



OPEN ACCESS

EDITED BY

Divya P. Kumar,
JSS Academy of Higher Education
and Research, India

REVIEWED BY

Prasanna K. Santhekadur,
JSS Academy of Higher Education
and Research, India
Carmine Fedele,
Novartis Institutes for BioMedical
Research, United States

*CORRESPONDENCE

Aejaz Sayeed
aejaz.sayeed@bblumberg.org

SPECIALTY SECTION

This article was submitted to
Gastrointestinal Cancers: Hepato
Pancreatic Biliary Cancers,
a section of the journal
Frontiers in Oncology

RECEIVED 07 June 2022

ACCEPTED 18 November 2022

PUBLISHED 13 December 2022

CITATION

Block T, Zezulinski D, Kaplan DE, Lu J,
Zanine S, Zhan T, Doria C and
Sayeed A (2022) Circulating
messenger RNA variants as a potential
biomarker for surveillance of
hepatocellular carcinoma.
Front. Oncol. 12:963641.
doi: 10.3389/fonc.2022.963641

COPYRIGHT

© 2022 Block, Zezulinski, Kaplan, Lu,
Zanine, Zhan, Doria and Sayeed. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

Circulating messenger RNA variants as a potential biomarker for surveillance of hepatocellular carcinoma

Timothy Block¹, Daniel Zezulinski¹, David E. Kaplan²,
Jingqiao Lu³, Samantha Zanine⁴, Tingting Zhan⁵,
Cataldo Doria⁶ and Aejaz Sayeed^{1*}

¹Department of Translational Medicine, Baruch S. Blumberg Institute, Doylestown, PA, United States,

²Division of Gastroenterology and Hepatology, University of Pennsylvania Perelman School of Medicine
and The Corporal Michael J. Crescenz Veterans Administration Hospital, Philadelphia, PA, United States,

³Ray Biotech Life Inc., Peachtree Corners, GA, United States, ⁴Department of Mechanical Engineering,
Pennsylvania State University, PA, United States, ⁵Division of Biostatistics, Department of Pharmacology
and Experimental Therapeutics, Thomas Jefferson University, Philadelphia PA, United States, ⁶CHS Liver
and Pancreas Centers of Excellence, Capital Health Cancer Center, One Capital Way, Pennington, NJ,
United States

Background and rationale: Liver derived messenger ribonucleic acid (mRNA) transcripts were reported to be elevated in the circulation of hepatocellular carcinoma (HCC) patients. We now report the detection of high-risk mRNA variants exclusively in the circulation of HCC patients. Numerous genomic alleles such as single nucleotide polymorphisms (SNPs), nucleotide insertions and deletions (called Indels), splicing variants in many genes, have been associated with elevated risk of cancer. Our findings potentially offer a novel non-invasive platform for HCC surveillance and early detection.

Approach: RNAseq analysis was carried out in the plasma of 14 individuals with a diagnosis of HCC, 8 with LC and no HCC, and 6 with no liver disease diagnosis. RNA from 6 matching tumors and 5 circulating extracellular vesicle (EV) samples from 14 of those with HCC was also analyzed. Specimens from two cholangiocarcinoma (CCA) patients were also included in our study. HCC specific SNPs and Indels referred as “variants” were identified using GATK HaplotypeCaller and annotated by SnpEff to filter out high risk variants.

Results: The variant calling on all RNA samples enabled the detection of 5.2 million SNPs, 0.91 million insertions and 0.81 million deletions. RNAseq analyses in tumors, normal liver tissue, plasma, and plasma derived EVs led to the detection of 5480 high-risk tumor specific mRNA variants in the circulation of HCC patients. These variants are concurrently detected in tumors and plasma samples or tumors and EVs from HCC patients, but none of these were detected in normal liver, plasma of LC patients or normal healthy individuals. Our results demonstrate selective detection of concordant high-risk HCC-specific mRNA variants in free plasma, plasma derived EVs and tumors of HCC patients. The variants comprise of splicing, frameshift, fusion and single nucleotide alterations and correspond to cancer and tumor

metabolism pathways. Detection of these high-risk variants in matching specimens from same subjects with an enrichment in circulating EVs is remarkable. Validation of these HCC selective ctmRNA variants in larger patient cohorts is likely to identify a predictive set of ctmRNA with high diagnostic performance and thus offer a novel non-invasive serology-based biomarker for HCC.

KEYWORDS

circulating mRNA variants, hepatocellular carcinoma, liver cirrhosis, extracellular vesicles, HCC surveillance and early detection

Introduction

Hepatocellular carcinoma (HCC) represents the fourth most deadly cancer worldwide and projected to become the 3rd leading cause of cancer related deaths by 2030 (1–3). Rising HCC incidence has been attributed to increasing prevalence of chronic liver disease (CLD) specifically non-alcoholic steatohepatitis (NASH) and viral hepatitis (4–6), which can progress to liver cirrhosis (LC) detected in up to 90% of HCC patients (7). Development of cirrhosis significantly increases risk of HCC (8). Though traditionally HCC is known to arise in the context of hepatitis B (HBV) and hepatitis C (HCV) viral infections, metabolic associated fatty liver disease (MAFLD) and NASH are becoming more prominent risk factors for HCC (9–11). Because early-stage tumor diagnosis improves options for potentially curative therapy and thereby improves overall survival (12), AASLD guidelines recommend surveillance in CLD patients at risk for HCC (8, 13) with liver ultrasound (US) and serum alpha-fetoprotein (AFP) test conducted every 6 months (14). Ultrasound plus AFP remains relatively insensitive for early detection identifying only 63% of early-stage cancers (15). Alternative surveillance biomarkers, Lens-culinaris agglutinin-reactive AFP, known more commonly as AFP-L3, and des-carboxy prothrombin (DCP), used in Japan for screening, are more specific but less sensitive biomarkers and therefore less utilized (16). Recently, GALAD score (Gender, Age, AFP-L3, AFP, DCP) has been proposed which combines simple demographic data with serum biomarkers and predicts the future probability of developing HCC in CLD patients (17, 18). Despite these advances, surveillance is suboptimal not only due to low sensitivity of existing methods of early detection but also due to barriers in access to quality imaging and biomarkers (13, 19, 20). When ultrasound or AFP detect abnormalities, AASLD and EASL guidelines recommend magnetic resonance imaging (MRI) using extracellular or hepatobiliary contrast media or dynamic computed tomography as diagnostic testing using the Liver Imaging Reporting and Data Systems (LI-RADS) grading system (21). However, vascular imaging provides little

information about tumor biology, patient prognosis or likelihood of response to therapy. Since a diagnosis of HCC is most often made radiologically without a biopsy, key tumor molecular information rarely is available to optimize therapy. The limitations of testing lead to a critical need for improving HCC surveillance and early detection (15, 22, 23).

Cell-free DNA (cfDNA) in the blood provides a useful non-invasive diagnostic analyte for cancer although for early detection, both technical and biological factors introduce challenges to the detection of mutant DNA in plasma and its interpretation (24–26). Detection of tumors is also possible through analysis of circulating tumor cells (CTCs). However, density of CTCs in circulation has been observed to be very low presumably because only large tumors release CTCs (27). cfDNA analysis in cancer clinical trials has suggested promising results in early detection, real time monitoring and management (28). Fragmentation patterns of cfDNA combined with mutations in cfDNA has been reported to increase the sensitivity of cancer detection (29). However, the mechanisms of release and degradation of cfDNA, and the factors that affect the representation of circulating tumor DNA (ctDNA) in plasma, are poorly understood. Additionally there are challenges in differentiating between potential tumors within a mixed liquid biopsy due to the low allelic frequency have been reported (25, 26). Circulating mRNA, without the copy number limitations of DNA, provides an alternative platform for use in cancer early detection and diagnostics.

Cancer is a multistep process driven by cumulative acquisition of both germline (inherited) and somatic (not inherited) mutations leading to the transformation of normal hepatocytes into malignant clones. Genetic variants can be classified into several categories, including single nucleotide polymorphisms (SNPs), small insertions and deletions (Indels), and structural variants (30). SNPs account for >90% of allelic disparities scattered throughout the human genome and nonsynonymous SNPs alter the amino acid sequence as well as potentially affect protein structure and function (31, 32). We have reported the detection of tumor specific mRNA transcripts

in the circulation of HCC patients by RNAseq (33). Here we report the detection of high-risk variants comprising of both SNPs and indels in the circulating mRNA of cancer patients using RNAseq. Mutation profiling by NGS platforms can potentially lead to false-positive results due to errors introduced during library preparation and subsequent sequencing steps. These challenges have necessitated the use of multiple mutation-enrichment methods like GATK (34) and SnpEff tools (35, 36). It has been suggested that comparison of paired tumor and plasma samples represents an important prerequisite to evaluate the diagnostic accuracy of analytical platforms, especially for variants with low allele frequency (37, 38). Profiling the mutational landscape reflected in mRNA of the tumor in parallel with that of the circulation from same patient may identify critical concordant RNA variants which could potentially be used to develop a liquid biopsy platform for surveillance and early HCC detection. Here we demonstrate the identification of high-risk HCC-specific mRNA variants present exclusively in tumors and plasma of HCC patients, but not detectable in LC patients without cancer or normal healthy individuals. These circulating tumor-specific mRNA variants are derived from RNAseq analysis of total RNA using GATK and SnpEff tools, further validation and characterization of these RNA variants is essential. Our study potentially offers the promise of a novel analyte to be used in non-invasive detection of high-risk cancer specific mRNA variants and may facilitate effective surveillance and screening of ever-increasing numbers of patients with chronic liver disease.

Materials and methods

Human subjects

Samples were acquired from commercial vendor, Biochemed Inc and from our collaborators at The University of Pennsylvania (UPENN) and Capital Health Cancer research

Institute. All samples were acquired before any therapeutic regimen was initiated. Patient plasma samples were collected at the time of diagnosis and acquired from treatment naïve adult consenting donors diagnosed with either hepatocellular carcinoma (HCC), cholangiocarcinoma or liver cirrhosis. Tumor tissues were harvested *via* biopsy or surgical resection before any treatment. HCC diagnosis was made by either 1) biopsy or 2) typical enhancement patterns on dynamic contrast-enhanced CT or MRI (later codified as Li-RADS criteria). Staging performed per Barcelona Clinic Liver Cancer system (BCLC) (0 very early, A early, B intermediate, C advanced, D terminal). Patients with CCA were diagnosed by histopathology of the liver resection. Patients with LC underwent standard clinical surveillance for HCC development with abdominal ultrasound and AFP testing every 6 months (optimally) per AASLD surveillance guidelines. Patient details are shown in Table 1 and Supplementary Table 7. Normal human liver tissue was a kind gift from Dr. Ramilla Philip, Immunotope Inc.

Isolation and characterization of EVs

Plasma samples from HCC or CCA patients (1 to 1.5ml) were spun at 2,000g for 30 minutes to remove the cellular debris before purification of EVs using ExoQuick kit (System biosciences, Palo Alto CA). EVs were further purified by ultracentrifugation at 100,000g for 2 hours at 4C, pellet resuspended in 100ul PBS. followed by NanoSight tracking and Chip analysis. Dilutions of samples with PBS (Zetaview NTA) were carried out as 1:1000 (HCC001), 1:100 (HCC004, HCC006, HCC007, CCA005, NHC66), 1:20,000 (HCC014) and 1:10,000 (LC19, LC34). Dilutions for Exoview chip analysis were carried out in solution A (Exoview tetraspanin plasma kit analysis) 1:1000 (HCC001), 1:100 (HCC004, CCA005), 1:100,000 (LC19). Manufacturer instructions were followed and EVs were stained for CD63, CD81 and CD19 at 1:1000 final working dilutions from the stock concentration. Antibodies

TABLE 1 Clinical and demographic information of patients used as a source of plasma and tumor tissues for RNAseq analyses.

	n	Age	Gender=M/F	HBV/HCV/NAFL/NASH	LC (n)	HCC stages(n) AJCC I, II or BCLC A, B	HCC Stages(n) AJCC III or BCLC C,D
NHC	6	60.6	3/3	0/0/0	N/A	N/A	N/A
LC	8	64	8/0	3/3/0	8	N/A	N/A
HCC	14	64.1	11/1/3	1/6/2	11	10	4
CCA	2	72.5	1/1	1/0/2	2	1	1
Normal liver	1	57	1/0	0/0/0	N/A	N/A	N/A

Early cancer (AJCC I,II;Barcelona A,B);Late (AJCC III;Barcelona C,D).
N/A, Not Applicable.

Plasma from 30 different individuals were used. N=6, from "Normal healthy Controls (NHC), n=8 from those with a diagnosis of Liver Cirrhosis (LC), n=14 from those with a diagnosis of hepatocellular carcinoma (HCC) and LC and n=2 from those with a diagnosis of cholangiocarcinoma (CCA). For the HCC group: 11 of the 14 were also diagnosed with LC, 2 with NAFL; 10 BCLC A, B or AJCC I, II; 4 with BCLC C, D or AJCC III A & B. All patient specimens were acquired at the time diagnosis before initiating any therapeutic treatment. Ages and gender (Male (M), Female (F)), as indicated. Individuals with chronic hepatitis B virus (HBV) or hepatitis C virus (HCV) or no evidence of viral hepatitis were included, as indicated. Diagnosis of HCC, LC and chronic viral hepatitis as in text.

against these markers are immobilized on chips called capture probes. Once they encounter sample, tetraspanin receptor proteins on EVs bind to the antibodies on the chip and after washing, the chips are treated with fluorescently labeled antibodies against tetraspanin receptors, just like a sandwich ELISA. After washing, fluorescent labels are detected to evaluate particle counts and relative concentrations of tetraspanin receptors.

RNA extraction

Total RNA was isolated from human plasma samples, tumor and normal tissues and plasma-derived extracellular vesicles as reported earlier (33).

RNAseq analysis

For characterizing the circulating transcriptome, as shown in Table 1, plasma samples from HCC patients (n=14) representing both early and late stage, LC patients without HCC (n=8) and normal healthy individuals (n=6) were analyzed by RNAseq analysis. Within the HCC patient group, we also investigated matching tumors from 6 patients and plasma derived EVs from 5 patients in parallel. One normal human liver tissue was also included in the study. Total RNA was isolated and RNAseq was carried out on either Illumina NextSeq 500 platform using the SMARTer® Stranded Total RNA-Seq Kit v2-Pico Input Mammalian (Takara #634411) on a highoutput flow cell 2x75 bp as reported earlier (25) or on the Illumina NovaSeq 6000 using a S1 Flow Cell with a paired end run, 2 x 150 cycles, generating more than 50 million reads each. Residual Pico v2 SMART adapters on paired end fastq files were trimmed. The expression of genes in each sample were called with RSEM with STAR (2.7.5a) as aligner against a human reference genome (GRCh38.p13).

Variant calling and derivation of high-risk variants

Briefly indel local realignment, base quality recalibration and variants calling (Haplotype caller algorithm) were performed using Broad Institutes Genome Analysis Tool Kit (GATK) version 4.1.7.0. Variants with 20 or higher quality score as well as depth of 5 or more were considered as true positive. Filtration and annotation of high-risk variants was carried out using SnpEff analysis (version 4.3t) to identify high risk variants in all sample subsets. The SnpEff uses sequence and structure-based bioinformatics tools to predict the effect of protein-coding SNPs on the structural phenotype of proteins. It includes intronic, untranslated region, upstream, downstream, splice site, or intergenic regions in its annotated genomic locations. It can predict coding effects such as synonymous or non-

synonymous amino acid replacement, start/stop codon gains or losses, or frame shift changes. Tumor-centric high-risk variants were sorted by concordant detection in tumors and plasma from HCC patients, proportion/frequency in samples, lack of detection in normal liver tissue and in the plasma of LC and NHC controls and established functional association with cancer. Pathway enrichment analysis was carried out using R package ClusterProfiler.

Statistics

The distinguishing high-risk variant transcripts were selected out of 5,480 tumor specific variants using Fisher's exact test with multiple testing adjustment *via* Benjamini-Hochberg method.

Results

Diversity of RNA variants and variant effects

Bioinformatic tools help determine the effect of variants (SNPs, indels, CNVs, structural variants) on genes, transcripts, protein sequences and regulatory regions. RNAseq variant analysis in 46 samples (Tumors, plasma, circulating EVs and normal liver tissue from 29 subjects) led to the detection of some 27.5M variant effects manifested due to 5.2 M SNPs, 0.91M insertions and 0.81 M deletions. Measuring by impact, 94.3% represented modifier variants, followed by high and low risk variants at 2.3% each and 1% of variants were of moderate risk. Among the effects by functional class, 62% represented missense variants, synonymous/silent 35% and stop gain/non-sense variants at 3%. Evaluating effects by type and region, 56% were detected in introns, upstream 12.5%, downstream 12.6%, intergenic 10.5%, exon 4.2%, splicing 2.5% and 0.05% are genic. No significant differences were observed when comparing these effects between sample groups. However, proportion of high and moderate risk, frame shift and missense variants was observed to increase as we moved from normal healthy, LC to HCC patients. Though the increase in sample subsets was not statistically significant, it did point to a possibility of progressive build-up of mutations during disease progression (Figure 1A). Figure 1B shows the average counts of variants corresponding to different categories in various clinical subsets.

Tumor-centric concordant variants

We then evaluated the occurrence of "high risk" variants in each category. Filtration and annotation of high-risk variants

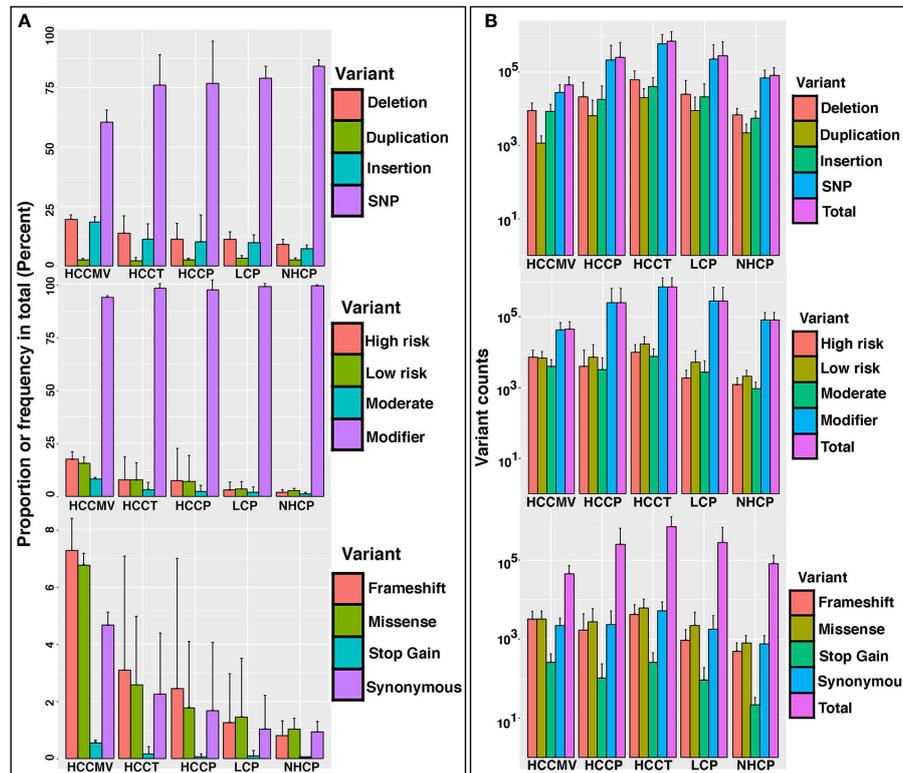


FIGURE 1

Diversity and profiles of circulating mRNA variants detected in the circulation, circulating extracellular vesicles and tumors of HCC patients. (A). Bar charts representing proportion or frequency of occurrence (% age) in patient and control specimens; HCCMV, HCC circulating microvesicles (n=6); HCCT, HCC tumors (n=6); HCCP, HCC plasma (n=14); LC, liver cirrhosis plasma (n=8); NHCP, healthy control plasma (n=6). Separate bar graphs are shown to represent profiles by type, impact, and functional class of variants. Error bars represent the variation within samples. (B). Bar graphs representing the total number and number of variant counts corresponding to different categories in the same sample subsets as 1A. The average number of variants per sample in RNA collected from different sources of subjects represented by the height of the bar, while the error bar showing the standard deviation within the group. Variant calling used a depth ≥ 5 and QUAL value ≥ 20 . The 'QUAL' value for each called SNP is calculated as $-\log_{10}(p)$ where p is the p value yielded from the statistical test carried out by ExactSNP. Therefore, a 'QUAL' value of 20 corresponds to a p value of 10^{-20} , which is extremely low.

was carried out using SnpEff analysis to identify high risk variants in all sample subsets (Figure 2A). "High risk" variants, derived by SnpEff are predicted to cause a significant effect on protein structure. As mentioned in methods, the SnpEff uses sequence and structure-based bioinformatics tools to predict the effect of protein-coding SNPs on the structural phenotype of proteins. High-risk variants in HCC tumors were further filtered to include variants detected in the plasma of HCC patients (concordant variants) and remove any variants detected in normal liver tissue and plasma samples of LC patients and NHC controls to derive a set of circulating tumor-centric RNA (ctrRNA) variants. Analysis of 14 plasma and 6 corresponding tumors from HCC patients identified 5480 high-risk concordant ctrRNA variants representing 3199 genes and detected in high frequency in both circulating plasma (7-50%) and tumors (30-100%). None of these tumor-centric

variants are detected in the plasma of either LC patients (n=8) or NHC (n=6) controls. Most of the high-risk concordant variants involve splice donor/acceptor regions (4933), frameshift (2506) changes and splicing variations. Comparing tumor centric variants with all SnpEff derived high-risk variants identified in HCCT, HCCP and HCCMV sample subsets point to the concordant variants shared between the different sample subsets (Figure 2B).

Prominent altered genes corresponding to variant transcripts.

Figure 3 represents ctrRNA variants detected in tumors and circulation of HCC patients in a clustering heat map comprised of splicing variants, SNPs, frameshift variants etc.

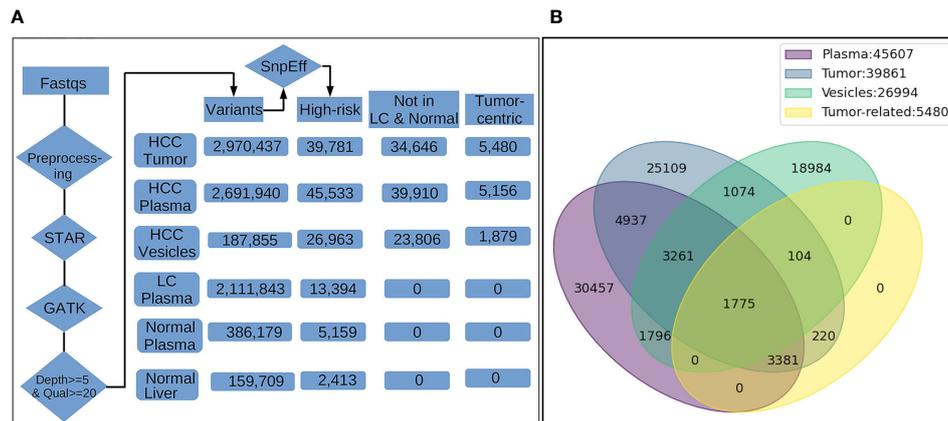


FIGURE 2

Filtration summary for derivation of high-risk HCC-specific mRNA variants and concordance of detection between different sample subsets from HCC patients. **(A)** The FASTQ files from NGS platform (illumina) were pre-processed to remove the residual of sequencing adapter in the files. The trimmed data were input into STAR, an aligner for RNA-Seq mapping, for alignment against human reference genome GRCh38. The alignments (BAM) were processed with GATK tools (v4.1.7.0) for splitting on reads with N cigar, base quality score recalibration, and variants calling by HaplotypeCaller. The variants reported by GATK were filtered by depth (≥ 5) and quality (≥ 20) and annotated with snpEff for location on genes/transcripts, mutation type and putative biological impact. The number in the boxes indicate the total unique variants identified in samples of different groups. Variants occurred in multiple samples within the same group were counted once only. Only high-risk variants and those with highest frequency in tumors and HCC plasma were considered during enrichment. High-risk variants also detected in the circulation of liver cirrhosis patients and normal healthy individuals were excluded leading us to cancer-associated high-risk variants referred to as tumor-centric variants. **(B)** Venn diagram showing the distribution and overlap of tumor-centric variants with total high-risk variants in various sample sources. The number of High-risk variants identified only in a single sample type either plasma, tumor and vesicles are 30,457, 25,109 and 18,894, respectively, reflecting the heterogeneity and difference in variant profiles attributed to the source of RNA. However, there are still 5,036 variants occurring in RNA from all three sources. Furthermore, 1,775 of them were among 5,480 tumor-related variants identified by tumor-control comparison, suggesting the plasma/vesicle RNA share abundant/significant transcripts carrying the tumor associated variants.

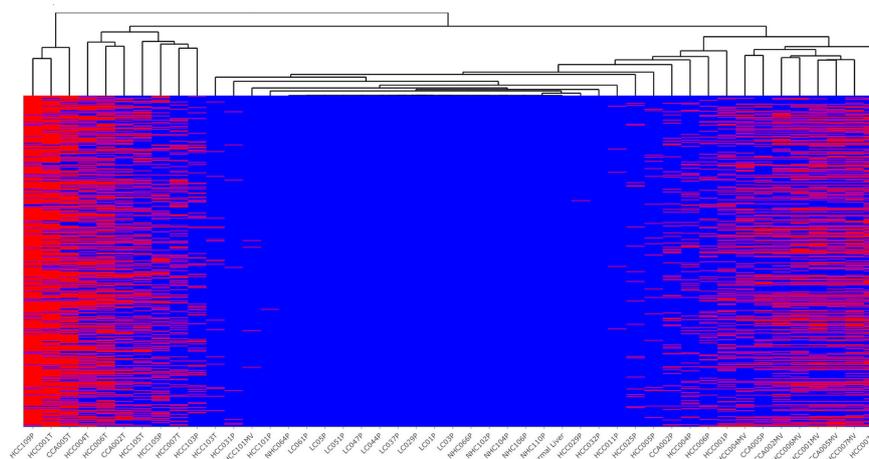


FIGURE 3

Profile of top high-risk concordant variants detected selectively in the plasma and tumors of HCC patients. Indicated occurrence (red color) of 1,501 high-risk variants identified in plasma of HCC patients but not in liver cirrhosis and healthy controls, with clustering dendrogram of samples on the top. Most of the tumor samples were clustered on the left part of the figure, while the vesicle samples on the right side. The liver cirrhosis and healthy controls not carrying the variants (blue) stay in the middle section. The variant annotations and profiles of these variants is shown in Supplementary Table 1.

(Supplementary Table 1). Notable high-risk variants present in tumors and plasma (concordant) correspond to TP53, CTNNB1, FAH and SF3B1 genes, which are already known to harbor driver mutations (reviewed in 39). Concordant variant transcripts corresponding to complement cascade like C2, C4A, C4B, C1R, C5, C8G, KLKB1, KNG1, CFHR5, MASP2, etc. which are known to be liver enriched (40) are represented among tumor-centric ctRNA variants. Additionally, among tumor-centric variants figure those ctRNA variants corresponding to coagulation factors F2, F11 and F12, which are not only liver enriched but also represent FDA approved drug targets (40).

Liver-derived transcripts reveal a high frequency of variants

With an interest to understand the variants in liver-associated transcripts, we calculated the number of total high-risk variants detected on these transcripts in various patient subsets. Some well-known liver-derived transcripts like FGL1, CYP2E1, SERPINA1, ALB, TF and SYCE1 display a huge variant load, particularly in the plasma and tumors from HCC patients. Several other liver-associated transcripts were observed to harbor significantly higher numbers of unique, high-risk variants in tumors, plasma or EVs of cancer patients compared to the levels detected in circulating plasma from LC patients and NHC controls. Bar graphs representing the total number of high-risk variants present in

liver-associated transcripts in various clinical subsets are shown in Figure 4. Focusing on certain selective tissue specific transcripts with associated high-risk variants can potentially offer increased specificity and sensitivity to identify high-risk patients.

Characterization of extracellular vesicles (EVs) isolated from plasma

It was interesting to know if the variant mRNA transcripts were present in the circulation within extracellular vesicles (EVs) since (EVs) in circulation are known to contain nucleic acids. Therefore, in five HCC subjects we compared the variant mRNA profiles in whole plasma with matching plasma derived EVs by RNAseq. EVs were purified from HCC, HCC-free LC or NHC plasma samples and analyzed for size and density by zeta view (Ncsi Spectradyne) analysis. Figure 5A shows the density peaks of circulating EVs from different patients and healthy control all centered around a 150 nm size. The EVs were further characterized for expression of tetraspanin receptors using Exoview chip arrays (Figure 5B). Pie charts represent the expression profiles of CD9, CD81 and CD63 in percentage of circulating EV particles from representative cancer patients and LC control. CD41a is a platelet marker, antibody against which was used together with IgG control. Data shows that all EVs seem to reflect comparable expression of the tetraspanin receptors authenticating their EV phenotype.

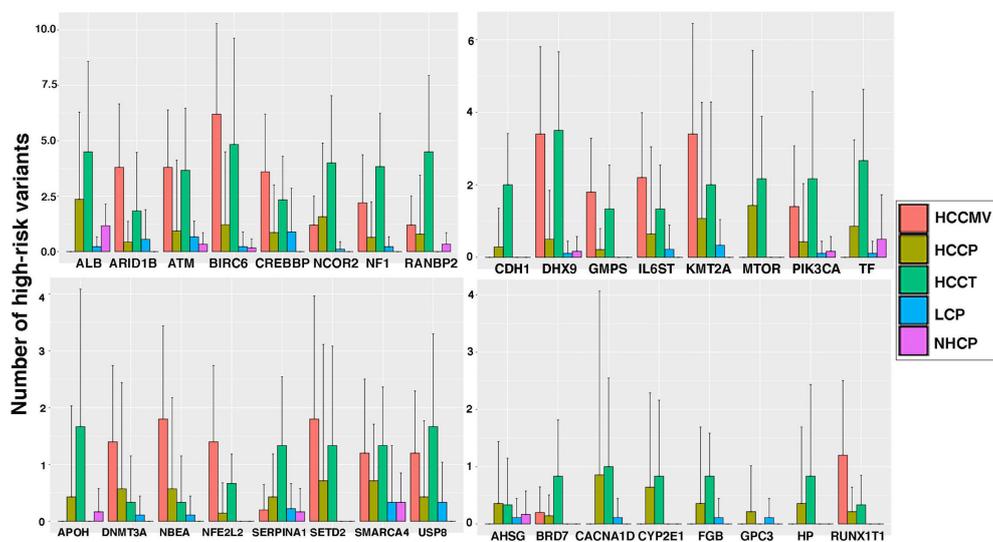
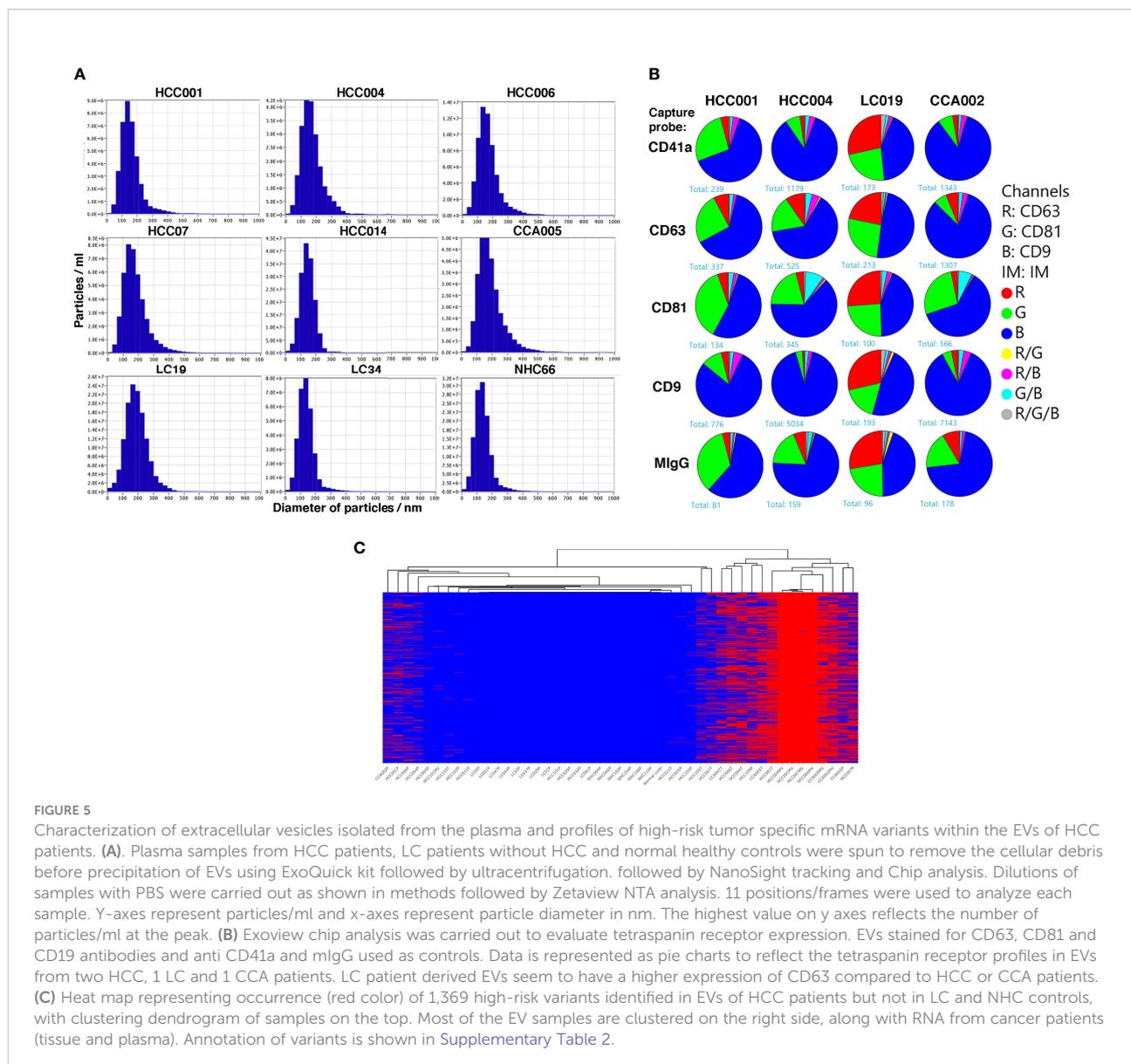


FIGURE 4

Liver specific transcripts in HCC patients reflect higher levels of high-risk variants than those from liver cirrhosis patients and healthy controls. The bar graphs represent the average counts of high-risk variants on 32 liver specific transcripts in several sample groups and error bar showing the standard deviation. Clearly plasma from HCC patients show significantly higher mutation load than that of LC patients.



Common concordant variants selectively enriched in circulating EVs of HCC patients

RNAseq analyses of circulating EVs from five HCC patients were performed and the results were compared with RNAseq from their corresponding tumors. A set of concordant variants selectively enriched in EVs of these patients are highlighted (Figure 5C; Supplementary Table 2). Briefly, as shown in Figure 5C, 218 high-risk concordant variants detected in tumors and plasma derived EVs of HCC patients were identified. Variants identified in this category correspond to some known HCC related transcripts like GOLGA2, ASAH1, ODC2, SAR1A, PIK3R1, CCNG1, MACF1, SERPINE1, CTNBN1, CREB3 and ADH5, among others. We have not investigated and compared these variants to the EVs from LC or NHC samples. However,

none of these EV-enriched variants are detected in the plasma of NHC controls or LC patients. The data indicates unique EV enrichment of these circulating variants.

Analyzing matching samples, whole plasma, circulating EVs and tumors from same cancer subjects indicated triple-concordant high-risk variants

Significantly, comparing three different specimens from the same cancer subjects identified a set of triple concordant variants not detected in the circulation of LC patients or normal healthy controls. Hundreds of such triple concordant variants were detected in a set of 8 cancer patients with matched specimens (Figure 6A). Venn diagrams show the distribution of these

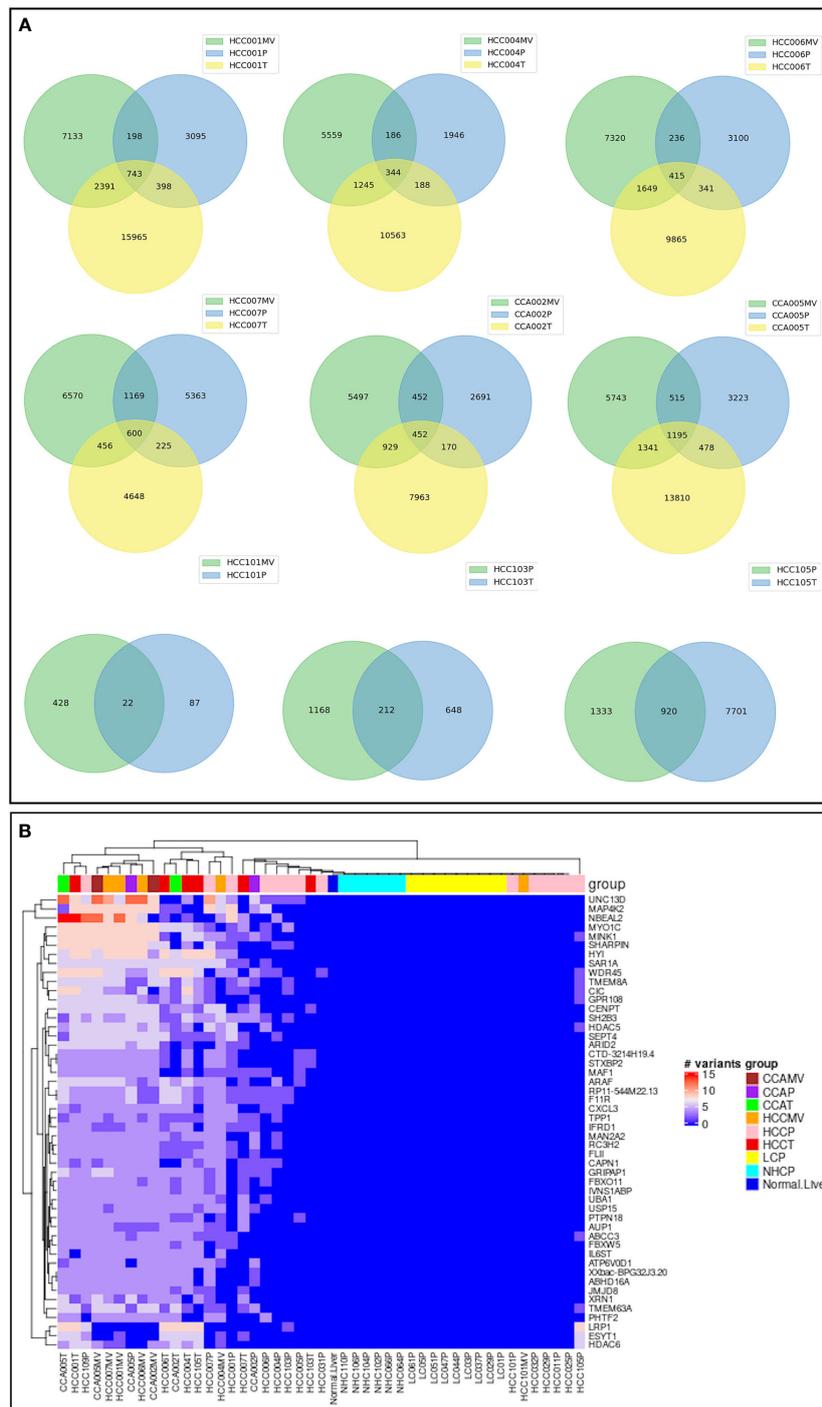


FIGURE 6 Concordance of high-risk variants in three matching samples from same cancer patients. **(A)** Venn diagrams corresponding to 6 cancer patients (4 HCC, 2 CCA) with three matching specimens, EVs, tumor and plasma from each subject and three subjects with two matching specimens. The numbers represent the distribution and concordance of high-risk mRNA variants in multiple samples from same patients. **(B)** Heat map representation of top genes carrying highest number of high-risk concordant variants. The number of high-risk variants, which were identified simultaneously in different types of samples (plasma, tumor, and/or vesicle) from same cancer patient, but not presented in any of liver cirrhosis or healthy control, were summarized by gene carrying them. The first 50 genes carrying the most of such high-risk concordant variants were selected and the numbers of variants on those genes were plotted against samples. UNC13D shown at the top, for example has the highest number of variants in HCC patient specimens. Annotation and profiles of these variants can be found in the Supplementary Table 3.

variants in matching specimens from same subjects and strongly suggest an association of these high-risk variants with cancer. We sorted triple concordant high-risk variants from all patients and organized the corresponding genes based on number of variants (Figure 6B; Supplementary Table 3). The heat map in Figure 6B represents the distribution of top genes sorted based on high number of high-risk triple concordant variants. The data highlight the top altered genes mutated frequently at several critical regions and underscores the authenticity and functional relevance of these ctRNA variants. For example, UNC13D is a calcium dependent cytoplasmic protein involved in vesicular and endocytic transport essential for vesicle maturation and docking and also known as an unfavourable prognostic marker in renal, endometrial and pancreatic cancer (40). Nurobeachin like 2 (NBEAL2) plays a role in development and secretion of alpha granules containing growth factors and is a prognostic marker for head and neck cancer (40). MAP4K2 is a serine/threonine protein kinase and an essential component of MAP kinase signaling. It acts as an upstream activator of stress activated protein kinase/c-jun N-terminal kinase (SAP/JNK) and p38 pathways. MAP4K2 can be activated by TNF α and interacts with TRAF2. MAP4K2 is known as an unfavourable prognostic marker of liver cancer (40). Presence of multiple high-risk variants on these transcripts reflected in all three specimens from same cancer subjects and lack of detection in non-cancer control subjects highlights their role in cancer pathology.

Variants associated with CCA

Cholangiocarcinoma is a liver cancer of cholangiocytes, and not hepatocytes. It was therefore of interest and opportunistic to examine a set of specimens of matched plasma and tumor for high-risk mRNA transcripts. Matched CCA plasma, CCA EVs and tumors from CCA patients were studied as additional liver cancer controls to test the specificity of HCC-specific variants. Surprisingly, most of the highly frequent high-risk variants in HCC patients were also detected in CCA samples. We sorted variants on the basis of their proportion in CCA patients (Figure 7; Supplementary Table 4). As in HCC, detection of high-risk variants in CCA patients was more frequent in tumors than plasma samples. EV-enriched HCC associated variants were also detected in the EVs of CCA patients. This concordance of ctRNA variations between HCC and CCA is interesting and underscores the functional relevance of such variants and validates their specificity in liver cancer.

Variants corresponding to transcripts involved in cancer and metabolic pathways suggest widespread metabolic alterations in HCC

The high-risk mRNA variants we have identified are selectively present in the people with HCC. Given that the

genes corresponding to these variants are well known players in tumor biology we hypothesize that these variants reflect the mutational environment of the tumor cells, from which we believe they are derived. To emphasize the importance of mutations in cancer biology, a multi-analyte blood test named as CancerSEEK based on circulating proteins and cfDNA mutations has been reported to detect 8 common cancers (41). Malignancies including HCC are characterized by the presence of mutated genes, possibly a result of altered tumor metabolism, DNA repair and splicing and our unique finding is that these genetic lesions can be detected in circulating transcripts.

Though we have not characterized any NGS derived concordant tumor-centric variants, high-risk variants identified in HCC patients using GATK and SnpEff tools correspond to transcripts of EGFR, CTNNA1, PTPA, AKT1, MTOR, ARAF, RAF1, HRAS, CREB3, SREBF1, ACY1, STAT3, etc. These genes correspond to well-known cancer pathways. We carried out KEGG pathway enrichment analysis on 578 genes corresponding to tumor-centric variants (Supplementary Figure 1) and were able to identify major pathways corresponding to variant genes. A profound and predominant representation of circulating tumor-centric variants correspond to genes involved in metabolism suggesting some interplay in altered tumor metabolism (Figure 8). Concordant variants of FASN, PC, SREBF1, ACADVL, PKLR, PCK1, ALDOC, PSAT1, MAT1A, MAT2A, GPT, etc. point to the altered genes associated with dysregulated tumor metabolism. In comparison, none of these tumor-centric variants were detected in the circulation of LC or NHC subjects. Variants corresponding to several other liver enriched transcripts like IDH2, PROC, PKLR, ACAA1, KLKB1, ACY1, ASL, MASP2, ECI1, ADH1A, ADH1B, ADH6, HADHA, etc. are well known metabolic enzymes facilitating tumor adaptation and some of these transcripts represent potential or FDA approved drug targets. Variants of GPT and PSAT1 are liver enriched and play role in glutamine metabolism and signaling. Variant transcripts associated with insulin signaling pathways (40) like PIK3R1, SLC2A2, SLC27A, PTPRF, OGT, NR1H, etc. are among the concordant ctRNA variants detected in both plasma and tumors of HCC patients.

The preliminary data are exciting and provocative. Though the number of samples analyzed is small, the data illustrate a high mutation milieu in HCC patients and highlight the diagnostic potential of circulating HCC-specific mRNA variants detectable in a non-invasive manner.

Discussion

Detection of tumor-derived genetic material from peripheral blood offers a non-invasive platform with an additional advantage that a full spectrum of mutations in tumors can be detected compared to tumor biopsies that may miss variants not uniformly distributed due to tumor/nodular heterogeneity (16).

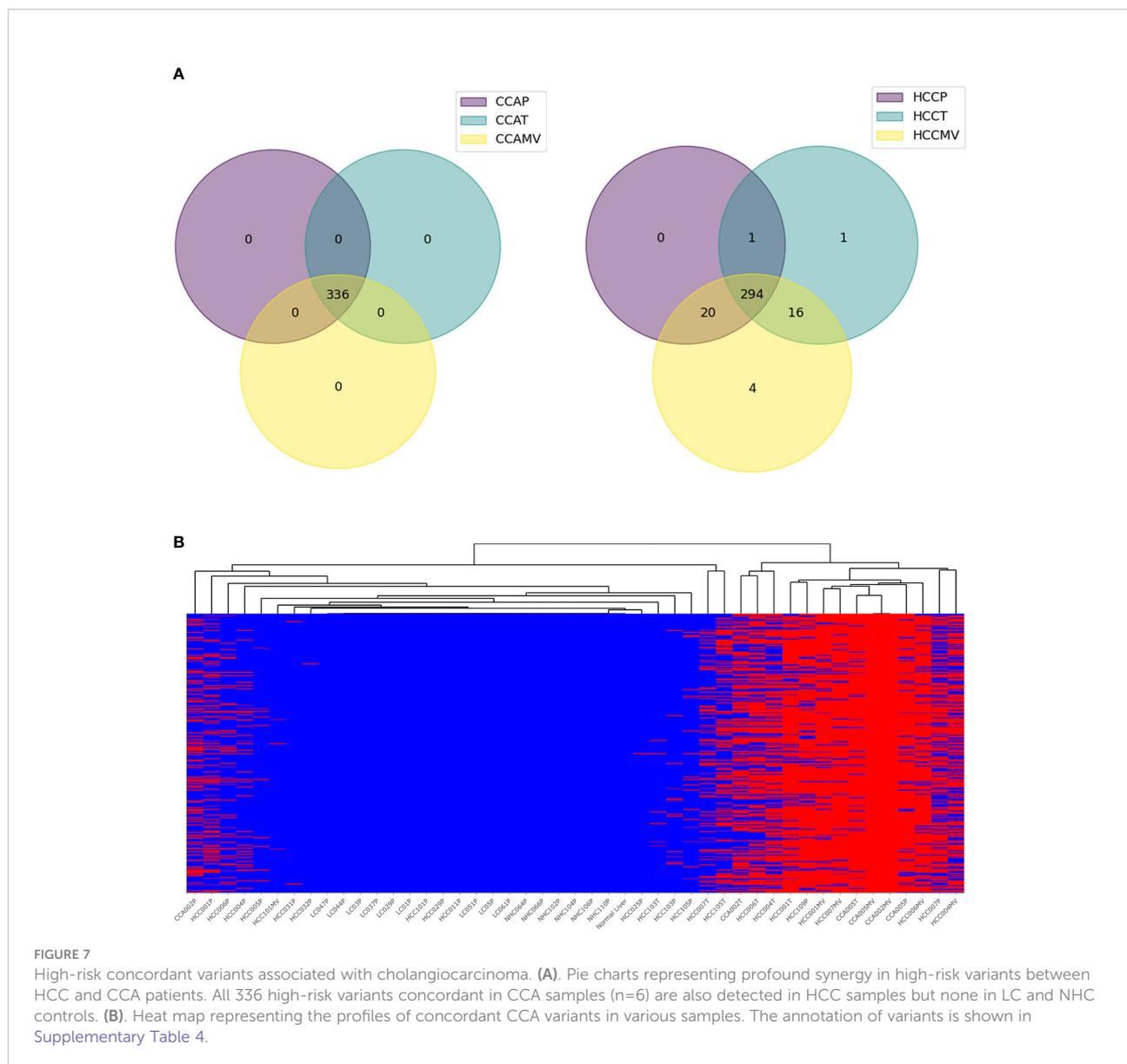


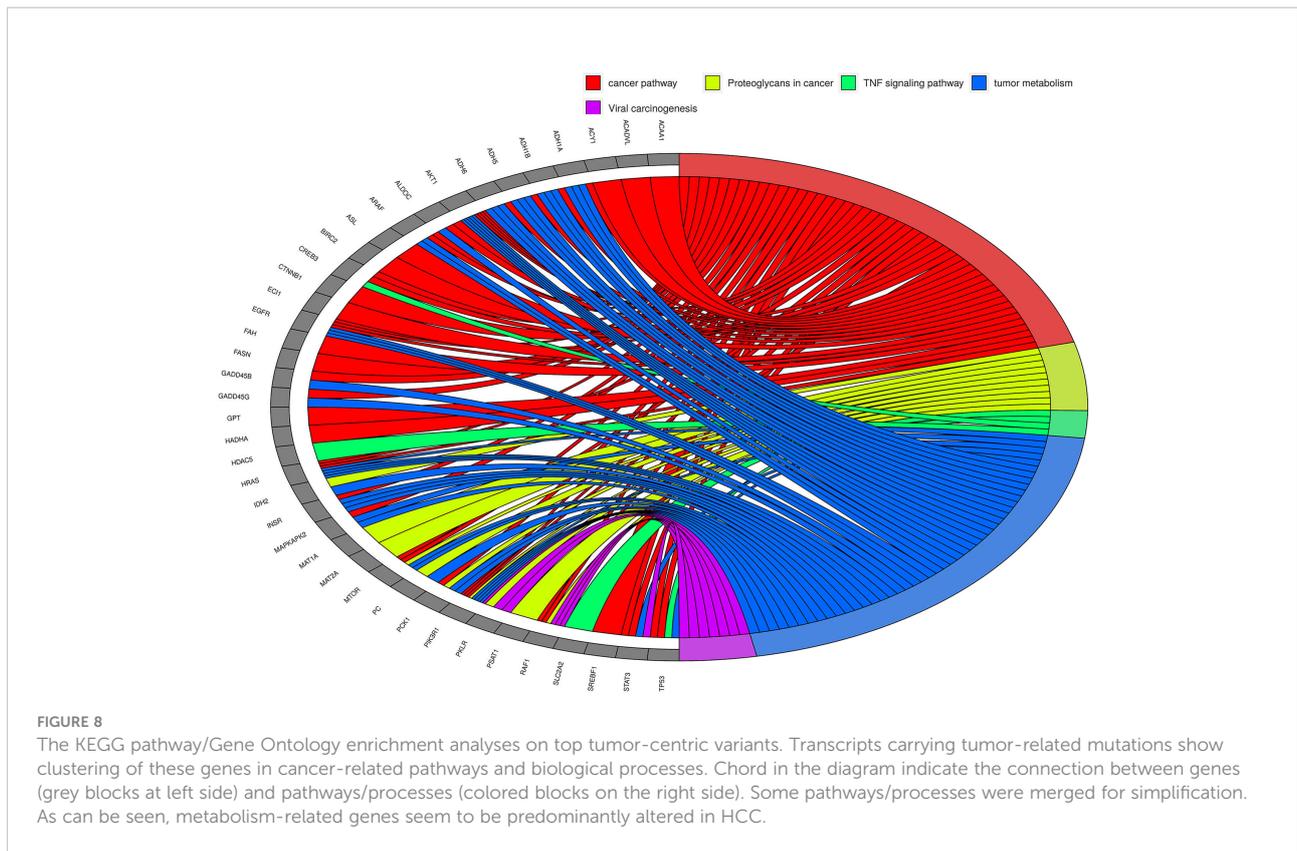
FIGURE 7

High-risk concordant variants associated with cholangiocarcinoma. (A) Pie charts representing profound synergy in high-risk variants between HCC and CCA patients. All 336 high-risk variants concordant in CCA samples ($n=6$) are also detected in HCC samples but none in LC and NHC controls. (B) Heat map representing the profiles of concordant CCA variants in various samples. The annotation of variants is shown in Supplementary Table 4.

Mutations are ubiquitous in cancer and the association between somatic mutations and cancer is well recognized (42). The inflammatory environment in hepatitis, MAFLD and cirrhosis may serve as a premalignant stage, the development of HCC requires progressive pre-neoplastic changes and further disease progression leads to enhanced tumor heterogeneity and increased mutation load (43, 44). Mutational signature is known to be largely conserved across the liver (45, 46).

An increasing number of studies demonstrate the potential use of cfDNA in diagnosis, prognosis, and monitoring of cancer. A multi-cancer early detection CancerSeek blood test based on specific DNA mutations and protein biomarkers when combined with PET-CT has shown progress in early detection (47). HCCscreen test (Genetron) claims detection of early-stage

HCC with high sensitivity based on DNA mutations and protein biomarkers (48). Mutations within a tumor can be clonal or subclonal, and the amount of available genome copies is a limiting factor for the detection of variants of low allele frequency (49, 50). Moreover, the tumor fraction of circulating DNA (ctDNA) varies between cancer types as well as between patients affected by the same cancer type (40). Even at the metastatic stage, some patients can yield a low amount of ctDNA (51, 52). Challenges in the optimization and standardization of preanalytical steps, significant gaps in our understanding of its origin, physical properties, and dynamics in circulation call for alternative approaches which can address and overcome low allelic frequency and limited copy number of target variants (24, 27, 53, 54).



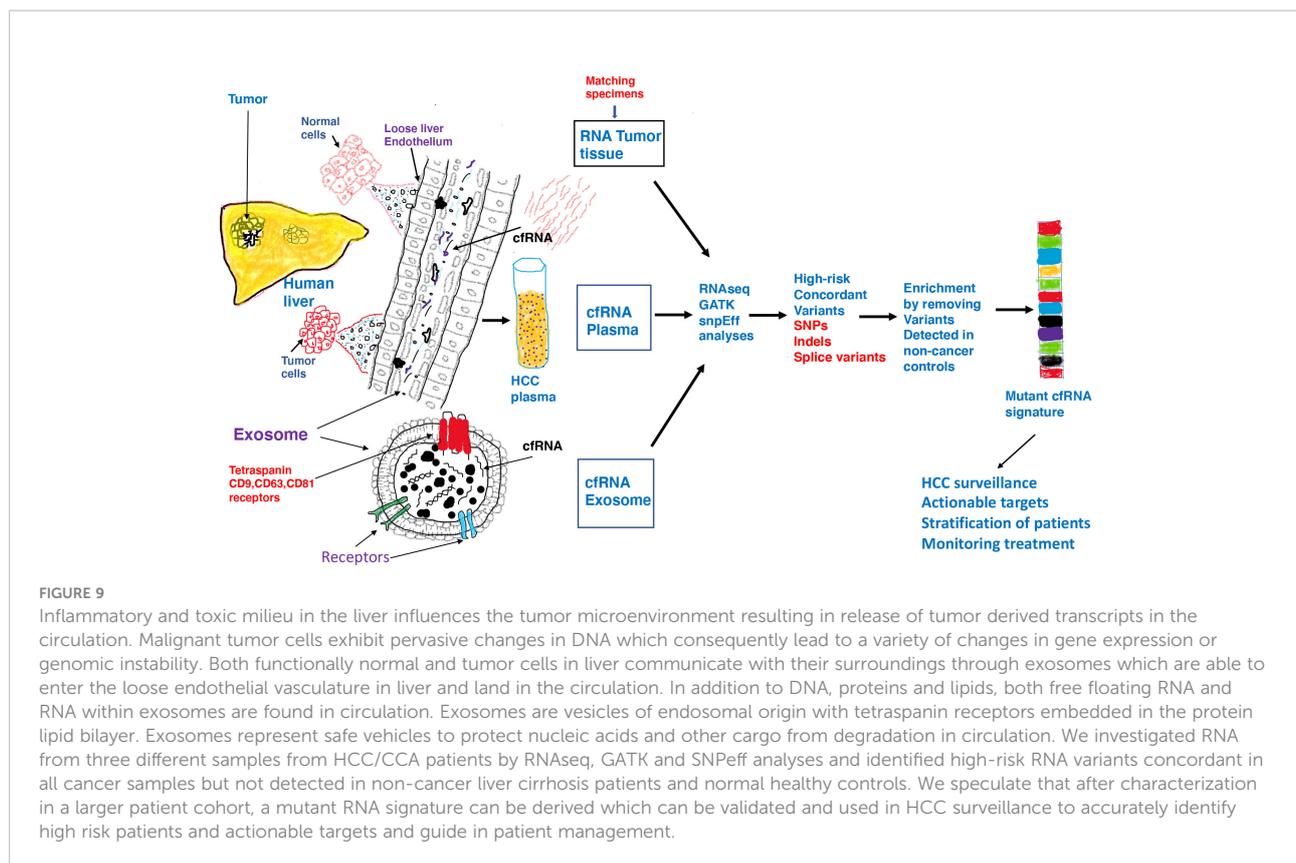
The presence of mRNA sequences in the circulation has begun to be recognized (55–58). Our laboratory is the first to demonstrate mRNA as a powerful analyte with significant upregulation of HCC-specific transcripts in the circulation of HCC patients compared to high-risk patients without cancer (33). Another study using circulating EVs reported a ten-gene RNA expression signature, consistent with our data, exhibiting high sensitivity in distinguishing early-stage HCC from LC patients (59). We now provide evidence of specific high-risk mutations in tumor-derived liver transcripts in the circulation of HCC patients (60) and identify a panel of 9 ctmutRNA with high diagnostic performance (data not shown). Studies on circulating RNA (ctRNA) is an emerging field in diagnostics (57, 58) and transcriptomics-based biomarkers are reported to exhibit higher efficiency than protein biomarkers of HCC (61).

Here, in this report, by profiling the mutational landscape of the tumor in parallel with that of circulation from HCC patients we identify some top highly frequent 1500 circulating high-risk mRNA variants concordant in tumors and plasma but not detected in the plasma of liver cirrhosis patients or normal healthy controls (Figure 9). GATK and SNPeff tools lead to the identity of high-risk variants comprised of both SNP's and indels most of which seem to be reported in catalogue of somatic mutations in cancer (COSMIC) and dbSNP data bases (Supplementary Table 6). Several high-risk variants correspond to cancer driver genes and cancer signaling pathways. Some of these variants correspond to genes

known as actionable targets with either potential or FDA approved therapies. Significantly, a large subset of HCC-specific high-risk variants were also detectable in the tumors, EVs and plasma of few cholangiocarcinoma patients but not in LC or normal healthy individuals.

Splicing variants represent a predominant category of high-risk mRNA variants detected in tumors, EVs and free plasma of cancer patients; none of these are detected in normal liver or plasma of LC patients and normal healthy individuals. Alternative isoforms and tumor-specific isoforms that arise from aberrant splicing during the liver tumorigenesis have been reported earlier (62) and long read RNA sequencing in liver cells and tumors have identified spliced variants corresponding to enzyme regulators, chromatin modifiers, RNA-binding proteins and receptors (63). A high level of differential splicing is known to occur in primary HCC tumor tissues compared with normal liver, and many of these changes have been shown to correlate with patient survival (64, 65). It is to be noted that any pathological changes associated with transcription and RNA splicing/editing aberrations cannot be identified through ctDNA. Splicing and fusion variants in ctRNA together with point mutations could be potentially more effective in identifying early stage and high-risk patients.

We demonstrate that plasma derived extracellular vesicles (EV) are enriched for a subset of high-risk cancer associated mRNA variants which were also detected in matching tumors.



EVs provide a mechanism of how tumor-derived variants land in the circulation, are protected from ubiquitous nucleases and variant enrichment provides better chances of detection in circulation. Significantly, overall high-risk variant density was highest in tumors, followed by EVs and lower in free plasma. This suggests that EVs might work as better tools as repositories of mRNA variants compared to free plasma.

Profiling the mutational landscape of the liver through circulating RNA in patients would be a paradigm shifting approach and potentially provide information about tumor biology, patient prognosis and likely response to therapy. Circulating variants of mRNA can offer a better understanding of genetic heterogeneity and if some constitutional variants predispose individuals to a specific molecular subtype of liver tumors. It could enable us to evaluate the exact function of the gene, if not already known, and the consequence of its loss of function in the liver. With regard to the efforts towards precision medicine in HCC, treatment decisions will become increasingly dependent upon genetic stratification and ctRNA variant profiles associated with different etiologies of chronic liver diseases could be very useful at the bedside. The identification of oncogenic activation of tyrosine kinases in some advanced NSCLC tumors, most notably mutations in the *EGFR*, *ALK* or *ROS1* gene, has led to a paradigm shift and the development of specific molecular treatments for patients (66). Moreover, any

pathological changes associated with transcription and RNA editing aberrations cannot be identified through DNA. It is becoming apparent that tumor biology (67) and treatment responses, even for immunotherapy (68), may be predictable based on mutational subtypes. Non-invasive and cost-effective access to these circulating mRNA variants could potentially revolutionize clinical management.

Since the fidelity of circulating RNA will be critical for clinical decision making, optimization and standardization of clinical sample collection, RNA isolation and analyses are critical for development of this platform. Our data clearly demonstrate the possibility of this platform to be used for clinical translation. Specific liver associated ctRNA biomarkers could potentially identify patients with liver cirrhosis, MAFLD and malignant pathology. Samples could be easily drawn linearly from patients and repeated for confirmation, at various time points of surveillance and treatment and treatment decisions could be tailored to mutational load of HCC-specific transcripts.

Though identification of these ctRNA variants in HCC patients was carried out using next generation sequencing (NGS) analyses, further validation of these need to be carried out in larger patient cohorts and variants characterized by other technical approaches. In summary, we demonstrate that circulating mRNA variants can potentially offer a viable liquid biopsy platform which on its own or in combination with

cfDNA or protein biomarkers can significantly aid in surveillance, early detection and patient management.

Data availability statement

All processed data in our study has been submitted in the manuscript in the form of figures and supplementary tables. Raw unprocessed FASTQ files can be accessed via a publicly available NCBI SRA database with accession number PRJNA907745 using this URL link: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA907745>.

Ethics statement

The studies involving human participants were reviewed and approved by 1. Human Studies Subcommittee, Deptt. of veteran affairs Med Center, Univ. of Pennsylvania, PA. 2. IRB Capital Health cancer Center, NJ. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AS and TB conceptualized and designed the study. AS, TB and DZ developed methodology, planned and carried out experiments, analyzed data and provided interpretive insights. SZ helped carry out extracellular vesicle isolation. JL carried out mapping, GATK and SNPeff and other RNAseq analyses. TZ played a role in biostatistical analyses. DK and CD provided the clinical specimens and reviewed the data. AS and JL created the figures and AS drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by a Commonwealth University research enhancement program grant with Pennsylvania department of Health (AS); The department disclaims responsibility for any analyses, interpretation, or conclusions. NIHR01 CA166111(DEK). This work was also supported by private funding sources viz Blake and Steve Miller foundation.

Conflict of interest

TB is on the board of the Directors of Hepion, a company developing medicines for liver diseases, Merlin, a company developing medicines for cancer and CIRNA, a startup

company created by the Baruch S Blumberg Institute developing early detection of cancer diagnostics. AS is a co-founder of CIRNA and an equity holder. DEK is the site investigator for Glycotest Inc and Exact Sciences biomarker studies.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.963641/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

The KEGG pathway enrichment analysis conducted with R package cluster Profiler using "Over-representation analysis" based on hypergeometric distribution in the subset of genes with tumor-centric mutations. Over-representation of 578 genes with tumor-centric mutations in KEGG pathway/Gene Ontology (GO) terms.

SUPPLEMENTARY TABLE 1

Annotation and characteristics of top 1500 concordant high-risk variants found in tumors and plasma of HCC patients corresponding to Figure 3. Gene and genomic location, dbSNP rs numbers if available, variant effects and proportion/frequency in various clinical subsets are shown. The order of variants in the table corresponds to the order in the clustering heat map in Figure 3.

SUPPLEMENTARY TABLE 2

Annotation and characteristics of 1,369 high-risk tumor specific mRNA variants within the EVs of HCC patients corresponding to Figure 5C showing clustering dendrogram of samples. Gene and genomic location, dbSNP rs numbers if available, variant effects and proportion/frequency in various clinical subsets are shown. The order of variants in the table corresponds to the order in the clustering heat map in Figure 5C.

SUPPLEMENTARY TABLE 3

Annotation of 716 concordant variants identified simultaneously in two to three matching specimens (plasma, tumor, and/or vesicle) from same cancer patients, but not present in any of liver cirrhosis or healthy controls. Gene and genomic location, dbSNP rs numbers if available, variant effects and proportion/frequency in various clinical subsets are shown. The first 50 genes carrying the most of such high-risk concordant variants are shown in Fig 6B.

SUPPLEMENTARY TABLE 4

Annotation and characteristics of top 336 concordant high-risk variants found in tumors and plasma of CCA patients corresponding to Figure 7B.

Gene and genomic location, dbSNP rs numbers if available, variant effects and proportion/frequency in various clinical subsets are shown. The order of variants in the table corresponds to the order in the clustering heat map in Figure 7B.

SUPPLEMENTARY TABLE 5

The frequency of tumor-related 5,155 variants in tumor/control samples were compared with Fisher's exact probability test under various group settings, i.e., HCC plasma vs. all the controls (normal plasma and cirrhosis plasma), HCC tumor vs all the controls (normal plasma and cirrhosis plasma), HCC vesicle vs all the controls (normal plasma and cirrhosis plasma), and all the HCC samples (plasma, tumor and vesicles) vs. all the controls (normal plasma and cirrhosis plasma). Variants are sorted on the basis of decreasing proportion in HCC tumors. Columns showing proportion of variants in HCC tumors (column DG) and p value between HCC tumors and all non tumor controls (column DQ) are highlighted in gray color. HCC - all includes all samples, tumors, plasma and extracellular vesicles from HCC patients. Tumor-all includes both HCC

and CCA tumors. Control-all includes all samples corresponding to non-cancer liver cirrhosis patients and those of normal healthy controls.

SUPPLEMENTARY TABLE 6

All the variants called from the 46 samples were filtered by depth (≥ 5) and quality score (≥ 20), and then annotated with SnpSift 5.1d and dbNSFP 4.3a to get clinvar information. There are 6,109 variants which had been reported in clinvar database. Clinical trait and clinical significance of each variant are indicated in the table. POS: location of variant; ID: dbSNP id; REF: Reference sequence; ALT: Altered sequence; QUAL: Quality of variant calling in phred score ($-\log_{10}$ [probability of false calling]); INFO: information of variant including annotation by SnpSift.

SUPPLEMENTARY TABLE 7

Characteristics of HCC and CCA patients whose specimens were used. CH, NJ: Capital Health Cancer Center, NJ. UPEN, PA: Veterans hospital University of Pennsylvania, PA. Biochemed.

References

- Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol*. (2019) 16(10):589–604. doi: 10.1038/s41575-019-0186-y
- American Cancer Society. *Cancer.org* (2019). Available at: <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2019>.
- Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the united states. *Cancer Res* (2014) 74(11):2913–21. doi: 10.1158/0008-5472
- Mittal S, Sada YH, El-Serag HB, Kanwal F, Duan Z, Temple S, et al. Temporal trends of nonalcoholic fatty liver disease-related hepatocellular carcinoma in the veteran affairs population. *Clin Gastroenterol Hepatol*. (2015) 13(3):594–601.e1. doi: 10.1016/j.cgh.2014.08.013
- Hirode G, Saab S, Wong RJ. Trends in the burden of chronic liver disease among hospitalized US adults. *JAMA Network Open* (2020) 3(4):e201997. doi: 10.1001/jamanetworkopen.2020.1997
- Dang K, Hirode G, Singal AK, Sundaram V, Wong RJ. Alcoholic liver disease epidemiology in the united states. *Am J Gastroenterology*. (2020) 115(1):96–104. doi: 10.14309/ajg.0000000000000380
- Yang JD, Harmsen WS, Slettedahl SW, Chaiteerakij R, Enders FT, Therneau TM, et al. Factors that affect risk for hepatocellular carcinoma and effects of surveillance. *Clin Gastroenterol Hepatol*. (2011) 9(7):617–23. doi: 10.1016/j.cgh.2011.03.027
- Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology* (2017) 67(1):358–80. doi: 10.1002/hep.29086
- Kanwal F, Hoang T, Kramer JR, Asch SM, Goetz MB, Zeringue A, et al. Increasing prevalence of HCC and cirrhosis in patients with chronic hepatitis c virus infection. *Gastroenterology* (2011) 140(4):1182–1188.e1. doi: 10.1053/j.gastro.2010.12.032
- El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* (2012) 142(6):1264–1273.e1. doi: 10.1053/j.gastro.2011.12.061
- Kanwal F, Kramer JR, Li L, Dai J, Natarajan Y, Yu X, et al. Effect of metabolic traits on the risk of cirrhosis and hepatocellular cancer in nonalcoholic fatty liver disease. *Hepatology* (2020) 71(3):808–19. doi: 10.1002/hep.31014
- Singal AG, Zhang E, Narasimhan M, Rich NE, Waljee AK, Hoshida Y, et al. HCC surveillance improves early detection, curative treatment receipt, and survival in patients with cirrhosis: A meta-analysis. *J Hepatol* (2022) 77(1):128–39. doi: 10.1016/j.jhep.2022.01.023
- European Association for the Study of the Liver, EASL organization. EASL clinical practice guidelines for the management of patients with decompensated cirrhosis. *J Hepatol*. (2018) 69(2):406–60. doi: 10.1016/j.jhep.2018.03.024
- Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American association for the study of liver diseases. *Hepatology* (2018) 68(2):723–50. doi: 10.1002/hep.29913
- Tzartzeva K, Obi J, Rich NE, Parikh ND, Marrero JA, Yopp A, et al. Surveillance imaging and alpha fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis: A meta-analysis. *Gastroenterology*. (2018) 154(6):1706–18. doi: 10.1053/j.gastro.2018.01.064
- Parikh AR, Leshchiner I, Elagina L, Goyal L, Levovitz C, Siravegna G, et al. Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med* (2019) 25:1415–21. doi: 10.1038/s41591-019-0561-9
- Best J, Bechmann LP, Sowa J-P, Sydor S, Dechêne A, Pflanz K, et al. GALAD score detects early hepatocellular carcinoma in an international cohort of patients with nonalcoholic steatohepatitis. *Clin Gastroenterol Hepatol*. (2020) 18(3):728–735.e4. doi: 10.1016/j.cgh.2019.11.012
- Yang JD, Addissie BD, Mara KC, Harmsen WS, Dai J, Zhang N, et al. GALAD score for hepatocellular carcinoma detection in comparison with liver ultrasound and proposal of GALADUS score. *Cancer Epidemiol Biomarkers Prev* (2018) 28(3):531–8. doi: 10.1158/1055-9965.epi-18-0281
- Bruix J, Sherman M. Management of hepatocellular carcinoma: An update. *Hepatology* (2011) 53(3):1020–2. doi: 10.1002/hep.24199
- Singal AG, Pillai A, Tiro J. Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: A meta-analysis. *Klenerman P. PLoS Med* (2014) 11(4):e1001624. doi: 10.1371/journal.pmed.1001624
- Kim TH, Kim SY, Tang A, Lee JM. Comparison of international guidelines for noninvasive diagnosis of hepatocellular carcinoma: 2018 update. *Clin Mol Hepatol* (2019) 25(3):245–63. doi: 10.3350/cmh.2018.0090
- Wolf E, Rich NE, Marrero JA, Parikh N, Singal AG. Utilization of hepatocellular carcinoma surveillance in patients with cirrhosis: A systematic review and meta-analysis. *Hepatology* (2020) 73(2):713–25. doi: 10.1002/hep.31309
- Ayuso C, Rimola J, Vilana R, Burrel M, Darnell A, García-Criado Á, et al. Corrigendum to “Diagnosis and staging of hepatocellular carcinoma (HCC): Current guidelines”. *Eur J Radiology*. (2019) 112:229. doi: 10.1016/j.ejrad.2019.01.018
- Wan JCM, Massie C, Garcia-Corbacho J, Moulriere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* (2017) 17:223–38. doi: 10.1038/nrc.2017.7
- Heitzer E, Perakis S, Geigl JB, Speicher MR. The potential of liquid biopsies for the early detection of cancer. *NPI Precis Oncol* (2017) 1(1):36. doi: 10.1038/s41698-017-0039-5
- Bronkhorst AJ, Ungerer V, Holdenrieder S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol Detect Quantif*. (2019) 17:100087. doi: 10.1016/j.bdq.2019.100087
- Abbosch C, Birkbak NJ and Swanton C. Early-stage NSCLC - challenges to implementing ctDNA-based screening and MRD detection. *Nat Rev Clin Oncol* (2018) 15(9):577–86. doi: 10.1038/s41571-018-0058-3
- Cisneros-Villanueva M, Hidalgo-Pérez L, Rios-Romero M, Cedro-Tanda A, Ruiz-Villavicencio CA, Page K, et al. Cell-free DNA analysis in current cancer

clinical trials: a review. *Br J Cancer* (2022) 126:391–400. doi: 10.1038/s41416-021-01696-0

29. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* (2019) 570:385–9. doi: 10.1038/s41586-019-1272-6
30. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* (2011) 12(9):628–40. doi: 10.1038/nrg3046
31. Krawczak M, Ball EV, Fenton I, Stenson PD, Abeyasinghe S, Thomas N, et al. Human gene mutation database—a biomedical information and research resource. *Hum Mutat* (2000) 15(1):45–51. doi: 10.1002/(SICI)1098-1004(200001)15:1<45::AID-HUMU10>3.0.CO;2-T
32. Wu J, Jiang R. Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases. *ScientificWorldJournal* (2013) 2013:675851. doi: 10.1155/2013/675851
33. Sayeed A, Dalvano BE, Kaplan DE, Viswanathan U, Kulp J, Janneh AH, et al. Profiling the circulating mRNA transcriptome in human liver disease. *Oncotarget* (2020) 11(23):2216–32. doi: 10.18632/oncotarget.27617
34. Vander Auwera GA. From fastq data to high confidence variant calls: the genome analysis tool kit best practices pipeline. *Curr Protoc Bioinf* 43(1110):11.10.1–11.10.33. doi: 10.1002/0471250953.b1110s43
35. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸; iso-2; iso-3. *Fly (Austin)*. (2012) 6(2):80–92. doi: 10.4161/fly.19695
36. Park KJ, Park JH. Variations in nomenclature of clinical variants between annotation tools. *Lab Med* (2021) 53(3):lmab074. doi: 10.1093/labmed/lmab074
37. Odegaard JJ, Vincent JJ, Mortimer S, Vowles JV, Ulrich BC, Banks KC, et al. Validation of a plasma-based comprehensive cancer genotyping assay utilizing orthogonal tissue- and plasma-based methodologies. *Clin Cancer Res* (2018) 24(15):3539–49. doi: 10.1158/1078-0432.CCR-17-3831
38. Strickler JH, Loree JM, Ahronian LG, Parikh AR, Niedzwiecki D, Pereira AAL, et al. Genomic landscape of cell-free DNA in patients with colorectal cancer. *Cancer Discovery* (2018) 8(2):164–73. doi: 10.1158/2159-8290.CD-17-1009
39. Muller M, Bird TG, Nault JC. The landscape of gene mutations in cirrhosis and hepatocellular carcinoma. *J Hepatology*. (2020) 72:990–1002. doi: 10.1016/j.jhep.2020.01.019
40. Nault JC. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cancer Genome Atlas Res Network. Hepatol* (2020) 71(1):164–82. doi: 10.1016/j.cell.2017.05.046
41. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* (2018) 359(6378):926–30. doi: 10.1126/science.aar3247
42. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. *Science* (2015) 349(6255):1483–9. doi: 10.1126/science.aab4082
43. Bettgowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Trans Med* (2014) 6(224):224ra24. doi: 10.1126/scitranslmed.3007094
44. Pelusi S, Baselli G, Pietrelli A, Dongiovanni P, Donati B, McCain MV, et al. Rare pathogenic variants predispose to hepatocellular carcinoma in nonalcoholic fatty liver disease. *Sci Rep* (2019) 9(1):3682. doi: 10.1038/s41598-019-39998-2
45. Blokzijl F, Ligt JD, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*; (2016) 538(7624):260–26. doi: 10.1038/nature19768
46. Brunner SF, Roberts ND, Wylie LA, Moore L, Aitken SJ, Davies SE, et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* (2019) 574(7779):538–42. doi: 10.1038/s41586-019-1670-9
47. Lennon AM, Buchanan AH, Kinde I, Warren A, Honushefsky A, Cohain AT, et al. Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. *Science* (2020) 369(6499):eabb9601. doi: 10.1126/science.abb9601
48. Qu C, Wang Y, Wang P, Chen K, Wang M, Zeng H, et al. Detection of early-stage hepatocellular carcinoma in asymptomatic HBsAg-seropositive individuals by liquid biopsy. *Proc Natl Acad Sci U S A*. (2019) 116(13):6308–12. doi: 10.1073/pnas.1819799116
49. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nature Genetics* (2012) 44(6):694–8. doi: 10.1038/ng.2256
50. Ally ACancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* (2017) 169(7):1327–1341.e23. doi: 10.1016/j.cell.2017.05.046
51. Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* (2012) 4(162):162ra154. doi: 10.1126/scitranslmed.3004742
52. McBride DJ, Orpana AK, Sotiriou C, Joensuu H, Stephens PJ, Mudie LJ, et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer*. (2010) 49(11):1062–9. doi: 10.1002/gcc.20815
53. Ye Q, Ling S, Zheng S, Xu X. Liquid biopsy in hepatocellular carcinoma: circulating tumor cells and circulating tumor DNA. *Mol Cancer*. (2019) 18(1):114. doi: 10.1186/s12943-019-1043-x
54. Su YH, Kim AK, Jain S. Liquid biopsies for hepatocellular carcinoma. *Transl Res* (2018) 201:84–97. doi: 10.1016/j.trsl.2018.07.001
55. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci* (2008) 105(30):10513–8. doi: 10.1073/pnas.0804549105
56. Enache L, Enache E, Ramière C, Diaz O, Bancu L, Sin A, et al. Circulating RNA molecules as biomarkers in liver disease. *Int J Mol Sci* (2014) 15(10):17644–66. doi: 10.3390/ijms151017644
57. Xue VW, Cheung MT, Chan PT, Luk LLY, Lee VH, Au TC, et al. Non-invasive potential circulating mRNA markers for colorectal adenoma using targeted sequencing. *Sci Rep* (2019) 9:12943. doi: 10.1038/s41598-019-49445-x
58. Akat KM, Lee YA, Hurlley A, Morozov P, Klaas EA, Max KEA, et al. Detection of circulating extracellular mRNAs by modified small-RNA-sequencing analysis. *JCI Insight* (2019) 4(9):e127317. doi: 10.1172/jci.insight.127317
59. Sun N, Lee YT, Zhang RY, Kao R, Teng PC, Yang Y, et al. Purification of HCC-specific extracellular vesicles on nanosubstrates for early HCC detection by digital scoring. *Nat Commun* (2020) 11(1):4489. doi: 10.1038/s41467-020-18311-0
60. Sayeed A, Doria C, Kaplan DE, Block T. Detection of tumor derived mutant mRNA transcripts in the circulation of HCC patients. *Hepatology* (2021) 683–4. doi: 10.1002/hep.32188
61. Gupta R, Kleinjans J, Caiment F. Identifying novel transcript biomarkers for hepatocellular carcinoma (HCC) using RNA-seq datasets and machine learning. *BMC Cancer*. (2021) 21(1):962. doi: 10.1186/s12885-021-08704-9
62. Lee SE, Alcedo KP, Kim HJ, Snider NT. Alternative splicing in hepatocellular carcinoma. *Cell Mol Gastroenterol Hepatol* (2020) 10(4):699–712. doi: 10.1016/j.jcmgh.2020.04.018
63. Chen H, Gao F, He M, Ding XF, Wong AM, Sze SC, et al. Long-read RNA sequencing identifies alternative splice variants in hepatocellular carcinoma and tumor-specific isoforms. *Hepatology* (2019) 70(3):1011–25. doi: 10.1002/hep.30500
64. Li S, Hu Z, Zhao Y, Huang S, He X. Transcriptome-wide analysis reveals the landscape of aberrant alternative splicing events in liver cancer. *Hepatology* (2019) 69(1):359–75. doi: 10.1002/hep.30158
65. Zhu GQ, Zhou YJ, Qiu LX, Wang B, Yang Y, Liao WT, et al. Prognostic alternative mRNA splicing signature in hepatocellular carcinoma: a study based on large-scale sequencing data. *Carcinogenesis* (2019) 40(9):1077–85. doi: 10.1093/carcin/bgz073
66. Jacquelyne P, Xiuning L, Heymach JV, Robichaux. structure-based classification predicts drug response in EGFR-mutant NSCLC. *Nature*. (2021) 597:732–7. doi: 10.1038/s41586-021-03898-1
67. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology* (2017) 152(4):745–61. doi: 10.1053/j.gastro.2016.11.048
68. Harding JJ, Nandakumar S, Armenia J, Khalil DN, Albano M, Ly M, et al. Prospective genotyping of hepatocellular carcinoma: Clinical implications of next-generation sequencing for matching patients to targeted and immune therapies. *Clin Cancer Res* (2019) 25(7):2116–26. doi: 10.1158/1078-0432.CCR-18-2293