Clinical application of molecular barcode sequencing for detection of low frequency variants in circulating tumor DNA of hepatocellular carcinoma

ESL KIM

Department of Medical Science

The Graduate School, Yonsei University

Clinical application of molecular barcode sequencing for detection of low frequency variants in circulating tumor DNA of hepatocellular carcinoma

ESL KIM

Department of Medical Science

The Graduate School, Yonsei University

Clinical application of molecular barcode sequencing for detection of low frequency variants in circulating tumor DNA of hepatocellular carcinoma

Directed by Professor JONG RAK CHOI

The Master's Thesis submitted to
the Department of Medical Science,
the Graduate School of Yonsei University in partial
fulfillment of the requirements for the degree of
Master of Medical Science

ESL KIM

June 2020

This certifies that the Master's Thesis
of ESL KIM is approved.

Thesis Supervisor : JONG RAK CHOI

Thesis Committee Member#1 : SEUNG-TAE LEE

Thesis Committee Member#2 : JAE MYUN LEE

The Graduate School

Yonsei University

June 2020

ACKNOWLEDGEMENTS

대학 졸업 후 직장생활을 하며 생물학 및 암 등 사람의 질병에 대한 보다 깊은 지식과 최신 과학 기술에 대한 이해도를 높이고자 시작된 석사학위 과정이 어느덧 마무리를 짓게 되었습니다. 학위 과정 동안 수많은 우여곡절이 있었지만, 뒤돌아보면 각 순간들이 의미가 있었다는 것을 깨닫게 되었고, 이 자리를 빌어 그 과정에서 도움을 주셨던 분들께 감사의 인사를 드리고자 합니다. 먼저, 제가 관심있는 ctDNA관련 연구를 할 수 있게 기회를 주신 최종락 교수님께 감사드립니다. 연구 진행 및 실험 관련하여 조언을 해주신 이승태 교수님과 따뜻한 격려를 해주신 신새암 교수님 그리고 논문 심사과정에서 조언과 격려를 해주신 이재면 교수님께도 감사드립니다. 또한 많은 관심과 좋은 논문이 될 수 있도록 격려해주고 조언을 아끼지 않으신 서지은 박사님께도 깊은 감사를 드립니다. 학위 과정을 시작할 때부터 지금까지 수많은 우여곡절 속에서 동료이자 친구로 그리고 실험실의 선배로써 더 발전된 연구원으로 성장할 수 있도록 함께 해준 이현아 선생님께도 감사드립니다. 그리고 이현아 선생님과 함께 같이 토론하고 실험에 대해 논의와 조언을 해준 윤우빈 선생님께도 감사드립니다. 또한, 반복되는 일상에도 항상 즐겁게 같이 일해주는 박미리 연구원과 각자 맡은 바를 잘 해주고 있는 랩 구성원들께도 감사를 드립니다. 그리고 언제나 저를 사랑해주시고 학업에 대한 조언을 아끼지 않으셨던 부모님과 멀리 있지만 항상 응원해준 동생 시원이에게 감사의 말을 전하고 싶습니다. 학위 과정은 새로운 소중한 인연들을 만나고 또 늘 곁에 있던 사람들의 소중함을 알게 된 시간들이었습니다. 이제 새로운 출발을 준비하는 시점에서 이곳에서 많은 도움을 받았듯이, 저도 베풀며 노력하는 사람이 되도록 노력하겠습니다.

TABLE OF CONTENTS

ABSTRACT 1

I. INTRODUCTION 3

II. MATERIALS AND METHODS 9

1. Patients enrollment and sample collection 9
2. Plasma separation from peripheral blood 9
3. Panel design and target selection 9
4. Cell free DNA extraction from plasma 13
5. Library preparation of cell free DNA 13
6. Hybridization capture-based target enrichment 16
7. Molecular barcode sequencing and data processing 17
8. Pilot study for detection of target genes under cell free DNA mimicking condition 18
9. Pilot study for detection of allelic frequency using cell free DNA reference materials 19
10. Amplicon preparation of TERT promoter region 21
11. Amplicon sequencing and data processing 21
12. Statistical analysis 21

III. RESULTS 22

1. Demographical and clinical-pathological features of hepatocellular carcinoma and liver diseases cohort 22
2. Molecular barcode sequencing quality analysis 26
3. Detection of target genes under circulating cell free DNA mimicking

condition 28
4. Limit of detection of molecular barcode sequencing using cell free DNA reference materials 28
5. Impact of circulating cell free DNA concentration on demographical and Cliopathological features in hepatocellular carcinoma and liver diseases patients 30
6. Genetic landscape of hepatocellular carcinoma and liver diseases .. 32
7. Circulating tumor DNA analysis of hepatocellular carcinoma patients and concordance with tissue 38
8. Survival analysis and hazard ratio depending on most frequent observed gene in hepatocellular carcinoma and liver diseases 40
IV. DISSCUSSION 43
V. CONCLUSION 47
REFERENCES 48
ABSTRACT (IN KOREAN) 54

LIST OF FIGURES

Figure 1. Graphical schematic flow chart of the study 8
Figure 2. Result of general next-generation sequencing and molecular barcode sequencing using custom fragmented DNA and cfDNA reference materials 29
Figure 3. Clinical features of cfDNA in hepatocellular carcinoma and liver diseases. 31
Figure 4. Population Frequency of hepatocellular carcinoma and liver disease patients 34
Figure 5. Variants Frequency in hepatocellular carcinoma and liver disease patients 35
Figure 6. Genetic landscape of hepatocellular carcinoma and liver diseases 36
Figure 7. Comparison of the number of variants detected using MBS based on demographical and cliopathological features. 37
Figure 8. Survival analysis of hepatocellular carcinoma related gene 39
Figure 9. Hazard ratio of three most frequently observed genes in hepatocellular carcinoma 41
Figure 10. Hazard ratio of genetic alteration in hepatocellular carcinoma 42

LIST OF TABLES

Table 1. List of selected genes for targeted next-generation sequencing for cell-free DNA of hepatocellular carcinoma . 11
Table 2. Selected regions of hepatocellular carcinoma cell free DNA target panel..... 12
Table 3. Mutation of Cell lines and genomic DNA of cell line used for circulating tumor DNA mimicking condition..... 20
Table 4. Demographical features and survival status of hepatocellular carcinoma and benign and chronic liver diseases patients..... 24
Table 5. Clinical and pathological characteristics of hepatocellular carcinoma patients 25
Table 6. Depth of coverage for each targeted genes in the HCC-cfDNA panel 27

ABSTRACT

Clinical application of molecular barcode sequencing for detection of low frequency variants in circulating tumor DNA of hepatocellular carcinoma

ESL KIM

Liver cancer is predicted to be the sixth most commonly diagnosed cancer and the fourth leading cause of cancer-related death worldwide in 2018 and hepatocellular carcinoma accounts for seventy-five to eighty-five percent of total liver cancer cases. Similar to other solid tumor cancer, the diagnosis of hepatocellular carcinoma is also required a tissue biopsy from a patient to determine the malignancy of tumors. The tissue biopsy is, however, an invasive procedure in patients and cannot depict various genetic characteristics of the tumors as it collects a sample from a limited region in the tumors. To decrease the risk in patients and to overcome challenges associated with tumor heterogeneity, liquid biopsy, a less invasive procedure than the tissue biopsy, has been widely studied and applied in both research and diagnostic levels. Among various biomarkers of liquid biopsy, circulating tumor DNA, cell-free DNA from tumors, is utilized as both diagnostic and prognostic biomarkers for cancer because it represents the current status of tumors at the molecular level as it is originated from cancer tumors. Comparing to the isolation of other liquid biopsy biomarkers such as cell-free RNA and exosome from blood, extraction

protocols of cell-free DNA are more straightforward and its concentration in blood is higher than that of other liquid biopsy biomarkers.

Next-generation sequencing has arisen in the past decade as a proficient technique for sequencing genetic materials and has been rapidly adopted in molecular diagnostic laboratories to detect genetic mutations in various diseases including cancers and rare diseases. Targeted sequencing, a type of next-generation sequencing, allows detecting variants with a limit of detection as low as five percent of allele frequency under a depth of coverage of five-hundred. To detect the lower variant allelic frequency of variants from circulating cell-free DNA by using next-generation sequencing, but still, an increasing amount of genetic materials and greater depth of coverage are required. Unique molecular barcode sequencing has been developed to overcome the lower compatibility of input amount, higher error rate, and less cost-effectiveness to detect rare and low allelic frequency variants in genetic materials.

A total of one-hundred forty-nine patients were enrolled in this study and whole blood samples were collected in cell-free DNA tubes for collection of plasma samples and twenty-seven of the total tissue samples were collected during surgery or tissue biopsy. Circulating tumor DNAs from the plasma samples were extracted, libraries were constructed with unique molecular barcode embedded adapters and captured and, targeted sequencing was performed to detect somatic variants in the circulating tumor DNA. The limit of detection of molecular barcode sequencing was as low as 0.5 percent of allelic frequency and the molecular barcode sequencing of circulating tumor DNA demonstrated matched somatic variants in both plasma and tumor samples. It suggests that liquid biopsy using circulating tumor DNA could be an alternative method of tissue biopsy and could enable a study of low allelic frequency variants or rare variants in plasma from malignant diseases using molecular barcode sequencing.

Clinical application of molecular barcode sequencing for detection of low frequency variants in circulating tumor DNA of hepatocellular carcinoma

ESL KIM

Department of Medical Science
The Graduate School , Yonsei University

(Directed by Professor JONG RAK CHOI)

I. INTROUDCUTION

Liver cancer is predicted to be the sixth most commonly diagnosed cancer and the fourth leading cause of cancer-related death in 2018[1]. Both incidence rate and mortality rate of liver cancer are two to three times higher among men than in women worldwide and the mortality rate of liver cancer in the Republic of Korea is the highest in the economically-active age group and the second highest across all age-groups in 2018[1,2]. The main types of primary liver cancer are hepatocellular carcinoma (HCC) accounting for 75- 85% of total liver cancer cases and intrahepatic cholangiocarcinoma (ICC) comprising 10- 15% of the cases as well as other rare types such as liver angiosarcoma (HA) and hepatoblastoma (HBL)[1,2]. HCC is also known for its high incidence and prevalence, and poor prognosis and survival rate, for example, the overall 5-year survival rate of HCC was 30% [4]. According to Barcelona clinic liver cancer (BCLC) staging system, the estimated survival time is longer than five years if a patient is diagnosed as very early stage such as BCLC stage 0 or A. HCC, however, is typically diagnosed at a late stage such as BCLC stage B, C and D and the estimated survival time is gradually decreased into two years, one year and three months, respectively[5].

Chronic infection with hepatitis B virus (HBV) or hepatitis C virus (HCV), heavy alcohol intake, obesity, aflatoxin-contaminate foodstuffs, smoking, and type 2 diabetes are the main risk factors for HCC[1,3]. As there are several main risk factors for HCC, the major risk factors of HCC vary from region to region. Among the highest rates of liver cancer regions, China, Eastern Africa, and Mongolia for instance, HBV infection and aflatoxin exposure are the key determinants whereas Japan and Egypt for example, HCV infection is the predominant cause of HCC[1,3]. Chronic hepatitis virus (HBV and HCV) infections are also the most frequent causes of HCC in Korea[2]. According to registry data and studies of the Korean Liver Cancer Association (KLCA) and the Korea Central Cancer Registry (KCCR), patients with HBV infected HCC were 62.2% and patients with HCV infected HCC were 10.4% of total HCC cases and alcoholic liver disease and nonalcoholic fatty liver disease (NAFLD) are presumed as unknown causes of HCC accounted for the remaining 27.4%[2].

For surveillance of HCC, several methods are used to detect tumors et early stages, increase the opportunity to use curative treatments, and improve survival rate in patients. The standard HCC surveillance tests are liver ultrasonography and serum-based tests[3,5]. For a long time, one of the serum biomarkers, alpha-fetoprotein (AFP) has been widely used with liver ultrasonography[6,7]. AFP, however, has a wide cut-off range between 7.7-112.0 ng/mL with 25-90% sensitivity diagnostic serum biomarker[6,7]. To overcome low sensitivity of the test, the other serum biomarker, Des-gamma-carboxy prothrombin (DCP) also known as a protein induced by vitamin K absence or antagonist-II (PIVKA-II) is also used with adjacent to AFP. PIVKA-II also has a wide cut-off range between 40-150 mAU/mL with 44-91% sensitivity[6,7]. Using these tests, overall sensitivity is 40-60% at cut-off 20ng/ml of AFP, which are still low for sensitivity for diagnostic markers for HCC[3,5]. Diagnosis of HCC, is furthermore, required tissue biopsy from patients to observe the histological characterization of the tumor, and reveal genetic profiling of tumor for prediction of disease progression and response to therapies[8]. Conventional tissue biopsy is still invasive to patients and

the result of tissue biopsy shows poor prognosis as it only represents the genetic characteristic of a biopsy site regardless of tumor heterogeneity[8,9]. It is also difficult to monitor patients' disease progression and therapeutic response as tissue biopsy is difficult to repeat and, the metastatic site and the primary site have extensive intertumoral and intratumoral evolution[8-10]. Thus, a single tissue biopsy is likely to underestimate the complexity of tumor genomics[8-10].

The liquid biopsy is a minimally invasive method that detects diagnostic, prognostic, and therapeutic biomarkers in various cancers such as non-small lung carcinoma and breast cancer[10,11]. It is currently investigated by many investigators, especially for a complementary approach to the solid tissue biopsy for cancer patients[8-11]. As it utilizes non-solid biological tissue and body fluids, primarily blood, liquid biopsy is an easily repeatable approach for sampling of patients who needs follow-up period for progression and treatment[8-11]. There are several types of liquid biopsy biomarkers such as exosomes, circulating tumor cells (CTCs), and cell-free DNA(cfDNA)[8-11]. Among the liquid biopsy biomarkers, cfDNA is currently widely investigated as the concentration of cfDNA is higher than that of CTC or exosome and the extraction method of cfDNA is easier and less time consuming than that of other biomarkers[10-14]. The cfDNA, discovered by Mandel and Metais in 1948[15], is circulating in the blood and is predominantly derived from apoptotic and necrotic cells but also released by living eukaryotic cells[8-13,15-16]. The size of cfDNA is 180-200 base pairs (bp) as when the macrophage phagocytosis is exhausted nucleosome amounts are increased and released into the bloodstream[16-18]. The average concentration of plasma cfDNA in the healthy and normal subjects is 10 ng per mL[19]. The half-life of cfDNA is between 15 minutes and 2.5 hours[18,19]. These characteristics can be both advantage and disadvantage of cfDNA. The concentration of cfDNA in the blood is known as higher in cancer patients, thus more cfDNA can be obtained from the patients[8-14]. It, however, is not always higher in the patients as the release of cfDNA is

affected by aging, the weight of subjects, and the status of treatment and surgery[8-14]. Thus, the concentration of cfDNA cannot be fully determined as a diagnostic biomarker for cancers[11-14]. Furthermore, the short half-life of cfDNA indicates that it consists of information about genetic alterations of primary and metastatic tumors at the time point, concurrently, it can be easily degraded if a plasma is not separated from peripheral blood within several hours and yield of cfDNA is gradually decreased over time[18,19].

Recently, next-generation sequencing (NGS) is widely applied as an experimental tool to successfully deal with the above problems, by producing a large amount of genomic data from cfDNA. NGS has arisen in the past decade as a proficient technique for sequencing genetic materials such as DNA, RNA, and single cells from various species and has been rapidly adopted in molecular diagnostic laboratories to detect genetic mutation in cancers and rare diseases[13,20]. Contrasting to other techniques such as Sanger sequencing and PCR based method, NGS can perform massively parallel sequencing that can process and generate data of several patients′ samples with different genomic regions in a short amount of time[13]. Another noteworthy advantage of NGS technology comparing to Sanger sequencing, it became more efficient, more sensitive, and is becoming less expensive as NGS is compatible with a low amount of DNA[20,21]. In clinical oncology, sequencing of targeted NGS panels can detect mutations as low as 5% of the allelic frequency with 500 depth of coverage[20,22]. Despite these aspects, to detect low variant allele frequency (VAF) less than 5%, the greater amount of DNA and the greater depth of coverage are required which means decreased compatibility of input amount, higher error rate, and less cost-effectiveness in processing NGS.

To detect rare and low allele frequency variants, unique molecular barcode sequencing (MBS) is developed in both experimental and data processing levels. MBS degenerates molecular barcode to label each fragmented DNA molecules with its own unique DNA sequence[23,24]. Molecular barcode embedded adapters

contain random yet complementary double-stranded nucleotide sequences and 6-12 bp in length and it becomes feasible to trace every sequence read back to one of the two strands of the original double-stranded DNA molecules[23-25]. When each parental DNA strand is labeled after adapter ligation and PCR-amplification, sequences with the same tag are amplified and can be eligible to perform NGS. After sequencing, generated data is processed by grouping reads of each tag as a family and establishing consensus sequence from each of two strands to form single-strand consensus sequences (SSCS)[23,24]. After making SSCS, the two complementary consensus sequences derived from the two strands are compared with each other and establish a duplex consensus sequence (DCS) if the two strands match perfectly at each position. If there is only one mutation of the two DNA strands, the mutation is assumed as a result of PCR errors or sequence errors during DNA damage and is not counted as true mutation[23,24]. This experimental and computational process of MBS yields higher accuracy to detect rare and low variant allele frequency[23-25].

In this study, a total of one-hundred forty-nine benign and malignant liver disease patients and hepatocellular carcinoma patients were enrolled and plasma samples from the patients were collected for targeted sequencing of cfDNA. Twenty-seven tissue samples of HCC patients were collected during surgery or tissue biopsy and targeted sequencing of tissue genomic DNA was performed whether liquid biopsy could replace tissue biopsy by comparing characteristics of somatic variants in both cfDNA from plasma and genomic DNA from tissue. The NGS of cfDNA was performed to detect somatic mutations and to analyze genetic characteristics of hepatocellular carcinoma comparing to benign and malignant liver diseases. Furthermore, MBS was utilized to identify low allelic frequency variants and to improve the detection rate of genetic alterations of the diseases under the condition of a limited amount of cfDNA from the samples and considering cost and time effectiveness of the process of the study (Figure 1).

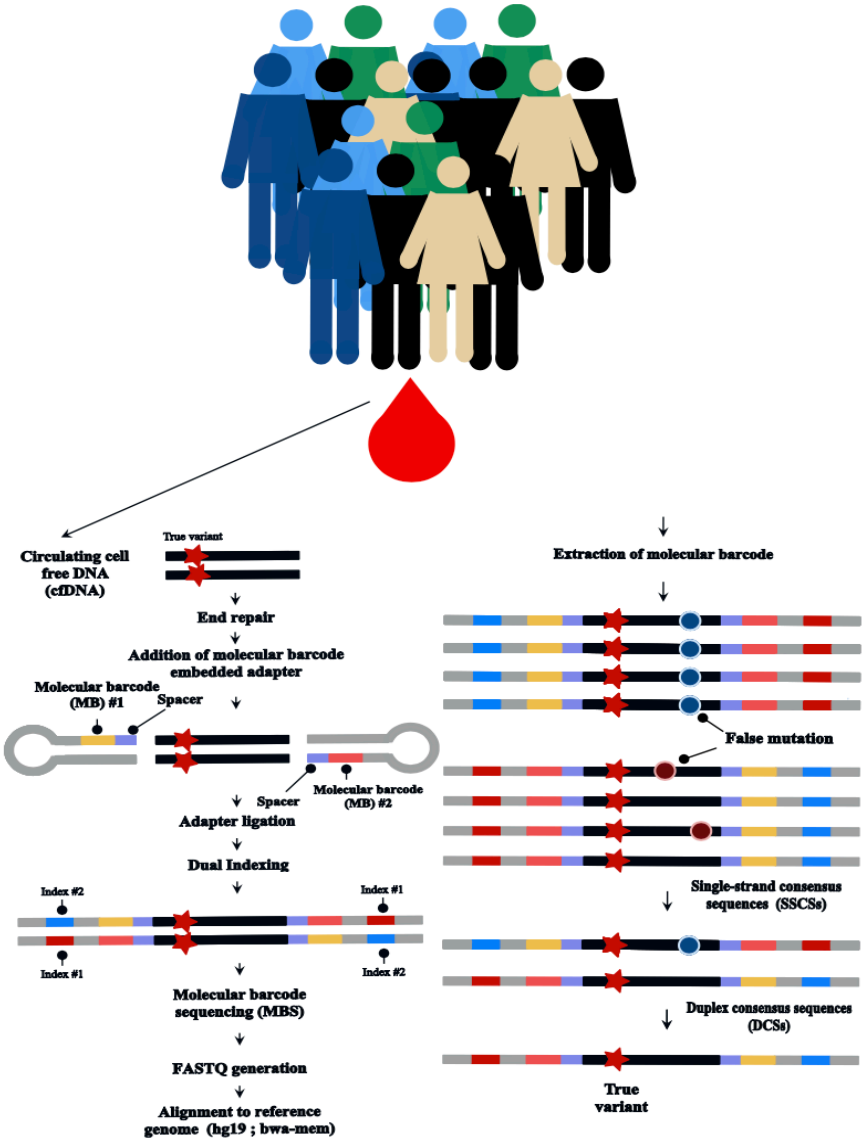

Figure 1. Graphical schematic flow chart of the study. Benign liver diseases, chronic liver diseases and HCC patients were enrolled and their whole blood samples were collected to extract cfDNA. The cfDNA libraries were prepared with the molecular barcodes embeded adapters and the MBS was performed. The sequencing data were processed to detect true genetic alterations in cfDNA via assembly of single-strand consensus sequences (SSCS) and duplex consensus sequences (DCS) and annotated by using public databases.

II. MATERIALS AND METHODS

1. Patient enrollment and sample collection

This study was conducted under approval from the Institutional Review Board at Yonsei University Severance Hospital, Seoul, Korea (IRB No. 4-2015-0184). To be eligible for inclusion into this study, subjects had to have met the following criteria: 1) age ≥ 18 years, and 2) histologically or radiologically confirmed as hepatocellular carcinoma or liver diseases. An 8 mL peripheral blood sample per patient was collected in a Cell-Free DNA BCT tube (Streck, La Vista, NE, USA).

2. Plasma separation from peripheral blood

Collected peripheral blood was processed for plasma separation within 48 hours. The plasma was separated from the whole blood by centrifugation under the following protocol. The whole blood was centrifuged at 1,600 g for 10 minutes at room temperature and the upper layer of plasma was transferred to a new conical tube. The upper layer of plasma was then centrifuged at 16,000 g for 10 minutes at room temperature to remove any cellular debris. The plasma supernatant was transferred to 1.5 mL microcentrifuge tubes and stored under refrigeration at -80°C until further use.

3. Panel design and target selection

To design an HCC-related cfDNA panel for NGS targeted sequencing, key genes and pathways in HCC were selected based on public databases and journal reviews. According to the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Catalogue of Somatic Mutations In Cancer (COSMIC), HCC-related pathways are hepatic differentiation, chromatin remodeling, Wnt signaling, oxidative stress, RAS-MAPK signaling, PIK3K-AKT signaling, telomere maintenance, p53 signaling, and cell

cycle. According to Fujimoto *et al* (2016) and Schulze K *et al* (2016) and Nojiri S *et al* (2014), albumin and apolipoprotein B are involved in hepatic differentiation and acted as tumor suppressors and mutations occurred in these genes result in the proliferation of HCC cells. According to Huang J *et al* (2012), Cancer Genome Atlas Research Network (2017) and Castelli G *et al* (2017), AT-rich interaction domain 1A and domain 2 are involved in chromatin remodeling and genetic alterations in these genes result in hepatocarcinogenesis. Axin1 and catenin beta 1 are known as a tumor suppressor and an oncogene, respectively and mutations on both genes cause hepatocarcinogenesis and progression of HCC [28,34-37]. Mutations occurred in nuclear factor, erythorid2 like that are involved in oxidative stress, kirsten ras proto-oncogene GTPase involving in RAS/MAPK signaling, and phosphatidylinositol-4,5-bisphophate3-kinase catalytic subunit alpha involving in PIK3K/mTOR signaling result in hepatocarcinogenesis according to Fujimoto *et al* (2016) and Schulze K *et al* (2016) and Cancer Genome Atlas Research Network (2017). Telomerase reverse transcriptase involves in telomere maintenance and immortalization pathway and genetic alterations in the gene cause hepatocarcinogenesis[29,30,32]. Tumor protein p53 controls the cell cycle and acts as a tumor suppressor and mutations in the gene result in hepatocarcinogenesis and proliferation of HCC cells [26-37] (Table 1). Based on above reviews, selected key genes are: albumin (*ALB*), apolipoprotein B (*APOB*), AT-rich interaction domain 1A (*ARID1A*), AT-rich interaction domain 2 (*ARID2*), Axin1 (*AXIN1*), catenin beta 1 (*CTNNB1*), nuclear factor, erythorid2 like (*NFE2L2*), kirsten ras proto-oncogene GTPase (*KRAS*), phosphatidylinositol-4,5-bisphophate3-kinase catalytic subunit alpha (*PIK3CA*), telomerase reverse transcriptase (*TERT*) and tumor protein p53 (*TP53*) and regions of genes were selected for HCC and liver diseases targeted sequencing (Table 2).

Table 1. List of selected genes for targeted next-generation sequencing for cell-free DNA of hepatocellular carcinoma.

Gene	Involved Pathway	Result of Mutations in Gene	Reference
ALB	Hepatic differentiation	Loss of function as suppressor of hepatocellular carcinoma	[33],[34],[35],[36],[37],[38]
APOB	Hepatic differentiation	Proliferation of HCC cells	[34],[35],[36],[37],[39]
ARID1A	Chromatin remodeling	Loss of function as tumor suppressor, Hepatocarcinogenesis	[26] ,[27] ,[28] ,[31] ,[33], [34],[35],[36],[37]
ARID2	Chromatin remodeling	Loss of function as tumor suppressor, Hepatocarcinogenesis	[28], [34],[35],[36],[37]
AXIN1	WNT signaling	Loss of function as tumor suppressor, Hepatocarcinogenesis and progression of HCC	[27],[30] ,[31],[32] ,[33],[34],[35], [36],[37]
CTNNB1	WNT signaling	Activation of oncogene, Hepatocarcinogenesis and progression of HCC	[26] ,[27] ,[30] ,[31] ,[32] ,[33], [34],[35],[36],[37]
NFE2L2	Oxidative stress	Activation of oncogene, Hepatocarcinogenesis and progression of HCC	[34],[35],[36],[37]
KRAS	RAS/MAPK signaling	Activation of oncogene, Hepatocarcinogenesis	[34],[35],[36],[37]
PIK3CA	PIK3K/mTOR signaling	Activation of oncogene, Hepatocarcinogenesis	[34],[35],[36],[37]
TERT	Telomere maintenance	Immortalization pathway and hepatocarcinogenesis	[29],[31],[33] ,[34],[35],[36],[37]
TP53	Cell cycle	Loss of function as tumor suppressor, Hepatocarcinogenesis	[26],[27],[28],[30] ,[31] ,[32],[33], [34],[35],[36],[37]

ALB: Albumin; APOB: Apolipoprotein B; ARID1A: AT-rich interaction domain 1A; ARID2: AT-rich interaction domain 2; AXIN1; Axin1;
CTNNB1: Catenin beta 1; KRAS: Kirsten ras proto-oncogene,GTPase; PIK3CA: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha;
TERT: Telomerase reverse transcriptase; TP53: Tumor protein p53

Table 2. Selected regions of hepatocellular carcinoma cell free DNA target panel.

Gene	Location	Chromosome Position (GRCh37/hg19)	Region	GC content (%)
ALB	4q13.3	Chr4: 74269972-74287129	All	35.39
APOB	2p24.1	Chr2: 21224301-21266945	All	42.96
ARID1A	1p36.11	Chr1: 27022522-27108601	All	45.93
ARID2	12q12	Chr12: 46123494-46301820	All	36.08
AXIN1	16p13.3	Chr16: 337440-403244	All	54.55
CTNNB1	3p22.1	Ch3: 41236401-41281939	All	35.61
NFE2L2	2q31.2	Chr2: 178098733-178099000	Exon 2	42.91
KRAS	12p12.1	Chr12: 25398280-25398285, 25380275-25380277	Exon 2, Exon 3	53.33
PIK3CA	3q26.32	Chr3: 178936082-178936103	Exon 9	45.45
TERT	5p15.33	Chr5: 1295228-1295250	Promoter	91.30
TP53	17p13.1	Chr17: 7571720-7590868	All	49.39

ALB: Albumin; APOB: Apolipoprotein B; ARID1A: AT-rich interaction domain 1A; ARID2: AT-rich interaction domain 2; AXIN1; Axin1;
CTNNB1: Catenin beta 1; KRAS: Kirsten ras proto-oncogene,GTPase; PIK3CA: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha;
TERT: Telomerase reverse transcriptase; TP53: Tumor protein p53

4. Cell free DNA extraction from plasma

Circulating cell-free DNA was extracted from 2 mL plasma with Bioo Scientific NextPrep-Mag cfDNA Isolation Kit (PerkinElmer, Austin, TX, USA) under the manufacturer's instruction. Briefly, in 50 mL conical tube, 2 mL plasma with 2.5 mL binding solution, 48 µL proteinase K, and 32 µL magnetic beads were added, vortexed, and incubated for 15 minutes at 55°C. The tube was placed on a magnetic stand until the beads were completed attracted and the supernatant was discarded. 1.5 mL of Wash buffer 1 was added to resuspend beads and the slurry was transferred to a 2 mL microcentrifuge tube. The 2 mL tube was placed on a magnetic stand, the supernatant was discarded and the wash was repeated but the tube was not changed. 1.5 mL of Wash buffer 2 was added, the tube was placed on the magnetic stand, and the supernatant was discarded. This step was repeated one more time and a 24 µL elution solution was added and incubated for 5 minutes at 55°C. The cfDNA elute was transferred to a new 1.5 mL microcentrifuge tube, measured using Qubit™ dsDNA HS assay kit (Invitrogen, Q32854, Carlsbad, CA, USA) with Qubit®2.0 fluorometer (Invitrogen) and stored at -80°C.

5. Library preparation of cell-free DNA

5a. Library preparation with regular adapter

New England BioLabs (NEB)′s NEBNext Ultra II DNA library prep kit for Illumina (New England BioLabs, E7645L, Ipswich, MA, USA) were used for library preparation under manufacturer′s protocol. In brief, in 0.2 mL PCR tube, 3 µL of NEBNext Ultra II End Prep Enzyme Mix per reaction and 7 µL of NEBNext Ultra II End Prep Reaction Buffer per reaction and 50 µL of PCR-graded water with 50 ng of each cfDNA reference material were mixed and placed in the thermocycler with the heated lid set to 75°C and performed PCR reaction under condition of 20°C for 30 minutes, 65°C

for 30 minutes and 4°C for hold. During the PCR reaction, adapter was diluted with 10 mN Tris-HCL containing 10 mM NaCl, pH 7.5 into 10-fold. Then, 30 μL of NEBNext Ultra II Ligation Master Mix per reaction, 1 μL of NEBNext Ligation Enhancer per reaction, and 2.5 μL of diluted NEBNext Adapter for Illumina per reaction were mixed and added to the End Prep reaction mixture. The PCR tube is transferred to a thermocycler with the heated lid off with 20°C for 15 minutes incubation. 3 μL of USER enzyme per reaction was added to the ligation mixture and transferred to a thermocycler with the heated lid set to 47°C with 37°C for 15 minutes incubation. After the incubation, 0.9X beads wash with NEBNext sample purification beads were performed and eluted with 17 μL of 0.1X TE. Then, 15 μL of elution was transferred to a new 0.2 mL PCR tube and 25 μL of NEBNext Ultra II Q5 Master mix, 5 μL of i7 Primer, 5 μL of i5 primer per reaction and PCR amplification reaction was processed under condition of initial denature of 98°C for 30 seconds, 7 cycles of denature and annealing/extension of 98°C for 10 seconds and 65°C for 1 minute 15 seconds and final extension for 65°C for 5 minutes, then 4°C for hold. After PCR amplification, 0.9X beads wash with NEBNext sample purification beads were performed and eluted with 33 μL of 0.1X TE. 30ul of elution was transferred to a new PCR tube and stored at -20°C for further use. The library quality check was performed using D1000 screentape assay kit (Agilent Technologies, Inc. Santa Clara CA USA) with 4150 TapesStation system (Agilent).

5b. Library preparation with molecular barcode embedded adapter

The 2-50 ng of cfDNA was used to construct next-generation sequencing library with ThruPLEX® Tag-seq 96D Kits (Takara Bio USA, Inc. Mountain View, CA, USA) and manufacturer's instruction was followed to

construct the library of the samples. In brief, 10 μL of cfDNA samples were dispensed into each PCR tube and, 4 μL of Template Preparation Buffer and 1 μL of Template Preparation Enzyme were added into the each PCR tube. Mix thoroughly with a pipette and tubes were tightly cap and briefly centrifuged to ensure the entire volume of the reaction is collected at the bottom of each tube. The PCR tubes were placed in a thermal cycler with heated lid set to 105°C and performed the Template preparation reaction under a conditions with 22°C for 25 minutes, 55°C for 20 minutes and 22°C for Hold less than 2 hours. After the thermal cycler reached 22°C, the PCR tubes were removed and briefly centrifuged to proceed Library synthesis step. In each PCR tube, 2.5 μL of Library synthesis buffer and 2.5 μL of Library synthesis enzymes were added, mixed with a pipette, and centrifuged briefly. The PCR tubes were placed in a thermal cycler with heated lid set to 105°C and performed the Library synthesis reaction under a conditions with 30°C for 40 minutes, 4°C for Hold less than 30 minutes. After the thermal cycler reached 4°C, the PCR tubes were removed and briefly centrifuged to proceed Library amplification. In each PCR tube, 21.5 μL of Library amplification buffer, 1.0 μL of library amplification enzyme and 5 μL of indexing reagent were transferred, mixed with a pipette and briefly centrifuged. The PCR tubes were placed in a thermal cycler with heated lid set to 105°C and performed the Library amplification reaction under a conditions with 72°C for 4 minutes, 85°C for 2 minutes, 98°C for 2 minutes, 4 cycles of 98°C for 20 seconds, 67°C for 20 seconds and 72°C for 40 seconds, 4 to 11 cycles of 92°C for 20 seconds and 72°C for 50 seconds and 4°C for Hold. The library samples were measured using D1000 screentape assay kit (Agilent) with 4150 TapesStation system (Agilent).

6. Hybridization capture-based target enrichment

An Hybridization capture was performed with 8-library pool according to Twist Custom Panel Hybridization Capture of DNA Libraries (Twist Bioscience, South San Francisco, CA, USA). Briefly, in a 1.5 mL microcentrifuge tube, 187.5 ng of each library was added and total 8 libraries were in each 1.5 mL tube as one hybridization reaction. Pooled libraries in 1.5 mL tubes were dried using a vacuum concentrator under low heat condition. In a clean thin-walled PCR 0.2 mL strip-tube, a probe solution per one hybridization reaction was prepared by mixing 20 μL of Hybridization mix, 4 μL of Twist Custom panel and 4 μL of water. When library pools were dried, 5 μL of blocker solution and 7 μL of universal blockers were added to the 1.5 mL tubes to resuspend library pools and resuspension was transferred to a new 0.2 mL thin-walled PCR 0.2 mL strip tube. The probe solution was heated at 95°C for 2 minutes in a thermal cycler with the lid at 105°C, cooled on ice for 5 minutes. While probe solution is cooled on ice, heat the PCR tube containing the resuspension at 95°C for 5 minutes in a thermal cycler with the lid at 105°C, then both probe solution and resuspension were equilibrated at room temperature on the benchtop for 5 minutes. Probe solution was mixed into resuspension, briefly vortexed, and incubated at 70°C for 16 hours in a thermal cycler with the lid at 85°C. 100 μL of room temperature equilibrated Streptavidin binding beads per hybridization reaction was transferred to a 1.5 mL microcentrifuge tube. The binding beads were washed with 200 μL binding buffer for three times and final 200 μL of binding buffer were added to the beads and re-suspended the beads by vortexing the tube. After hybridization was completed, hybridization reaction was transferred into the resuspended beads and tube was placed on a rotator for 30 minutes. The tube was placed on a magnetic stand for 1 minute, supernatant was discarded and add 200 μL Wash buffer 1. The tube was vortexed and the

entire volume was transferred into a new 1.5 mL microcentrifuge tube. 200 µL of 48°C Wash buffer 2 was added to the new tube and incubated at 48°C for 5 minutes, the tube was placed on a magnetic stand for 1 minute, supernatant was discarded. 48°C Wash step was repeated two more times and the tube was placed on the magnetic stand and 45 µL of water was added for resuspension of streptavidin binding beads. In a 0.2 mL PCR tube, 22.5 µL of the streptavidin binding bead slurry, 2.5 µL of amplification primers, and 25 µL of KAPA HiFi hotstart readymix (KAPA biosystems, KK2602, Wilmington, MA, USA) were added and performed PCR under following condition: 98°C for 45 seconds, 11 cycles of 98°C for 15 seconds, 60°C of 30 seconds and 72°C for 30 seconds, and final extension of 72°C for 1 minute, 4°C for hold. With room temperature equilibrated DNA purification beads, 1.0x beads wash was performed by adding 50 µL of DNA purification beads in the PCR tube, incubating for 5 minutes at room temperature, two washed of 200 µL of 80% ethanol and incubated with 32 µL of water for 2 minutes. The tube was placed on a magnetic stand and 30 µL of supernatant was transferred into a new tube. The hybridization samples were measured using D1000 screentape assay kit (Agilent) with 4150 TapesStation system (Agilent).

7. Molecular barcode sequencing and data processing

Molecular barcode sequencing was performed with a Nextseq 550 illumina platform using NextSeq 500/550 High output kit v2.5 with 300PE (Illumina, San Deigo, CA USA). Raw data from Nextseq 550 was converted from bcl file to fastq file using bcl2fastq software (Illumina, Inc. San Diego, CA, USA). Data processing for unique molecular barcoded libraries were followed with Kennedy *et al* (2014). Then, genomic variants in data were discovered, filtered and analyzed with genomic analysis toolkit (GATK)[40]

and annotated using public database such as 1000 Genomes[41], Genome Aggregation Database (gnomAD)[42] and COSMIC[43] for somatic mutation.

8. Pilot study for detection of target genes under cell-free DNA mimicking condition

Nine cell line from the Korean Cell Line Bank (KCLB, Seoul, Republic of Korea) and two genomic DNA (gDNA) of cell line from the American Type Culture Collection (ATCC, Manassas, VA, USA) were selected to establish cancer hotspot mutations containing-circulating tumor DNA (ctDNA) mimicking condition (Table 3). The cell pellet of cell lines from KCLB were used for gDNA extraction using Qiagen's QIAamp DNA Mini and blood Mini kit (cat no. 51104, Hilden, Germany) under following protocol. In brief, cell pellet was resuspended in 200 μL PBS in a 1.5 mL microcentrifuge tube, and 20 μL of proteinase K and 200 μL of Buffer AL was added to the 1.5 mL tube. The tube was vortexed and incubated at 56°C for 10 minutes. 200 μL 100% ethanol was added to the sample and the mixture transferred to a QIAamp Mini spin column. It was centrifuged at 8,000 rpm for 1 minute and changed to a new 2 mL collection tube. Then 500 μL Buffer AW1 was added to the spin column and was centrifuged at 8,000 rpm for 1 minute and changed to a new 2 mL collection tube. Then, 500 μL of Buffer AW2 was added to the column and was centrifuged at 14,000 rpm for 3 minutes. The column was placed in a clean 1.5 mL microcentrifuge tube and 200 μL of distilled water was added to the column and was incubated at room temperature for 1 minute. Then the column was centrifuged for 8,000 rpm for 1 minute. The gDNA in 1.5 mL microcentrifuge tube and gDNA of cell line from ATCC were measured with Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit™ dsDNA BR assay kit (Invitrogen) with Qubit®2.0 fluorometer (Invitrogen) and stored at -20°C for further use. The

1.0 μg of gDNA of cell lines were sonicated using Bioruptor Pico (Diagenode Inc, Denville NJ, USA) with the manufacturer's protocol. The sonication condition for target size 150 bp was 30 cycles of 30 second on and 30 second off per cycle.

9. Pilot study for detection of allelic frequency using cell-free DNA reference materials

To determine the limit of detection of variant allele frequency, cfDNA reference materials with different allelic frequency (AF) were utilized for targeted NGS. Horizon's Multiplex I cfDNA Reference Standard (Horizon Diagnostics, HD780, Cambridge, UK) that consists of Wild type, 0.1% AF, 1% AF, 5% AF, Horizon's Structural Multiplex cfDNA Reference Standard (Horizon) that consists of copy number variation (CNV) and indel variants with 5-20% AF and 4 to 10 copies, and Seraseq ctDNA Mutation Mix v2 (SeraCare, Milford, MA, USA) that has different AF values per kit including WT (cat no. 0710-0144), AF 0.125% (cat no.0710-0143), AF 0.25% (cat no.0710-0142), AF 0.5% (cat no. 0710-0141), AF 1% (cat no. 0710-0140), AF 2% (cat no. 0710-0139) were selected for further use. Horizon's cfDNA products are all derived from human cell line with DNA fragmented to an average size of 160 bp and SeraCare's cfDNA products are derived from GM24385 cell line gDNA with DNA fragmented to an average size of 170 bp. Both cfDNA reference materials mimic the length of plasma cfDNA from cancer patients. The cfDNA reference materials were processed with two different types of library preparation: regular library preparation and unique molecular barcoded library preparation. The equal amount of 50 ng cfDNA reference materials were used for library preparation step that is in *5. Library preparation of cell free DNA* and target enrichment step that is in *6. Hybridization capture-based target enrichment*.

Table 3. Mutation of Cell lines and genomic DNA of cell line used for circulating tumor DNA mimicking condition.

Name	Supplier	Type	Gene	Amino Acid change (Protein change/promoter)
NCI-H1975	ATCC	gDNA	*EGFR*	c.2369C>T (p.Thr790Met); c.2573T>G (p.Leu858Arg)
SW1271	ATCC	gDNA	*NRAS*	c.182A>G (p.Gln61Arg)
HCT-15	KCLB	Cell line	*KRAS*	c.38G>A (p.Gly13Asp)
			PIK3CA	c.1633G>A p.Glu545Lys); c.1645G>A (p.Asp549Asn)
			TP53	c.722C>T (p.Ser241Phe)
HCT 116	KCLB	Cell line	*CTNNB1*	c.133_135delTCT (p.Ser45del)
			KRAS	c.38G>A (p.Gly13Asp)
			PIK3CA	c.3140A>G (p.His1047Arg)
NCI-H1650	KCLB	Cell line	*EGFR*	c.2236_2250del (p.Glu746_Ala750del)
NCI-N87	KCLB	Cell line	*TP53*	c.743G>A (p.Arg248Gln)
SNU-387	KCLB	Cell line	*TERT*	c.228C>T (-124C>T)
			NRAS	c.181C>A (p.Gln61Lys)
SNU-61	KCLB	Cell line	*KRAS*	c.35G>A (p.Gly12Asp)
			TP53	c.524G>A (p.Arg175His)
SW620	KCLB	Cell line	*KRAS*	c.35G>T (p.Gly12Val)
			TP53	c.818G>A (p.Arg273His); c.925C>T (p.Pro309Ser)
T24	KCLB	Cell line	*TERT*	c.228C>T (-124C>T)
T98G	KCLB	Cell line	*PTEN*	c.125T>G (p.Leu42Arg)
			TP53	c.711G>T (p.Met237Ile)

10. Amplicon preparation of TERT promoter region

Due to high GC content of *TERT* promoter region, *TERT* promoter region targeted amplicons were constructed with KAPA 2G Robust HotStart PCR kit (KAPA biosystems). Briefly, 5 μL of 5X KAPA 2G Robust HotStart Buffer A, 5 μL of 5X KAPA Enhancer 1, 1.25 μL of DMSO, 1.5 μL of primers, 0.5 μL of 10 mM KAPA dNTP Mix, 0.2 μL of 5U/μL KAPA2G Robust hotstart DNA polymerase, 10 μL of cfDNA, 1.0 μL of 10 pM primers and 0.55 μL of PCR-grade water were mixed in 0.2 mL PCR tube. PCR was performed in a thermal cycler with a heated lid of 105°C under following cycle condition: 95°C for 3minutes, 35 cycles of 95°C for 15 seconds, 60°C for 15 seconds and 72°C for 30 seconds, and final extension of 72°C for 1 minute. The amplicon samples were measured using D1000 screentape assay kit (Agilent) with 4150 TapesStation system (Agilent).

11. Amplicon sequencing and data processing

Amplicon sequencing was performed with a Miseq illumina platform using Miseq Reagent kit v2 with 300PE (Illumina). The sequencing data was processed with BWA for alignment, SAMtools for sorted data without removing duplication and genomic variants in data was discovered filtered and analyzed with GATK[40] and annotated using public database such as 1000 Genomes[41], Genome Aggregation Database (gnomAD)[42] and COSMIC[43] for somatic mutation.

12. Statistical analysis

Statistical association among the varied data types in this study was evaluated by comparison of pairs of features. Hypothesis testing was performed by testing against null models for absence of association and yielded a p-value. The p values for the association between and among clinical and molecular data types were computed using Kruskal-Wallis test and Wilcoxon test. The p values from the two tests were significant when $p < 0.05$.

III. RESULTS

1. Demographical and clinical-pathological features of hepatocellular carcinoma and liver diseases cohort

A total of 149 Korean non-cancerous liver diseases and HCC patients were enrolled in this study. The non-cancerous liver disease group could be divided into two groups: benign liver disease (BLD) and chronic liver disease (CLD). In the BLD group, there were angiomyolipoma (AML), focal nodular hyperplasia (FNH), hepatocellular adenoma (HA), and hepatic nodule (HN) while in the CLD group, there were autoimmune hepatitis (AIH) and/or liver cirrhosis (LD). Demographically, the age range of the total was from the 20s to 80s and the age of mean with standard deviation was 60.3 ± 11.47. The range of age in HCC, BLD, and CLD was from the 20s to 80s, from the 20s to 70s and from the 40s to 80s, respectively. The HCC group consisted of 19 females (17.6%) and CLD group comprised of 21 females (72.4%) while the BLD group contained only males; and in total, 44 females (29.5%) were involved in this study. While in the BLD and the CLD groups, there was no death reported, in the HCC group, 30 patients (27.8%) were reported as dead and it took up 20.1% of the total (Table 4). Histologically, the total was composed of 108 HCCs, 29 BLD, and 12 CLD cases. Etiologically, 74 cases from an HBV^{+} hepatitis background, 6 were from HCV+ hepatitis background, 9 were from heavy alcohol uptake, 18 were negative for HBV and HCV in serological tests (NBNC) and 2 were from either HBV or HCV infection with heavy alcohol uptake background. The stage of the HCC was determined via two systems: Union for international cancer control (UICC) stage and Barcelona Clinic liver cancer (BCLC) stage. Pathologically, 49 cases were from stage I and II and 59 cases were from stage III and IV according to the UICC stage while 49 cases were from

stage A, 15 were from stage B and 44 were from stage C and no cases from stage D in BCLC stage system. Protein serum markers AFP and PIVKA-II were measured in all HCC patients. The recommended normal AFP and PIVKA-II levels in Severance Hospital, Seoul Korea were less than 9.0 ng/mL and less than 36 mAU/mL, respectively. There were 33 cases (30.6%) of HCC patients and 34 cases (31.5%) of HCC patients reported as in the normal range of AFP and PIVKA-II, respectively (Table 5).

Table 4. Demographical features and survival status of hepatocellular carcinoma and benign and chronic liver diseases patients

Characteristic	Total (N=149)	HCC (N=108)	CLD* (N=29)	BLD** (N=12)
Age at diagnosis				
Mean [SD]	60.3 [11.47]	59.9 [10.30]	65.4 [10.50]	50.3 [17.56]
Gender				
Female (%)	44 (29.5)	19 (17.6)	21 (72.4)	0 (0.0)
Survival status				
Alive (%)	119 (79.9)	78 (72.2)	29 (100)	12 (100)
Dead (%)	30 (20.1)	30 (27.8)	0 (0.0)	0 (0.0)

HCC: Hepatocellular carcinoma; *CLD: Chronic liver diseases including Autoimmune hepatitis (AIH), Liver cirrhosis (LC);
**BLD: Benign Liver disease including Angiomyolipoma (AML), Focal nodular hyperplasia (FNH), Hepatocellular Adenomas (HA); Hepatic nodule (HN)
SD: standard deviation

Table 5. Clinical and pathological characteristics of hepatocellular carcinoma patients (N=108)

Features	Number of HCC patients (%)	Features	Number of HCC patients (%)
UICC (TNM) stage		AFP	
I (T1N0M0)	17 (15.7)	< 9.0 ng/mL	33 (30.6)
II (T2N0M0)	32 (29.6)	> 9.0 ng/mL	75 (69.4)
III (T3 or T4 N0M0)	14 (13.0)	PIVKA-II	
IV-A (AnyT N1M0)	36 (33.3)	< 35 mAU/mL	34 (31.5)
IV-B (any for all)	9 (8.4)	> 35 mAU/mL	74 (68.5)
BCLC stage		Cause	
A (Early)	49 (45.4)	HBV	74 (68.5)
B (Intermediate)	15 (13.9)	HCV	6 (5.6)
C (Advanced)	44 (40.7)	Alcohol	9 (8.7)
D (Terminal)	0 (0.0)	NBNC	18 (16.7)
		HV with Alcohol	2 (0.5)

UICC: Union for international cancer control; TNM: Tumor, Lymph node, Metastasis; BCLC: Barcelona clinical liver cancer; HBV: Hepatitis B virus; HCV: Hepatitis C virus; NBNC: non HBV and non HCV; HV: hepatitis virus AFP: Alpha fetoprotein; PIVKA-II: Protein induced by vitamin K/absence/antagonist-II;

2. Molecular barcode sequencing quality analysis

The total of 149 molecular barcodes embedded libraries were formed and were captured with a HCC-cfDNA Panel probe. The captured libraries were sequenced with Nextseq 550 Illumina platform using NextSeq 500/550 High output kit v2.5 with 300PE (Illumina). The raw files from the sequencing platform were converted to fastq files via bcl2fastq (Illumina) and Bam files were formed via BWA and Samtools. The depth of coverage was calculated using bam files and a bed file that includes the target regions. The target region of the panel was 37,312 bp. An average depth of coverage was 2,309.4x and on-target was 57.11%. The percent covered regions of targeted genes were calculated based on three different depths of coverage of each gene (Table 6).

Table 6. Depth of coverage for each targeted genes in the HCC-cfDNA panel

Gene	Depth of coverage 30x (%, stdev)	Depth of coverage 50x (%, stdev)	Depth of coverage 100x (%, stdev)
ALB	100 (0)	99.4 (0.70)	99.1 (0.78)
APOB	99.5 (0.21)	99.5 (0.21)	99.5 (0.21)
ARID1A	94.8 (3.66)	92.9 (2.96)	90.4 (1.33)
ARID2	100 (0)	100 (0)	100 (0)
AXIN1	99.8 (0.35)	99.7 (0)	99.7 (0)
CTNNB1	100 (0)	100 (0)	100 (0)
KRAS	100 (0)	100 (0)	100 (0)
NFE2L2	100 (0)	100 (0)	100 (0)
PIK3CA	100 (0)	100 (0)	100 (0)
TERT	76.1 (44.25)	57.9 (45.04)	12.5 (35.36)
TP53	100 (0)	100 (0)	100 (0)

ALB: Albumin; *APOB*: Apolipoprotein B; *ARID1A*: AT-rich interaction domain 1A; *ARID2*: AT-rich interaction domain 2; *AXIN1*; Axin1; *CTNNB1*: Catenin beta 1; *KRAS*: Kirsten ras proto-oncogene,GTPase; *PIK3CA*: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; *TERT*: Telomerase reverse transcriptase; *TP53*: Tumor protein p53

3. **Detection of targeted genes under cell free DNA mimicking condition**

The nine cell-lines from KCLB and two genomic DNA of cell lines from ATCC were selected to establish customized cfDNA mimicking fragmented DNA that includes cancer hotspot mutations (Table 3). The custom fragmented DNA was captured with the HCC-cfDNA panel and was expected to detect variants in KRAS, PIK3CA, CTNNB1, TERT, and TP53. The custom fragmented DNA was sequenced and detected all variants except TERT promoter variants with r^2 of 0.92 (Figure 2a).

4. **Limit of detection of molecular barcode sequencing using cell free DNA reference materials.**

To determine the limit of detection of variant allele frequency using molecular barcoded sequencing, cfDNA reference materials with different allele frequency (AF) were utilized for targeted NGS. Horizon's Multiplex I cfDNA Reference Standard (Horizon Diagnostics) that consists of Wild type, AF 0.1%, AF 1.0%, AF 5.0%, and Seraseq ctDNA Mutation Mix v2 (SeraCare) that has different AF values per kit including WT (cat no. 0710-0144), AF 0.125% (cat no.0710-0143), AF 0.25% (cat no.0710-0142), AF 0.5% (cat no. 0710-0141), AF 1.0% (cat no. 0710-0140), AF 2.0% (cat no. 0710-0139) were sequenced and compared the limit of detection of non-molecular barcode NGS and molecular barcode sequencing. In non-molecular barcode NGS, only Horizon's cfDNA reference with AF 5% was detected and r^2 value was 0.832 (Figure 2b). In MBS, Horizon's cfDNA reference standards with AF 5.0%, AF 1.0%, and Seraseq cfDNA reference standards with AF 2.0%, AF 1.0%, AF 0.5% were detected with r^2 of 0.997 (Figure 2c).

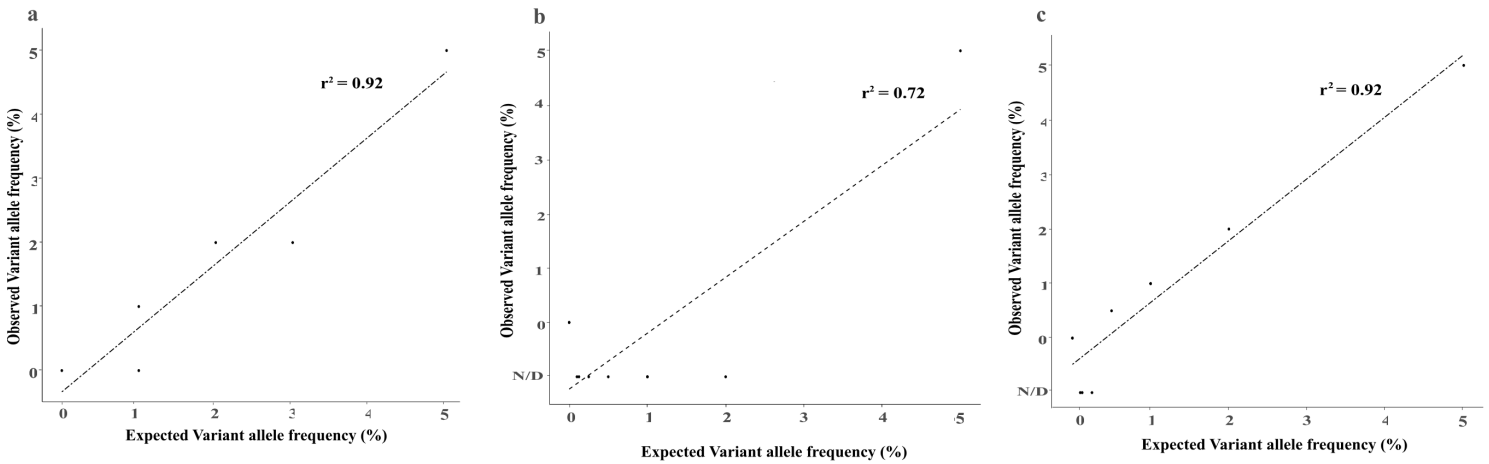

Figure 2. Result of general next-generation sequencing and molecular barcode sequencing using custom fragmented DNA and cfDNA reference materials. (a) Observed variant allele frequency versus expected variant allele frequency of MBS of custom fragmented DNA using cell line gDNA. (b) Observed variant allele frequency versus expected variant allele frequency of general NGS of cfDNA reference materials. (c) Observed variant allele frequency versus expected variant allele frequency of MBS of cfDNA reference materials. N/D means not detected

5. Impact of circulating cell-free DNA concentration on demographical and Cliopathological features in hepatocellular carcinoma and liver disease patients

The concentrations of a total of 149 cfDNA samples were measured and characterized by age groups, types of diseases, and survival status. Two non-parametric analyses including the Kruskal–Wallis test and the Wilcoxon test were utilized to define the significance of the relationship between the concentration of cfDNA and clinical features. The concentration of cfDNA in HCC and liver diseases population were ranged from 0.150 ng/mL to 761.9 ng/mL. The age distribution of the population was ranged from 28 to 83 years old and the age groups were divided into seven groups. Kruskal-Wallis test and Wilcoxon test were utilized to determine if there are statistically significant differences among the groups based on cfDNA concentration. The p-values of age groups and gender were 0.46 and 0.21, respectively (Figure 3a and Figure 3b). Two more Kruskal-Wallis tests were processed for comparison of malignancy of diseases and BCLC stages based on the concentration of cfDNA. For the malignancy of diseases, the total population was analyzed while for the stages of the BCLC system, 108 of HCC cfDNA samples were selected. For the malignancy of diseases and BCLC stages, the p-values were 0.38 and 0.27, respectively (Figure 3c and Figure 3d).

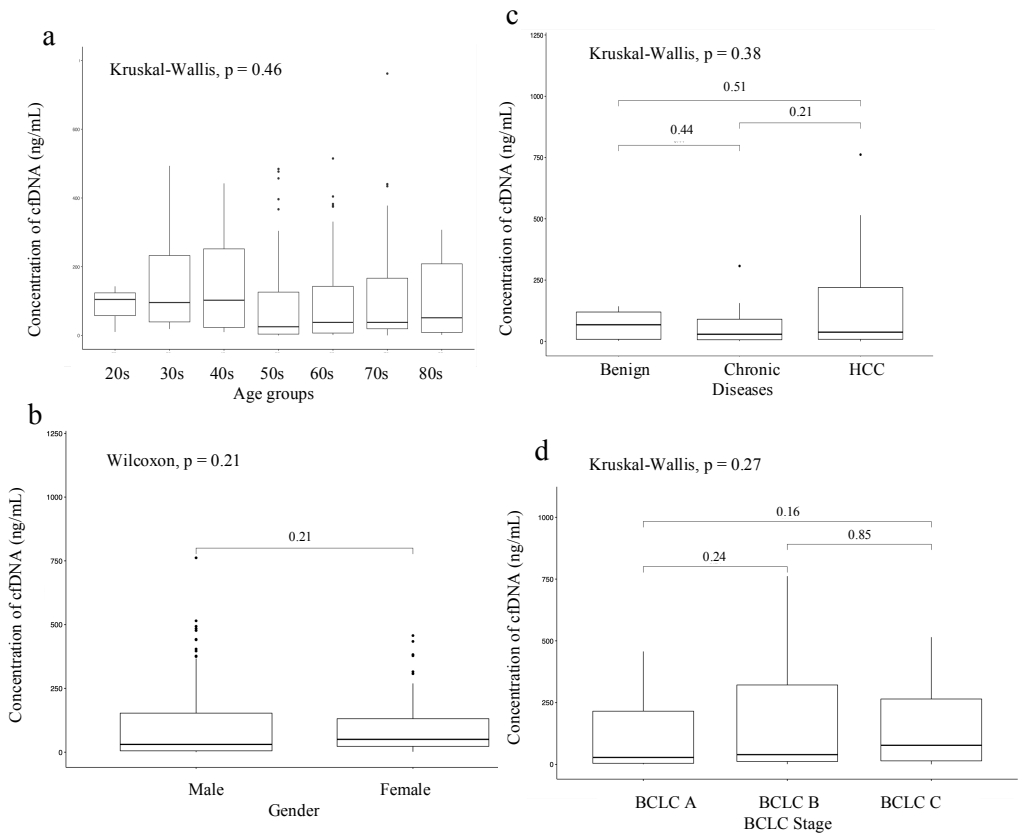

Figure 3. Clinical features of cfDNA in hepatocellular carcinoma and liver diseases. cfDNA concentration were compared with (a) age groups (p = 0.46). (b) gender (p = 0.21). (c) malignancy of liver diseases. (d) BCLC stages of HCC. For age groups, malignancy of liver diseases, BCLC stages of HCC comparison with cfDNA concentration, p values were calculated using Kruskal-Wallis test while for gender comparison with cfDNA concentration, p value was calculated using Wilcoxon test.

6. Genetic landscape of hepatocellular carcinoma and liver diseases

The MBS of a total of 149 cfDNAs from Korean non-cancerous liver diseases and HCC plasma samples were performed. The number of enrolled patients of the non-cancerous benign liver diseases chronic liver diseases were 12 and 29, respectively while that of enrolled HCC patients were 108. The variants detected from the sequencing were divided into two groups according to the presence of the variants in public databases or not. The public databases utilized were gnomAD, 1000 genomes, and COSMIC. In total variant population frequency, novel singleton variants of HCC were accounted for 33.6% and known singleton variants of HCC were 19.2% while both novel and known singleton variants of liver diseases were 5.6% (Figure 4). Then the number of variants according to types of genetic alterations was counted. In total variants frequency of HCC and liver diseases, missense was accounted for 42.53%, synonymous 18.18%, frameshift 16.24%, inframe indel 8.90%, 3′UTR 6.60%, splicing 2.70% and 5′UTR 2.13%. The missense of HCC and that of liver diseases were accounted for 29.7% and 13.2%, respectively (Figure 5). The overall MBS ratio of transition and transversion (Ti/Tv) was 2.07 and synonymous to missense ratio in the liver diseases and the HCC were 1.92 and 2.33, respectively. The demographical, clinical, and pathological features of the patients were malignancy of liver disease and BCLC stage of HCC, causes of the disease, and gender. In total eleven targeted genes, the proportion of variants detected in each gene was calculated. The frequencies of TP53, TERT, CTNNB1, AXIN1, ARID1A, ARID2, NFE2L2, KRAS, ALB and APOB were 34.0%, 34.0%, 15.0%, 10.0%, 10.0%, 7.0%, 1.0%, and 21.0% respectively and no PIK3CA variants were detected (Figure 6). The number of variants observed via MBS was compared with demographical and cliopathological features. Two non-parametric analyses including

the Kruskal–Wallis test and the Wilcoxon test were utilized to define the significance of the relationship between the number of variants and clinical features. The age distribution of the population was ranged from 28 to 83 years old and the age groups were divided into seven groups. Kruskal-Wallis test and Wilcoxon test were utilized to determine if there are statistically significant differences among the groups based on the number of variants. The p-values of age groups and gender were 0.57 and 0.99, respectively (Figure 7a and Figure 7b). Two more Kruskal-Wallis tests were processed for comparison of malignancy of diseases and BCLC stages of HCC based on the number of variants. For the malignancy of diseases, the total population was analyzed while for the stages of the BCLC system, 108 of HCC cfDNA samples were selected. For the malignancy of diseases and BCLC stages, the p-values were 0.80 and 0.00052, respectively (Figure 7c and Figure 7d).

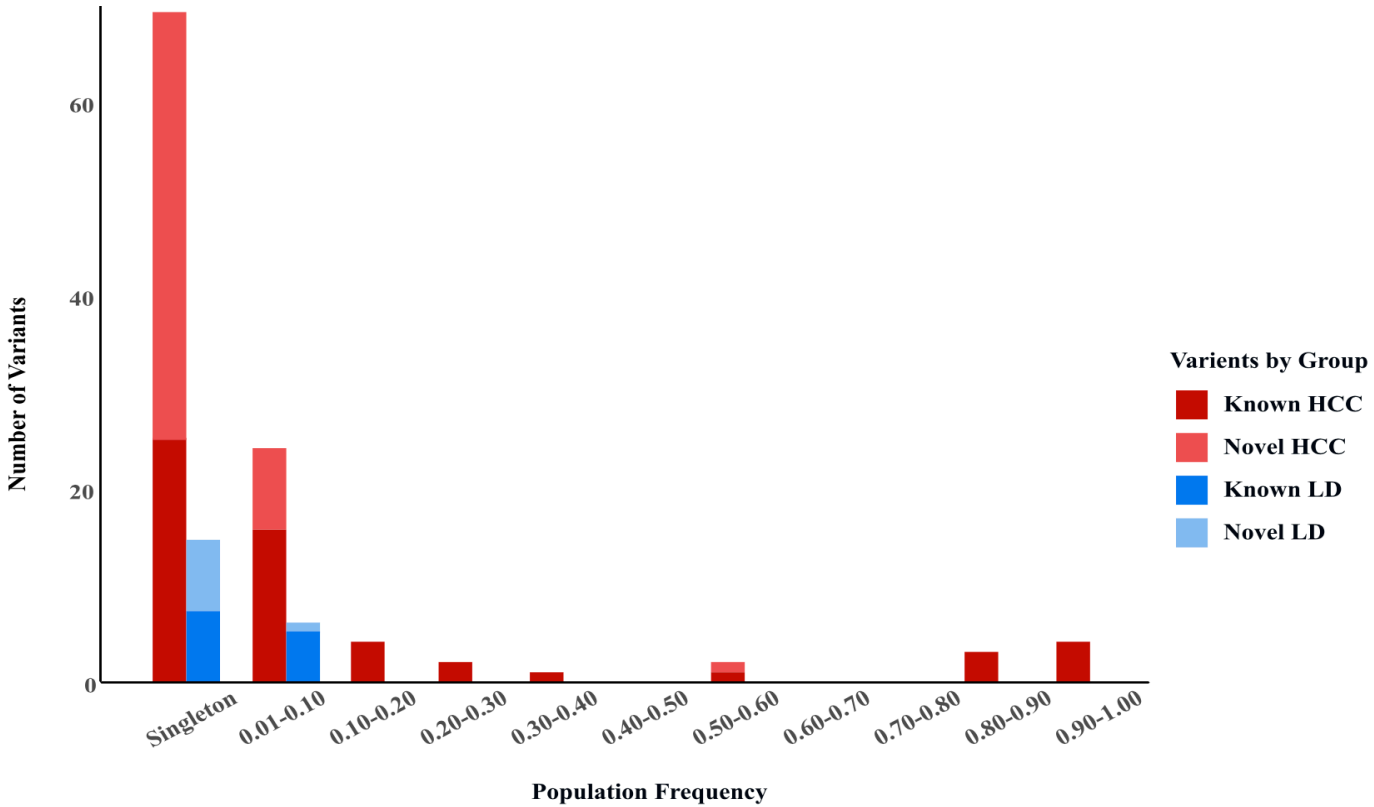

Figure 4. Population Frequency of hepatocellular carcinoma and liver disease patients. Known HCC and Known LD groups are variants that are registered in public databases while Novel HCC and Novel LD groups are variants that are not registered in public databases.

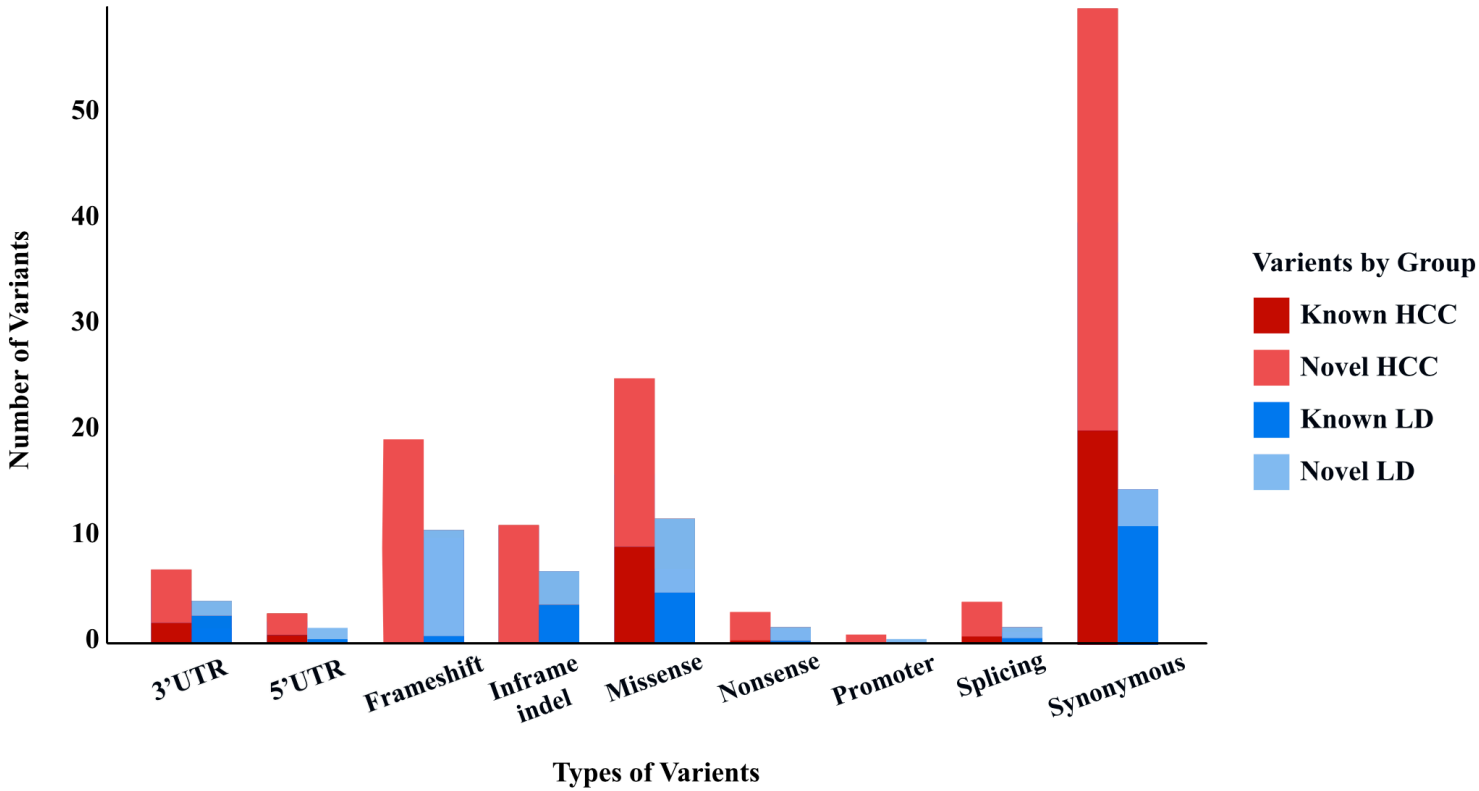

Figure 5. Variants Frequency in hepatocellular carcinoma and liver diseases. Different types of variants were detected using MBS. Known HCC and Known LD groups are variants that are registered in public databases. Novel HCC and Novel LD groups are variants that are novel and not registered in public databases.

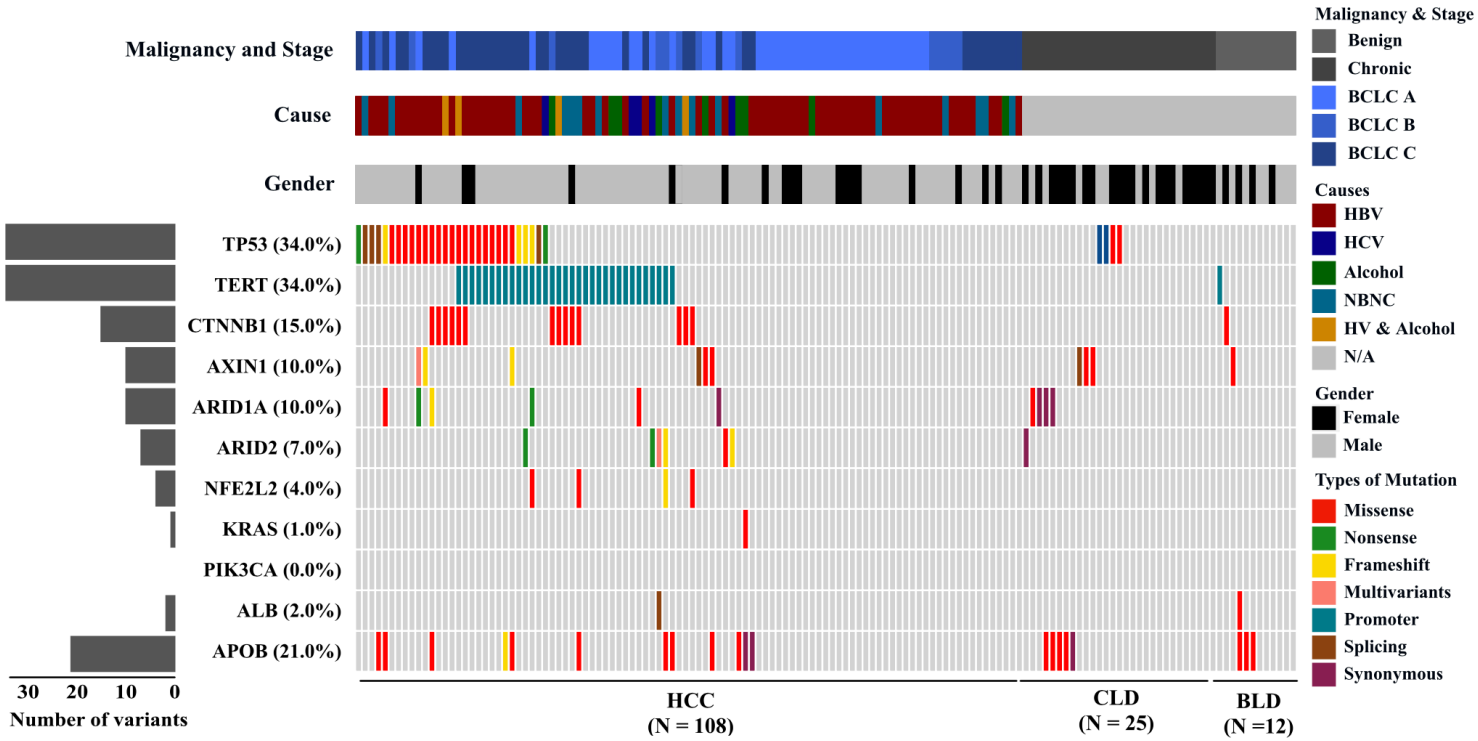

Figure 6. Genetic landscape of Hepatocellular carcinoma and benign and chronic liver diseases. All variants were displayed according to genes, malignancy of liver diseases and stage of HCC, causes of liver diseases and HCC and gender. The number of variants in each genes were showed in bar graph and percentages of each genes in the study were presented next to name of the genes.

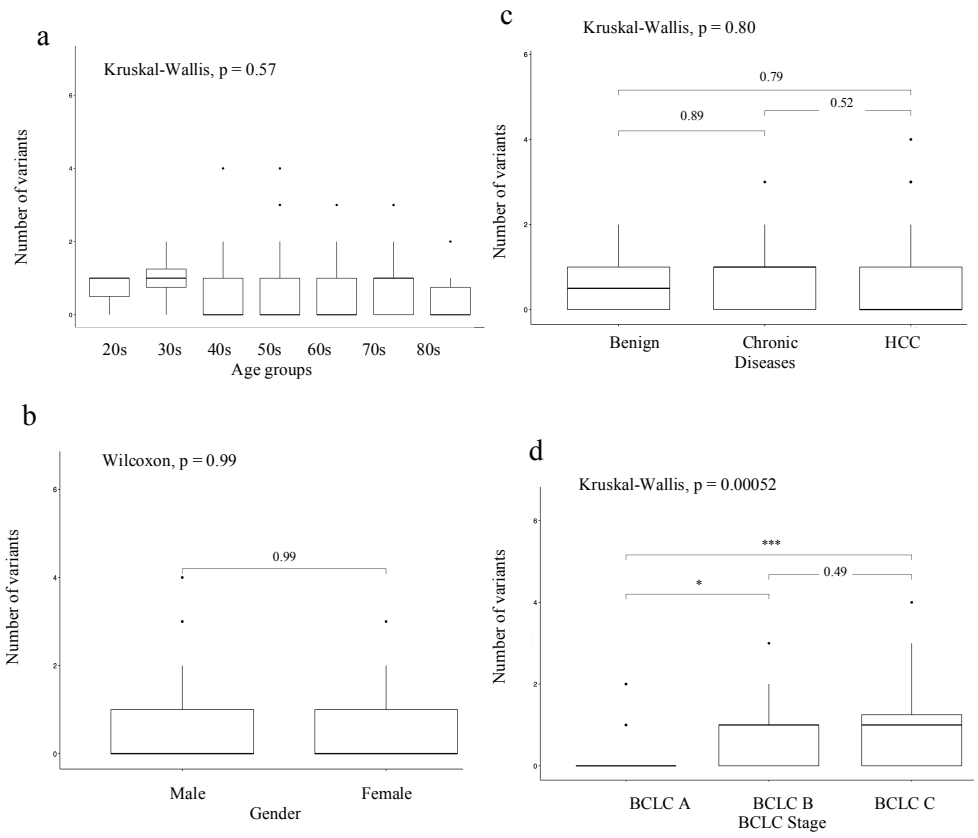

Figure 7. Comparison of the number of variants detected using MBS based on demographical and cliopathological features. The number of variants detected compared with (a) age groups (p = 0.46). (b) gender (p = 0.21). (c) malignancy of liver diseases. (d) BCLC stages of HCC. For age groups, malignancy of liver diseases, BCLC stages of HCC comparison with number of variants, p values were calculated using Kruskal-Wallis test while for gender comparison with number of variants, p value was calculated using Wilcoxon test.

7. Mutational landscape of cfDNA in hepatocellular carcinoma and liver diseases

Among the one-hundred eight HCC patients cohort, twenty-seven tissue biopsy samples were collected. NGS targeted sequencing was performed after tissue genomic DNA samples were extracted. The result of tissue samples were then compared with that of MBS of matched cfDNA samples. The tissue samples consists of genetic alterations in TP53, TERT, CTNNB1, AXIN, ARID1A, ARID2, ALB and APOB. All genetic variants of the targeted genes except TERT were detected in both tissue gDNA and plasma cfDNA. Among the TERT promoter variants occurred in tissue gDNA, promoter variants of tissue gDNA with VAF above 25.0% were also detected in matched plasma cfDNA while the promoter variants with VAF less than 25.0% were not able to detect the variants in matched plasma cfDNA (Figure 8).

Case#	Gene	Coding	Mutation	Tissue	Plasma
#1 (HCC)	TP53	c.817C>T	p.Arg273Cys	36.48%	43.70%
	CTNNB1	c.94G>C	p.Asp32His	25.56%	4.70%
	APOB	c.5834A>T	p.Asp1945Val	48.89%	55.20%
#2 (HCC)	TERT	-124C>T	C228T	30.00%	3.00%
	CTNNB1	c.101G>A	p.Gly34Glu	31.50%	12.00%
	ARID1A	c.3996_4001dupGCAGCA	p.Gln1333_Gln1334dup	17.85%	53.80%
#3 (HCC)	ARID1A	c.880A>C	p.Thr294Pro	11.00%	2.00%
	ARID1A	c.863A>C	p.Gln288Pro	28.29%	25.00%
	APOB	c.12175G>A	p.Glu4059Lys	27.15%	25.40%
	APOB	c.12175G>A	p.Glu4059Lys	24.38%	14.80%
#4 (HCC)	TP53	c.742C>T	p.Arg248Trp	42.46%	4.60%
	TERT	-124C>T	C228T	22.00%	N/D
#5 (HCC)	TP53	c.725G>A	p.Cys242Tyr	43.87%	7.80%
	TP53	c.721T>G	p.Ser241Ala	43.34%	8.20%
	CTNNB1	c.95A>T	p.Asp32Val	27.06%	3.30%
#6 (HCC)	ARID1A	c.880A>C	p.Thr294Pro	17.22%	0.80%
	ARID1A	c.863A>C	p.Gln288Pro	7.46%	0.30%
	APOB	c.12175G>A	p.Glu4059Lys	43.53%	49.60%
#7 (HCC)	TERT	-124C>T	C228T	16.00%	N/D
	AXIN1	c.1953C>T	p.His651His	42.70%	44.80%
	APOB	c.7223C>T	p.Ser2408Phe	42.79%	20.00%
#8 (HCC)	TERT	-124C>T	C228T	5.00%	N/D
	AXIN1	c.2405delG	p.Gly803Alafs Ter13	12.98%	3.90%
	APOB	c.8414A>G	p.Asn2805Ser	44.87%	50.40%
	APOB	c.6194A>G	p.Asp2065Gly	50.33%	43.80%
#9 (HCC)	TERT	-124C>T	C228T	36.00%	3.00%
	AXIN1	c.1911T>C	p.Ile637Ile	45.72%	71.40%
#10 (HCC)	TERT	-124C>T	C228T	29.00%	10.00%
	CTNNB1	c.94G>T	p.Asp32Tyr	20.98%	46.60%

Figure 8. Mutational landscape of cfDNA with matched tissue genomic DNA in hepatocellular carcinoma. Genetic alterations in both cfDNA and matched tissue gDNA were displayed with variant allele frequency. The numbers presented in the boxes are VAF of genetic alterations per samples. N/D stands for not detected.

8. Survival analysis and hazard ratio of hepatocellular carcinoma

The research recorded the length of time from study entry to an endpoint of the study to perform time-to event analysis. Total eleven targeted genes were analyzed under Kaplan-Meier's survival analysis and *TP53* showed differences in survival curves between *TP53* negative position group and *TP53* positive mutations group (Figure 9). Time-to-event curves were further analyzed using Cox proportional hazards model. The hazard function can be interpreted as the risk of dying at time. Hazard ratio of 1.0 represents there were no effect, hazard ratio less than 1.0 indicates reduction in the hazard ratio greater than 1.0 suggests increase in hazard. The three most mutated genes were performed Kaplan-Meier survival analysis and Cox proportional hazard model. According to the statistical analysis, the hazard ratio of *TP53, TERT* and *CTNNB1* were 1.80, 1.30 and 1.10. Furthermore, the three genes were analyzed with pairing each other gene. Variants present in one or the other gene in *TP53* and *TERT* had hazard ratio of 1.70, in *TP53* and *CTNNB1* had hazard ratio of 1.0 and in *TERT* and *CTNNB1* had hazard ratio of 1.2. Variants present in both *TP53* and *TERT* had hazard ratio of 1.50, that in both *TERT* and *CTNNB1* had hazard ratio of 1.50, and that in both *TP53* and *CTNBB1* had hazard ratio of 3.00 (Figure 10).

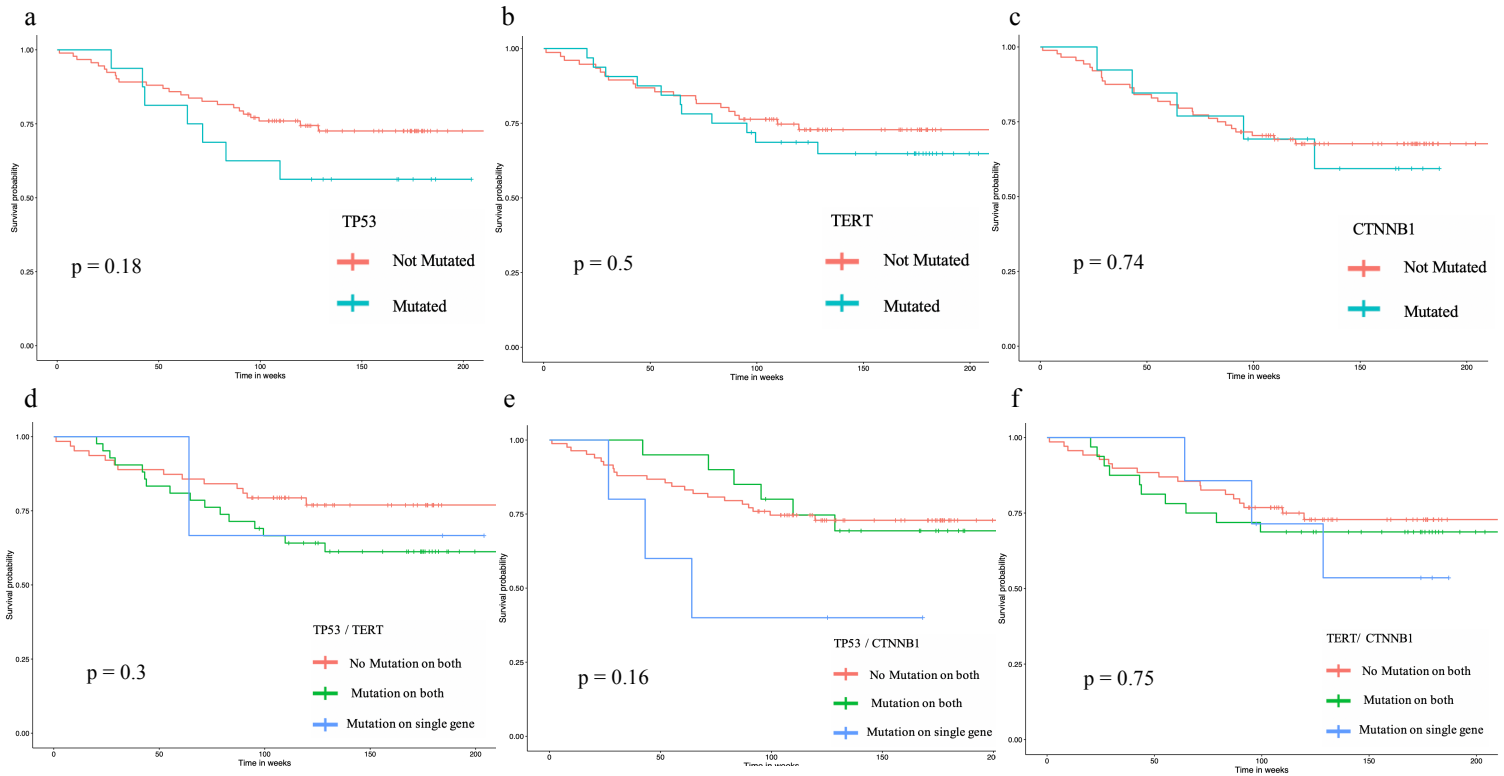

Figure 9. Survival analysis of three most frequent genes detected in hepatocellular carcinoma. The graphs are survival analysis of (a) TP53 mutated. (b) TERT mutated (c) CTNNB1 mutated (d) TP53 and/or TERT mutated (e) TP53 and/or CTNNB1 mutated (f) TERT and/or CTNNB1 mutated for HCC.

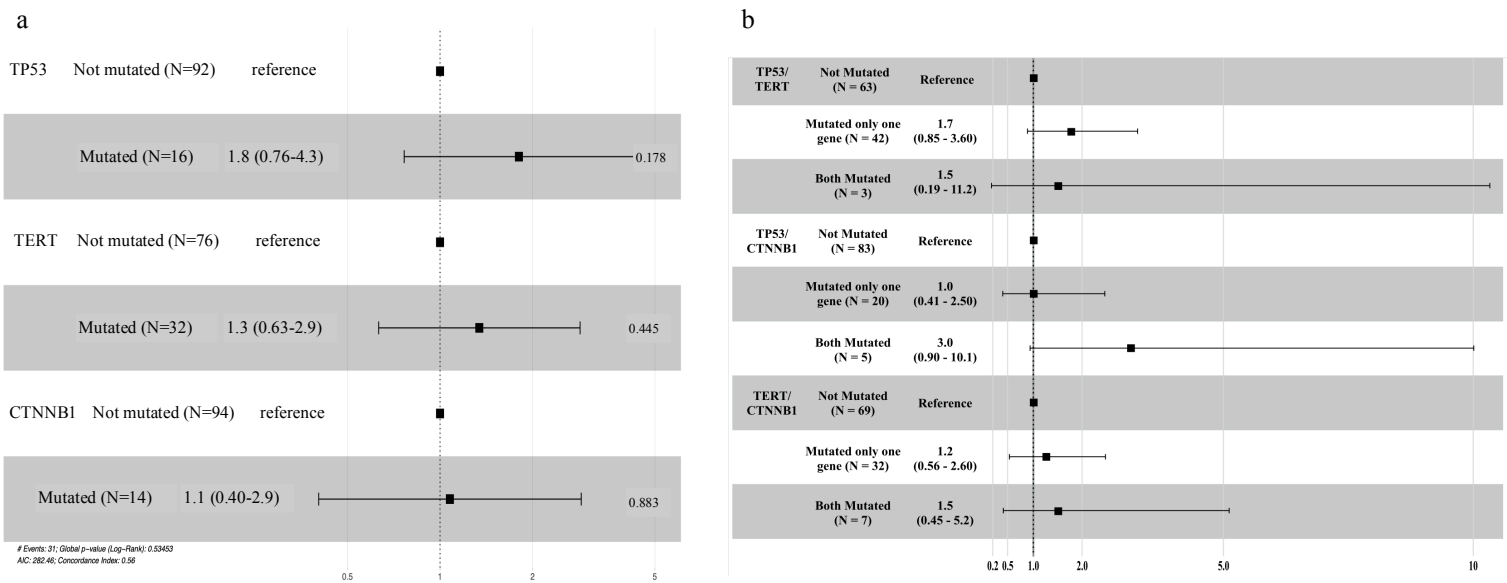

Figure 10. Hazard ratio of genetic alterations in hepatocellular carcinoma. (a) single gene variant hazard ratio. Hazard ratio of *TP53* mutation is 1.8, that of *TERT* mutation is 1.3 and that of *CTNNB1* mutation is 1.1 (b) two gene variants hazard ratio is depends on whether one or other gene was mutated or both gene variants were occurred in a same sample. Hazard ratio of single variant occurred either of one gene for *TP53/TERT* pair was 1.7 while both mutated was 1.5. that of single variant and both variants for *TP53/CTNNB1* were 1.0 and 3.0, respectively. Hazard ratio of single variant and both variants for *TERT/CTNNB1* were 1.2 and 1.5 respectively.

IV.DISCUSSION

The incidence rate of liver cancer is known as three times higher among men than in women, and the population aged fifty to sixty years old has the highest risk of liver cancer worldwide and in the Republic of Korea[1-2]. In this study, a total of one-hundred forty-nine patients were enrolled and the total cohort consisted of one-hundred eight HCC patients. A ratio of males to females in the HCC cohort was 4 to 1 and the mean age of HCC cohort was fifty-nine years old, which indicates that the enrolled HCC cohort was similarly represented as a general population of liver cancer in worldwide and Republic of Korea (Table 4). The cancer stage of the HCC cohort was distributed in balance as the frequency of patients with stage I to stage II was 45.4% and that of patients with stage III to stage IV was 54.6%. AFP and PIVKA-II protein markers from blood were widely used as one of HCC surveillance tests and their sensitivities are known as 25-90% and 44-91%, respectively[3,5-7]. In this HCC cohort, serum protein markers AFP and PIVKA-II levels were collected and the sensitivities of AFP and PIVKA-II were 69.4% and 68.5%, respectively (Table 5). This result indicates that about one-third of HCC patients with a normal range of protein markers were failed to surveil and distinguish as HCC patients and surveillance and diagnosis of HCC with serum protein markers levels suggest low detection rate of HCC.

Liquid biopsy is currently widely applied in a research basis for study and discovery of driver mutations of human diseases and clinical basis for diagnosis, prognosis, and monitoring of cancer and rare diseases as an alternative procedure for tissue biopsy. At the beginning of the liquid biopsy era, cfDNA concentration was a considerable diagnostic indicator of malignancy of tumors and monitoring for the progression of cancer and the stage of tumors[44]. The concentration of cfDNA is, however, not presently correlated with the recurrence of cancer or malignancy of tumors[44] while it is associated with the stage of tumors[45]. In this study, the concentration of cfDNA was increased as the stage of HCC was higher but it was not statistically significant ($p \geq 0.05$, Figure 3).

Next-generation sequencing has been advanced the analysis of genetic and structural alterations of human diseases. Currently, the cost of sequencing per sample has been lower and sample compatibility of NGS has been increased and the application of NGS in diverse fields is increasing. The NGS of cfDNA, however, requires higher cfDNA input for the detection of variant allele frequency less than 5%. By using a molecular barcode, a short random sequence embedded in an adapter, a higher correlation between observed and expected VAF in fragmented custom cell line gDNA and cfDNA reference materials as R square values of general NGS and MBS were 0.72 and 0.92, respectively (Figure 2). In this study, HCC patients and liver disease patients were compared depends on population frequency and types variants observed. In the population frequency of the total cohort, the highest detected variants were singleton variants. Among singleton variants in HCC cohort, frequency of novel variants that are not registered in public databases was higher than that of known variants that are registered in public databases such as 1000 Genomes, gnomAD, and COSMIC while among singleton variants in liver diseases cohort, novel and known variants frequency were equivalent (Figure 4). It implies that the genetic alterations are closely related to the development of cancer. Total nine types of genetic variants were observed and frequency of missense mutation was the second-highest variant type followed by that of synonymous. The loss of function mutations in HCC cohort such as frameshift, inframe indel, missense and nonsense, promoter variants consist of a higher rate of novel variants than that of known variants while the loss of function mutation in liver disease cohort consists of a higher rate of known variants than that of novel variants. (Figure 5).

Furthermore, the NGS including MBS has difficulties in the analysis of some regions of the genome as it is affected by sequence characteristics of the genome such as GC-poor or GC-rich region[46]. To discover genetic alterations and structural variation via the NGS, sufficient depth of coverage is required. The NGS, nonetheless, yields a low depth of coverage in both the GC-poor region with

a mean GC content of less than 25% and the GC-rich region with a mean GC content of higher than 60%[46-47]. One of GC-rich region is widely known as *TERT* promoter region and many studies including Pan-cancer analysis of whole genomes (PCAWG) study have had difficulties to obtain high depth of coverages on *TERT* promoter region using NGS and tried alternative procedures such as sanger sequencing, droplet digital PCR (ddPCR) and amplicon sequencing[48-49]. In this study, the MBS was performed with the HCC-cfDNA panel for target enrichment. Among the targeted genes, the *TERT* promoter region had the GC content of 91.30% (Table 2). The result of MBS showed that the *TERT* promoter region had a low depth of coverage contrasting to other target genes (Table 6). Thus, the amplicon sequencing of the *TERT* promoter region was required and performed, and was able to obtain sufficient depth of coverage to discover two promoter region variants c.-146 C>T (C250T) and c.-124C>T (C228T).

In hepatocellular carcinoma, most commonly mutated genes are known as *TERT* promoter, *TP53, CTNNB1, ARID1A, AXIN1,* and *ARID2* and the mutation rate were 60%, 30%, 30%, 10%, 10%, and 5%, respectively[50]. In this study, the most commonly occurred genes were equivalent to most commonly occurred and known genes in the HCC but the mutation rates of the *TERT* promoter (34.0%) and *CTNNB1* (15.0%) were different than previously reported. The mutation rate of *TERT* promoter variants in the HCC greater than 50% is observed in Europe and America and the occurrence of *TERT* promoter region variants in the HCC Asian population is known as 20% to 40%[51]. Thus, the mutation rate of the *TERT* promoter region in this cohort was 34.0%, which is matched with Asian studies as the ethnicity of the total cohort in this study is the Korean. Furthermore, the *CTNNB1* mutation rate is lower than previously reported as most causes of HCC in the Korean population are HBV not HCV and in this cohort, HCV-infected HCC patients account for 5.6% of the total HCC cohort[2]. Currently, proteins that make up the HCV virus are believed that they independently stimulate a Wnt/β-catenin signaling pathway, which significantly increases the frequency of *CTNNB1*

mutations compared with the HBV-infected and the non-viral HCC tumors[52]. The number of variants observed in the HCC subsequently compared with demographical and clio-pathological characteristics. Although the number of variants had no statistically significant relationship with age, gender, and malignancy of liver diseases, it was statistically significant with the cancer stage of HCC with a p-value of 0.00052 (Figure 7) and it denotes that a higher cancer stage consists of greater the number of genetic alterations. The genetic mutations from the tissue gDNA and the plasma ctDNA were compared using the MBS and the amplicon sequencing of the *TERT* promoter region. All genetic alterations of the targeted genes in the tissue gDNA were equivalent to that in the plasma ctDNA and it suggests that the plasma cfDNA can resemble the characteristic of the tissue gDNA(Figure 8). The most frequently occurred genes in the HCC were selected to uncover the association between the genetic alterations and the patient survival status. The Kaplan-Meier survival curve and the Cox proportional hazard ratio were utilized for survival analysis. According to the Kaplan-Meier survival curve, the association between the presence of genetic alterations and survival was not statistically significant($p \geq 0.05$, Figure 9). The Cox proportional hazard ratio defines the relative risk of the complication based on the comparison of event rates[53]. A hazard ratio of 1.0 represents there were no effect, a hazard ratio less than 1.0 indicates reduction of risk, a hazard ratio greater than 1.0 suggests increase in risk[53]. The patients with the *TP53* mutations had a hazard ratio of 1.80, which is the highest rate among that of other gene alterations and it implies that *TP53* mutations cause a greater risk in survival of the patients. The patients with both *TP53* and *CTNNB1* mutations had a hazard ratio of 3.00 that is the highest rate among that of other gene mutation pairs. The hazard ratio indicates that *TP53* and *CTNNB1* mutations and a greater number of genetic alterations in driver genes are critical to the survival of patients. The overall hazard model shows that the occurrence of genetic alterations in driver genes increases the risk of survival of HCC patients (Figure 10).

V. CONCLUSION

With the advances in NGS technology, comprehensive exploration of somatic alterations within cfDNA has been increasingly accessible in various malignancies including hepatocellular carcinoma. Targeted sequencing detects variants with limit of detection as low as five percent of allele frequency under depth of coverage of five-hundred. To detect lower variant allelic frequency of variants from circulating tumor DNA however, several factors including increasing amount of genetic materials and greater depth of coverage should fulfill to uncover low variant allelic frequency variants and rare variants. In this study, molecular barcode sequencing was utilized as an alternative method for detection of low frequency and rare variants. Comparison of performance of both general NGS and MBS of cfDNA reference materials suggests that allelic frequency as low as 0.3-0.5% could be detected by MBS in plasma cfDNA of hepatocellular carcinoma and liver diseases. Furthermore, matched genetic alterations in tissue gDNA and plasma cfDNA indicates that cfDNA could represent tissue and its molecular characteristics and could be an substitute method of NGS of tissue biopsy and enable the investigation of low variant allele frequency mutations in plasma from malignant diseases.

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 Cancer J Clin 2018;68:394-424.
2. Korean Liver Cancer A, National Cancer C. 2018 Korean Liver Cancer Association-National Cancer Center Korea Practice Guidelines for the Management of Hepatocellular Carcinoma. Gut Liver 2019;13:227-99.
3. 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;16:589-604.
4. Abdelaziz AO, Elbaz TM, Shousha HI, Ibrahim MM, Rahman El-Shazli MA, Abdelmaksoud AH, et al. Survival and prognostic factors for hepatocellular carcinoma: an Egyptian multidisciplinary clinic experience. Asian Pac J Cancer Prev 2014;15:3915-20.
5. Villanueva A. Hepatocellular Carcinoma. N Engl J Med 2019;380:1450-62.
6. Van Hees S, Michielsen P, Vanwolleghem T. Circulating predictive and diagnostic biomarkers for hepatitis B virus-associated hepatocellular carcinoma. World J Gastroenterol 2016;22:8271-82.
7. Park SJ, Jang JY, Jeong SW, Cho YK, Lee SH, Kim SG, et al. Usefulness of AFP, AFP-L3, and PIVKA-II, and their combinations in diagnosing hepatocellular carcinoma. Medicine (Baltimore) 2017;96:e5811.
8. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol 2013;10:472-84.
9. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer 2011;11:426-37.
10. Ilie M, Hofman V, Long E, Bordone O, Selva E, Washetine K, et al. Current challenges for detection of circulating tumor cells and cell-free circulating

nucleic acids, and their characterization in non-small cell lung carcinoma patients. What is the best blood substrate for personalized medicine? Ann Transl Med 2014;2:107.

11. Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. Clin Chem 2015;61:112-23.
12. Palmirotta R, Lovero D, Cafforio P, Felici C, Mannavola F, Pelle E, et al. Liquid biopsy of cancer: a multimodal diagnostic tool in clinical oncology. Ther Adv Med Oncol 2018;10:1-24.
13. W.Bennett C, Berchem G, Kim YJ, El-Khoury V. Cell-free DNA and next-generation sequencing in the service of personalized medicine for lung cancer. Oncotarget 2016;7:71013-35.
14. Liao W, Yang H, Xu H, Wang Y, Ge P, Ren J, et al. Noninvasive detection of tumor-associated mutations from circulating cell-free DNA in hepatocellular carcinoma patients by targeted deep sequencing. Oncotarget 2016;7:40481-90.
15. Mandel, P. & Métais, P. Les acides nucléiques du plasma sanguin chez l′homme. C. R. Acad. Sci. Paris 1948;142:214-243
16. Stroun M, Maurice P, Vasioukhin V, Lyautey J, Lederrey C, Lefort F, et al. Ann N Y Acad Sci 2000; 906:161-8
17. Choi JJ, Reich CF 3rd, Pisetsky DS. The role of macrophages in the in vitro generation of extracellular DNA from apoptotic and necrotic cells. Immunology 2005;115:55-62
18. Mouliere F, Robert B, ArnauPeyrotte E, Del Rio M, Ychou M, Molina F, et al. High fragmentation characterizes tumor-detrived circulating DNA. PloS 2011;6:e23418
19. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer -A survey. Biochim Biophys Acta 2007; 1775:181-232

20. Cheng YW, Stefaniuk C, Jakubowski MA. Real-time PCR and targeted next-generation sequencing in the detection of low level EGFR mutations: Instructive case analyses. Respir Med Case Rep 2019;28:100901.
21. Meldrum C, Doyl MA, Tothil RW. Next-generation sequencing for cancer diagnostics: apractical perspective. Clin Biochem Rev 2011; 32:177-195
22. Petrackova A, Vasinek M, Sedlarikova L, Dyskova T, Schneiderova P, Novosad T, et al. Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics. Front Oncol 2019;9:851
23. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A 2012;109:14508-13.
24. Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. Nat Protoc 2014;9:2586-606.
25. Hirotsu Y, Otake S, Ohyama H, Amemiya K, Higuchi R, Oyama T, et al. Dual-molecular barcode sequencing detects rare variants in tumor and cell free DNA in plasma. Sci Rep 2020;10:3391.
26. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet 2012;44:750-764.
27. 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. Nat Genet 2012; 44:694-698.
28. Huang J, Deng Q, Wang Q, Li KY, Dai JH, Li N, et al. Exom sequencing of hepatitis B virus-associated hepatocellular carcinoma. Nat Genet 2012; 44:1117-21.

29. Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. Nat Genet 2012;44:765-769.
30. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. Genome Res 2013; 23:1422-33.
31. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma. Nat Genet 2014;46:1267-73.
32. Ahn SM, Jang SJ, Shin JH, Kim DH, Hong SM, Sung CO, et al. Genomic portrait of resectable hepatocellular carcinomas: Implications of RB1 and FGF19 aberrations for patient stratification. Hepatology 2014;60:1972-82.
33. Schulze K, Imbeaud S, Letouze E, Alexandrov L. B, Calderaro J, Rebouissous S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 2015;47:505-511.
34. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat Genet 2016;48:500-9.
35. Schulze K, Nault JC, Villanueva A. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. J Hepatol 2016;65:1031-42.
36. Cancer Genome Atlas Research Network. Electronic address wbe, Cancer Genome Atlas Research N. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell 2017;169:1327-41.
37. Castelli G, Pelosi E, Testa U. Liver Cancer: Molecular Characterization, Clonal Evolution and Cancer Stem Cells. Cancers Basel 2017;9.
38. Nojiri S and Joh T, Albumin suppresses human hepatocellular carcinoma proliferation and the cell cycle. Int J Mol Sci 2014;15:5163-74.

39. Lee G, Jeong YS, Kim DW, Kwak MJ, Koh J, Joo EK, et al. Clinical significance of APOB inactivation in hepatocellular carcinoma. Exp Mol Med 2018; 50:1-12.
40. PoplinR, Ruano-Rubio Va, DePristo M, Fennell T, Carneiro M, Vader Auwera G, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv 2017.
41. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015; 526:68-74.
42. Lek M, Kaczewski K, Minikel E, Samocha K, Banks E, Fennell T, et al. Analysis of protein0coding genetic variation in 60,706 humans. Nature 2016; 536:285-291
43. Tate J, Bamford S, Jubb H, Sondka Z, Beare D, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nuclic Acids Res 2019. 47: 941-947
44. Mirtavoos-Mahyari H, Ghafouri-Fard S, Khosravi A, Motevaseli E, Esfahani-Monfared Z, Seifi S et al. Circulating free DNA concentration as a marker of disease recurrence and metastatic potential in lung cancer. Clin Trans Med 2019; 8:14
45. Chicard M, Boyault S, Daage L, Richer W, Gentien D, Pierron G et al. Genomic copy number profiling using circulating free tumor DNA highlights heterogeneity in neuroblastoma. Clin Cancer Res 2016; 22:5564-5573
46. Chen Y, Liu T, Yu C, Chiang T, Hwang C. Effects of GC bias in next-generation sequencing data on de novo genome assembly. PLosOne 2013; 8:1-20
47. Tilak M, Botero-Castro F, Galtier N, Nabholz B. Illumina Library Preparation for sequencing the GC-rich of heterogeneous genomic DNA. Genome Bio Evol 2018; 10:616-622

48. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature 2020; 579:82-93
49. Ward D, Baxter L, Gordon N, Ott S, Savage R, Beggs A et al. Multiplex PCR and next generation sequencing for the non-invasive detection of bladder cancer. PLos One 2016; 11:e0149756
50. Villanueva A. Hepatocellular carcinoma. N Engl J Med 2019; 380:1450-1462
51. Pezzuto F, Buonaguro L, Buonaguro F, Tornesello M. Frequency and geographic distribution of TERT promoter mutation in primary hepatocellular carcinoma. Infect Agent Cancer 2017; 12:27
52. Khalaf A, Fuentes D, Morshid A, Burke M, Kaseb A, Hassan M et al. Role of Wnt/β-catenin signaling in hepatocellular carcinoma, pathogenesis, and clinical significance. J Hepatocell Carcinoma 2018; 5:61-73
53. Spruance S, Reid J, Grace M, Samore M. Hazard ratio in clinical trials. Antimicrob Agents Chemother 2004: 48:2787-2792

ABSTRACT (IN KOREAN)

분자 바코드를 이용한 차세대 염기서열 분석법을 통하여 간세포암종의 세포 유리 DNA의 낮은 대립유전자 빈도 검출 및 임상적 응용

<지도교수 최 종 락 >

연세대학교 대학원 의과학과
김 이 슬

간암은 한국인에게 높은 발병률을 나타내는 암 종으로서, 발병률과 사망률이 꾸준히 감소해오고 있음에도 불구하고, 여전히 여섯 번째로 많이 검진되는 암이며, 암으로 인한 사망률 중 네번째로 치명적이다. 특히 간세포암종의 경우, 50 대와 60 대 남성에서 가장 많이 발생하며, 국내 간암의 주요원인은 B 형과 C 형 간염 바이러스가 대표적이다. 간암을 진단하기 위해서는 직접 암 조직을 획득하여 검사해야하는 조직학적 진단인 조직생체검사가 필수적이나, 바늘이나 내시경 등을 이용해 인체에 침습적으로 시행되어 환자에게 불안감을 초래할 수 있으며 반복적으로 조직 생검이 필요한 경우 부작용 및 환자나 의사 모두에게 부담스럽다는 단점이 있다. 또한 조직 생검을 통하여 얻은 종양 샘플의 경우, 종양의 비균질성으로 인하여 전체 암 조직의 특성을 파악하기가 힘들다. 이에 비해 혈액이나 복수 등 인체 내 체액을 이용한 액체 생검은 체액 내에 떠다니는 세포유리 DNA 를 이용하여, 보다 덜 침습적이고 전반적인 암 조직의 특성을 잘 발견할 수 있다는 장점이 있으나, 체내에서의 반감기가 짧아서 세포유리 DNA 의 양이 제한적이다. 이를 극복하기 위하여, 본 연구에서는 차세대 염기서열방식 (NGS)을 이용하여 간세포암종의 세포유리 DNA 를 이용한 간세포암종의 체세포

돌연변이를 검출하고자 하였으며, 적은양의 세포유리 DNA 로 낮은 대립유전자 빈도의 변이를 검출하기 위해 분자바코드가 포함되어 있는 어뎁터를 라이브러리에 부착하여 시퀀싱을 실시하였다. 총 149 명의 환자의 혈장에서 세포유리 DNA 를 추출하여 분자바코드를 이용한 NGS 를 진행하였으며, 최소 0.5%의 대립유전자 빈도를 가진 변이를 검출하였다. 또한 혈장에서 분리한 세포유리 DNA 와 생체 종양 조직 DNA 를 비교하여, 세포유리 DNA 와 종양 조직 DNA 의 동일한 변이를 검출할 수 있었다. 이를 근거로 하여, 두 검사결과를 비교하여 더욱 확실한 진단 결과를 담보할 수 있을 뿐만 아니라 액체 생검이 조직 생검의 대체적인 방법으로 제시될 수 있음을 시사한다.

핵심되는 말 : 차세대 염기 서열 분석법, 분자 바코드, 세포 유리 DNA, 간세포암종