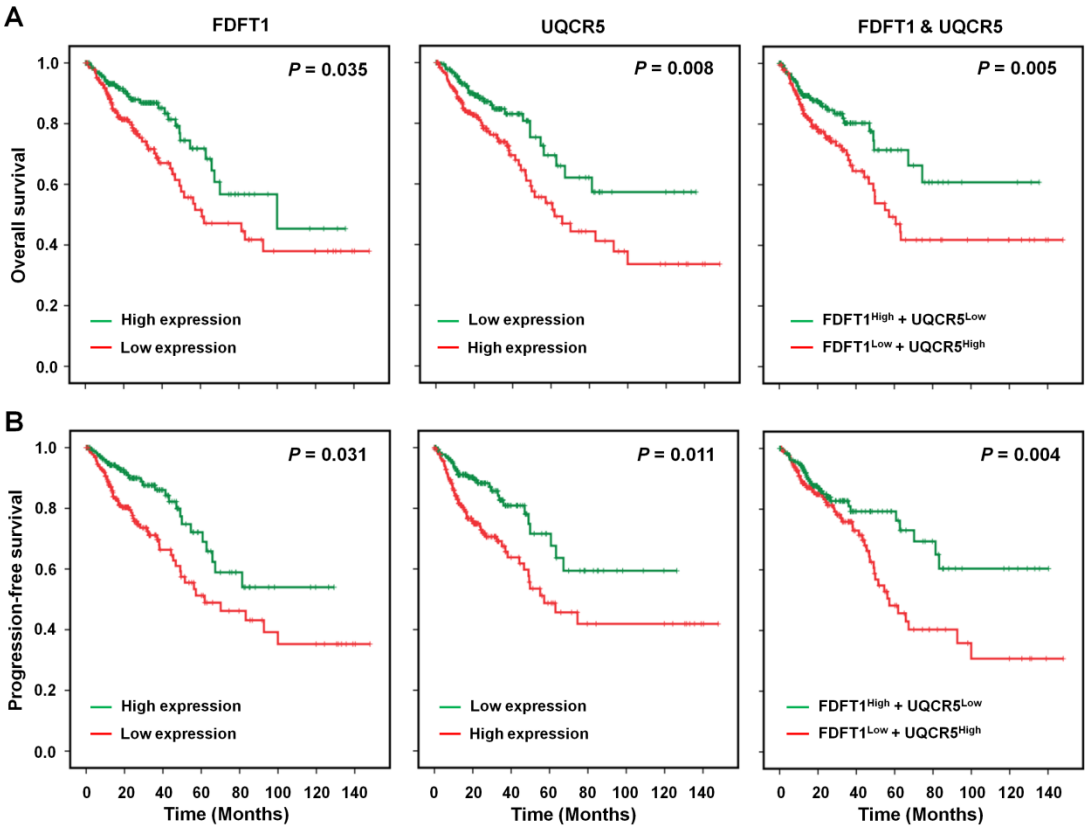

Figure S7. Clinical significance of FDFT1 and UQCR5 in 379 CRC patients from TCGA portal. Kaplan-Meier survival analysis was used to evaluate the prognostic value of FDFT1 and/or UQCR5 expression in 379 CRC patients from TCGA portal for OS (A) and PFS (B).

Table S1. Summary of CRC patients demographic and clinical characteristics.

Factor	Variables	Non-metastatic (N = 21)	Metastatic to liver (N = 23)
		Number (%)	Number (%)
Age	≥ 60	14 (66.7%)	11 (47.8%)
	< 60	7 (33.3%)	12 (52.2%)
Gender	Male	11 (52.4%)	10 (43.5%)
	Female	10 (47.6%)	13 (56.5%)
Primary site	Colon	9 (42.9%)	14 (60.9%)
	Rectum	12 (57.1%)	9 (39.1%)
Differentiation	Well	0 (0.0%)	0 (0.0%)
	Moderately	16 (76.2%)	16 (69.6%)
	Poorly	5 (23.8%)	7 (30.4%)
Completeness	R0	21 (100.0%)	23 (100.0%)
	R1	0 (0.0%)	0 (0.0%)
Diameter	≥ 5 cm	10 (47.6%)	10 (43.5%)
	< 5 cm	11 (52.4%)	13 (56.5%)
Number of primary foci	Multiple	8 (38.1%)	9 (39.1%)
	Single	13 (61.9%)	14 (60.9%)
TNM stage	I	10 (47.6%)	2 (8.7%)
	II	6 (28.6%)	3(13.0%)
	III	5 (23.8%)	12 (52.2%)
	IV	0 (0.0%)	6 (26.1%)

Necrosis	Yes	11 (52.4%)	10 (43.5%)
	No	10 (47.6%)	13 (56.5%)

Table S2. Prognosis analysis based on clinical characteristics of 374 CRC patients.

Factor	Variable	N	Overall survival			Progression free survival			Hazard Ratio		
			Months (Mean)	95% CI (Mean)	P value	Months (Mean)	95% CI (Mean)	P value	Hazard Ratio	95% CI (univariate)	P value
Age	≥ 60	245	30.89	26.32 - 34.21	0.825	30.26	26.89 - 34.54	0.779	1.21	1.03 - 1.61	0.438
	< 60	129	31.52	26.12 - 34.93		31.02	26.55 - 34.83				
Gender	Male	206	31.11	27.27 - 33.94	0.147	30.12	26.33 - 35.04	0.196	1.11	0.93 - 1.22	0.798
	Female	168	31.87	26.01 - 35.74		31.29	25.76 - 34.83				
Lymph node metastasis	Positive	169	27.34	25.43 - 29.58	<0.001	28.02	26.04 - 30.96	<0.001	2.31	1.49 - 3.61	0.021
	Negative	205	31.28	29.72 - 33.87		31.33	29.26 - 33.41				
TNM stage	III-IV	307	26.69	24.36 - 29.83	0.012	26.04	24.33 - 29.98	<0.001	2.52	1.94 - 4.21	<0.001
	I-II	67	31.58	26.19 - 33.48		31.96	26.87 - 34.56				
Lympho-vascular invasion	Positive	101	29.67	27.89 - 32.26	0.058	30.09	26.89 - 32.76	0.066	1.89	0.93 - 2.52	0.068
	Negative	228	31.03	26.74 - 34.52		31.53	26.24 - 33.75				
	Unknown	45	29.31	27.65 - 35.43		29.31	27.65 - 35.43				
Vascular invasion	Positive	76	28.39	26.45 - 30.33	0.045	27.95	24.36 - 30.85	0.038	2.59	1.69 - 3.97	<0.001
	Negative	247	31.04	28.71 - 34.31		30.83	26.77 - 33.61				
	Unknown	51	29.33	27.58 - 33.46		29.34	26.62 - 33.41				
Diameter	≥ 2 cm	168	26.86	24.39 - 29.01	< 0.001	26.56	24.21 - 30.04	< 0.001	2.69	1.88 - 4.05	<0.001
	< 2 cm	206	29.86	27.39 - 33.01		29.35	27.86 - 33.21				

Table S3. Association between deregulated genes expression and clinicopathological characteristics in 374 CRC patients.

			BGN		EHD2		UQCR5		COL1A2		FDFT1	
Factor	**Variable**	**N**	**Exp.**	**P val**	**Exp.**	**P val**	**Exp.**	**P val**	**Exp.**	**P val**	**Exp.**	**P val**
Age	≥ 60	245	7382.77 ± 57.51	0.059	988.15 ± 106.822	0.096	2306.19 ± 77.25	0.355	44266.64 ± 4471.07	0.143	4859.76 ± 278.89	0.691
	< 60	129	5913.18 ± 83.25		1158.61 ± 89.22		2218.22 ± 105.94		36716.91 ± 5477.49		4973.92 ± 273.91	
Gender	Male	206	6482.14 ± 685.82	0.593	996.25 ± 94.87	0.087	2312.88. ± 98.64	0.229	40475.24 ± 4459.14	0.967	4820.44 ± 283.99	0.449
	Female	168	6886.32 ± 776.09		1171.99 ± 104		2197.98 ± 82.15		40691.79 ± 5589.24		5037.77 ± 266.55	
Lymph node metastasis	Positive	169	8056.72 ± 672.64	0.005	1242.81 ± 96.23	0.009	2396.53 ± 98.44	0.005	47442.41 ± 4783.17	0.042	4249.89 ± 272.39	0.001
	Negative	205	5769.78 ± 804.92		953.42 ± 106.12		2100.17 ± 84.66		36080.96 ± 5487.47		5241.01 ± 246.66	
TNM stage	III-IV	307	7180.22 ± 700.48	0.002	1138.28 ± 104.19	0.008	2380.39 ± 81.34	0.033	43495.06 ± 5095.37	0.014	4748.09 ± 283.47	0.012
	I-II	67	4116.95 ± 754.57		784.39 ± 91.94		2141.84 ± 99.18		26943.58 ± 4842.12		5682.63 ± 267.95	
Lympho-	Positive	101	9164.55 ±	0.001	1387.15 ±	0.001	2319.57 ±	0.018	53918.21 ±	0.008	4117.85 ±	0.001

vascular invasion			752.61		92.55		82.19		4884.51	284.67	
	Negative	228	5989.06 ± 744.28		979.59 ± 112.35		2070.79 ± 103.63		37457.02 ± 5547.11	5236.09 ± 254.64	
	Unknown	46	7718.94 ± 955.53		1218.94 ± 95.53		2217.64 ± 95.88		39918.94 ± 9855.53	4418.94 ± 335.53	
Vascular invasion	Positive	76	8075.33 ± 685.82	0.048	1305.52 ± 94.87	0.037	2295.64 ± 82.15	0.039	49634.54 ± 4459.14	0.034	4264.56 ± 283.99	0.031
	Negative	247	6411.98 ± 776.09		1026.79 ± 104.01		2001.96 ± 98.64		40229.51 ± 5589.24	5022.76 ± 266.55	
	Unknown	57	6920.36 ± 855.45		1120.36 ± 95.45		2112.36 ± 95.46		45520.36 ± 5455.45	5020.36 ± 265.45	
Diameter	≥ 2 cm	168	8716.08 ± 558.84	0.001	1406.11 ± 71.39	0.001	2394.95 ± 83.58	0.022	50931.51 ± 3572.48	0.017	2802.45 ± 312.14	0.023
	< 2 cm	206	5776.66 ± 859.58		970.41 ± 117.26		2010.82 ± 214.38		37502.59 ± 5960.01	5200.36 ± 251.85	

Supplementary Table 4. SAAVs in NM-specific sample.

Protein Variants	Accession number	Protein_Change	Exon	RefSeq	Chrs	Position	cDNA_Change	SIFT Score	PolyPhenV2
ABCA4	P78363	p.W1479X	exon30	NM_000350	1	94495103	c.4437G>A	0	0.735
ACTR10	Q9NZ32	p.I258V	exon10	NM_018477	14	58690558	c.772A>G	1	0
ACTRT1	Q8TDG2	p.L70M	exon1	NM_138289	X	127185978	c.208T>A	0.01	0.004
ETNPPL	Q8TBG4	p.G118V	exon5	NM_001146627	4	109674142	c.353G>T	0	0.997
AKAP1	Q92667	p.S651N	exon4	NM_003488	17	55189262	c.1952G>A	0.95	0
ALDH18A1	P54886	p.Q4X	exon2	NM_001017423	10	97413125	c.10C>T	0.13	0.728
ANKRD11	Q6UB99	p.S56C	exon4	NM_013275	16	89371674	c.166A>T	0	0.053
ANO1	Q5XXA6	p.G154R	exon3	NM_018043	11	69949190	c.460G>A	0.37	
ARHGAP20	Q9P2F6	p.A995V	exon16	NM_020809	11	110450686	c.2984C>T	0.26	0.001
ARHGAP36	Q6ZRI8	p.T126P	exon4	NM_144967	X	130217764	c.376A>C	0.02	0.899
ARMC9	Q7Z3E5	p.R405C	exon14	NM_025139	2	232137670	c.1213C>T	0	1
BAI1	O14514	p.A214T	exon1	NM_001702	8	143546199	c.640G>A	1	
BMP10	O95393	p.E420K	exon2	NM_014482	2	69092780	c.1258G>A	0.07	0.206
C16orf70	Q9BSU1	p.G113A	exon5	NM_025187	16	67166411	c.338G>C	0	0.998
CCDC189	A1A4V9	p.C145F	exon4	NM_001014979	16	30771640	c.434G>T	0.98	0.999
TMEM249	Q2WGJ8	p.R206H	exon5	NM_001252402	8	145577099	c.617G>A	0	
CABIN1	Q9Y6J0	p.G75S	exon5	NM_001199281	22	24437599	c.223G>A	0.25	0.405
CACNA1H	O95180	p.T1615P	exon27	NM_021098	16	1263845	c.4843A>C	0	
CCDC120	Q96HB5	p.A559G	exon10	NM_001163323	X	48925467	c.1676C>G	0.41	0
CCDC39	Q9UFE4	p.E81K	exon3	NM_181426	3	180379765	c.241G>A	0	
CCDC39	Q9UFE4	p.Q660H	exon14	NM_181426	3	180349275	c.1980G>T	0	
CCNK	O75909	p.H406P	exon11	NM_001099402	14	99976593	c.1217A>C	0.09	
CD109	Q6YHK3	p.K556X	exon14	NM_133493	6	74477946	c.1666A>T	0.13	0.693
CD3G	P09693	p.L110R	exon4	NM_000073	11	118221288	c.329T>G	0	0.992
CETP	P11597	p.D487N	exon16	NM_000078	16	57017555	c.1459G>A	0.18	0.726
CHST6	Q9GZX3	p.R155Q	exon3	NM_021615	16	75513263	c.464G>A	0.83	0.027

CLSPN	Q9HAW4	p.D64E	exon3	NM_001190481	1	36230257	c.192C>A	0	0.187
CNOT4	O95628	p.R46H	exon2	NM_001190850	7	135122943	c.137G>A	0	0.999
COL12A1	Q99715	p.P1565H	exon26	NM_004370	6	75853101	c.4694C>A		0.998
COL8A1	P27658	p.G659S	exon4	NM_020351	3	99514720	c.1975G>A	0	0.999
CPN2	P22792	p.D415Y	exon2	NM_001080513	3	194062189	c.1243G>T	0.01	0.989
CPZ	Q66K79	p.R219H	exon3	NM_003652	4	8605895	c.656G>A	0.09	0.951
CREBBP	Q92793	p.D1665N	exon30	NM_004380	16	3781372	c.4993G>A	0.01	0.783
CST7	O76096	p.C110Y	exon3	NM_003650	20	24939649	c.329G>A	0	0.727
CUX2	O14529	p.A863T	exon17	NM_015267	12	111758400	c.2587G>A	0.2	0
DDX31	Q9H8H2	p.S303I	exon7	NM_022779	9	135527875	c.908G>T	0.02	0.89
DGKB	Q9Y6T7	p.S172P	exon6	NM_145695	7	14741308	c.514T>C	0.01	
EFNB1	P98172	p.R154C	exon3	NM_004429	X	68059560	c.460C>T	0.05	0.717
EML5	Q05BV3	p.D1295N	exon27	NM_183387	14	89123841	c.3883G>A	0	
ERI2	A8K979	p.P623L	exon9	NM_001142725	16	20809254	c.1868C>T	0	
FAM69C	Q0P6D2	p.R111W	exon2	NM_001044369	18	72114386	c.331C>T	0.05	
FANCG	O15287	p.S383F	exon10	NM_004629	9	35075747	c.1148C>T	0.03	0
FYCO1	Q9BQS8	p.G973S	exon8	NM_024513	3	46007909	c.2917G>A	0.75	0
GCM2	O75603	p.R143W	exon3	NM_004752	6	10876707	c.427C>T	0	1
GLI2	P10070	p.P906L	exon13	NM_005270	2	121746207	c.2717C>T	0	0.998
GZMA	P12544	p.I191V	exon4	NM_006144	5	54404166	c.571A>G	0	0.149
HIF3A	Q9Y2N7	p.E582K	exon13	NM_152795	19	46834444	c.1744G>A	0.28	0
HMCN1	Q96RW7	p.R3793H	exon74	NM_031935	1	186084052	c.11378G>A	0.01	0.999
HUWE1	Q7Z6Z7	p.A2634V	exon57	NM_031407	X	53586329	c.7901C>T	0.07	0.125
IGSF3	O75054	p.V489M	exon7	NM_001542	1	117146465	c.1465G>A	0.01	0.003
IKZF4	Q9H2S9	p.R489L	exon8	NM_022465	12	56428823	c.1466G>T	0.01	
KAT6A	Q92794	p.Q1335X	exon18	NM_001099413	8	41791735	c.4003C>T	0.48	0.735
KCNF1	Q9H3M0	p.R299C	exon1	NM_002236	2	11053447	c.895C>T	0	1
LY86	O95711	p.G37S	exon1	NM_004271	6	6589076	c.109G>A	0.34	0.081
MAGEC1	O60732	p.W717L	exon4	NM_005462	X	140995340	c.2150G>T	0	0.489
MGA	Q8IWI9	p.T2716M	exon24	NM_001164273	15	42058427	c.8147C>T	0	

MUC12	Q9UKN1	p.D868N	exon2	NM_001164462	7	100636446	c.2602G>A		
MUC12	Q9UKN1	p.D868G	exon2	NM_001164462	7	100636447	c.2603A>G		
NDST1	P52848	p.L182F	exon3	NM_001543	5	149907396	c.544C>T	0.03	0.02
NID1	P14543	p.D171Y	exon2	NM_002508	1	236212004	c.511G>T	0.04	0.001
NPDC1	Q9NQX5	p.A232V	exon6	NM_015392	9	139934979	c.695C>T	0.51	0.253
NRK	Q7Z2Y5	p.Q554K	exon13	NM_198465	X	105153293	c.1660C>A	0.03	
NRK	Q7Z2Y5	p.N1282S	exon23	NM_198465	X	105183911	c.3845A>G	0.08	
NWD1	Q149M9	p.V1279I	exon18	NM_001007525	19	16918495	c.3835G>A	0.87	0.085
OR1N2	Q8NGR9	p.M175I	exon1	NM_001004457	9	125315973	c.525G>A	0.49	0.291
PABPC1	P11940	p.L126V	exon2	NM_002568	8	101730326	c.376C>G	0	0.51
PABPC1	P11940	p.G123C	exon2	NM_002568	8	101730335	c.367G>T	0	0.998
PABPC1	P11940	p.R419C	exon9	NM_002568	8	101721442	c.1255C>T	0.01	0
PABPC1	P11940	p.T416S	exon9	NM_002568	8	101721451	c.1246A>T	0.46	0
PAPLN	O95428	p.C626Y	exon15	NM_173462	14	73726226	c.1877G>A	0	0.999
PHF20	Q9BVI0	p.R882C	exon16	NM_016436	20	34526962	c.2644C>T	0.01	0.992
PQLC1	Q8N2U9	p.T112I	exon3	NM_001146345	18	77703331	c.335C>T	0.11	0
PRSS3	P35030	p.F150S	exon3	NM_002771	9	33798075	c.449T>C	0.45	0
PRSS3	P35030	p.W222R	exon5	NM_002771	9	33799098	c.664T>C	0.22	0
PRSS3	P35030	p.W222X	exon5	NM_002771	9	33799099	c.665G>A	0.19	0.589
PTPN21	Q16825	p.Y569H	exon13	NM_007039	14	88946070	c.1705T>C	0	0.997
NECTIN4	Q96NY8	p.I398V	exon7	NM_030916	1	161043551	c.1192A>G	0.13	0.74
PXDNL	A1KZ92	p.D828H	exon17	NM_144651	8	52321702	c.2482G>C	0	
QKI	Q96PU8	p.T317M	exon7	NM_206853	6	163986993	c.950C>T	0.01	0.535
RFC4	P35249	p.K63Q	exon3	NM_002916	3	186518929	c.187A>C	0.1	0.001
RILPL1	Q5EBL4	p.R293W	exon5	NM_178314	12	123970277	c.877C>T	0	
RIMS1	Q86UR5	p.E23K	exon1	NM_001168407	6	72922892	c.67G>A		
RPA4	Q13156	p.S22G	exon1	NM_013347	X	96139373	c.64A>G	0.17	0.042
SAFB2	Q14151	p.T950P	exon21	NM_014649	19	5587268	c.2848A>C	0	0.557
SBF1	O95248	p.A5P	exon1	NM_002972	22	50913257	c.13G>C	0	0.957
SDR9C7	Q8NEX9	p.R276H	exon4	NM_148897	12	57317732	c.827G>A	0.02	0.716

SELPLG	Q14242	p.A131V	exon2	NM_003006	12	109017692	c.392C>T	0.29	
SELPLG	Q14242	p.A131T	exon2	NM_003006	12	109017693	c.391G>A	0.3	
SLC11A1	P49279	p.S511N	exon14	NM_000578	2	219259498	c.1532G>A	0.18	0
SLC2A8	Q9NY64	p.A81T	exon3	NM_014580	9	130160205	c.241G>A	0.07	0.988
SLC30A3	Q99726	p.P216L	exon5	NM_003459	2	27480152	c.647C>T	0.15	0.049
SLC8A3	P57103	p.R488H	exon2	NM_033262	14	70633677	c.1463G>A	0	0.925
SPAG8	Q99932	p.G472R	exon7	NM_001039592	9	35809979	c.1414G>A	0	0.394
SPATC1	Q76KD6	p.A146V	exon2	NM_001134374	8	145095035	c.437C>T	0.02	0.205
SVIL	O95425	p.G532D	exon13	NM_003174	10	29812670	c.1595G>A	0.22	0.001
TBR1	Q16650	p.A547S	exon6	NM_006593	2	162280328	c.1639G>T	1	0.002
TLL2	Q9Y6L7	p.A302S	exon7	NM_012465	10	98180732	c.904G>T	0.04	0.892
TMEM114	B3SHH9	p.S97L	exon2	NM_001146336	16	8619856	c.290C>T		
TMEM132C	Q8N3T6	p.A821T	exon9	NM_001136103	12	129189974	c.2461G>A	0.45	
TP53	P04637	p.V172D	exon5	NM_001126114	17	7578415	c.515T>A	0	0.998
UNC13C	Q8NB66	p.R182X	exon1	NM_001080534	15	54305644	c.544C>T	1	
USP6	P35125	p.I67M	exon5	NM_004505	17	5036210	c.201T>G	0.22	0.369
VEGF	P15692	p.F183L	exon1	NM_001025369	6	43738992	c.549T>G	0.02	0.991
VGF	O15240	p.A325G	exon2	NM_003378	7	100807151	c.974C>G	0	0.512
ZNF331	Q9NQX6	p.D10N	exon4	NM_001253801	19	54074876	c.28G>A	0	0.547
ZNF438	Q7Z4V0	p.Q594X	exon7	NM_001143771	10	31137524	c.1780C>T	0.07	0.735
ZNF717	Q9BY31	p.Y758F	exon5	NM_001128223	3	75786501	c.2273A>T	0.4	
ZNF717	Q9BY31	p.H386Q	exon5	NM_001128223	3	75787616	c.1158C>G	1	
ZXDA	P98168	p.S136T	exon1	NM_007156	X	57936449	c.406T>A	0.6	0
ZXDA	P98168	p.C135S	exon1	NM_007156	X	57936451	c.404G>C	0.1	0.251

Supplementary Table 5. SAAVs in NM & MT-share sample.

Protein Variants	Accession number	Protein_Change	Exon	RefSeq	Chrs	Position	cDNA_Change	SIFT Score	PolyPhenV2
A1BG	P04217	p.T278P	exon5	NM_130786	19	58862835	c.832A>C	0	0.735
ADGRG6	Q86SQ4	p.W789C	exon17	NM_001032394	6	1.43E+08	c.2367G>C	1	0
AGTR2	P50052	p.R182X	exon3	NM_000686	X	1.15E+08	c.544C>T	0.01	0.004
AKAP13	Q12802	p.T2795P	exon36	NM_006738	15	86287035	c.8383A>C	0	0.997
ANAPC1	Q9H1A4	p.Q451H	exon11	NM_022662	2	1.13E+08	c.1353G>C	0.95	0
APC2	O95996	p.F2256S	exon15	NM_005883	19	1470067	c.6767T>C	0.13	0.728
AQP7	O14520	p.E202D	exon7	NM_001170	9	33385784	c.606G>C	0	0.053
ASB7	Q9H672	p.H91R	exon5	NM_024708	15	1.01E+08	c.272A>G	0.37	
ASL	P04424	p.E214D	exon8	NM_001024944	7	65552360	c.642G>T	0.26	0.001
ATP8A2	Q9NTI2	p.L624F	exon22	NM_016529	13	26153950	c.1872G>T	0.02	0.899
ATRNL1	Q5VV63	p.C961G	exon18	NM_207303	10	1.17E+08	c.2881T>G	0	1
BRD8	Q9H0E9	p.G298C	exon10	NM_006696	5	1.38E+08	c.892G>T	1	
BTBD16	Q32M84	p.T297N	exon10	NM_144587	10	1.24E+08	c.890C>A	0.07	0.206
C20orf96	Q9NUD7	p.P8H	exon2	NM_153269	20	270946	c.23C>A	0	0.998
C22orf39	Q6P5X5	p.R25G	exon1	NM_001166242	22	19435250	c.73C>G	0.98	0.999
C2CD2	Q9Y426	p.A79G	exon1	NM_015500	21	43373522	c.236C>G	0	
C2CD3	Q4AC94	p.Y1745D	exon27	NM_015531	11	73760510	c.5233T>G	0.25	0.405
CACNA1H	O95180	p.T1615P	exon27	NM_021098	16	1263845	c.4843A>C	0	
CAPN12	Q6ZSI9	p.R413G	exon10	NM_144691	19	39227921	c.1237C>G	0.41	0
CASK	O14936	p.G306C	exon10	NM_001126054	X	41485956	c.916G>T	0	
CCDC180	Q9P1Z9	p.L536M	exon14	NM_020893	9	1E+08	c.1606C>A	0	
CCDC60	Q8IWA6	p.D33E	exon2	NM_178499	12	1.2E+08	c.99C>A	0.09	
CENPV	Q7Z7K6	p.S35G	exon1	NM_181716	17	16256648	c.103A>G	0.13	0.693
CFAP91	Q7Z4T9	p.Q533K	exon13	NM_033364	3	1.19E+08	c.1597C>A	0	0.992
CFH	P08603	p.P346T	exon8	NM_000186	1	1.97E+08	c.1036C>A	0.18	0.726

CGNL1	Q0VF96	p.E898D	exon11	NM_001252335	15	57810674	c.2694G>T	0.83	0.027
CHIT1	Q13231	p.A359G	exon10	NM_003465	1	2.03E+08	c.1076C>G	0	0.187
CHIT1	Q13231	p.W358X	exon10	NM_003465	1	2.03E+08	c.1073G>A	0	0.999
CLIP1	P30622	p.L271F	exon5	NM_198240	12	1.23E+08	c.813G>T		0.998
CMTR1	Q8N1G2	p.R672Q	exon19	NM_015050	6	37443167	c.2015G>A	0	0.999
CNN3	Q15417	p.L65Q	exon3	NM_001839	1	95368730	c.194T>A	0.01	0.989
COL15A1	P39059	p.P653T	exon15	NM_001855	9	1.02E+08	c.1957C>A	0.09	0.951
COL7A1	Q02388	p.S2564I	exon103	NM_000094	3	48606284	c.7691G>T	0.01	0.783
CRY2	Q96524	p.P508T	exon11	NM_001127457	11	45893705	c.1522C>A	0	0.727
CUL1	Q13616	p.A467S	exon13	NM_003592	7	1.48E+08	c.1399G>T	0.2	0
CYP2J2	P51589	p.W322C	exon6	NM_000775	1	60373495	c.966G>T	0.02	0.89
CYP4F2	P78329	p.T472A	exon13	NM_001082	19	15989730	c.1414A>G	0.01	
DBF4	Q9UBU7	p.E77D	exon3	NM_006716	7	87514305	c.231A>C	0.05	0.717
DMAP1	Q9NPF5	p.H362N	exon10	NM_001034023	1	44685721	c.1084C>A	0	
DOCK8	Q8NF50	p.Y1969X	exon45	NM_001190458	9	463655	c.5907C>A	0	
DPP9	Q86TI2	p.T144A	exon6	NM_139159	19	4704313	c.430A>G	0.05	
DPY19L1	Q2PZI1	p.P254H	exon9	NM_015283	7	35009079	c.761C>A	0.03	0
DSPP	Q9NZW4	p.N934D	exon5	NM_014208	4	88536614	c.2800A>G	0.75	0
DYNC1H1	Q14204	p.D937E	exon10	NM_001376	14	1.02E+08	c.2811C>A	0	1
EP400	Q96L91	p.H31P	exon2	NM_015409	12	1.32E+08	c.92A>C	0	0.998
EPPK1	P58107	p.R2239H	exon1	NM_031308	8	1.45E+08	c.6716G>A	0	0.149
ERN2	Q76MJ5	p.T202P	exon6	NM_033266	16	23718102	c.604A>C	0.28	0
FAM104B	Q5XKR9	p.S51N	exon3	NM_001166700	X	55172716	c.152G>A	0.01	0.999
FAM71D	Q8N9W8	p.A113E	exon4	NM_173526	14	67669989	c.338C>A	0.07	0.125
FERMT3	Q86UX7	p.V45G	exon2	NM_178443	11	63974970	c.134T>G	0.01	0.003
FDFT1	P37268	p.R110W	exon5	NM_004462	8	11809797	c.328C>T	0.01	
FDFT1	P37268	p.Y174C	exon7	NM_004462	8	11826034	c.521A>G	0.48	0.735
FDFT1	P37268	p.N205S	exon7	NM_004462	8	11826127	c.614A>G	0	1
FDFT1	P37268	p.N215S	exon7	NM_004462	8	11826157	c.644A>G	0.34	0.081
FDFT1	P37268	p.R317W	exon9	NM_004462	8	11831587	c.949C>T	0	0.489

FDFT1	P37268	p.I345M	exon10	NM_004462	8	11838390	c.1035T>G	0	
FLG	P20930	p.T2235R	exon3	NM_002016	1	1.52E+08	c.6704C>G		
FOXK1	P85037	p.T683P	exon9	NM_001037165	7	4801940	c.2047A>C		
FSTL1	Q12841	p.A7P	exon2	NM_007085	3	1.2E+08	c.19G>C	0.03	0.02
GAS8	O95995	p.S57G	exon1	NM_001214	16	90095582	c.169A>G	0.04	0.001
GAS8	O95995	p.V44A	exon1	NM_001214	16	90095620	c.131T>C	0.51	0.253
GBP7	Q8N8V2	p.T482S	exon9	NM_207398	1	89607253	c.1444A>T	0.03	
GCKR	Q14397	p.V102F	exon4	NM_001486	2	27721140	c.304G>T	0.08	
GIGYF2	Q6Y7W6	p.P1211Q	exon26	NM_001103148	2	2.34E+08	c.3632C>A	0.87	0.085
GPER	Q99527	p.E218Q	exon2	NM_001505	7	1132016	c.652G>C	0.49	0.291
GSN	P06396	p.S511R	exon14	NM_001127664	9	1.24E+08	c.1533C>A	0	0.51
HERC2	O95714	p.F544L	exon13	NM_004667	15	28511089	c.1630T>C	0	0.998
HLA-A	P30443	p.A48S	exon2	NM_002116	6	29910602	c.142G>T	0.01	0
HLA-DRB1	P01911	p.H87L	exon2	NM_001243965	6	32551970	c.260A>T	0.46	0
HMCN1	Q96RW7	p.Y2941D	exon57	NM_031935	1	1.86E+08	c.8821T>G	0	0.999
HMX1	P32339	p.A269P	exon2	NM_018942	4	8869661	c.805G>C	0.01	0.992
HRCT1	Q6UXD1	p.R108H	exon1	NM_001039792	9	35906607	c.323G>A	0.11	0
IFNA7	P01567	p.V11L	exon1	NM_021057	9	21202134	c.31G>T	0.45	0
IFNK	Q9P0W0	p.V154F	exon1	NM_020124	9	27524794	c.460G>T	0.22	0
IGSF3	O75054	p.E446D	exon7	NM_001542	1	1.17E+08	c.1338G>C	0.19	0.589
IGSF3	O75054	p.E434G	exon7	NM_001542	1	1.17E+08	c.1301A>G	0	0.997
IGSF3	O75054	p.R680Q	exon8	NM_001542	1	1.17E+08	c.2039G>A	0.13	0.74
IGSF9	Q9P2J2	p.C395X	exon10	NM_001135050	1	1.6E+08	c.1185C>A	0	
IQSEC2	Q5JU85	p.S1444P	exon15	NM_001111125	X	53263538	c.4330T>C	0.01	0.535
ITGB4	P16144	p.Q478H	exon12	NM_001005731	17	73728300	c.1434G>T	0.1	0.001
ITPR1	Q14643	p.G1159E	exon29	NM_001099952	3	4725969	c.3476G>A	0	
KCNIP4	Q6PIL6	p.A5T	exon1	NM_025221	4	20884321	c.13G>A		
KIAA0226	RUBCN	p.A943G	exon20	NM_014687	3	1.97E+08	c.2828C>G	0.17	0.042
KIF18A	Q8NI77	p.R227H	exon5	NM_031217	11	28112183	c.680G>A	0	0.557
KMT2C	Q8NEZ4	p.R284Q	exon7	NM_170606	7	1.52E+08	c.851G>A	0	0.957

KMT2D	O14686	p.K2185N	exon28	NM_014727	19	36224005	c.6555A>C	0.02	0.716
KNL1	Q8NG31	p.S25X	exon3	NM_170589	15	40897346	c.74C>A	0.29	
KRT4	P19013	p.G87A	exon1	NM_002272	12	53207583	c.260G>C	0.3	
KRT8	P05787	p.M406I	exon8	NM_001256293	12	53292288	c.1218G>T	0.18	0
KRTAP4-3	Q9BYR4	p.S103T	exon1	NM_033187	17	39324117	c.308G>C	0.07	0.988
KRTAP4-8	Q9BYQ9	p.S68R	exon1	NM_031960	17	39254133	c.204C>A	0.15	0.049
KRTAP5-1	Q6L8H4	p.A166G	exon1	NM_001005922	11	1605983	c.497C>G	0	0.925
LILRA6	Q6PI73	p.Y350X	exon6	NM_024318	19	54744358	c.1050T>G	0	0.394
LILRB3	O75022	p.Q53L	exon3	NM_006864	19	54726347	c.158A>T	0.02	0.205
LILRB3	O75022	p.T6A	exon1	NM_006864	19	54726833	c.16A>G	0.22	0.001
MAGEB6	Q8N7X4	p.Y124S	exon2	NM_173523	X	26212334	c.371A>C	1	0.002
MAGEC1	O60732	p.S185R	exon4	NM_005462	X	1.41E+08	c.555T>A	0.04	0.892
MAGEC1	O60732	p.P189T	exon4	NM_005462	X	1.41E+08	c.565C>A		
MAP1B	P46821	p.D591E	exon5	NM_005909	5	71490955	c.1773C>A	0.45	
MAP2K7	O14733	p.R276W	exon7	NM_145185	19	7976015	c.826C>T	0	0.998
MATR3	P43243	p.S807X	exon14	NM_001194955	5	1.39E+08	c.2420C>A	1	
MEGF10	Q96KG7	p.G795S	exon19	NM_001256545	5	1.27E+08	c.2383G>A	0.22	0.369
MFSD14B	Q5SR56	p.F79V	exon3	NM_032558	9	97191501	c.235T>G	0.02	0.991
MTCL1	Q9Y4B5	p.N958K	exon13	NM_015210	18	8818975	c.2874C>A	0	0.512
MTHFR	P42898	p.L150M	exon3	NM_005957	1	11861245	c.448C>A	0	0.547
MTMR9	Q96QG7	p.R77X	exon2	NM_015458	8	11152749	c.229C>T	0.07	0.735
MUC16	Q8WXI7	p.K13414E	exon51	NM_024690	19	9002576	c.40240A>G	0.4	
MUC4	Q99102	p.H3709Q	exon2	NM_018406	3	1.96E+08	c.11127C>G	1	
MUC4	Q99102	p.T3710A	exon2	NM_018406	3	1.96E+08	c.11128A>G	0.6	0
MUC6	Q6W4X9	p.T1987P	exon31	NM_005961	11	1016842	c.5959A>C	0.1	0.251
MUC6	Q6W4X9	p.G1581E	exon31	NM_005961	11	1018059	c.4742G>A	0.04	0.001
UQCR5	P47985	p.R92H	exon2	NM_005961	13	29208098	c.275G>A	0	0.735
UQCR5	P47985	p.R177C	exon2	NM_005961	13	29207844	c.529C>T	1	0
NAF1	Q96HR8	p.D158Y	exon2	NM_138386	4	1.64E+08	c.472G>T	0.01	0.004
NFKB1	P19838	p.R186W	exon7	NM_001165412	4	1.03E+08	c.556C>T	0	0.997

NFKB1	P19838	p.E206X	exon8	NM_001165412	4	1.04E+08	c.616G>T	0.95	0
NKX2-8	O15522	p.V89G	exon2	NM_014360	14	37050561	c.266T>G	0.13	0.728
NOC4L	Q9BVI4	p.D292Y	exon9	NM_024078	12	1.33E+08	c.874G>T	0	0.053
NOTCH2NL	Q7Z3S9	p.T158I	exon4	NM_203458	1	1.45E+08	c.473C>T	0.37	
NPEPPS	P55786	p.S653I	exon17	NM_006310	17	45682781	c.1958G>T	0.26	0.001
OR2L3	Q8NG85	p.G196C	exon1	NM_001004687	1	2.48E+08	c.586G>T	0.02	0.899
OR2T29	Q8NH02	p.S61T	exon1	NM_001004694	1	2.49E+08	c.182G>C	0	1
UQCR5	P47985	p.R177H	exon2	NM_006003	13	29207843	c.530G>A	1	
UQCR5	P47985	p.D201H	exon2	NM_006003	13	29207772	c.601G>C	0.07	0.206
OR8U1	Q8NH10	p.H229R	exon1	NM_001005204	11	56143785	c.686A>G	0	0.998
OTOP1	Q7RTM1	p.K56R	exon1	NM_177998	4	4228425	c.167A>G	0.98	0.999
PABPC1	P11940	p.T48P	exon1	NM_002568	8	1.02E+08	c.142A>C	0	
PAPLN	O95428	p.G620V	exon15	NM_173462	14	73726208	c.1859G>T	0.25	0.405
PHF21A	Q96BD5	p.S324R	exon9	NM_001101802	11	45986887	c.972C>A	0	
PIEZO2	Q9H5I5	p.S2546L	exon49	NM_022068	18	10677849	c.7637C>T	0.41	0
PLEKHA6	Q9Y2H5	p.A801D	exon17	NM_014935	1	2.04E+08	c.2402C>A	0	
PLIN4	Q96Q06	p.V661A	exon3	NM_001080400	19	4511948	c.1982T>C	0	
PLXNB3	Q9ULL4	p.A1010S	exon19	NM_001163257	X	1.53E+08	c.3028G>T	0.09	
POLR2E	P19388	p.V209G	exon7	NM_002695	19	1089492	c.626T>G	0.13	0.693
POLR3F	Q9H1D9	p.E315Q	exon9	NM_006466	20	18464194	c.943G>C	0	0.992
POT1	Q9NUX5	p.V52G	exon8	NM_001042594	7	1.24E+08	c.155T>G	0.18	0.726
POU4F2	Q12837	p.A168V	exon2	NM_004575	4	1.48E+08	c.503C>T	0.83	0.027
PRG4	Q92954	p.K482E	exon5	NM_001127709	1	1.86E+08	c.1444A>G	0	0.187
PRG4	Q92954	p.T367A	exon5	NM_001127709	1	1.86E+08	c.1099A>G	0	0.999
PRR21	Q8WXC7	p.G58S	exon1	NM_001080835	2	2.41E+08	c.172G>A		0.998
PRSS3	P35030	p.G208R	exon5	NM_002771	9	33799056	c.622G>A	0	0.999
PTAR1	Q7Z6K3	p.E110X	exon4	NM_001099666	9	72349166	c.328G>T	0.01	0.989
PTPN21	Q16825	p.Y569S	exon13	NM_007039	14	88946069	c.1706A>C	0.09	0.951
RECQL4	O94761	p.S280N	exon5	NM_004260	8	1.46E+08	c.839G>A	0.01	0.783
ROBO1	Q9Y6N7	p.P683H	exon15	NM_002941	3	78711183	c.2048C>A	0	0.727

RTKN2	Q8IZC4	p.P413Q	exon11	NM_145307	10	63959569	c.1238C>A	0.2	0
SCRIB	Q14160	p.R733L	exon17	NM_015356	8	1.45E+08	c.2198G>T	0.02	0.89
SDK1	Q7Z5N4	p.T256P	exon11	NM_001079653	7	4247821	c.766A>C	0.01	
SELENOT	P62341	p.G121V	exon3	NM_016275	3	1.5E+08	c.362G>T	0.05	0.717
SLC16A14	Q7RTX9	p.L92F	exon3	NM_152527	2	2.31E+08	c.276G>T	0	
SLC2A3	P11169	p.E482G	exon10	NM_006931	12	8074055	c.1445A>G	0	
SLC7A7	Q9UM01	p.S386R	exon9	NM_001126106	14	23243650	c.1158C>A	0.05	
SLC9A3R1	O14745	p.V60G	exon1	NM_004252	17	72745164	c.179T>G	0.03	0
SMPD4	Q9NXE4	p.F484V	exon15	NM_017951	2	1.31E+08	c.1450T>G	0.75	0
SNED1	Q8TER0	p.T1241P	exon25	NM_001080437	2	2.42E+08	c.3721A>C	0	1
SNRNP40	Q96DI7	p.A57S	exon2	NM_004814	1	31766168	c.169G>T	0	0.998
SOGA3	Q5TF21	p.E212G	exon2	NM_001012279	6	1.28E+08	c.635A>G	0	0.149
SORCS2	Q96PQ0	p.L181M	exon2	NM_020777	4	7398075	c.541C>A	0.28	0
SPEG	Q15772	p.T3254P	exon41	NM_005876	2	2.2E+08	c.9760A>C	0.01	0.999
SPPL3	Q8TCT6	p.R182P	exon7	NM_139015	12	1.21E+08	c.545G>C	0.07	0.125
SSC5D	A1L4H1	p.Y1312H	exon14	NM_001144950	19	56029577	c.3934T>C	0.01	0.003
SSPO	A2VEC9	p.G3700R	exon78	NM_198455	7	1.5E+08	c.11098G>C	0.01	
STK4	Q13043	p.P19S	exon1	NM_020998	3	49726070	c.55C>T	0.48	0.735
SVEP1	Q4LDE5	p.A634D	exon9	NM_153366	9	1.13E+08	c.1901C>A	0	1
TAS2R31	P59538	p.E151Q	exon1	NM_176885	12	11183484	c.451G>C	0.34	0.081
TBC1D22B	Q9NU19	p.S224X	exon5	NM_017772	6	37250727	c.671C>A	0	0.489
TBX21	Q9UL17	p.T180P	exon2	NM_013351	17	45820022	c.538A>C	0	
TIGAR	Q9NQ88	p.L217F	exon6	NM_020375	12	4461695	c.651A>C		
TLL2	Q9Y6L7	p.C588G	exon14	NM_012465	10	98146800	c.1762T>G		
TMCO3	Q6UWJ1	p.S109Y	exon2	NM_017905	13	1.14E+08	c.326C>A	0.03	0.02
TMEM128	Q5BJH2	p.V134I	exon4	NM_032927	4	4239589	c.400G>A	0.04	0.001
TMEM259	Q4ZIN3	p.A387P	exon9	NM_001033026	19	1011424	c.1159G>C	0.51	0.253
TMEM268	Q5VZI3	p.Q270K	exon8	NM_153045	9	1.17E+08	c.808C>A	0.03	
TMTC1	Q8IUR5	p.Y41X	exon1	NM_001193451	12	29936562	c.123C>A	0.08	
TMTC2	Q8N394	p.R139G	exon2	NM_152588	12	83251120	c.415C>G	0.87	0.085

TNPO1	Q92973	p.R805T	exon22	NM_153188	5	72196824	c.2414G>C	0.49	0.291
TPSD1	Q9BZJ3	p.I87V	exon3	NM_012217	16	1306802	c.259A>G	0	0.51
TREML2	Q5T2D2	p.V50M	exon2	NM_024807	6	41166075	c.148G>A	0	0.998
TRIM65	Q6PJ69	p.A132S	exon1	NM_173547	17	73892625	c.394G>T	0.01	0
TRIM72	Q6ZMU5	p.P320L	exon7	NM_001008274	16	31235601	c.959C>T	0.46	0
TTC37	Q6PGP7	p.A830S	exon23	NM_014639	5	94852403	c.2488G>T	0	0.999
UBE2D1	P51668	p.Q34K	exon3	NM_003338	10	60121266	c.100C>A	0.01	0.992
UBR4	Q5T4S7	p.Q3358K	exon68	NM_020765	1	19447752	c.10072C>A	0.11	0
UBXN11	Q5T124	p.G372S	exon11	NM_001077262	1	26608879	c.1114G>A	0.45	0
UNC13D	Q70J99	p.V63F	exon3	NM_199242	17	73839314	c.187G>T	0.22	0
UNC5B	Q8IZJ1	p.R755H	exon14	NM_170744	10	73055656	c.2264G>A	0.19	0.589
USB1	Q9BQ65	p.H208N	exon6	NM_024598	16	58052888	c.622C>A	0	0.997
USP43	Q70EL4	p.G739R	exon14	NM_153210	17	9615329	c.2215G>C	0.13	0.74
USP8	P40818	p.R763W	exon16	NM_001128611	15	50784950	c.2287C>T	0	
USP8	P40818	p.N764K	exon16	NM_001128611	15	50784955	c.2292C>A	0.01	0.535
VGF	O15240	p.A325G	exon2	NM_003378	7	1.01E+08	c.974C>G	0.1	0.001
VPS39	Q96JC1	p.R650L	exon19	NM_015289	15	42456633	c.1949G>T	0	
WDR88	Q6ZMY6	p.L72F	exon1	NM_173479	19	33623291	c.216G>T		
WNK2	Q9Y3S1	p.V1801G	exon22	NM_006648	9	96055149	c.5402T>G	0.17	0.042
ZBTB48	P10074	p.H425Y	exon7	NM_005341	1	6647586	c.1273C>T	0	0.551
ZBTB9	Q96C00	p.R109P	exon2	NM_152735	6	33423203	c.326G>C	0	0.957
ZFPM1	Q8IX07	p.E444D	exon10	NM_153813	16	88599698	c.1332G>C	0.02	0.716
ZNF185	O15231	p.D400Y	exon15	NM_007150	X	1.52E+08	c.1198G>T	0.29	
ZNF2	Q9BSG1	p.E17K	exon3	NM_021088	2	95843243	c.49G>A	0.3	
ZNF492	Q9P255	p.A40T	exon3	NM_020855	19	22836805	c.118G>A	0.18	0
ZNF717	Q9BY31	p.M51I	exon3	NM_001128223	3	75790792	c.153G>T	0.07	0.988
ZNF83	P51522	p.A277V	exon4	NM_001105552	19	53116988	c.830C>T	0.15	0.049
ZSCAN18	Q8TBC5	p.S224T	exon3	NM_001145542	19	58600105	c.671G>C	0	0.925

Supplementary Table 6. SAAVs in MT-specific sample.

Protein Variants	Accession number	Protein_Change	Chrs	Exon	Position	cDNA_Change	SIFT Score	PolyPhenV2
ABCA7	Q8IZY2	p.T1883M	19	exon42	1062248	c.5648C>T	0	0.999
ACTN1	P12814	p.Q634H	14	exon16	69349226	c.1902G>T	0.04	0.007
AEBP2	Q6ZN18	p.C297X	12	exon3	19626193	c.891C>A	0	
AFF1	P51825	p.Q733K	4	exon12	88036182	c.2197C>A	0.55	0.444
ANKRD33B	A6NCL7	p.Q115P	5	exon1	10564923	c.344A>C	0.1	
ANKS1A	Q92625	p.G49S	6	exon1	34857324	c.145G>A	0.65	0
APC	P25054	p.Q775X	5	exon14	1.12E+08	c.2323C>T	0.17	0.733
ARMC12	Q5T9G4	p.L172P	6	exon3	35706285	c.515T>C	0.12	0.877
ASTN2	O75129	p.A226S	9	exon5	1.19E+08	c.676G>T	0.01	0.445
BRE	O75150	p.T284R	2	exon10	28464260	c.851C>G	0.06	0.121
BTRC	Q9Y297	p.L556F	10	exon13	1.03E+08	c.1666C>T	0.32	0
C16orf62	Q7Z3J2	p.V71G	16	exon1	19566996	c.212T>G	0	
CACHD1	Q5VU97	p.I1035M	1	exon24	65145291	c.3105C>G	0.24	0.032
CASKIN2	Q8WXE0	p.E487G	17	exon14	73500303	c.1460A>G	0	0.932
CCDC63	Q8NA47	p.D13N	12	exon3	1.11E+08	c.37G>A	0.41	0.006
CES5A	Q6NT32	p.E475K	16	exon11	55883536	c.1423G>A	0.41	
CNTD1	Q8N815	p.Q94K	17	exon3	40956277	c.280C>A	0.75	0.365
CNTN5	O94779	p.R53X	11	exon4	99690376	c.157C>T	1	
COL5A2	P05997	p.G321V	2	exon15	1.9E+08	c.962G>T	0	0.996
COPS6	Q7L5N1	p.T294S	7	exon10	99689309	c.881C>G	0.01	0.002
COPZ2	Q9P299	p.G22A	17	exon1	46115075	c.65G>C		
CRELD2	Q6UXH1	p.A243T	22	exon7	50316920	c.727G>A	0.35	0.023
CUBN	O60494	p.L2370I	10	exon46	16957922	c.7108C>A	0.23	0.085
CYP27B1	O15528	p.D418E	12	exon8	58157553	c.1254C>A	0.02	0.998
DDX51	Q8N8A6	p.G114E	12	exon2	1.33E+08	c.341G>A	0.45	0.001
DIAPH2	O60879	p.L751I	X	exon20	96354696	c.2251C>A	0.01	0.782
DPP6	P42658	p.A20T	7	exon1	1.54E+08	c.58G>A	0.01	
DPYSL3	Q14195	p.P305T	5	exon9	1.47E+08	c.913C>A	0	1

DRP2	Q13474	p.R492W	X	exon14	1.01E+08	c.1474C>T	0	1
EIF5AL1	Q6IS14	p.V137L	10	exon1	81272814	c.409G>C		
ELL2	O00472	p.G161V	5	exon5	95242486	c.482G>T	0.06	0.932
EMB	Q6PCB8	p.D95Y	5	exon3	49707131	c.283G>T	0.01	0.838
ESYT1	Q9BSJ8	p.D1097E	12	exon31	56537622	c.3291C>A	0.47	0.875
FAM104B	Q5XKR9	p.E40V	X	exon2	55185563	c.119A>T	0	0.849
FAM131C	Q96AQ9	p.R107Q	1	exon5	16386495	c.320G>A	0.92	0.998
FAM135B	Q49AJ0	p.V67M	8	exon4	1.39E+08	c.199G>A	0	0.762
FAM26F	Q5R3K3	p.A107V	6	exon2	1.17E+08	c.320C>T	0.56	0.01
FBXL6	Q8N531	p.R402Q	8	exon7	1.46E+08	c.1205G>A	0.5	0
FOXH1	O75593	p.W77R	8	exon2	1.46E+08	c.229T>A	0	1
FOXRED1	Q96CU9	p.Y325D	11	exon9	1.26E+08	c.973T>G	0	0.621
FRAS1	Q86XX4	p.P2606T	4	exon54	79396725	c.7816C>A	0.14	
FRY	Q5TBA9	p.R1875Q	13	exon42	32808807	c.5624G>A	0	1
FZD8	Q9H461	p.P644R	10	exon1	35928427	c.1931C>G	0.08	0.011
GABBR2	O75899	p.Q386X	9	exon7	1.01E+08	c.1156C>T	1	0.735
GCC1	Q96CN9	p.V699A	7	exon2	1.27E+08	c.2096T>C	0.01	0.024
GCNT4	Q9P109	p.R388Q	5	exon1	74324700	c.1163G>A	0	1
GEMIN2	O14893	p.C221F	14	exon8	39601190	c.662G>T	0	1
GPR150	Q8NGU9	p.P202A	5	exon1	94956583	c.604C>G	0.15	0.063
GREB1	Q4ZG55	p.S1598I	2	exon27	11772216	c.4793G>T	0.05	0.485
GRIK3	Q13003	p.L386F	1	exon8	37319270	c.1158G>T	0.27	0.954
GXYLT1	Q4G148	p.D153V	12	exon3	42512830	c.458A>T	0.01	0.957
GXYLT1	Q4G148	p.H148Q	12	exon3	42512844	c.444T>G	0.04	0.999
HOXA3	O43365	p.S354I	7	exon3	27147805	c.1061G>T	0.01	0.737
HR	O43593	p.L673M	8	exon8	21980110	c.2017C>A	0.03	0.191
HS6ST1	O60243	p.D87E	2	exon1	1.29E+08	c.261C>A	0.02	0.999
IGFBP2	P18065	p.L22P	2	exon1	2.17E+08	c.65T>C	0.1	0.483
IGSF3	O75054	p.E434G	1	exon7	1.17E+08	c.1301A>G	0.11	0.067
KAZN	Q674X7	p.S310T	1	exon6	15386680	c.929G>C	0	0.664
KCNK18	Q7Z418	p.L258F	10	exon3	1.19E+08	c.774G>T	0.33	0.024
KCNT2	Q6UVM3	p.A176S	1	exon7	1.96E+08	c.526G>T	0.28	0
KIF25	Q9UIL4	p.C220X	6	exon7	1.68E+08	c.660C>A	0.08	0.395
KLF11	O14901	p.D2N	2	exon2	10186289	c.4G>A	0.01	0.99

KRT4	P19013	p.G87A	12	exon1	53207583	c.260G>C	0.54	0.406
LILRA6	Q6PI73	p.Y350X	19	exon6	54744358	c.1050T>G	0.57	0.4011
MAP2K3	P46734	p.L186W	17	exon8	21207813	c.557T>G	0	1
MAPKAP1	Q9BPZ7	p.M46I	9	exon2	1.28E+08	c.138G>T	0.3	0.002
MARK2	Q7KZI7	p.D338E	11	exon12	63669681	c.1014C>A	0.04	0.003
MAS1L	P35410	p.C194F	6	exon1	29455099	c.581G>T	0	1
MCC	P23508	p.G18S	5	exon1	1.13E+08	c.52G>A	0	0.221
MCTP1	Q6DN14	p.L129F	5	exon4	94278063	c.387G>T	0	0.926
MICAL2	O94851	p.L99F	11	exon4	12225829	c.297G>T	0.03	0.057
MICAL2	O94851	p.L99F	11	exon4	12225829	c.297G>T	0.03	0.057
MIR205HG	G1QLW0	p.R98G	1	exon4	2.1E+08	c.292A>G		
MITF	O75030	p.A258T	3	exon8	70008503	c.772G>A	0.17	0
KMT2C	Q8NEZ4	p.N729D	7	exon14	1.52E+08	c.2185A>G	0.47	0
KMT2C	Q8NEZ4	p.S730Y	7	exon14	1.52E+08	c.2189C>A	0.1	0.011
MPG	P29372	p.R12W	16	exon2	129433	c.34C>T	0	0.986
MUC12	A1L198	p.D868N	7	exon2	1.01E+08	c.2602G>A		
MUC2	Q02817	p.P1495T	11	exon30	1092664	c.4483C>A	0.35	
MUC2	Q02817	p.T1498S	11	exon30	1092674	c.4493C>G	0.77	
MUC2	Q02817	p.S1513T	11	exon30	1092719	c.4538G>C	0.89	
NIPBL	Q6KC79	p.L1257I	5	exon16	37003363	c.3769C>A	1	0
NKX2-8	O15522	p.V89G	14	exon2	37050561	c.266T>G	0	0.995
NRXN2	P58401	p.Q256R	11	exon5	64390493	c.767A>G	0.39	0.005
OBSCN	Q5VST9	p.A1274G	1	exon13	2.28E+08	c.3821C>G		
OR2T27	Q8NH04	p.Y120C	1	exon1	2.49E+08	c.359A>G	0.02	0.999
OTOP1	Q7RTM1	p.K56R	4	exon1	4228425	c.167A>G	0	0.98
PCMTD1	Q96MG8	p.K307E	8	exon6	52733066	c.919A>G	0.97	0
JADE3	Q92613	p.R137C	X	exon5	46884250	c.409C>T	0	0.371
KLHL23	F7D1B4	p.R52X	2	exon4	1.71E+08	c.154C>T	1	0.724
PITRM1	Q8K411	p.Q59P	10	exon3	3208567	c.176A>C	0.22	
PKMYT1	Q99640	p.R81W	16	exon3	3026802	c.241C>T	0.01	0.512
PKN3	Q6P5Z2	p.L705I	9	exon18	1.31E+08	c.2113C>A	0	0.785
PLEKHN1	Q494U1	p.E331G	1	exon10	907758	c.992A>G	0.36	0
PLIN4	Q96Q06	p.G736R	19	exon3	4511724	c.2206G>C	0.33	
PLIN4	Q96Q06	p.V661A	19	exon3	4511948	c.1982T>C	0.29	

POLR2E	P19388	p.V209G	19	exon7	1089492	c.626T>G	0	0.961
POLRMT	O00411	p.G923C	19	exon12	620077	c.2767G>T	0	1
PPP1R3A	Q16821	p.E292D	7	exon3	1.14E+08	c.876G>C	0.03	0.104
RBMXL1	Q96E39	p.K9N	1	exon2	89449483	c.27G>C	0	0.11
RBMXL3	Q8N7X1	p.R393G	X	exon1	1.14E+08	c.1177A>G		
RERE	Q9P2R6	p.S1019P	1	exon19	8420512	c.3055T>C	0.24	0.263
RERE	Q9P2R6	p.A1018P	1	exon19	8420515	c.3052G>C	0.04	0.401
RP1L1	Q8IWN7	p.E1305G	8	exon4	10467694	c.3914A>G	0.01	0.462
SAA1	P0DJI8	p.A56S	11	exon3	18290816	c.166G>T	0	0.003
SFI1	A8K8P3	p.S237R	22	exon8	31957324	c.711C>A	0.3	0.11
SFRP4	Q6FHJ7	p.K324N	7	exon6	37947150	c.972A>C	0.05	0.514
SIRT6	Q8N6T7	p.R291P	19	exon8	4174810	c.872G>C	0.15	0.989
SRGAP3	O43295	p.A510S	3	exon12	9074345	c.1528G>T	0.82	0
STARD8	Q92502	p.P918T	X	exon12	67943660	c.2752C>A	0	0.997
STK19	P49842	p.S307G	6	exon7	31948448	c.919A>G	0.11	0
SZT2	Q5T011	p.P919H	1	exon19	43891255	c.2756C>A	0.09	0.998
TAS2R19	P59542	p.F290S	12	exon1	11174302	c.869T>C	0.07	0.298
TAS2R19	P59542	p.F290S	12	exon1	11174302	c.869T>C	0.07	0.298
TBCD	Q9BTW9	p.A422E	17	exon13	80772757	c.1265C>A	0	
TBL1X	O60907	p.D516N	X	exon17	9683035	c.1546G>A	0.39	0.034
TCEAL5	Q5H9L2	p.D82E	X	exon3	1.03E+08	c.246C>A	0.02	0
TCF7	P36402	p.R339W	5	exon8	1.33E+08	c.1015C>T	0	1
TCHH	Q07283	p.L507V	1	exon2	1.52E+08	c.1519C>G		0.363
TFAP2A	P05549	p.N50H	6	exon2	10410466	c.148A>C	0.02	0.899
TMEM163	Q8TC26	p.I157L	2	exon5	1.35E+08	c.469A>C	0.34	0.366
TPRX1	Q8N7U7	p.S244P	19	exon2	48305538	c.730T>C	0.37	0.164
TRIM49C	P0CI26	p.S298N	11	exon8	89774252	c.893G>A	1	
TRPC4AP	Q8TEL6	p.C438F	20	exon10	33603824	c.1313G>T	0	1
TSHR	P16473	p.T441P	14	exon10	81609723	c.1321A>C	0	0.999
TSHZ1	Q6ZSZ6	p.T112P	18	exon2	72997831	c.334A>C	0.02	0.025
TTC37	Q6PGP7	p.G515X	5	exon17	94859448	c.1543G>T	0.49	0.735
TXNRD2	Q9NNW7	p.M263I	22	exon11	19883090	c.789G>T	0.26	0.067
TXNRD2	Q9NNW7	p.V162F	22	exon6	19903332	c.484G>T	0.01	0
UNKL	Q9H9P5	p.G443C	16	exon11	1421541	c.1327G>T	0.07	

VANGL1	Q8TAA9	p.S336X	1	exon6	1.16E+08	c.1007C>A	0.01	0.735
VEGF	P15692	p.H218Q	6	exon2	43742125	c.654C>A	0.08	0.831
VPS13D	Q5THJ4	p.T3861N	1	exon60	12460260	c.11582C>A	0.05	0.453
VPS37A	Q8NEZ2	p.A386E	8	exon11	17143899	c.1157C>A	0.04	0.004
ZBTB45	Q96K62	p.T138P	19	exon2	59028629	c.412A>C	0.96	0
ZFP1	Q6P2D0	p.E297K	16	exon4	75203897	c.889G>A	0.29	0.956
ZNF141	Q15928	p.E315K	4	exon4	367169	c.943G>A	0.06	0.13
ZNF410	Q86VK4	p.A90S	14	exon5	74363066	c.268G>T	0	0.001
ZNHIT1	O43257	p.R114C	7	exon4	1.01E+08	c.340C>T	0	1
ZP3	P21754	p.H287R	7	exon8	76069881	c.860A>G	0.15	0.046

Extended Experimental Procedures

Patient specimen acquisition

The study was examined and approved by the Ethics Committee of the Shanghai Tenth People's Hospital, Tongji University School of Medicine (SHSY-IEC-PAP-16-24). This study was registered with ClinicalTrials.gov, number NCT02917707. Follow-up data and statistics were recorded for all patients through Dec. 31, 2017.

DNA and RNA extraction.

Using a co-isolation protocol, DNA and RNA were purified simultaneously using the QIAGEN All Prep DNA/RNA Micro Kit (Qiagen, CA, USA) according to the manufacturer's instructions [1]. The nucleic acid concentration was determined using a Nanodrop1000 spectrophotometer (Thermo Fisher Scientific; Waltham, MA, USA) [2], and the RNA purity was verified using 1.5% denaturing agarose gels.

Protein extraction and analysis by LC-MS/MS.

Forty four paired fresh primary tumors from the colorectal and paired normal colorectal tissues (PN) independently confirmed by two experienced pathologists Z.Y.C. and H.Q.Y., were used for proteogenomic analysis (Table S1). Three different parts of the same lesions for every sample were compared for data analysis and measurement of the variation caused by random biological effects. The tissues were homogenized in 4 % SDS and 0.1 M DTT in 0.1 M Tris-HCl, pH 7.6 on ice, sonicated 10 times (80 w; 10 s sonication/5 s suspension), incubated for 3 min at 95 ℃, and briefly sonicated. The protein concentrations of clarified lysates were determined using a fluorescence assay and then 200 μg of clarified lysates were proteolyzed on a 10 kDa filter (PALL Life Sciences, Shanghai, China) using the filter-aided sample preparation method [3]. The peptide samples were then desalted onto a solid-phase extraction cartridge. The lyophilized peptide mixture was re-suspended in water with 0.1% formic acid (v/v), and its content was estimated by ultraviolet light spectral density at 280 nm [4]. Then, 3 μg of the digest sample was analyzed by

nano-liquid chromatography-tandem mass spectrometry on a LTQ Orbitrap Velos Pro mass spectrometer as previously described [5].

The acquired data from mass spectrometry runs were combined and searched against the UniProt Human database (05/2016, 153652 entrys) using Maxquant software (version 1.3.0.5; http://maxquant.org/) as described [6]. Proteins were identified using the Andromeda peptide search engine integrated into the Maxquant platform [7]. The required minimum peptide length for identification was 7 amino acids, and the false discovery rate at the protein level, peptide level and site were set to 0.01. The normalized spectral protein intensity values were calculated for each protein group.

The Maxquant peptide and protein quantification results files were imported into Perseus software (version 1.5.1.6) to identify the differentially expressed proteins [8-10]. After importing the quantitative data from ProteinGroups.txt into Perseus, a filtering criterion is set to keep the identified proteins with the quantified values of all ten reporter ions (no missing value) in the final identification list. The protein intensities are $\log_2$-transformed and normalized by subtracting the median intensity in each column/sample. Principal component analysis (PCA) is performed based on protein intensities to differentiate groups [11]. Two-samples t-tests coupled with Benjamini–Hochberg (FDR cutoff of 0.05) correction are performed to identify the differentially expressed proteins [12].

Validation of point mutations by PCR and Sanger sequencing.

The reliability of the exome analysis and somatic variant identification strategies was assessed using PCR and Sanger sequencing. PCR was performed using the GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). About 20 ng template DNA from each sample was used per reaction. The products were sequenced, and all sequences were analyzed with the Sequencing Analysis Software Version 5.2 (Applied Biosystems) [13].

RNA sequencing analysis.

Of the 44 samples described in this study, 9 specimens from 3 CRC patients with metastasis (comprised of triplet sets of PN, primary MT, and synchronously matched liver metastases sample) and 3 specimens from 3 CRC patients without liver metastasis were obtained for RNA sequencing analysis. The mRNA libraries were separately generated from total RNA and constructed according to the standard Illumina RNA library preparation protocol (Illumina Inc, USA) [14]. Sequencing was performed on the Illumina Nextseq 500 platform according to the manufacturer's instructions [15]. Images generated by Nextseq 500 were converted into nucleotide sequences using a base call pipeline and stored in FASTQ format, and the raw reads were filtered prior to analyzing the data [16]. Clean reads were mapped to reference Homo sapiens transcriptome sequences from the UCSC website (hg19) using Bowtie 2 and Tophat 2.0.1 software [17-19].

To annotate gene expression, reads per kilobases per million read values of each gene were calculated, and differentially expressed genes were extracted using this value. The formula for calculating these values was defined as: reads per kilobases per million read values = total exon reads / (mapped reads [millions] × exon length [kbp]). RNA-Seq generated from 7,256,241 to 368,134,260 reads that were aligned to the human reference GRCh37.p5/hg19, representing an average genome-coverage of 105x.

Hierarchical clustering, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.

Hierarchical clustering was performed using MEV software (http://mev.tm4.org/, v4.7.0, TIGR) [20]. The matrix was presented graphically by colouring each expression result on the basis of measured colour range: lower limit '0.0' was coloured green, upper limit '369.5' was coloured red and midpoint value '37.5' was coloured black. Pearson correlation was used as distance metric and the complete linkage method was used.

To identify genes/proteins that are specifically dysregulated in CLM, we fixed the cutoff at 2-fold with a P value less than 0.05 with a FDR q-value < 0.05. Dysregulated genes/proteins were

subjected to GO analysis and KEGG pathway analysis by DAVID (http://david.ncifcrf.gov) and STRING (https://string-db.org/) online analysis software [21-23].

TCGA data acquisition and processing.

We downloaded RNA-sequencing data from 379 CRC patients from TCGA portal (https://cancergenome.nih.gov/), 12 of which had liver metastasis at the time of diagnosis or during the five-year follow-up period, and 367 of which had CRC without metastasis to the liver. The mRNA expression levels were investigated in 379 CRC tissues and 32 PN tissues in TCGA datasets by Illumina HiSeq 2000 RNA Sequencing Version 2 analysis and normalized by the RSEM algorithm. Whole-exome sequencing mutation datasets were downloaded from TCGA data set to create a customized CRC mutation database. The clinical information recorded, including the patient's characteristics, tumor characteristics, and overall and progression-free survival was assessed.

Cell lines and transfection

Human CRC cell line SW480 were purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China) and cultured in DMEM media (Invitrogen, Carlsbad, USA) and supplemented with 10 % (v/v) fetal bovine serum, 100 U/ml penicillin, and 100 mg/ml streptomycin. SW480 cell lines were routinely tested for mycoplasma contamination, and have been authenticated with short-tandem repeat analysis. Cell culture was conducted at 37 ℃ in a humidified 5% CO_2 incubator.

FDFT1 shRNA sequences were cloned into the plvx-shRNA plasmid. The non-silencing lentiviral shRNA vector was used as a control. The lentivirus was packaged using psPAX2 and pMD2G, a three-plasmid system. To obtain stable cell lines, lentivirus supernatant was added to SW480 cells, followed by screening with 1 μg/ml puromycin for 2 weeks. For UQCR5 over-expression, the human full length *UQCR5* cDNA was cloned into the pMSCV-hygro vector. The SW480 cells with stable over-expression of UQCR5 were polyclonal derivatives with

hygromycin selection to avoid clonal variations in functional assays.

Scratch-Wound Assay

The human CRC SW480 cells were conducted at 37 ℃ in a humidified 5% CO_2 incubator and cells were grown into confluency in 6-well plates. The monolayer was artificially injured by scratching across the plate with a 200 ul pipette tip. The wells were washed 3 times to remove detached cells or cell debris. After 12 hours, digital images were captured using a camera-equipped, inverted microscope (Carl Zeiss, Inc., Thorwood, NY, USA) and wound width measurements were subtracted from wound width at time zero to obtain the net wound closure.

***In vitro* invasion assays**

Corning Costar Transwell 24-well plates with 8-um-pore-size polycarbonate membrane filters (Costar, Cambridge, MA) coated with BD Matrigel matrix (Becton Dickinson, Bedford, MA) were maintained for 1 h at 37 ℃, followed by the addition of 1×10^5 transfected cells suspended in 200 μl medium with 1% serum into the top of each well insert [24-26]. Normal growth medium was added to the bottom wells. The cells were allowed to migrate for 24 h at 37 ℃. The migrated cells were fixed with 10% methanol for 15 min. The invading cells on the lower surface of the membrane were stained with 0.5% crystal violet for 5 min at room temperature. Random fields were photographed and the stained cells were counted under a microscope (Nikon Corporation).

Statistical analysis.

Data were expressed as means ± standard deviations. Categorical data were reported as numbers and percentages. F tests were used to assess the equality of variances for comparable groups. Paired t test, One-way analysis of variance (ANOVA), Kruskal-Wallis test, and χ^2 tests were used to analyze mRNA expression. For time-to-event analyses, survival estimates were calculated by the Kaplan-Meier analysis, and groups were compared with the log-rank test. Clinical variables that were considered for single variable analyses were previously identified as confounding variables with impact on the prognosis of patients with colorectal cancer. The

Spearman's correlation coefficient was used to test the relationship of two independent groups. To identify genes/proteins that are specifically dysregulated in CLM, we fixed the cutoff at 2-fold with a P value less than 0.05 with a FDR q-value < 0.05. All calculations were performed with SPSS 20.0 software (SPSS Inc, Chicago, IL, USA).

Supplementary References

1. Ma Y.S., Wu Z.J., Bai R.Z., Dong H., Xie B.X., Wu X.H., Hang X.S., Liu A.N., Jiang X.H., Wang G.R., Jiang J.J., Xu W.H., Chen X.P., Tan G.H., Fu D., Liu J.B., Liu Q. DRR1 promotes glioblastoma cell invasion and epithelial-mesenchymal transition via regulating AKT activation. Cancer Lett. 2018;423:86-94.
2. Xie R.T., Cong X.L., Zhong X.M., Luo P., Yang H.Q., Lu G.X., Luo P., Chang Z.Y., Sun R., Wu T.M., Lv Z.W., Fu D., Ma Y.S. MicroRNA-33a downregulation is associated with tumorigenesis and poor prognosis in patients with hepatocellular carcinoma. Oncol. Lett. 2018;154571-4577.
3. Neilson K.A., Ali N.A., Muralidharan S., Mirzaei M., Mariani M., Assadourian G., Lee A., van Sluyter S.C., Haynes P.A. Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics. 2011;11:535-553.
4. Bouckaert R.R., Drummond A.J. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. 2017;17:42.
5. Polisetty R.V., Gautam P., Sharma R., Harsha H.C., Nair S.C., Gupta M.K., Uppin M.S., Challa S., Puligopu A.K., Ankathi P., Purohit A.K., Chandak G.R., Pandey A., Sirdeshmukh R. LC-MS/MS analysis of differentially expressed glioblastoma membrane proteome reveals altered calcium signaling and other protein groups of regulatory functions. Mol. Cell. Proteomics. 2012;11: M111.013565.

6. Luber C.A., Cox J., Lauterbach H., Fancke B., Selbach M., Tschopp J., Akira S., Wiegand M., Hochrein H., O'Keeffe M., Mann M. Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. Immunity. 2010;32: 279-289.
7. Ma Y.S., Huang T., Zhong X.M., Zhang H.W., Cong X.L., Xu H., Lu G.X., Yu F., Xue S.B., Lv Z.W., Fu D. Proteogenomic characterization and comprehensive integrative genomic analysis of human colorectal cancer liver metastasis. Mol. Cancer. 2018;17:139.
8. Imperial R., Ahmed Z., Toor O.M., Erdoğan C., Khaliq A., Case P., Case J., Kennedy K., Cummings L.S., Melton N., Raza S., Diri B., Mohammad R., El-Rayes B., Pluard T., Hussain A., Subramanian J., Masood A. Comparative proteogenomic analysis of right-sided colon cancer, left-sided colon cancer and rectal cancer reveals distinct mutational profiles. Mol. Cancer. 2018;17:177.
9. Lim S.Y., Lee J.H., Diefenbach R.J., Kefford R.F., Rizos H. Liquid biomarkers in melanoma: detection and discovery. Mol. Cancer. 2018;17:8.
10. Monteleone F., Taverna S., Alessandro R., Fontana S. SWATH-MS based quantitative proteomics analysis reveals that curcumin alters the metabolic enzyme profile of CML cells by affecting the activity of miR-22/IPO7/HIF-1α axis. J. Exp. Clin. Cancer Res. 2018;37:170.
11. Antoury L., Hu N., Balaj L., Das S., Georghiou S., Darras B., Clark T., Breakefield X.O., Wheeler T.M. Analysis of extracellular mRNA in human urine reveals splice variant biomarkers of muscular dystrophies. Nat. Commun. 2018;9:3906.
12. Wu S., Ou T., Xing N., Lu J., Wan S., Wang C., Zhang X., Yang F., Huang Y., Cai Z. Whole-genome sequencing identifies ADGRG6 enhancer mutations and FRS2 duplications as angiogenesis-related drivers in bladder cancer. Nat. Commun. 2019;10:720.
13. Downes N.L., Laham-Karam N., Kaikkonen M.U., Ylä-Herttuala S. Differential but complementary HIF1α and HIF2α transcriptional regulation. Mol. Ther. 2018;26:1735-1745.

14. Chen H., Pan H., Qian Y., Zhou W., Liu X. MiR-25-3p promotes the proliferation of triple negative breast cancer by targeting BTG2. Mol. Cancer. 2018;17:4.
15. Xing F., Wang S., Zhou J. The expression of MicroRNA-598 inhibits ovarian cancer cell proliferation and metastasis by targeting URI. Mol. Ther. Oncolytics. 2019;12:9-15.
16. Li J., Song P., Jiang T., Dai D., Wang H., Sun J., Zhu L., Xu W., Feng L., Shin V.Y., Morrison H., Wang X., Jin H. Heat shock factor 1 epigenetically stimulates glutaminase-1-dependent mTOR activation to promote colorectal carcinogenesis. Mol. Ther. 2018;26:1828-1839.
17. Liu J.J., Zhang X., Wu X.H. miR-93 promotes the growth and invasion of prostate cancer by upregulating its target genes TGFBR2, ITGB8, and LATS2. Mol. Ther. Oncolytics. 2018;11:14-19.
18. Wang Q., Ju X., Wang J., Fan Y., Ren M., Zhang H. Immunogenic cell death in anticancer chemotherapy and its impact on clinical studies. Cancer Lett. 2018;438:17-23,
19. Ma Y.S., Lv Z.W., Yu F., Chang Z.Y., Cong X.L., Zhong X.M., Lu G.X., Zhu J., Fu D. MiRNA-302a/d inhibits the self-renewal capability and cell cycle entry of liver cancer stem cells by targeting the E2F7/AKT axis. J. Exp. Clin. Cancer Res. 2018;37(1):252.
20. Yang H., Li Y., Zhong X., Luo P., Luo P., Sun R., Xie R., Fu D., Ma Y., Cong X., Li W. Upregulation of microRNA-32 is associated with tumorigenesis and poor prognosis in patients with hepatocellular carcinoma. Oncol. Lett. 2018;15:4097-4104.
21. 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., Xie W.T., Jiang J.J., Song Y.C., Chai L., Jia C.Y., Lu G.X., Zhong X.J., Hou L.K., Wu C.Y., Shi M.X., Liu J.B, Fu D. Prognostic implications of decreased microRNA-101-3p expression in patients with non-small cell lung cancer. Oncol. Lett. 2018;16:7048-7056.
22. Long H.D., Ma Y.S., Yang H.Q., Xue S.B., Liu J.B., Yu F., Lv Z.W., Li J.Y., Xie R.T., Chang Z.Y., Lu G.X., Xie W.T., Fu D., Pang L.J. Reduced hsa-miR-124-3p levels are associated with

the poor survival of patients with hepatocellular carcinoma. Mol. Biol. Rep. 2018;45:2615-2623.

23. Chen F., Long Q., Fu D., Zhu D., Ji Y., Han L., Zhang B., Xu Q., Liu B., Li Y., Wu S., Yang C., Qian M., Xu J., Liu S., Cao L., Chin Y.E., Lam E.W., Copp éJ.P., Sun Y. Targeting SPINK1 in the damaged tumour microenvironment alleviates therapeutic resistance. Nat. Commun. 2018;9:4315.
24. Wang Z.Y., Hu M., Dai M.H., Xiong J., Zhang S., Wu H.J., Zhang S.S., Gong Z.J. Upregulation of the long non-coding RNA AFAP1-AS1 affects the proliferation, invasion and survival of tongue squamous cell carcinoma via the Wnt/β-catenin signaling pathway. Mol. Cancer. 2018;17:3.
25. Sun M., Nie F.Q., Zang C., Wang Y., Hou J., Wei C., Li W., He X., Lu K.H. The Pseudogene DUXAP8 promotes non-small-cell lung cancer cell proliferation and invasion by epigenetically silencing EGR1 and RHOB. Mol. Ther. 2017;25:739-751.
26. Bialk P., Wang Y., Banas K., Kmiec E.B. Functional gene knockout of NRF2 increases chemosensitivity of human lung cancer A549 cells in vitro and in a xenograft mouse model. Mol. Ther. Oncolytics. 2018;11:75-89.