Research Article

Whole-Exome Sequencing Characterized the Landscape of Somatic Mutations and Pathways in Colorectal Cancer Liver Metastasis

Liuxing Feng (ID),[1] **Shifu Hong** (ID),[2] **Jin Gao** (ID),[3] **and Jiayi Li** (ID)[4]

[1]*Department of Hepato-Biliary-Pancreato-Vascular Surgery, The First Affiliated Hospital of Xiamen University, Xiamen, Fujian, China*
[2]*Department of Gastrointestinal Surgery, Cancer Hospital, The First Affiliated Hospital of Xiamen University, Teaching Hospital of Fujian Medical University, Xiamen, China*
[3]*Department of Gastrointestinal Surgery, Xiamen Cancer Hospital, The First Affiliated Hospital of Xiamen University, Xiamen, Fujian 361003, China*
[4]*Department of Medical Oncology, Cancer Hospital, The First Affiliated Hospital of Xiamen University, Teaching Hospital of Fujian Medical University, Xiamen, China*

Correspondence should be addressed to Jiayi Li; jiayi_li01@163.com

Purpose. Liver metastasis remains the leading cause of cancer-related mortality in colorectal cancer. The mechanism of occurrence and development of liver metastasis from colorectal cancer is unclear. *Methods*. The primary tumor tissues and blood samples of 8 patients with liver metastasis of colorectal cancer were collected, followed by nucleic acid extraction and library construction. Whole-exome sequencing was performed to detect the genomic variations. Bioinformatics was used to comprehensively analyze the sequencing data of these samples, including the differences of tumor mutation burden, the characteristics of gene mutations, and signaling pathways. *Results*. The results showed that the top three genes with the highest mutation frequency were *TP53*, *APC*, and *KRAS*. Tumor mutation burden of this study, with a median of 8.34 mutations per MB, was significantly different with The Cancer Genome Atlas databases. Analysis of molecular function and signaling pathways showed that the mutated genes could be classified into five major categories and 39 signaling pathways, involving in Wnt, angiogenesis, P53, Alzheimer disease-presenilin pathway, notch, and cadherin signaling pathway. *Conclusions*. In conclusion, we identified the extensive landscape of altered genes and pathways in colorectal cancer liver metastasis, which will be useful to design clinical therapy for personalized medicine.

1. Introduction

Colorectal cancer (CRC) is the third most common type of malignancy and leading cause of cancer-related death worldwide [1]. Metastasis is still the main cause of cancer-related morbidity, and mortality of colorectal cancer due to liver metastasis accounts for about 25% [2]. Although, early detection and prevention or surgical resection of primary and metastatic lesions can reduce the risk of CRC and improve survival of CRC [3–5], metastatic CRC is still the leading cause of cancer-related deaths, and treatment options are not as selective.

A previous study suggested that the frequency rates of mutations such as *KRAS*, *NRAS*, *BRAF*, and *PIK3CA* in CRC differ among population [6]. *AMER1* is a frequently mutated gene in CRC comprising 553 samples [7]. *TMEM9*, as a novel human transmembrane protein, transactivated by β-catenin functions as a positive feedback regulator of WNT signaling in CRC and mTOR signaling, has been suggested to be an important factor involved in tumorigenesis [8, 9]. Therefore, a better understanding of the biological and phenotypic evolution of CRC and its molecular and genetic mechanisms during the transfer process is crucial.

To further investigate the genetic characteristics of colorectal cancer liver metastasis (CRLM), we performed whole-exome sequencing (WES) in 8 patients with CRLM. Somatic mutations, tumor mutation burden (TMB), molecular functions of mutational genes, and signaling pathways were analyzed. It is expected to provide clinical help for the treatment of patients with liver metastasis from colorectal cancer.

2. Patients and Methods

2.1. Patient Specimen Acquisition. Blood and primary CRC tumor tissue samples were collected from 8 patients with CRLM in the Oncology Department of the First Affiliated Hospital of Xiamen University during the period May 2018 to November 2018. Informed written consent was obtained from all patients before inclusion in the study. Respective tumor tissue samples which were a histologically confirmed adenocarcinoma by two molecular pathologists matched with the inclusion criteria. The study was conducted in accordance with the Helsinki Declaration and was approved by the Ethics Committee of the First Affiliated Hospital of Xiamen University [10].

2.2. DNA Extraction. DNA was extracted from serial sections cut from tumor samples and matched peripheral blood leukocytes as germline DNA control. The cases with tumor cell populations were estimated by pathologists to ensure more than 70% of cells were tumor cells. The DNA was isolated from the FFPE and blood samples using the DNeasy Blood and Tissue Kit (69504, QIAGEN, Venlo, Netherlands) according to the manufacturer's instructions. DNA quantity was assessed by using Agilent's Bioanalyzer (USA).

2.3. Whole-Exome Sequencing and Data Processing. The targeted capture pulldown and exon-wide libraries from genomic DNA were generated through the xGen® Exome Research Panel (Integrated DNA Technologies, Inc., Illinois, USA) and the TruePrep DNA Library Prep Kit V2 for Illumina (#TD501, Vazyme, Nanjing, China). The sequences of captured libraries were performed as pair-end reads on sequences on the Illumina HiSeq 2500 platform. Sequencing reads were processed and mapped to the reference GRCh37/hg19 human genome assembly and to the identified germline variations. Further local rearrangements were performed to improve the alignment of individual reads [11].

2.4. Variant Annotation and Mutation Signature Analysis. Somatic mutations identification and indels were annotated through Mutect [12] and Somatic Indel Detector [13]. The variant data were annotated using ANNOVAR [14] and Oncotator [15] and converted to MAF files by maf tools [16]. The cancer driver genes were analyzed using Intogen [17], including Oncodrive FM and Oncodrive CLUST. Both tools detect signals of positive selection, which appear in genes whose mutations are selected during tumor development

Table 1: Patient characteristics.

Characteristic	No. of cases	Proportion (%)
Total number	$n = 8$	
Age, years (mean)	66.6 (46–83)	
Sex		
Male	5	62.5
Female	3	37.5
Smoking history		
Smoker	1	12.5
Nonsmoker	7	87.5
Drinking	1	12.5
Metastsis	7	87.5
Anatomical classification		
Right hemicolon	2	25.0
Left hemicolon	6	75.0
Stage		
IV	8	100
Chemotherapy	8	100

and are therefore likely drivers. The landscape of top driver mutation spectrum predicted by Intogen for tumors was visulized via R Script, including mutation rate and mutation subclass/subtypes (filtering cutoff, ONCODRIVE FM P value ≤ 0.1).

2.5. Statistical Analyses. All the correlate clinical and biological variables were employed using the SPSS Statistics 22.0 package and ggpubr package [18] in R [19] by means of Fisher's test or a nonparametric test when necessary. The Kruskal–Wallis test was used to analyze whether TMB differ between different data sets.

3. Results

3.1. Patient Characteristics. We collected tumor tissue and matched blood from 8 patients with CRLM at the time of diagnosis, including 5 males and 3 females, with an average age of 66.6 years (range, 46–83 years). One of the patients was a former smoker, and the other seven were nonsmokers. Additionally, one male patient was also alcoholic. According to the anatomical classification system, 75.0% (6/8) of samples were classified as left hemicolon carcinoma, and the other 2 patients were right hemicolon carcinoma. All the patients were in stage IV and treated with chemotherapy. 37.5% (3/8) of the patients had a history of chronic disease, including diabetes, hypertension, coronary heart disease, and hyperuricemia. No patients received radiation therapy before surgery. The detailed clinical characteristics of the patients are shown in Table 1 and Supplemental Table 1.

3.2. Whole-Exome Sequencing and Identification of Somatic Mutations. We performed WES on DNA from 8 tumor tissues along with blood matched and then analyzed successfully with a mean depth of 244x. Somatic mutations were identified by comparing significant changes in nonreference alleles in the tumor and control groups. Overall, 1151 nonsynonymous single nucleotide variants (SNV) were
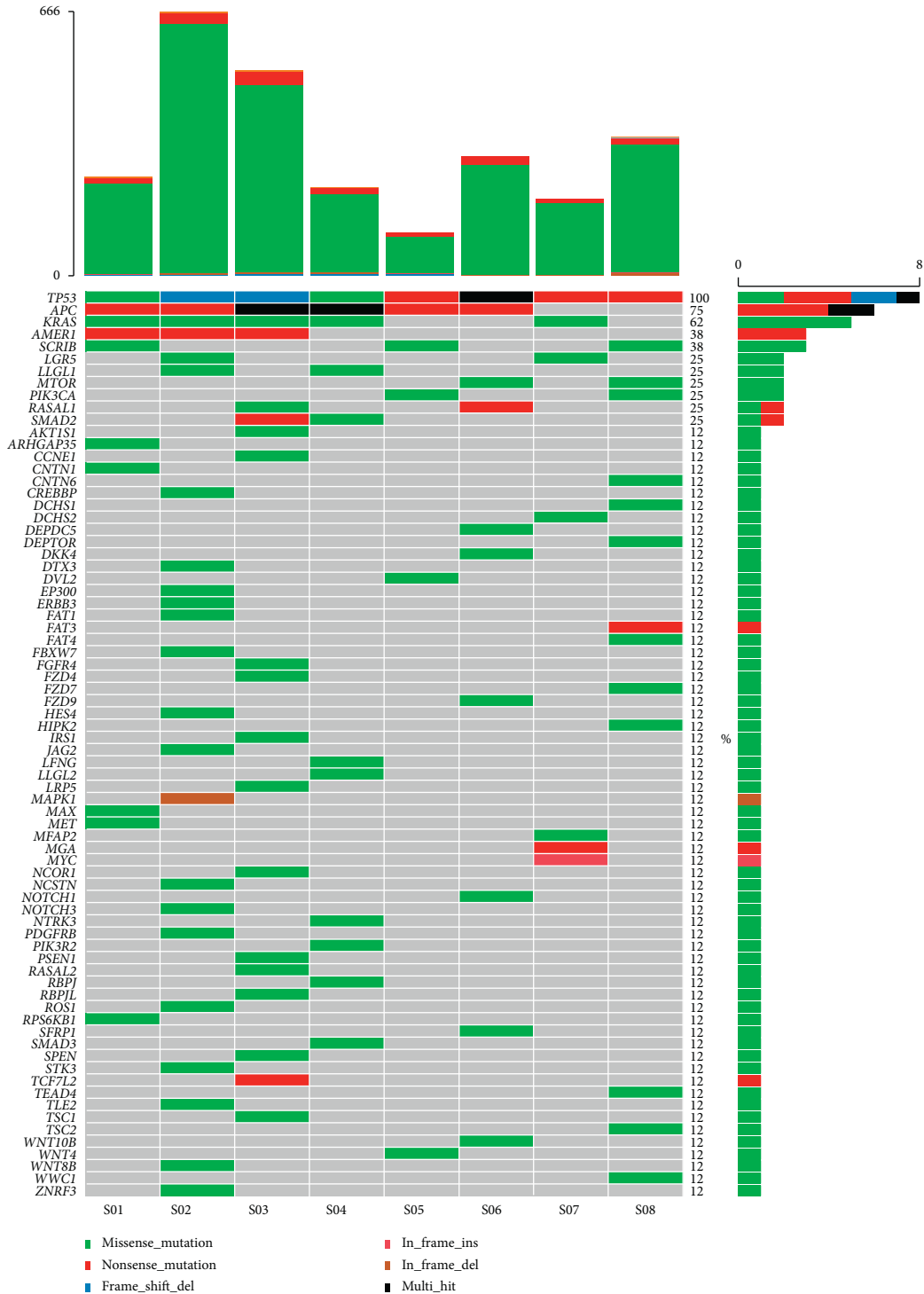
FIGURE 1: Landscape of somatic mutations in CRLM. The different colored tables represent different types of mutations (middle bars). We also calculated somatic mutations SNV using only somatic nonsynonymous mutations sequenced with WES for each sample (top bars), and the right bars represent the absolute number of mutations observed per gene across all samples.

identified (Supplemental Table 2). An overview of the whole-exome sequencing results and the algorithm-generated arm-length copy number alterations are shown in Figure 1. Each gene with a nonsynonymous SNV was reviewed against

known mutations identified in prior studies and subjected to Mutsig analysis. As shown in Figure 1, S02 have the most SNVs, following S03. We listed the top 75 genes based on the frequency of mutations. Among them, *TP53* (100%), *APC*

(75%), and *KRAS* (62%) were the genes with the highest mutation rates. Missense mutation was the most common type of mutation, along with frame shift del, in frame ins, frame del, and so on (Figure 1).

We also calculated TMB using only somatic non-synonymous mutations sequenced with WES. On the whole, we found that the TMBs of different samples were significantly different, with a median of 8.34 mutations per MB (range, 2.79–17.04 mutations/MB) (Figure 2).

In order to compare the differences in TMB between CRLM and TCGA database (COAD and READ), we used the Kruskal–Wallis nonparametric test to test the anova of multiple groups of data after homogeneity of variance test (therefore, anova cannot be used) and found significant difference between multiple database cohorts ($P = 5.9e - 05$) (Figure 3).

3.3. The Landscape of Mutational Signatures. In principle, all types of mutations (such as substitutions, indels, and rearrangements) and any accessory mutation characteristic, for example, the sequence context of the mutation or the transcriptional chain where the mutation occurred, can be incorporated into the set of features by which a mutational signature is defined.

We extracted mutational signatures using base substitutions, and six classes of substitutions ($C > A$, $C > G$, $C > T$, $T > C$, $T > A$, and $T > G$) were referred to by the pyrimidine of the mutated Watson–Crick base pair. In this study, the six mutation types were compared with the TCGA database, and it was found that the proportion of these six mutation replacement types was roughly the same. The mutation percent of $C > T$ was the highest in all substitutions, and this study has no significant difference with COAD and READ in this substitution. $T > G$ substitution have significant difference between CRLM with COAD and READ (Figures 4(a) and 4(b)).

3.4. CRLM-Related Gene Molecular Function and Pathway Analyses. In order to further characterize the functions of mutational genes and their involved regulatory pathways, we used PANTHER classification system [20], an Ontology-Based Pathway Database Coupled with Data Analysis Tools. The results showed that molecular functions were divided into five categories, namely, binding, catalytic activity, molecular function regulator, molecular transducer activity, and transcription regulator activity. Of these, the category of binding (40) and catalytic activity (32) have the most function hits (Figure 5).

Through the PANTHER classification system pathway analysis, it was found that 74 pathway-related genes were involved in a total of 39 primarily signaling pathways, among which the pathways with higher frequency were Wnt signaling pathway (P00057), angiogenesis (P00005), P53 pathway (P00059), Alzheimer disease-presenilin pathway (P00004), notch signaling pathway (P00045), and cadherin signaling pathway (P00012). The other involved pathways and the genes involved in each pathway were referred to Figure 6 and Supplemental Table 3.
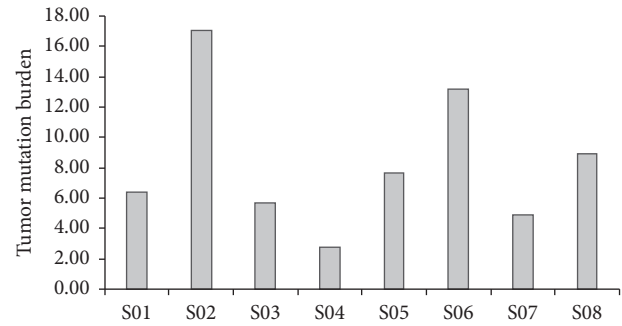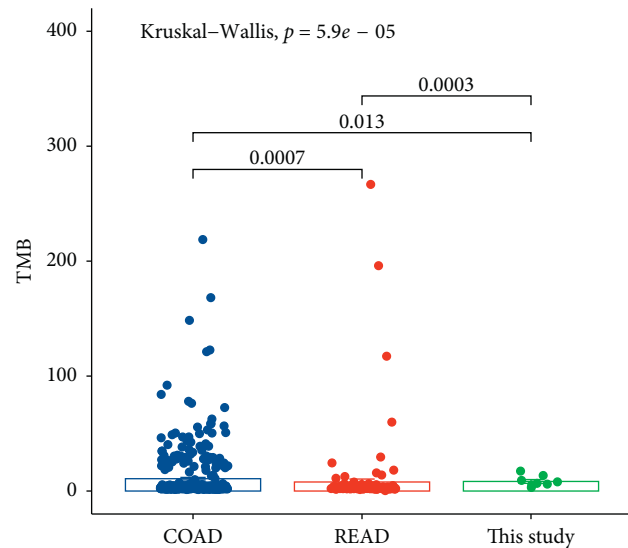

FIGURE 2: TMB analysis in CRLM patients.

FIGURE 3: Comparison of TMB between the date in this study and COAD with READ by thwe Kruskal–Wallis test. The primary sites of COAD and READ are colon and rectum.

4. Discussion

CRC is the third most common malignancy in many countries and the second leading cause of cancer death. It develops from benign adenomatous polyp to invasive cancer, and nearly 50% of CRC patients develop into CRLM [21]. Without treatment, the median survival period of patients with colorectal liver metastasis is only 5–10 months, and the survival rate of over 5 years is less than 0.5% [22].

The molecular pathogenesis of CRC is related to a variety of genetic changes that result in abnormal activation of proto-oncogenes and inactivation of tumor suppressor genes [23]. We briefly described the characteristics of the CRLM by WES and had important insights into the genes and mechanisms of cancer occurrence and development. We found 1151 SNVs and the prevailing mutations were *APC*, *KRAS*, and *TP53* (Figure 1, Supplemental Table 2), which is in accordance with data reported by The Cancer Genome Atlas Network [24]. Currently, there are dozens of biomarkers related to checkpoint inhibitors, among which TMB, PD-L1, and MSI/dMMR have been verified by phase III clinical trials and are widely used in clinical practice.

(a)
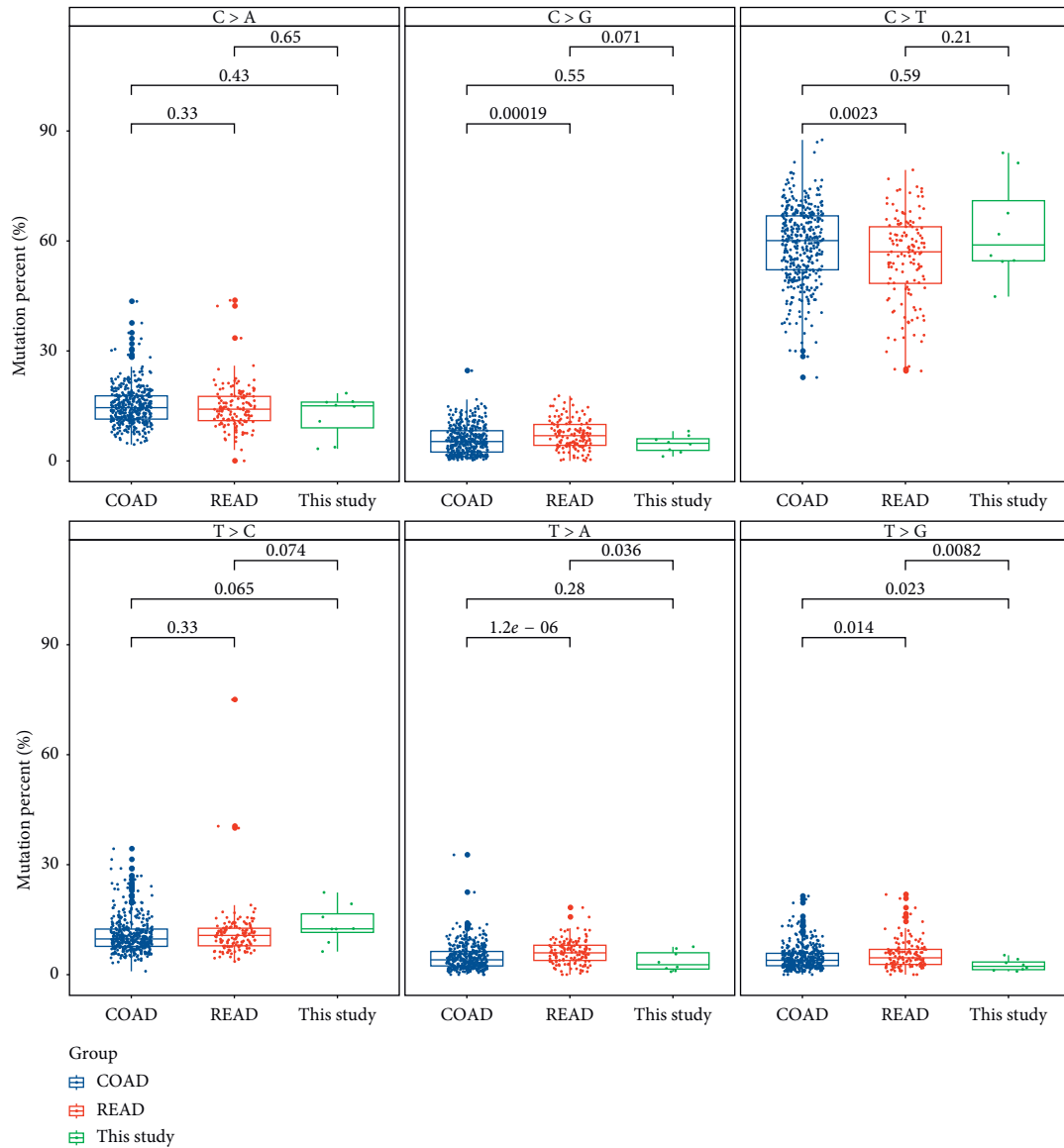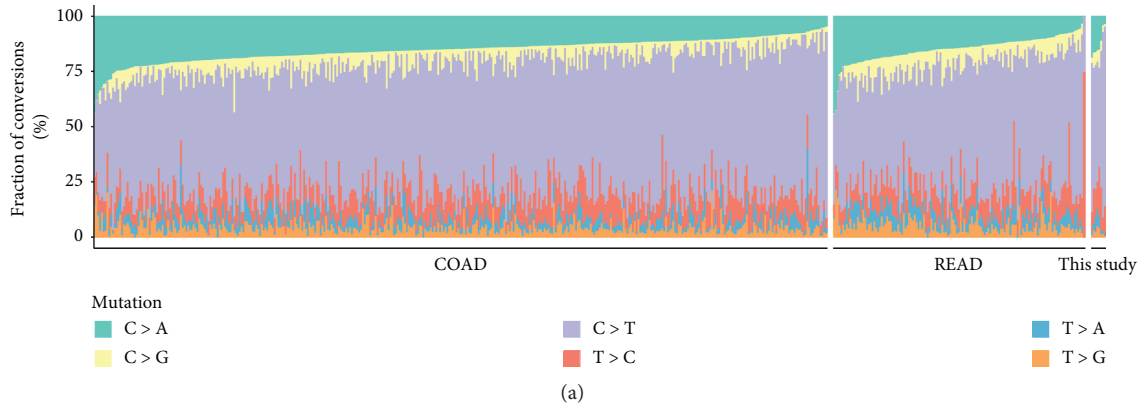
(b)

FIGURE 4: Mutational signature difference in multiple groups. (a) Transition and transversion proportions for six nucleotide changes. The stacked proportion bar chart is sorted by increasing the transition/transversion fraction. (b) Transition/transversion (Ti/Tv) ratios.
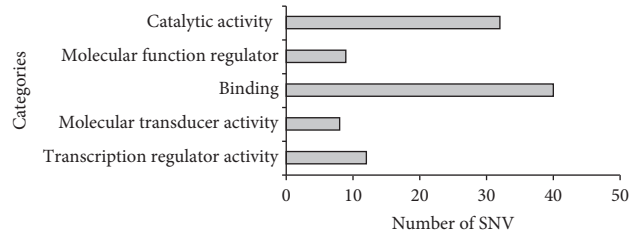
FIGURE 5: Go-slim molecular function by PANTHER for single nucleotide variants.
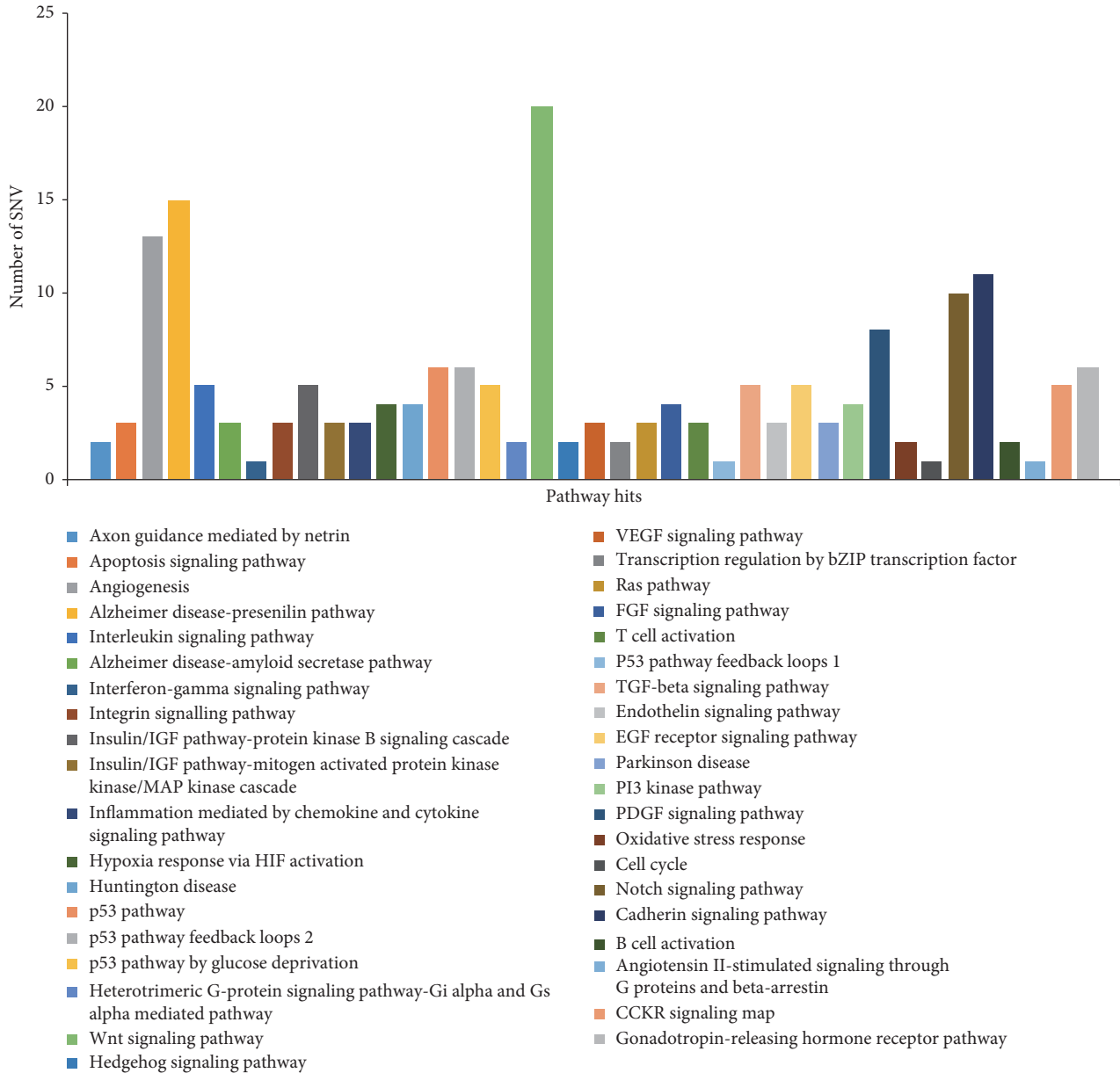

FIGURE 6: Analysis of pathways involving nonsynonymous genes in CRLM.

Tumor mutation load (TMB) is a new biomarker for predicting PD-1/PD-L1 immune response [25]. Even though it has been reported that TMB ≥20 mutation/Mb (TMB-H) alone is not suitable for predicting the immunotherapy effect of each solid tumor type [26], we found that there was a significant difference in TMB between CRLM and colon and rectum, but the TMB did not exceed 20 mutations per MB (mean 8.34) (Figures 2 and 3). For different cancer types, the

setting of high TMB threshold may need more clinical studies and a large number of patient information statistics.

The signatures can be understood as different mutation processes often generate different combinations of mutation types. Thousands of somatic mutations can be identified in a single cancer sample, making it possible to decipher the mutant signature, even if several mutations are operative [27]. The $C > A$ mutational signature, is associated with smoking and chewing tobacco. Six classes of substitutions were extracted, and there was no significant difference in mutation percent between CRLM and colon with rectum cohorts (Figure 4). The genetic characteristics of liver metastasis may be more similar to that of the primary tumor, and the treatment strategy should be more similar to that of the primary tumor colorectal cancer.

Through pathway analysis, we found that oncogenes represented by *KRAS*, *PIK3CA*, *AKT1*, *PIK3R*, and tumor suppressor genes represented by *TP53*, *APC*, *EP300*, *CREBBP*, and *PIK3R1* were mutated, which may lead to changes in angiogenesis, TGF-β, Wnt signaling pathway, notch signaling pathway, and other pathways (Figure 6, Supplemental Table 3). The pathway is complex, mainly reflected in the fact that one mutation gene is involved in multiple pathways [28], and various pathways are also cross-regulated, such as angiogenesis and notch signaling pathway.

5. Conclusion

Our study identified changes in driver gene mutations, TMB, and base Ti/Tv ratios in CRC with liver metastasis compared with rectal or colon cancer, although our study has some limitations, such as small sample size and lack of matched CRC liver metastasis samples. In conclusion, the current findings help define the genomic landscape of CRLM and identify specific pathways that are frequently altered, providing direction for research of targeted therapies against these tumors.

Data Availability

All the related software and scripts used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors have declared no conflicts of interest.

Authors' Contributions

Liuxing Feng, Shifu Hong, and Jin Gao contributed equally to this work. Liuxing Feng, Shifu Hong, Jin Gao, and Jiayi Li designed the project and devised the experiments. Liuxing Feng, Shifu Hong, and Jin Gao performed the experiments and provided tumor samples and clinical information. Jiayi Li dealt with the figures and prepared the main manuscript. All authors contributed to the discussions and manuscript preparation.

Supplementary Materials

Supplemental Table 1: clinical characteristics of the patients. Supplemental Table 2: the number of SNV in the CRLM samples. Supplemental Table 3: list of pathway genes in CRLM. (*Supplementary Materials*)

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA: A Cancer Journal for Clinicians*, vol. 60, no. 5, pp. 277–300, 2015.

[2] T.-M. Kim, S.-H. Jung, C. H. An et al., "Subclonal genomic architectures of primary and metastatic colorectal cancer based on intratumoral genetic heterogeneity," *Clinical Cancer Research*, vol. 21, no. 19, pp. 4461–4472, 2015.

[3] J. 't Lam-Boer, C. Al Ali, R. H. A. Verhoeven et al., "Large variation in the utilization of liver resections in stage IV colorectal cancer patients with metastases confined to the liver," *European Journal of Surgical Oncology*, vol. 41, no. 9, pp. 1217–1225, 2015.

[4] A. Ahmad, J. Reha, P. Somasundar, N. J. Espat, and S. C. Katz, "Predictors of surgical non-referral for colorectal liver metastases," *Journal of Surgical Research*, vol. 205, no. 1, pp. 198–203, 2016.

[5] R. Ciria, S. Ocaña, I. Gomez-Luque et al., "A systematic review and meta-analysis comparing the short- and long-term outcomes for laparoscopic and open liver resections for liver metastases from colorectal cancer," *Surgical Endoscopy*, pp. 1–12, 2019.

[6] H. O. Al-Shamsi, J. Jones, Y. Fahmawi et al., "Molecular spectrum of KRAS, NRAS, BRAF, PIK3CA, TP53, and APC somatic gene mutations in Arab patients with colorectal cancer: determination of frequency and distribution pattern," *Journal of Gastrointestinal Oncology*, vol. 7, no. 6, pp. 882–902, 2016.

[7] R. Sanz-Pamplona, A. Lopez-Doriga, L. Pare-Brunet et al., "Exome sequencing reveals AMER1 as a frequently mutated gene in colorectal cancer," *Clinical Cancer Research*, vol. 21, no. 20, pp. 4709–4718, 2015.

[8] Y.-J. Zhang, Q. Dai, D.-F. Sun et al., "mTOR signaling pathway is a target for the treatment of colorectal cancer," *Annals of Surgical Oncology*, vol. 16, no. 9, pp. 2617–2628, 2009.

[9] W. Wei, F. Jiang, X.-C. Liu, and Q. Su, "TMEM9 mediates IL-6 and IL-1β secretion and is modulated by the Wnt pathway," *International Immunopharmacology*, vol. 63, pp. 253–260, 2018.

[10] World Medical Association (WMA), "Declaration of Helsinki. Ethical principles for medical research involving human subjects," *Jahrbuch für Wissenschaft und Ethik*, vol. 14, no. 1, pp. 233–238, 2009.

[11] C. Chiang, R. M. Layer, G. G. Faust et al., "SpeedSeq: ultra-fast personal genome analysis and interpretation," *Nature Methods*, vol. 12, no. 10, pp. 966–968, 2015.

[12] K. Cibulskis, M. S. Lawrence, S. L. Carter et al., "Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples," *Nature Biotechnology*, vol. 31, no. 3, pp. 213–219, 2013.

[13] M. A. DePristo, E. Banks, R. Poplin et al., "A framework for variation discovery and genotyping using next-generation DNA sequencing data," *Nature Genetics*, vol. 43, no. 5, pp. 491–498, 2011.

[14] K. Wang, M. Li, and H. Hakonarson, "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data," *Nucleic Acids Research*, vol. 38, no. 16, p. e164, 2010.

[15] A. H. Ramos, L. Lichtenstein, M. Gupta et al., "Oncotator: cancer variant annotation tool," *Human Mutation*, vol. 36, no. 4, pp. E2423–E2429, 2015.

[16] A. Mayakonda, D.-C. Lin, Y. Assenov, C. Plass, and H. P. Koeffler, "Maftools: efficient and comprehensive analysis of somatic variants in cancer," *Genome Research*, vol. 28, no. 11, pp. 1747–1756, 2018.

[17] C. Rubio-Perez, D. Tamborero, M. P. Schroeder et al., "In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities," *Cancer Cell*, vol. 27, no. 3, pp. 382–396, 2015.

[18] A. Kassambara, "Ggpubr: 'ggplot2' based publication ready plots," 2018, http://www.sthda.com/english/rpkgs/ggpubr.

[19] R. C. R. Team, *A Language and Environment for Statistical Computing*, 2018, https://www.R-project.org/.

[20] M. Huaiyu, A. Muruganujan, J. T. Casagrande, and P. D. Thomas, "Large-scale gene function analysis with the PANTHER classification system," *Nature Protocols*, vol. 8, no. 8, pp. 1551–1566, 2013.

[21] O. Halit, C. L. Andersen, J. B. Bramsen et al., "Putative cis-regulatory drivers in colorectal cancer," *Nature*, vol. 512, no. 7512, pp. 87–90, 2014.

[22] G. M. de Jong, F. Aarts, T. Hendriks, O. C. Boerman, and R. P. Bleichrodt, "Animal models for liver metastases of colorectal cancer: research review of preclinical studies in rodents," *Journal of Surgical Research*, vol. 154, no. 1, pp. 167–176, 2009.

[23] Y.-S. Ma, T. Huang, X.-M. Zhong et al., "Proteogenomic characterization and comprehensive integrative genomic analysis of human colorectal cancer liver metastasis," *Molecular Cancer*, vol. 17, no. 1, p. 139, 2018.

[24] M. B. Mogensen, M. Rossing, O. Østrup et al., "Genomic alterations accompanying tumour evolution in colorectal cancer: tracking the differences between primary tumours and synchronous liver metastases by whole-exome sequencing," *BMC Cancer*, vol. 18, no. 1, p. 752, 2018.

[25] J. J. Havel, D. Chowell, T. A. Chan et al., "The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy," *Nature Reviews Cancer*, vol. 19, no. 3, pp. 133–150, 2019.

[26] Z. Xu, J. Dai, D. Wang et al., "Assessment of tumor mutation burden calculation from gene panel sequencing data," *OncoTargets and Therapy*, vol. 12, pp. 3401–3409, 2019.

[27] L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge et al., "Signatures of mutational processes in human cancer," *Nature*, vol. 500, no. 7463, pp. 415–421, 2013.

[28] R. Lim, T. Sugino, H. Nolte et al., "Deubiquitinase USP10 regulates notch signaling in the endothelium," *Science*, vol. 364, no. 6436, pp. 188–193, 2019.