

Human Hepatocellular Carcinomas with a Periportal Phenotype Have the Lowest Potential for Early Recurrence after Curative Resection

Romain Désert,¹ Florian Rohart,² Frédéric Canal,³ Marie Sicard,¹  Mireille Desille,¹ Stéphanie Renaud,¹ Bruno Turlin,¹ Pascale Bellaud,¹ Christine Perret,³ Bruno Clément,¹ Kim-Anh Lê Cao,² Orlando Musso^{1,#}

¹INSERM, INRA, Univ Rennes 1, Univ Bretagne Loire, CHU Pontchaillou, Nutrition Metabolisms and Cancer, CRB-Santé, Biosit, Biogenouest, Rennes, France.

²The University of Queensland, Diamantina Institute, The University of Queensland, Translational Research Institute, Brisbane, QLD, Australia;

³INSERM, CNRS, Univ Paris Descartes, Sorbonne Paris Cité, Institut Cochin, Paris, France.

Running head: *Low Early Recurrence of HCCs with Periportal Phenotype*

E-mail addresses:

RD, rdesert87@gmail.com

FR, f.rohart@uq.edu.au

FC, frederic.canal94@gmail.com

MS, sicard.marie.i@gmail.com

MD, mireille.desille-dugast@chu-rennes.fr

SR, stephanie.renaud39@gmail.com

BT, bruno.turlin@chu-rennes.fr

PB, pascale.bellaud@univ-rennes1.fr

CP, christine.perret@inserm.fr

BC, bruno.clement@inserm.fr

K-A LC, k.lecao@uq.edu.au

OM, orlando.musso@inserm.fr

Keywords: Wnt/ β -catenin; liver zonation; HNF4A; sparse multivariate discriminant analysis; HCC classification.

Contact information:

#Corresponding author

Phone: 33-(0)2 23 23 45 65

Fax: 33-(0)2 99 54 01 37

E-mail: orlando.musso@inserm.fr

List of abbreviations: APC, adenomatous polyposis coli; AUROC, area under the receiver operating curve; BIO, 6-bromoindirubin-3'-oxime; DF, disease-free survival; ECM, extracellular matrix; GSEA, gene set enrichment analysis; HAL, histidine ammonia lyase; HCC, hepatocellular carcinoma; ODAM, odontogenic ameloblast-associated protein; PCA, principal component analysis; PP, periportal; PV, perivenous; sPLS-DA, sparse partial least squares discriminant analysis; TCF, T-cell factor; VNN1, vanin 1.

Title: 115 characters,

Abstract: 275 words,

Manuscript, 5842 words,

References: 49,

Figures: 7,

Tables: 1

Supporting Figures: 15,

Supporting Tables: 12,

Pages (main manuscript): 36.

ABSTRACT

Hepatocellular carcinomas (HCCs) exhibit a diversity of molecular phenotypes, raising major challenges in clinical management. HCCs detected by surveillance programs at an early stage are candidates for potentially curative therapies (local ablation, resection or transplantation). In the long term, transplantation provides the lowest recurrence rates. Treatment allocation is based on tumor number, size, vascular invasion, performance status, functional liver reserve and on the prediction of early (< 2 years) recurrence, which reflects the intrinsic aggressiveness of the tumor. Well-differentiated, potentially low-aggressiveness tumors form the heterogeneous molecular class of non-proliferative HCCs, characterized by an approximate 50% β -catenin (*CTNNB1*) mutation rate. To define the clinical, pathological, molecular features and the outcome of non-proliferative HCCs, we constructed an 1133-HCC transcriptomic metadata set and validated findings in a publically available 210-HCC RNAseq set. We show that non-proliferative HCCs preserve the zonation program that distributes metabolic functions along the porto-central axis in normal liver. More precisely, we identified two well-differentiated, non-proliferation subclasses, namely *Periportal-type* (wild-type *CTNNB1*) and *Perivenous-type* (mutant *CTNNB1*), which expressed negatively correlated gene networks. The new *Periportal-type* subclass represented 29% of all HCCs; expressed an HNF4A-driven gene network, which was down-regulated in mouse *Hnf4a*-KO mice; were early-stage tumors by BCLC, CLIP and TNM staging systems; had no macrovascular invasion and showed the lowest metastasis-specific gene expression levels and *TP53* mutation rates. Also, we identified an 8-gene *Periportal-type* HCC signature, which was independently associated with the highest 2-year recurrence-free survival by multivariate analyses in two independent cohorts of 247 and 210 patients.

Conclusion: Well-differentiated HCCs display mutually exclusive periportal or perivenous zonation programs. Among all HCCs, *Periportal-type* tumors have the lowest intrinsic potential for early recurrence after curative resection.

Accepted Article

INTRODUCTION

Hepatocellular carcinoma (HCC) is the 3rd leading cause of cancer-related death worldwide. HCC is endemic in East Asia and sub-Saharan Africa, and its incidence has doubled over the past 20 years in Western countries. Projections anticipate a further increase in incidence, despite recent breakthroughs in the management of chronic hepatitis.(1) More than 80% of HCCs arise in a field of chronic liver disease resulting from viral hepatitis, alcohol, hemochromatosis, obesity and metabolic syndrome or genotoxins. This diversity in etiology and natural history results in high HCC heterogeneity, raising major challenges in clinical management.(1)

Current therapeutic strategies are based upon tumor number, size, vascular invasion, performance status and functional reserve of the liver. These variables have been integrated in the Barcelona Clinic Liver Cancer (BCLC) classification to match the best candidates with the best therapies available.(1) Surveillance programs currently detect early-stage tumors (single or ≤ 3 nodules ≤ 3 cm) that are candidates for potentially curative therapies (local ablation, resection or transplantation) with 5-year survival rates of 50-70%. Five-year recurrence rates after HCC resection (up to 70%) or even higher after percutaneous ablation make liver transplantation the best possible treatment, with a recurrence rate of 10% and the additional advantage of eliminating the underlying liver disease.(1) Treatment allocation for early-stage HCCs might be improved by identifying homogeneous molecular subclasses with predictable outcomes.(2, 3)

The molecular landscape of HCCs is emerging as a result of global gene expression analyses and the discovery of crucial driver tumor mutations.(1, 2) At present, HCCs are split into two molecular classes, each representing 50% of tumors: proliferative

and non-proliferative HCCs.(2) Proliferative HCCs include two subclasses enriched in WNT/TGFB signals and stem/progenitor cell markers, respectively.(1, 2) Among non-proliferative HCCs, half of the tumors form a homogeneous subclass with a high rate of activating β -catenin (*CTNNB1*) exon 3 mutations.(2, 4-6) These tumors are well-differentiated,(2, 4-8) but patient survival rates do not differ from HCCs carrying wild-type *CTNNB1* (reviewed in Supporting Table S1). The rest of non-proliferative HCCs carry wild-type *CTNNB1*,(5) but no evidence to date allowed them to qualify as a homogeneous HCC subclass.

This study was conducted to determine whether non-proliferative HCCs carrying wild-type *CTNNB1* warrant consideration as a distinct, clinically relevant tumor subclass.

To analyze a large and statistically powerful population, we constructed an 1133-HCC transcriptomic metadata set and identified four HCC subclasses by discriminant analyses and hierarchical clustering. As the 1133 HCCs were not annotated for *CTNNB1* mutations, we developed a reliable method to predict *CTNNB1* mutations in an independent set of 225 β -catenin-sequenced HCCs. *CTNNB1* mutation prediction, HCC classification and survival analyses were further validated in an external RNAseq dataset of 210 HCC patients (Fig. 1).

Altogether, analysis of data from 1568 HCC patients identified two new well-differentiated, mutually exclusive, low-proliferation subclasses of HCCs. Both subclasses showed favorable outcomes and preserved metabolic liver zonation programs. They respectively displayed *periportal* (wild-type β -catenin) or *perivenous* (mutant β -catenin) phenotypes. *Periportal-type HCCs* showed the highest 2-year recurrence-free survival rates by multivariate analysis, suggesting that these tumors have the lowest potential for early recurrence among all HCCs.

Patients and Methods

Patients, study design and data management

The study design is presented in Fig. 1. Nine public transcriptomic datasets passing the quality control analyses were integrated into a metadata set totalizing 1133 HCCs and 9542 genes (Supporting Tables S2 and S3). As HCCs in these datasets were not annotated for *CTNNB1* mutations, we developed a robust prediction method. Complying with REMARK guidelines, identification of *CTNNB1* mutation predictors in a training set (n=87) was confirmed in a validation set (n=56) and in our recently described in-house independent cohort (n=82).⁽⁹⁾ The three datasets (Supporting Table S2) were annotated for Sanger-sequenced *CTNNB1* exon 3 mutations (Supporting Table S4). We thus predicted *CTNNB1* mutational status in the 1133-HCC transcriptomic metadata set. Independently, hierarchical cluster analysis was used in this metadata set to identify robust HCC subclasses. Then, the resulting tumor classification was validated in a 210-HCC patient dataset from The Cancer Genome Atlas (TCGA) consortium⁽¹⁰⁾ (normalized RNAseq genome-wide mRNA expression, whole genome sequencing and clinical data downloaded in June 2016 from the TCGA website). Out of 294 HCCs with available RNAseq and clinical annotations, 84 met our exclusion criteria (Supporting Fig. S1A). Quality control of clinical annotations was performed by studying the relationship between tumor staging and overall/disease-free survival by Kaplan-Meier plots and Log Rank tests (Supporting Fig. S1B).

Raw data from publically available transcriptome profiling experiments were extracted from Gene Expression Omnibus. Probes detected over background in at least one HCC were quantile normalized (R package preprocessCore) and log₂ intensity

expression values for each probe set were calculated by Robust Multi-array Average Cross-platform and batch-dependent variances were corrected with COMBAT(11) (R package *sva*) in the 1133-HCC metadata set and with YuGene(12) (R package *YuGene*) in the datasets used for *CTNNB1* mutation prediction.

Statistical analyses

To optimize the identification of stable gene predictors of *CTNNB1* activating mutations, we applied Sparse Partial Least Squares Discriminant Analysis (sPLS-DA) with bootstrap subsampling (Supporting Fig. S2), as described.(13) After further marker validation in the in-house cohort by real-time PCR, we established a score to predict *CTNNB1* mutations. Independently, robust HCC subclasses were identified using hierarchical cluster analysis defined by a stepwise algorithm (see Supporting Materials and Methods). Survival analyses were performed using the Log Rank test, Kaplan-Meier curves and Cox models. Statistical analyses were performed with R (version 3.3.0).

See Supporting Materials and Methods for: tissue banking, fully detailed statistics, functional genomics, detection of β -catenin exon 3 mutations/deletions, tissue microarray-based immunohistochemistry, antibodies, cell culture, nucleic acid extraction, real-time PCR, primers, and RNA interference.

RESULTS

A robust 5-gene score predicts *CTNNB1* mutations in large transcriptomic datasets

As described in the *Patients and Methods* section and Fig. 1, nine public transcriptomic datasets were integrated into a metadata set of 1133 HCCs and 9542 genes (Supporting Tables S2 and S3). The 1133-HCC transcriptomic dataset was not annotated for *CTNNB1* mutational status; thus, we set up a robust pipeline to predict *CTNNB1* activating mutations (Fig. 1). The model achieved a prediction accuracy of 87% in the training set and 93% in the independent validation sets (Supporting Fig. S2). The analysis identified Histidine Ammonia-Lyase (*HAL*) and Vanin 1 (*VNN1*) (selected in >85% of the models after 300 bootstrap runs), which were both downregulated in HCCs carrying mutant *CTNNB1*. In addition, a search for the top upregulated genes in HCCs carrying *CTNNB1* mutations identified *ODontogenic AMeloblast-associated protein (ODAM)*. Then, to construct a score predicting *CTNNB1* mutations, we combined *HAL*, *VNN1* and *ODAM* with two well-known markers of *CTNNB1* activating mutations in HCCs: *GLUL* and *LGR5*.⁽¹⁴⁾ RNA expression levels of the five biomarkers were significantly associated with mutant *CTNNB1* in the three datasets (Fig. 2A,B). *AXIN2*, which reflects Wnt/ β -catenin pathway activation,⁽¹⁵⁾ was positively correlated with *GLUL*, *LGR5* and *ODAM* and negatively correlated with *HAL* and *VNN1* (Fig. 2A,B; Supporting Fig. S3A). Cluster analyses confirmed the association of the five markers with *CTNNB1* mutation (Fig. 2, B1,B2,B3). In the external validation set, *CTNNB1* mutation was negatively associated with invasive (Gamma= -0.83, $p < 10^{-3}$) and multinodular (Gamma= -0.56, $p = 0.02$) tumors (Fig. 2B3). In turn, nuclear β -catenin immunostaining was positively correlated with *CTNNB1* mutations, *GLUL*, *ODAM* and *LGR5* and negatively with

VNN1 and *HAL* and (Supporting Fig. S3B). Combination of the five markers into an arithmetic score resulted in a reliable prediction in the three datasets (AUROC=0.87-0.90; Sensitivity=0.86-0.91; Specificity=0.83-1; Fig. 2C-E).

Tissue microarray-based immunohistochemistry in 20 HCCs carrying mutant *versus* 20 HCCs carrying wild-type *CTNNB1* revealed a clear-cut difference in GLUL expression between HCCs showing mutant *versus* wild-type *CTNNB1* at low power microscopic examination, whereas *HAL* and *VNN1* were globally higher in wild-type samples (Fig. 2F). *ODAM* protein expression was low in all samples (not shown), but detected in HCCs carrying mutant *CTNNB1* and expressing high GLUL (Fig. 2G).

In the external validation set, analysis of matching HCC and non-tumor samples showed that tumor/non-tumor ratios were highly predictive of *CTNNB1* mutations (AUROC=0.90, Se 0.87; Sp 0.88) (Supporting Fig. S3C-F). As data from non-tumor tissues were not available in all the transcriptomic datasets, further predictions of *CTNNB1* mutations were performed with tumor samples. Thus, we predicted *CTNNB1* mutations in the 1133 HCCs of the metadata set (Supporting Table S5). Predicted mutation rates between 20-40% matched the previously described *CTNNB1* mutation rates in HCC.(16)

To better understand the association of the newly discovered biomarkers (*HAL*, *VNN1* and *ODAM*) with *CTNNB1* activating mutations, we analyzed human HCC cell lines (Supporting Fig. S4). *HAL* was expressed at higher levels in HCC cell lines carrying wild-type *CTNNB1*, *i.e.*, Huh7 or HepaRG, than in those carrying activating *CTNNB1* mutations, *i.e.*, B16, BC2, HepG2 and Huh-6. *VNN1* was expressed at much higher levels in HepaRG cells than in the cell lines carrying mutant *CTNNB1*. By contrast, Huh7 expressed *VNN1* at low levels. Last, *ODAM* was clearly expressed

at higher levels in cell lines carrying mutant than in those carrying wild-type *CTNNB1*. The GSK3 β inhibitor 6-bromoindirubin-3'-oxime (BIO), which activates β -catenin signaling,(17) strongly upregulated β -catenin-dependent transcriptional activity (Supporting Fig. S5A). BIO upregulated *GLUL*, *LGR5* and *ODAM* (Supporting Fig. S5B) and attenuated the increase in *HNF4A*, *ALDOB*, *HAL* and *VNN1* mRNA expression over the 30 days required for differentiation of HepaRG progenitors to hepatocyte-like cells (Supporting Fig. S5C-F). Conversely, HepaRG cells transfected with β -catenin targeting siRNA downregulated *AXIN2* and *ODAM* and upregulated *HAL* and *VNN1* mRNA expression (Supporting Fig. S5G). In addition, *in silico* analysis of the 5000 base pairs upstream the transcription start site of *ODAM* DNA (NM_017855; PROMO program,(18) TRANSFAC database), revealed three putative TCF-4/LEF-1 consensus transcription factor binding sites in the *ODAM* 5'UTR region at positions -99; -1262 and -3956. Taken together, these data suggest that β -catenin signaling regulates *HAL*, *VNN1* and *ODAM* mRNA expression.

Predicted *CTNNB1* mutations cluster within a homogeneous tumor subclass after analysis of transcriptomic data from 1343 HCCs.

We integrated the nine datasets into an 1133 HCC metadata set with 9542 common genes and performed cross-platform normalization with the COMBAT algorithm(11) (Supporting Fig. S6). Identification of the 1618 most discriminant genes by partial least squares discriminant analysis (PLS-DA), using a VIP score threshold >1, followed by stepwise hierarchical cluster analysis, identified four subclasses. Subclasses were named *red*, *green*, *blue* and *purple* before further functional characterization (Supporting Fig. S7). The *red* subclass had a predicted *CTNNB1*

mutation rate of 84% (Supporting Fig. S7). The predicted *CTNNB1* mutation rates in this subclass were high in all datasets (65% to 100%, Supporting Table S6). These findings are consistent with previous evidence that HCCs carrying *CTNNB1* mutation are a homogeneous subclass.(4, 6)

Reproducibility (Supporting Fig. S8) and robustness (Supporting Fig. S9A,B) analyses showed that the *red* and *green* subclasses had similarly low intra-cluster inertia, indicating high homogeneity of HCC samples within each cluster. In terms of gene expression profiles, the *purple* cluster was highly different from the *red* and *green* ones, whereas the *blue* cluster was intermediate between the green and purple ones (Supporting Fig. S9C). To further test the robustness of the HCC classification, we repeated PLS-DA-based identification of the most discriminant genes with a more stringent VIP score threshold (>1.5), which yielded 550 top discriminant genes and confirmed the four tumor subclasses (Supporting Fig. S9D). Further, the value of the 550-geneset as a classifier was confirmed in the external validation RNAseq dataset (n= 210 HCCs; Supporting Fig. 9E). Thus, using the top 550 genes, both the 1133-HCC transcriptomic (Fig. 3A) and the 210-HCC RNAseq (Supporting Fig. 9E) datasets confirmed four HCC subclasses and three gene sets. In the *red* subclass, *CTNNB1* was predicted to be mutated in 89% of HCCs (Fig. 3A). In consistency with the high rate of predicted *CTNNB1* mutations, the red subclass showed the highest levels of *GLUL*, *LGR5* and *ODAM* and the lowest levels of *VNN1* and *HAL* (Fig. 3B).

By immunohistochemistry in human liver (Fig. 3C-D), *GLUL* showed a clear perivenous localization, as expected.(19) *HAL* showed a clear-cut periportal distribution, consistently with its expression by periportal hepatocytes in mice.(20) *VNN1* was panlobular and was abundantly detected in the cytoplasm and in bile

canaliculi in normal liver and in well-differentiated HCCs (Fig. 3C,D), which goes well with the evidence that VNN1 is one of the 59 major proteins in human bile.(21). Further, anti-VNN1 staining of bile canaliculi was totally suppressed by the immunogen polypeptide (Fig. 3D). In consistency with the above findings, full genome sequencing data from the independent RNAseq 210-HCC TCGA dataset confirmed that 81% of HCCs within the *red* subclass carried *CTNNB1* mutations (Supporting Fig. S9E).

Well-differentiated HCCs display a preserved metabolic liver zonation program.

Functional genomics studies by gene set enrichment analysis (GSEA) (Fig. 4; Supporting Table S7) revealed that the HCC subclass containing the highest rate of predicted *CTNNB1* mutations (84%) was enriched in signatures of HCCs carrying mutant *CTNNB1*, such as the G6(6) and the *CTNNB1* subclasses,(4) as well as perivenous hepatocyte signatures, such as fatty acid and bile salt metabolism.(20) We thus called this subclass “Perivenous-type HCCs” (*PV*). The HCC subclass enriched in signatures of differentiated periportal hepatocytes (gluconeogenesis, amino acid catabolism, HNF4A-induced genes),(20) good prognosis, (hepatocyte-like (S3) subclass HCCs)(5) and low recurrence,(5, 22) was called “Periportal-type HCCs” (*PP*). Both *PP* and *PV* subclasses were enriched in signatures of favorable survival(23) and low proliferation in HCCs.(4) The HCC subclass featuring signatures of cancer extracellular matrix (ECM) remodeling and epithelial mesenchymal transition(24) was called “ECM-type HCCs”. It was characterized by integrin cell-surface interactions, *KRAS*, *TGFB*, *IL6*, and the *MMP14* network typical of invasive tumors. The last HCC subclass was associated with high grade metastasis

signatures, tumor aggressiveness with decreased patient survival,(23) the cancer stem cell program predicting metastasis and death(25) and upregulation of p53 mutation markers (Supporting Fig. S10). It was also enriched in signatures of cell cycle progression (*MYC*, *RAC*, *AURKA*, *RB1* and *PDGFRB* pathways), HCCs displaying stem cell features(26) and extrahepatic undifferentiated cancers of various origins, as well as cancer cell resistance to doxorubicin and vincristine. Thus, we called this subclass “STEM-type HCCs”. Both ECM- and STEM-type HCCs shared signatures of high tumor cell proliferation(4) and were associated with the Wnt/TGF- β (S1) HCC subclass,(5) indicating bad prognosis.(27) Also, they were both associated with early recurrence(28). Expression levels of representative genes in HCC subclasses and in gene clusters are illustrated in Fig. 4B and in Supporting Fig. S11, respectively. Functional genomics findings were confirmed by gene ontology analysis of gene clusters (Supporting Table S8). The mean \pm 95%CI mRNA expression level of the top 550 genes in the four HCC subclasses is shown in Supporting Table S9.

We observed differences in the percentage of samples in each subclass according to the dataset of origin (Supporting Table S10). The sampling technique (1060 resections *versus* 73 biopsy specimens) had no impact on tumor classification (Supporting Table S11). By contrast, cohorts containing >95% HCV (+) HCCs showed a higher rate of PV samples than cohorts with >95% HBV (+) ($p= 10^{-3}$), in agreement with the well-known positive association of *CTNNB1* mutation with HCV and its negative association with HBV.(2)

The metabolic program of Periportal-type HCCs is regulated by HNF4A

The interplay between TCF4, β -catenin and HNF4A(20) governs the differential distribution of metabolic liver functions along the porto-central axis of the liver, which is known as “liver zonation”.(29) Using the ortholog genes of the mouse liver periportal and perivenous signatures,(20) we clustered the 326 Periportal-type and the 210 Perivenous-type HCCs from the 1133 HCC metadata set. We observed a remarkable matching of mouse normal liver perivenous genes with Perivenous-type HCCs and mouse normal liver periportal genes with Periportal-type HCCs, thus confirming the identity of both HCC subclasses described above (Supporting Fig. S12A). Subsequently, network analysis in the 1133 patient metadata set revealed that periportal and perivenous genes were negatively correlated and respectively upregulated in Periportal-type and Perivenous-type HCCs (Fig. 5A). Predictive markers of *CTNNB1* mutation showed high connectivity within this network. *GLUL*, *HAL* and *LGR5* were connected respectively with 60 (42%), 54 (38%) and 54 (38%) of genes in the network. Although mouse *Odam* and *Vnn1* were not present in the original periportal and perivenous signatures,(20) human *ODAM* and *VNN1* were connected respectively with 47 (33%) and 28 (20%) of the genes in human HCCs.

Suppression of HNF4A in mouse liver has profound effects on zoned metabolic functions,(30, 31) cell proliferation(30) and oncogenesis.(32, 33) Consistent with this, the top 550 genes representative of the four HCC subclasses perfectly discriminated *Hnf4a*-KO and *Hnf4a*-WT(31) mouse livers (Fig. 5B). Among the 550 genes, a first cluster was strongly upregulated in *Hnf4a*-KO mice and clearly enriched in genes upregulated in ECM/STEM HCCs. A second cluster was downregulated in *Hnf4a*-KO mice and enriched in genes upregulated in Periportal-type HCCs. The rest of the genes (20%) were poorly affected by HNF4A. Finally, we extracted from the *Hnf4a*-KO dataset the genes showing the highest modulation (fold change >2 or <0.5 and

Bonferroni adjusted P value <0.01) to look at the expression of their orthologs in the 1133 HCC metadata set (Supporting Fig. S12B). Hierarchical clustering of these genes showed a good discrimination of the four subclasses (80% of samples correctly classified), with a strong enrichment of genes highly downregulated by *Hnf4a*-KO in Periportal-type HCCs. Conversely, genes highly upregulated by *Hnf4a*-KO were remarkably enriched in the STEM-type HCC subclass. These findings are in line with the notion that HNF4A induces hepatocyte differentiation of cancer stem cells, suppressing hepatocyte proliferation and oncogenesis.(30, 32, 33)

Periportal-type HCCs show the most favorable clinical features and the highest early (2-year) disease-free and overall survival rates after resection.

Comparative analysis of the clinical outcome of the four HCC subclasses in a 247-HCC transcriptomic dataset(34) (dataset C, Supporting Table S2) revealed increasingly aggressive tumor phenotypes from Periportal through Perivenous → ECM → STEM HCCs (Fig. 6A). These subclasses showed a progressive increase in metastatic signature(34) rates, CLIP and BCLC scores as well as serum AFP concentrations. Survival curves (Fig. 6B) showed a progressive decrease in overall and disease free survival (DF) rates from Periportal through Perivenous → ECM → STEM HCCs. Thus, the Periportal-type is a well-differentiated, favorable-outcome HCC subclass carrying wild-type *CTNNB1*, displaying a periportal liver metabolic program and expressing HNF4A target genes. In turn, the Perivenous-type is a well-differentiated HCC subclass carrying mutant *CTNNB1*, displaying a perivenous liver metabolic program and expressing β -catenin target genes.

The above data were further validated using an external 210-HCC RNAseq dataset. Tumor aggressiveness was inferred from increasing AFP serum levels, tumor grade, TNM staging, vascular invasion, TP53 mutation rates, tumor onset in younger patients (Fig. 6C), as well as survival (Supporting Fig. S13). This analysis confirmed the existence of increasingly aggressive tumor phenotypes from Periportal-type through Perivenous-type → ECM-type → STEM-type HCCs. The overall survival of Perivenous-type HCCs was significantly more favorable than that of STEM-type HCCs in both the 247-HCC transcriptomic (Fig. 6B) and in the 210-HCC RNAseq (Supporting Fig. S13) datasets.

To obtain a *Periportal-type HCC gene signature*, we searched for genes meeting the following criteria: >2 fold change between the Periportal and the other HCC subclasses; high connectivity within the periportal gene network in HCCs (>0.30 correlation coefficient with >50% of the genes) and association with survival ($p < 0.05$ in $\geq 7/8$ tests; Supporting Materials and Methods). The resulting 8-gene *Periportal-type HCC signature* (Fig. 6D) was associated with favorable overall and disease-free outcome in the 247-HCC transcriptomic (34) (Fig. 6E) and in the 210-HCC RNAseq (Fig. 6F) datasets. Periportal-type HCCs showed the lowest early (2-year) recurrence and the highest overall survival rates after resection among all other HCCs taken together. Univariate and multivariate survival analysis revealed that the *Periportal-type HCC signature* was independently associated with low early recurrence after HCC resection (Table 1).

Of note, the *Periportal-type HCC signature* was predominantly expressed in normal liver, as shown by analysis of their mRNA expression levels in 84 tissues,(35, 36) from the GSE1133 microarray dataset (Supporting Fig. S14,A-H). In addition, the expression levels of the eight genes constituting the signature were highly correlated

(Supporting Table S12). These genes were expressed at higher levels in Periportal-type HCCs than in the other HCC subclasses (Fig. 6D). The expression levels of these genes in 70 Periportal HCCs were closer to those detected in 232 non-tumor livers than in 167 non-Periportal HCCs. Two of these genes (*AGXT* and *OTC*) were not significantly different in non-tumor livers from Periportal HCCs (Supporting Fig. S15).

DISCUSSION

Our study revealed that preservation of the liver zonation program is an important feature of well-differentiated HCCs. Periportal-type HCCs showed the most favorable patient outcome in two independent cohorts of 247 and 210 patients. Further, they displayed the lowest rates of *Metastasis Signature*,⁽³⁴⁾ BCLC, CLIP, TNM staging, vascular invasion and TP53 mutations. We identified an 8-gene *Periportal-type HCC signature* composed of liver-enriched genes involved in periportal liver functions, such as amino-acid catabolism (GLS2, AGXT and GNMT), biliary acid transport (SLC10A1 and SCL22A7), lipid (SLC27A5) and glucose (FETUB) metabolism, as well as the urea cycle (OTC). These genes were expressed at higher levels in non-tumor livers and in Periportal-type than in non-Periportal-type HCCs.

In normal liver, β -catenin is transiently activated in perivenous zones to meet homeostatic needs.⁽¹⁵⁾ Therefore, turning *OFF* β -catenin in adult mouse liver constitutively leads to a default periportal metabolic program, with a normal liver function. Conversely, turning *ON* β -catenin leads to liver hyperplasia, hyperammonemia and death.⁽²⁰⁾ Considered in this context, Periportal-type HCCs preserve the default metabolic program of normal liver. Therefore, converging evidence indicates that the periportal program embodies the lowest level of aggressiveness in HCCs.

Recently, no association was detected between β -catenin mutations and survival in 362 HCC patients,⁽⁷⁾ in line with data summarized in Supporting Table S1. We confirmed these findings by analyzing two independent cohorts of 247 and 210 HCC patients (Fig. 7A). Noteworthy, subclass-specific survival analyses in these two cohorts revealed that Periportal-type and Perivenous-type HCC patients had a more favorable outcome than STEM-type HCC patients (Fig. 6B; Supporting Fig. S13).

Also, discriminant analysis confirmed that Perivenous-type HCCs lie between Periportal-type and ECM-type HCCs in terms of aggressiveness (Fig. 7B). This body of evidence suggests that HCCs carrying β -catenin mutations are moderately aggressive (Fig. 7C). *De novo* expression of cancer stem/progenitor cell genes such as *ODAM* and *LGR5* in *CTNNB1*-mutated HCCs goes well with the moderate aggressiveness of this tumor subclass. *ODAM* is a basement membrane component of the periodontal junctional epithelium secreted by ameloblasts, repressed in adult tissues, but expressed *de novo* in cancers.(37) *LGR5*, in turn, is expressed by liver stem cells or progenitors in response to stimuli calling for tissue renewal or tumor cell growth.(38)

Perivenous-type HCCs match the G6(6) and *CTNNB1*(4) subclasses that carry *CTNNB1* mutations, but not the S3(5) subclass. In turn, the S3 subclass matches Periportal-type HCCs. Indeed, S3 HCCs are well-differentiated, hepatocyte-like, and approximately 50% of them carry *CTNNB1* mutations.(5) Further, S3 HCC patients have a more favorable survival than S1 and S2 HCC patients.(27) As the main feature of S3 HCCs is preserved hepatocyte differentiation, S3's may include both Periportal-type and Perivenous-type HCCs.

GLUL, *HAL* and *VNN1* are involved in metabolic liver zonation. *HAL* is an enzyme that degrades the amino acid histidine(39) and we showed that it was restricted to periportal hepatocytes in normal liver. Likewise, Periportal-type HCCs expressed a host of amino acid degrading enzymes, such as *ARG1* (arginase 1) and *GLS2* (glutaminase 2), which were major hubs in the periportal gene network in HCCs (Fig. 5A). Some of these amino acid degrading enzymes, such as *SDS* (serine dehydratase) and *GLDC* (glycine decarboxylase), yield the end-product oxaloacetate, which is used for gluconeogenesis, a periportal metabolic pathway.(20, 40) *VNN1* is

a key activator of hepatic gluconeogenesis, transcriptionally activated by HNF4A.(41) By immunohistochemistry, VNN1 had a panlobular distribution, in consistency with unsupervised transcriptomic data,(20, 40) suggesting that *VNN1* probably responds to dynamic zonation.(29)

Cross-talks between HNF4A and β -catenin control the transcription of genes involved in periportal and perivenous zonation in normal liver.(29) Our data suggest that this cross-talk is preserved in well-differentiated HCCs. The 5-gene score predicting *CTNNB1* mutation was composed of genes known to be upregulated by β -catenin (*LGR5*, *GLUL* and *ODAM*) or by HNF4A (*VNN1* and *HAL*). (20, 41) By *in vitro* experiments, we confirmed that β -catenin activation enhanced *GLUL*, *LGR5* and *ODAM*, but inhibited *HAL* and *VNN1* expression. Conversely, β -catenin knockdown silenced *AXIN2* and *ODAM*, but enhanced *HAL* and *VNN1* expression. Further, we showed that β -catenin activation attenuated the gain in the expression of *HNF4A*, *ALDOB*, *HAL* and *VNN1* upon the 30-day time course of differentiation of HepaRG progenitors into hepatocyte-like cells (Supporting Fig. S5C-F). These observations suggest that β -catenin mutations divert the default periportal metabolic program of well-differentiated HCC cells toward a perivenous one (Fig. 7B).

In summary, we identified a new well-differentiated HCC subclass, Periportal-type HCCs, which has the most favorable clinical outcome among all HCCs. Accordingly, the Periportal-type HCC signature was associated with low early HCC recurrence (< 2 years after resection) independently of tumor size and vascular invasion in multivariate analyses. Given the heterogeneity of HCC phenotypes shown in Fig 7B, the question arises whether early, well-differentiated HCC types (Periportal and Perivenous) ultimately progress to a more aggressive, progenitor-like phenotype spontaneously or in response to treatments that may spare cancer progenitor cells.

Prediction of patient survival impacts the indication of liver transplantation for HCC. In particular, early recurrence is one of the principles governing patient staging within the spectrum of eligibility for liver transplantation.(3) Thus, it would be interesting to test whether identification of Periportal-type HCCs contributes to adapt indications of liver transplantation beyond current criteria. This issue will gain importance in the near future, because the control of the hepatitis C virus epidemics with direct antiviral agents will increase donor liver availability.

The major limitation to the direct translation of these basic findings to the clinical arena is the exclusion of diagnostic liver biopsy from current guidelines for HCC diagnosis and allocation to treatments. However, in view of exciting results demonstrating the clinical relevance of metabolic imaging in early-stage HCCs(42) and of the association of the transporter of gadoxetic acid OATP1B3 with Wnt/ β -catenin signaling in HCCs(43), the Periportal-type molecular phenotype raises a new challenge to metabolic imaging of HCCs.

FIGURE LEGENDS

Fig. 1. Flowchart of the study. (A) Construction of a transcriptomic metadata set for subclass-specific analyses after curative resection of 1133 HCCs. Nine public transcriptomic datasets (*A to I*) were merged into a metadata set. As the 1133-HCCs were not annotated for *CTNNB1* mutations, we developed a *CTNNB1* mutation prediction score from transcriptomic data. The most stable markers were identified by sPLS-DA modeling in training and validation sets. Mutation prediction was validated in an external in-house HCC collection(9); then, *CTNNB1* mutations were predicted in 1133 HCCs after cross-platform normalization. Subclass-specific survival analyses were performed in dataset C (n= 247 HCCs)(34) and validated in an external RNAseq, full-genome sequenced, dataset (LIHC-US; TCGA consortium, n= 210 HCCs).(10) (B) Transcriptomic datasets used: 1, yes or present; 0, no or absent. References: Training set(4); Validation set(6); A(44); B(45); C(34); D(46); E(47); F(48); G(49); H(49); I(28); RNAseq dataset(10). Major endpoints are indicated by bold arrows. External validation sets are represented by black boxes.

Fig. 2. A 5-gene score reliably predicts *CTNNB1* mutations using mRNA expression data in HCCs. (A) Gene expression of the five biomarkers in the training (n=87), validation (n=56) and the external validation (n=82) sets, according to Sanger-sequenced *CTNNB1* mutational status. *AXIN2* was included as a Wnt pathway activation marker (missing in *A2*). (B) Heatmaps showing mRNA expression of the five biomarkers plus *AXIN2* and *CTNNB1* mutational status (B1, training set; B2, validation set; B3, external HCC collection). Low and high gene expressions (median as cut-off) are represented in green and red, respectively. (C) Formula; (D)

Performance and (E) ROC curves of the *CTNNB1* mutation prediction score. (F) Tissue micro array-based immunostainings (*brown*) for GLUL, HAL and VNN1 in 20 *CTNNB1* mutant and 20 *CTNNB1* wild-type HCCs. Insets on the right: β -catenin cell membrane staining in *CTNNB1* wild-type (*arrowheads, upper right*), or cytoplasmic and nuclear staining in *CTNNB1* mutant HCCs (*arrowheads, lower right*). (G) Immunostaining (*brown*) for ODAM and GLUL in serial sections in a *CTNNB1* mutant HCC. Images were acquired with a tissue slide scanner at 20X original magnification. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (Wilcoxon tests). AUROC, Area under the receiver operating curve; qPCR, real-time PCR (a pool of five normal liver samples was used as calibrator); CI, Confidence interval; Se, Sensitivity; Sp, Specificity.

Fig. 3. (A) Hierarchical clustering of transcriptomic metadata from 1133 HCCs reveals four tumor subgroups (*green, red, purple, & blue*) with different proportions of predicted *CTNNB1* mutations. Lowest and highest gene expression values are represented in *green* and *red*, respectively. (B) Gene expression levels of the five *CTNNB1* mutation predictors. Results are relative to the green subclass, arbitrarily set to $1 \pm 95\%$ CI. Green, $n=326$; Red, $n=210$; Blue, $n=420$; Purple, $n=177$. NS, non-significant. ** $P < 0.01$, *** $P < 0.001$, after Student's *t* tests. (C) Immunostainings (*brown*) of β -catenin, GLUL, VNN1 and HAL in serial sections of normal human liver illustrates clear-cut perivenous and periportal zonation of GLUL and HAL, respectively. β -catenin is predominantly membranous and highlights bile ducts (*BD*). Arrows show portal tracts (*PT*) and central veins (*CV*). VNN1 is panlobular. (D) VNN1 is detected in bile canaliculi in normal liver and in a well-differentiated HCC. VNN1 signal is totally suppressed by the immunogen polypeptide.

Fig. 4. Well-differentiated HCCs display a preserved metabolic liver zonation program. (A) Gene Set Enrichment Analysis (GSEA) in the four HCC subclasses. (B) Expression of relevant genes in the 1133-HCC metadata set. Color code keys are shown at the bottom of the heatmap. Samples (columns) are ordered by HCC subclass. *Green*, low; *red*, high expression. *PP*, Periportal-type; *PV*, Perivenous-type; *ECM*, Extracellular matrix-type; *STEM*, Stem cell-type.

Fig. 5. The metabolic program of Periportal-type HCCs is regulated by HNF4A. (A) Periportal (*PP*, *green*) and Perivenous (*PV*, *red*) gene networks constructed by Weighted Gene Correlation Analysis and shown with Cytoscape graphics. Networks consist of human orthologs of mouse liver periportal and perivenous gene signatures,⁽²⁰⁾ detected in 326 Periportal-type and 210 Perivenous-type HCCs from the 1133 HCC set. *Node border colors (Green/red)* denote periportal/perivenous attributes of mouse genes, respectively.⁽²⁰⁾ *Node core colors* are proportional to the PV/PP fold-change in the HCC set (*Color key: lower right*). *Link thickness* is proportional to correlation coefficients (>0.30 in all networks; *blue*, positive; *red*, negative correlation). (B) The 550 genes representative of the four HCC subclasses discriminate between *Hnf4a*-KO and *Hnf4a*-WT mice.⁽³¹⁾ Three gene clusters (*black nodes on the left*) denote: 1, upregulation in *Hnf4a*-KO's and enrichment in human HCC STEM/ECM signatures; 2, upregulation in *Hnf4a*-WT's and enrichment in HCC PP genes; 3, genes poorly affected by *Hnf4a* status. *Red*, high expression, *green*, low expression.

Fig. 6. Periportal-type HCCs show the most favorable clinical features and the highest early (2-year) disease-free and overall survival rates after resection. (A) Clinical features of HCC subclasses in a 247-patient dataset(34). (B) Kaplan-Meier plots of subclass-specific overall and disease-free survival; *P<0.05, **P<0.01, ***P<0.001. (C) Subclass-specific clinical features, *CTNNB1* and *TP53* mutation rates (full genome sequencing) in an external validation 210-HCC RNAseq dataset (TCGA-LIHC-US). (D) Subclass-specific mRNA expression levels of *Periportal HCC signature* genes in 1133 HCCs. (E, F) Kaplan-Meier plots of overall and disease-free survival in HCC patients with and without the *Periportal HCC signature* in two datasets. AFP, serum alpha-fetoprotein; NA, not available; BCLC (Barcelona Clinic Liver Cancer); CLIP (Cancer of the Liver Italian Program). Statistics: Fisher exact test (categorical variables); Student's *t* test (continuous variables); Log-rank test (survival analyses).

Fig. 7. Periportal-type HCCs and the survival paradox between *CTNNB1*-mutated and wild-type HCCs. (A) Kaplan-Meier plots and log-rank tests in two independent HCC datasets confirm similar outcome of HCCs carrying mutated (*red*) or wild-type (*black*) *CTNNB1*. Mutational status results from prediction in dataset C and full-genome sequencing in TCGA-LIHC-US. (B) Multivariate discriminant analysis of the four HCC subclasses. In the first component, there is a grading from Periportal-type (*green*), through Perivenous- (*red*) and ECM-types (*blue*), to STEM-type (*purple*). In the second component, Perivenous-type differs from other HCC subclasses. (C) HCCs with predominantly mutated *CTNNB1* are moderately aggressive. (D) HCCs with predominantly wild-type *CTNNB1* include mildly, moderately and highly aggressive subclasses. The % of HCCs in each subclass is indicated in parentheses.

Acknowledgements: We thank the following core facilities: « *Structure Fédérative de Recherche en Biologie et Santé de Rennes* » UMS CNRS 3480/US INSERM 018 (Health Genomics, Jean Moser; High Precision Histopathology, Alain Fautrel); SynNanoVect electroporation facility (Pascal Loyer). We are indebted to Patricia Jouas and Michèle Le Guennec for secretarial support, to Edouard Curran for *CTNNB1* deletion analysis, to Sihem Mebarki for mRNA extraction, to Gaëlle Angenard and Catherine Ribault for tissue banking. We thank Marie-Jo Brion (The University of Queensland Diamantina Institute, the University of Queensland) for her thorough editing contribution. We thank Laurent Sulpice and Damien Bergeat (Dept of Gastrointestinal and Hepatobiliary Surgery, Rennes, France) for helpful discussion.

Financial support: Inserm, Univ Rennes 1, Région Bretagne, Agence Nationale de la Recherche, Institut National du Cancer, Fondation Recherche Médicale, Ligue Nationale Contre le Cancer (Comité des Côtes d'Armor), Feder, Contrat de Plan Etat Région 2007-2013, projet Cancéropôle.

Contributors: Study design, RD, OM, FR, K-A LC, CP, BC; statistics, RD, FR, K-A LC; experiments, RD, FC, MS, SR, PB, CP, OM; anatomic pathology data, BT, OM; tissue banking and quality control, MD, OM; manuscript preparation, RD, OM; manuscript editing, all authors.

Competing interests: None

Sample collection: Sample collection was reported to the Ministry of Education and Research (No. DC-2008-338).

Ethics approval: The study protocol complied with French laws and regulations and was approved by INSERM's Institutional Review Board (number 01-036) in the context of the National Network of Liver Biological Resource Centers.

REFERENCES

1. Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, Gores G. Hepatocellular carcinoma. *Nature reviews. Disease primers* 2016;2:16018.
2. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology* 2015;149:1226-1239 e1224.
3. Sapisochin G, Bruix J. Liver transplantation for hepatocellular carcinoma: outcomes and novel surgical approaches. *Nature reviews. Gastroenterology & hepatology* 2017;14:203-217.
4. Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, LeBlanc AC, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. *Cancer research* 2008;68:6779-6788.
5. Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, Villanueva A, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer research* 2009;69:7385-7392.
6. Boyault S, Rickman DS, de Reynies A, Balabaud C, Rebouissou S, Jeannot E, Herault A, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 2007;45:42-52.
7. Rebouissou S, Franconi A, Calderaro J, Letouze E, Imbeaud S, Pilati C, Nault JC, et al. Genotype-phenotype correlation of CTNNB1 mutations reveals different β -catenin activity associated with liver tumor progression. *Hepatology* 2016.
8. Lachenmayer A, Alsinet C, Savic R, Cabellos L, Toffanin S, Hoshida Y, Villanueva A, et al. Wnt-pathway activation in two molecular classes of hepatocellular

carcinoma and experimental modulation by sorafenib. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012;18:4997-5007.

9. Mebarki S, Desert R, Sulpice L, Sicard M, Desille M, Canal F, Schneider HD, et al. De novo HAPLN1 expression hallmarks Wnt-induced stem cell and fibrogenic networks leading to aggressive human hepatocellular carcinomas. *Oncotarget* 2016.

10. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nature communications* 2015;6:8971.

11. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;28:882-883.

12. Le Cao KA, Rohart F, McHugh L, Korn O, Wells CA. YuGene: a simple approach to scale gene expression data derived from different platforms for integrated analyses. *Genomics* 2014;103:239-251.

13. Rohart F, Mason EA, Matigian N, Mosbergen R, Korn O, Chen T, Butcher S, et al. A molecular classification of human mesenchymal stromal cells. *PeerJ* 2016;4:e1845.

14. Zucman-Rossi J, Benhamouche S, Godard C, Boyault S, Grimber G, Balabaud C, Cunha AS, et al. Differential effects of inactivated Axin1 and activated beta-catenin mutations in human hepatocellular carcinomas. *Oncogene* 2007;26:774-780.

15. Wang B, Zhao L, Fish M, Logan CY, Nusse R. Self-renewing diploid Axin2(+) cells fuel homeostatic renewal of the liver. *Nature* 2015;524:180-185.

16. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, Calderaro J, et al. Integrated analysis of somatic mutations and focal copy-number

changes identifies key genes and pathways in hepatocellular carcinoma. *Nature genetics* 2012;44:694-698.

17. Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature medicine* 2004;10:55-63.

18. Farre D, Roset R, Huerta M, Adsuara JE, Rosello L, Alba MM, Messeguer X. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic acids research* 2003;31:3651-3653.

19. Benhamouche S, Decaens T, Godard C, Chambrey R, Rickman DS, Moinard C, Vasseur-Cognet M, et al. Apc tumor suppressor gene is the "zonation-keeper" of mouse liver. *Dev Cell* 2006;10:759-770.

20. Gougelet A, Torre C, Veber P, Sartor C, Bachelot L, Denechaud PD, Godard C, et al. T-cell factor 4 and beta-catenin chromatin occupancies pattern zonal liver metabolism in mice. *Hepatology* 2014;59:2344-2357.

21. Kristiansen TZ, Bunkenborg J, Gronborg M, Molina H, Thuluvath PJ, Argani P, Goggins MG, et al. A proteomic analysis of human bile. *Molecular & cellular proteomics : MCP* 2004;3:715-728.

22. Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, Gupta S, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *The New England journal of medicine* 2008;359:1995-2004.

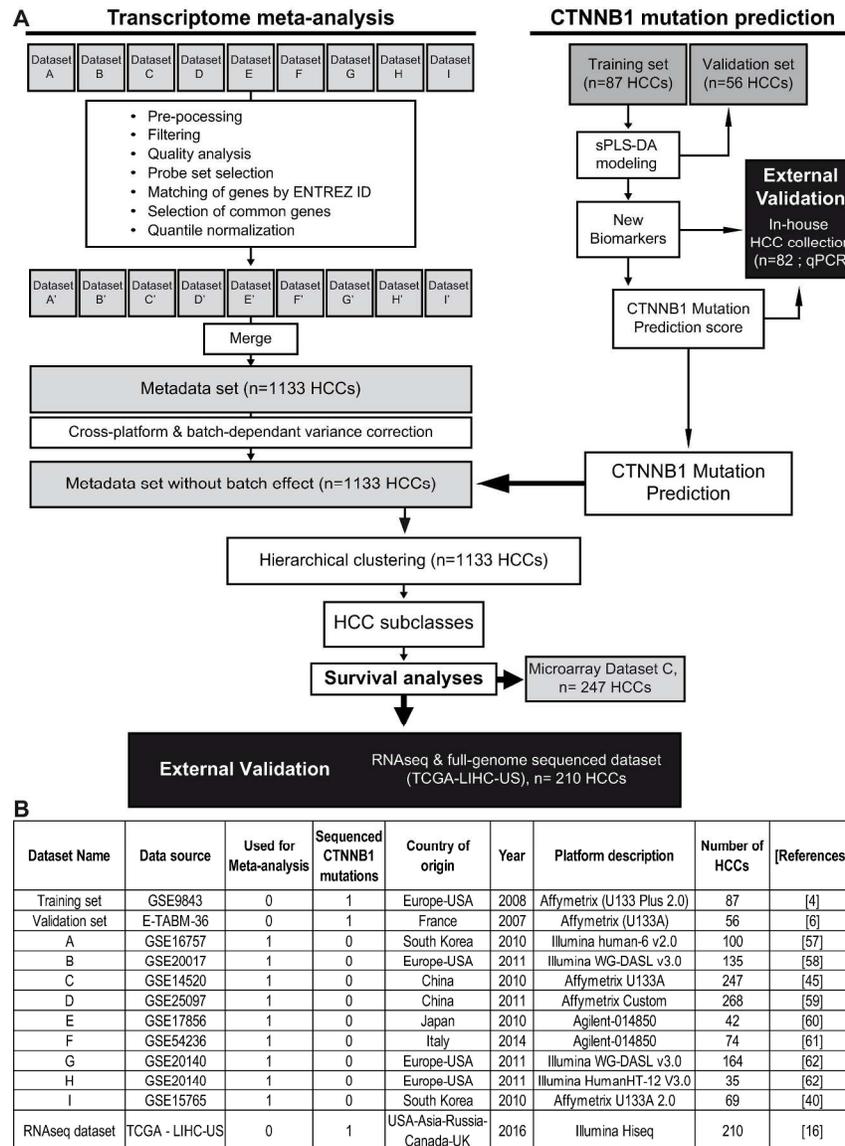
23. Lee JS, Chu IS, Heo J, Calvisi DF, Sun Z, Roskams T, Durnez A, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* 2004;40:667-676.

24. Naba A, Clauser KR, Hoersch S, Liu H, Carr SA, Hynes RO. The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular & cellular proteomics* : MCP 2012;11:M111 014647.
25. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell stem cell* 2008;2:333-344.
26. Yamashita T, Ji J, Budhu A, Forgues M, Yang W, Wang HY, Jia H, et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. *Gastroenterology* 2009;136:1012-1024.
27. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, Gao H, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome research* 2013;23:1422-1433.
28. Woo HG, Park ES, Cheon JH, Kim JH, Lee JS, Park BJ, Kim W, et al. Gene expression-based recurrence prediction of hepatitis B virus-related human hepatocellular carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008;14:2056-2064.
29. Berasain C, Avila MA. Deciphering liver zonation: new insights into the beta-catenin, Tcf4, and HNF4alpha triad. *Hepatology* 2014;59:2080-2082.
30. Bonzo JA, Ferry CH, Matsubara T, Kim JH, Gonzalez FJ. Suppression of hepatocyte proliferation by hepatocyte nuclear factor 4alpha in adult mice. *The Journal of biological chemistry* 2012;287:7345-7356.
31. Holloway MG, Miles GD, Dombkowski AA, Waxman DJ. Liver-specific hepatocyte nuclear factor-4alpha deficiency: greater impact on gene expression in male than in female mouse liver. *Molecular endocrinology* 2008;22:1274-1286.

32. Ning BF, Ding J, Yin C, Zhong W, Wu K, Zeng X, Yang W, et al. Hepatocyte nuclear factor 4 alpha suppresses the development of hepatocellular carcinoma. *Cancer research* 2010;70:7640-7651.
33. Hatziapostolou M, Polytarchou C, Aggelidou E, Drakaki A, Poultsides GA, Jaeger SA, Ogata H, et al. An HNF4alpha-miRNA inflammatory feedback circuit regulates hepatocellular oncogenesis. *Cell* 2011;147:1233-1247.
34. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, Thorgeirsson SS, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer research* 2010;70:10202-10212.
35. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101:6062-6067.
36. Wu C, Jin X, Tsueng G, Afrasiabi C, Su AI. BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic acids research* 2016;44:D313-316.
37. Foster JS, Fish LM, Phipps JE, Bruker CT, Lewis JM, Bell JL, Solomon A, et al. Odontogenic ameloblast-associated protein (ODAM) inhibits growth and migration of human melanoma cells and elicits PTEN elevation and inactivation of PI3K/AKT signaling. *BMC cancer* 2013;13:227.
38. Mebarki S, Desert R, Sulpice L, Sicard M, Desille M, Canal F, Dubois-Pot Schneider H, et al. De novo HAPLN1 expression hallmarks Wnt-induced stem cell and fibrogenic networks leading to aggressive human hepatocellular carcinomas. *Oncotarget* 2016;7:39026-39043.

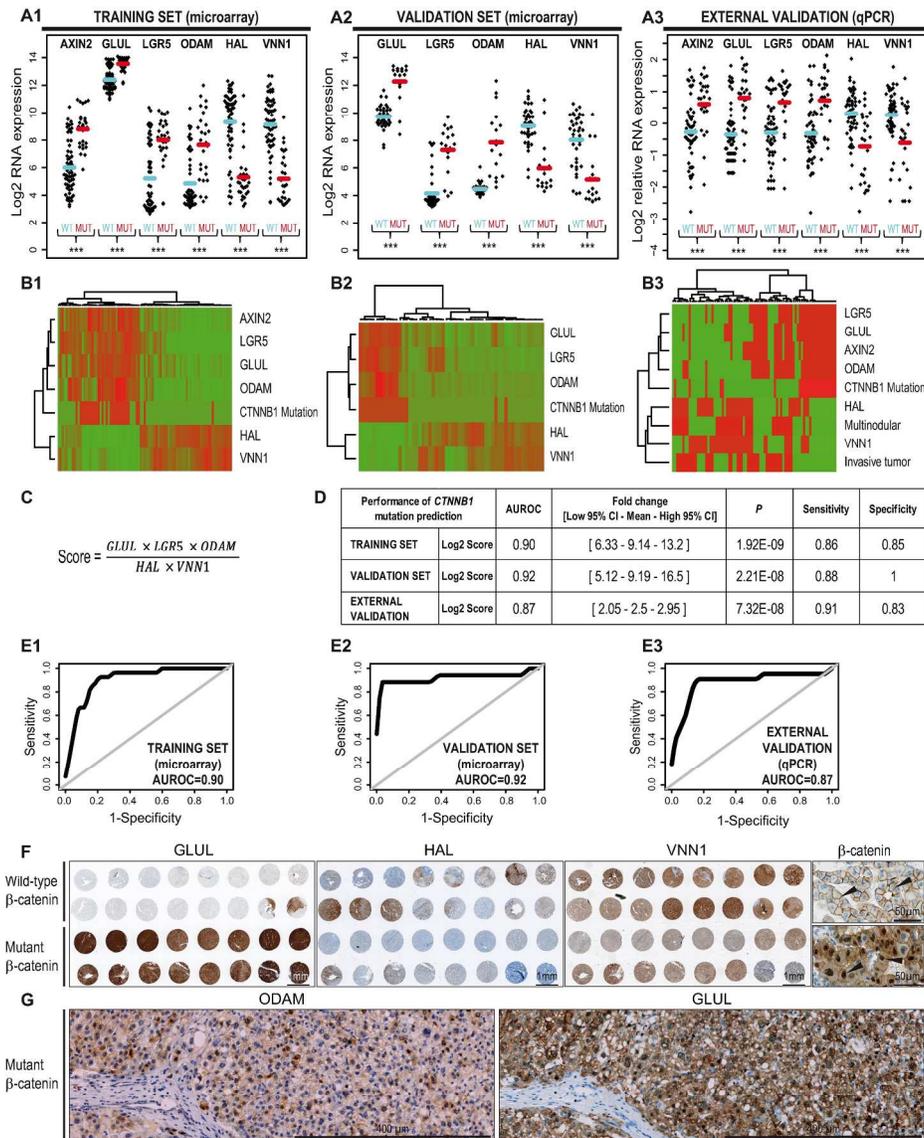
39. Aleman G, Ortiz V, Contreras AV, Quiroz G, Ordaz-Nava G, Langley E, Torres N, et al. Hepatic amino acid-degrading enzyme expression is downregulated by natural and synthetic ligands of PPARalpha in rats. *The Journal of nutrition* 2013;143:1211-1218.
40. Braeuning A, Ittrich C, Kohle C, Hailfinger S, Bonin M, Buchmann A, Schwarz M. Differential gene expression in periportal and perivenous mouse hepatocytes. *The FEBS journal* 2006;273:5051-5061.
41. Chen S, Zhang W, Tang C, Tang X, Liu L, Liu C. Vanin-1 is a key activator for hepatic gluconeogenesis. *Diabetes* 2014;63:2073-2085.
42. Yamashita T, Kitao A, Matsui O, Hayashi T, Nio K, Kondo M, Ohno N, et al. Gd-EOB-DTPA-enhanced magnetic resonance imaging and alpha-fetoprotein predict prognosis of early-stage hepatocellular carcinoma. *Hepatology* 2014;60:1674-1685.
43. Ueno A, Masugi Y, Yamazaki K, Komuta M, Effendi K, Tanami Y, Tsujikawa H, et al. OATP1B3 expression is strongly associated with Wnt/beta-catenin signalling and represents the transporter of gadoteric acid in hepatocellular carcinoma. *Journal of hepatology* 2014;61:1080-1087.
44. Kim SM, Leem SH, Chu IS, Park YY, Kim SC, Kim SB, Park ES, et al. Sixty-five gene-based risk score classifier predicts overall survival in hepatocellular carcinoma. *Hepatology* 2012;55:1443-1452.
45. Minguéz B, Hoshida Y, Villanueva A, Toffanin S, Cabellos L, Thung S, Mandeli J, et al. Gene-expression signature of vascular invasion in hepatocellular carcinoma. *Journal of hepatology* 2011;55:1325-1331.
46. Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, Lee NP, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nature genetics* 2012;44:765-769.

47. Tsuchiya M, Parker JS, Kono H, Matsuda M, Fujii H, Rusyn I. Gene expression in nontumoral liver tissue and recurrence-free survival in hepatitis C virus-positive hepatocellular carcinoma. *Molecular cancer* 2010;9:74.
48. Villa E, Critelli R, Lei B, Marzocchi G, Camma C, Giannelli G, Pontisso P, et al. Neoangiogenesis-related genes are hallmarks of fast-growing hepatocellular carcinomas and worst survival. Results from a prospective study. *Gut* 2016;65:861-869.
49. Villanueva A, Hoshida Y, Battiston C, Tovar V, Sia D, Alsinet C, Cornella H, et al. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology* 2011;140:1501-1512 e1502.



Désert et al., Figure 1

Fig. 1. Flowchart of the study. (A) Construction of a transcriptomic metadata set for subclass-specific analyses after curative resection of 1133 HCCs. Nine public transcriptomic datasets (A to I) were merged into a metadata set. As the 1133-HCCs were not annotated for CTNNB1 mutations, we developed a CTNNB1 mutation prediction score from transcriptomic data. The most stable markers were identified by sPLS-DA modeling in training and validation sets. Mutation prediction was validated in an external in-house HCC collection(9); then, CTNNB1 mutations were predicted in 1133 HCCs after cross-platform normalization. Subclass-specific survival analyses were performed in dataset C (n= 247 HCCs)(34) and validated in an external RNAseq, full-genome sequenced, dataset (LIHC-US; TCGA consortium, n= 210 HCCs).(10) (B) Transcriptomic datasets used: 1, yes or present; 0, no or absent. References: Training set(3); Validation set(5); A(45); B(46); C(34); D(47); E(48); F(49); G(50); H(50); I(28); RNAseq dataset(10). Major endpoints are indicated by bold arrows. External validation sets are represented by black boxes.



Désert et al., Figure 2

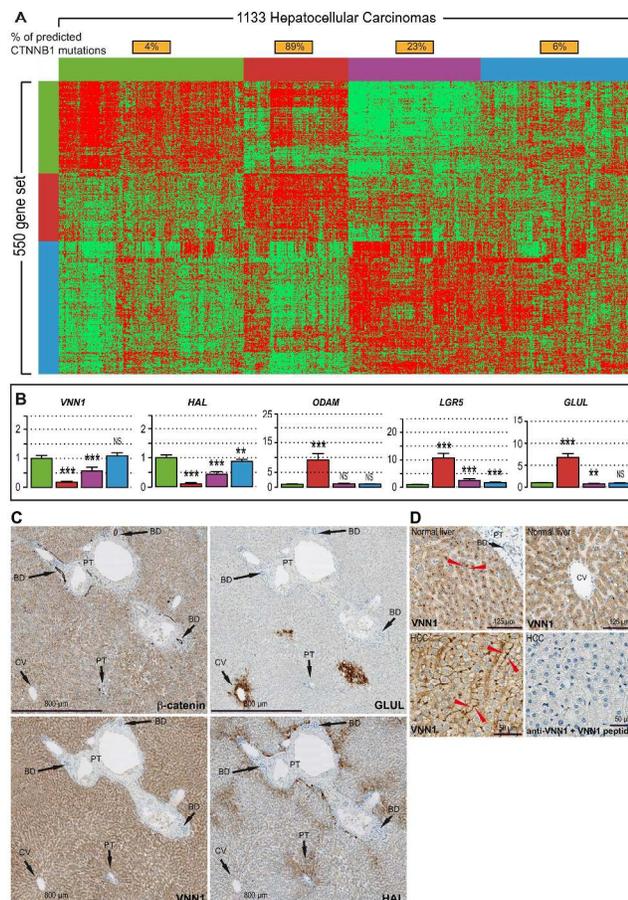
Fig. 2. A 5-gene score reliably predicts CTNNB1 mutations using mRNA expression data in HCCs. (A) Gene expression of the five biomarkers in the training (n=87), validation (n=56) and the external validation (n=82) sets, according to Sanger-sequenced CTNNB1 mutational status. AXIN2 was included as a Wnt pathway activation marker (missing in A2). (B) Heatmaps showing mRNA expression of the five biomarkers plus AXIN2 and CTNNB1 mutational status (B1, training set; B2, validation set; B3, external HCC collection). Low and high gene expressions (median as cut-off) are represented in green and red, respectively. (C) Formula; (D) Performance and (E) ROC curves of the CTNNB1 mutation prediction score. (F) Tissue micro array-based immunostainings (brown) for GLUL, HAL and VNN1 in 20 CTNNB1 mutant and 20 CTNNB1 wild-type HCCs. Insets on the right: β -catenin cell membrane staining in CTNNB1 wild-type (arrowheads, upper right), or cytoplasmic and nuclear staining in CTNNB1 mutant HCCs (arrowheads, lower right). (G) Immunostaining (brown) for ODAM and GLUL in serial sections in a CTNNB1 mutant HCC. Images were acquired with a tissue slide scanner at 20X original magnification. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (Wilcoxon tests). AUROC, Area under the receiver operating curve; qPCR, real-time PCR (a pool of five

normal liver samples was used as calibrator); CI, Confidence interval; Se, Sensitivity; Sp, Specificity.

219x282mm (300 x 300 DPI)

Accepted Article

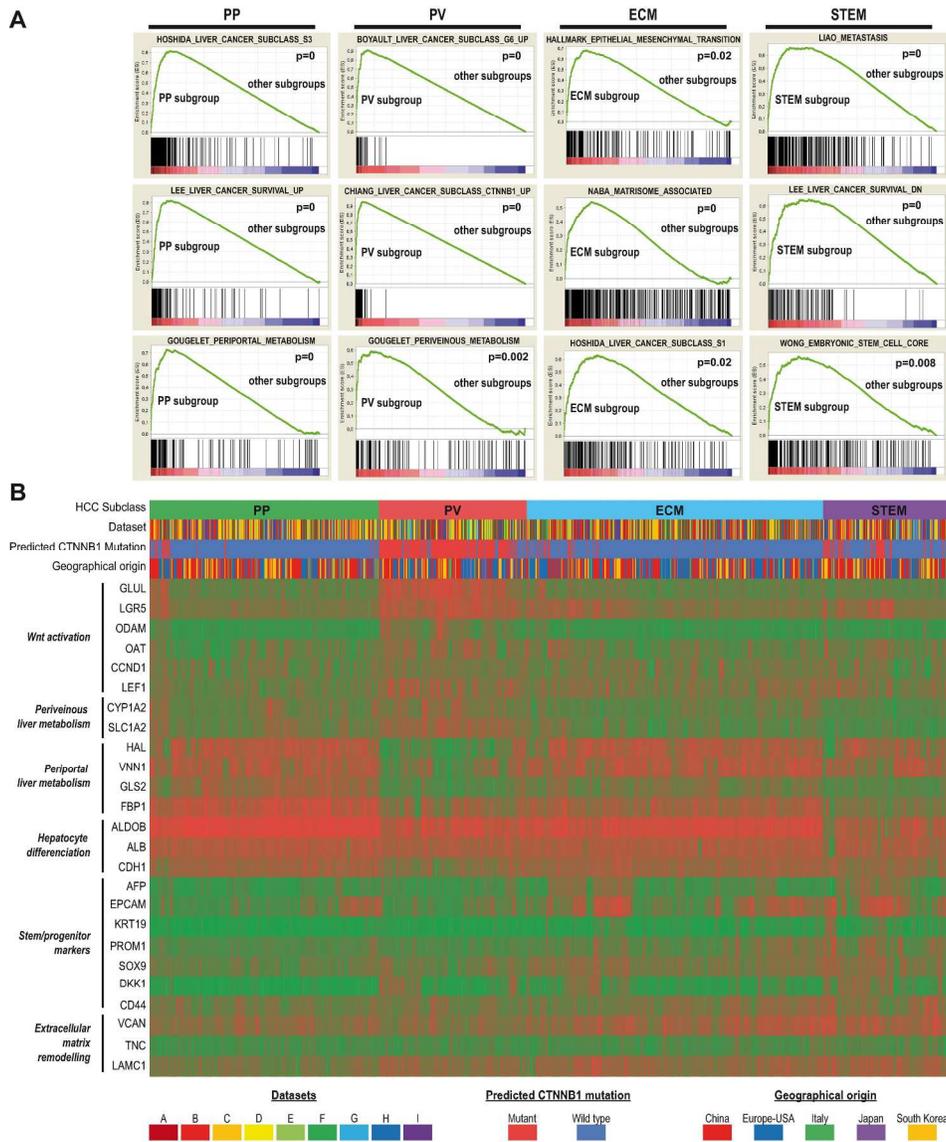
A



Désert et al., Figure 3

Fig. 3. (A) Hierarchical clustering of transcriptomic metadata from 1133 HCCs reveals four tumor subgroups (green, red, purple, & blue) with different proportions of predicted CTNNB1 mutations. Lowest and highest gene expression values are represented in green and red, respectively. (B) Gene expression levels of the five CTNNB1 mutation predictors. Results are relative to the green subclass, arbitrarily set to $1 \pm 95\%$ CI. Green, $n=326$; Red, $n=210$; Blue, $n=420$; Purple, $n=177$. NS, non-significant. $**P<0.01$, $***P<0.001$, after Student's t tests. (C) Immunostainings (brown) of β -catenin, GLUL, VNN1 and HAL in serial sections of normal human liver illustrates clear-cut perivenous and periportal zonation of GLUL and HAL, respectively. β -catenin is predominantly membranous and highlights bile ducts (BD). Arrows show portal tracts (PT) and central veins (CV). VNN1 is panlobular. (D) VNN1 is detected in bile canaliculi in normal liver and in a well-differentiated HCC. VNN1 signal is totally suppressed by the immunogen polypeptide.

304x463mm (300 x 300 DPI)

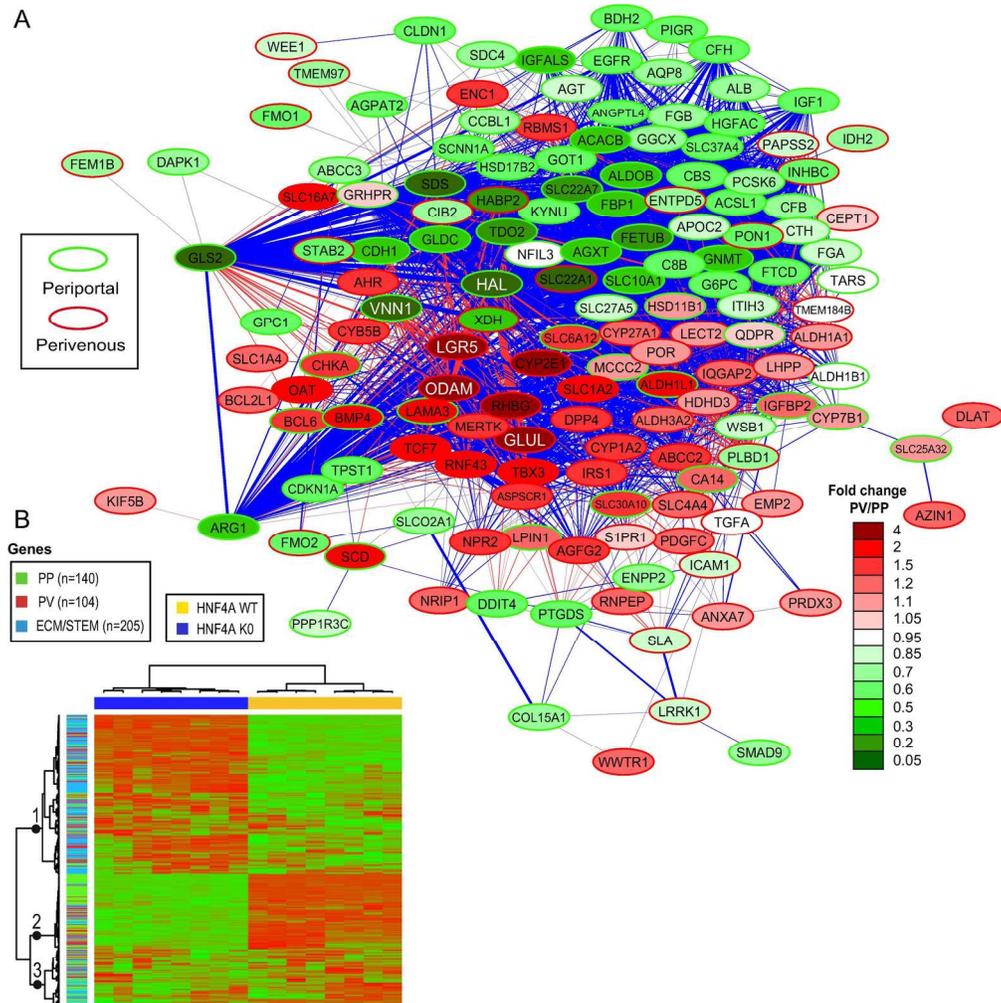


Désert et al., Figure 4

Fig. 4. Well-differentiated HCCs display a preserved metabolic liver zonation program. (A) Gene Set Enrichment Analysis (GSEA) in the four HCC subclasses. (B) Expression of relevant genes in the 1133-HCC metadata set. Color code keys are shown at the bottom of the heatmap. Samples (columns) are ordered by HCC subclass. Green, low; red, high expression. PP, Periportal-type; PV, Perivenous-type; ECM, Extracellular matrix-type; STEM, Stem cell-type.

212x269mm (300 x 300 DPI)

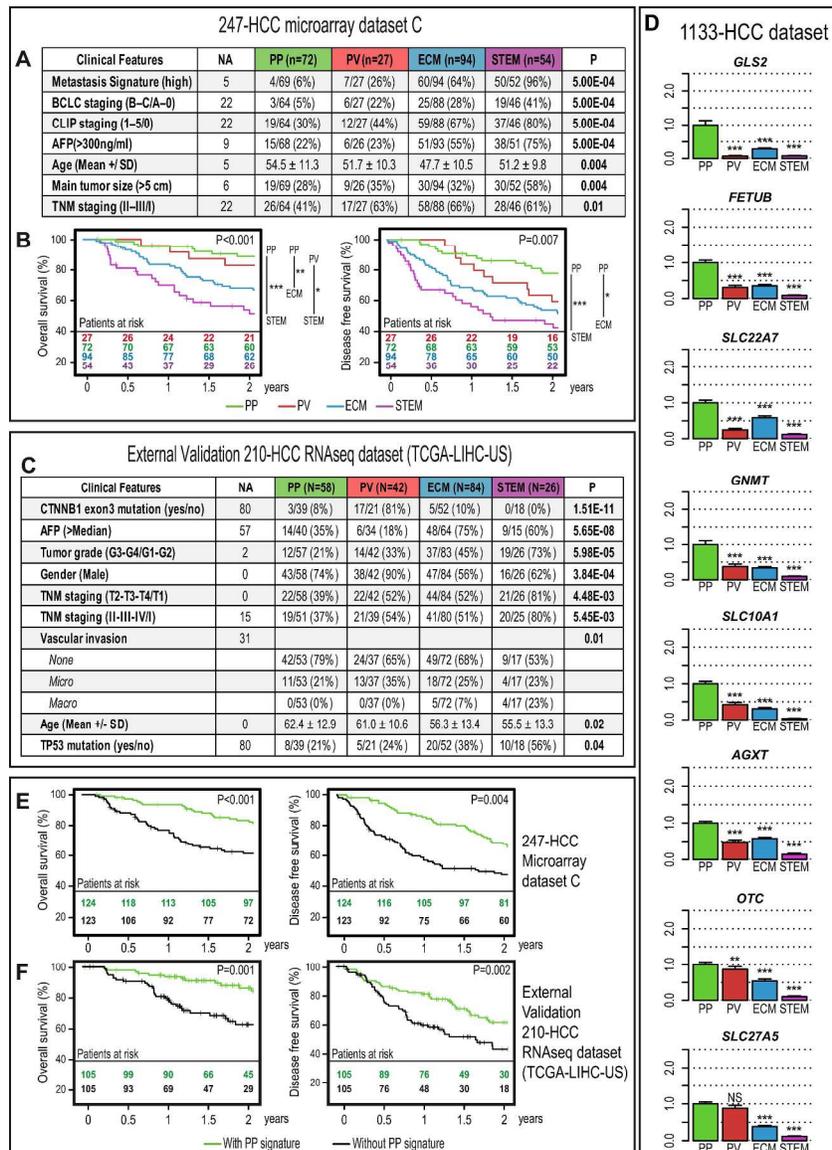
A



Désert et al., Figure 5

Fig. 5. The metabolic program of Periportal-type HCCs is regulated by HNF4A. (A) Periportal (PP, green) and Perivenous (PV, red) gene networks constructed by Weighted Gene Correlation Analysis and shown with Cytoscape graphics. Networks consist of human orthologs of mouse liver periportal and perivenous gene signatures, (20) detected in 326 Periportal-type and 210 Perivenous-type HCCs from the 1133 HCC set. Node border colors (Green/red) denote periportal/perivenous attributes of mouse genes genes, respectively. (20) Node core colors are proportional to the PV/PP fold-change in the HCC set (Color key: lower right). Link thickness is proportional to correlation coefficients (>0.30 in all networks; blue, positive; red, negative correlation). (B) The 550 genes representative of the four HCC subclasses discriminate between Hnf4a-KO and Hnf4a-WT mice. (31). Three gene clusters (black nodes on the left) denote: 1, upregulation in Hnf4a-KO's and enrichment in human HCC STEM/ECM signatures; 2, upregulation in Hnf4a-WT's and enrichment in HCC PP genes; 3, genes poorly affected by Hnf4a status. Red, high expression, green, low expression.

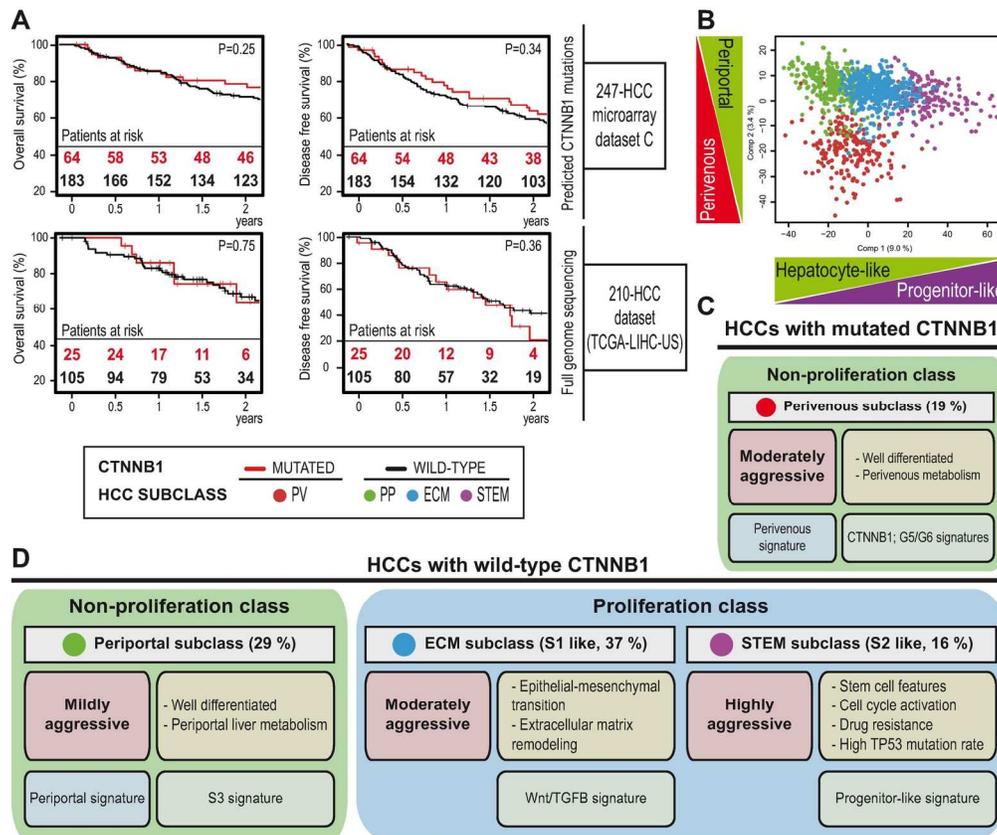
179x191mm (300 x 300 DPI)



Désert et al., Figure 6

Fig. 6. Periportal-type HCCs show the most favorable clinical features and the highest early (2-year) disease-free and overall survival rates after resection. (A) Clinical features of HCC subclasses in a 247-patient dataset(34). (B) Kaplan-Meier plots of subclass-specific overall and disease-free survival; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. (C) Subclass-specific clinical features, CTNNB1 and TP53 mutation rates (full genome sequencing) in an external validation 210-HCC RNAseq dataset (TCGA-LIHC-US). (D) Subclass-specific mRNA expression levels of Periportal HCC signature genes in 1133 HCCs. (E, F) Kaplan-Meier plots of overall and disease-free survival in HCC patients with and without the Periportal HCC signature in two datasets. AFP, serum alpha-fetoprotein; NA, not available; BCLC (Barcelona Clinic Liver Cancer); CLIP (Cancer of the Liver Italian Program). Statistics: Fisher exact test (categorical variables); Student's t test (continuous variables); Log-rank test (survival analyses).

227x327mm (300 x 300 DPI)



Désert et al., Figure 7

Fig. 7. Periportal-type HCCs and the survival paradox between CTNNB1-mutated and wild-type HCCs. (A) Kaplan-Meier plots and log-rank tests in two independent HCC datasets confirm similar outcome of HCCs carrying mutated (red) or wild-type (black) CTNNB1. Mutational status results from prediction in dataset C and full-genome sequencing in TCGA-LIHC-US. (B) Multivariate discriminant analysis of the four HCC subclasses. In the first component, there is a grading from Periportal-type (green), through Perivenous-type (red) and ECM-types (blue), to STEM-type (purple). In the second component, Perivenous-type differs from other HCC subclasses. (C) HCCs with predominantly mutated CTNNB1 are moderately aggressive. (D) HCCs with predominantly wild-type CTNNB1 include mildly, moderately and highly aggressive subclasses. The % of HCCs in each subclass is indicated in parentheses.

155x139mm (300 x 300 DPI)

ACI

Table 1. Univariate and multivariate Cox's 2-year disease-free survival analyses in two independent HCC datasets.

Variables	Missing values	N (%)	Univariate analysis		Multivariate analysis	
			HR (95% CI)	Wald P value	HR (95% CI)	Wald P value
<i>247-HCC microarray dataset C</i>						
Sex (male)	5	211 (87%)	2.38 (1.25-4.53)	8.41E-03	2.25 (1.16-4.33)	0.01
Age (> 60 y)	5	46 (19%)	1.02 (0.66-1.56)	0.94		
HBV viral status (AVR)	23	58 (26%)	1.34 (0.91-1.96)	0.14		
ALT (>50 U/L.)	5	100 (41%)	1.4 (1-1.96)	0.05		
Main tumor size (>5 cm)	6	88 (37%)	1.38 (0.97-1.94)	0.07	1.14 (0.79-1.65)	0.47
Multinodular	5	52 (21%)	1.33 (0.9-1.98)	0.15	1.10 (0.73-1.66)	0.63
Cirrhosis	5	223 (92%)	1.99 (0.93-4.27)	0.08		
AFP (>300ng/ml)	9	110 (46%)	1.31 (0.93-1.83)	0.12	1.22 (0.85-1.74)	0.26
Periportal 8-gene signature	0	124 (50%)	0.62 (0.45-0.88)	6.37E-03	0.69 (0.48-0.97)	0.03
<i>External validation 210-HCC RNAseq dataset (TCGA-LIHC-US)</i>						
Sex (male)	0	144 (69%)	0.84 (0.55-1.27)	0.40		
Age (> 60 y)	0	104 (50%)	0.8 (0.53-1.19)	0.26		
Patients from Asia	6	71 (35%)	0.96 (0.63-1.47)	0.87		
Patients from USA	6	101 (50%)	1.09 (0.73-1.63)	0.67		
Obesity	18	32 (17%)	0.85 (0.48-1.51)	0.59		
Tumor sample weight (>median)	0	69 (33%)	1.73 (1.11-2.7)	0.02	0.98 (0.54-1.8)	0.96
Vascular invasion	31	55 (31%)	1.68 (1.05-2.69)	0.03	1.52 (0.92-2.5)	0.10
Child pugh (>A)	68	15 (11%)	1.44 (0.65-3.18)	0.36		
AFP (>median)	57	76 (50%)	1.16 (0.72-1.85)	0.55		
Platelet count (>median)	34	87 (49%)	1.49 (0.95-2.34)	0.08		
Hepatic inflammation of adjacent tissue	68	76 (54%)	1.03 (0.64-1.65)	0.92		
Performance Status (>0)	40	77 (45%)	1.53 (0.99-2.35)	0.05		
CTNNB1 exon3 mutation	80	25 (19%)	1.33 (0.74-2.36)	0.34		
TP53 mutation	80	43 (33%)	1.4 (0.84-2.32)	0.20		
Periportal 8-gene signature	0	105 (50%)	0.53 (0.36-0.8)	0.003	0.59 (0.37-0.94)	0.02