Genes & Diseases (2020) 7, 380–391

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: http://ees.elsevier.com/gendis/default.asp

FULL LENGTH ARTICLE

Whole-exome mutational landscape of metastasis in patient-derived hepatocellular carcinoma cells

Qian Zhou [a,b,c,1], Zuli Li [a,b,c,1], Linlan Song [a,b,c], Di Mu [a,b,c], Jin Wang [a,b,c], Li Tian [a,b,c], Yong Liao [a,b,c,*]

[a] *Key Laboratory of Molecular Biology on Infectious Diseases, Ministry of Education, Chongqing, China*
[b] *Institute for Viral Hepatitis, Chongqing Medical University, Chongqing, China*
[c] *Department of Infectious Diseases, Second Affiliated Hospital, Chongqing Medical University, Chongqing, China*

Received 10 February 2020; received in revised form 29 April 2020; accepted 9 May 2020
Available online 15 May 2020

KEYWORDS
Clonal evolution;
Encyclopedia of genes and genomes (KEGG);
Gene ontology (GO);
Genome-wide association;
Hepatocellular carcinoma;
Metastatic potentiality;
Somatic gene mutation;
Whole exome sequencing

Abstract In order to explore the genomic basis for liver cancer metastasis, whole-exome sequencing (WES) was performed on patient-derived hepatocellular carcinoma (HCC) cell lines with differential metastatic potentials and analyzed their clonal evolution relationships. An evolutionary tree based on genomic single nucleotide polymorphism (SNP) was constructed in MegaX software. The WES data showed that the average percentage of heterogeneous mutations in each HCC cell lines was 16.55% (range, 15.38%–18.17%). C: G > T: A and T: A > C: G somatic transitions were the two most frequent substitutions. In these metastatic HCC cell lines, non-silent gene mutations were found in 21.88% of known driver genes and 10 classical signaling pathways. The protein interaction network was constructed by STRING, and hub genes were found in the shared trunk mutation genes and the heterogeneous branch mutations respectively. In cBioPortal database, some of the selected hub genes were found to be associated with poor overall survival (OS) of HCC patients. Among the mutated HCC driver genes, a novel KEAP1 mutation with a homozygous frameshift truncation at the c-terminal Nrf2 binding region was detected and verified in MHCC97-H and HCC97LM3 cells. In conclusion, WES data demonstrate that HCC cell lines from tumor biopsy specimens of the same patient have

Abbreviations: bZIP, basic-region leucine zipper; BTB, Broad-complex, Tramtrack, and Bric-a-brac; CDS, coding for amino acids in protein; CNC, cap'n'collar; CTR, C-terminal region; CUL3, Cullin3; DGR, DC domain harboring six Kelch-repeat domain; FA, fatty acid; GO, Gene Ontology; HCC, hepatocellular carcinoma; IVR, intervening region; KEGG, Kyoto Encyclopedia of Genes and Genomes; NTR, N-terminal region; OS, overall survival; SNP, single nucleotide polymorphism; WES, whole exome sequencing.

* Corresponding author. Key Laboratory of Molecular Biology on Infectious Diseases, Ministry of Education, Chongqing, China.
E-mail addresses: yongliao@cqmu.edu.cn, y8982000@yahoo.com (Y. Liao).
Peer review under responsibility of Chongqing Medical University.
[1] The two authors have equal contribution.

https://doi.org/10.1016/j.gendis.2020.05.003

obtained different metastatic potentials through repeated selection in rodents in vivo, and they do indeed have a genetic relationship at the genomic level.

Introduction

Hepatocellular carcinoma (HCC) is a common malignancy in men and a leading cause of cancer death, second only to lung cancer in males.[1] In 2018, approximately 841,080 new HCC cases and 781,631 deaths worldwide were reported due to HCC.[1,2] Although our understanding of the etiology and genetic events leading to HCC has been improved, in addition to the latest advances in diagnosis and precision microsurgery, HCC 's overall 5-year survival rate is still less than 18%.[3] The major obstacle for the improvement of patients' overall survival is the intra-neoplastic heterogeneity of HCC cells, which have been attributed to the anti-cancer resistance, local recurrence and distant metastasis in patients with liver cancer.[4]

According to Peter Novell's clonal evolutionary theory, cancer cells originate from a single cell with accumulated germ-line or somatic mutations and epigenetic alterations that lead to transformation of a normal cell into a malignant cell.[5] From an evolutionary perspective, a neoplasm is viewed as a large, genetically and epigenetically heterogeneous population of individual cells. Whilst fitness selection enables the elimination of random mutations that are harmful to neoplastic cells, genetic and epigenetic alterations beneficial to neoplastic clone survival are positively selected, leading to the expansion of heterogeneous subpopulations with differing invasive and metastatic capabilities.[6,7]

Previous reports demonstrate a higher genetic complexity of HCC amongst different patients (inter-tumor heterogeneity) and within the same tumors (intra-tumor heterogeneity).[8,9] One of the widely used high metastatic HCC model systems, MHCC97-L, MHCC97-H and HCC97LM3 cell lines, was originally derived from a patient xenograft model LCI-D20.[10] Through step-wise and repeated *in vivo* selection in rodents, these HCC cell lines demonstrate different metastatic potentiality and organ-tropic proclivity when inoculated orthotopically into mouse livers, with MHCC97-H showing lung metastasis in 100% of cases, MHCC97-L showing 40%, and HCC97LM3 showing the capacity to metastasis to the lungs when inoculated subcutaneously in mice.[11,12] However, although these cell lines are used worldwide, the genetic background that gives them differential metastatic traits remains unclear. In this study, we utilized whole exome-sequencing (WES) technology to reveal the genetic basis of liver cancer metastasis. Based on WES data, we explored the kinship and evolution of these HCC cell lines.

Materials and methods

Cell culture

MHCC97-L, MHCC-97H and HCC97LM-3 cells were obtained from the Liver Cancer Institute of Fudan University. Cells were cultured in DMEM (Hyclone) medium supplemented with 10% fetal bovine serum (Gibco), 1% streptomycin and penicillin at 37 °C in a 5% CO_2 atmosphere.

DNA extraction

Cells grown in 10 cm plates were rinsed with phosphate buffered saline (PBS) and lysed in 1 ml of DNA extract buffer (10 mmol/L Tris (hydroxymethyl) aminomethane, 0.1 mol/L Ethylene Diamine Tetraacetic Acid, 0.5% Sodium dodecyl sulfate). Proteinase K (final concentration of 100 ug/ml) was added to the samples at 56 °C for 1 h followed by the addition of an equal volume of saturated phenol. Samples were centrifuged at 12,000 rpm for 10 min and added 1/10 to sodium acetate (3 M) for 1 h at −20 °C. Samples were centrifuged at 12,000 rpm for 10 min and washed in 75% ethanol on ice for 15 min. Samples were centrifuged at 12,000 rpm for 5 min, air dried and dissolved in RNAase free water or Tris—EDTA (TE) buffer at a final concentration of 20 ug/ml. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added and centrifuged at 12,000 rpm for 10 min. The purification procedures were then repeated and samples were finally dissolved in 20 μL of TE buffer after air-drying.

Quantitative, qualitative DNA assessment and whole-exome sequencing

DNA samples were assessed using the Applied Biological Materials Inc (ABM) and DNA quality was confirmed on a Nanodrop. Accurate DNA quantifications were performed using TaqMan® RNase P Detection qPCR Assays (Life Technologies) and Ion AmpliSeqTM Exome Kits (Life Technologies) for exome targeted PCR amplification, primer digestion, adaptor ligation, and qPCR library quantification. Emulsion PCRs were performed using the Ion PI™ Template OT2 200 Kit v2 (Life Technologies) and semiconductor sequencing was performed using Ion PI™ Sequencing 200 Kit v2 with XT Reagents (Life Technologies). Samples were subjected to exome target amplification, primer digestion, adaptor ligation, qPCR library quantification, emulsion PCR, and ion torrent semiconductor sequencing. Two sequencing

runs were performed and BAM alignment data were generated immediately after the run.

Annotation of variants with oncotator

Oncotator is an open-source and broad scope database that permits the annotation of genomic point mutations and short nucleotide insertions/deletions (indels) with variant- and gene-centric information.[13] To annotate the variants in the HCC cell lines, Oncotator was applied to annotate the alternative alleles for variants. WES data were imported into Oncotator in the formats of "Chromosome", "Position", "Reference" and "Variant"(Reference sequence: GRCh37.p13, *Homo sapiens*), and mutation annotations were provided by Oncotator (MAF format mutation annotation file) for "Hugo Symbol", "Chromosome", "Variant Classification", "Reference Allele", "Tumor Seq Allele1" and "Tumor Seq Allele2", "dbSNP RS". Non-silent mutations are of great significance for the functional analysis of mutated genes, so subsequent analysis focuses only on non-silent mutations. According to UCSC Genome Browser (http://genome.ucsc.edu), non-silent mutations including "Stop Codon Ins", "Stop Codon Del", "Start Codon Ins", "Start Codon Del", "Splice Site", "Nonsense Mutation", "Missense Mutation", "In Frame Ins", "Frame Shift Ins", and "Frame Shift Del".

Constructing evolutionary tree and venn diagram

To illustrate the clonal heterogeneity of MHCC97-L, MHCC97-H and HCC97LM3 cells, an evolutionary tree based on genomic single nucleotide polymorphism (SNP) was constructed in MegaX software.[14] Neighbor-Joining algorithm was used to construct evolutionary tree, and P-distance model was used to calculate distance. Venn diagram was utilized to demonstrate the relationship between mutation sites and gene composition in the three HCC cell lines. Gene names, mutation positions and mutation bases were used to compile the sequence number of the mutation position. For example, the gene name is *ARVCF*, the position is 19,959,473, the mutation base is T, and the sequence number of the mutation position is "*ARVCF*19959473T", and then venn diagram was drawn by using all the sequence number of the mutation position in each cell line. Private branch mutations were classified as those observed in one particular cell line only. Mutations in known driver genes and 10 canonical pathway in HCC were further analyzed.[15] According to the mutation data annotated by Oncotator, non-silent mutation types were selected in excel software, and then non-silent mutant genes in driver genes and 10 signal pathway genes were screened.

Functional enrichment analysis

Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using Annotation, Visualization, and Integrated Discovery (DAVID) online tools. DAVID is an online database and analytic tool that provides information regarding the functional annotation of large lists of genes or proteins.[16] The selected gene sets were organized based on GO enrichment analysis and KEGG biochemical pathways.[17] The molecular functions, biological processes, cellular components and pathways with P-values $\leq$ 0.05 amongst the selected genes were used to explore potential information.[18] All information is visualized using DAVID.

PPI network generation/visualization and module analysis

Based on the gene variants, shared trunk gene mutation and branch private gene mutations with different numbers and sites of mutations in MHCC97-L, MHCC97-H and HCC97LM3 cell lines were collected. Selected protein maps were drawn using online STRING database (Search Tool for the Retrieval of Inter-acting Genes/Proteins).[19] Enter the gene names in STRING database, select Organism as "*Homo sapiens*", select the minimum interaction score as "medium confidence (0.400)", and then export the data as "TSV format". The data in TSV format is integrated into the gene network related to protein targets and visualized with Cytoscape (version 3.6.0). The centiscape2.2 app was used to count the degrees of the PPI network in Cytoscape. Hub genes were collected into STRING with confidence scores $\geq$0.4. GO and KEGG pathway analysis was used to study the potential functions of genes.

Cancer genomics and survival analysis using the cBio Cancer Genomics Portal

The cBio Cancer Genomics Portal (http://www.cBioPortal.org) provides information regarding cancer mutations through exploring, visualizing, and analyzing multidimensional cancer genomic data. cBio portal was used to explore the percentage of alterations with hub genes in these three tumor cell lines and datasets with "Liver cancer" was performed. The details of the genetic changes are displayed in OncoPrint. In addition, the P value of OS was evaluated by Kaplan Meier-plot.

Expression and survival analysis of hub genes in liver cancer

UALCAN database (http://ualcan.path.uab.edu/analysis.html), is a comprehensive web resource for analyzing cancer OMICS data,[20] and is used for the expression and survival analysis of hub genes in liver cancer. Enter the gene name in UALCAN database, select the cancer type as "liver hepatocellular carcinoma ", and click "Expression" and "Survival" to obtain the expression level and survival analysis. The p value is provided accordingly in the database.

KEAP1 gene mutation analysis and PCR verification

The DNA sequence of the mutated gene is derived from exon sequencing data. PRIMER PLUS 3 (http://www.primer3plus.com/cgi-bin/dev/primer3plus.cgi) and PRIMER 5 were used to design primers for mutation sites, and the specificity of primers was verified in PRIMER-BLAST (https://www.ncbi.

nlm.nih.gov/tools/primer-blast/). In the mutations of *KEAP1* gene, a homozygous frameshift deletion mutation at c.1334delC (p.P445fs) in exon 4 was observed and PCR further verified by PCR. Specific PCR primers for the mutated *KEAP1* gene were designed based on the location and mutation sequence of the *KEAP1* gene. 300 ng of genomic DNA per sample were subjected to PCR amplification with specific primers and 2X Power Taq PCR Mastermix (Bioteke,China) in the T100 Thermal Cycler (BIO-RAD,USA). The PCR cycling conditions were 94 °C (4 min) for one cycle, 94 °C (45s), 57 °C (45 s), and 72 °C (1 min) for 39cycles, and a final extension of 72 °C (5 min). All reactions were performed in triplicate. The primers used in this study are as follows: Mutated *KEAP1* forward 5′-ccatactcgtctcgccctac-3′ and reverse 5′-ggaggacccaagttcgtcaa -3′.

To further analysis the mutated *KEAP1* gene verified by PCR, wild type human *KEAP1* gene sequence was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.25/). The coding sequence (CDS) for the wild type *KEAP1* protein in FASTA format was used for the query and aligned with the mutated *KEAP1* sequence obtained from the WES data. The translation product from the CDS region of the mutated *KEAP1* was analyzed using an open reading frame finder (https://www.ncbi.nlm.nih.gov/orffinder/). The amino acid sequence similarity between the mutated *KEAP1* protein and the wild type *KEAP1* protein in the NCBI database was aligned by using PROTEIN BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Results

Genetic profile analysis of metastatic HCC cell lines by WES

Genomic DNA samples were extracted from metastatic HCC cell lines and the quality of purified DNA preparations was assessed on Nanodrop. DNA concentrations were confirmed by TaqManTM RNase qPCR (Supplemental Table 1). The coverage depth of the samples was up to 100 X (Supplemental Table 2). The HCC cell lines were sequenced with WES on the Illumina platform. Among the non-silent mutations, 9293, 9608 and 9364 mutations (5255, 5372, 5297 mutation genes) were detected in MHCC97-L, MHCC97-H and HCC97LM3 cell lines respectively. More than 99% of the variations is in the coding regions, and 0.75%—0.80% is in the no-coding regions. Sequencing data were annotated using Oncotator and 10 non-silent mutation gene mutation types were identified and summarized in Supplemental Table 3. The main alterations observed in the WES data are missense mutations and frameshift del mutations.

To elucidate the evolutionary trajectory of the HCC cells, the venn diagram was mapped according to the mutation site. There are 4691 shared trunk mutant genes in these three HCC cell lines, regardless of their metastatic potential (Fig. 1A). Mutational landscapes indicated that HCC97LM3 and MHCC97-H retain more common genomic mutations than MHCC97-L, suggesting that both cell lines may have been evolved from MHCC97-L. Evolutionary tree (Fig. 1B) demonstrates the evolutionary relationship of the three HCC cell lines, MHCC97-H and HCC97LM3 appear to be closer in evolutionary relationship. The average percentage of heterogeneous mutations in each HCC cell line was 16.55% (range: 15.38%—18.17%). Among these three HCC cell lines, MHCC97-H has the highest branch private mutation (18.17%) (Fig. 1B).

Next, the types of base substitution in three HCC cell lines were analyzed. Amongst all the substitution mutations, C: G > T: A and T: A > C: G somatic mutations are predominant (Fig. 1C). Among the three cell lines, C: G> T: A accounted for 34.43% of the shared trunk mutations, and about 17.74% of the branch private mutations. T: A > C: G mutations accounts for 31.24% of the base substitution mutations shared by the three cell lines and 17.31% of the branch private mutations, respectively.

Alterations in driver mutations and oncogenic signaling pathways

Previous report showed that 32 driver genes play an important role in the occurrence and development of HCC based on analysis of exon group data of cancer genome atlas (TCGA).[15] To further explore the genomic alterations in these metastatic HCC cells, driver gene mutations were retrieved and analyzed. Seven out of the 32 reported driver gene mutations (21.88%) were observed in the three HCC cells, including APOB, P53, Keap1, CDKN1A,XPO1,ARID1A and HIST1H1E. No mutations were detected in the remaining 25 driver genes (*IDH1*, *NUP133*, *NFE2L2*, *ARID2*, *PIK3CA*, etc.) in the MHCC97 series (Fig. 2A). Among driver gene mutations, the P53 mutation patterns detected in these HCC cells include one heterozygous nonsense mutation (c.151G > T, p.E51*), and one heterozygous missense mutation (c.747G > T, p.R249S). The non-sense and missense mutations of P53 overlapped with the database Catalogue of Somatic Mutations in Cancer (COSMIC).[21] Interestingly, *ARID1A* heterozygous missense mutation (c.850G > C, p.G284R) was detected only in MHCC97-L, and HIST1H1E homozygous frameshift deletion mutation (c.426delG, p.T142fs) was only in HCC97LM3, which may reflect their respective clonal evolution. On average, 87.10% (range, 81.82%—90.00%) of the driver mutations are trunk gene mutations common to the three cell lines, while 12.90% (range, 10.00%-18.18%) are in the branched private mutations.

Functionally, the mutated genes detected in the HCC cell lines were categorized into 10 classical oncogenic signaling pathways, including cell cycle regulation, hippo signaling, β-catenin/Wnt signaling, Myc signaling, Notch signaling, p53 activity, oxidative stress responses, Nrf2, PI-3-Kinase signaling, TGF β signaling and receptor-tyrosine kinase (RTK)/RAS/MAP-Kinase signaling (Fig. 2B).[22] According to TCGA program, the mutations in receptor-tyrosine kinase (RTK)/RAS/MAPK signaling had the highest mutation frequency in HCC cells. TGFβ signaling however, had the lowest overall frequency of alterations.

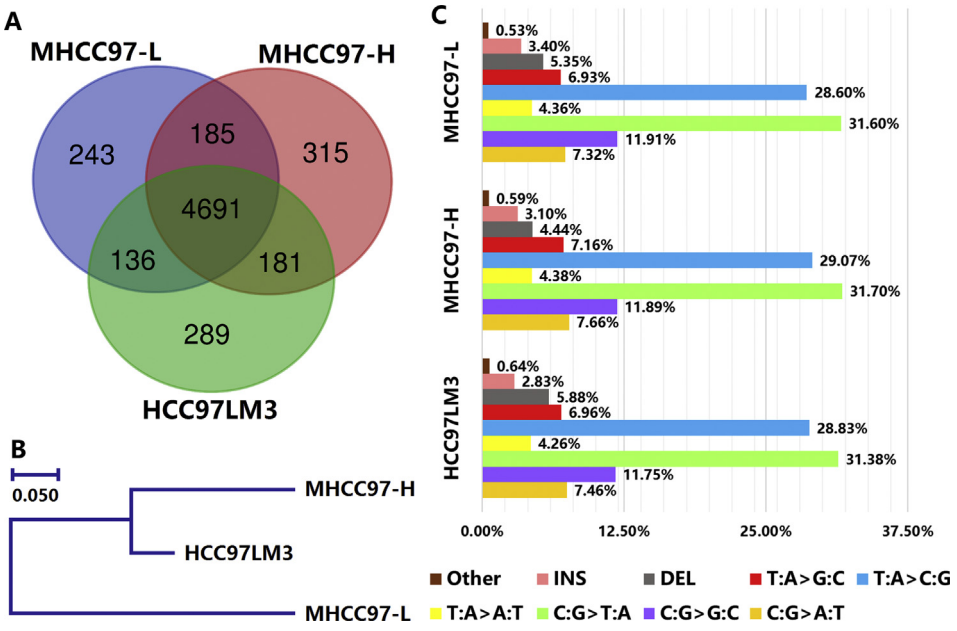

Figure 1 Evolutionary Relationship of Metastatic HCC Cell Lines Revealed By WES. (A) The data of Venn diagram are extracted from WES, which demonstrates the total number of overlapping non-silent mutation genes in the three metastatic HCC cell lines. (B) A SNP-based phylogenetic tree illustrates the heterogeneity and evolutionary relationship between the three HCC cell lines. (C) Distribution of somatic substitution patterns among the three HCC cell lines.

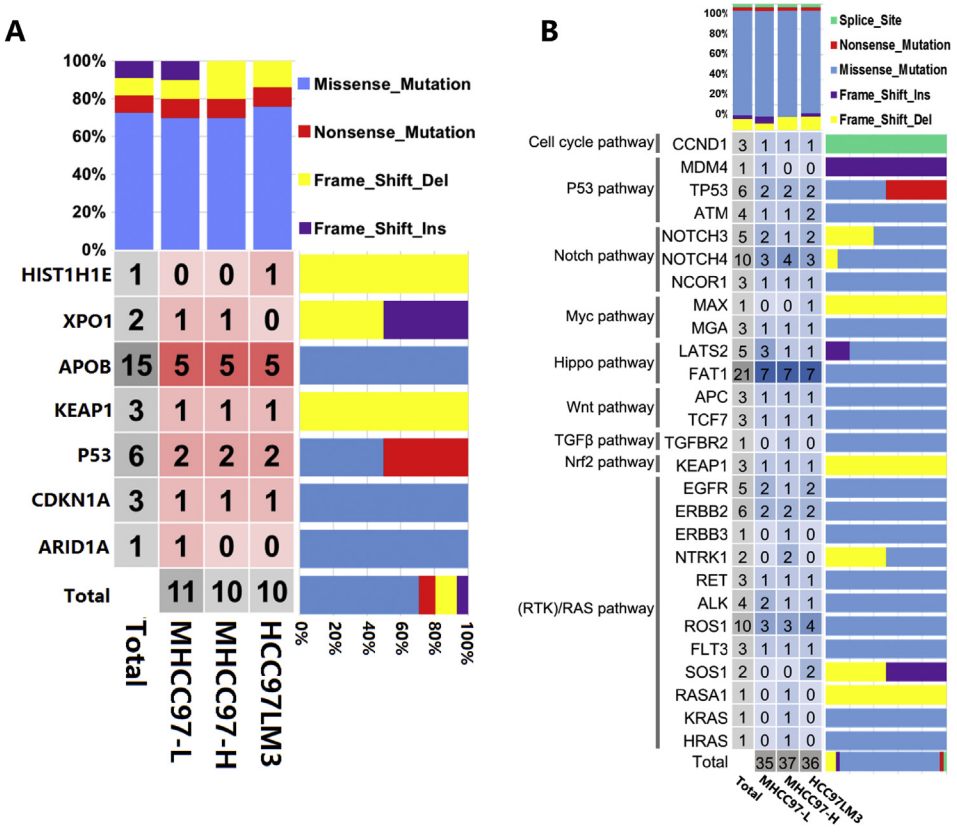

Figure 2 Non-silent Mutations of Liver Cancer Driver Genes and 10 Classical Signaling Pathway Genes in Metastatic HCC Cell Lines. (A) Distribution and mutation types of driver genes in the metastatic HCC cell lines. (B) The distribution and mutation types of mutant genes in 10 classical signaling pathways.

Functional annotation and feature analysis of shared trunk gene mutations

A total of 1054 mutant genes with different mutation sites or number in each of the genes categorized as the shared trunk mutations in the three cell lines were further analyzed, and their mutation frequency, mutation incidence and mutation pattern were re-evaluated by functional enrichment analysis. Some of these genes are classified as biological processes (BP), such as extracellular matrix tissue and collagen catabolism. And some of these genes are enriched in extracellular matrix structural components and microtubule motor activity through molecular function (MF) analysis. Whereas, GO cell composition (CC) analysis found that these shared trunk mutation genes involved microtubules, integrin complexes and collagen trimers (Supplemental Table 4). The most significant enrichment of differentially mutated genes in the KEGG pathway shared by the metastatic HCC cells is shown in Fig. 3A. These mutated genes are mainly enriched in PI3K-Akt signaling, focal adhesions, extracellular matrix—receptor interactions, amoebiasis and protein digestion and absorption, and so on.

In the PPI networks, the top 14 hub genes with the highest degree of connectivity were identified (degree $>$ 30) (Fig. 3B). KEGG pathway enrichment analysis demonstrated that the 14 hub genes are related to ECM-receptor interaction, focal adhesion and PI3K-Akt

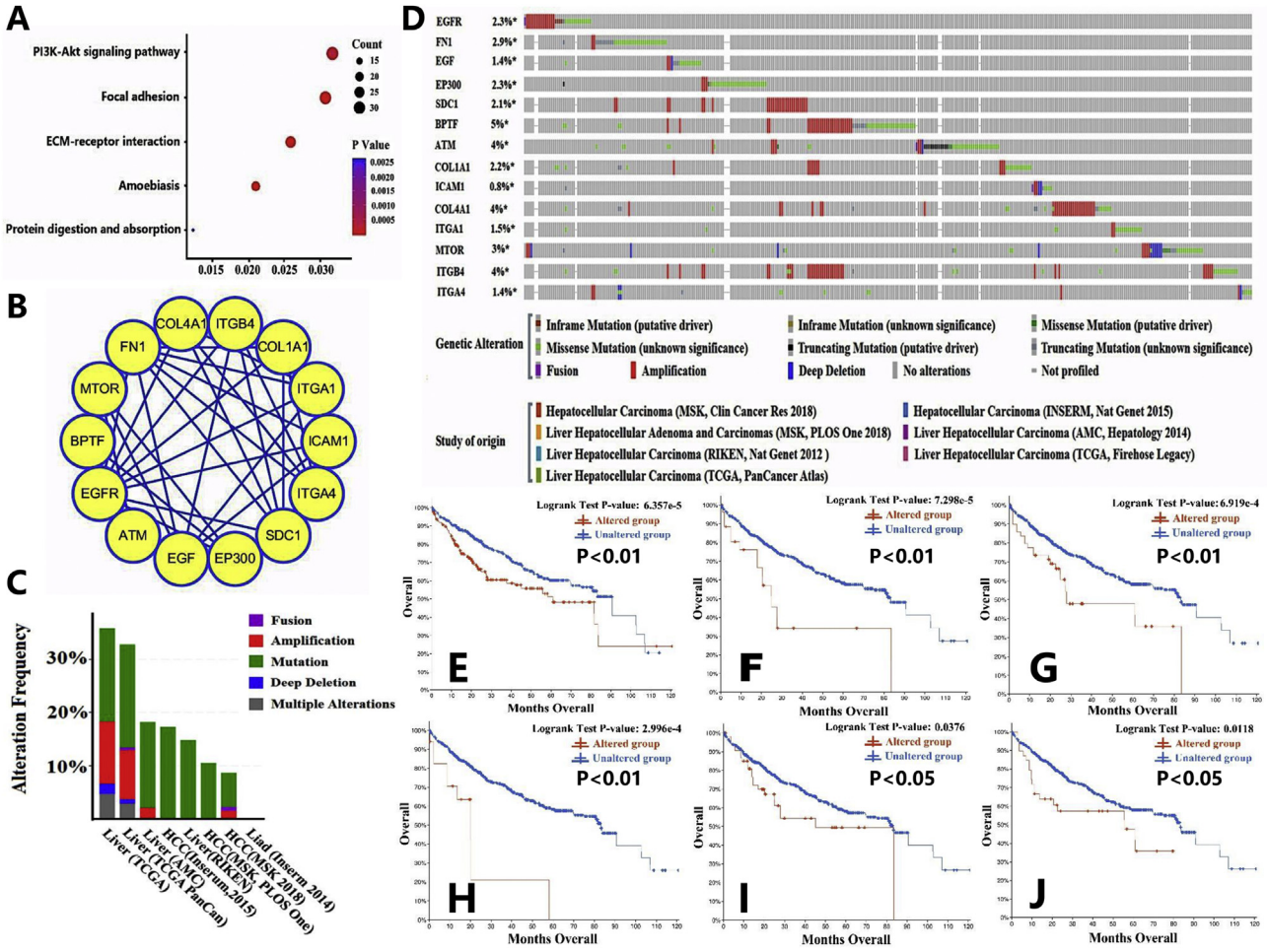

Figure 3 Trunk Mutations Shared in Metastatic HCC Cell Lines. (A) Scatter plots of 1054 differentially mutated genes analyzed through the KEGG pathway. The X axis represents the proportional variable, and the Y axis represents the terms of the KEGG pathway. The size and color of the dots represent the number of mutant genes and the range of P value, respectively. (B) A protein—protein interaction network of 14 hub genes identified from 1054 differentially mutated genes. (C) Statistical analysis of 14 hub gene mutations from 7 different hepatocellular carcinoma cohorts in cBio Cancer Genome Portal database. (D) Query the 14 hub genes from 7 different hepatocellular carcinoma cohorts in cBio Cancer Genome Portal database. Different genomic changes, such as mutation, amplification, fusion and deep deletion, were summarized, and the percentage changes of particularly affected genes in individual tumor samples were shown by color coding. Each column represents a tumor sample and each row represents a hub gene, respectively. Red bars designate gene amplifications, blue bars represent deletions, green squares indicate missense mutations, black bars stand for truncating mutations, purple bars demonstrate fusion and grey bars show no alteration or not profiled. (E-J) The survival curves demonstrate the overall survival of hepatocellular carcinoma patients with or without mutations of the 14 hub genes (E) and some of the individual hub genes, such as *SDC1* (F), *ITGB4* (G), *ITGA4* (H), *BPTF* (I) or *COL4A1* (J).

signaling, including EGFR, EGF and FN1. In order to explore the correlation between these mutation data and clinical parameters, cBio portal analysis was performed and it was found that the mutated hub gene had an effect on the overall survival rate (OS) of patients with hepatocellular carcinoma. According to the query of these 14 hub genes in the 7 databases of the HCC research cohorts, 0.8%−5% of HCC cases show that these hub genes have alterations (Fig. 3C). Next, OncoPrint was used to evaluate multiple genetic changes in each tumor sample. Of the 1507 sequenced cases, a total of 357 cases (24%) demonstrate mutations in at least one of the 14 hub genes queried. Fig. 3D shows the frequency of alterations in each of the 14 hub genes.

cBioportal database was used to plot survival analysis outputs for patients with or without gene mutations. Among patients with HCC, the OS of patients with these 14 hub gene mutations was significantly shorter than those without these hub gene mutations (median, 60.84 months versus 90.64 months; P = 6.357e-5) (Fig. 3E). Moreover, in HCC patients with 5 hub gene mutations, including SCD1, ITGB4, ITGA4, BPTF and COL4A1, their OS was shorter than patients lacking these 5 hub gene mutations (SCD1, median, 24.89 months vs. 90.64 months, P = 7.298 e−5; ITGB4, median, 27.88 months vs. 90.64 months, P = 6.919 e−4; ITGA4, median, 20.11 months vs. 90.64 months, P = 2.996 e−4; BPTF, median, 45.07 months vs. 83.18 months, P = 0.0376; COL4A1, median, 55.69 months vs. 83.18 months, P = 0.0118, respectively) (Fig. 3F−J).

Functional annotation and feature analysis of branch private gene mutations

For functional characterization of mutated genes, David software was used to cluster private gene mutations in the three cell lines (Fig. 4A). Around 243 private gene mutations were identified in MHCC97-L cells, which are functionally enriched in chromatin DNA binding and core promoter binding through MF analysis. These genes may be involved in kinesin complex formation as demonstrated by CC analysis ($P < 0.05$). Interestingly, KEGG pathway analysis found that these private gene mutations are mainly concentrated in oxytocin signaling pathway. About 315 private gene mutations were identified in MHCC97-H and they were mainly enriched in DNA binding, actin-dependent ATPase activity and microfilament motor activity by MF analysis. Whereas, private gene mutations enriched in pore complex and in Huntington's disease were observed by CC analysis and KEGG pathway analysis, respectively. Analysis of 289 private gene mutations in HCC97LM3 by David software revealed that these private mutant genes were mainly enriched in learning, positive regulation of excitatory postsynaptic potential and scaffold protein binding. Although, enrichment in chronic myeloid leukemia and Toll-like receptor signaling pathway were observed by KEGG pathway analysis, no significant enrichment was obtained by CC analysis.

Through the PPI network constructed by STRING, the top 8 hub genes were obtained in these liver cancer cell lines (Fig. 4B). KEGG pathway analysis showed that only the hub gene identified in HCC97LM3 cells was significantly enriched in chronic myeloid leukemia, and it was significantly associated with tumorigenesis. OncoPrint analysis showed that the mutation rates of hub genes of MHCC97-L, MHCC97-H and HCC97LM3 were 0.2%−2.5%, 0.4%−2.2% and 0.3%−5%, respectively (Fig. 4C).

Through a cBio portal analysis of the relationship between clinical characteristics of liver cancer patients and changes in the expression of these private hub genes, we found that patients with hub gene mutations identified in MHCC97-L cells had significantly lower OS than patients without these hub gene mutations (median, 45.57 months vs. 90.64 months, P = 4.825e-3) (Fig. 4D−F). However, the same analysis of the private hub genes identified in MHCC97-H and HCC97LM3 cells demonstrates that there was no significant difference in survival rates between patients regardless of whether they carried these hub gene mutations or not. The cBio potal data also demonstrated that patients with mutations of hub gene AR (P = 7.728e-5) and EFTUD2 (P = 0.036) as was identified in HCC97LM3, and MYOD1 (P = 0.026) as was identified in MHCC97-L, were associated with significantly poorer OS than patients without the hub gene mutations in liver cancer (Fig. 4G−I).

Expression and survival analysis of hub genes in UALCAN database

Chen et al demonstrated that gene mutations affect mRNA levels.[23] Next, to further explore the effect of hub gene mutations on overall survival of liver cancer patients, all hub genes detected in the HCC cell lines (including both branch private hub gene mutations and 1054 shared trunk mutations) were retrieved in liver cancer tissues in UALCAN database. The expression of most hub gene was significantly different between liver cancer tissues and normal tissues (Supplemental Table 5). Among them, altered expression of CDC27, RELA, EIF4E, FOXO3, EZH2, POLR2G, WDTC1 and DLG4 genes were associated with significantly poorer OS in HCC patients (Fig. 5A−H).

In depth sequence analysis and verification of KEAP1 gene mutation

Data from the WES show a deletion mutation of KEAP1 gene, which is related to previously reported truncation mutation of the C-terminal region of KEAP1 protein.[24] Intriguing, the current KEAP1 mutation detected in the MHCC97 cells is a distinctive mutation from that was reported by Qiu et al, with an insertion of 12 amino acid fragment in the C-terminal region after the amino acid residue 445 (Fig. 6A and B). Interestingly, sequence analysis of the 12 residue insertion fragment indicates that it matches to no known protein in the NCBI protein database (Fig. 6A and B). Due to the truncation mutation of KEAP1 gene, the proline at position 445 (Protein Change, p.P445fs) of KEAP1 protein was changed into glutamine, and the Terminator was introduced at position 457 resulting in an insertion of extra12-aa residue fragment in the C-terminal region of KEAP1 protein (RefSeq KEAP1 proteins, NP_987096.1 GI: 45269145) (Fig. 6B). To verify if this unique mutation did exist in the MHCC97 cells, PCR primers were designed for KEAP1 gene based on the WES data (Fig. 6C).

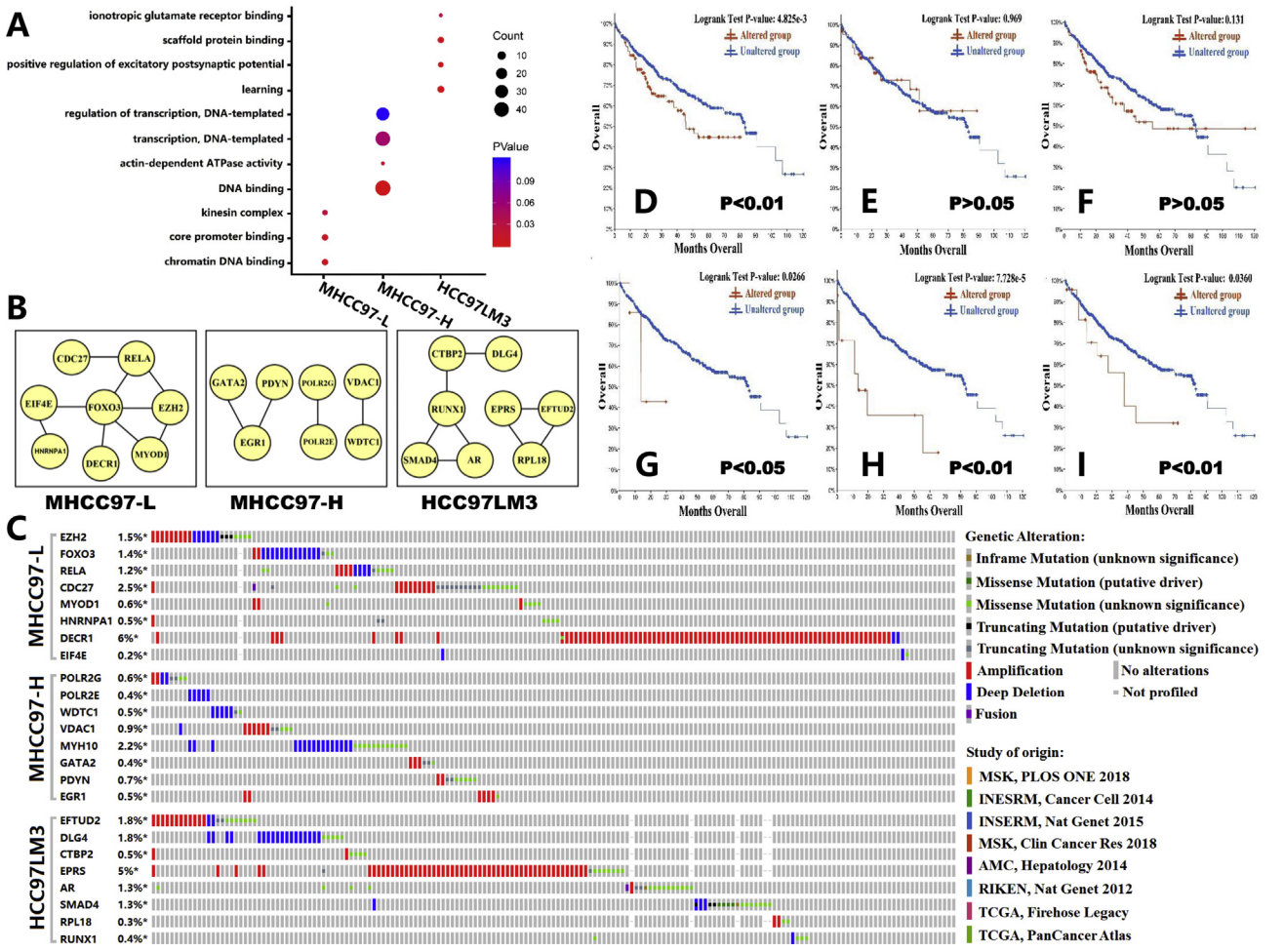

Figure 4 Private Mutated Genes in Metastatic HCC Cell Lines. (A) Scatterplot of GO enrichment and KEGG pathway analysis of private gene sets in metastatic HCC cell lines. (B) The protein—protein interaction network of 8 hub genes in private mutated genes among the three HCC cell lines. The MYH10 gene in MHCC97-H has no interaction with other hub genes, so it does not show in Fig4B. (C) Query 8 hub genes of the three HCC cell lines from 8 different hepatocellular carcinoma cohorts in the cBio Cancer Genome Portal database. Distinct genetic alterations and color codes shown with the percentage of inquired hub genes in individual tumor samples. (D—F) Bioinformatic analysis of the overall survival of HCC patients with or without 8 hub gene mutations. MHCC97-L (D); MHCC97-H (E); HCC97LM3 (F). (G—I) Altered expression of MYOD1 (G) identified from MHCC97-L cell line and other two genes, AR (H) and EFTUD2 (I), identified from HCC97LM3 cells and their association with OS in patients with HCC.

As shown in Fig. 6C, the mutation primer of KEAP1 showed no specific product in refseq reparative genome database (Organization limited to homosapiens). PCR verification of mutated KEAP1 was carried out in cell lines MHCC97-L, MHCC97-H, HCC97LM3, Hep G2, L-O2, Huh7 and SMMC-7721, and observed amplified target DNA fragments with the insertion mutation (Estimated product size of mutated KEAP1:587bp) in MHCC97-H and HCC97LM3 only (Fig. 6D).

Discussion

In this study, WES detected a total of 9293, 9608 and 9364 variations in MHCC97-L, HCC97LM3 and MHCC97-H cells, respectively. The data demonstrated a shared genomic landscape with less than 20% of heterogeneous mutations observed among these cell lines. MHCC97-H had the highest heterogeneous mutational burden (18.17%). These heterogeneous mutations may originate from the continuous passage of cell lines or xenografts in mice *in vivo* and subsequent culture of cells *in vitro*. Although there are heterogeneous mutations in the three cell lines, according evolutionary tree, we found that HCC97LM3 and MHCC97-H cells shared similar mutational patterns and evolutionary traits than MHCC97-L. This is consistent with the screening process of the three HCC cell lines, i.e., HCC97LM3 was further selected by continuous passage of MHCC97-H cells in mice.[11,12]

Intriguingly, it is unclear as to why the mutational burden of HCC97LM3, which possessed both lung and abdominal metastatic potentialities after subcutaneous inoculation into NOD-SCID mice, is less than that of MHCC97-H and MHCC97-L cells (The major difference between the MHCC97-H and MHCC97-L cell lines is in the propensity and degree of lung metastasis after orthotopic inoculation into mouse liver). We speculate that the

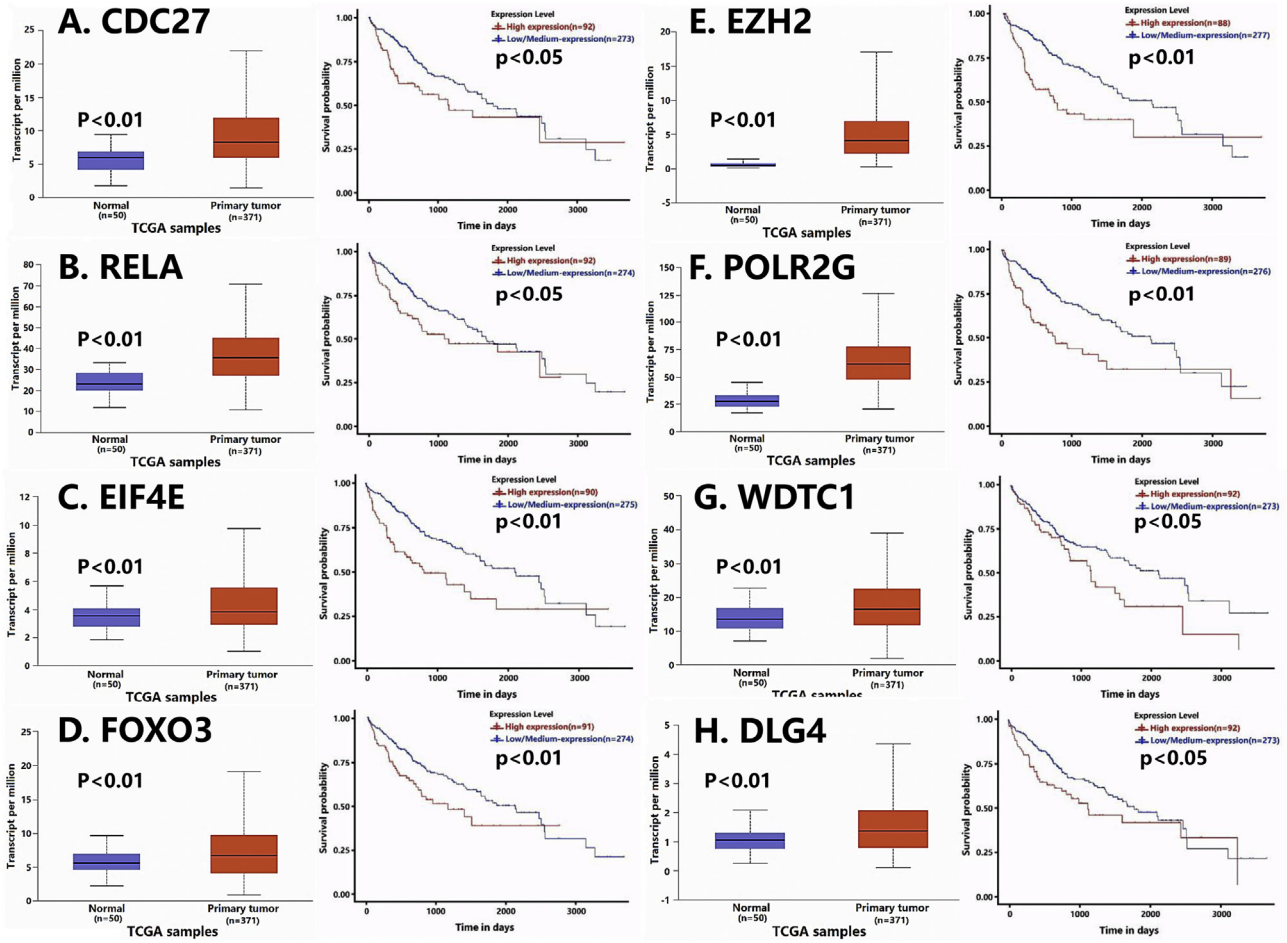

Figure 5 Expression and Survival Analysis of Hub Genes in UALCAN database. Hub genes *CDC27* (A), *RELA* (B), *EIF4E* (C), *FOXO3* (D), *EZH2* (E), *POLR2G* (F), *WDTC1* (G), and *DLG4* (H) are highly expressed in liver cancer tissues, which is associated with a poorer overall survival in liver cancer patients.

differences may in part reflect the environmental fitness selection pressure posed to the cells after subcutaneous inoculation and or negative selection of clones with less mutational burdens during repeated in vivo passages in the rodents. Obviously, intrinsic intra-neoplastic heterogeneity may also account for a significant portion of the differences in metastatic potentiality and organo-tropic proclivity manifested in these HCC cell lines.[25]

In the HCC cell lines, C: G > T: A and T: A > C: G somatic transitions were the two most frequent substitutions.[26] In previous studies, Totoki and colleagues reported that C > T/G > A, T > C/A > G and C > A/G > T transitions were over-represented in the HCC genome.[27] They further confirmed that C > T/G > A and T > C/A > G somatic transitions were associated with HCV- and HBV-positive HCC cases.[26,27] The most frequent somatic substitutions in these HCC cell lines were comparable to that of HCC patients, highlighting their feasibility as a representative model system.

Around 21.88% of the reported driver genes was identified, 87.10% of which are shared trunk mutations in three HCC cell lines and 12.90% of them are private mutations in one of the sub-clones. Most of the driver genes we have detected occurred in the share trunk mutations, which is consistent with the previous reports.[28,29] Clonal sweeps can lead to the homogeneity of driver gene mutations.[30] In these metastatic cell lines, the top two mutated driver genes are *P53* and *APOB*. The tumor suppressor P53 is frequently mutated in HBV-related HCC. P53 mutations are associated with worse OS in Chinese patients with HBV background.[31,32] APOB mutations have also been shown to occur in 10% of HCC tumors.[33–36] These HCC cell lines also demonstrate accumulative mutations in receptor-tyrosine kinase (RTK)/RAS/MAP-Kinase signaling pathways. The overall frequency of RTK-RAS alterations in HCC are reported as 22%.[22]

Among 1054 differentially mutated genes shared by the three HCC cells, the top 5 KEGG pathways enriched are ECM-receptor interactions, focal adhesion, Amoebiasis, PI3K-Akt signaling and protein digestion and absorption. These data indicated the differentially mutated genes shared in the three HCC cell lines are likely associated with HCC invasion and metastatic progression. The hub gene *SCD1* selected from the three HCC cell lines is involved in regulation of fatty acid (FA) metabolism and is proposed as a potential therapeutic target in HCC.[37–39] Previous studies demonstrated that upregulated expression of the hub gene ITGB4 promoted the lung metastases in HCC via activating the AKT signaling pathway.[40]

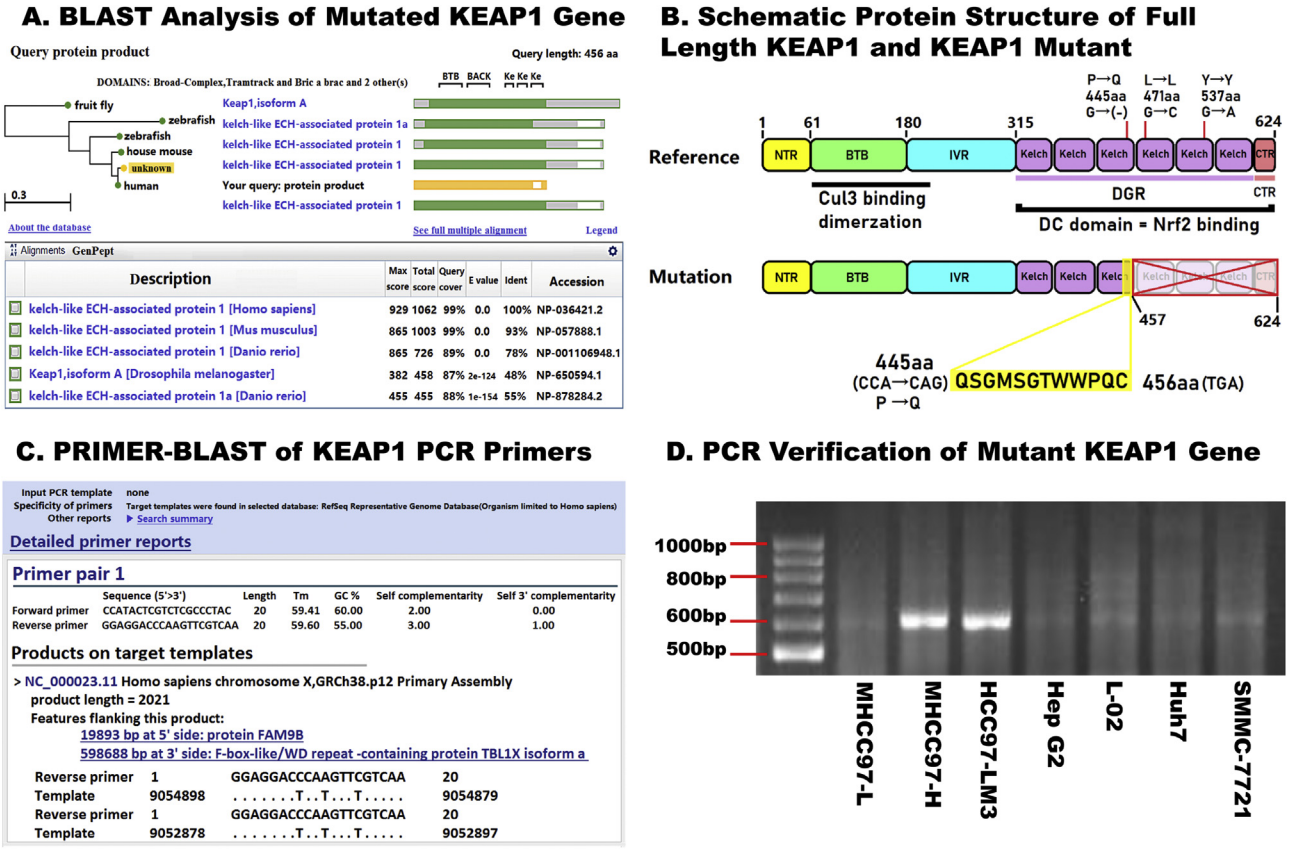

Figure 6 Verification of *KEAP1* Gene Mutation in the HCC Cell Lines. (A) According to WES data, the protein sequence of mutant KEAP1 was predicted, and the protein sequence was compared in NCBI-BLAST. (B) Schematic illustration of the WES data showing that one of the *KEAP1* gene mutation (c.1334delC, p.P445fs) will cause an amino acid substitution from Proline to Glutamine at the residue 445 of *KEAP1* protein, while the other two changes are synonymous mutations. The full-length protein of *Keap1* consists of NTR, BTB, IVR, DGR (DC domain harboring six Kelch-repeats) and CTR. (C) Primers designed for analysis of *KEAP1* mutations were verified in PRIMER-BLAST database. (D) PCR verification of mutant *KEAP1* in cell lines MHCC97-L, MHCC97-H, HCC97LM3, Hep G2, L-O2, Huh7 and SMMC-7721.

Moreover, the hub gene *AR* mutation has been reported previously as a causative factor of HCC and is therefore regarded as a potential therapeutic target for HCC.[41,42]

KEAP1 is well-established cancer driver gene, also known as Nrf2 inhibitors.[15] The schematic diagram of the structure of *KEAP1* is shown in Fig. 6B. The full-length protein of *KEAP1* consists of 624 amino acids.[43] The domain of *KEAP1* protein consists of the following five parts: NTR, BTB, intervening region (IVR), DC domain harboring six Kelch-repeat domain (DGR) and C-terminal region (CTR) (Fig. 6B). The DGR and CTR domains of *KEAP1* located in the C-terminal region are the Nrf2 bonding domains.[43] Biochemical and structural analysis showed that the complete *KEAP1* homodimer formed a cherry wave structure by using two binding sites in the Neh2 domain of Nrf2, in which one Nrf2 molecule bound with two *KEAP1* molecules.[44] Binding of *KEAP1* with Nrf2 recruits the Cullin3 (CUL3)-based E3 ligase complex and mediates Nrf2 ubiquitination and subsequent proteasome degradation of Nrf2 protein.[45,46]

Thus, deletion mutation of the Nrf2 bonding region (DGR and CTR) of *KEAP1* protein, as detected in these MHCC97 cells, will likely lead to a loss of *KEAP1* function, thus resulting in stabilization and constitutive activation of Nrf2 pathway.[47] Nrf2/KEAP1 pathway plays an important role in the occurrence and development of a variety of cancers, including cervical cancer, lung cancer, esophageal cancer, papillary thyroid carcinoma, pancreatic cancer, colon cancer, breast cancer, laryngeal squamous cell carcinoma and liver cancer.[48–50] Nrf2 plays an extremely important role in tumorigenesis and development, such as promoting the growth and proliferation of tumor cells, promoting the self-renewal and angiogenesis of tumor stem cells, enhancing the anti-chemotherapy and anti-radiation ability of tumor cells, and inhibiting the apoptosis of tumor cells.[51] NRF2 is abnormally accumulated in many types of cancer, and is associated with aggressiveness, metastasis and the poor prognosis of patients.[52,53]

The current WES data showed that *KEAP1* mutations in MHCC97H and HCC97LM3 resulted in a deletion of the C-terminal DCR domain of *KEAP1* protein. The mutated *KEAP1* gene DNA fragment was amplified from MHCC97-H and HCC97LM3. Similar to the results of Zhang et al, WES data show that the *KEAP1* gene of MHCC97H cells and HCC97LM3 cells lacks cytosine at position 1334 (cDNA Change: c.1334delC). This will lead to a mutation in the reading

frame shift of the *KEAP1* protein, the presence of a terminator at position 1368, and translation of the *KEAP1* protein will be prematurely terminated, resulting in the C-terminal truncation of the *KEAP1* protein.[24] Based on the WES data, the amino acid sequence of the current *KEAP1* mutant was predicted. Unlike that of *KEAP1* full-length protein consisting of 624 amino acids,[45] the current truncation of *KEAP1* mutant contains only 456 amino acids. Thus, further functional and mechanistic studies of this truncation *KEAP1* mutant, its interaction with Nrf2 and its oncogenic effect in particular, during the development of liver cancer are warranted.

In summary, the WES data indicate that the three metastatic cell lines from the same patient's tumor do indeed have kinship at the genomic level. The hub genes identified here may play an important role during the development and progression of liver cancer.

Conflict of Interest

The authors have no conflict of interest.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC, NO. 81172066; NO. 81472858; NO. 91529103) and Innovation Team Fund of Second Affiliated Hospital of Chongqing Medical University. The authors would like to thank Dr. Zhou-You Tang, Professor & Director, Liver Cancer Institute, Fudan University, for providing the three HCC cell lines(MHCC97-L, MHCC97-H, HCC97LM3).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gendis.2020.05.003.

References

1. 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 J Clin*. 2018;68(6):394–424.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(1):7–30.
3. Kulik LM, Chokechanachaisakul A. Evaluation and management of hepatocellular carcinoma. *Clin Liver Dis*. 2015;19(1):23–43.
4. Greaves M. Evolutionary determinants of cancer. *Canc Discov*. 2015;5(8):806–820.
5. Nowell PC. The clonal evolution of tumor cell populations. *Science (New York, NY)*. 1976;194(4260):23–28.
6. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature*. 2013;501(7467):328–337.
7. Martelotto LG, Ng CK, Piscuoglio S, Weigelt B, Reis-Filho JS. Breast cancer intra-tumor heterogeneity. *Breast Cancer Res*. 2014;16(3):210.
8. Harris JF, Chambers AF, Hill RP, Ling V. Metastatic variants are generated spontaneously at a high rate in mouse KHT tumor. *Proc Natl Acad Sci USA*. 1982;79(18):5547–5551.
9. Takai A, Dang HT, Wang XW. Identification of drivers from cancer genome diversity in hepatocellular carcinoma. *Int J Mol Sci*. 2014;15(6):11142–11160.
10. Tang ZY, Sun FX, Tian J, et al. Metastatic human hepatocellular carcinoma models in nude mice and cell line with metastatic potential. *World J Gastroenterol*. 2001;7(5):597–601.
11. Li Y, Tang ZY, Ye SL, et al. Establishment of cell clones with different metastatic potential from the metastatic hepatocellular carcinoma cell line MHCC97. *World J Gastroenterol*. 2001;7(5):630–636.
12. Li Y, Tang Y, Ye L, et al. Establishment of a hepatocellular carcinoma cell line with unique metastatic characteristics through in vivo selection and screening for metastasis-related genes through cDNA microarray. *J Canc Res Clin Oncol*. 2003;129(1):43–51.
13. Ramos AH, Lichtenstein L, Gupta M, et al. Oncotator: cancer variant annotation tool. *Hum Mutat*. 2015;36(4):E2423–E2429.
14. Noorpisheh Ghadimi S, Abedini MR, Sarkari B, Savardashtaki A, Mikaeili F. Neobalantidium coli: first molecular identification from the Eurasian wild boar, Sus scrofa in Bushehr province, southwestern Iran. *Veterinary Med. Sci*. 2020;6(1):142–146.
15. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*. 2018;174(4):1034–1035.
16. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
17. Song J, Wu S, Xia X, Wang Y, Fan Y, Yang Z. Cell adhesion-related gene somatic mutations are enriched in aggressive papillary thyroid microcarcinomas. *J Transl Med*. 2018;16(1):269.
18. Kennedy-Shaffer L. Before $p<0.05$ to beyond $p<0.05$: using history to contextualize p-values and significance testing. *Am Statistician*. 2019;73:82–90.
19. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43(D1):D447–D452.
20. Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia*. 2017;19(8):649–658.
21. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 2015;43:D805–D811 (Database issue).
22. Sanchez-Vega F, Mina M, Armenia J, et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell*. 2018;173(2):321–337. e310.
23. Chen S, Li K, Cao W, et al. Codon-Resolution analysis reveals a direct and context-dependent impact of individual synonymous mutations on mRNA level. *Mol Biol Evol*. 2017;34.
24. Qiu L, Wang M, Zhu Y, Xiang Y, Zhang Y. A naturally-occurring dominant-negative inhibitor of Keap1 competitively against its negative regulation of Nrf2. *Int J Mol Sci*. 2018;19(8).
25. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell*. 2017;168:613–628.
26. Totoki Y, Tatsuno K, Yamamoto S, et al. High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet*. 2011;43(5):464–469.
27. Fujimoto A, Totoki Y, Abe T, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet*. 2012;44(7):760–764.
28. Torrecilla S, Sia D, Harrington AN, et al. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. *J Hepatol*. 2017;67(6):1222–1231.
29. Birkbak NJ, McGranahan N. Cancer genome evolutionary trajectories in metastasis. *Canc Cell*. 2020;37(1):8–19.

30. Reiter JG, Baretti M, Gerold JM, et al. An analysis of genetic heterogeneity in untreated cancers. *Nat Rev Canc*. 2019; 19(11):639—650.
31. Cleary SP, Jeck WR, Zhao X, et al. Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology*. 2013;58(5):1693—1702.
32. Woo HG, Wang XW, Budhu A, et al. Association of TP53 mutations with stem cell-like gene expression and survival of patients with hepatocellular carcinoma. *Gastroenterology*. 2011; 140(3):1063—1070.
33. Schulze K, Imbeaud S, Letouze E, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet*. 2015; 47(5):505—511.
34. Fujimoto A, Furuta M, Totoki Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet*. 2016;48(5): 500—509.
35. Fernandez-Banet J, Lee NP, Chan KT, et al. Decoding complex patterns of genomic rearrangement in hepatocellular carcinoma. *Genomics*. 2014;103(2—3):189—203.
36. Wheeler DA, Roberts LR, The Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*. 2017;169(7): 1327—1341. e1323.
37. Miyazaki M, Flowers M, Sampath H, et al. Hepatic stearoyl-CoA desaturase-1 deficiency protects mice from carbohydrate-induced adiposity and hepatic steatosis. *Cell Metabol*. 2008; 6:484—496.
38. Ntambi J, Miyazaki M. Recent insights into stearoyl-CoA desaturase-1. *Curr Opin Lipidol*. 2003;14:255—261.
39. Ma M, Lau E, Leung D, et al. Stearoyl-CoA Desaturase regulates sorafenib resistance via modulation of ER stress induced differentiation. *J Hepatol*. 2017;67.
40. Leng C, Zhang ZG, Chen WX, et al. An integrin beta4-EGFR unit promotes hepatocellular carcinoma lung metastases by enhancing anchorage independence through activation of FAK-AKT pathway. *Canc Lett*. 2016;376(1):188—196.
41. Ao J, Meng J, Zhu L, et al. Activation of androgen receptor induces ID1 and promotes hepatocellular carcinoma cell migration and invasion. *Molecular Oncol*. 2012;6(5):507—515.
42. Kanda T, Yokosuka O. The androgen receptor as an emerging target in hepatocellular carcinoma. *J Hepatocell Carcinoma*. 2015;2:91—99.
43. Taguchi K, Motohashi H, Yamamoto M. Molecular mechanisms of the Keap1-Nrf2 pathway in stress response and cancer evolution. *Gene Cell*. 2011;16(2):123—140.
44. Padmanabhan B, Tong KI, Ohta T, et al. Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol Cell*. 2006;21(5):689—700.
45. Furukawa M, Xiong Y. BTB protein Keap1 targets antioxidant transcription factor Nrf2 for ubiquitination by the Cullin 3-Roc1 ligase. *Mol Cell Biol*. 2005;25(1):162—171.
46. Liu Q, Gao Y, Ci X. Role of Nrf2 and its activators in respiratory diseases. *Oxidative medicine and cellular longevity*. 2019; 2019:7090534.
47. Zimta AA, Cenariu D, Irimie A, et al. The role of Nrf2 activity in cancer development and progression. *Cancers*. 2019;11(11).
48. Renaud CO, Ziros PG, Chartoumpekis DV, Bongiovanni M, Sykiotis GP. Keap1/Nrf2 signaling: a new player in thyroid pathophysiology and thyroid cancer. *Front Endocrinol*. 2019; 10:510.
49. Qin JJ, Cheng XD, Zhang J, Zhang WD. Dual roles and therapeutic potential of Keap1-Nrf2 pathway in pancreatic cancer: a systematic review. *Cell Commun Signal*. 2019;17(1):121.
50. Hua CC, Chang LC, Fan CW, Tseng WK. Associations between the Nrf2/Keap1 pathway and mitochondrial functions in colorectal cancer are affected by metastasis. *Canc Biomarkers*. 2020;27(2):163—171.
51. Wu S, Lu H, Bai Y. Nrf2 in cancers: a double-edged sword. *Cancer Med*. 2019;8(5):2252—2267.
52. Mizumura K, Maruoka S, Shimizu T, Gon Y. Role of Nrf2 in the pathogenesis of respiratory diseases. *Respir Investig*. 2020; 58(1):28—35.
53. Hanafi H. *NRF2 as a Central Point in Cancer Metastasis*. 2018.