TITLE: **Mixed hepatocellular-cholangiocarcinoma tumors: cholangiolocellular carcinoma is a distinct molecular entity.**

AUTHORS: Agrin Moeini[1,2], Daniela Sia[2], Zhongyang Zhang[3,4], Genis Camprecios[2], Ashley Stueck[2], Hui Dong[1], Robert Montal[1], Laura Torrens[1], Iris Martinez-Quetglas[1], M. Isabel Fiel[2], Ke Hao[3,4], Augusto Villanueva[2], Swan N. Thung[2], Myron E. Schwartz[2], Josep M. Llovet[1,2,5].

AFFILIATIONS:

1. Liver Cancer Translational Research Laboratory, Liver Unit, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic, CIBERehd, Universitat de Barcelona, Barcelona, Catalonia, Spain.
2. Mount Sinai Liver Cancer Program, (Divisions of Liver Diseases, Hematology and Medical Oncology, Department of Medicine, Department of Pathology, Recanati Miller Transplantation Institute), Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, USA.
3. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA
4. Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, USA.
5. Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain.

KEYWORDS:

liver cancer, molecular subclass, progenitor-like origin, TGF-β signaling

CORRESPONDING AUTHOR:

Josep M Llovet, MD

Professor of Medicine.

Director, Mount Sinai Liver Cancer Program.

Division of Liver Diseases.

Tisch Cancer Institute

Icahn School of Medicine at Mount Sinai.

Madison Ave 1425. 11F-70. Box:1123

New York, NY10029. USA

Phone: 212-6599503

FAX: 212-849-2574

E-mail: Josep.Llovet@mssm.edu

ABBREVIATIONS:

HCC-CCA: Hepatocellular cholangiocarcinoma

HCC: Hepatocellular Carcinoma

iCCA: Intrahepatic Cholangiocarcinoma

WHO: World Health Organization

CLC: Cholangiolocellular Carcinoma

TGF-beta: Transforming Growth Factor, Beta 1

TERT: Telomerase Reverse Transcriptase

KRAS: Kirsten Rat Sarcoma Viral Oncogene Homolog

IDH1/2: Isocitrate Dehydrogenase 1/2

SNP: Single Nucleotide polymorphism

WES: Whole Exome Sequencing

SALL4: Spalt-Like Transcription Factor 4

IRB: Institutional Review Board

FFPE: Formalin-Fixed, Paraffin-Embedded

TACE: Transarterial Chemoembolization

FF: Fresh-Frozen

OCT: Optimal Cutting Temperature compound

KRT7: Keratin 7

KRT19: Keratin 19

HEP1: Hep Par 1

CK7: Cytokeratin 7

CK19: Cytokeratin 19

NCAM: Neural Cell Adhesion Molecule 1

GPC3: Glypican 3

EpCAM: Epithelial Cell Adhesion Molecule

CNV: Copy Number Variation

CCND1: Cyclin D1

FGF19: Fibroblast Growth Factor 19

EMT: Epithelial to Mesenchymal Transition

INF: Interferon

IPA: Ingenuity Pathway Analysis

IGF: Insulin-like Growth Factor

IGF1R: Insulin-Like Growth Factor 1 Receptor

MYC: V-Myc Avian Myelocytomatosis Viral Oncogene Homolog

BRAF: B-Raf Proto-Oncogene, Serine/Threonine Kinase

CTNNB1: Catenin Beta 1

TP53: Tumor Protein P53

FGFR2: Fibroblast Growth Factor Receptor 2

CN-LOH: Copy Neutral Loss of Heterozygosity

NTP: Nearest Template Predictions

GSEA: Gene Set Enrichment Analysis

FDR: False Discovery Rate

SD: Standard Deviation

RNA: Ribonucleic acid

DNA: Deoxyribonucleic acid

dsDNA: Double Stranded Deoxyribonucleic Acid

PCR: Polymerase Chain Reaction

qRT-PCR: Quantitative Real Time Polymerase Chain Reaction

HBV: Hepatitis B virus

HCV: Hepatitis C virus

PSC: Primary Sclerosing Cholangitis

WORD COUNT: 6100

NUMBER OF FIGURES AND TABLES: 8

CONFLICT OF INTEREST: The authors disclose no conflicts of interest.

FINANCIAL SUPPORT:

JML is supported by grants from the European Commission Horizon 2020 (HEPTROMIC, proposal number 259744; HEPCAR, proposal number 667273-2), the U.S. Department of Defense (CA150272P1), The Samuel Waxman Cancer Research Foundation, the Spanish National Health Institute (SAF-2013-41027-R) and the Asociación Española contra el Cáncer (AECC).AM is recipient of grants from theSpanish National Health Institute (FPI program), Asociación Española contra el Cáncer (AECC) and Asociación Española para el Estudio del Hígado (AEEH). DS is supported by the Andrea Marie Fuquay Memorial Research Fellowship (Cholangiocarcinoma Foundation). RM is supported by a Rio Hortega grant from Sociedad Española de Oncología Médica – Instituto de Salud Carlos III. LT and IM-Q are funded by CIBEREHD and HEPCAR, respectively. AV is the recipient of the American Association of the Study of the Liver Foundation Alan Hofmann Clinical and Translational Award, and the U.S. Department of Defense (CA150272P3).

AUTHOR CONTRIBUTION:

AM (study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; statistical analysis), DS (study concept and design; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; statistical analysis; study supervision), ZZ (acquisition of data; analysis and interpretation of data; statistical analysis), GC (acquisition of data; technical support), AS (analysis and interpretation of data), HD (acquisition of data), RM (acquisition of data), LT(acquisition of data), IMQ (acquisition of data), MIF (analysis and interpretation of data), HK (analysis and interpretation of data), AV (study concept and design; drafting of the manuscript; critical revision of the manuscript for important intellectual content); SNT (study concept and design; acquisition of data; analysis and

interpretation of data), MES (study concept and design), 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: Mixed hepatocellular-cholangiocarcinoma (HCC-CCA) is a rare and poorly understood type of primary liver cancer. We aimed to perform a comprehensive molecular characterization of this malignancy.

Methods: We performed gene expression profiling, DNA copy number detection, and exome sequencing using formalin-fixed samples from 18 patients with mixed HCC-CCA encompassing the whole histological spectrum of the disease. Comparative genomic analysis was performed with independent datasets of HCC (n=164) and iCCA (n=149).

Results: Integrative genomic analysis of HCC-CCAs revealed that cholangiolocellular carcinoma (CLC) represents a distinct biliary-derived entity compared with the stem-cell and classical types. CLC tumors were NCAM positive (6/6 vs 1/12, P<0.001), chromosomally stable (mean chromosomal aberrations 5.7 vs 14.1, P=0.008), showed significant upregulation of TGF-beta signaling and enrichment for inflammation-related and immune response signatures (P<0.001). Stem-cell tumors were characterized by SALL4 positivity (6/8 vs 0/10, P<0.001), enrichment of progenitor-like signatures, activation of specific oncogenic pathways (i.e. MYC and IGF), and signatures related to poor clinical outcome. Regarding classical type, a significant correlation in the copy number aberrations of the iCCA and HCC components suggested a clonal origin. Exome sequencing revealed an average of 63 non-synonymous mutations per tumor (mean driver mutations:2). Among those, *TP53* was the most frequently mutated gene (6/21, 29%) in HCC-CCAs.

Conclusions: Mixed HCC-CCA represents a heterogeneous group of tumors, with stem-cell type characterized by features of poor prognosis and classical type with common lineage for HCC and iCCA components. CLC stands alone as a distinct biliary-derived entity associated with chromosomal stability and TGF-beta signaling.

LAY SUMMARY

The molecular characterization of mixed hepatocellular cholangiocarcinoma (HCC-CCA) defined that cholangiolocellular carcinoma (CLC) is a distinct molecular entity with biliary-derived origin and no traits of HCC. On the other hand, within the mixed HCC-CCA, the stem-cell type shared aggressive phenotype and poor outcome whereas the classic type shows a common cell lineage for both the HCC and the iCCA component. These data supports re-defining the pathological classification of mixed HCC-CCA in light of the novel molecular data provided.

INTRODUCTION

Liver cancer is the second leading cause of cancer-related deaths, with more than 850,000 new cases annually worldwide [1]. Mixed hepatocellular-cholangiocarcinoma (HCC-CCA) is a rare type of primary liver cancer accounting for less than 1% of all primary liver malignancies [2,3]. Diagnosis is based on histological examination and requires unequivocal presence of both hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (iCCA) elements intimately admixed [2]. Due to its low incidence and the lack of an established consensual pathological diagnosis, the demographic features and clinical behavior of these tumors remain ill-defined. Roughly, the age, sex specific incidence and geographical distribution are similar to those for HCC [2,4,5]. Median overall survival rates of HCC-CCA are similar to iCCA [3,6–8]. To date, clinical practice guidelines do not include a specific treatment recommendation for HCC-CCA and surgical resection, when feasible, remains the standard of practice.

Histologically, mixed HCC-CCA is a heterogeneous group of primary liver tumors. According to the 2010 World Health Organization (WHO) classification [2], they are divided in two main categories: the classical type and the stem-cell features type. The classical type is characterized by areas of typical HCC and iCCA with an intermediate transition which holds mixed features of both entities. The category of stem-cell features is further subdivided into typical, intermediate and cholangiolocellular carcinoma (CLC). Subtypes with stem-cell features are composed of tumor cells with intermediate histological features between hepatocytes and cholangiocytes. In addition, recent studies have suggested the presence of distinct properties for each subtype of HCC-CCA with stem-cell features given their association with different clinicopathological factors [9,10].

Unlike HCC or iCCA, there is no genome-wide characterization of mixed HCC-CCA tumors. Indeed, it is unclear whether histologic subtypes have a well-defined correlate at

the molecular level. Gene expression profiling on a small series of HCC-CCA samples suggested that HCC-CCA might share common characteristics with poorly differentiated HCC and iCCA with stem-cell traits [11–15]. Furthermore, WNT/beta-catenin and TGF-β signaling were reported to be significantly activated in mixed HCC-CCA when compared to progenitor-like HCC[13]. Mutational analysis has suggested common recurrent driver mutations in HCC and HCC-CCA in comparison to iCCA, such as larger frequency of *TERT* promoter mutations and a lower frequency of *KRAS* and *IDH1/2* mutations [14]. On the other hand, genome wide allelotyping analyses of classical HCC-CCA suggested a closer genomic proximately to iCCA than to HCC [16].

Herein, we provide a comprehensive molecular characterization of mixed HCC-CCA including histological characterization, whole-genome expression profiling, single-nucleotide polymorphism (SNP) array, and whole-exome sequencing (WES). Integrated analysis to evaluate the genomic overlap with a large independent set of HCC and iCCA samples was also performed. Overall, integrative genomic analysis indicates that CLC is a distinct entity with a biliary molecular profile, low chromosomal instability, and enrichment of TGF-β and immune-related signaling. The other mixed tumors can be molecularly distinguished in two main subclasses: the stem-cell subclass characterized by the presence of the early progenitor marker (SALL4) and signatures of more aggressive phenotype, and the classical subclass, constituted by components of both HCC and iCCA with a clonal origin. Thus, we propose a molecular classification that encompasses two groups within the mixed HCC-CCA tumors (stem-cell and classical). In addition, our data suggest that CLC stands alone as an independent biliary-derived entity not sharing any molecular traits of HCC.

MATERIAL & METHODS

Human samples and nucleic acid extraction

For the purpose of this study, we evaluated 4728 consecutive patients who underwent surgery for primary liver cancer between 1994 and 2013 at the Icahn School of Medicine at Mount Sinai [HCC (4307, 91%), iCCA (360, 7.7%), and mixed HCC-CCA (61, 1.3%)], following local Institutional Review Board (IRB) approval. Among the 61 mixed HCC-CCA cases, 43 cases were excluded due to several reasons: a) pre-treated with locoregional therapies (such as transarterial chemoembolization, TACE), b) lack of sample availability or c) low tumor cell viability. The diagnosis of mixed HCC-CCA was confirmed by two expert hepato-pathologists (MIF and ST). A final set of 18 patients with available fixed paraffin-embedded (FFPE) samples and clinical data were selected and classified according to the latest WHO classification [2], as classical (n=4) or with stem-cell features (n=14). The stem-cell features subgroup included Typical (n=2), Intermediate (n=6), and CLC (n=6). For the purpose of molecular profiling in the case of the classical mixed tumors, nucleic acids were extracted separately from the HCC-like and CCA-like components (4 patients, 8 tumor samples total). **Table 1** summarizes the main clinico-pathological features of the 18 patients included in the study. For the purpose of integrative genomic analysis, molecular data previously reported by our group on HCC (n=164) and iCCA (n=149) was used [17,18]. In addition, for the identification of driver mutations in stem-cell features subtype, fresh frozen optimal cutting temperature compound (FF-OCT) embedded tumor tissues and corresponding normal tissue (n=6 pairs including 3 overlapping cases with above) were provided by the Mount Sinai Institutional Biorepository after IRB committee approval. For detailed description of nucleic acid extraction see supplementary Material and Methods section.

Immunohistochemistry, Whole-genome gene expression profiling, Genome-wide analysis of DNA copy number alteration, Whole-exome sequencing, and Statistical analyses

See Supplementary Materials and Methods section.

RESULTS

CLC, stem-cell and classical types are distinct entities.

In order to understand if the different histological subtypes of mixed HCC-CCA represent distinct subgroups of the disease, we performed gene expression-based unsupervised clustering. The unsupervised clustering analysis (**Figure 1**, **Supplementary Figure 1**) revealed three distinct clustered groups: 1) CLC tumors in cluster C, (*P*=0.0013), 2) stem-cell feature tumors in cluster D (*P*=0.0062), 3) Classical tumors depending on its components (HCC-component in clusters A and B, *P*=0.024). In addition, the fact that the iCCA-like components of the classical subtype co-clustered with either stem-cell feature tumors or CLC suggest the presence of common molecular traits among these (**Figure 1A**). Differential molecular profile of CLC with respect to other stem-cell feature tumors was further confirmed by integrative genomic analysis with an independent set of HCC (n=164) and iCCA (n=149) samples (**Figure 2**). CLC tumors significantly co-clustered together suggesting high genomic similarity among them in comparison to other primary liver tumors (*P*<0.001). Significant genomic proximity was also observed for stem-cell HCC-CCA (*P*<0.001). Moreover, CLC tumors co-clustered with iCCA from the proliferation class, whereas the stem-cell mixed tumors co-clustered with HCC with progenitor-like traits (**Figure 2**, *P*<0.001). Subsequent analysis of cell lineage with specific marker genes further corroborated the observation that CLC may represent a separate entity, as indicated by the expression of a biliary phenotype with significant up-regulation of biliary-specific genes (e.g. *KRT7*, *KRT19*, *ITGB4*) and down-regulation of hepatocyte-related genes (e.g. *ADH1A*, *ALB*, *APOB*, *HNF1A*)[19] (**Figure 1**). These findings were in concordance with the immunostaining profile (**Figure 1B and Figure 3**), which defined CLC tumors as negative for the hepatocyte marker HepPar1 (0/6 in CLC vs 10/12 in others, *P*=0.015, **Supplementary table 1**), but positive for biliary markers (CK7, CK19)

and more specifically the progenitor-like marker NCAM (6/6 in CLC vs 1/12 in others, *P*<0.0001, **Supplementary table 1**).

The stem-cell molecular subclass was characterized by the expression of both hepatocyte and biliary markers (**Figure 1B and Figure 3**), and the early progenitor cell marker SALL4 (6/8 vs 0/10 in rest of mixed tumors, *P*=0.0004, **Supplementary Table 1**). On the other end of the spectrum, the HCC-like and iCCA-like components of the classical subtype showed a biphenotypic profile with simultaneous expression of hepatocytic and biliary markers (**Figure 1B**) despite their distinct histological features (**Figure 3**). Comparison of gene expression levels and immunostaining grading scores showed significant correlation for all markers used (**Supplementary Figure 2**).

The different genomic profile of CLC tumors was further confirmed by the DNA copy number variation (CNV) analysis which revealed significantly higher chromosomal stability in CLC compared to the non-CLC tumors (5.7 mean alterations in CLC vs 14.1 for others, *P*=0.008, **Figure 4**). In contrast, classical and stem-cell subclasses presented frequent broad chromosomal aberrations, recapitulating those previously reported in both HCC and iCCA, including gains of 1q, and 8q, and losses of 4q, 8p, 9q, 16q and 17p (**Figure 4**). High-level amplifications of 11q13, harboring the oncogenes *CCND1* and *FGF19,* were detected in 3 cases of mixed HCC-CCA (**Supplementary Table 2**). Thus, the above findings support that mixed HCC-CCA tumors can be classified into 2 distinct molecular subclasses (stem-cell and classical). CLC tumors can be defined as a separate entity with biliary phenotype and no traits of HCC. We, then, further characterized the molecular traits of each of these tumor subtypes.

CLC subclass: characterized by TGF-β signaling and immune-related response signaling

When evaluating previously reported prognostic gene signatures in liver cancer, CLC showed a significant enrichment of S1 subclass [20] characterized by higher TGF-β activation, and Late TGF-β induced signature [21] (**Figure 5A**, **Supplementary Table 3**). Furthermore, unsupervised hierarchical clustering with independent set of iCCA and HCC samples, suggested a shared molecular profile of CLC and progenitor-like iCCA tumors(**Figure 2,** *P*<0.001). On the other hand, gene set enrichment analysis (GSEA) of canonical and hallmark oncogenic pathways showed that the CLC subclass was significantly enriched with TGF-β signaling, epithelial to mesenchymal transition (EMT), inflammation and immune response related signaling (i.e. TNF-α, lNF-ɣ, IL2/STAT5, IL6/STAT3, T cell response) (**Supplementary Table 4**). Notably, significant enrichment of immune cells [22] including T cells (i.e. effector memory, CD8+ PD1 high), cytotoxic lymphocytes, T cell helper 1 (Th1), natural killer (NK) and neutrophils was also detected in CLC (**Figure 5A-middle panel**, *P*<0.05). In addition, specific chemokines (e.g. *CXCL12*, *CCL2*, *CCL21*) and several cytokine receptors (i.e. *CXCR4, IL8RB, IL10RA, IL17RC*) were also identified to be significantly up-regulated in CLC (**Supplementary Figure 3, Supplementary Figure 4**). Furthermore, the Ingenuity Pathway Analysis (IPA) of the top deregulated genes in CLC (**Supplementary Table 5**), predicted TGF-β as the major activated upstream effector molecule in these tumors (**Supplementary Figure 5A**, z-score>2, p<0.0001). Specifically, the ligand *TGFB2* was found to be significantly up-regulated among the different TGF-β superfamily of ligands (**Supplementary Figure 3**). Other significantly deregulated candidate upstream effectors and top bio-functions in CLC included activation of TP53 (**Supplementary Figure 5A)**, and DNA damage response checkpoint regulation and Ataxia telangiectasia mutated (ATM) signaling (**Supplementary Figure 5B)**, respectively. Overall, TGF-β and pro-inflammatory response related signaling pathways were found to be specifically up-regulated in CLC tumors.

Stem-cell subclass: associated with progenitor-like phenotype and proliferative signaling pathways

We next sought to assess dominant genomic traits in the stem-cell subclass. GSEA of the SALL4 positive stem-cell tumors showed enrichment of progenitor-like liver cancer subtypes (i.e Stem cell [12], CK19 [17], EpCAM [23]) and more aggressive HCC — including HCC Proliferation subclass [24], cell-cycle deregulated G3 subclass [25], poor survival cluster A [26], and S2 subclass [20]— together with activation of IGF1R [27] and NOTCH [28] signaling (**Figure 5A**, **Supplementary Table 3**). Consistently, MYC, mTORC and NOTCH signaling were identified as the main canonical pathways associated with stem cell subclass (**Supplementary Table 6**). IGF2 signaling was further confirmed as one of the main deregulated signaling pathway networks in these tumors (**Figure 5C, Supplementary Table 7**). Another core node up-regulated in stem cell subclass included genes implicated in the hepatic specification of liver progenitor cells (i.e. *PROX1, HNF1B, FOXA1, FOXA3*), suggesting a more hepatocyte committed lineage in contrast to the biliary phenotype observed in CLC. In addition, consistent with SALL4 expression, the pluripotent embryonic related OCT4 signaling was found as one of the top activated canonical pathways (**Supplementary Figure 6**). On the other hand, in regards to CNV profiling (**Figure 4**), the most frequent chromosomal alterations reported in HCC showed an enrichment trend in stem-cell subclass, including 1q gains (6/8 vs 2/10), 8q gains (5/8 vs 1/10), 1p losses (4/8 vs 1/10), and 4q losses (6/8 vs 2/10). Altogether, these data suggest that the stem-cell subclass shares common molecular features with more aggressive and progenitor-like HCCs.

Classical subclass: HCC and iCCA components are derived from the same clone

The classical subclass showed enrichment of the poor prognosis iCCA subclass [29] and the chromosome 7 polysomy HCC subclass [24]. By performing separate analysis of the

classical iCCA and HCC components, we found enrichment of several gene sets in the iCCA component whereas none reached our pre-specified statistical threshold in the HCC component (**Figure 5A**, **Supplementary Table 3**). Moreover, the classical iCCA component seemed to have an intermediate molecular profile to stem-cell and CLC subclasses, as it showed association with liver derived gene signatures enriched in either stem-cell (i.e. HCC Proliferation, G3, S2, cluster A, CK19, IGF1R) or CLC (i.e. Late TGF-β induced signature). GSEA of canonical signaling pathways in the iCCA component (**Supplementary Table 8)** and HCC component (**Supplementary Table 9)** of the classical tumors further suggested a more aggressive phenotype for the iCCA component. Pathways enriched in the iCCA component included pro-mitotic DNA replication related signaling, proliferative signals such as MYC and mTOR, and pro-inflammatory pathways such as INF-ɣ and downstream IL2/STAT5 signaling (**Supplementary Table 8)**. These data indicate that the iCCA component of the classical subclass may have a more aggressive molecular profile within classical mixed tumors.

Clonality analysis based on CNV profiling of the HCC-like and iCCA-like components of the classical HCC-CCA showed a significant correlation and remarkable similarity (mean 51%, $P<0.001$) between both components in 3 out of 4 cases (**Supplementary Figure 7 and 8**, **Supplementary Table 10**). Specifically, focal chromosomal aberrations harboring known driver genes (i.e. 11q13 gain and 9p21 deletion), were identified in both components in 1 classical subclass case (**Supplementary Table 2, Supplementary Figure 8**). These findings suggest that the HCC and iCCA components of the classic subtype may share a common cell of origin that later undergoes clonal divergent expansion.

Landscape of mutations in mixed HCC-CCA

Next, we evaluated the mutational landscape of mixed HCC-CCA tumors (4 CLC and 2 stem-cell, including 3 overlapping cases with above analysis) using exome sequencing. Somatic substitutions were predominantly G>A and C>T transitions in both CLC and stem cell tumors (**Supplementary Figure 9A**), as previously reported in iCCA [30,31], HCC [32,33] and other cancers [34]. Globally, an average of 63 non-synonymous mutations (range 10-129) and 3 small insertions and deletions (indels, range: 1-5, **Supplementary Table 11**) were identified per mixed HCC-CCA tumor (**Supplementary Figure 9B**). Non-synonymous mutations included 93% missense (average 59 per tumor, range: 9-115) and 7% nonsense (average 4 per tumor, range: 1-14,) mutations. Overall, both CLC and stem-cell subclasses presented similar percentage of missense, silent, and nonsense mutations (**Figure 6A**). Among the non-synonymous mutations, 10 affected known cancer driver genes [35,36] (average of 2 driver mutations per tumor), which were confirmed by independent PCR and sequencing in each tumor (**Figure 6B**, **Supplementary table 12**).

We further explored the incidence of hot-spot mutations identified by exome sequencing in known oncogenic drivers (*BRAF*, *DNMT3A*, *IDH1*) together with recurrent mutations reported in HCC (*TERT* promoter, *CTNNB1*, *TP53*) and iCCA (*KRAS*, *FGFR2-BICC1*, *FGFR2-PPHLN1*, *IDH2*) in the remaining FFPE samples. *TP53* mutations emerged as the most frequent alteration (6/21, 29%), regardless of the subclass (**Figure 6C**). The mutational profile of CLC included *TP53* and *IDH1,* while the stem-cell subclass (i.e. *TERT* promoter, *TP53*, *AXIN1*, *BRAF*, *FGFR2-BICC1*) seemed to recapitulate those characteristic of both HCC and iCCA (**Figure 6B and 6C**). Interestingly, the most frequent HCC driver mutation, *TERT* promoter mutation, co-occurred with *TP53* mutation in 2 cases (1 classical case and 1 stem-cell SALL4 negative case). The co-occurrence of *TERT* promoter and *TP53* mutations (**Figure 6C**), in both the HCC and iCCA component

of the classical subclass case further supported the model of single cell of origin suggested by CNV profiling. Other screened mutations such as *CTNNB1, KRAS*, *FGFR2-PPHLN1* and *IDH2* were absent in our cohort. In summary, mutational profiling of mixed HCC-CCA revealed common drivers in typical HCC and iCCA tumors. However, the CLC subclass seems to have a distinct mutational profile in comparison to stem-cell subclass.

External validation confirms CLC as a distinct entity from stem-cell HCC-CCA

We validate the molecular subclasses in a publicly available dataset including 20 HCC-CCAs described with CLC features [13]. The CLC main molecular characteristics, such as biliary committed phenotype and activation of TGF-beta and immune-response signaling, were successfully reproduced. (**Supplementary Figure 10A**). In addition, a CLC-derived 156-gene signature (**Supplementary Table 13**) was significantly enriched in ~60% (11/20) of these CLC tumors. Subclass mapping approach further reinforced the notion that CLC tumors have a sole biliary-like phenotype, with no traits of HCC [37] (**Supplementary Figure 10B**). Moreover, stem-cell HCC-CCAs shared molecular traits of HCC but not of CLC tumors (**Supplementary Figure 10B**) suggesting a hepatocyte lineage. These results confirm that CLC and stem-cell mixed tumors represent distinct molecular entities.

DISCUSSION

Mixed HCC-CCAs are a heterogeneous group of primary liver cancers. The current histological classification describes two main subtypes, classical and with stem-cell features. Our study provides molecular evidence confirming a biphenotypical fingerprint for stem-cell and classical types, while defining CLC as a separate biliary-derived entity with no genomic features of HCC. We were able to characterize these differences based on integrated genomic analysis of gene expression, DNA copy number alteration, signaling pathway deregulation and mutational profiling (**Figure 7**).

CLC was characterized by chromosomal stability, up-regulation of TGF-beta signaling, prominent enrichment in immune-related pathways and defined biliary features. Histologically, CLC were characterized by being strongly embedded in fibrous stroma and by presenting positive NCAM immune-staining, as previously described [5,11,38,39]. Characteristically, TGF-β signaling was activated in CLC tumors as opposed to other mixed HCC-CCA. TGF-β is a pleiotropic cascade with different functions depending on the cellular context. It exerts pro-tumorigenic effects by enhancing tumor growth and invasion through the induction of EMT, activation of myofibroblasts and collagen deposition [40]. TGF-β has also been associated with the transformation of hepatic progenitor cells (HPCs) into tumor initiating cells [41] and biliary differentiation during early development [42]. From the signaling stand point, CLC tumors also showed up-regulation of EMT-related markers, down-regulation of hepatocyte specific gene markers and enrichment in immune cells (T cells, cytotoxic lymphocytes, Th1, NK) and immune-response related gene signatures. Particularly, the CXCL12-CXCR4 axis, which has been described to be instrumental for the chemoattraction of myeloid and lymphoid cells in to the tumor [43], was also shown to be significantly up-regulated in CLC. All these findings -coordinated Th1 cell and cytotoxic immune infiltration- are consistent with those reported in other

tumor types [44–46]. It is intriguing to understand how activation of TGF-β signaling, which is a known immune-suppressor, coexists with activation of T cells in the same tumor, a finding that requires further studies. In our view, all these specific molecular traits support the concept that CLC should stand alone as a molecularly differentiated subtype of intrahepatic biliary carcinomas.

Tumors belonging to the stem-cell subclass share completely distinct molecular traits. First, they showed enrichment for signatures defining activation of proliferative signals such as MYC, IGF2, mTOR and NOTCH and HCC poor prognosis. Histologically, they were characterized by the presence of SALL4 – an early progenitor-like marker- in 75% (6/8) of cases compared to none of other mixed subclasses. Of note, re-expression of the oncofetal SALL4 transcription factor has also been described in a subset of HCCs with progenitor features and poor prognosis [47].

The classical subtype represents a completely distinct entity, since it shares features from both typical HCC and iCCA. Our findings support a single cell of origin model for these tumors based on similarities in CNV aberrations in HCC and iCCA components in 3 out of 4 paired cases. These results are aligned with previous findings based on LOH analysis where 70% (8 out of 11) of the cases studied showed significant similarities [48]. Moreover, the presence of a characteristic transition area between the HCC and iCCA components with biphenotypic features [2], further supports the model of a single clonal process in which genetic divergence within the tumor parallels the histological diversity. From the molecular stand point, each component retained genomic and biomarker traits resembling either HCC or iCCA. Similarly, mutational profiling data showed common oncogenic driver mutations characteristic of either HCC or iCCA in classical HCC-CCA subclass. Finally, we did not identify WNT/β-catenin signaling activation or presence of

CTNNB1 mutations in our classical HCC-CCA samples, in agreement with previous studies [13,16].

After thoroughly exploring all genomic results, we can speculate on the diverse cellular lineage of this heterogeneous group of tumors (**Supplementary Figure 11**). Certainly, our results support the possible existence of multiple cells of origin [49]. Mixed HCC-CCA may share a common ancestor, the HPCs, but also might derive from more mature progenitor cells. Our gene expression profiling data suggested a biliary committed precursor in the CLC cases and a biphenotypic progenitor-like precursor in the stem-cell and classical subtypes. First, CLC characteristically express NCAM, and has a loss of hepatocyte markers only retaining cholangiocyte markers, both features suggesting a more mature biliary progenitor ancestor. In addition, IDH1 mutation, a known inhibitor of hepatocyte differentiation and gatekeeper for iCCA generation [19], was only detected in CLC. Conversely, the stem cell subclass might derive directly from HPCs, or from transdifferentiation of more mature ancestors. Specifically, those tumors were enriched with stem-cell signatures and SALL4 positive staining and, thus are the logical candidates to derive from HPC lineage. Finally, the classical subtype retained markers of both HCC and iCCA, and since it seems to have a clonal origin, the cells of origin should be mature enough to have lost early progenitor markers (SALL4), albeit retaining biphenotypical markers.

In summary, our study provides a comprehensive molecular characterization of mixed HCC-CCA. First, from the molecular standpoint both the stem-cell and classical types retain biliary and HCC components, and thus fit within the HCC-iCCA definition. Conversely, our data supports defining CLC as a distinct biliary-derived molecular entity with no HCC traits. These results provide the rationale for re-defining the current pathological classification [2], and for establishing more precise therapeutic approaches.

ACKNOWLEDGMENTS

We would like to thank the Biorepository Tissue Bank at the Icahn School of Medicine at Mount Sinai for providing the FF-OCT samples, and Dr. Mireia Castillo-Martin and Yayoi Kinoshita for their generous help in reviewing and processing the samples. We would also like to thank Dr. Yujin Hoshida for his support with gene expression data normalization and advice. 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.

REFERENCES

[1] Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, et al. Hepatocellular carcinoma. Nat Rev Dis Prim 2016;2:16018.

[2] Theise ND, Nakashima O, Park YN, Nakanuma Y. Combined hepatocellular-cholangiocarcinoma. WHO Classification of Tumours of the Digestive System, 4th, BosmanFT, Carneiro F, Hruban RH, Theise ND. (Eds), IARC, Lyons 2010. p. 225–7.

[3] Jarnagin WR, Weber S, Tickoo SK, Koea JB, Obiekwe S, Fong Y, et al. Combined hepatocellular and cholangiocarcinoma: demographic, clinical, and prognostic factors. Cancer 2002;94:2040–6.

[4] Wachtel MS, Zhang Y, Xu T, Chiriva-Internati M, Frezza EE. Combined hepatocellular cholangiocarcinomas; analysis of a large database. Clin Med Pathol 2008;1:43–7.

[5] Sempoux C, Fan C, Singh P, Obeidat K, Roayaie S, Schwartz M, et al. Cholangiolocellular carcinoma: an innocent-looking malignant liver tumor mimicking ductular reaction. Semin Liver Dis 2011;31:104–10.

[6] Lee J-H, Chung GE, Yu SJ, Hwang SY, Kim JS, Kim HY, et al. Long-term prognosis of combined hepatocellular and cholangiocarcinoma after curative resection comparison with hepatocellular carcinoma and cholangiocarcinoma. J Clin Gastroenterol 2011;45:69–75.

[7] Yin X, Zhang B-H, Qiu S-J, Ren Z-G, Zhou J, Chen X-H, et al. Combined hepatocellular carcinoma and cholangiocarcinoma: clinical features, treatment modalities, and prognosis. Ann Surg Oncol 2012;19:2869–76.

[8] Lee W-S, Lee K-W, Heo J-S, Kim S-J, Choi S-H, Kim Y-I, et al. Comparison of combined hepatocellular and cholangiocarcinoma with hepatocellular carcinoma and intrahepatic cholangiocarcinoma. Surg Today 2006;36:892–7.

[9] Sasaki M, Sato H, Kakuda Y, Sato Y, Choi JH, Nakanuma Y. Clinicopathological significance of “subtypes with stem-cell feature” in combined hepatocellular-cholangiocarcinoma. Liver Int 2015;35:1024–35.

[10] Ikeda H, Harada K, Sato Y, Sasaki M, Yoneda N, Kitamura S, et al. Clinicopathologic significance of combined hepatocellular-cholangiocarcinoma with stem cell subtype components with reference to the expression of putative stem cell

markers. Am J Clin Pathol 2013;140:329–40.

[11] Komuta M, Spee B, Vander Borght S, De Vos R, Verslype C, Aerts R, et al. Clinicopathological study on cholangiolocellular carcinoma suggesting hepatic progenitor cell origin. Hepatology 2008;47:1544–56.

[12] Oishi N, Kumar MR, Roessler S, Ji J, Forgues M, Budhu A, et al. Transcriptomic profiling reveals hepatic stem-like gene signatures and interplay of mir-200c and EMT in intrahepatic cholangiocarcinoma. Hepatology 2012;56:1792–803.

[13] Coulouarn C, Cavard C, Rubbia-Brandt L, Audebourg A, Dumont F, Jacques S, et al. Combined hepatocellular-cholangiocarcinomas exhibit progenitor features and activation of Wnt and TGFβ signaling pathways. Carcinogenesis 2012;33:1791–6.

[14] Fujimoto A, Furuta M, Shiraishi Y, Gotoh K, Kawakami Y, Arihiro K, et al. Whole-genome mutational landscape of liver cancers displaying biliary phenotype reveals hepatitis impact and molecular diversity. Nat Commun 2015;6:6120.

[15] Woo HG, Lee JH, Yoon JH, Kim CY, Lee HS, Jang JJ, et al. Identification of a cholangiocarcinoma-like gene expression trait in hepatocellular carcinoma. Cancer Res 2010;70:3034–41.

[16] Cazals-Hatem D, Rebouissou S, Bioulac-Sage P, Bluteau O, Blanché H, Franco D, et al. Clinical and molecular analysis of combined hepatocellular-cholangiocarcinomas. J Hepatol 2004;41:292–8.

[17] Villanueva A, Hoshida Y, Battiston C, Tovar V, Sia D, Alsinet C, et al. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. Gastroenterology 2011;140:1501–12.e2.

[18] Sia D, Hoshida Y, Villanueva A, Roayaie S, Ferrer J, Tabak B, et al. Integrative Molecular Analysis of Intrahepatic Cholangiocarcinoma Reveals 2 Classes That Have Different Outcomes. Gastroenterology 2013;144:829–40.

[19] Saha SK, Parachoniak CA, Ghanta KS, Fitamant J, Ross KN, Najem MS, et al. Mutant IDH inhibits HNF-4α to block hepatocyte differentiation and promote biliary cancer. Nature 2014;513:110–4.

[20] Hoshida Y, Nijman SMB, Kobayashi M, Chan JA, Brunet J-PP, Chiang DY, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res 2009;69:7385–92.

[21] Coulouarn C, Factor VM, Thorgeirsson SS. Transforming growth factor-beta gene expression signature in mouse hepatocytes predicts clinical outcome in human cancer. Hepatology 2008;47:2059–67.

[22] Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity 2013;39:782–95.

[23] Yamashita T, Forgues M, Wang W, Kim JW, Ye Q, Jia H, et al. EpCAM and alpha-fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. Cancer Res 2008;68:1451–61.

[24] Chiang DYY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res 2008;68:6779–88.

[25] Boyault S, Rickman DS, de Reyniès A, Balabaud C, Rebouissou S, Jeannot E, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology 2007;45:42–52.

[26] Lee J-S, Chu I-S, Heo J, Calvisi DF, Sun Z, Roskams T, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology 2004;40:667–76.

[27] Tovar V, Alsinet C, Villanueva A, Hoshida Y, Chiang DYY, Solé M, et al. IGF activation in a molecular subclass of hepatocellular carcinoma and pre-clinical efficacy of IGF-1R blockage. J Hepatol 2010;52:550–9.

[28] Villanueva A, Alsinet C, Yanger K, Hoshida Y, Zong Y, Toffanin S, et al. Notch signaling is activated in human hepatocellular carcinoma and induces tumor formation in mice. Gastroenterology 2012;143:1660–9 e7.

[29] Andersen JB, Spee B, Blechacz BR, Avital I, Komuta M, Barbour A, et al. Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. Gastroenterology 2012;142:1021–31 e15.

[30] Sia D, Losic B, Moeini A, Cabellos L, Hao K, Revill K, et al. Massive parallel sequencing uncovers actionable FGFR2-PPHLN1 fusion and ARAF mutations in intrahepatic cholangiocarcinoma. Nat Commun 2015;6:6087.

[31] Ong CK, Subimerb C, Pairojkul C, Wongkham S, Cutcutache I, Yu W, et al. Exome

sequencing of liver fluke-associated cholangiocarcinoma. Nat Genet 2012;44:690–3.

[32] Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 2015;47:505–11.

[33] Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet 2014;46:1267–73.

[34] Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. Nature 2007;446:153–8.

[35] Garraway LA, Lander ES. Lessons from the cancer genome. Cell 2013;153:17–37.

[36] Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science 2013;339:1546–58.

[37] Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: identifying common subtypes in independent disease data sets. PLoS One 2007;2:e1195.

[38] Komuta M, Govaere O, Vandecaveye V, Akiba J, Van Steenbergen W, Verslype C, et al. Histological diversity in cholangiocellular carcinoma reflects the different cholangiocyte phenotypes. Hepatology 2012;55:1876–88.

[39] Akiba J, Nakashima O, Hattori S, Tanikawa K, Takenaka M, Nakayama M, et al. Clinicopathologic analysis of combined hepatocellular-cholangiocarcinoma according to the latest WHO classification. Am J Surg Pathol 2013;37:496–505.

[40] Massagué J. TGFbeta in Cancer. Cell 2008;134:215–30.

[41] Wu K, Ding J, Chen C, Sun W, Ning B-F, Wen W, et al. Hepatic transforming growth factor beta gives rise to tumor-initiating cells and promotes liver cancer development. Hepatology 2012;56:2255–67.

[42] Karkampouna S, Ten Dijke P, Dooley S, Julio MK. TGFβ signaling in liver regeneration. Curr Pharm Des 2012;18:4103–13.

[43] Chen Y, Huang Y, Reiberger T, Duyverman AM, Huang P, Samuel R, et al. Differential effects of sorafenib on liver versus tumor fibrosis mediated by stromal-

derived factor 1 alpha/C-X-C receptor type 4 axis and myeloid differentiation antigen-positive myeloid cell infiltration in mice. Hepatology 2014;59:1435–47.

[44] Chew V, Chen J, Lee D, Loh E, Lee J, Lim KH, et al. Chemokine-driven lymphocyte infiltration: an early intratumoural event determining long-term survival in resectable hepatocellular carcinoma. Gut 2012;61:427–38.

[45] Gao Q, Qiu S-J, Fan J, Zhou J, Wang X-Y, Xiao Y-S, et al. Intratumoral balance of regulatory and cytotoxic T cells is associated with prognosis of hepatocellular carcinoma after resection. J Clin Oncol 2007;25:2586–93.

[46] Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 2006;313:1960–4.

[47] Yong KJ, Gao C, Lim JSJ, Yan B, Yang H, Dimitrov T, et al. Oncofetal gene SALL4 in aggressive hepatocellular carcinoma. N Engl J Med 2013;368:2266–76.

[48] Fujii H, Zhu XG, Matsumoto T, Inagaki M, Tokusashi Y, Miyokawa N, et al. Genetic classification of combined hepatocellular-cholangiocarcinoma. Hum Pathol 2000;31:1011–7.

[49] Moeini A, Sia D, Bardeesy N, Mazzaferro V, Llovet JM. Molecular Pathogenesis and Targeted Therapies for Intrahepatic Cholangiocarcinoma. Clin Cancer Res 2016;22:291–300.

TABLES

Table 1: Clinico-pathological characteristics of mixed HCC-CCA cohort according to mixed molecular subclass.

Variable	Total	CLC	Stem-cell	Classical	p-value
Patients, n	18	6	8	4	
Gender, n					0.376
Male	14	5	7	2	
Female	4	1	1	2	
Age, years					0.976
Median (range)	55 (15-82)	56 (15-82)	57 (33-71)	54 (49-66)	
Etiology, n					0.599
Hepatitis C	8	4	3	1	
Hepatitis B	7	1	4	2	
PSC	1	0	0	1	
Others	1	1	0	0	
None	1	0	1	0	
Cirrhosis, n					0.235
Absent	8	1	5	2	
Present	10	5	3	2	
Tumor size, cm					0.225
Median (range)	3.2 (0.5-13.5)	1.5 (0.5-8)	3.1 (1.0-13.5)	6.3 (2.5-9)	
Satellites, n					0.235
Absent	10	5	3	2	
Present	8	1	5	2	
Microvascular invasion, n					0.559
Absent	3	2	1	0	
Present	15	4	7	4	
Macrovascular invasion, n					0.275
Absent	15	4	8	3	
Present	3	2	0	1	
AFP, ng/mL[1]					0.953
Median (range)	19 (1-1573)	33 (1-1573)	18 (4-842)	114 (5-223)	
CA 19-9, UI/mL[1]					0.461
Median (range)	186 (8-945)	124 (8-318)	244 (8-945)	440 (20-472)	
Albumin, g/dL[2]					***0.049***
Median (range)	4.1 (3.3-4.8)	4.0 (3.7-4.3)	4.5 (3.3-4.8)	4.0 (3.4-4.1)	
Bilirrubin, mg/dL[2]					0.104
Median (range)	0.6 (0.3-3.4)	0.9 (0.6-1.3)	0.5 (0.3-3.4)	0.5 (0.4-0.6)	

CLC: cholangiolocellular carcinoma. Stem cell subclass includes the histological stem cell feature typical and intermediate subtypes.

[1] Not available in 5 patients

[2] Not available in 1 patient

p-value corresponds to statistical analysis of the 3 molecular subclasses. Fisher Exact Probability Test was used for categorical variables and The Kruskal-Wallis Test for the continuous factors.

FIGURE LEGENDS

Figure 1: Molecular classes of mixed HCC-CCA and correlation with histopathological features. A) Non-negative matrix factorization–based clustering analysis of 22 tumor samples from 18 patients with mixed HCC-CCA. **B)** Heat-map representing expression of specific cell lineage gene markers (upper panel) and immunostainig profiling data (lower panel) of HCC-CCA tumors. The iCCA and HCC components of classical subclass were analyzed separately.

Figure 2: Integrative genomic analysis of HCC-CCA with HCC and iCCA. A) Unsupervised hierarchal clustering analysis was performed on merged gene expression profile cohorts of HCC-CCA (6 CLCs and 8 stem-cell subclass), HCC (n=164) and iCCA (n=149). Integration of the datasets was based on the Z-score transformation of the differentially expressed gene in each independent cohort. Nearest prediction method was used for the association with liver cancer derived gene signatures. **B)** Schematic representation of overlapping molecular profile and genomic proximity of CLC and stem-cell subclass mixed tumors with iCCA and HCC, respectively, based on unsupervised clustering analysis. Statistically significant association was calculated by Fisher Exact Probability Test ($p<0.05$).

Figure 3: Immunostaining profile of mixed HCC-CCA subclasses. Representative morphological and positive immunohistochemical staining features observed for each mixed HCC-CCA subclass.

Figure 4: Broad chromosomal alterations detected in mixed HCC-CCA subclasses. Chromosome arms are displayed in descending order along the vertical axis. Detected broad chromosomal gains, losses, and copy neutral loss of heterozygosity (CN-LOH) per tumor sample have been highlighted. The iCCA and HCC components of each classical

case were analyzed separately. Bars indicate the total number of broad chromosomal gains and losses. CLC showed higher chromosomal stability in comparison to other mixed HCC-CCAs ($p<0.01$, two-sided T-test).

Figure 5: Whole-genome gene expression analysis of different mixed HCC-CCA subclasses. A) Heat-map representing prediction of liver cancer derived molecular classification and gene signatures (upper panel), immune-related gene signatures for cell type and activated signaling (lower panel). **B)** Network analysis of deregulated genes in CLC showed TGF-β signaling as one of the main activated signaling nodes. **C)** Network analysis of deregulated genes in SALL4 stem-cell subclass showed activation of IGF2 signaling and hepatic specification of progenitor cells. In the network analyses, a node symbolizes a gene or gene product, and direct and indirect interactions are indicated by *solid lines* and *dotted lines*, respectively. Statically significant associated features in CLC are highlighted with (*) and in stem-cell subclass with (¥) using Fisher Exact Probability Test for categorical variables and two-sided T-test for continuous variables. *$p<0.05$, **$p<0.01$, ***$p<0.001$. The iCCA and HCC components of each classical case were analyzed separately.

Figure 6: Mutational profile of mixed HCC-CCA. A) Histogram of the number of mutations in each primary tumor sample (upper panel) and pie chart representation of the percentage of non-synonymous somatic mutations for CLC and stem-cell subclasses (lower panel). **B)** Heat-map representing the individual mutation in known cancer driver genes identified by exome-sequencing. **C)** Heat-map representing the validation of exome-sequencing results and screening of the most prevalent oncogenic mutations reported in HCC and iCCA in the study cohort. Overlapping cases from which fresh-frozen

and FFPE samples were analyzed separately by exome-sequencing and PCR validation are highlighted with an asterisk.

Figure 7: Summary of molecular characterization of mixed HCC-CCAs and CLC tumors as distinct entities. CLC only share biliary-derived features, as opposed to HCC-iCCA tumors. Specific cell lineage markers, liver cancer derived gene signatures, pathway signaling, chromosomal stability and driver mutations are depicted for each entity.

Figure 1

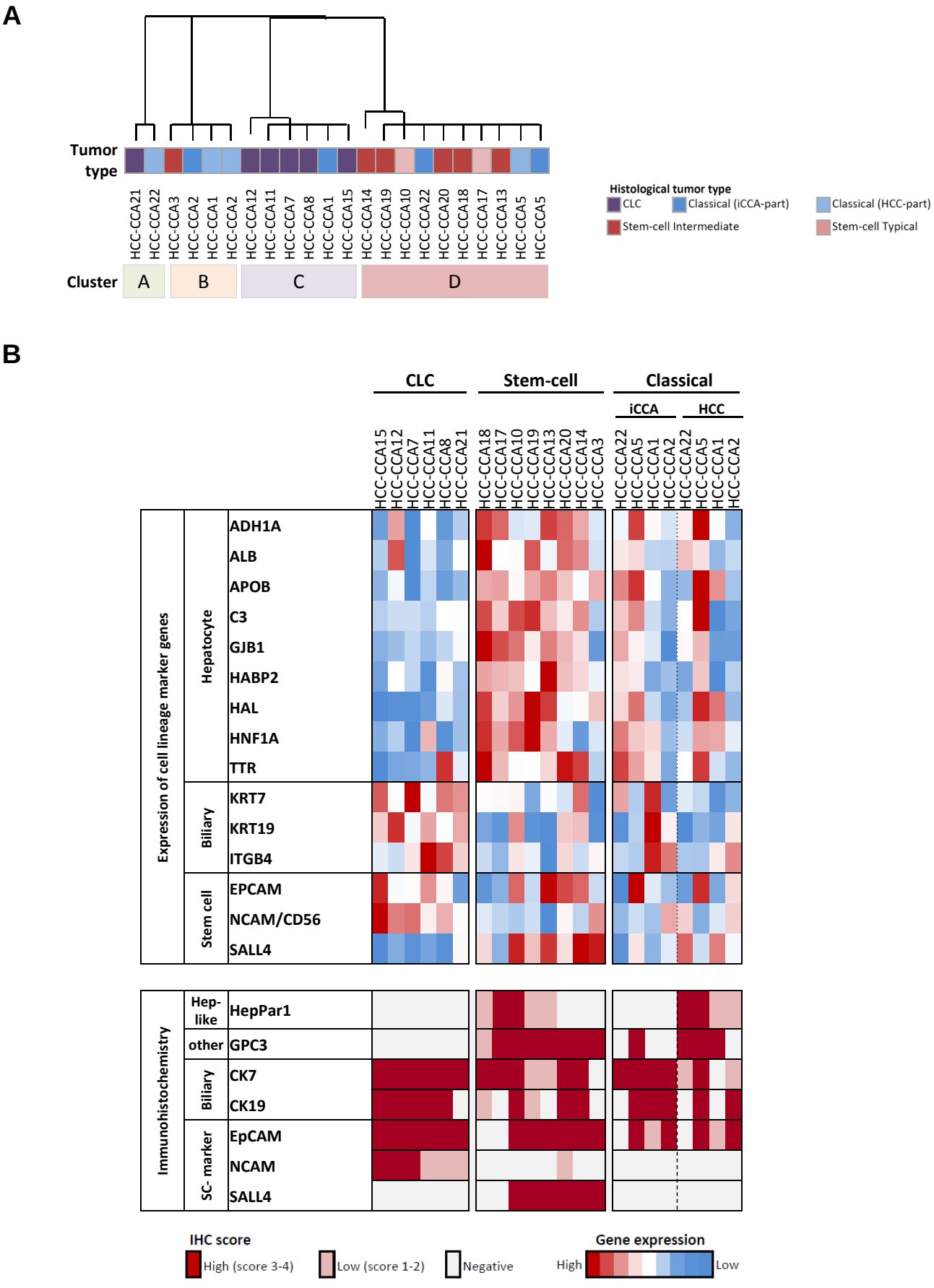

A
Tumor type
HCC-CCA21
HCC-CCA22
HCC-CCA3
HCC-CCA2
HCC-CCA1
HCC-CCA2
HCC-CCA12
HCC-CCA11
HCC-CCA7
HCC-CCA8
HCC-CCA1
HCC-CCA15
HCC-CCA14
HCC-CCA19
HCC-CCA10
HCC-CCA22
HCC-CCA20
HCC-CCA18
HCC-CCA17
HCC-CCA13
HCC-CCA5
HCC-CCA5
Cluster
A
B
C
D
Histological tumor type
CLC
Classical (iCCA-part)
Classical (HCC-part)
Stem-cell Intermediate
Stem-cell Typical
B
CLC
Stem-cell
Classical
iCCA
HCC
HCC-CCA15
HCC-CCA12
HCC-CCA7
HCC-CCA11
HCC-CCA8
HCC-CCA21
HCC-CCA18
HCC-CCA17
HCC-CCA10
HCC-CCA19
HCC-CCA13
HCC-CCA20
HCC-CCA14
HCC-CCA3
HCC-CCA22
HCC-CCA5
HCC-CCA1
HCC-CCA2
HCC-CCA22
HCC-CCA5
HCC-CCA1
HCC-CCA2
Expression of cell lineage marker genes
Hepatocyte
ADH1A
ALB
APOB
C3
GJB1
HABP2
HAL
HNF1A
TTR
Biliary
KRT7
KRT19
ITGB4
Stem cell
EPCAM
NCAM/CD56
SALL4
Immunohistochemistry
Hep-like
HepPar1
other
GPC3
Biliary
CK7
CK19
SC- marker
EpCAM
NCAM
SALL4
IHC score
High (score 3-4)
Low (score 1-2)
Negative
Gene expression
High
Low

Figure 2

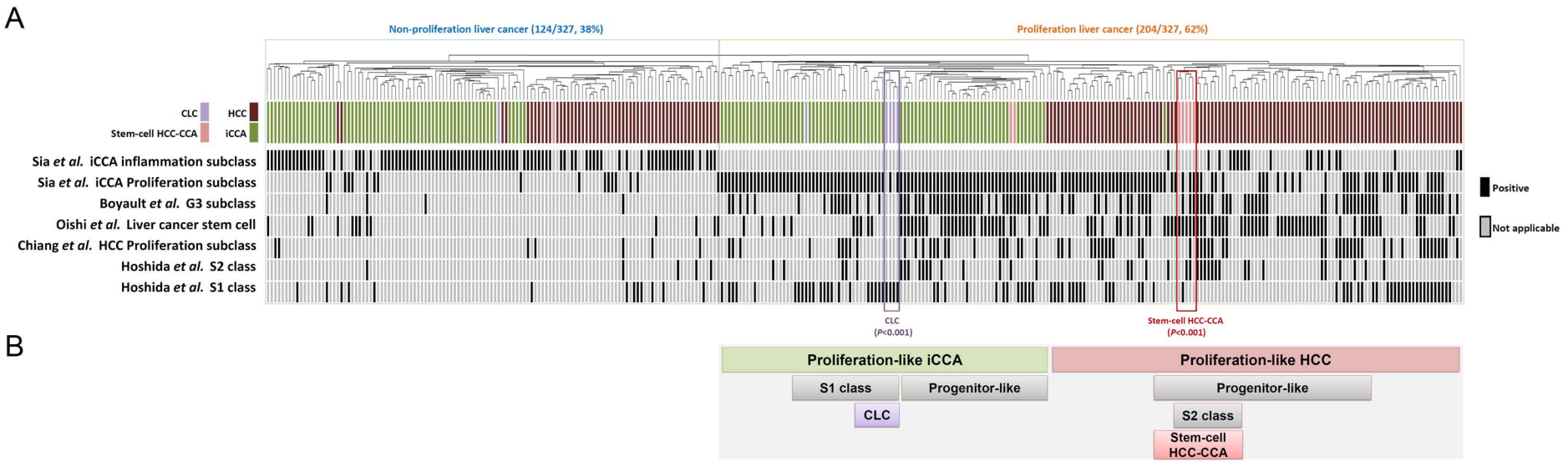

A
Non-proliferation liver cancer (124/327, 38%)
Proliferation liver cancer (204/327, 62%)
CLC
HCC
Stem-cell HCC-CCA
iCCA
Sia et al. iCCA inflammation subclass
Sia et al. iCCA Proliferation subclass
Boyault et al. G3 subclass
Oishi et al. Liver cancer stem cell
Chiang et al. HCC Proliferation subclass
Hoshida et al. S2 class
Hoshida et al. S1 class
Positive
Not applicable
CLC (P<0.001)
Stem-cell HCC-CCA (P<0.001)
B
Proliferation-like iCCA
Proliferation-like HCC
S1 class
Progenitor-like
Progenitor-like
CLC
S2 class
Stem-cell HCC-CCA

Figure 3

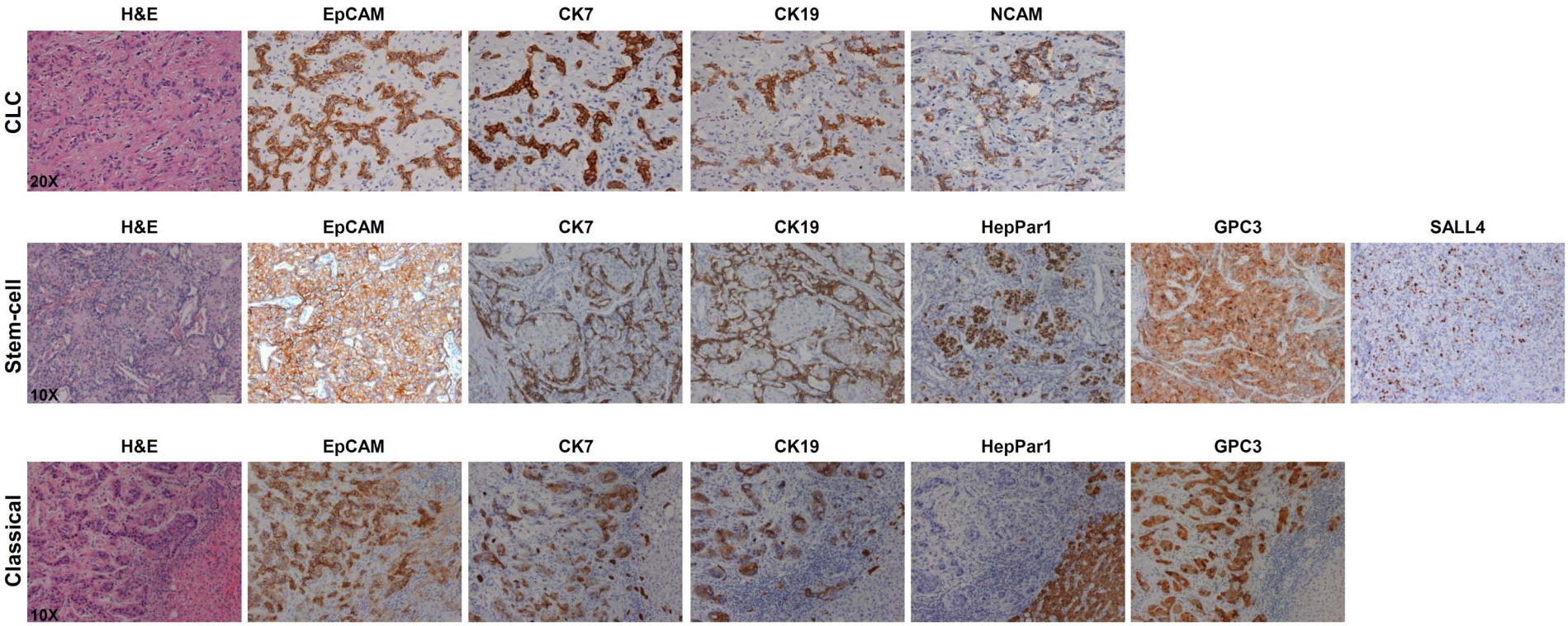

H&E
EpCAM
CK7
CK19
NCAM
CLC
20X
H&E
EpCAM
CK7
CK19
HepPar1
GPC3
SALL4
Stem-cell
10X
H&E
EpCAM
CK7
CK19
HepPar1
GPC3
Classical
10X

Figure 4

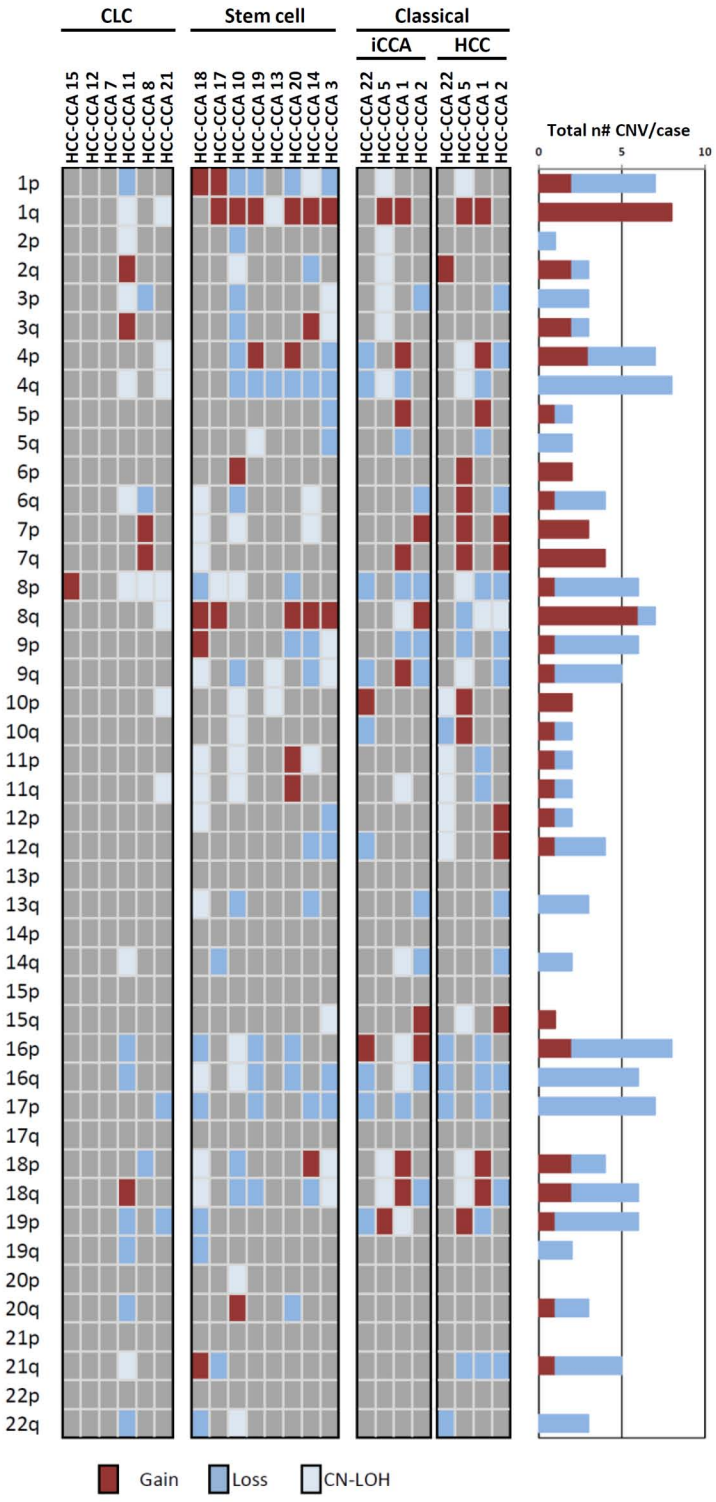
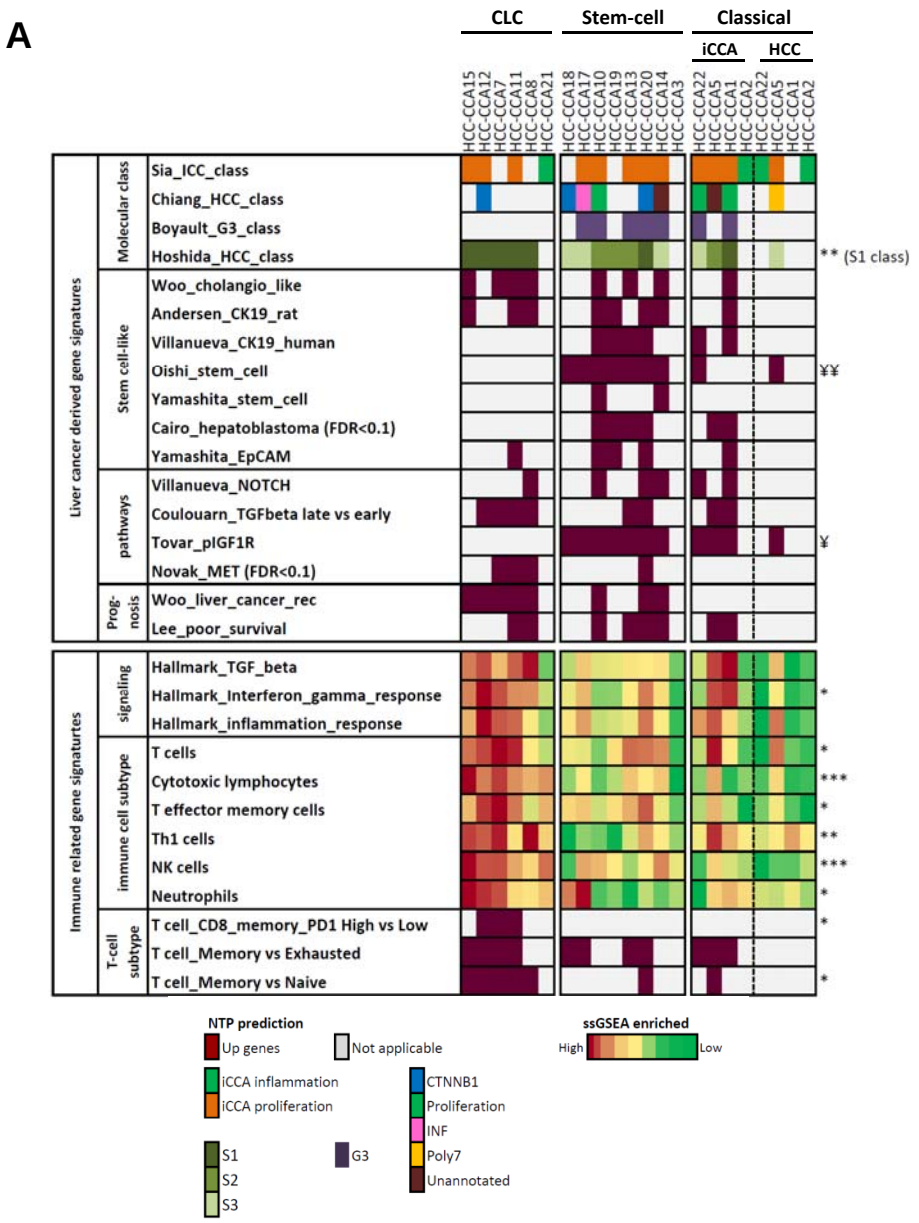
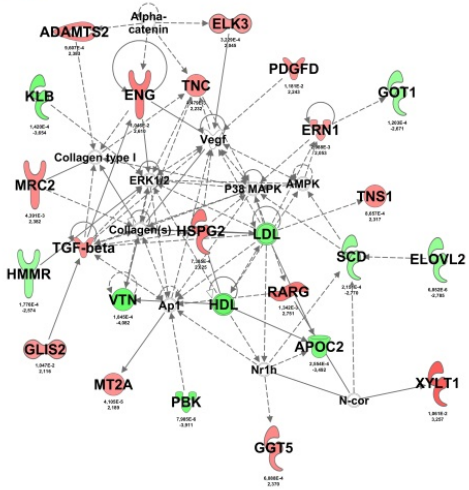

CLC
Stem cell
Classical
iCCA
HCC
HCC-CCA 15
HCC-CCA 12
HCC-CCA 7
HCC-CCA 11
HCC-CCA 8
HCC-CCA 21
HCC-CCA 18
HCC-CCA 17
HCC-CCA 10
HCC-CCA 19
HCC-CCA 13
HCC-CCA 20
HCC-CCA 14
HCC-CCA 3
HCC-CCA 22
HCC-CCA 5
HCC-CCA 1
HCC-CCA 2
HCC-CCA 22
HCC-CCA 5
HCC-CCA 1
HCC-CCA 2
Total n# CNV/case
0
5
10
1p
1q
2p
2q
3p
3q
4p
4q
5p
5q
6p
6q
7p
7q
8p
8q
9p
9q
10p
10q
11p
11q
12p
12q
13p
13q
14p
14q
15p
15q
16p
16q
17p
17q
18p
18q
19p
19q
20p
20q
21p
21q
22p
22q
Gain
Loss
CN-LOH

Figure 5


A
CLC
Stem-cell
Classical
iCCA
HCC
HCC-CCA15
HCC-CCA12
HCC-CCA7
HCC-CCA11
HCC-CCA8
HCC-CCA21
HCC-CCA18
HCC-CCA17
HCC-CCA10
HCC-CCA19
HCC-CCA13
HCC-CCA20
HCC-CCA14
HCC-CCA3
HCC-CCA22
HCC-CCA5
HCC-CCA1
HCC-CCA2
HCC-CCA22
HCC-CCA5
HCC-CCA1
HCC-CCA2
Liver cancer derived gene signatures
Molecular class
Sia_ICC_class
Chiang_HCC_class
Boyault_G3_class
Hoshida_HCC_class
** (S1 class)
Stem cell-like
Woo_cholangio_like
Andersen_CK19_rat
Villanueva_CK19_human
Oishi_stem_cell
¥¥
Yamashita_stem_cell
Cairo_hepatoblastoma (FDR<0.1)
Yamashita_EpCAM
pathways
Villanueva_NOTCH
Coulouarn_TGFbeta late vs early
Tovar_pIGF1R
¥
Novak_MET (FDR<0.1)
Prog-nosis
Woo_liver_cancer_rec
Lee_poor_survival
Immune related gene signatures
signaling
Hallmark_TGF_beta
Hallmark_Interferon_gamma_response
*
Hallmark_inflammation_response
immune cell subtype
T cells
*
Cytotoxic lymphocytes
***
T effector memory cells
*
Th1 cells
**
NK cells
***
Neutrophils
*
T-cell subtype
T cell_CD8_memory_PD1 High vs Low
*
T cell_Memory vs Exhausted
T cell_Memory vs Naive
*
NTP prediction
Up genes
Not applicable
iCCA inflammation
iCCA proliferation
S1
S2
S3
G3
CTNNB1
Proliferation
INF
Poly7
Unannotated
ssGSEA enriched
High
Low


B
ADAMTS2
Alpha-catenin
ELK3
PDGFD
KLB
ENG
TNC
GOT1
ERN1
Vegf
Collagen type I
MRC2
ERK1/2
P38 MAPK
AMPK
TNS1
Collagen(s)
HSPG2
LDL
TGF-beta
SCD
ELOVL2
HMMR
VTN
Ap1
HDL
RARG
GLIS2
APOC2
Nr1h
MT2A
PBK
N-cor
XYLT1
GGT5

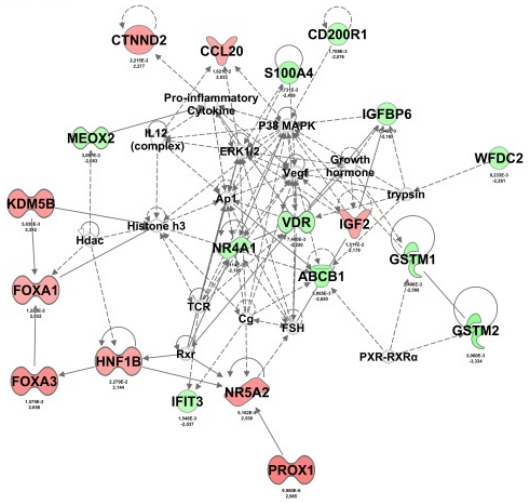

C
CTNND2
CCL20
CD200R1
S100A4
Pro-inflammatory Cytokine
IGFBP6
MEOX2
IL12 (complex)
P38 MAPK
ERK1/2
Growth hormone
Vegf
WFDC2
Ap1
trypsin
KDM5B
VDR
IGF2
Hdac
Histone h3
NR4A1
GSTM1
FOXA1
ABCB1
TCR
Cg
FSH
GSTM2
PXR-RXRα
FOXA3
HNF1B
Rxr
NR5A2
IFIT3
PROX1

Figure 6

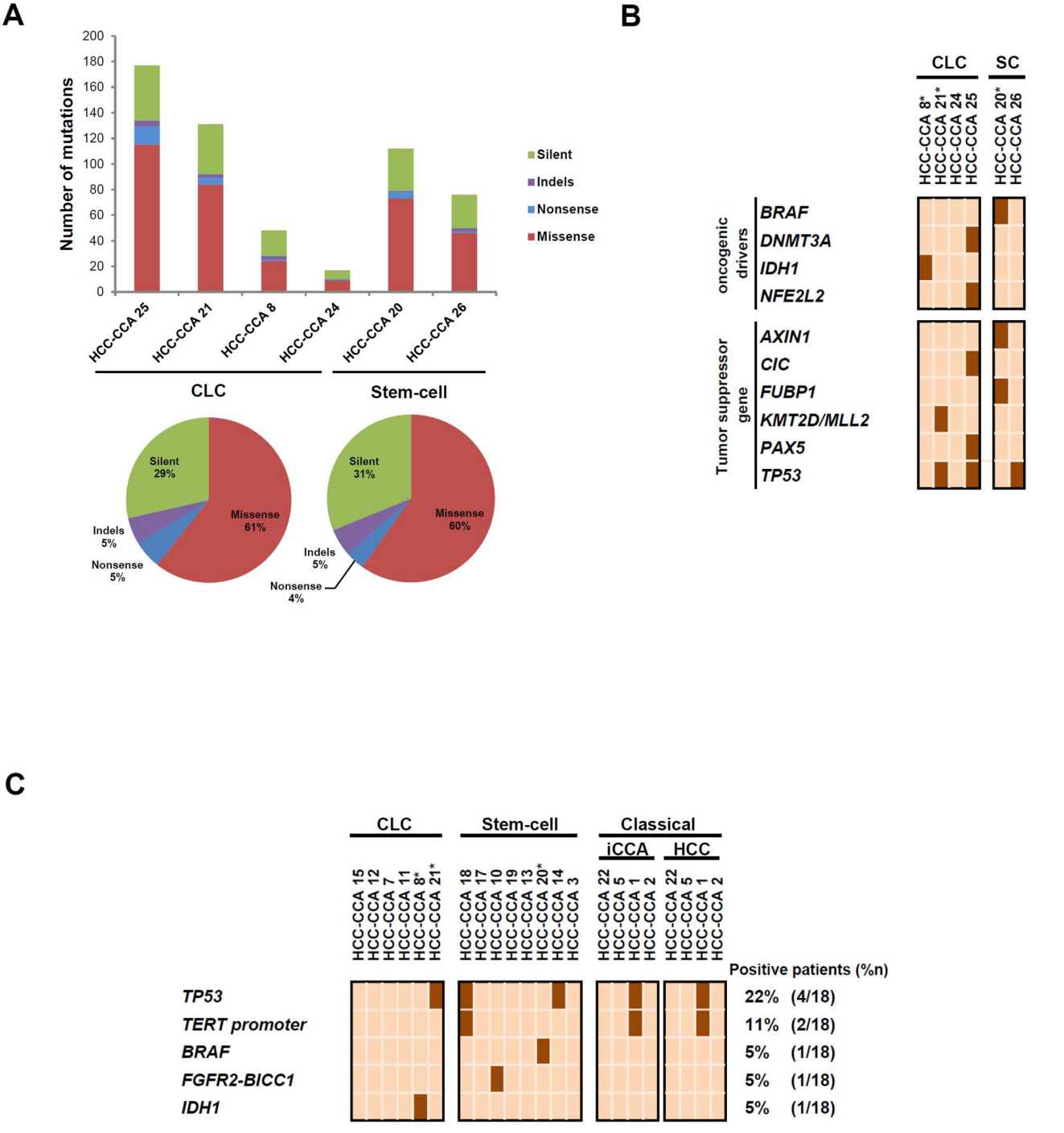
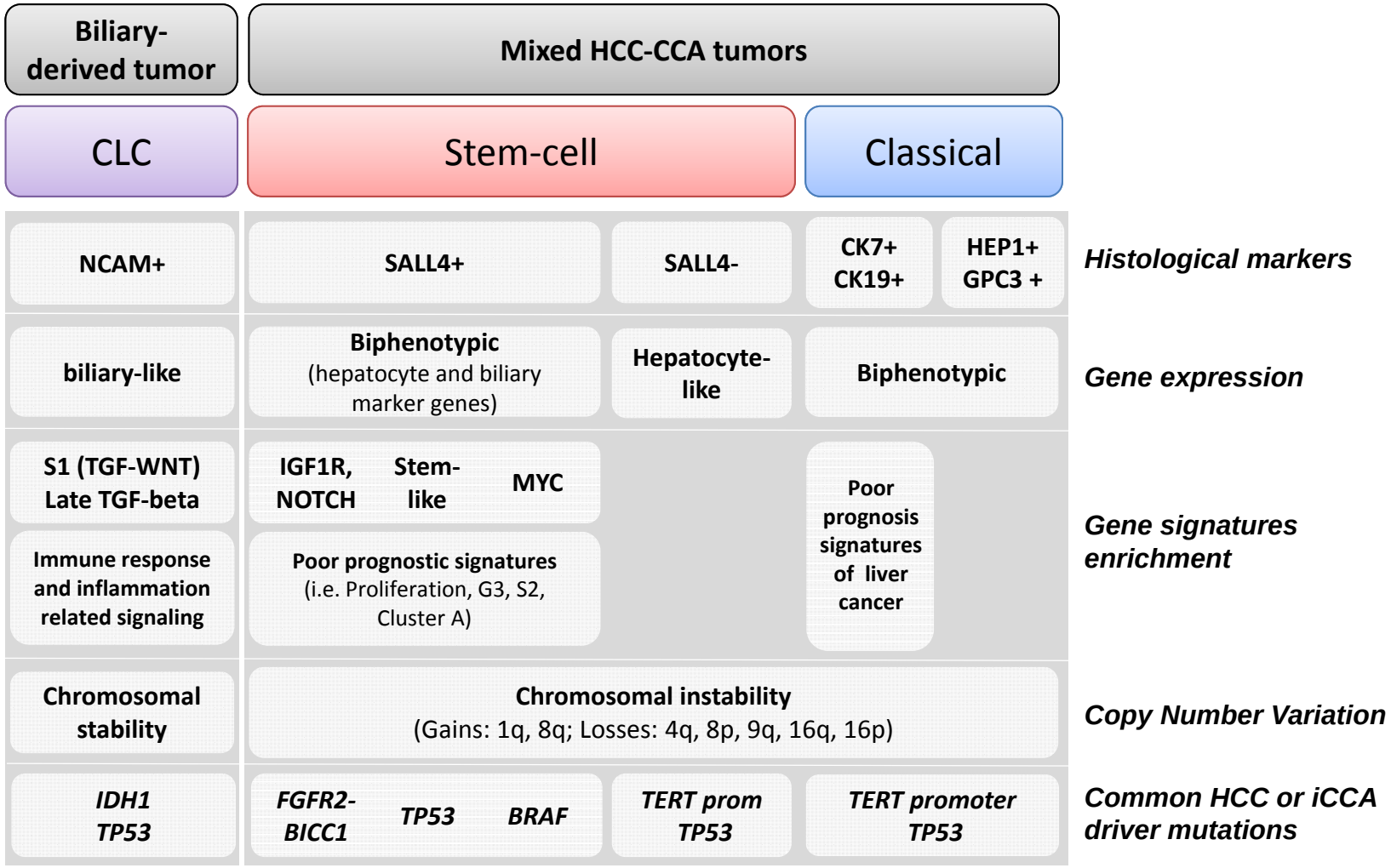

A
Number of mutations
200
180
160
140
120
100
80
60
40
20
0
Silent
Indels
Nonsense
Missense
HCC-CCA 25
HCC-CCA 21
HCC-CCA 8
HCC-CCA 24
HCC-CCA 20
HCC-CCA 26
CLC
Stem-cell
Silent 29%
Missense 61%
Indels 5%
Nonsense 5%
Silent 31%
Missense 60%
Indels 5%
Nonsense 4%
B
CLC
SC
HCC-CCA 8*
HCC-CCA 21*
HCC-CCA 24
HCC-CCA 25
HCC-CCA 20*
HCC-CCA 26
oncogenic drivers
BRAF
DNMT3A
IDH1
NFE2L2
Tumor suppressor gene
AXIN1
CIC
FUBP1
KMT2D/MLL2
PAX5
TP53
C
CLC
Stem-cell
Classical
iCCA
HCC
HCC-CCA 15
HCC-CCA 12
HCC-CCA 7
HCC-CCA 11
HCC-CCA 8*
HCC-CCA 21*
HCC-CCA 18
HCC-CCA 17
HCC-CCA 10
HCC-CCA 19
HCC-CCA 13
HCC-CCA 20*
HCC-CCA 14
HCC-CCA 3
HCC-CCA 22
HCC-CCA 5
HCC-CCA 1
HCC-CCA 2
HCC-CCA 22
HCC-CCA 5
HCC-CCA 1
HCC-CCA 2
Positive patients (%n)
TP53
22% (4/18)
TERT promoter
11% (2/18)
BRAF
5% (1/18)
FGFR2-BICC1
5% (1/18)
IDH1
5% (1/18)

Figure 7


Biliary-derived tumor
Mixed HCC-CCA tumors
CLC
Stem-cell
Classical
NCAM+
SALL4+
SALL4-
CK7+ CK19+
HEP1+ GPC3 +
Histological markers
biliary-like
Biphenotypic (hepatocyte and biliary marker genes)
Hepatocyte-like
Biphenotypic
Gene expression
S1 (TGF-WNT) Late TGF-beta
IGF1R, NOTCH
Stem-like
MYC
Immune response and inflammation related signaling
Poor prognostic signatures (i.e. Proliferation, G3, S2, Cluster A)
Poor prognosis signatures of liver cancer
Gene signatures enrichment
Chromosomal stability
Chromosomal instability (Gains: 1q, 8q; Losses: 4q, 8p, 9q, 16q, 16p)
Copy Number Variation
IDH1 TP53
FGFR2-BICC1
TP53
BRAF
TERT prom TP53
TERT promoter TP53
Common HCC or iCCA driver mutations