The Potential and Analysis of ctDNA Sequencing in Hepatocellular Carcinoma

Yubo Ding

University of South China

Jingwei Yao

University of South China

Jialu Huang

University of South China

Minghui Zhang

University of South China

Yu Zhang

University of South China

Yufan Lv

University of South China

Zhuoyi Xie

University of South China

JianHong Zuo (✉ 632138414@qq.com)

University of South China

Primary research

Keywords: Liver cancer, High-throughput sequencing, Gene mutation, Pathway analysis, Circulating tumor DNA

Posted Date: July 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-693901/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The Genome map of hepatocellular carcinoma (HCC) is complex. We used the next generation of targeted sequencing technology to decipher the mutations in patients with liver cancer

Methods: The circulating tumor cell DNA (ctDNA) of 10 patients with hepatocellular carcinoma (Including 8 cases of primary hepatocellular carcinoma and 2 cases of metastatic hepatocellular carcinoma) were sequenced. We used SAMtools to detect and screen single nucleotide polymorphism (SNP) and insertion deletion (INDEL) mutations, and ANNOVAR to annotate the structure and function of the detected mutations.

Results: Targeted capture and deep sequencing of 560 cancer-related genes in 10 liver cancer ctDNA samples showed 8950 single nucleotide variations (SNVs) mutations and 70 INDELS. The most common mutation gene was PDE4DIP, followed by SYNE1, KMT2C, PKHD1 and FN1. According to the American College of Medical Genetics and Genomics (ACMG) guidelines, we authenticated 54 pathogenic and possible pathogenic mutations in 39 genes in exons and splice regions of 10 patients with HCC.

Conclusion: Our research provides the gene mutation map of Chinese hepatocarcinoma patients, and enriches the understanding of the pathogenesis of HCC.

1. Introduction

The morbidity of HCC is the sixth place of the world's cancer incidence, and the mortality rate is the fourth place of the global cancer deaths, most of which occur in the 50–60 years old population[1]. According to the global cancer statistics report in 2018, there are 841000 new liver cancer patients and 781000 deaths, accounting for 4.7% and 8.2% of the total cancer Morbidity and Mortality: in the world respectively[2]. The pathological types of liver cancer include hepatocyte type, bile duct type and mixed cell type, among which hepatocyte type accounts for about 70%[3]. HCC is one of the most frequent fatal malignant tumors in the world. It is the fifth most common cancer in the world. The standardized mortality rate ranks third among all kinds of malignant tumors, next to gastric cancer and esophageal cancer [4]. In Japan and parts of China, the main risk factors of hepatic carcinoma are chronic hepatitis C virus (HCV) or hepatitis B virus (HBV) [5]. The incidence of hepatocellular carcinoma is increasing, with a particularly high incidence in sub-Saharan Africa and the south-east.[6]. Although the prognosis and treatment of hepatocellular carcinoma have been developing over the years, the morbidity of HCC is still rising. In the past 20 years, progress has been made in elucidating the mechanism of cancer, early diagnosis of disease and improving local and systemic treatment of HCC[7].

The diagnosis of HCC can be realized by using ultrasound and tumor markers to monitor high-risk groups. Until now, alpha fetoprotein (AFP) has obvious positive reaction, but studies have found that AFP sensitivity is low, AFP does not rise in many patients with liver cancer, and even AFP level can be normal in patients with advanced liver cancer[8]. At present, the European Liver Association and the American Association for the Study of Liver Diseases no longer recommend the determination of AFP levels for the diagnosis of liver cancer, and there are some doubts about the diagnostic sensitivity of AFP[9]. The presence or absence of cirrhosis with tumor features, including vascular invasion and portal vein thrombosis, tumor size and alpha-fetoprotein, are important prognostic indicators of HCC, which affects treatment decisions and outcomes[10].

The occurrence of HCC is a progressive, dynamic and multi gene regulated pathological process. Therefore, understanding the molecular mechanism of its pathogenesis and screening according to the relevant factors of high-risk groups can prevent and treat HCC. It is well known that environmental risk factors that may lead to potential cirrhosis include hepatitis B virus, hepatitis C virus, exposure to toxins (such as aflatoxin) and alcohol consumption. Specific gene mutations have been isolated for each cause of HCC[11]. In recent years, in order to improve the prognosis of liver cancer, many researchers are committed to exploring tumor biomarkers. For example, Wei Lu thought that the expression level of TCF21 in HCC was significantly decreased, and it was negatively correlated with the invasive progression of the disease. TCF21 may be a biomarker to predict the prognosis of HCC[12]. Jianguo Qiu et al found that the expression level of LncRNA LOC285194 in tumor tissues was significantly lower than that in adjacent normal tissues, and it was often down regulated in HCC. LncRNA LOC285194 expression is closely related to the occurrence, development, invasion and metastasis of tumor, and has anti-tumor effect, which can be used as a potential target for the development of new therapies for liver cancer[13]. Although many biomarkers have been proposed, the prognosis evaluation of HCC is still a great challenge in clinic. In addition, it is necessary to find new and specific biomarkers for this malignant tumor.

NGS (Next Generation Sequencing) technology is a new gene screening, prognosis and diagnosis technology, and also an effective and acceptable clinical gene detection method[14]. Targeted sequencing is based on high-throughput sequencing, which can simultaneously detect multiple gene mutation types (Including point mutation, insertion / deletion, copy number change and so on.) of multiple cancer species. It is suitable for the detection of any tissue samples and liquid samples. The sequencing included the whole exon region of 560 genes and the hot spot region of TERT gene promoter mutation. All the genes were derived from thousands of classic literatures, the cancer genome census, the authoritative commercial cancer panels, and most of the driving genes in the databases closely related to clinical medication guidance High frequency mutation genes and susceptibility genes are from three classical reviews[15–17], and are supplemented and sorted out by literature collection and reading. This method is based on the Agilent SureSelect targeted sequence capture system, combined with repeatedly optimized probe design and powerful capture efficiency. The target gene has high coverage and strong specificity, which can realize the accurate detection of gene mutation and accurately screen cancer-related mutations.

2. Materials And Methods

2.1 Patients and DNA extraction

This study was approved by the ethics committee of Nanhua hospital. All patients agreed to participate and signed the corresponding informed consent. A total of 10 subjects were included in this study. Nine patients with primary hepatocellular carcinoma, aged 33-68, were collected from January 2019 to December 2019 in our hospital. The diagnostic criteria of primary liver cancer (meeting any of the following three criteria): 1, It has two typical imaging manifestations of liver cancer (ultrasound, CT, MRI or selective hepatic arteriography), with the lesion > 2cm. 2, A typical imaging manifestation, lesions > 2cm, AFP > 400ng / ml. 3, Liver biopsy was positive. Among the patients with primary liver cancer, 1 case was diagnosed by pathological biopsy, and the other 8 cases were diagnosed by clinical diagnosis. Two patients with liver metastases (one with gallbladder cancer and liver metastasis, one with rectal cancer and liver metastasis) were collected. The diagnostic criteria were implemented according to the guidelines for diagnosis and treatment of gallbladder cancer (2015 Edition) and the Chinese code for clinical diagnosis and treatment of colorectal cancer. Patients with primary liver cancer complicated with other tumors such as gastric cancer, lung cancer, cervical cancer, ovarian cancer and prostate cancer, human immunodeficiency virus infection or autoimmune liver disease, alcoholic liver disease, nonalcoholic fatty liver disease, history of other chronic liver diseases, and samples with hemolysis during the test were excluded. According to the inclusion and exclusion criteria, 9 cases of primary HCC group and 2 cases of liver metastasis group were collected. However, 1 case of primary liver cancer group whose ctDNA was not qualified after blood extraction was excluded, and the rest were qualified. There were 8 cases of primary HCC group and 2 cases of liver metastasis group. 5ml venous blood was collected by a special nurse using EDTA anticoagulant tube. After balancing, it was put into a centrifuge (within half an hour of blood drawing), and centrifuged at room temperature for 5 minutes. After centrifugation, plasma was collected and stored in 2 ml EP tube in ultra-low refrigerator at - 80 °C until ctDNA was extracted. Use GeneRead DNA FFPE Kit Qiagen kit to extract ctDNA from plasma (strictly follow the instructions of the kit). Agarose gel electrophoresis was used to analyze the extent of DNA degradation and whether there was contamination of RNA, or protein, and Qubit was used to quantify the DNA concentration (DNA samples with DNA concentration above 20ng/ul and 0.2ug above 0.2ug were used to build the database).

2.2 Target sequencing

Nuohe Zhiyuan uses Agilent's liquid chip capture system to efficiently enrich human specific target region DNA, and then conducts high-throughput and high-depth sequencing on Illumina HiSeq platform. The Agilent SureSelectXT Custom kit was used in the database building and capture experiments. The reagents and consumables recommended in the manual were strictly used, and the latest optimized experimental process was referred to for operation. Genomic DNA fragments with 180-280 BP in length were randomly interrupted by Covaris fragmentation apparatus. After terminal repair and adding A-tail, the ends of the fragments were connected to prepare DNA library. After the library with specific index was pooled, it was hybridized with up to 500000 biotin labeled probes in liquid phase, and then the target gene fragment was captured by magnetic beads with streptomycin. After PCR linear amplification, the quality of the library was tested, and the qualified library could be sequenced. After building the library, we used Qubit2.0 for preliminary quantification, diluted the library to 1ng / μL, and then used Agilent 2100 to detect the insert size of the library. After the insertion size reached the expected value, Q-PCR method was used to accurately quantify the effective concentration of the library (the effective concentration of the library >2nm) to ensure the quality of the library. The Illumina-HiSeq platform was sequenced according to the library's effective concentration and data output requirements.

2.3 Data analysis

The raw image data files obtained by sequencing were filtered by raw reads for low quality, base uncertainty and other factors to get clean reads. The wrong sequencing data was screened out by the sequencing Phred value of each base. The quality of sequencing data was mainly above Q20. The effective sequencing data were aligned to the reference genome (B37) by BWA (Li h et al.) and Samblaster, and the initial alignment results in Bam format were obtained. The BAM file is marked and repeated by Samblaster to get the final comparison result of BAM format. ANNOVAR software was used to annotate the mutation sites, including gene structure annotation, genome feature annotation, nonsynonymous mutation hazard prediction, known mutation database annotation and mutation related gene function annotation. RefSeq and Gencode were used to annotate the gene structure of the mutation site, including mRNA, noncoding RNA, small RNA and microRNA; The genomic characteristics of the mutation sites included CG Island, cell karyotype, phastconsequences46way conserved region, genome repeat, transcription factor binding site and encode annotation of Gm12878 cell lines; Sift, Polyphen, Mutationassessor, LRT and other methods were used to comprehensively evaluate the impact of nonsynonymous mutations on disease / tumor; dbSNP, thousand human genome SNP database, HapMap database, cosmic known tumor somatic mutation database and esp6500 mutation database were provided to screen any combination of mutation results; Go biological process, Go cell components, Go analysis function, KEGG, Reactome, Biocarta, PID and other functional annotation databases were used to interpret signal transduction and metabolic pathways.

3. Results

3.1 Clinical characteristics of patients with liver cancer

Ten patients with liver cancer were identified, including 8 primary liver cancer patients, 1 gallbladder cancer with liver metastasis and one with colorectal cancer complicated with liver metastasis. Table 1 of the patient characteristics summarizes the clinical and pathological data of all patients in this study. The clinical indicators include average age, sex, age range, stage, tumor size, lymph node, distant metastasis, AFP, hepatitis, cirrhosis.

3.2 Gene mutation spectrum

We sequenced the whole exon region of 560 genes in 10 patients with liver cancer and the promoter mutation hotspot of the TERT gene. These 560 genes are the hot genes in cancer research and clinical treatment related genes, including 194 clearly functional drivers, 127 cancer signaling pathway genes (12 genes enriched by KEGG pathway There are 86 high frequency mutation genes, 113 genes, 245 drug sensitivity genes, drug resistance related genes and 15 drug metabolism related genes. Four authoritative drug-related databases (Drugbank, MyCancerGenome, PharmGKB and PG_FDA) and manually sorted drug-

resistant genes were used for annotation, of which 202 genes were labeled as drug target genes by at least one drug database, and 43 genes were labeled drug resistance genes. 15 genes related to drug metabolism were found by drug metabolism pathway enrichment.

The target capture depth sequencing of 560 tumor related genes in 10 ctDNA samples of hepatoma showed that 8950 SNVs mutations were found in 481 genes, while only 20 genes had 70 insertions and deletions. There were 281 gene mutations over 10, 145 gene mutations more than 20. Supplemental Table 2 describes all mutated genes in targeted sequencing of HCC patients. Among them, the most common mutation genes are *pde4dip* (450), followed by *SYNE1* (214), *kmt2c* (119), *PKHD1* (110), *FN1* (105), *LRP1B* (98), *ALK* (97), *FANCA* (95), *Notch1* (92), *ABCC4* (79), *rnf213* (79), *HNF1A* (77), *LAMA2* (76). We analyzed the SNV single nucleotide changes detected and found that in all patients with liver cancer, the changes of G > A, T > C, A > G, C > T were more common than other changes, as shown in Figure 1a.

In addition, we also analyzed the regions of these SNV. Among these variants (Fig. 1b), exon variants (80.91%) were most common, followed by intron variants (13.31%), splicing variants (2.51%), and UTR variants (1.21%). In addition, we detected the genetic effects of these variations in the exon region, including missense variation, synonymous mutation, stop gain / loss mutation and unknown mutation.

3.3 Analysis of the mutation of high frequency mutation gene in primary liver cancer samples after screening

We detected a large number of cancer related gene mutation sites. Comparing with top20 mutation gene in the COSMIC database, we found that 6 gene loci (*PREX2*, *ZFH3*, *FHIT*, *CAMTAL*, *GPHN* and *SND1*) in the high frequency gene did not find mutation sites, and the other 14 genes in the samples were as shown in the Table below. According to the mutation, most of the high frequency mutations of HCC have found mutation sites in ctDNA of primary liver cancer patients, but the mutation frequency is not exactly the same as the reference frequency in cosmic database, which may be related to the insufficient amounts of experimental samples, and may be affected by other factors (somatic release) in the blood fluid of the body.

TP53 and *FOXP1* were the mutation sites after screening, indicating that these mutations may be related to the development of liver cancer. It has certain diagnostic significance and provides possible target drug treatment direction, and has further research significance.

Table. Mutations of HCC high-frequency mutations in patients with primary liver cancer

sample	TP53	CTNNB1	AXIN1	LRP1B	ARID2	KMT2C	ALK	ERBB4	PTPRT	ZNF521	PTPRK	NTRK3	FOXP1	LPP
1	+	-	+	+	+	+	+	+	+	-	+	+	+	-
2	+	-	+	+	+	+	+	+	+	-	+	+	-	+
3	+	-	-	+	+	+	+	+	+	-	+	+	-	+
4	-	-	+	+	-	+	+	-	+	+	+	+	-	+
5	-	+	+	+	+	+	+	+	+	-	+	+	+	+
6	+	-	-	+	-	+	+	-	+	+	+	+	-	+
7	+	+	+	+	-	+	+	-	+	-	+	+	+	+
8	+	-	+	+	-	+	+	+	+	-	-	+	-	+
Mutation frequency	75%	21%	75%	100%	50%	100%	100%	62.5%	100%	25%	90%	100%	37.5%	12.5%
Reference frequency	28%	19%	9%	28%	12%	18%	17%	19%	16%	18%	17%	17%	14%	15%

Note: The blue mutant genes represent the significant genes after sample screening

3.4 Analysis of the gene mutation of high frequency mutation gene in circulating tumor DNA of patients with liver metastasis

In patients with liver metastases, 12 high frequency mutation genes were detected in liver cancer tissues with mutation sites, as shown in the following Table. *FOXP1* did not find mutation site, only gene mutation occurred in primary liver cancer, and the specificity was relatively high in the diagnosis of liver cancer. *KMT2C*/*TP53* was screened in the gene samples of liver metastasis patients, and *KMT2C*/*TP53* met the screening requirements. The mutations of gallbladder cancer and rectal cancer were found in cosmic database, among which *KMT2C* was the high frequency mutation gene site of gallbladder cancer and *TP53* was the high frequency mutation site of rectal cancer, which was in line with the mutation of circulating tumor DNA gene in blood. Therefore, it can be speculated that *KMT2C*/*TP53* may be related to the development of gallbladder cancer / rectal cancer. *KMT2C* may be used as a reference gene for the diagnosis of gallbladder cancer, and *TP53* is a reference gene for the diagnosis of colorectal cancer.

Table. HCC high-frequency mutation in patients with liver metastases

Note: The blue mutant genes represent the significant genes after sample screening

3.5 Pathogenic genes and possible pathogenic genes in filtrate

sample	TP53	CTNNB1	AXIN1	LRP1B	ARID2	KMT2C	ALK	ERBB4	PTPRT	ZNF521	PTPRK	NTRK3	FOXP1	LPP
1	+	-	-	+	+	+	+	+	+	+	+	+	-	+
2	+	-	+	+	-	+	+	+	+	+	+	+	-	+
frequency	28%	19%	9%	28%	12%	18%	17%	19%	16%	18%	17%	17%	14%	15%

According to the guidelines of the American Academy of medical genetics and genomics (ACMG), based on the preliminary analysis of NGS data, through the analysis of Sorting Tolerant From Intolerant(SIFT), PolyPhen, Mutation Taster and Combination Annotation Dependent Deletion (CADD), zygosity, variation type, variation effect, location and filter coverage rate, the results showed that there was no significant difference between them, Small allele frequency and other conditions were screened to further classify and analyze the genes. We found 39 possible mutations related to the development of liver cancer, including RET, TP53 (40% (4/10), LAMA2 (30%), KAT6B, DNMT1, FGFR3, PRKDC, FOXP1, PDGFRB and SYNE1 (20% (2/10).

We also found that some mutation sites do not exist in dbSNP database or cosmic database. 90 significant nonsynonymous mutation were identified, including 72 missense mutations, 2 deletion mutations, 1 code shifting deletion mutation, 8 non shift code deletion mutations, 7 non shift code insertion mutations, which were located in 87 genes. After we summarized these loci, we extracted representative genes, which were AXIN1, CHD7, FN1, LAMA2, NIN, PRKDC, CEPPA, PTCH1 (2/10), PDE4DIP (4/10), SYNE1 (3/10)). The mutation of these genes has further significance for the development of cancer.

We used Davis to analyze the gene ontology (Go) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. The term go is annotated from three aspects: cell component (Go CC), molecular function (Go MF) and biological process (Go BP). Based on KEGG pathway database <https://www.kegg.jp/kegg/> The pathway enrich of is applied to making pathway annotations. The term Go and KEGG pathway were considered statistically significant at $P < 0.05$. PPI network forecast is provided by string online software (<https://string-db.org/>).

3.6 Go notes on the filtered pathogenic genes and possible pathogenic genes

According to ACMG guidelines, we identified 54 pathogenic and possibly pathogenic mutations in 39 genes in exons and splicing regions in 10 patients with liver cancer. (Table 3). The gene ontology annotation and path analysis were carried out on 39 pathogenic genes and possible pathogenic genes by GO annotation of pathogenic genes and possible pathogenic genes (Figure 2 and Supplemental Table 4). BP of these overlapping genes may be related to positive regulation of transcription of RNA polymerase σ promoter, phosphorylation of peptide tyrosine, positive regulation of transcription, positive regulation of GTPase activity, signal transduction, MAPK cascade, negative regulation of apoptosis process, protein phosphorylation, and signal pathway of transmembrane receptor protein tyrosine activator. The results of go MF annotation indicate that these genes are involved in protein binding, ATP binding, protein tyrosine kinase activity, transcription factor activity, sequence specific DNA binding, RAS guanosine exchange factor activity, transcription factor binding, enzyme binding, protein kinase activity, receptor binding, protein heterodimerization activity. In addition, CC of these pathogenic and possibly pathogenic genes is mainly related to the cytoplasmic perinuclear region, receptor complex, cell adhesion connection, cytoskeleton, extracellular components of cytoplasmic side, replication fork, flotillin complex and catenin complex.

3.7 Pathway analysis and PPI protein prediction

Further enrichment analysis based on KEGG database shows that these pathogenic and possibly pathogenic genes are highly enriched in the liver cancer and cancer related pathways, as shown in Table 5. The approaches include: cancer related pathways, such as central carbon metabolism and protein polysaccharides, small RNA in cancer, and various cancers, including thyroid, prostate and pancreatic. The PPI network of 39 pathogenic and possibly pathogenic genes was constructed by using string database. The string map shows that these genes contain 44 nodes and 143 edges. Please refer to <https://www.string-db.org/cgi/network?taskId=bydYaAww2Fdf&sessionId=b8FnnMXJsCm6> for more information.

4. Discussion

In this experiment, a large number of liver cancer related mutation genes were detected by ctDNA extraction, and the level of ctDNA was significantly increased, which was similar to the contrast gene, which was in line with the experimental results. Most of the high frequency mutation genes in HCC group can be found mutation sites, which indicates that the mutation genes in blood ctDNA are basically the same as those in liver cancer tissues. It is of diagnostic value to verify that blood ctDNA does carry the information of the mutation gene in the liver cancer tissue.

In this study, 560 tumor related genes from 10 hepatoma samples were sequenced in depth. We identified 8950 SNV and 70 INDEL in 481 genes, among which PDE4DIP, SYNE1, KMT2C, PKHD1, FN1, LRP1B, ALK, FANCA, NOTCH1, ABCC4, RNF213, HNF1A, LAMA2, GF2R, NIN, PIK3C2B, APC and DOCK8 had more than 70 mutations. 39 pathogenic genes and possible pathogenic genes were identified by ACMG. The results showed that the pathogenic genes and the possible pathogenic genes were analyzed by go enrichment, path analysis and PPI network analysis. Our results may provide some molecular data for the mapping of genetic variation of Chinese hepatoma patients.

The mutation analysis was used to draw genetic variation of liver cancer, and it was found that PDE4DIP mutation was the highest in almost all patients with liver cancer, followed by SYNE1, KMT2C, PKHD1 and FN1. PDE4DIP is a protein encoding gene. The protein encoded by this gene anchors phosphodiesterase 4D to the dictyosome/centrosome region of the cell, participate in microtubule dynamics, promote microtubule assembly, and play a role in the level of dictyosome or centrosome[18, 19]. The gene defect may be one of the reasons for the correlation between myelodysplasia (MBD) and eosinophil proliferation, and also one of the risk factors of adult pineal blastoma[20, 21]. The results of database showed that PDE4DIP protein may interact with

PDE4D, PRLAR2A and AKAP9. These proteins can be combined with cAMP or PKA, suggesting that PDE4DIP may participate in cAMP/PKA signaling pathway. My study suggests that PDE4DIP may be involved in the pathogenesis of liver cancer. In addition, SYNE1 is a coding gene, which encodes an anchor protein, expressed in skeletal muscle, smooth muscle and peripheral blood lymphocytes, and is located in the nuclear membrane, and it participates in the relationship between nuclear layer and cytoskeleton [22]. It has been reported that the gene has higher transcription expression in the beginning, progress and stage of HCC. The encoded protein plays a potential role in the development of HCC and tumor[23]. KMT2C is a protein coding gene, which is a member of ASC-2 / NCOA 6 complex (AsCOM), it has histone methylation activity and participates in transcriptional CO activation[24]. It is reported that KMT2C is frequently mutated in a variety of human cancers, which is crucial for the occurrence and development of most cancers [25]. PRKDC is a member of the PI3 / pi4 kinase family. PRKDC encodes the catalytic subunit of DNA dependent protein kinase (DNA-PK), it plays a role in DNA double strand break repair and recombination together with Ku 70 / Ku 80 heterodimer protein[26]. The expression of PRKDC was significantly correlated with the overall survival rate of HCC, FN1 gene encodes fibronectin, a glycoprotein in the form of soluble dimer in plasma and dimer or polymer in cell surface and extracellular matrix[27] [28]. Fibronectins are involved in cell adhesion, cell motility, opsonization, wound healing, and maintenance of cell shape[29]. Fibronectin is a known biomarker for inchoate diagnosis of HCC, its changes may be an alternative indicator to evaluate the response of patients with early HCC after therapy[30]. Significant mutations of SYNE1, KMT2C, PKHD1 and FN1 were found in 10 HCC patients.

We used the method based on the ACMG mutation classification guidelines in 481 mutation genes, and found 54 pathogenic and possible pathogenic mutations in 39 genes, which may be the cause of liver cancer in these individuals. Go and KEGG analysis were performed to further understand the role of pathogenic mutant genes. GO annotation showed that BP included peptide tyrosine phosphorylation, positive regulation of RNA polymerase II promoter transcription, positive regulation of GTPase activity, protein autophosphorylation, phosphatidylinositol phosphorylation and so on, which may provide a basis for further understanding of the occurrence and development of liver cancer. In addition, the enriched KEGG pathway was found to be involved in hepatitis (hsa05161) and tumor (hsa05200, hsa05230, hsa05206, hsa05205) related pathways, and BCR, RB1, PDGFRB, RET, CASP8, APC, LAMA2, CDH1, ERBB2, CBL, FGFR3, TP53 played a key role in carcinogenesis and tumor progression.

5. Conclusion

Inevitably, there are several shortcomings in this study. Above all, this is a straightforward center research, and requires a large-scale multi agency study to verify the consequences. Secondly, the potential functions and pathways are predicted by bioinformatics merely, which need test check out. Third, because our sample size is small and the ctDNA sequencing is used, there may be errors. But our target sequencing can be used to screen out the hot genes of liver cancer efficiently, quickly and conveniently, and provide new ideas for the diagnosis and prognosis of patients with liver cancer. In conclusion, we sequenced the samples of liver cancer, identified some new mutation gene sites and screened some genes which are of significance for the diagnosis and prognosis of liver cancer. Our research provides the gene mutation map of Chinese hepatocarcinoma patients, and enriches the understanding of the pathogenesis of liver cancer. In addition, we screened some pathogenic genes according to ACMG guidelines, and carried out Go analysis, pathway analysis and PPI network analysis. Notwithstanding, the greater cohort research and experimental research need to be explored in the potential mechanism of HCC.

Abbreviations

INDEL: insertion deletion;

SNP: single nucleotide polymorphism;

ctDNA: circulating tumor cell DNA;

HCC: hepatocellular carcinoma;

ACMG: American College of Medical Genetics and Genomics;

HCV: hepatitis C virus;

HBV: hepatitis B virus;

AFP: alpha fetoprotein;

GO: gene ontology;

CC: cell component;

MF: molecular function;

BP: biological process;

KEGG: Kyoto Encyclopedia of Genes and Genomes;

SIFT: Sorting Tolerant From Intolerant;

CADD: Combined Annotation Dependent Depletion.

Declarations

Ethics approval

This study was approved by the ethics committee of Nanhua affiliated Nanhua hospital.

Availability of data and material

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no conflicts of interest to this manuscript.

Funding: This work is supported by the Key Research Program from the Science and Technology Department of Ningxia Hui Autonomous Region, China (2019BFH02012); the Key Research Program of Hunan Health Committee (20201909); the Program of Hengyang science and Technology Bureau (2017-1, 2020-67);

Authors'contributions

Yubo Ding: Conceptualization, Investigation, Writing - original draft. **Jinwei Yao:** Writing - original draft. **Jialu Huang, Minghui Zhang, Yu Zhang, Yufan Lv, Zhuoyi Xie:** review & editing, cooperater. **Jianhong Zuo:** Supervision, Writing - review & editing, Funding acquisition, Project administration.

Acknowledgments

We thank all the authors for their contributions and support and the study participants who provided blood samples. We also thank the hospital staff for their contribution to the data collection for this study. We thank all programs and organizations for their support

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*. 2015;65(2):87-108. doi:10.3322/caac.21262.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018;68(6):394-424. doi:10.3322/caac.21492.
3. Massarweh NN, El-Serag HB. Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma. *Cancer Control*. 2017;24(3):1073274817729245. doi:10.1177/1073274817729245.
4. Budny A, Kozłowski P, Kaminska M, Jankiewicz M, Kolak A, Budny B et al. [Epidemiology and risk factors of hepatocellular carcinoma]. *Pol Merkuriusz Lekarski*. 2017;43(255):133-9.
5. Kulik L, El-Serag HB. Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology*. 2019;156(2):477-91 e1. doi:10.1053/j.gastro.2018.08.065.
6. Mohammadian M, Allah Bakeshei K, Mohammadian-Hafshejani A. International epidemiology of liver cancer: geographical distribution, secular trends and predicting the future. *J Prev Med Hyg*. 2020;61(2):E259-E89. doi:10.15167/2421-4248/jpmh2020.61.2.1244.
7. Page AJ, Cosgrove DC, Philosophe B, Pawlik TM. Hepatocellular carcinoma: diagnosis, management, and prognosis. *Surg Oncol Clin N Am*. 2014;23(2):289-311. doi:10.1016/j.soc.2013.10.006.
8. Bruix J, Sherman M, American Association for the Study of Liver D. Management of hepatocellular carcinoma: an update. *Hepatology*. 2011;53(3):1020-2. doi:10.1002/hep.24199.
9. Kim SH, Moon DB, Kim WJ, Kang WH, Kwon JH, Jwa EK et al. Preoperative prognostic values of alpha-fetoprotein (AFP) and protein induced by vitamin K absence or antagonist-II (PIVKA-II) in patients with hepatocellular carcinoma for living donor liver transplantation. *Hepatobiliary Surg Nutr*. 2016;5(6):461-9. doi:10.21037/hbsn.2016.11.05.
10. Al-Shamsi HO, Abdel-Wahab R, Hassan MM, Shalaby AS, Dahbour I, Lacin S et al. Natural History of T1N0M0 Hepatocellular Carcinoma: Large-Scale Study in the United States. *Oncology*. 2017;93(4):233-42. doi:10.1159/000455957.
11. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018;391(10127):1301-14. doi:10.1016/S0140-6736(18)30010-2.
12. Lu W, Yang C, Du P, Zhang JL, Zhang JC. Expression tendency and prognostic value of TCF21 in hepatocellular carcinoma. *Artif Cells Nanomed Biotechnol*. 2019;47(1):1466-70. doi:10.1080/21691401.2019.1601102.
13. Qiu J, Pan G, Li M. Decreased expression of long non-coding LOC285194 predicts tumour progression and poor prognosis of hepatocellular carcinoma after curative hepatectomy. *Postgrad Med J*. 2020;96(1132):79-83. doi:10.1136/postgradmedj-2019-136905.

14. Chen Y, Yan W, Xie Z, Guo W, Lu D, Lv Z et al. Comparative analysis of target gene exon sequencing by cognitive technology using a next generation sequencing platform in patients with lung cancer. *Mol Clin Oncol.* 2021;14(2):36. doi:10.3892/mco.2020.2198.
15. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. *Science.* 2013;339(6127):1546-58. doi:10.1126/science.1235122.
16. Rahman N. Realizing the promise of cancer predisposition genes. *Nature.* 2014;505(7483):302-8. doi:10.1038/nature12981.
17. Kandath C, McLellan MD, Vandin F, Ye K, Niu B, Lu C et al. Mutational landscape and significance across 12 major cancer types. *Nature.* 2013;502(7471):333-9. doi:10.1038/nature12634.
18. Yang C, Wu J, de Heus C, Grigoriev I, Liv N, Yao Y et al. EB1 and EB3 regulate microtubule minus end organization and Golgi morphology. *J Cell Biol.* 2017;216(10):3179-98. doi:10.1083/jcb.201701024.
19. Bouguenina H, Salaun D, Mangon A, Muller L, Baudalet E, Camoin L et al. EB1-binding-myomegalin protein complex promotes centrosomal microtubules functions. *Proc Natl Acad Sci U S A.* 2017;114(50):E10687-E96. doi:10.1073/pnas.1705682114.
20. Chang YC, Lin HC, Chiang YH, Chen CG, Huang L, Wang WT et al. Targeted next-generation sequencing identified novel mutations in triple-negative myeloproliferative neoplasms. *Med Oncol.* 2017;34(5):83. doi:10.1007/s12032-017-0944-z.
21. Snuderl M, Kannan K, Pfaff E, Wang S, Stafford JM, Serrano J et al. Recurrent homozygous deletion of DROSHA and microduplication of PDE4DIP in pineoblastoma. *Nat Commun.* 2018;9(1):2868. doi:10.1038/s41467-018-05029-3.
22. Indelicato E, Nachbauer W, Fauth C, Krabichler B, Schossig A, Eigentler A et al. SYNE1-ataxia: Novel genotypic and phenotypic findings. *Parkinsonism Relat Disord.* 2019;62:210-4. doi:10.1016/j.parkreldis.2018.12.007.
23. Faraj Shaglouf LH, Ranjpour M, Wajid S, Jain SK. Elevated expression of cellular SYNE1, MMP10, and GTPase1 and their regulatory role in hepatocellular carcinoma progression. *Protoplasma.* 2020;257(1):157-67. doi:10.1007/s00709-019-01423-w.
24. Cho SJ, Yoon C, Lee JH, Chang KK, Lin JX, Kim YH et al. KMT2C Mutations in Diffuse-Type Gastric Adenocarcinoma Promote Epithelial-to-Mesenchymal Transition. *Clin Cancer Res.* 2018;24(24):6556-69. doi:10.1158/1078-0432.CCR-17-1679.
25. Chen X, Zhang G, Chen B, Wang Y, Guo L, Cao L et al. Association between histone lysine methyltransferase KMT2C mutation and clinicopathological factors in breast cancer. *Biomed Pharmacother.* 2019;116:108997. doi:10.1016/j.biopha.2019.108997.
26. Chaplin AK, Hardwick SW, Liang S, Kefala Stavridi A, Hnizda A, Cooper LR et al. Dimers of DNA-PK create a stage for DNA double-strand break repair. *Nat Struct Mol Biol.* 2021;28(1):13-9. doi:10.1038/s41594-020-00517-x.
27. Qi Z, Yan F, Chen D, Xing W, Li Q, Zeng W et al. Identification of prognostic biomarkers and correlations with immune infiltrates among cGAS-STING in hepatocellular carcinoma. *Biosci Rep.* 2020;40(10). doi:10.1042/BSR20202603.
28. Liu B, Ding Y, Li P, Wang T, He S, Jia Z et al. MicroRNA-219c-5p regulates bladder fibrosis by targeting FN1. *BMC Urol.* 2020;20(1):193. doi:10.1186/s12894-020-00765-5.
29. Cai X, Liu C, Zhang TN, Zhu YW, Dong X, Xue P. Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion. *J Cell Biochem.* 2018;119(6):4717-28. doi:10.1002/jcb.26651.
30. Kim SA, Cho EJ, Lee S, Cho YY, Kim B, Yoon JH et al. Changes in serum fibronectin levels predict tumor recurrence in patients with early hepatocellular carcinoma after curative treatment. *Sci Rep.* 2020;10(1):21313. doi:10.1038/s41598-020-78440-w.

Tables

Due to technical limitations, table 1, 3 and 5 is only available as a download in the Supplemental Files section.

Figures

Figure 1

Targeted cell SNVs by next-generation sequencing. (A) Percentages of different transitions and transversions in SNV. (B) Proportion of SNV types by gene region. SNVs, single nucleotide variants.

Figure 2

Top 10 enrichment fractions in GO enrichment analysis of 39 pathogenic and potential pathogenic genes. (a) Pathogenic genes and biological processes of possible pathogenic genes; (b) Pathogenic genes and cellular components of possible pathogenic genes; (c) Molecular functions of pathogenic genes and potential pathogenic genes.

Figure 3

PPI networks of 39 pathogenic and potential pathogenic genes. Protein interaction

Supplementary Files

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

- [SupplementalTable2.xlsx](#)
- [SupplementalTable4.xlsx](#)
- [Table.pdf](#)
- [Table3Supplementarymaterials.xlsx](#)
- [Table5Supplementarymaterials.xlsx](#)