Trunk events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma

Authors: Sara Torrecilla[1,2*], Daniela Sia[1*], Andrew N. Harrington[1,3*], Zhongyang Zhang[4,5], Laia Cabellos[2], Helena Cornella[2], Agrin Moeini[2], Genis Camprecios[1], Wei-Qiang Leow [1,6], Maria Isabel Fiel[1], Ke Hao[4,5], Laia Bassaganyas[2], Milind Mahajan[1], Swan N. Thung[1], Augusto Villanueva[1], Sander Florman[1], Myron E. Schwartz[1], Josep M. Llovet[1,2,7]

Affiliations:

(1) Mount Sinai Liver Cancer Program (Divisions of Liver Diseases, Department of Medicine, Department of Pathology, Recanati Miller Transplantation Institute), Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, USA.

(2) Liver Cancer Translational research laboratory, BCLC Group, IDIBAPS, Liver Unit, Hospital Clinic, University of Barcelona, Barcelona, Catalonia, Spain.

(3) Mount Sinai West/St Lukes, Department of Surgery, Residency Program, New York, USA.

(4) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA

(5) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, USA.

(6) Department of Anatomical Pathology, Singapore General Hospital

(7) Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain.

1

2 * These authors contributed equally to this work

3

4 **Corresponding Author:**

5 Josep M. Llovet, MD. Professor of Medicine

6 Icahn School of Medicine at Mount Sinai. Mount Sinai Liver Cancer Program - Division of

7 Liver Diseases.

8 1425 Madison Avenue, Box 11-23. New York 10029, NY. Phone: 1-2126599503. Fax:

9 212-849-2574. Email: Josep.Llovet@mssm.edu

10

11 **Key Words:** liver cancer, gatekeeper, driver, tumor evolution, clonality.

12

13 **List of abbreviations:** HCC: hepatocellular carcinoma; SNP: single nucleotide

14 polymorphism; DN: dysplastic nodule; sHCC: small-HCC; IM: intra-hepatic metastasis;

15 HBV: hepatitis B virus; HCV: hepatitis C virus; LGDN: low-grade dysplastic nodule;

16 HGDN: high-grade dysplastic nodule; eHCC: early HCC; pHCC: progressed HCC; VAF:

17 variant allele frequency; CNV: copy-number variation; FFPE: formalin-fixed paraffin

18 embedded; gDNA: genomic deoxyribonucleic acid; PCR: polymerase chain reaction;

19 LRR: log R ratio; BAF: B allele frequency; Mb: megabase; DASL: cDNA-mediated

20 Annealing, Selection, Extension, and Ligation; GEO: Gene Expression Omnibus; GSEA:

21 gene set enrichment analysis; FDR: false discovery rate; LOH (loss-of-heterozygosity);

22 FDA: food and drug administration.

23

Word count: 6090

Number of Figures and Tables: 6 Figures

Disclosures: The authors declare no conflict of interest

Financial support: Sara Torrecilla is supported by the Spanish Ministry of Economy and Competitiveness (BES-2014-068300 and EEBB-I-16-11251). Laia Bassaganyas is supported by the Juan de la Cierva Fellowship. Augusto Villanueva is supported by the U.S. Department of Defense (CA150272P3), The Tisch Cancer Institute, and the American Association for the Study of Liver Diseases Foundation (AASLDF) Alan Hofmann Clinical and Translational Award. Josep M. Llovet is supported by grants from the U.S. Department of Defense (CA150272P3), European Commission Framework Program 7 (HEPTROMIC, proposal number 259744) and Horizon 2020 Program (HEPCAR, proposal number 667273-2), the Recanati / Miller Transplantation Institute, the Asociación Española Contra el Cáncer (AECC), Samuel Waxman Cancer Research Foundation, Spanish National Health Institute (SAF2013-41027 and SAF 2016-76390) and Grup de Recerca Consolidat –- Recerca Translacional en Oncologia Hepàtica. AGAUR (Generalitat de Catalunya), SGR 1162.

Authors Contributions: ST (study concept and design, acquisition of data, analysis and interpretation of data, statistical analysis, drafting of the manuscript); DS (study concept and design, acquisition of data, analysis and interpretation of data, statistical

analysis, critical revision of the manuscript for important intellectual content, study supervision); ANH (study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript); ZZ (acquisition of data, analysis and interpretation of data, statistical analysis); LC (acquisition of data); HC (acquisition of data); AM (analysis and interpretation of data, critical revision of the manuscript for important intellectual content); GC (acquisition of data); WQL (acquisition of data , analysis and interpretation of data), MIF (acquisition of data); KH (analysis and interpretation of data); LB (analysis and interpretation of data, critical revision of the manuscript for important intellectual content); MM (acquisition of data); SNT (acquisition of data, analysis and interpretation of data); AV (acquisition of data , critical revision of the manuscript for important intellectual content); SF (acquisition of data); MES (acquisition of data); JML (study concept and design, drafting of the manuscript, critical revision of the manuscript for important intellectual content, obtained funding, study supervision).

Abstract

Background and Aims: According to the clonal model of tumor evolution, trunk alterations arise at early stages and are ubiquitous. Through the characterization of early stages of hepatocarcinogenesis, we aimed to identify trunk alterations in HCC and study their intra- and inter-tumor distribution in advanced lesions.

Methods: 151 samples representing the multi-step process of hepatocarcinogenesis were analyzed by targeted-sequencing and SNP array. Genes altered in early lesions [31 dysplastic nodules (DNs) and 38 small HCCs (sHCC)] were defined as trunk. Their distribution was explored in: a) different regions of large tumors (43 regions, 21 tumors), and b) different nodules of the same patient [39 tumors, 17 patients]. Multinodular lesions were classified as intrahepatic metastases (IMs) or synchronous tumors based on chromosomal aberrations.

Results: *TERT* promoter mutations (10.5%) and broad copy-number aberrations in chromosomes 1 and 8 (3-7%) were identified as trunk gatekeepers in DNs and were maintained in sHCCs. Trunk drivers identified in sHCCs included *TP53* (23%) and *CTNNB1* (11%) mutations, and focal amplifications or deletions in known drivers (6%). Overall, *TERT, TP53* and *CTNNB1* mutations were the most frequent trunk event and at least one was present in 51% of sHCCs. Around 90% of mutations in these genes were ubiquitous among different regions of large tumors. In multinodular HCCs, 35% of patients harbored IMs; 85% of mutations in *TERT, TP53* and/or *CTNNB1* were retained in primary and metastatic tumors.

Conclusions: Trunk events in early stages (*TERT, TP53*, *CTNNB1* mutations) were ubiquitous across different regions of the same tumor and between primary and metastatic nodules in >85% of cases. This concept supports the knowledge that single biopsies would suffice to capture trunk mutations in HCC.

Lay summary

Trunk alterations arise at early stages of cancer and are shared among all malignant cells of the tumor. In order to identify trunk alterations in HCC, we characterized early stages of hepatocarcinogenesis represented by dysplastic nodules and small lesions. Mutations in *TERT*, *TP53* and *CTNNB1* genes were the only ones found at this early stage. Analyses in more advanced lesions showed that mutations in these same genes were shared between different regions of the same tumor and between primary and metastatic tumors, suggesting their trunk role in this disease^{0}

Introduction

Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer and the second leading cause of cancer-related mortality worldwide[1]. Unlike other neoplasms, HCC usually arises in a previously damaged organ. Liver cirrhosis caused by well-known etiologies (i.e. HBV, HCV, alcohol consumption and metabolic syndrome) is the underlying disease in more than 80% of cases[2]. The development of HCC from cirrhosis follows a multistep process with a defined sequence of lesions, starting with the generation of low-grade and high-grade dysplastic nodules (LGDNs and HGDNs) that eventually develop into early HCC and culminate into advanced malignancy[2]. Although there is a progressive natural history of the disease, most HCC patients are still diagnosed at intermediate or advanced stages, when curative approaches are not feasible. In patients with advanced disease, the only therapeutic options able to improve survival include the multikinase inhibitors sorafenib in first line[3], and regorafenib in second line[4]. In this scenario, the discovery of effective targeted therapies remains an important challenge underscoring the need to better understand the mechanisms driving tumor progression and dissemination.

It is well-known that tumors evolve by acquiring a series of genomic and epigenomic alterations over time, following a sequential process of clonal expansion and selection[5]. As a result of this process, solid tumors may be comprised of subpopulations of cells with distinct genomic alterations and divergent biological behaviors. In addition to such intra-tumor heterogeneity, the existence of multinodular HCCs and the resulting inter-

1 tumor heterogeneity adds a new level of complexity that, in the era of personalized
2 medicine, is likely to impact targeted therapies and biomarker discovery. Although many
3 biological aspects of intra- and inter-tumor heterogeneity remain obscure, it is overall
4 accepted that tumors arise from a single cell and trunk alterations are the first pro-
5 oncogenic molecular events arising during tumor evolution. Therefore, they would be
6 clonally dominant and present in all tumor cells. Trunk events include both “gatekeeper”
7 alterations in pre-malignant stages, necessary but no sufficient for tumor onset (i.e APC
8 mutations in colorectal cancer), as well as “driver” events (i.e Her2/neu amplifications in
9 breast cancer), able to confer a selective growth advantage to malignant cells[6]. In
10 contrast, alterations appearing in more advanced stages occur in only a subpopulation
11 of tumor cells, for which they are called “branch”, and are subjected to selection
12 pressures such as hypoxic conditions and anti-tumoral treatment[7,8].
13 Over the last years, there has been an improvement in understanding the molecular
14 pathogenesis of HCC. Next-generation sequencing technologies have provided a clear
15 picture of the main alterations driving this disease[2]. Nevertheless, little effort has been
16 directed to discriminate trunk gatekeeper and driver alterations that arise early in HCC
17 development. To date, mutations in *TERT* promoter represent the only gatekeeper
18 alterations described in early stages of hepatocarcinogenesis, being already present in
19 25% of dysplastic nodules[9,10]. Nonetheless, other trunk gatekeepers and driver events
20 occurring during early stages of liver carcinogenesis remain currently ill-defined.
21
22 In this study we aimed to characterize trunk gatekeepers and drivers by identifying
23 molecular alterations in early hepatocarcinogenesis and studying their distribution

1 through the different stages of the disease. By using targeted-deep-sequencing and
2 SNP array in dysplastic nodules, we identified *TERT* promoter mutations (10.5%) and
3 broad copy-number alterations in chromosomes 1 and 8 (13.8%) as the only trunk
4 gatekeepers. Mutations in *TP53* and *CTNNB1* represent trunk driver alterations arising
5 in small HCC tumors (23% and 11.4%, respectively). Finally, deep-sequencing analysis
6 of the candidate trunk genes confirmed that *TERT, TP53* and *CTNNB1* mutations are
7 ubiquitous events across different regions of the same tumor and between primary and
8 metastatic lesions in 85-90% of cases. This concept supports the knowledge that single
9 biopsies would suffice to capture trunk driver mutations in HCC.
10

1 # Materials and Methods

2

3 ## Human samples, pathology review and extraction

4 Upon institutional review board approval, a total of 151 samples were collected from
5 patients treated with surgical resection or liver transplantation at three hospitals
6 belonging to the HCC Genomic Consortium: Hospital Clínic, Barcelona (Spain) -leading
7 institution-, Mount Sinai School of Medicine, NY (USA); and Instituto Nazionale dei
8 Tumori, Milan (Italy). Samples used in this study represented the stepwise process of
9 hepatocarcinogenesis and were distributed in 3 different cohorts: **1)** preneoplastic and
10 initial lesions, **2)** single large HCC tumors for the study of intra-tumor heterogeneity, and
11 **3)** multinodular HCC cases to assess inter-tumor heterogeneity (**Fig. 1**). The first cohort
12 comprised early stages of the disease including 31 pre-malignant dysplastic nodules
13 (DNs) and 38 small HCCs (sHCC). Dysplastic diagnosis was confirmed by two expert
14 liver pathologists (M.I.F. and S.N.T) following previously published guidelines[11,12].
15 Accordingly, 31 DNs obtained from cirrhotic transplanted patients were classified as
16 LGDNs (n=15) and HGDNs (n=16). sHCC lesions were defined as carcinoma-in-situ
17 lesions and/or single tumors ≤2cm in diameter, without satellites or vascular invasion
18 and were further sub-classified as early or progressed by two expert pathologists
19 (S.N.T. and W.Q.L) according to the histopathological criteria proposed by the
20 International Consensus Group for Hepatocellular Neoplasia[11,12]. Among the 38 sHCCs,
21 14 lesions were sub-classified as early sHCCs (eHCC) and 22 as progressed sHCCs
22 (pHCCs); 2 small tumors could not be sub-classified because H&E slides were not
23 available (**Supplementary Table 1**).

Cohort #2, previously reported[13], comprised 21 resected large HCC tumors (> 4cm in diameter). For each of the large HCCs, surrounding non-tumoral liver, center and periphery fresh-frozen samples (separated by a minimum distance of 2 cm) were collected, for a total of 43 sample regions. Cohort #3 included 39 formalin-fixed paraffin embedded (FFPE) HCCs from 17 patients presenting multinodularity (2 or 3 non-satellite foci HCCs) (**Supplementary Table 2**). FFPE slides of multinodular tumors have been examined by two expert pathologists (S.N.T. and W.Q.L.) to assess differentiation grade, architectural pattern and cytological characteristics. In tumors presenting different patterns for each variable, the pattern displayed in ≥50% of the examined area was considered dominant and used for comparisons. Non-tumoral cirrhotic liver was collected for all cases. Trunk genes were defined as those with alterations present in (a) early stages of the disease -dysplastic nodules and small tumors from cohort #1, and (b) ubiquitous in all regions of large tumors (cohort #2) or intra-hepatic metastases (cohort#3).

For further details regarding deep targeted-sequencing, genome-wide analysis of DNA copy number alteration, whole-genome gene expression profiling and statistical analyses, see the **Supplementary Materials and Methods section**.

1 # Results

2

3 ## Identification of candidate trunk gatekeeper alterations in pre-neoplastic lesions

4 To identify potential gatekeeping trunk alterations in HCC, we analyzed 31 pre-
5 malignant DNs –including 15 LGDNs and 16 HGDNs- collected from cirrhotic patients at
6 the time of transplant (cohort #1, **Fig. 1, Supplementary Table 4**). First, we
7 characterized early CNV events in these samples using SNP array data. The study of
8 gross structural alterations showed that DNs presented a stable genome with an
9 average percentage of aberrant chromosomal arms of 0.5% (**Fig. 2A**). No significant
10 differences between LGDN and HGDN (0.5% each) were observed. Broad gains and
11 losses were detected in few cases. In particular, large gains (6.9%, 2/29) and losses
12 (6.9%, 2/29) in chromosome 8 were the most frequent events in dysplastic lesions,
13 including both LGDNs and HGDNs (**Fig. 2B, Supplementary Fig. 1)**. Broad gains in 1q
14 and losses in 22q were also identified (3.4% each). Interestingly, well-known oncogenes
15 such as *MYC* and *MDM4* or *PARP1* were included in all the regions 8q and 1q affected
16 by chromosomal aberrations (**Fig. 2B**). No high-level focal alterations in already known
17 driver genes were identified in dysplastic lesions.

18

19 Additionally, DNs were submitted to deep targeted-sequencing to unravel gatekeeper
20 mutational events. A specific discovery panel was designed to cover 20 well-
21 characterized and frequently mutated genes in HCC (**Supplementary Table 3**).
22 Targeted regions in each sample were sequenced to a median depth of 952× (48-2694),
23 with ~96% coverage. After applying the filtering criteria described in Materials and

Methods, a total of 2 *TERT* promoter mutations were identified in 14 DNs (14.3%) (**Fig. 2C, Supplementary Table 5**). Besides *TERT* mutations, no other mutations were identified in the remaining 19 genes included in the panel. Therefore, the characterization of trunk gatekeepers in DNs confirmed that, among the 20 explored HCC drivers, the only recurrent mutations present in premalignant lesions occur in *TERT* promoter. Nonetheless, broad CNVs in chromosomes 1 and 8 were identified in 13.8% (4/29) of dysplastic tissues containing potential trunk gatekeeper genes, such as *MYC*, *MDM4* and *PARP1*.

Identification of candidate trunk driver alterations in small HCCs

Following the natural history of HCC progression, we next sought to identify trunk alterations with a potential driving role in small tumoral lesions. To this purpose, we collected early and progressed small HCCs which can both develop from pre-existing dysplastic foci or nodule,[12] (cohort #1, **Fig. 1**). A total of 38 sHCCs -including 14 eHCCs and 22 pHCCs (2 non-classified tumors)- were submitted for molecular analysis (**Supplementary Table 4**). SNP array data showed a significant increase in the percentage of aberrant chromosomal arms compared to DNs (8.9% vs 0.5%, $p<0.0001$), with no differences between eHCCs and pHCCs (3.4% vs 6.9%, $p=0.4$) (**Fig. 2A**). Broad gains in 1q (47.1%, 8/17) and 8q (29.4%, 5/17), as well as broad losses in 8p (35.3%, 6/17) were frequent events in sHCCs (**Fig. 2B, Supplementary Fig. 1**). Interestingly, significantly higher frequency of chr 1q gains was observed in sHCCs compared to DNs [3.4% vs 47.1%, ($p=0.03$)] (**Supplementary Table 6**). Since these broad CNVs were already present in DNs, these data might support their potential

gatekeeper role in hepatocarcinogenesis. Similarly to DNs, all the regions presenting broad gains in chromosomes 1q and 8q contained well-known oncogenes such as *MYC*, *PARP1* and/or *MDM4* (**Fig. 2B**). Additional recurrent chromosomal aberrations were identified in sHCCs and included broad losses in 17p (35.3%, 6/17), 4q (23.5%, 4/17), 16p (23.5%, 4/17) and 16q (23.5%, 4/17) (**Fig. 2B, Supplementary Fig. 1**). Although absent in DNs, high-level focal amplifications in 6p21.1 (*VEGFA)* and 8q24.21 (*MYC*), as well as homozygous-deletions in 9p21.3 (*CDKN2A)*, 16p13.3 (*AXIN1)* and 4q35.1 (*IRF2*), were detected in sHCCs (5.8% each, 1/17), pointing them as potential trunk drivers (**Supplementary Table 7**).

The identification of candidate trunk driver mutations was performed by evaluating the 20 genes included in the previously described discovery panel (**Supplementary Table 3**) in 35 sHCCs (14 eHCCs, 19 pHCCs and 2 small tumors non-classified). The average number of mutations per sample was significantly higher in small tumors than DNs (0.7 vs 0.1, p=0.004), confirming that genomic complexity progressively increases from pre-neoplastic lesions to small tumors. As expected, *TERT* promoter mutations (35.5%) were the most frequent events in small tumors confirming their trunk-gatekeeper role (**Fig. 2C, Supplementary Table 5**). In addition, we identified *TP53* (23%) and *CTNNB1* (11.4%) as the most commonly mutated genes in sHCCs. Taking into account an average 60% of tumoral cells in the sequenced regions[14], and the presence of the mutation in a single allele of tumoral cells, we would expect at least 30% VAF for trunk mutations. Accordingly, the average VAF for the identified mutations was 52.2%, (70.4% for *TERT*, 47% for *TP53* and 30.7% for *CTNNB1*). Overall, 51.6% of our cohort

(16 out of the 31 sHCCs with TERT promoter data available) presented at least one potential trunk driver mutation (**Fig. 2C, Supplementary Table 7**). In small HCC, we confirmed that the presence of TP53 mutations was significantly associated with the molecular subclasses G1-2-3 (p=0.04), whereas CTNNB1 mutations were enriched in G5-6 subclasses (p=0.04), as previously described[15,16]. However, no significant association was observed between the previously reported HCC molecular classes G1-G6, and the histological type (eHCC vs pHCC)[15,16]. Co-occurrence of loss-of-heterozygosity (LOH) and mutations in *TP53* gene was not detected in any case.

In summary, the molecular events identified in DNs -broad CNVs in chromosomes 1 and 8 and mutations in *TERT* promoter region- were also observed at higher rates in sHCCs, confirming its gatekeeper role. Molecular events arising exclusively in sHCCs included focal CNVs (amplifications in *VEGFA* and *MYC;* deletions in *CDKN2A, AXIN1* and *IRF2*) and mutations in *TERT*, *TP53* and *CTNNB1*, pointing them as trunk drivers in hepatocarcinogenesis. Overall, mutations in *TERT*, *TP53* and *CTNNB1* were the most frequent potential trunk alterations in early stages of the disease.

Mutations in the identified trunk genes are ubiquitously distributed in different regions of large tumors

After identifying the most recurrent alterations that arise in the first stages of tumor development, we next sought to study their intra-tumoral distribution in HCC to confirm their trunk role. We first designed a deep-sequencing validation panel to further study the most frequent events (*TERT*, *TP53* and *CTNNB1* mutations) identified in cohort #1

(**Supplementary Table 3**). To study the spatial intra-tumoral heterogeneity of these potential trunk events, we sequenced 2-3 regions of 21 large single HCC tumors (>4cm in average from cohort #2). A total of 43 tumoral regions were submitted to deep-sequencing with the validation panel [median depth: 695x (192-1656)]. After applying the pre-defined filtering criteria, 63 total mutations were identified in the 5 analyzed genes (**Fig. 3A, Supplementary Table 8**). In this cohort, *TERT* (80.9%, 17/21) was the most frequently mutated gene, followed by *TP53* (47.6%, 10/21) and *CTNNB1* (19%, 4/21). *AXIN1* and *ARID1A* genes were mutated in only 1 patient each (4.8%). When assessing the spatial distribution of the identified mutations, we found that 89% (56/63) of the mutations were common to the tumor center and periphery (**Fig. 3A, Supplementary Table 8**). In particular, *TERT* promoter mutations were shared between center and periphery of the tumor in 88% of the 17 patients (15/17) harboring the event (**Fig. 3B)**. Similarly, *TP53* and *CTNNB1* were found to be ubiquitous in 70% (7/10) and 75% (3/4) of patients, respectively. Mutations in *AXIN1* and *ARID1A* were also shared between regions in their respective patients. On the other hand, 7 mutations were found in only one of the tumoral regions.

Overall, 90% (17/19) of patients harbored at least 1 ubiquitous mutation (**Fig. 3A**). Only in 2 patients (9.5%) we did not identify mutations in the candidate trunk genes. Altogether, these results show that the most frequent early alterations identified – mutations in *TERT*, *TP53* and *CTNNB1* - are shared across different regions of large tumors in ~90% of cases, reinforcing their potential role as trunk alterations in HCC.

Copy number profile confidently differentiates intra-hepatic metastases and synchronous tumors

To further delineate the inter-tumoral heterogeneity of the identified candidate trunk events, we aimed to study their distribution in different tumors of the same patient. To this purpose, we collected 39 multinodular tumors from 17 patients (2-3 non-satellite foci HCCs) (cohort #3) (**Supplementary Table 2**). Since presence of multinodularity can reflect the growth of independent synchronous tumors or intrahepatic metastases (IMs), we first classified the 39 multinodular cases accordingly. Tumor-clonality was defined by measuring the similarity of genome-wide CNV profiles between nodules as described in Materials and Methods. CNV profiles predicted that 33.3% (13/39) of analyzed nodules were IMs (clonal tumors) and 66.7% (26/39) were synchronous (non-clonal) (**Fig. 4A-B, Supplementary Table 9**). According to the number of nodules and the clonality prediction, 3 different profiles of patients were captured: 1) patients harboring 2 or 3 IMs (23.5%, 4/17), 2) patients harboring 2 or 3 synchronous tumors (64.7%, 11/17) and 3) patients harboring 2 IMs and 1 independent tumor (11.8%, 2/17) (**Fig. 4A)**.

CNV-based classification was further validated with whole genome expression data. To this end, hierarchical clustering was performed to measure genetic distance between single nodules. The analysis revealed that IMs showed proximity to its paired tumor and clustered around the same node in 100% of cases, as opposed to the 17% (4/23) of the non-clonal tumors (**Fig. 4C**). Interestingly, GSEA comparing the two groups showed significant enrichment of functions related to cell cycle, proliferation and metastasis in the IMs group ($p<0.05$, $FDR<0.10$) (**Supplementary Table 10**). In addition, among the 350 genes up-regulated in IMs versus synchronous tumors, we found *KIF15*, *MAGEA1*

and *MAGEC3*, genes previously reported to be highly expressed in metastatic HCC and melanoma (Fold change ≥2.4, p<0.001, FDR<0.10, **Supplementary Table 10**)[17,18]. Interestingly, higher percentage of aberrant chromosomal arms was observed in the group of IMs compared to synchronous tumors (22.9% vs 8.9%, p=0.003). Pathological review of different histomorphological characteristics of multinodular cases did not show any correlation between morphological patterns and molecular classification. In fact, in ~80% of all possible pairs (22/27), at least 2 out of 3 histomorphological characteristics were maintained, regardless of the tumor clonality, suggesting that although pathological review could be helpful, molecular analysis is required to differentiate IMs from versus synchronous tumors (**Supplementary Table 11**).

Broad gains in 1q, previously identified as trunk in DNs and sHCCs, were the most frequent CNV identified in IMs (76.9%, 10/13) (**Supplementary Fig. 2**) and, as such, were shared between IMs pairs in 90% of the cases (9/10). Similarly, 8p losses and 8q gains were ubiquitous in 75% (4/5) and 100% (4/4) of IM cases, respectively. No focal CNVs in known HCC driver genes were found in IMs. Finally, in terms of clinico-pathological parameters, presence of IM tumors was significantly associated with presence of satellites (p=0.022) and recurrence (4/6 vs 0/9, p=0.004, **Supplementary Table 12**). In contrast, all patients with synchronous tumors were HBV-infected HCC patients (0/6 vs 6/10, p=0.03, **Supplementary Table 12**).

Overall, clonality prediction based on CNV profiles confidently classified multinodular HCCs in IMs (33%) and synchronous tumors (67%). Genetic proximity was observed in all IMs. Previously identified trunk CNVs in chr1 and 8 were shared among all nodules in ~90% of IMs.

Mutations in the identified trunk genes are ubiquitously distributed in primary tumors and intrahepatic metastases

Once multinodular cases were assigned to metastatic or non-metastatic groups, we next assessed the distribution of candidate trunk alterations in these samples. To this purpose, 22 multinodular tumors (11 IMs and 11 synchronous tumors) from 9 patients of cohort #3 were submitted to targeted-deep-sequencing using the validation panel (**Supplementary Table 3**). With a median sequencing depth of 852x (162-22270), 120 non-synonymous somatic SNV [range 1-38, 5.45 average per sample] were identified. Out of 120, 22 mutations were further selected for validation (**Supplementary Table 13**). The mutational analysis revealed that 81.8% (9/11) of the IMs presented ubiquitous mutations in the primary and metastatic tumor, while only 1/11 (9%) of synchronous tumors shared mutational events (**Fig. 5A**, **Supplementary Table 13**). For example, as represented in **Fig. 5A** and **Supplementary Table 13**, all lesions of patient #459 showed mutations in TP53, but only the 2 IMs shared the same mutated position (p.Y220H), while the independent tumor presented a different altered locus (p.R249S).

Overall, in IMs a total of 13 mutations were identified in the 5 evaluated genes, and 85% of them (11/13) were shared between primary and intrahepatic metastasis (**Fig. 5A**, **Supplementary Table 13**). *TERT* and *TP53* genes were mutated in 53.8% (17/13) and 15.4% (2/13) of IMs, respectively. In all cases, these mutations were shared between paired IM tumors (**Fig. 5B**). *CTNNB1* was found mutated in 4 IMs and ubiquitous in 2 of these cases (50%). No mutations were found in *AXIN1* and *ARID1B*. In the synchronous tumor group, patients #457 and #460 harbored mutations in *TP53* and

1 *CTNNB1*, respectively. Interestingly, the mutated position was different in the nodules of
2 the same patient (**Supplementary Table 12**). No presence of TP53-LOH was detected
3 in tumors with *TP53* mutations.
4 Overall, 85% of the mutations identified in *TERT*, *TP53* and *CTNNB1* genes were
5 ubiquitous in primary and intra-hepatic metastases, confirming their trunk role in HCC
6 development. On the other hand, only 9% of synchronous tumors shared trunk events,
7 indicating that they arise independently.
8

Discussion

Our study represents a comprehensive characterization of trunk events in HCC. The analysis of a large cohort of samples recapitulating the stepwise process of hepatocarcinogenesis has identified molecular gatekeeping events in preneoplastic stages of liver cancer, including broad CNVs in chromosomes 1q, 8p and 8q (3-7%) as well as oncogenic mutations in *TERT* promoter (10.5%). Additionally, focal CNVs and mutations in *TP53* and *CTNNB1* (11-23%) were found to arise in small HCC lesions. Overall, *TERT, TP53* and *CTNNB1* mutations were the most recurrent candidate trunks and, accordingly, mutations in these genes were ubiquitously distributed between different regions of the same tumor and between intra-hepatic metastases in 90% of cases (**Fig. 6**).

Over the past decade, the landscape of molecular alterations in HCC has been thoroughly explored. However, relatively little is known about the molecular events driving the early stages of the disease. The study of the temporal order of molecular alterations is of high interest in those tumor types that present multistage development. Indeed, the characterization of the different tumor stages represents an invaluable resource to further understand the disease. In the current study, taking advantage of the multistep process of hepatocarcinogenesis, we have systematically applied targeted-sequencing and SNP array to premalignant and small HCC lesions to identify trunk gatekeeper and trunk driver alterations. A similar approach has led to the identification of trunk alterations in lung adenocarcinoma[19]. To our knowledge, this is the first time

1 that the combination of temporal dissection of mutations and the study of spatial intra-
2 and inter-tumoral heterogeneity is used to characterize trunk alterations in liver cancer.
3 The characterization of the chromosomal aberrations confirmed that DNs present a very
4 stable genome compared to early HCC lesions with broad gains found only in 1q and 8q
5 and losses 8p (**Fig. 6**). Few previous publications have reported broad CNVs in liver
6 DNs[20,21]. However, this is the first study using high-resolution SNP array in ~30
7 samples. Although these alterations cannot be considered trunk *per se*, we hypothesize
8 that they could harbor potential gatekeeper genomic hits, such as *MYC* (8q), *PARP1*
9 (1q) and/or *MDM4* (1q). In this regard, the role of *MYC* in HCC malignant conversion
10 has been previously described [22,23]. Interestingly, a recent sequencing study in multiple
11 HCC lesions also reported gains in chr 1q and 8q and deletions in 8q as common trunk
12 events further supporting our observations[24]. All together, these data suggest that
13 chromosomal instability might represent an early hallmark in hepatocarcinogenesis.
14 In terms of mutations, *TERT* promoter was confirmed to be the only frequent
15 gatekeeper mutation in DNs (10.5%) and the most frequent in sHCCs (35.5%), a finding
16 consistent with previous studies[9,10,25,26]. Overall, these findings confirm the gatekeeper
17 role of *TERT* in hepatocarcinogenesis. Additionally, *TP53* and *CTNNB1* mutations were
18 frequent events in sHCCs (23% and 11.4%, respectively) with high VAFs (average
19 52%) (**Fig. 6**). Although these mutations have been previously reported in HCC, the
20 disease stage at which these alterations first appear remain poorly understood and they
21 were even suggested to be late genomic events[25]. Interestingly, our data suggest that
22 *TP53* and *CTNNB1* are trunk driver genes occurring in early malignant lesions.
23 Similarly, focal amplifications in *VEGFA* and *MYC* and homozygous-deletions in

1 *CDKN2A*, *AXIN1* and *IRF2* are already present in initial tumors (~6%), suggesting a
2 trunk driver role. However, this observation should be further validated in more
3 advanced HCCs.
4
5 Although extensive intra-tumor heterogeneity has been previously reported in
6 HCC[24,27,28] the most recurrent identified trunk alterations -*TERT*, *TP53*, *CTNNB1*
7 mutations- were shared between different regions of the same tumor and between intra-
8 hepatic metastases in around 90% of cases (ranging 50%-100%) (**Fig. 6**). Similarly,
9 broad CNVs in chromosomes 1 and 8 were ubiquitous events in 75%-100% in
10 multinodular tumors. These results support the theory of branched evolutionary tumor
11 growth by which early events are shared among all malignant cells within the same
12 tumor and between clonal lesions. Due to their ubiquitous distribution, trunk alterations
13 would be potentially captured with single biopsies, simplifying the problem of intra- and
14 inter-tumor heterogeneity[8], and might represent ideal therapeutic targets as shown in
15 other cancer types[7,8]. To date, FDA approved drugs for biomarker-selected populations
16 specifically target trunk alterations. Clear examples include EGFR and ALK
17 rearrangements in lung cancer [29], BRAF in melanoma [30,31]; ERBB2 in breast cancer[32].
18 Interestingly, these cases were sampled with single biopsies, no multisampling
19 approaches.
20
21 When analyzing the intra and inter-tumoral distribution of selected genes at more
22 advanced stages a small fraction of these events were not confirmed as trunk (15-10%
23 on average). This could be due to technical difficulties or could imply that in few cases

these driver events occur at later stages of the disease and thus, might represent branch drivers. Indeed, among all the studies assessing the presence of trunk alterations in different tumor types, the identification of “obligatory” early events is not frequent. The one example can be found in clear cell renal cell carcinoma, in which mutations in the von Hippel–Lindau (VHL) gene, together with the loss of chromosome 3p, are always trunk[33]. Nevertheless, further studies with more sophisticated techniques, such as single-cell-sequencing, are required to thoroughly understand the intra-tumoral distribution of *TERT*, *TP53* and *CTNNB1* alterations. Since our targeted-sequencing panel was designed to interrogate driver genes mutated in ≥1% of patients according to previously published genome-wide studies[25,26], we can conclude that *TERT*, *TP53* and *CTNNB1* mutations are the most frequent trunk alterations in HCC. Nevertheless, we cannot rule out that other genes could be trunk drivers in a small proportion of patients. In addition, we cannot exclude that other molecular events, such as miRNAs and aberrant methylation in driver genes[34,35], could be responsible of initial tumor onset and growth.

Our study offers a complete understanding of the genetic alterations that initiate and drive the progression of HCC. In particular, we identified gatekeeping (broad CNVs in chromosomes 1q, 8p and 8q, mutations in *TERT* promoter) and driver trunk alterations including focal CNVs (in *VEGFA*, *MYC*, *CDKN2A*, *AXIN1* and *IRF2*) and *TP53* and *CTNNB1* mutations. Overall, *TERT, TP53* and *CTNNB1* mutations were the most recurrent trunk alterations identified and, accordingly, showed limited intra- and inter-tumoral heterogeneity in >80% of cases. Therefore, these early mutations could be

captured with single biopsies and could represent ideal therapeutic targets in the near future.

Acknowledgements

This work was supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai. Authors acknowledge the Genomics core facility of IDIBAPS for the technical help, Dr Robert Sebra for technical support and the CERCA Programme / Generalitat de Catalunya.

References

1. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2015;385(9963):117-171.
2. Llovet JM, Zucman-Rossi J, Pikarsky E, et al. Hepatocellular carcinoma. *Nat Rev Dis Prim*. 2016;2:16018.
3. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med*. 2008;359(4):378-390.
4. Bruix J, Qin S, Merle P, et al. Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2017;389(10064):56-66.
5. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481(7381):306-313.
6. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW.

Cancer genome landscapes. *Science*. 2013;339(6127):1546-1558.

7. Alizadeh AA, Aranda V, Bardelli A, et al. Toward understanding and exploiting tumor heterogeneity. *Nat Med*. 2015;21(8):846-853.

8. McGranahan N, Swanton C. Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution. *Cancer Cell*. 2015;27(1):15-26.

9. Nault JC, Mallet M, Pilati C, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat Commun*. 2013;4:2218. doi:10.1038/ncomms3218.

10. **Nault JC, Calderaro J**, Tommaso L Di, et al. TERT promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. *Hepatology*. August 2014. doi:10.1002/hep.27372.

11. International Consensus Group for Hepatocellular Neoplasia. Pathologic diagnosis of early hepatocellular carcinoma: A report of the international consensus group for hepatocellular neoplasia. *Hepatology*. 2009;49(2):658-664.

12. Roncalli M, Terracciano L, Di Tommaso L, David E, Colombo M. Liver precancerous lesions and hepatocellular carcinoma: The histology report. *Dig Liver Dis*. 2011;43:S361-S372.

13. Villanueva A, Hoshida Y, Battiston C, et al. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology*. 2011;140:1501-1512.e2.

14. Carter SL, Cibulskis K, Helman E,. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012;30.

1 15. **Boyault S, Rickman DS, De Reyniès A,** et al. Transcriptome classification of
2 HCC is related to gene alterations and to new therapeutic targets. Hepatology.
3 2007;45(1):42-52.
4 16. Calderaro J, Couchy G, Imbeaud S, et al. Histological Subtypes of Hepatocellular
5 Carcinoma Are Related To Gene Mutations and Molecular Tumour Classification.
6 J Hepatol. 2017. doi:10.1016/j.jhep.2017.05.014.
7 17. **Miao R, Luo H**, Zhou H, et al. Identification of prognostic biomarkers in hepatitis
8 B virus-related hepatocellular carcinoma and stratification by integrative multi-
9 omics analysis. *J Hepatol*. 2014;61(4):840-849
10 18. Riker AI, Enkemann SA, Fodstad O, et al. The gene expression profiles of primary
11 and metastatic melanoma yields a transition point of tumor progression and
12 metastasis. *BMC Med Genomics*. 2008;1(1):13.
13 19. Izumchenko E, Chang X, Brait M, et al. Targeted sequencing reveals clonal
14 genetic changes in the progression of early lung neoplasms and paired circulating
15 DNA. *Nat Commun*. 2015;6:8258.
16 20. **Moinzadeh P, Breuhahn K,** Stützer H, Schirmacher P. Chromosome alterations in
17 human hepatocellular carcinomas correlate with aetiology and histological grade--
18 results of an explorative CGH meta-analysis. *Br J Cancer*. 2005;92:935-941.
19 21. **Raidl M, Pirker C,** Schulte-Hermann R, et al. Multiple chromosomal abnormalities
20 in human liver (pre)neoplasia. *J Hepatol*. 2004;40:660-668.
21 22. Kaposi-Novak P, Libbrecht L, Woo HG, et al. Central role of c-Myc during
22 malignant conversion in human hepatocarcinogenesis. *Cancer Res*.
23 2009;69(7):2775-2782.

1 23. Hunecke D, Spanel R, Länger F, Nam SW, Borlak J. MYC-regulated genes
2 involved in liver cell dysplasia identified in a transgenic model of liver cancer. *J*
3 *Pathol*. 2012;228(4):520-533.
4 24. **Xue R, Li R, Guo H, Guo L**, et al. Variable Intra-Tumor Genomic Heterogeneity of
5 Multiple Lesions in Patients With Hepatocellular Carcinoma. *Gastroenterology*.
6 2016;150(4):998-1008.
7 25. **Schulze K, Imbeaud S, Letouzé E**, et al. Exome sequencing of hepatocellular
8 carcinomas identifies new mutational signatures and potential therapeutic targets.
9 *Nat Genet*. 2015.
10 26. **Totoki Y, Tatsuno K, Covington KR**, et al. Trans-ancestry mutational landscape
11 of hepatocellular carcinoma genomes. *Nat Genet*. 2014;46(12):1267-1273.
12 27. Gao Q, Wang Z-C, Duan M, et al. Cell Culture System for Analysis of Genetic
13 Heterogeneity Within Hepatocellular Carcinomas and Response to Pharmacologic
14 Agents. *Gastroenterology*. 2017;152(1):232-242.e4.
15 28. **Zhai W, Lim TK-H, Zhang T**, et al. The spatial organization of intra-tumour
16 heterogeneity and evolutionary trajectories of metastases in hepatocellular
17 carcinoma. *Nat Commun*. 2017;8:4565.
18 29. Gridelli C, Rossi A, Carbone DP, et al. Non-small-cell lung cancer. *Nat Rev Dis*
19 *Prim*. 2015;1:15009.
20 30. Van Allen EM, Wagle N, Sucker A, et al. The genetic landscape of clinical
21 resistance to RAF inhibition in metastatic melanoma. *Cancer Discov*. 2014;4(1):94-
22 109.
23 31. Schadendorf D, Fisher DE, Garbe C, et al. Melanoma. *Nat Rev Dis Prim*. April

2015:15003.

32. Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012;149(5):979-993.

33. **Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP**, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet*. 2014;46(3):225-233.

34. Villanueva A, Portela A, Sayols S, et al. DNA Methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology*. February 2015.

35. Martinez-Quetglas I, Pinyol R, Dauch D, et al. IGF2 Is Up-regulated by Epigenetic Mechanisms in Hepatocellular Carcinomas and Is an Actionable Oncogene Product in Experimental Models. *Gastroenterology*. 2016;151(6):1192-1205.

Author names in bold designate shared co-first authorship.

FIGURE LEGENDS

Fig. 1. Study design. Three cohorts of samples representing the multistep process of hepatocarcinogenesis- were collected for the study. **Cohort #1** was used to identify trunk alterations, including gatekeepers and oncogenic drivers. The spatial distribution of trunk alterations was then assessed in different regions (center and periphery) of the same tumor (**Cohort #2**) and in primary tumors and intrahepatic metastases of multinodular HCCs (**Cohort #3**). Total number of samples per cohort and main clinical features of samples in each cohort are indicated.

Fig. 2. Identification of trunk alterations in early stages of hepatocarcinogenesis. A) Dot plot representation of the distribution of chromosome instability relative to tumor stage. Chromosomal instability was defined by measuring the percentage of aberrant chromosomal arms (PAA) in each patient. Middle bar indicates mean, and error bars indicate SEM. **B)** Frequency of broad CNVs in dysplastic nodules (LGDN and HGDN, left graph) and sHCCs (eHCCs and pHCCs, right graph). The bottom axis indicates the frequency of broad gains (red bars) and losses (blue bars). Candidate gatekeepers and drivers contained in regions are indicated. **C)** Heatmap summarizing the trunk mutations identified by deep-sequencing in DNs (blue) and trunk drivers in sHCCs (red). ns: non-significant; **: <0.01; ***: <0.001. LGDN: low-grade dysplastic nodules; HGDN: high-grade dysplastic nodules; eHCC: early HCC; pHCC: progressed HCC; NA: small tumors unclassified.

Fig. 3. Intra-tumor distribution of mutations in trunk candidates. A) Mutations identified by deep-sequencing in cohort #2 (large HCC, >4 cm) are shown in the heatmap. Each column indicates the region (center or periphery) sequenced per tumor; tumor regions are grouped per patient. Colors indicate the status of the mutations: green indicates ubiquitous mutations shared between regions whereas red indicates private mutations. **B)** Schematic representation of the distribution of the 5 trunk genes characterized in **cohort #2**. Thick bars indicate the distribution of the sum of events for all genes.

Fig. 4. CNV profiles confidently classify multinodular tumors in IMs and synchronous tumors. 39 tumors from 17 HCC patients presenting with 2 or 3 non-satellite foci HCCs were profiled for genome-wide CNV and gene expression landscapes. **A)** Similarity of CNV profiles was used to classify tumors as intrahepatic metastasis (IM) or synchronous tumors. The different clonality statuses identified in the multinodular patients are indicated. **B)** Representative case of IMs (upper panel) and synchronous tumors (lower panel) are shown. LRR and mBAF signals comparison is here represented. **C)** Gene expression-based hierarchical clustering was used to calculate genetic proximity for each tumor. Red areas indicate co-clustering of tumors belonging to the same patient. CNV-based classification as IMs and synchronous tumors (Sync) is indicated at the bottom.

Fig. 5. Distribution of mutations in trunk candidates in multinodular HCCs. A) Mutations identified by deep-sequencing in cohort #3 are shown in the heatmap. Each

column indicates the tumor nodule sequenced per patient; tumor nodules are grouped per patient. Intrahepatic-metastases (IMs) or synchronous tumors (ST) are indicated. Colors indicate the status of the mutations: green boxes indicate ubiquitous mutations shared between tumors whereas red boxes indicate private mutations. **B)** Representation of the distribution of the characterized trunk genes included in cohort #3 (intrahepatic metastases, left panel; synchronous tumors, right panel). Thick bars indicate the distribution of the sum of events for all genes.

Fig. 6. **Graphic overview of the identified trunk alterations and their onset along the hepatocarcinogenesis process**.