Patients with polyclonal hepatocellular carcinoma are at a high risk of early recurrence and have a poor recurrence-free survival period

Masaki Kaibori

Kansai Medical University Hirakata Hospital: Kansai Ika Daigaku Fuzoku Hirakata Byoin

Kazuko Sakai

Kindai University Faculty of Medicine

Hideyuki Matsushima

Kansai Medical University Hirakata Hospital: Kansai Ika Daigaku Fuzoku Hirakata Byoin

Hisashi Kosaka

Kansai Medical University Hirakata Hospital: Kansai Ika Daigaku Fuzoku Hirakata Byoin

Kosuke Matsui

Kansai Medical University Hirakata Hospital: Kansai Ika Daigaku Fuzoku Hirakata Byoin

Mitsugu Sekimoto

Kansai Medical University Hirakata Hospital: Kansai Ika Daigaku Fuzoku Hirakata Byoin

Kazuto Nishio (✉ knishio@med.kindai.ac.jp)

Kindai University Faculty of Medicine <https://orcid.org/0000-0002-8275-0846>

Research Article

Keywords: Clonal composition, Hepatocellular carcinoma, Recurrence-free survival, Copy number variation, Gene expression, Gene set enrichment analysis, Early recurrence, Single nucleotide polymorphism array, B-allele frequency, Proliferation

Posted Date: August 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-782031/v1>

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

Version of Record: A version of this preprint was published at Hepatology International on January 1st, 2022. See the published version at <https://doi.org/10.1007/s12072-021-10278-4>.

Abstract

Background/purpose of the study

Tumor heterogeneity based on copy number variations is associated with the evolution of cancer and its clinical grade. Clonal composition (CC) represents the number of clones based on the distribution of B-allele frequency (BAF) obtained from a genome-wide single nucleotide polymorphism (SNP) array. A higher CC number represents a high degree of heterogeneity. We hypothesized and evaluated that the CC number in hepatocellular carcinoma (HCC) tissues might be associated with the clinical outcomes of patients.

Methods

Somatic mutation, whole transcriptome, and copy number variations of 36 frozen tissue samples of operably resected HCC tissues by targeted deepsequencing, RNAseq and SNP array.

Results

The samples were classified the heterogeneous tumors as poly-CC (n = 26) and the homogeneous tumors as mono-CC (n = 8). The patients with poly-CC had a higher rate of early recurrence and a significantly shorter recurrence-free survival period than the mono-CC patients (7.0 vs. not reached, p = 0.0084). No differences in pathogenic non-synonymous mutations, such as *TP53*, were observed between the two groups when targeted deep sequencing was applied. A transcriptome analysis showed that cell cycle-related pathways were enriched in the poly-CC tumors, compared with the mono-CC tumors. poly-CC HCC is highly proliferative and has a high risk of early recurrence.

Conclusion

CC is a candidate biomarker for predicting the risk of early postoperative recurrence and warrants further investigation.

Introduction

Liver cancer is one of the most common malignancies, with more than 800,000 new cases diagnosed globally each year [1]. Hepatocellular carcinoma (HCC) accounts for approximately 75% of liver cancers. Although surgical resection improves the survival of HCC patients, early recurrence after hepatic resection is a poor prognostic factor for patients with HCC. HCC is a highly heterogeneous cancer [2–4]. Molecular heterogeneity and a lack of biomarkers for recurrence have contributed to the poor prognosis of HCC patients.

The intratumoral heterogeneity of cancer cells remains largely unexplored. Copy number variations reflect genomic structural changes that give rise to gene amplification, deletion, or copy number gain. Chromosomal

microarray and other array-based approaches have been widely adopted for the detection of whole-genome copy number variations[5, 6]. The OncoScan FFPE Assay Kit (Thermo Fisher Scientific) relies on molecular inversion probe (MIP) technology to detect genome-wide copy number alterations, loss of heterozygosity, and somatic mutation[7]. This assay provides the B-allele frequency (BAF), the log2 ratio (log2R), and the copy number for each of over 220,000 analyzed polymorphic genomic locations. The copy number is derived from the log2R and BAF values. The clonal composition of a tumor can be analyzed based on the BAF and log2R, which can be determined from the whole-genome copy number profiles obtained using the OncoScan FFPE Assay Kit. We previously demonstrated that the whole-genome single nucleotide polymorphism (SNP) array could be applied to the detection of the clonal composition of human ovarian cancer. Genome-wide segmentation data consisting of the log2R and BAF have been previously used to estimate the clonal composition (CC) number for individual tumors[8]. A similar approach resulted in a report indicating that the clonal composition can be derived from both somatic mutations as well as the log2R and BAF for loci with an aberrant copy number[9]. The CC number is not equivalent to the number of clones in the tumor tissue, but instead represents and estimates the number of clones that show copy number changes.

We hypothesized that HCC with a high CC would exhibit highly aggressive behaviors when compared with HCC with a low CC. We examined the CC numbers of previously collected frozen tissue samples of HCC and investigated the association with early recurrence in post-operative HCC patients. In addition, we also used NGS-based deep sequencing to detect somatic mutations and to examine the whole transcriptome of HCC tissues so as to investigate any associations with CC.

Materials And Methods

Hepatocellular carcinoma samples

Formalin-fixed, paraffin-embedded (FFPE) tumor tissue samples were obtained from 36 patients who underwent surgical resection at Kansai Medical University Hospital between August 2011 and December 2013. All the patients provided written informed consent to participate in the study, including the collection of tumor tissues for analysis. The early recurrence of postoperative HCC has been previously defined as recurrence within one year of resection[10, 11]. The recurrence-free survival (RFS) and overall survival (OS) were defined as the time interval between the date of hepatectomy and the date of recurrence or death, respectively. After surgery, patients were followed in the outpatient clinic every 3 months for the first 2 years and every 6 months thereafter. Early recurrence was defined as the appearance of a new lesion compatible with HCC on radiologic examination during the follow-up period.

Isolation of genomic DNA and RNA

DNA and RNA was isolated from frozen tissues using an AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The quality and quantity of the nucleic acid were verified using a NanoDrop 2000 device, PicoGreen dsDNA Reagent, and RiboGreen RNA reagent (all from Thermo Scientific, Wilmington, DE).

Microarray-based comparative genomic hybridization assay

A MIPs array for the genome-wide estimation of copy number aberrations was performed using the OncoScan FFPE Assay Kit (Thermo Fisher Scientific K.K., Tokyo, Japan), as previously reported.⁸ Briefly, 80 ng of DNA was subjected to annealing with these MIPs for 16 to 18 h, followed by enzyme digestion and two separate gap-fill reactions. The circular MIPs were then separately linearized for each gap fill with a cleavage enzyme and amplified using PCR. The PCR products were subjected to enzymatic cleavage and fragmentation, followed by hybridization for 16 to 18 h with two OncoScan arrays (one for each gap fill). The arrays were then stained and washed using a GeneChip Fluidics Station 450 and loaded into a GeneChip Scanner 3000 7G (Affymetrix). Array fluorescence intensity (CEL) files were generated using Affymetrix GeneChip Command Console (AGCC) software, version 4.0, and the CEL files were converted to OSCHP files using OncoScan Console software 1.3 and visualized with Chromosome Analyses Suite software (version 4.0). A representative whole-genome view is shown in Fig. S1.

Clonal composition analysis

Clonal composition numbers were calculated using BAF and log₂R information obtained from the OncoScan FFPE Assay using the Onco Clone Composition program, as described previously^[8]. A given copy number segment can be associated with a percentage of aberrant tumor cells (%AC) and is assumed to result from a single underlying event attributable to a single clone; in other words, the copy number segments are associated with the same %AC and belong to the same clone. The clonal composition number is calculated by identifying the number of different %AC detected among the aberrant segments. The number of minor alleles (NOMA) at a heterozygous site can be defined as the BAF of a segment. The calculation of clone composition number only uses NOMA = 0, which represents a segment with 100% loss of heterozygosity and a normal copy number. If only one of such alleles is present, the BAF for such a segment will be zero. The Onco Clone Composition program provides analytical estimations of clonal composition using a clustering approach that clusters all segments with the same log₂R and BAF and then combines clusters corresponding to the same %AC. A graphical representation of the aberrated segments and their association with %AC is shown in Supplementary Fig. S2.

Targeted DNA sequencing

A targeted DNA library comprising approximately 1.2-Mb of the coding regions of 409 genes for panel sequencing was constructed using an Ion AmpliSeq Comprehensive Cancer Panel (CCP) (Thermo Fisher Scientific), as described previously ^[12]. Briefly, 40 ng of DNA was subjected to multiplex PCR amplification using an Ion AmpliSeq Library Kit 2.0 and the Ion AmpliSeq Comprehensive Cancer Panel (Thermo Fisher Scientific), which covers all exons in 409 genes. After multiplex PCR, Ion Xpress Barcode Adapters (Thermo Fisher Scientific) were ligated to the PCR products, which were then purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA). The purified libraries were pooled and then sequenced using an Ion Torrent S5 instrument and the Ion 550 Chip Kit (all from Thermo Fisher Scientific). DNA sequencing data were accessed through the Torrent Suite ver. 5.10 program (Thermo Fisher Scientific). Reads were aligned against the hg19 human reference genome, and variants were called using Variant Caller, ver. 5.10. Raw variant calls were filtered with a quality score of < 100 and were manually checked using the integrative genomics viewer (IGV; Broad Institute, Cambridge, MA). Germline mutations were excluded using the Human Genetic Variation Database (<http://www.genome.med.kyoto-u.ac.jp/SnpDB>) and the Exome Aggregation

Consortium database. Variants with FATHMM scores greater than 0.7 were classified as pathogenic [13]. The tumor mutation burden (TMB mutations/Mb) was assessed using Ion Reporter Software, ver. 5.10 (Thermo Fisher Scientific).

Whole transcriptome analysis

The whole transcriptome analysis was performed using the AmpliSeq Transcriptome Human Gene Expression Kit (Thermo Fisher Scientific). For library preparation, a barcoded cDNA library was first generated using the SuperScript VILO cDNA Synthesis kit (Thermo Fisher Scientific) from 10 ng of total RNA. Then, cDNA was amplified for 12 cycles by adding PCR Master Mix and the AmpliSeq human transcriptome gene expression primer pool (Thermo Fisher Scientific). After multiplex PCR, Ion Xpress Barcode Adapters (Thermo Fisher Scientific) were ligated to the PCR products, which were then purified using Agencourt AMPure XP beads (Beckman Coulter). The purified libraries were pooled and then sequenced using an Ion Torrent S5 instrument and the Ion 550 Chip Kit (all from Thermo Fisher Scientific). The Ion Torrent Suite v5.10 software (Thermo Fisher Scientific) was used for base calling, alignment to the human reference genome (hg19), and quality control. Raw reads were then analyzed automatically using the AmpliSeqRNA plugin to generate gene-level expression values for all 20802 RefSeq human genes.

Gene set enrichment analysis (GSEA)

A Gene Set Enrichment Analysis (GSEA) was performed to identify pathways enriched in the Molecular Signatures Database (MSigDB) Hallmark gene set [14, 15]. A nominal p value of < 0.05 and an FDR (false discovery rate) q value of < 0.05 were considered statistically significant.

Gene selection and pathway analysis

A total of 2,408 genes that were differentially expressed according to CC numbers were selected using the following criteria: p value < 0.05 and FDR q value < 0.5 . Among these 2,408 genes, the top 500 markers of differential expression were identified based on the signal-to-noise ratio. Hierarchical clustering was performed using one minus the Pearson correlation coefficient as a distance measure and the average linkage method. To explore the potential biological pathways, genes in a cluster were submitted to the Metascape tool (<https://metascape.org/gp/index.html>) [16].

Statistical analysis

Categorical variables were compared using the Fisher exact test. Continuous variables were compared between groups with the Mann-Whitney U test. Survival was estimated using the Kaplan-Meier method and the log-rank test method. All the statistical analyses were performed using JMP software, version 14.2 (SAS Institute, Cary, NC), and Prism software, version 8.4 (GraphPad Software, San Diego, CA). A P value of < 0.05 was considered statistically significant.

Results

Clonal composition number of resectable hepatocellular cancer

The genome-wide copy number screen identified a heterogeneous pattern based on the BAF and the log2R profile. In BAF plots, a value of 0.5 indicates a heterozygous genotype (AB), whereas values of 0 and 1 indicate homozygous genotypes (AA and BB, respectively). The BAF distribution in the segments with copy number changes suggests the presence of different clones with different BAF patterns intratumorally. In segments with copy number gains, the percentage of aberrant cells with different BAF patterns due to BAF variation is difficult to evaluate because the copy number gain is not constant. On the other hand, there is no such limitation in evaluating the BAF in segments with copy number loss, and the clonal composition number can be derived from the presence of aberrant cells with a different BAF pattern. We analyzed 36 frozen samples of HCC using a whole-genome SNP array (Fig. S1) and estimated the clonal composition number using the Onco Clone Composition estimation program (Fig. S2). The mean $\pm$ SD value for the CC number was 1.0 ± 0.8 (range, 0 to 3). It should be noted that a case with a CC number of 0 indicates a homogenous tumor mass. The CC number could not be estimated in two cases because the copy number data did not plot onto the fitting curve in the program. Representative plots for the CC profile are shown in Fig. S2. The HCC samples were categorized into mono-CC (CC = 0, n = 8) and poly-CC (CC $\geq$ 1, n = 26) groups that respectively reflected homogenous and heterogenous clones.

Clinicopathological features associated with clonal composition status

The clinicopathological characteristics of patients with mono-CC or poly-CC are summarized in Table 1. No significant differences were observed in sex, mean age, liver function status (Child-Pugh A or B), etiology, histological grade, or stage as determined using the Fisher exact test. The RFS and OS of the mono- and poly-CC groups were compared (Fig. 1). Interestingly, early recurrence was significantly more frequent in the poly-CC group than in the mono-CC group ($p = 0.0127$). The median RFS was not reached in the mono-CC HCC groups and 7.0 months in the poly-CC HCC groups. The RFS of patients with poly-CC HCC was significantly shorter than that of the mono-CC patients (Fig. 1A, $p = 0.0084$). The OS of patients with poly-CC tended to be shorter than that of patients with mono-CC, although the difference was not significant (Fig. 1B). These results suggest that the CC number might predict early recurrence and prognosis after curative resection for HCC.

Table 1
Clinicopathological features associated with clonal composition in surgical resected hepatocellular carcinoma.

		CC = 0 (n = 8)	CC ≥ 1 (n = 26)	<i>p</i>
Sex	Male/Female	6/2 (75.0/25.0)	15/11 (57.7/42.3)	0.4438
Mean age (years)		72 (57–81)	71 (50–87)	
Liver function status	Child-Pugh A/B	7/1 (87.5/12.5)	24/2 (92.3/7.7)	1.0000
Etiology	HBV/HCV/alcoholic/normal	1/5/0/2 (12.5/62.5/0/25.0)	5/8/2/11 (19.2/30.8/7.7/42.3)	0.5036
Histologic grade	well/moderate/poor/unknown	4/3/0/1 (50.0/37.5/0/12.5)	4/19/2/1 (15.4/73.1/7.7/3.9)	0.1340
Stage	I/II/III/IVa	1/2/3/2 (12.5/25.0/37.5/25.0)	0/5/18/3 (0/19.2/69.2/11.5)	0.1735
Early recurrence	Yes/No	2/6 (25.0/75.0)	20/6 (76.9/23.1)	0.0127*
<i>P</i> value by Fisher's exact test for categorical data. *: significant (< 0.05)				

Association of clonal composition status with somatic mutations

Somatic non-synonymous mutations in tissue samples (n = 36) were successfully analyzed using targeted deep sequencing. Figure 2 summarizes the pathogenic variants with a FATHMM score ≥ 0.7 among the somatic variants. Pathogenic mutations of *TP53* (9/36, 25.0%) and *CTNNB1* (5/36, 13.9%) were the most frequently identified mutations seen in the 36 samples that were analyzed, as reported previously [17]. No significant association was seen between the presence of a pathogenic *TP53* gene mutation and grouping in the mono- and poly- CC groups (12.5% vs. 30.8%, $p = 0.4030$). Pathogenic *CTNNB1* mutations tended to occur more frequently in the mono-CC group, but the difference was not significant (37.5% vs. 7.7%, $p = 0.0721$). The tumor mutation burden (TMB) has been reported as the total number of non-synonymous variants or single nucleotide variants per tumor genomic region [18]. The TMB level of HCC is relatively low, compared with other types of solid cancers [19]. The TMB score of the presently reported cohort was also not high (median TMB = 6.7), and no difference in TMB was observed between the poly-CC (median, 6.7; range, 2.5–12.7) and the mono-CC (median 6.7; range 4.2–12.6) groups.

Gene expression profile and functional enrichment analysis

Thus far, our findings have indicated that the biological behavior of poly- and mono-CC tumors may differ. To investigate the biological functional differences between mono- and poly-CC tumors, we performed gene-set enrichment and pathway analysis using gene expression data from the Ion AmpliSeq Transcriptome Human

Gene Expression assay. GSEA was performed for the enrichment of 50 hallmark gene sets from the MSigDB database. The enriched hallmarks with $p < 0.05$ and a false discovery rate (FDR) with $p < 0.05$ are shown in Table 2. Notably, the top five enriched gene set in the poly-CC group were related to E2F targets, the G2M checkpoint, *Myc* targets, the mitotic spindle, and DNA repair (Fig. 3). These results suggest that cell cycle progression is enhanced in poly-CC tumors.

Table 2
Gene Set Enrichment Analysis (GSEA) results according to the MSigDB Hallmark gene sets.

MSigDB Hallmark Pathway	NES	NOM p-value	FDR q-value
E2F targets	3.20	0	0
G2M checkpoint	3.12	0	0
Myc targets v1	2.38	0	0
Mitotic spindle	2.31	0	0
DNA repair	2.02	0	0
Myc targets v2	1.82	0	0.0003
mTORC1 signaling	1.81	0	0.0004
Spermatogenesis	1.69	0.0015	0.0020
Oxidative phosphorylation	1.58	0	0.0077
Unfolded protein response	1.58	0.0044	0.0069
Protein secretion	1.52	0.0104	0.0136
Cholesterol homeostasis	1.46	0.0162	0.0205
NES; normalized enrichment score, NOM; nominal, FDR; false discovery rate.			

To further explore this notion, we analyzed significant gene sets in the mono- and poly-CC tumors. A hierarchical clustering analysis of 500 differentially expressed genes revealed the enrichment of cluster I and cluster II genes in mono- and poly-CC tumors, respectively (Fig. 4A). Early recurrent patients exhibited the enrichment of cluster II genes. We conducted a gene ontology (GO) analysis using Metascape and found significant enrichment of the cellular pathways associated with the immune response and cytokine-related pathways in cluster I, which was enriched in mono-CC tumors (Fig. 4B). On the other hand, cell cycle-related pathways including those related to E2F targets, the G2M checkpoint, and DNA replication were enriched in cluster II, which was enriched in poly-CC tumors (Fig. 4C). These results suggest that poly-CC and mono-CC tumors could be classified as “proliferation class” and “non-proliferation class,” respectively[20].

Discussion

Tumor heterogeneity is considered to be associated with a poor prognosis and outcome in cancer patients. In this study, we evaluated the CC number, a surrogate marker of intratumor heterogeneity, to investigate its

clinical relevance in HCC. In general, HCC is reported to be highly heterogeneous [2–4]. In our sample set, the rate of homogeneous tumors with a CC number of 0 and the rate of heterogeneous tumors with a CC number of greater than 1 were 23.5% and 76.5%, respectively, suggesting that HCC was often a heterogeneous tumor in our cohort. Heterogeneous tumors with poly-CC had a significantly higher rate of early recurrence. In addition, patients with poly-CC tumors were significantly associated with a poor postsurgical RFS. This finding suggests that HCC with high heterogeneity has a higher risk of recurrence after surgery. The CC number may be useful for predicting the risk of recurrence, but further studies are essential.

A gene mutation analysis was performed on the same sample set, and pathogenic mutations were selected using the FATHMM score. Therefore, the frequencies of *TP53* and *CTNNB1* mutations were lower than previously reported [17]. The gene mutation analysis showed no association between the CC number and the gene mutations status, including the TMB. The calculation of CC number is based on the copy number loss, suggesting that the diversity of copy number variations is reflected by the CC number. In lung cancer, subclonal evolution based on copy number changes is reportedly associated with the RFS and OS to a greater degree than those based on single nucleotide mutations [21]. Evaluation of the CC number based on copy number loss may be more informative for predicting clinical outcome than the detection of single nucleotide variants in hepatocellular cancer.

In an enriched transcriptome analysis using a supervised approach, cell cycle-related pathways were among the top 5 enriched pathways in poly-CC tumors. Oshi et al. showed that G2/M scoring is associated with cell proliferation-related factors and the breast cancer grade using an enrichment analysis based on a gene expression analysis of metastatic ER-positive breast cancer [22]. The pathways enriched in our cohort were related to the entire cell cycle progression in the G1, S and G2/M phases, suggesting active involvement in cell cycle progression and proliferation. Recently, Llovet *et al.* [20]. reviewed the molecular classification of HCC that classifies HCC into “proliferation” and “non-proliferation” classes based on a previous transcriptome analysis [23]. The proliferation class is characterized by enriched pathways related to the cell cycle, TP53 mutation, and HBV-positive populations. Compared with the mono-CC group, the poly-CC group in our sample cohort tended to exhibit enriched E2F and other cell cycle pathways, a high *TP53* mutation rate (8/26 [30.8%] vs. 1/8 [12.5%]), and HBV positivity (5/26 [19.2%] vs. 1/8 [12.5%]), although these differences were not significant because of the small sample size. Thus, poly-CC tumors (cluster II) are consistent with these molecular and etiological characteristics of the “proliferative class.” On the other hand, the non-proliferation class is associated with alcohol use, HCV, and NASH and a molecular profile characterized by enhanced IL6-JAK-STAT signaling or *CTNNB1* mutation-mediated Wnt-beta-catenin signaling associated with inflamed pathogenicity [24–26]. Mono-CC tumors, compared with poly-CC tumors, are characterized by non-HBV (7/8 [87.5%] vs. 21/26 [80.8%]), *CTNNB1* mutation (3/8 [37.5%] vs. 2/26 [7.7%]), and enriched cytokine related pathways. Thus, mono-CC tumors (cluster I) are consistent with these molecular and etiological characteristics of the “non-proliferative class” .

In conclusion, HCCs with poly-CCs, a surrogate of heterogeneity based on copy number variation, are enriched in genes involved in the cell cycle pathway, have a proliferative phenotype, and have a higher risk of early postoperative recurrence. The limitations of this study include its retrospective design and small sample size. Future studies examining larger sample sizes are needed to confirm the findings of this study.

Declarations

Acknowledgement We thank the participating patients and their families as well as all investigators and operations staff. The authors also thank Mr. Yoshihiro Mine (Center for Instrumental Analyses Central Research Facilities, Kindai University Faculty of Medicine) and Ayaka Kitano (Department of Genome Biology, Kindai University Faculty of Medicine) for technical assistance provided during the study.

Author Contributions M. Kaibori contributed substantially to the conception, design, and planning of the study, acquisition of data, analysis of the data, interpretation of the results, and drafting the manuscript; K. Sakai contributed substantially to the conception, design, and planning of the study, acquisition of data, analysis of the data, interpretation of the results, and drafting the manuscript; K. Nishio contributed substantially to the conception, design, or planning of the study, analysis of the data, interpretation of the results, and drafting the manuscript; H. Matsushima, H. Kosaka, , K.Matsui, and M. Sekimoto contributed substantially to the acquisition of data, and interpretation of the results. All authors contributed substantially to critically reviewing or revising the manuscript for important intellectual content, and approved the final manuscript.

Animal Research (Ethics) Not applicable

Plant Reproducibility Not applicable

Clinical Trials Registration Not applicable

Funding

This work was supported in part by a Grant-in Aid for Scientific Research (C) from the Japan Society for the Promotion of Science Grant Numbers JP19K07722 (KS).

Data Availability Yes. Data generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflict of interests Kazuko Sakai reports personal fees from AstraZeneca, Bio-Rad Laboratories, Chugai Pharmaceutical, Roche Diagnostics, Hitachi, outside the submitted work. Kazuto Nishio reports grants and personal fees from Eli Lilly, Nippon Boehringer Ingelheim, grants from Ignyta, Korea Otsuka Pharmaceutical, Thoracic Oncology Research Group, North East Japan Study Group, and personal fees from Chugai Pharmaceutical, Eisai, Pfizer, Novartis Pharma, MSD, Ono Pharmaceutical, Bristol-Myers Squibb, SymBio Pharmaceuticals, Life Technologies Japan, Solasia Pharma, Yakult Honsha, Roche Diagnostics, AstraZeneca, Otsuka Pharmaceutical, Sanofi, Guardant Health, Amgen, outside the submitted work. Masaki Kaibori, Hideyuki Matsushima, Hisashi Kosaka, Kosuke Matsui and Mitsugu Sekimoto declare no conflicts of interest.

Consent to participate This study was conducted in compliance with the Helsinki Declaration and the ethical guidelines for medical and health research involving human subjects in Japan. This study was approved by the institutional ethics review boards of Kansai Medical University Hospital (#1429) and Kindai University, Faculty of Medicine (27-033).

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(6):394–424.
2. Chan LK, Tsui YM, Ho DW. & Ng IO Cellular heterogeneity and plasticity in liver cancer. *Semin Cancer Biol.* 2021.
3. Shimada S, Mogushi K, Akiyama Y, Furuyama T, Watanabe S, Ogura T, Ogawa K, Ono H, Mitsunori Y, Ban D, Kudo A, Arii S, Tanabe M, Wands JR. & Tanaka S Comprehensive molecular and immunological characterization of hepatocellular carcinoma. *EBioMedicine.* 2019; 40457–470.
4. Wu Y, Liu Z, Xu X. Molecular subtyping of hepatocellular carcinoma: A step toward precision medicine. *Cancer Commun (Lond).* 2020;40(12):681–93.
5. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet.* 2007;39(7 Suppl):16–21.
6. Davies JJ, Wilson IM. & Lam WL Array CGH technologies and their applications to cancer genomes. *Chromosome Res.* 2005;13(3):237–48.
7. Foster JM, Oumie A, Togneri FS, Vasques FR, Hau D, Taylor M, Tinkler-Hundal E, Southward K, Medlow P, McGreeghan-Crosby K, Halfpenny I, McMullan DJ, Quirke P, Keating KE, Griffiths M, Spink KG. & Brew F Cross-laboratory validation of the OncoScan(R) FFPE Assay, a multiplex tool for whole genome tumour profiling. *BMC Med Genomics.* 2015; 85.
8. Sakai K, Ukita M, Schmidt J, Wu L, De Velasco MA, Roter A, Jevons L, Nishio K. & Mandai M Clonal composition of human ovarian cancer based on copy number analysis reveals a reciprocal relation with oncogenic mutation status. *Cancer Lett.* 2017; 40522–28.
9. Miller CA, White BS, Dees ND, Griffith M, Welch JS, Griffith OL, Vij R, Tomasson MH, Graubert TA, Walter MJ, Ellis MJ, Schierding W, DiPersio JF, Ley TJ, Mardis ER, Wilson RK. & Ding L SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol.* 2014;10(8):e1003665.
10. Yu B, Liang H, Ye Q, Wang Y. Establishment of a Genomic-Clinicopathologic Nomogram for Predicting Early Recurrence of Hepatocellular Carcinoma After R0 Resection. *J Gastrointest Surg.* 2021;25(1):112–24.
11. Zhang XP, Chen ZH, Zhou TF, Li LQ, Chen MS, Wen TF, Shi J, Guo WX, Wu MC, Lau WY. & Cheng SQ A nomogram to predict early postoperative recurrence of hepatocellular carcinoma with portal vein tumour thrombus after R0 liver resection: A large-scale, multicenter study. *Eur J Surg Oncol.* 2019;45(9):1644–51.
12. Sakai K, Tsuboi M, Kenmotsu H, Yamanaka T, Takahashi T, Goto K, Daga H, Ohira T, Ueno T, Aoki T, Nakagawa K, Yamazaki K, Hosomi Y, Kawaguchi K, Okumura N, Takiguchi Y, Sekine A, Haruki T, Yamamoto H, Sato Y, Akamatsu H, Seto T, Saeki S, Sugio K, Nishio M, Okabe K, Yamamoto N. & Nishio K

- Tumor mutation burden as a biomarker for lung cancer patients treated with pemetrexed and cisplatin (the JIPANG-TR). *Cancer Sci.* 2021;112(1):388–96.
13. Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day IN, Gaunt TR. & Campbell C An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics.* 2015;31(10):1536–43.
 14. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D. & Groop LC PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34(3):267–73.
 15. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES. & Mesirov JP Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–50.
 16. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C. & Chanda SK Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* 2019;10(1):1523.
 17. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology.* 2015;149(5):1226–39 e1224.
 18. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, Barron DA, Zehir A, Jordan EJ, Omuro A, Kaley TJ, Kendall SM, Motzer RJ, Hakimi AA, Voss MH, Russo P, Rosenberg J, Iyer G, Bochner BH, Bajorin DF, Al-Ahmadie HA, Chaft JE, Rudin CM, Riely GJ, Baxi S, Ho AL, Wong RJ, Pfister DG, Wolchok JD, Barker CA, Gutin PH, Brennan CW, Tabar V, Mellinghoff IK, DeAngelis LM, Ariyan CE, Lee N, Tap WD, Gounder MM, D'Angelo SP, Saltz L, Stadler ZK, Scher HI, Baselga J, Razavi P, Klebanoff CA, Yaeger R, Segal NH, Ku GY, DeMatteo RP, Ladanyi M, Rizvi NA, Berger MF, Riaz N, Solit DB. Chan TA & Morris LGT Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet.* 2019;51(2):202–6.
 19. Fessas P, Spina P, Boldorini RL, Pirisi M, Minisini R, Mauri FA, Simpson F, Olivieri P, Gennari A, Wong CN, Siddique A, Goldin RD, Akarca AU, Marafioti T. & Pinato DJ Phenotypic Characteristics of the Tumour Microenvironment in Primary and Secondary Hepatocellular Carcinoma. *Cancers (Basel).* 2021; 13(9).
 20. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J. & Finn RS Hepatocellular carcinoma. *Nat Rev Dis Primers.* 2021;7(1):6.
 21. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, Shafi S, Johnson DH, Mitter R, Rosenthal R, Salm M, Horswell S, Escudero M, Matthews N, Rowan A, Chambers T, Moore DA, Turajlic S, Xu H, Lee SM, Forster MD, Ahmad T, Hiley CT, Abbosh C, Falzon M, Borg E, Marafioti T, Lawrence D, Hayward M, Kolvekar S, Panagiotopoulos N, Janes SM, Thakrar R, Ahmed A, Blackhall F, Summers Y, Shah R, Joseph L, Quinn AM, Crosbie PA, Naidu B, Middleton G, Langman G, Trotter S, Nicolson M, Remmen H, Kerr K, Chetty M, Gomersall L, Fennell DA, Nakas A, Rathinam S, Anand G, Khan S, Russell P, Ezhil V, Ismail B, Irvin-Sellers M, Prakash V, Lester JF, Kornaszewska M, Attanoos R, Adams H, Davies H, Dentre S, Tanriere P, O'Sullivan B, Lowe HL, Hartley JA, Iles N, Bell H, Ngai Y, Shaw JA, Herrero J, Szallasi Z, Schwarz RF, Stewart A, Quezada SA, Le Quesne J, Van Loo P, Dive C. Hackshaw A & Swanton C Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med.* 2017;376(22):2109–21.

22. Oshi M, Takahashi H, Tokumaru Y, Yan L, Rashid OM, Matsuyama R, Endo I. & Takabe K G2M Cell Cycle Pathway Score as a Prognostic Biomarker of Metastasis in Estrogen Receptor (ER)-Positive Breast Cancer. *Int J Mol Sci.* 2020; 21(8).
23. Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, Villanueva A, Newell P, Ikeda K, Hashimoto M, Watanabe G, Gabriel S, Friedman SL, Kumada H, Llovet JM. & Golub TR Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res.* 2009;69(18):7385–92.
24. Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, Gupta S, Moore J, Wrobel MJ, Lerner J, Reich M, Chan JA, Glickman JN, Ikeda K, Hashimoto M, Watanabe G, Daidone MG, Roayaie S, Schwartz M, Thung S, Salvesen HB, Gabriel S, Mazzaferro V, Bruix J, Friedman SL, Kumada H, Llovet JM. & Golub TR Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med.* 2008;359(19):1995–2004.
25. Hou J, Zhang H, Sun B. & Karin M The immunobiology of hepatocellular carcinoma in humans and mice: Basic concepts and therapeutic implications. *J Hepatol.* 2020;72(1):167–82.
26. Lachenmayer A, Alsinet C, Savic R, Cabellos L, Toffanin S, Hoshida Y, Villanueva A, Minguez B, Newell P, Tsai HW, Barretina J, Thung S, Ward SC, Bruix J, Mazzaferro V, Schwartz M, Friedman SL. & Llovet JM Wnt-pathway activation in two molecular classes of hepatocellular carcinoma and experimental modulation by sorafenib. *Clin Cancer Res.* 2012;18(18):4997–5007.

Figures

Figure 1

. Kaplan-Meier curves of recurrence-free survival (RFS) (A) and overall survival (OS) (B) for patients with mono-CC (CC=0) (blue line) and poly-CC (CC≥1) (red line).

Figure 2

Profiling of pathogenic mutations detected in HCC tissue samples. The different mutation types, CC numbers, and cases with recurrences are color-coded as indicated in the figure.

Figure 3

Enrichment plots of gene expression signatures for top five pathways in poly-CC tumors. NES, normalized enrichment score; NOM p value, Nominal p value; FDR q value: false discovery rate.

Figure 4

Hierarchical clustering analysis of differentially expressed genes. (A) Hierarchical clustering analysis of 250 upregulated and 250 downregulated genes in poly-CC tumors compared with mono-CC tumors. Clustering was based on average linkage and the one minus Pearson correlation distance using Morpheus, and two major clusters (clusters I and II) were formed. (B) Plot shows the top biological pathways enriched in cluster I from the GO Biological Processes, Hallmark gene sets, and KEGG database using Metascape. (C) Plot shows the top biological pathways enriched in cluster II from the GO Biological Processes, Hallmark gene sets, and KEGG database using Metascape.

Supplementary Files

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

- [figS1.pdf](#)
- [figsS2.pdf](#)