



Integrated genomic analysis identifies driver genes and cisplatinresistant progenitor phenotype in pediatric liver cancer

Theo Z Hirsch, Jill Pilet, Guillaume Morcrette, Amelie Roehrig, Benedict Je Monteiro, Laura Molina, Quentin Bayard, Eric Trepo, Lea Meunier, Stefano Caruso, et al.

► To cite this version:

Theo Z Hirsch, Jill Pilet, Guillaume Morcrette, Amelie Roehrig, Benedict Je Monteiro, et al.. Integrated genomic analysis identifies driver genes and cisplatinresistant progenitor phenotype in pediatric liver cancer. Cancer Discovery, 2021, pp.candisc.1809.2020. 10.1158/2159-8290.CD-20-1809 . inserm-03212814

HAL Id: inserm-03212814

<https://inserm.hal.science/inserm-03212814>

Submitted on 30 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Integrated genomic analysis identifies driver genes and cisplatin-resistant progenitor phenotype in pediatric liver cancer

Theo Z Hirsch^{1*}, Jill Pilet^{1*}, Guillaume Morcrette^{1,2*}, Amélie Roehrig¹, Benedict JE Monteiro¹, Laura Molina^{1,3}, Quentin Bayard¹, Eric Trépo^{1,4}, Léa Meunier¹, Stefano Caruso¹, Victor Renault⁵, Jean-François Deleuze^{5,6}, Brice Fresneau⁷, Christophe Chardot⁸, Emmanuel Gonzales⁹, Emmanuel Jacquemin⁹, Florent Guerin¹⁰, Monique Fabre¹¹, Isabelle Aerts¹², Sophie Taque¹³, Véronique Laithier¹⁴, Sophie Branchereau¹⁵, Catherine Guettier¹⁶, Laurence Brugières⁷, Sandra Rebouissou¹, Eric Letouzé^{1**} & Jessica Zucman-Rossi^{1,17**}.

¹ Centre de Recherche des Cordeliers, Sorbonne Université, Université de Paris, INSERM, F-75006, Paris France.

² Department of Pediatric Pathology, APHP, Robert Debré Hospital, 75019 Paris, France.

³ Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA.

⁴ Department of Gastroenterology, Hepatopancreatology and Digestive Oncology, C.U.B. Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium.

⁵ Laboratory for Bioinformatics, Fondation Jean Dausset - CEPH, 75010 Paris, France.

⁶ Université Paris-Saclay, Commissariat à l'Énergie Atomique et aux Énergies Alternatives, Centre National de Recherche en Génomique Humaine, Évry, France.

⁷ Gustave Roussy, Université Paris-Saclay, Department of Children and Adolescents oncology, Villejuif, F-94805, France

⁸ Université de Paris, APHP, Necker Hospital, F-75015 Paris, France.

⁹ Pediatric Hepatology and Liver Transplantation Unit, National Reference Centre for Rare Pediatric Liver Diseases, FILFOIE, ERN RARE LIVER, APHP, Bicêtre University Hospital, University of Paris Saclay, Le Kremlin Bicêtre, and INSERM UMR_S 1193, Hepatinov, University of Paris Saclay, Orsay, France.

¹⁰ Department of pediatric surgery, Bicêtre Hospital, APHP, Paris-Saclay University, France.

¹¹ Department of Pathology, Hôpital Universitaire Necker-Enfants malades, AP-HP.Centre – Université de Paris, 75743 Paris Cedex 15, France - Université Paris Descartes.

¹² Oncology Center SIREDO, Institut Curie, PSL Research University, Paris, France.

¹³ Département de Pédiatrie, CHU Fontenoy, 35000, Rennes, France.

¹⁴ Department of children oncology, Centre Hospitalier Universitaire Besançon, 25030 Besancon, France.

¹⁵ Department of pediatric surgery, Bicêtre Hospital, APHP, Paris-Saclay University, France.

¹⁶ Department of Pathology Hôpital Bicêtre - APHP, 94270 Le Kremlin-Bicêtre- INSERM U1193 _ Paris Saclay University, France.

¹⁷ Hôpital Européen Georges Pompidou, F-75015, Assistance Publique-Hôpitaux de Paris, Paris, France.

* These authors contributed equally to this work.

** These authors jointly supervised this work.

Running title: Driver mechanisms and cisplatin-resistant phenotype in PLC.

Keywords: Pediatric liver cancer; Integrated genomic analysis; Chemotherapy resistance; Cancer stem cells; Metastasis.

Correspondence should be addressed to Jessica Zucman-Rossi: +33 6 01 07 78 75
15 rue de l'Ecole de Médecine, 75006 Paris, France, jessica.zucman-rossi@inserm.fr

Financial support: This work was supported by France Génomique (GEPELIN project), Institut National du Cancer (INCa, PELICAN.resist project), SFCE, association Etoile de Martin, Fédération Enfants et Santé (FES), association Hubert Gouin « Enfance et Cancer », INSERM with the HTE (plan Cancer), Cancéropôle Ile-de-France « ThéBioCaPe » project, Inserm ITMO Cancer 3R “PEILICANS” project. The group is supported by the Ligue Nationale contre le Cancer (Equipe Labellisée), Labex OncoImmunology (investissement d’avenir), grant IREB, Coup d’Elan de la Fondation Bettencourt-Shueller, the SIRIC CARPEM, FRM prix Rosen, Ligue Contre le Cancer Comité de Paris (prix René et André Duquesne) and Fondation Mérieux. TZH was supported by a fellowship from Cancéropôle Ile de France, Fondation d’Entreprise Bristol-Myers Squibb pour la Recherche en Immuno-Oncologie and CARPEM, GM was supported by CARPEM, the Ligue Nationale contre le Cancer and Assistance Publique Hôpitaux de Paris. QB was funded by a fellowship from the Ministry of Education and Research. This work was supported by the Fondation pour la Recherche Médicale, grant number "ECO201906008977", to AR and grant number "ECO20170637540" to JP. LaM was funded by the Chateaubriand Fellowship from the Embassy of France in the USA. SC was funded by Labex OncoImmunology and CARPEM. HEPATOBIO was supported by SFCE, Etoile de Martin, Enfant Sans Cancer, INCA/DGOS (Grant INCA-TRANSLA N° 2013-209).

Conflict of interest disclosure statement: The authors declare no potential conflicts of interest.

Word count: 4907

Main figures: 6

Supplementary figures: 17

Supplementary tables: 11

ABSTRACT

Pediatric liver cancers (PLCs) comprise diverse diseases affecting infants, children and adolescents. Despite overall good prognosis, PLCs display heterogeneous response to chemotherapy. Integrated genomic analysis of 126 pediatric liver tumors showed a continuum of driver mechanisms associated with patient age, including new targetable oncogenes. In 10% of hepatoblastoma patients, all before 3 years old, we identified a mosaic premalignant clonal expansion of cells altered at the 11p15.5 locus. Analysis of spatial and longitudinal heterogeneity revealed an important plasticity between 'Hepatocytic', 'Liver Progenitor' and 'Mesenchymal' molecular subgroups of hepatoblastoma. We showed that during chemotherapy, 'Liver Progenitor' cells accumulated massive loads of cisplatin-induced mutations with a specific mutational signature, leading to the development of heavily mutated relapses and metastases. Drug screening in PLC cell lines identified promising targets for cisplatin-resistant progenitor cells, validated in mouse xenograft experiments. These data provide new insights into cisplatin resistance mechanisms in PLC and suggest alternative therapies.

SIGNIFICANCE

Pediatric liver cancers (PLCs) are deadly when they resist to chemotherapy, with limited alternative treatment options. Using a multiomics approach, we identified PLC driver genes and the cellular phenotype at the origin of cisplatin resistance. We validated new treatments targeting these molecular features in cell lines and xenografts.

INTRODUCTION

Pediatric liver cancers (PLCs) are rare tumors and therefore have not been molecularly characterized in large series. Hepatoblastomas (HB) represent about 67%-80% of all pediatric liver cancers worldwide, generally developing before 5 years of age on non-fibrotic liver (1,2). While some rare syndromes such as Familial Adenomatous Polyposis, Beckwith-Wiedemann syndrome or Simpson Golabi Behmel predispose to HB development, the etiology of HB is poorly understood since most of these tumors are sporadic. HB are characterized by their histological heterogeneity, with 3 main histology patterns - fetal, embryonal and mesenchymal - that often coexist within a single tumor. A handful of genomics studies (3-7) have established HB as genetically simple tumors, with the smallest mutation burden among 24 pediatric cancer types (8). Beyond activating *CTNNB1* (β -catenin) alterations found in most HB (70-90%), only few recurrent driver mutations have been described, including *NFE2L2* (5-10%) and *TERT* promoter mutations (2-5%), and no potentially druggable event. Pediatric hepatocellular carcinomas (HCC, incidence=0.24-0.65 per 1,000,000) resemble adult HCC in their histology and frequently develop on fibrotic/cirrhotic liver as a consequence of hepatitis B virus infection or rare congenital disorders (1,9). Previous genomic studies identified recurrent alterations in Wnt signaling and telomerase pathways, but no oncogenic driver in pediatric HCC with underlying liver disease (10,11). Fibrolamellar carcinomas (FLC), a rare subtype of HCC, develop in adolescent and young adults on healthy liver and have been characterized by a recurrent *DNAJB1-PRKACA* driver gene fusion (12). Finally, benign lesions such as hepatocellular adenomas (HCA) and focal nodular hyperplasias are usually related to congenital malformation associated with vascular abnormalities, metabolic genetic diseases or occur after chemotherapy (2,13).

HB are usually treated by cisplatin-based neo-adjuvant chemotherapy and subsequent surgical removal of the tumor, leading to >80% 5-years survival (14,15). However, some HB develop resistance to chemotherapy during the initial neo-adjuvant chemotherapy or after tumor recurrence, and the molecular determinants of cisplatin resistance are yet to be discovered. In contrast to HB, pediatric HCC respond poorly to chemotherapy and as in adults, they have a poor prognosis if not completely removed by surgery.

We aimed to establish the detailed driver landscape of 126 pediatric liver tumors, then we analyzed the plasticity of HB tumors in relation to cisplatin resistance, and explored new therapeutic strategies to overcome this resistance.

RESULTS

The driver landscape of pediatric liver cancers (PLC)

We analyzed a cohort of 126 pediatric liver tumors comprising 104 hepatoblastomas (HB) that developed in 65 patients (87 primary tumors including 18 pre-chemotherapy, and 17 relapses/metastases), 10 hepatocellular carcinomas (HCC, 9 patients), 7 fibrolamellar carcinomas (FLC) and 5 hepatocellular adenomas (HCA) sequenced by Whole Genome (WGS, 65), Whole Exome (WES, 57), RNA (RNAseq, 120) and Reduced Representation Bisulfite (RRBS, 92) sequencing (**Supplementary Fig. S1** and **Supplementary Table S1**). Among the 65 HB patients, 14 were older than 5 years old at diagnosis. HCC developed in fibrotic or cirrhotic liver related to various constitutional liver diseases (tyrosinemia, mitochondrial cytopathy, progressive familial intrahepatic cholestasis, associated to germline mutations of *FAH*, *NDUFA11*, *NDUFB9*, *TJP2*, *ABCB11* **Supplementary Table S2**) while the other tumors arose on normal liver. Germline truncating mutations of *BRCA1* and *BRCA2* were also identified in 2 HB and 1 FLC. In WGS analyses, primary HB displayed a small number of somatic mutations (median=886, 0.3 mutations/Mb) whereas HB metastases and relapses showed a massive mutation load (median=12,824, 4.3 mutations/Mb) with a high proportion of doublet-base substitutions (8%, **Fig. 1a**). Overall, HCC displayed a higher mutation rate than HB (median 5,318, $P=0.002$)

The Wnt/ β -catenin pathway was the most frequently altered oncogenic pathway in PLCs (84.5%), with different activating mechanisms across diagnoses. In HB, *CTNNB1* alterations activating β -catenin were identified in 92% of the tumors (**Supplementary Tables S3-4**), with missense mutations exclusively observed in young patients (<4y) whereas exon 3 inframe deletions were observed later in life (**Fig. 1b** and **Supplementary Fig. S2**). The remaining 5 HB patients without *CTNNB1* mutation showed germline truncating mutations of *APC* (n=3) or *AXIN1* (n=1) with somatic inactivation of the second allele in the tumor, and only one HB remained without an identified alteration in the pathway. *CTNNB1* was also altered in 40% of HCA. In HCC, no *CTNNB1* mutations were detected, but bi-allelic inactivation of *AXIN1* and/or *AMER1* occurred in 6/9 patients (67%).

The 11p15.5 imprinted locus, containing the *IGF2* oncogene, was the second most frequently altered locus in HB and HCC (84% and 89%, respectively), mostly through copy-neutral loss of heterozygosity (cn-LOH, 51-56%, **Fig. 1c**). Adding to LOH, we found epimutations of the imprinting control regions IC1 (gain of methylation, 22% of HB and 33% of HCC) and IC2 (loss of methylation, 5% of HB), and recurrent somatic mutations of *CDKN1C* in 4 HB patients. In 6 HB patients (10%), imbalanced B allele frequency profiles revealed that the cn-LOH found in tumor cells was present as a mosaic in the adjacent normal liver (**Fig. 2a-b**). These young patients (median=8.4 months, $P=0.045$) were not diagnosed with Beckwith–Wiedemann syndrome (BWS). Yet, the cn-LOH of 11p15.5 was detected in a significant fraction of normal liver cells (6%-58%) without the other driver alterations identified in the corresponding tumor, indicating pre-malignant clonal expansions of hepatocytes. Oncogenic transformation involved the acquisition of *CTNNB1* missense activating mutation in the 6 cases (**Fig. 2b**). Accordingly, *IGF2* was highly overexpressed in the non-tumor liver samples of two patients harboring a mosaic 11p15.5 alteration whereas Wnt/ β -catenin target genes were only overexpressed in tumor cells

(**Fig. 2c**). In patient #3559 (30% mosaic cn-LOH in the non-tumor liver), RNAscope *in situ* hybridization revealed a massive over expression of IGF2 in both the tumor and adjacent non-tumor tissue, whereas Glutamine synthetase and β -catenin immunostainings demonstrated an oncogenic activation of Wnt/ β -catenin pathway specifically in tumor cells (**Fig. 2d**). These data support the idea that premalignant clonal expansions of “normal” hepatocytes with 11p15.5 cn-LOH overexpress IGF2 and can lead to HB formation after oncogenic *CTNNB1* mutation. Finally, one patient with BWS (#3180) displayed IC2 mosaic epimutation in both blood cells and the liver (27% of cells). Overall, the mechanism of 11p15.5 alteration in PLC was related to age, with cn-LOH occurring more frequently in young patients and gain of methylation of IC1 mostly observed in older patients (**Fig. 1b** and **Supplementary Fig. S2**).

A paucity of structural variants (SV) was observed in primary HB (median=6/tumor) except for 6 tumors with chromoplexy, which mostly developed in older patients (**Fig. 1a** and **Supplementary Table S5**). In HCC, we identified an unusual number of focal deletions (median=47/tumor) of small size (median=27 kb). This remarkable deleter phenotype led to recurrent cancer driver alterations, particularly on chromosome X with complete inactivation of *AMER1* (44%), *GPC3* (44%), *RPS6KA3* (33%), *SMARCA1* (22%) and *BCORL1* (11%), but also on chromosome 16 with homozygous deletion or combined deletion and truncating mutation of *AXIN1* (44%) and *CREBBP* (11%) (**Fig. 1b-c** and **Supplementary Fig. S3**). Of note, *GPC3*, *BCORL1*, *CREBBP*, *RPS6KA3* and *AXIN1* were also inactivated in HB through truncating or damaging mutations, including one germline *GPC3* mutation in a patient with Simpson-Golabi syndrome. Recurrent copy-number alterations were identified in two genes controlling p53 degradation: focal deletions at 4q35 pinpointed inactivation of the tumor suppressor *IRF2* in 30% of HB and 56% of HCC, whereas focal amplifications at 1q32.1 led to a high expression of *MDM4* in 4 HB patients (**Fig. 1c**, **Supplementary Fig. S3** and **Supplementary Table S6**). Interestingly, gains of the entirety of 1q were also observed in 50% of HB and HCC (**Fig. 1c**). At the chromosome arm level, HCC and HB had a roughly similar profile of gains but HCC harbored more losses, including loss of the 13q arm (encompassing *RB1*) in 33% of cases (**Supplementary Fig. S4**). Finally, focal amplifications of *CCND1/FGF19* were found in 3 HB patients leading to the overexpression of both genes. Other recurrent driver mutations in HB involved *TERT* promoter in 9 patients (all older than 40 months, $P=1.5 \times 10^{-5}$, **Supplementary Fig. S2**), *NFE2L2* (4 pts) and *ARID1A* (2 pts). No recurrent gene fusion was identified in the cohort except the *PRKACA-DNAJB1* fusion pathognomonic of FLC. We also identified recurrent mutations of *ERBB4* in 2/7 FLC (29%) and of *HNF1A* in 3 HCA (60%), including one germline mutation. Overall, HB and HCC shared common pathways altered by diverse genes and mechanisms, while HCA and FLC had specific driver alterations (**Fig. 1b**).

Phenotypic plasticity of hepatoblastoma cells across three differentiation states

Unsupervised transcriptomic classification of 100 HB samples from 64 patients revealed 4 robust molecular groups characterized by diverse differentiation states, cell proliferation and immune infiltration levels (**Fig. 3a**). The differentiated ‘Hepatocytic’ HB group comprised 44 samples, 73% of which had only ‘fetal’ histological component in their corresponding mirror block (versus 10%, $P=1.5 \times 10^{-10}$). They strongly expressed transcription factors (TFs) involved in hepatic differentiation (*HNF1A*, *HNF4A*, **Fig. 3a**).

'Hepatocytic' HB were divided in two clusters: the 'Hepatocytic hot' subgroup was defined by a strong signature of polymorphic immune infiltration, including a mixture of T, B, NK cells and macrophages, with 20/21 of these samples collected after chemotherapy; the 'Hepatocytic cold' subgroup was characterized by lower levels of immune infiltrates, and it was enriched in pre-chemotherapy samples (39%, $P=0.010$, **Fig. 3a** and **Supplementary Fig. S5a-c**). The 'Liver Progenitor' group comprised 44 samples enriched in highly proliferative, immune-cold tumors with embryonal compartments, 11p15.5 alterations and *MDM4* amplifications. They expressed TFs involved in hepatic differentiation but also TFs involved in self-renewal and pluripotency maintenance (*MYCN*, *MIXL1*, *LIN28B*). The last group, 'Mesenchymal' HB, included all samples with a mesenchymal histology which were derived from mixed hepatoblastomas with both epithelial and mesenchymal components, and was found mostly in younger patients (**Supplementary Fig. S2**). These mesenchymal samples displayed a distinct differentiation program, with no expression of liver differentiation genes but a strong expression of mesenchymal stem cell TFs (*TWIST1*, *TBX5*, *MSX2*), and variable levels of immune infiltration. We validated these 4 transcriptomic subgroups and their associations with histology and immune infiltration in an independent RNA-seq cohort of 34 HB samples (16) (**Supplementary Fig. S6a-c**).

Strikingly, 14/24 patients (58%) with multiple synchronous and/or metachronous samples displayed transcriptomic group switches (**Fig. 3b** and **Supplementary Fig. S7**). Spatial transcriptomic heterogeneity was identified in the primary tumors of 7 patients and matched histological heterogeneity between fetal and embryonal parts of the tumor (**Fig. 3b**). Longitudinal transcriptomic group changes were also identified between pre- and post-chemotherapy samples (6 patients), and between primary tumors and relapses/metastases (8 patients). Phylogenetic trees revealed private driver alterations in some cases, but no recurrent gene associated with specific transcriptional group changes. In contrast, DNA methylation profiles were closely associated with differentiation states, and transcriptomic group changes were associated with DNA methylation reprogramming (**Fig. 3c** and **Supplementary Fig. S8a**). In particular, mesenchymal samples displayed coordinated hypermethylation of HNF4 and PPAR binding sites, both in our cohort (**Supplementary Fig. S8b-c**) and in an independent data set (16) (**Supplementary Fig. S6b**).

The molecular plasticity of HB recapitulated the heterogeneity observed at the histological level (**Fig. 3b**). A systematic histological review at the sample (mirror block) and whole tumor levels revealed that 80% of primary HB displayed spatial heterogeneity with a mixture of embryonal, fetal, or mesenchymal areas. The 'Liver Progenitor' molecular group was associated with increased intra-sample histological heterogeneity ($P=1.0 \times 10^{-6}$, **Fig. 3d**), with frequent coexistence of fetal and embryonal cells (**Fig. 3a-b**). Primary tumors with alterations in the 11p15.5 locus also displayed more histological heterogeneity (89% vs. 46%, $P=0.0021$, **Fig. 3e**). Thus, the phenotypic plasticity of HB may relate to the multipotency of progenitor cells with 11p15.5 alteration.

Consistent with this plasticity, the proportion of molecular groups evolved across disease stages ($P=0.0014$, **Fig. 3b**). The 'Hepatocytic hot' group was enriched in post-versus pre-chemotherapy samples (30% vs 6%), whereas the 'Liver Progenitor' group was enriched in metastases and relapses (64% vs 40%). Wnt/ β -catenin alterations were always trunk in phylogenetic trees and already present in 100% of pre-chemotherapy

primary samples. In contrast, 11p15 alterations occurred late in 8 out of the 20 affected phylogenetic trees and their frequency increased in later disease stages, which was also the case for *IRF2* deletions (**Fig. 3b**).

To validate this phenotypic plasticity at the single-cell level, we analyzed matched non-tumor liver, primary HB and lung metastasis from one patient (#3981) by single-nucleus RNA-seq. Virtual copy-number profiles were consistent with the copy-number changes observed in bulk whole exome sequencing data, with gains of 1q, 2q, 5q and 15q identified in all tumor cells, and an extra gain of Xq in all metastatic cells (**Fig. 4a**). In contrast to this genetic homogeneity, we identified clusters of tumor cells with distinct transcriptomic profiles, corresponding to the 'Mesenchymal', 'Liver progenitor' and 'Hepatocytic' signatures (**Fig. 4b**). The 3 populations were present in different proportions in each sample, with a larger 'Mesenchymal' contingent in the primary tumor, consistent with the molecular classification of bulk samples ('Mesenchymal' for the primary, 'Liver Progenitor' for the metastasis). These data demonstrate the existence of the three differentiation states at single-cell level, and the plasticity of hepatoblastoma cells across these differentiation states. Indeed, both the last common ancestor (LCA) of the primary tumor and the LCA of the metastasis (with additional Xq gain) were able to generate the three cell types repeatedly during tumor evolution in this patient (**Fig. 4c**).

Thus, dynamic switches between 3 differentiation states operate in hepatoblastoma cells, resulting in high intra-patient heterogeneity in space and time. This phenotypic plasticity is more frequent in tumors with 11p15.5 alterations and involves transcriptional and epigenetic reprogramming of TF modules. Molecular phenotypes match well with histological cell types and display drastically different immune infiltrates.

Cisplatin resistance results from the expansion of progenitor cell clones acquiring massive mutation load

Mutational signature analysis of 65 pediatric liver cancer genomes identified 4 single base substitution (SBS), 2 doublet base substitution (DBS), and 5 indel signatures, most of which matched signatures previously described in pan-cancer studies (17,18) (**Fig. 5a, Supplementary Fig. S9a-b and Supplementary Table S7**). The majority of signatures, including the clock-like signatures SBS1 and SBS5, were found ubiquitously in the tumors, while the signature SBS18, commonly found in neuroblastoma (17) and related to oxidative DNA damage (19), was identified in a subset of hepatoblastomas. Finally, we identified a massive load of mutations due to the signatures SBS35, DBS5 and ID3 in 20/66 primary HB resected after neoadjuvant chemotherapy (vs. 0/17 HB sampled before cisplatin treatment, $P=0.009$). Consistently, SBS35 and DBS5 are known to reflect the diverse mutation types caused by cisplatin adducts on DNA (18,20). In primary HB after neoadjuvant-chemotherapy, cisplatin mutations were subclonal and almost exclusively found in the 'Liver Progenitor' tumor subgroup (18/27 vs 2/38; $P=1.2 \times 10^{-7}$) (**Fig. 5b-c**). Furthermore, in 4 tumors with spatial heterogeneity, cisplatin signature was restricted to the 'Liver Progenitor' component even though the non-'Progenitor' samples were exposed to the same chemotherapy regimen (**Supplementary Fig. S7**).

Interestingly, all 16 HB relapses and metastases that developed after chemotherapy displayed a deluge of clonal cisplatin-induced mutations (**Fig. 5c**), leading to high mutation loads (median = 4.3 mutations/Mb), comparable to those of adult tumors and 15-times higher than those of primary liver tumors of the same age (**Supplementary**

Fig. S9c-e). We analyzed in detail 8 patients with chemoresistant HB at different stages of the disease including primary tumors, relapses and metastases operated between 4 months and 12 years after chemotherapy. All relapses were clonally related to their matched primary tumors, even the lung metastasis of patient #3538 (germline *APC*), detected 12 years after initial surgery (**Fig. 5d**). Phylogenetic trees revealed that a sudden burst of mutations is acquired during cisplatin treatment. Every relapse/metastasis was derived from a single common ancestor cell that acquired between 5,000 and 13,000 mutations during cisplatin treatment, with various modes of metastatic seeding. In patients #3370 and #3981, SBS35 mutations were private to each sample, indicating that metastatic seeding involved independent resistant cells. In patient #3694, the two relapses displayed a mixture of shared and private SBS35 mutations, indicating successive rounds of cisplatin mutagenesis followed by clonal expansion of resistant cells. Finally, patients #3529 and #3949 displayed late branching of relapse/metastasis samples derived from a same cisplatin-resistant clone. The seeding abilities of these clones is clearly illustrated by patient #3529 whose primary tumor was treated by liver transplantation: the same cisplatin resistant clone gave rise first to a metastasis in the spleen, followed by 2 subsequent metastases on the grafted liver (**Fig. 5d**).

Overall, these data suggest that ‘Liver Progenitor’ cells are able to bypass cisplatin-DNA adducts and proliferate under chemotherapy while accumulating SBS35 mutations. Accordingly, we found an enrichment of genes related to cisplatin resistance (21) or DNA repair among genes overexpressed in ‘Liver Progenitor’ HB (63 genes, $P < 2.2 \times 10^{-16}$). Among those, 20 genes were also up-regulated in post- versus pre-chemotherapy ‘Liver Progenitor’ HB (**Supplementary Fig. S10a-d**), including genes involved in inhibition of apoptosis (*BIRC5*, coding for survivin) and DNA repair through homologous recombination (*BRCA1*, *RAD54L*, *EXO1*), Fanconi anemia (*FANCA*, *FANCB*, *FANCD2*, *FANCI*) or base excision repair pathways (*LIG1*, *LIG3*, *POLE*, *POLE2*). Of note, most of these genes were also overexpressed in fetal liver samples (13th to 30th weeks of amenorrhea) vs. post-natal liver.

The subclonal presence of the signature SBS35 in primary tumors after neo-adjuvant chemotherapy is thus a marker of cisplatin-resistant cell proliferation, and was associated with poor progression-free and overall survival (log-rank $P=0.012$ and $P=0.032$ respectively, **Supplementary Fig. S11**). Heavily mutated resistant cells later give rise to relapses and metastases. A median of 70 coding sequence mutations per relapse/metastasis occurred due to the extra mutation load attributed to cisplatin. While some of these mutations affected known cancer genes including *NF1*, *BRAF* (L485W), *KMT2C*, *KMT2D*, *BCORL1* and *NOTCH1*, no recurrent driver gene associated with tumor progression was identified.

New therapeutic strategies targeting hepatoblastoma based on molecular features

To investigate new therapeutic options, we characterized a panel of 9 pediatric liver cancer cell lines PL-CCL (8 HB, 1 HCC) (**Fig. 6a** and **Supplementary Table S8**). Our panel of cell lines, which are mainly derived from older patients and of ‘Liver Progenitor’ type with multiple alterations in cancer driver genes, reproduced the major driver events identified in PLC. In particular, 3 cell lines with a matched primary or relapsed HB displayed genomic alterations globally similar to the original tumors (**Supplementary Table S9**). All HB cell lines carried *CTNNB1* alteration whereas Hep3B, derived from an

HCC, carried *TP53*, *AXIN1* and *RPS6KA3* alterations with frequent homozygous deletions in agreement with the deleter phenotype that we observed in pediatric HCC. Furthermore, all cell lines exhibited *11p15.5* locus alterations through cn-LOH (5/9), GOM IC1 (3/9) or *CDKN1C* mutation (1/9).

We tested our panel of cell lines with conventional therapeutic regimens (cisplatin, carboplatin and doxorubicin treatments, **Supplementary Table S10**). In the 9 cell lines analyzed, resistance to cisplatin and carboplatin correlated with a high expression of 'Liver Progenitor' markers (**Fig. 6b** and **Supplementary Fig. S12a**). Additionally, we treated 5 cell lines with long term exposure to low doses of cisplatin (0.5 μ M, 4 weeks, **Supplementary Fig. S12b**). Only B6-2 was completely sensitive after 2 weeks. In the 4 remaining cell lines, WES of resistant clones revealed an accumulation of the SBS35 cisplatin mutational signature. These results demonstrate that the SBS35 signature is directly induced by cisplatin exposure in resistant cells (**Fig. 6c**). Interestingly, doxorubicin, which is commonly used in resistant or high-risk HB patients, exhibited a significant anti-tumor effect in all cell lines.

Next, we tested treatments targeting specific genomic alterations and their consequences. Three out of 5 tested cell lines were sensitive to anti-IGF-2 antibodies, and sensitivity to this drug correlated with the level of *IGF2* expression ($P=0.03$, $R=-0.91$, Pearson correlation, **Fig. 6d**). We also tested the monoclonal therapeutic antibody Xentuzumab and we obtained between 14% and 47% growth inhibition (**Supplementary Fig. S12c**), consistent with a previous report in mice xenografts (22). Surprisingly, sensitivity to Xentuzumab was not correlated with *IGF2* expression, possibly because it neutralizes both IGF-1 and IGF-2. We then tested our cell lines with the MEK1/2 inhibitor Trametinib (**Fig. 6e**). Remarkably, 5/9 cell lines were sensitive to Trametinib and harbored alterations within the MAP kinase pathway through *RPS6KA3* homozygous deletions, *NRAS* activating mutation, or *FGF19* amplifications as previously shown in adult HCC cell lines (23). Of note, one of our patients with a metastatic HB (#3538) displayed a hotspot *MAPK1* E322K trunk mutation and acquired a *BRAF* L485W mutation in the metastasis which has previously been associated with response to an ERK inhibitor (24).

To identify new targets in cisplatin-resistant 'Progenitor' cells, we performed a differential expression analysis between the 'Liver Progenitor' and 'Hepatocytic' HB subgroups. Of the genes over-expressed in 'Liver Progenitor' samples, we selected 3 key targetable genes: *PLK1*, *BIRC5* (Survivin), and *CHEK1* which are involved in cell cycle, apoptosis and DNA damage response respectively (**Fig. 6f-g**). In HB patients high expression of *PLK1*, *BIRC5* and *CHEK1* was associated with (1) poor response to chemotherapy as assessed with AFP reduction in the sera ($P=4.1\times10^{-4}$, 2.0×10^{-4} , and 4.7×10^{-4} , respectively), (2) embryonal histology ($P=1.2\times10^{-5}$, 5.1×10^{-6} and 7.1×10^{-6}), (3) the cisplatin resistance signature SBS35 ($P=8.8\times10^{-10}$, 1.3×10^{-9} , 7.6×10^{-9}) and (4) a high expression of proliferation genes ($P\leq2.2\times10^{-16}$) (**Supplementary Fig. S13a-d**). We thus tested inhibitors of PLK1 (BI-2536), Survivin (YM155) and CHEK1 (AZD-7762) in our collection of 9 cell lines. Interestingly, the 3 drugs were more efficient than both cisplatin and carboplatin in all the cell lines except B6-2, which has the least 'Liver Progenitor'-like phenotype. All cell lines were sensitive to BI-2536, including those which are resistant to cisplatin (**Fig. 6h**). Notably, the cell lines most resistant to cisplatin were also the most sensitive to YM155 and AZD-7762 (**Supplementary Fig. S14a-c**). Finally, we tested the most effective drug, BI-2536, *in vivo* in 24 nude mice xenografted with HepG2 cells

randomized in 4 groups (cisplatin, cisplatin+doxorubicin, BI-2536 and control). HepG2 xenografts were resistant to cisplatin treatment, both in isolation and in combination with doxorubicin. In contrast, tumors in mice receiving BI-2536 responded to treatment ($P=0.003$) without significant toxicity unlike with cisplatin (**Fig. 6i-j**).

In conclusion, the PLK1 inhibitor BI-2536 showed efficacy in both *in vitro* and *in vivo* models and appears to be a good candidate drug for the treatment of PLC. Trametinib and anti-IGF-2 antibodies show promise for the treatment of resistant HB with *CCND1/FGF19* amplification or high *IGF2* expression, respectively.

DISCUSSION

Altogether, our findings suggest a major role of cell plasticity in the spatio-temporal evolution of hepatoblastomas and their resistance to cisplatin-based chemotherapy. In this study, genomic analysis of multiple parts of primary tumors and of longitudinal samples at various timepoints allowed us to profile the clonal evolution of hepatoblastoma along treatment. Cisplatin induced hundreds of mutations in progenitor tumor compartments, a phenomenon reproduced in cell lines, and targeting specific oncogenic processes with different inhibitors could be used to target resistant cells.

Primary liver cancers showed a low mutation burden, as seen in other pediatric cancers (25). In PLC, β -catenin activation, mainly caused by *CTNNB1* mutations, acted as the major early and common mechanism of liver tumorigenesis. However, we identified an earliest recurrent premalignant alteration with mosaic cn-LOH of 11p15 in the liver of very young HB patients (10%). This finding is in line with similar pre-malignant clonal expansions recently identified as precursors of Wilms tumors, associated with hypermethylation of IC1 at the 11p15 locus (26). Our data reinforce the role of 11p15 alterations in pre-malignant stages that occur during pregnancy or early in life. Systematic searches of 11p15 locus alterations in the liver and other organs of pediatric cancer patients could reveal additional mosaic cases.

The present molecular transcriptomic classification revealed 4 robust HB subgroups defined by both their differentiation states and immune infiltration levels. We verified that our clustering was not driven by patients with multiple samples (**Supplementary Fig. S15**), and we validated our classification in an independent RNA-seq cohort of 34 HB (**Supplementary Fig. S6**). Our transcriptomic classification overlaps the previously published subtypes C1/C2 (27), C1/C2A/C2B (28) and MRS1/MRS2/MRS3a/MRS3b (16), with a good match between 'Hepatocytic' and C1, 'Liver progenitor' and C2A, and 'Mesenchymal' and C2B subgroups (**Fig. 3a** and **Supplementary Fig. S16**). Our 'Liver progenitor' subgroup also overlaps with the very high-risk MRS-3b subgroup identified by Carrillo-Reixach et al. (16) ($P=2.0 \times 10^{-6}$). The most original findings of our classification are (1) the definition of clear transcriptomic signature of 'Mesenchymal' histological cell type validated by careful histological review of mirror blocks, and (2) the identification of two subgroups of 'Hepatocytic' HB showing varying levels of immune infiltration, which is potentially enhanced by cisplatin-based chemotherapy. HB subgroups are not defined by specific driver events, as opposed to medulloblastoma subgroups (29), for example. In turn, each state of HB differentiation is defined by specific transcription factors and a characteristic DNA methylation landscape. Our methylation-based classification revealed two main HB clusters reminiscent of the previously described Epi-CA/CB groups (16): a normal-like cluster matching the Epi-CA group comprising most 'Hepatocytic' samples, and a differentially methylated cluster matching the Epi-CB group, which we further divided into 3 subgroups associated to 'Liver Progenitor', 'Mesenchymal' phenotype and older age. Importantly, HB cells are able to switch between differentiation states along the course of the disease, leading to striking spatial and temporal heterogeneity. Our single-cell data support the notion that HB cells have the inherent ability to switch across 3 differentiation states. Yet, clinical and molecular features associated with transcriptomic subgroups suggest that the cell of origin, patient age, chemotherapy and driver alterations can favor the predominance of a

given subtype in each tumor (**Supplementary Table S11** and **Supplementary Fig. S17**). This molecular plasticity, driven by specific TF modules and epigenetic landscapes, is reminiscent of neuroblastoma cells that can trans-differentiate between committed adrenergic cells and undifferentiated mesenchymal cells (30,31), the latter being chemoresistant and enriched in relapse tumors. Similarly, we showed that HB ‘Progenitor’ cells proliferate under neo-adjuvant chemotherapy and are enriched in relapses and metastases.

Moreover, PLC development follows evolutionary trajectories that vary according to age, involving similar pathways but diverse mechanisms of activation. In contrast to HB, pediatric HCC occur in older children, on cirrhotic liver due to various causes. Like HB, HCC displayed frequent 11p15.5 alterations, but also a characteristic chromosome deleter phenotype leading to WNT/ β -catenin activation through *AXIN1* or *AMER1* deletions, as well as frequent *GPC3*, *RPS6KA3* and *SMARCA1* alterations. Since a deleter phenotype is only seen in 3% of adult HCC (32), this oncogenic mechanism could be associated with young age. Interestingly, many HCC driver genes are located on the X chromosome (*AMER1*, *RPS6KA3*, *GPC3*, *SMARCA1*), which may partly explain the enrichment of pediatric HCC in male patients described previously (9) and retrieved in the present cohort (7/9 males, 78%). Pediatric HCC are commonly resistant to cisplatin, but their mechanism of resistance likely differs from that involved in HB since they do not exhibit similar progenitor and stem cell features.

‘Progenitor’ cells are able to proliferate under treatment, bypass cisplatin-induced adduct formation, acquire a huge number of cisplatin-induced mutations (median 11,142) with a characteristic SBS35 signature, and give rise to highly mutated relapse tumors. Various DNA polymerases and DNA repair genes involved in double-strand break repair, homologous recombination and Fanconi pathway are already over-expressed both in ‘Liver Progenitor’ HB before chemotherapy and in the fetal liver, showing that these mechanisms are constitutively active in liver progenitor cells. Relatedly, a previous work had shown that Fanconi anemia inhibitors were able to block the growth of HB cells *in vitro* and *in vivo* (28). Pich *et al.* estimated that ~15 of every 1,000 cisplatin-induced mutations affect the sequence of coding genes, of which ~0.7 are expected to affect the sequence of known cancer genes (33). Thus, an average HB relapse or metastasis is at risk of acquiring 167 coding mutations, including 7.8 in cancer genes due to cisplatin treatment. Although most cancer gene mutations are unlikely to play a functional role in HB cells, this increased mutational burden provides an opportunity for resistant cells to acquire additional oncogenic capabilities. These data highlight the necessity to identify alternative treatments for cisplatin-resistant HB, in which more aggressive chemotherapy regimens may just promote the selection of aggressive progenitor cells with a massive extra mutational burden.

Our findings revealed new potential therapies to combat HB resistant to cisplatin. Targeting the different mechanisms involved in cisplatin resistance (21) such as apoptosis (YM155 targeting survivin), DNA repair (AZD-7762 targeting CHEK1) or cell cycle control (BI-2536 as a PLK1 inhibitor) is very promising since an efficient anti-tumor effect was observed in PLC cell lines with a progenitor molecular signature. In the same line, other drugs targeting the proteasome (bortezomib) could also be efficient in resistant HB(28). Immunotherapy could be another appealing type of treatment for HB since we showed that cisplatin treatment can induce an intra-tumor polymorphic immune

response leading to 'Hepatocytic hot' HB, consistent with our previous observation in HB with germline *APC* mutations (34). Reinforcement of an efficient intra-tumor immune response could be improved by the use of oxaliplatin to induce an immune cell death (35) or by treatments with immunomodulators. However, 'Liver Progenitor' HB are immune cold despite showing multiple cisplatin induced mutations. In these cases, penetrating the progenitor compartments with immune cells will be challenging. Ongoing and planned clinical trials aim to test immunomodulators or targeted therapies in high-risk hepatoblastoma; however these trials are limited by the small number of patients.

In conclusion, PLC showed various mechanisms of tumorigenesis related to age and cell of origin. Hepatoblastomas demonstrated a striking spatial and longitudinal phenotypic plasticity related to the progenitor compartment associated with cisplatin resistance and the mutational signature of DNA adduct bypass. And finally, the identification of drugs targeting 'progenitor' cells opens new avenues to treat children at high risk of resistance.

METHODS

Clinical samples

A series of 126 liver tumor samples and their non-tumor counterparts were collected from 86 patients surgically treated in various French hospitals. The study was approved by the local Ethics Committee (CCPRB Paris Saint-Louis). Written informed consent was obtained in accordance with French legislation. All samples were immediately frozen in liquid nitrogen and stored at -80°C . Tumors included in this study comprised 87 primary hepatoblastomas (HB, 18 prior chemotherapy and 69 after chemotherapy) and 17 HB relapses and metastases (from 65 HB patients including 14 older than 5 years old), 5 hepatocellular adenomas (HCA, 5 patients), 7 fibrolamellar carcinomas (FLC, 7 patients) and 10 hepatocellular carcinomas (HCC, 9 patients). HCC were developed in various etiological contexts including tyrosinemia (2 patients), mitochondrial cytopathy (3 patients) and progressive familial intrahepatic cholestasis (3 patients). Samples were analyzed by a combination of whole Genome (WGS, $n=65$), whole exome (WES, $n=57$), RNA (RNAseq, $n=120$) and reduced representation bisulfite sequencing (RRBS, $n=92$). A summary of the cohort is provided in **Supplementary Fig. S1**, and detailed clinical characteristics of each sample are provided in **Supplementary Table S1**. Also, 7 human fetal liver samples between the 13th and 30th weeks of amenorrhea were analyzed by RNAseq.

Pathological reviewing

All tumors were reviewed by 3 expert pathologists specialized in pediatric liver tumors. For hepatoblastoma, fractions of histological components (fetal, embryonal, mesenchymal, cholangioblastic, small cell undifferentiated) were estimated according to the consensus classification (14) for the whole tumor as well as for mirror blocks corresponding to frozen samples when available. Thus spatial heterogeneity was defined by the coexistence of at least 2 histological components, either at the whole tumor level or at the intra-sample level (mirror block).

Whole genome sequencing (WGS)

We extracted DNA using Maxwell DNA extraction kit (Promega) or AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). Sixty-five tumors and matched non-tumor samples were sequenced at the Centre National de Génomique (Evry, France) on an Illumina HiSeqX5 as paired-end 150 bp reads, with an average depth of 90X for tumors and 30X for non-tumor liver samples. Sequence reads were aligned to the hg19 version of the human genome using BWA (36). We used Picard tools (<http://broadinstitute.github.io/picard/>) to remove PCR duplicates and GATK (37) for local indel realignment and base quality recalibration, as recommended in GATK best practices (38).

Whole exome sequencing (WES)

Whole exome data from 57 tumors and matched non-tumor samples were analyzed in this study, as well as 9 pediatric liver tumor cell lines. Sequence capture, enrichment and elution of genomic DNA was performed by IntegraGen (Evry, France). Agilent in-solution enrichment was used with the manufacturer's biotinylated oligonucleotide probe library

SureSelect Clinical Research Exome V2 or Twist Bioscience Human Core Exome Enrichment System, according to the manufacturer's instructions. The eluted enriched DNA samples were sequenced on an Illumina HiSeq4000 as paired-end 75 bp reads (n=35) or Illumina NovaSeq as paired-end 100 bp reads (n=31), with an average depth of 100X for tumors and 65X for non-tumor liver samples. We used BWA to align reads on the hg38 version of the human genome and sambamba to remove duplicate reads.

Somatic mutation calling

We used MuTect2 to call somatic mutations from WES and WGS data by comparing each tumor sample with its matched non-tumor counterpart. The 3 cell lines derived from patient tumors within our cohort were compared to the matched non-tumor samples, while the 6 other cell lines were compared to a panel of normal samples. We excluded mutations belonging to the ENCODE Data Analysis Consortium blacklisted regions (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable.bed.gz>) and regions covered by < 6 reads in the tumor or normal sample. We then selected only single nucleotide variants (SNVs) with a MuTect2 flag among "PASS", "clustered_events", "t_lod_fstar", "alt_allele_in_normal" or "homologous_mapping_event" and small insertions and deletions (indels) with a MuTect2 flag among "PASS", "clustered_events" or "str_contraction". To improve specificity in the calling of mutations with low variant allele frequency (VAF), we quantified the number of high quality variant reads in the tumor (mapping quality ≥ 20 , base quality ≥ 20) and the number of variant reads in the non-tumor sample with no quality threshold using bamreadcount (<https://github.com/genome/bam-readcount>). Only variants matching the following criteria were finally retained: VAF $\geq 2\%$ in the tumor with ≥ 3 variant reads, VAF $\leq 5\%$ in the non-tumor samples with ≤ 2 variant reads, and a VAF ratio ≥ 5 between the tumor and non-tumor sample.

Germline mutation calling

Germline variant calling was performed independently for WGS and WES data according to GATK (version 4.0) best practices recommendations (39). For each dataset, we performed SNVs and indels discovery using HaplotypeCaller and joint genotyping across all non-tumor samples simultaneously using GenomicsDBImport followed by GenotypeGVCFs. We used hard filtering for WES data with different parameters for SNVs (ExcessHet > 54.69, FS > 10.0, MQ < 50.0, MQRankSum < -5.0, QD < 5.0, QUAL < 50.0, ReadPosRankSum < -5.0, SOR > 2.0, DP < 8.0, GQ < 20.0) and indels (ExcessHet > 54.69, FS > 200.0, QD < 2.0, QUAL < 30.0, ReadPosRankSum < -20.0, DP < 8.0, GQ < 20.0), and we used variant quality score recalibration for WGS (truth sensitivity level: SNVs = 99.6%, indels = 95.0%). Resulting high quality variants were annotated using the Variant Effect Predictor toolset (40). We selected variants with an allele frequency lower than 0.01 or lacking in the gnomAD database (41). All candidate pathogenic variants were manually verified using the Integrative Genomics Viewer (42).

Copy-number and structural rearrangement analysis

We used MANTA (43) software to identify somatic structural rearrangements in WGS data. To keep only the most reliable events, we selected only rearrangements supported by ≥ 15 reads and with a variant allele fraction $\geq 10\%$ in the tumor, and not more than 1

variant read in the non-tumor counterpart. We used *cgpBattenberg* (44) algorithm to reconstruct absolute copy-number profiles from WGS data and estimate tumor purity. We used the Genome Alteration Print method (45) to infer absolute copy-number profiles and tumor purity from WES data, and the circular binary segmentation algorithm implemented in the Bioconductor package *DNAcopy* (46) to identify focal homozygous deletions and high-level amplifications, as previously described (47).

Identification of driver genes

We used two different approaches to identify driver genes in pediatric liver cancers, using both WES and WGS data. First, the *MutSigCV* tool (48) was used to identify genes with significantly recurrent mutations while taking into account gene size and genomic covariates. High-level amplifications and homozygous deletions were included in this analysis as additional mutation categories. Second, we used *Oncodrive* (49) to identify genes with a significant enrichment of mutations with functional impact. Finally, we defined as putative drivers genes with damaging alterations in ≥ 2 patients and either (1) a p -value < 0.05 with both *MutSigCV* and *Oncodrive* tools or (2) identified as driver genes in previous pediatric pan-cancer studies (25,50). After removing 3 genes not expressed in normal liver nor in tumors (95th FPKM quantile < 0.1), we obtained 22 candidate driver genes (**Supplementary Table S2**). Only genes with damaging alterations in ≥ 3 patients were represented in **Fig. 1b**. The frequency of driver alterations per tumor type was estimated by patient and not by sample, in order to avoid biases due to patients with multiple samples. To this end, a patient was considered as altered for a gene as long as one of its samples harbored an alteration in this gene.

Identification of mosaic copy-neutral LOH of 11p15.5 locus in non-tumor liver

For each tumor with a copy-neutral LOH (cn-LOH) of 11p15.5 locus, we searched for the presence of the same cn-LOH in matched non-tumor liver tissue. To that aim, we calculated the B-allele frequencies (*BAF*) of common single nucleotide polymorphisms (SNPs), obtained from WGS or WES data, in the tumor (*BAF_T*) and in the non-tumor (*BAF_{NT}*) sample. We considered a cn-LOH to be present in the non-tumor liver if there was a significant overlap between SNPs with the B allele retained in the tumor (*BAF_T* > 0.5) and those with a *BAF_{NT}* > 0.5 (binomial test). We then estimated the proportion of non-tumor liver cells carrying the cn-LOH as $2 * \text{RAF}_{NT} - 1$ with *RAF_{NT}* the median *BAF* of retained alleles (those with a *BAF_T* > 0.5) in the non-tumor sample.

Mutational signature analysis

We used *Palimpsest* (51) to extract signatures of single base substitutions (SBS), doublet base substitutions (DBS) and indels in WGS data of 65 pediatric liver cancers, and to compare them with known signatures from COSMIC database (v3). *De novo* analysis revealed 4 SBS signatures corresponding to the COSMIC SBS1, 5, 18 and 35 (cosine similarity scores > 0.8), 2 DBS signatures corresponding to the COSMIC DBS5 and a new signature (DBSnew), and 5 indel signatures corresponding to the COSMIC ID1, 2, 3, 5 and 8. To ensure comparability with other data sets, signature exposures in each tumor were recalculated using the COSMIC versions of signatures, except for signature SBS35 (we kept our own version as it was based on a larger number of mutations than the COSMIC version) and DBSnew (absent in COSMIC). SBS signature exposures were also calculated

in WES data. Finally, we estimated the cancer cell fraction (CCF), i.e. the proportion of tumor cells carrying each mutation (see the 'Clonality' section below). Mutations with an upper bound of their CCF confidence interval < 0.95 were considered subclonal. We then quantified SBS signature exposures among clonal and subclonal mutations separately.

Clonality analysis and phylogenetic tree reconstruction

We used Palimpsest to estimate the CCF of each mutation from its variant allele fraction, taking into account tumor purity and absolute copy-number, as previously described (52). For 23 patients with multiple tumor samples, we used a Bayesian Dirichlet process in multiple dimensions (53) to identify clusters of mutations with a similar CCF distribution across all samples, hence belonging to the same branch of the phylogenetic tree. We then manually reconstructed the phylogenetic tree of each patient by organizing branches to fit the observed clonal composition of each sample. Damaging mutations affecting candidate driver genes were annotated on the trees, and we quantified the contribution of each mutational process on each branch with Palimpsest.

RNA sequencing (RNA-seq)

We performed RNA-seq for 120 tumors, 4 non-tumor liver, 7 fetal liver samples and 9 pediatric liver tumor cell lines. RNA samples were enriched for polyadenylated RNA from 1 µg of total RNA, and the enriched samples were used to generate sequencing libraries with the Illumina TruSeq Stranded mRNA kit or NEBNext Ultra II Directional RNA Library Prep kit and associated protocol as provided by the manufacturer. Libraries were sequenced by IntegraGen (Evry, France) on an Illumina HiSeq 4000 as paired-end 75 bp reads or Illumina NovaSeq as paired-end 100 bp reads. Full Fastq files were aligned to the reference human genome hg38 using TopHat2 (54). We removed reads mapping to multiple locations, and we used HTSeq (55) to obtain the number of reads associated to each gene in the Gencode database. We used DESeq2 (56) to import raw read counts into R statistical software and apply variance stabilizing transformation (VST) to the raw count matrix. FPKM scores (number of fragments per kilobase of exon model and millions of mapped reads) were calculated by normalizing the count matrix for the library size and the coding length of each gene.

Gene fusion detection

Fusions detected by TopHat2 (54) (`--fusion-search --fusion-min-dist 2000 --fusion-anchor-length 13 --fusion-ignore-chromosomes chrM`) were filtered using the TopHatFusion-post algorithm and validated using FusionInspector (57) (<https://github.com/FusionInspector>). We kept only fusions validated by BLAST, with at least 10 split-reads or read pairs spanning the fusion event, and an FFPM (Fusion Fragments Per Million reads) ≥ 0.1 . We removed fusions identified recurrently in a cohort of 36 normal liver samples, involving genes with inconsistent orientations, non-coding genes and putative read-through transcripts.

Detection of large in-frame *CTNNB1* deletions

Large in-frame *CTNNB1* deletions were screened in WGS, WES and RNA-seq data using dedicated approaches. We used MANTA software (43) to identify structural variations spanning *CTNNB1* region in WGS and WES (with the `--exome` option) data, with the

following parameters to increase sensitivity: *minEdgeObservations*=2, *minScoreVariantSize*=10. For RNA-seq, we analyzed the *junctions.bed* output files from TopHat2 (54) to identify abnormal junctions involving *CTNNB1* exons 3 or 4. We also screened for paired reads with one mate in exon 2 and the other in exon 3/4, and an abnormally long insert size, greater than the mean insert size of the library + 1 standard deviation.

Gene expression-based classification of hepatoblastoma

We restricted the transcriptomic classification to HB samples, excluding other diagnoses (HCC, FLC, HCA) with a limited number of samples, in order to identify robust subgroups. Hierarchical clustering was performed on 100 hepatoblastomas and 4 non-tumor liver samples, based on the vst-normalized expression of the 3000 autosomal genes with sufficient expression (95th FPKM percentile ≥ 0.1) and the highest standard deviation. The *hclust* R function was used with Euclidean distance and Ward.D2 linkage method, after median-centering the data. To verify that HB subgroups were not driven by patients with multiple samples, we reproduced the clustering with only one sample by patient and we obtained similar results (**Supplementary Fig. S15**). For external validation of HB clusters, we downloaded RNA-seq data set from Carrillo-Reixach *et al.* (16) comprising 34 HB (Gene Expression Omnibus accession number GSE132219) and performed a hierarchical clustering based on the expression of the 19 markers of hepatic differentiation, progenitor and mesenchymal subgroups represented in **Fig. 3**, with Cosine distance and Ward.D2 linkage method (**Supplementary Fig. S6**). We used the Bioconductor *limma* (58) package to identify differentially expressed genes between HB subgroups. We applied a q-value threshold of ≤ 0.05 to define differentially expressed genes.

Gene expression signatures

We used previously established molecular signatures from MSigDB database to quantify the level of hepatic differentiation ('Hsiao liver specific genes'), cell proliferation ('Hallmark REACTOME cell cycle') and inflammation ('Hallmark inflammatory response') in each sample. We used MCPcounter (59) to estimate the infiltration by diverse immune cell types and classify HB samples accordingly. We also classified each HB sample according to the previously described C1/C2 (27), C1/C2A/C2B (28) and RMS-1/2/3a/3b (16) classifications. For C1/C2 (respectively C1/C2A/C2B) classification, we generated a consensus clustering (1000 resampling iterations of hierarchical clustering with Euclidean distance and Ward.D2 linkage) of our cohort in 2 (resp. 3) groups based on the 16 marker genes defined by Cairo *et al.* (27) (resp. the 4 markers *VIM*, *HSD17B6*, *TOP2A* and *ITGA6* defined by Hooks *et al.* (28)), and we assigned to each consensus cluster to the relevant subgroup based on the expression of their respective markers. For the RMS classification, we first performed a hierarchical clustering based on the expression of 68 genes of the 14q32 imprinted locus to define 2 groups (moderate vs strong 14q32-gene signature). We verified that our two main DNA methylation-based clusters were consistent with Carrillo Reixach's Epi-CA/CB groups (16). Finally, we combined the 14q32 signature, Epi-CA/CB methylation groups and C2A transcriptomic subgroup to define the RMS classification as in Carrillo-Reixach *et al.* (16). A signature of the 'Liver Progenitor'

subgroup of HB comprising the 7 representative genes indicated in **Fig. 3a** was used to quantify this phenotype in cell lines.

Single-nucleus isolation from frozen tissues

Single nuclei were isolated from 4 samples for single-nucleus RNA-seq as previously described (60), using EZ Lysis buffer workflow with slight modifications. Briefly, tissue samples were thawed in PBS and cut into pieces < 0.5 cm. Approximately 35 mg of tissue were poured in a glass Dounce tissue grinder (Sigma, cat. no. D8938) and homogenized 25 times with pestle A and 25 times with pestle B in 1.5 mL of ice-cold nuclei EZ lysis buffer. Samples were then incubated on ice for 5 min with an additional 3 mL of cold EZ lysis buffer. Nuclei were centrifuged at 500 g for 5 min at 4 °C, washed with 5 mL ice-cold EZ lysis buffer and incubated on ice for 5 min. After centrifugation, the nucleus pellet was washed with 5 mL of Nuclei Wash buffer containing 1× PBS, 0.1%, non-acetylated BSA (Thermo AM2618) and 200 units/mL RNase inhibitor (NEB M0307L). Isolated nuclei were resuspended in 2 mL of Nuclei Suspension Buffer containing 1× PBS, 1% non-acetylated BSA (Thermo AM2618) and 200 units/mL RNase inhibitor (NEB M0307L), filtered through a 70 µm and then a 30 µm MACS SmartStrainers (Miltenyibiotec 130-098-462 & 130-098-458), and counted under microscope using C-chip disposable hemocytometer. A final concentration of 1,000 nuclei per µL was used for loading on a 10x channel.

Single-nucleus RNA sequencing (snRNA-seq)

Single-nucleus RNA-sequencing was performed by Integrage SA (Evry, France) on matched non-tumor liver, primary HB and lung metastasis samples from one patient (#3981), following the Chromium Next GEL Single Cell 3' V3.1 protocol. In short, about 8,800 single nuclei were loaded into each channel of a Chromium single-cell 3' chip. Single nuclei were partitioned into droplets with gel beads in the Chromium Controller. After emulsions were formed, barcoded reverse transcription of RNA took place, followed by cDNA amplification, fragmentation, and adapter and sample index ligation, according to the manufacturer's recommendations. Libraries from the 10X channels were pooled together and sequenced as paired-end 100b reads on an Illumina NovaSeq. We used 10x Genomics Cell Ranger 5.0 (61) to align snRNA-seq reads to the human genome (GrCh38/hg38) and generate UMI counts for each sample, including intronic reads. We obtained respectively 9142, 5859 and 2875 nuclei, with a median of 2042, 2334 and 3446 genes per nuclei from samples #3981N (non-tumor liver), #3984T (primary HB) and #3988T (lung metastasis). We then filtered the feature-barcode matrix to retain only good-quality nuclei and reliable genes. We removed nuclei with <1000 read counts, <500 genes detected or >5% of UMI counts in mitochondrial genes, and we removed genes detected in <3 nuclei as well as ERCC and mitochondrial genes. After QC we kept respectively 9140, 5685 and 2826 nuclei, with a total of 16470, 18384 and 18294 genes for samples #3981N, #3984T and #3988T. All secondary analyses were performed using Seurat v3 (62). We normalized each dataset using the SCTransform function with default parameters, performed principal component analysis on the 3,000 most variable genes and ran Louvain graph-based clustering on the 30 principal components with a resolution of 0.5. We used a Uniform Manifold Approximation and Projection (UMAP) with default settings to visualize the results. We used inferCNV v1.6 (Tickle T *et al.*, available from <https://github.com/broadinstitute/inferCNV>) to reconstruct virtual copy-number profiles, using healthy hepatocytes from the non-tumor liver sample as reference, and

genes with an average read count >0.1 in reference nuclei. We performed a clustering of copy-number profiles in each sample separately. We identified tumor cells using both inferCNV clusters and Seurat classification, and we redid the whole analysis (SCTransform normalization, dimensionality reduction, clustering and UMAP visualization) restricted to tumor cells. To characterize tumor cell clusters, we selected marker genes of each differentiation state ('Hepatocytic', 'Liver Progenitor' and 'Mesenchymal'). To that end, we selected the top 70 genes overexpressed in the corresponding group in bulk RNA-seq data, restricted to genes identified in snRNA-seq data. We obtained ~50 marker genes for each subgroup and we computed for each nucleus the mean log-normalized expression (NormalizeData pipeline) over each set of markers. We then assigned a status to each cluster based on the expression of these marker genes.

Reduced representation bisulfite sequencing (RRBS)

RRBS was performed in two distinct projects (PJ17 (n=82) and PJ20 (n=23)) by Integrage SA (Evry, France) as described by Gu *et al.* (63), with the Diagenode Premium RRBS kit. In brief, 100 ng of qualified genomic DNA were digested with MspI. After end-repair, A-tailing and ligation to methylated and indexed adapters, the size selected library fragments were subjected to bisulfite conversion and PCR amplified. Samples of the PJ17 project (res. PJ20) were then sequenced on an Illumina HiSeq4000 (resp. NovaSeq) sequencer as paired-end 75 bp (resp. 100 bp) reads. Image analysis and base calling was performed using Illumina Real Time Analysis with default parameters. Reads were aligned to the hg38 version of the human genome using BS_Seeker2 (64). Sorted bam files were converted into CGmap files using CGmaptools (65), and the methylation level of each CpG site was defined as the ratio between the number of effective cytosines after bisulfite conversion (=methylated cytosines) and the total number of cytosines and thymines after bisulfite conversion (=methylated + unmethylated cytosines). On average ~ 7 million CpG sites were covered in each sample after discarding CpG sites located in ENCODE blacklisted genomic regions (*wgEncodeDacMapabilityConsensusExcludable* tract from UCSC genome browser). We next integrated methylation levels across 100 bp-long genomic regions (tiles) using the *tileMethylCounts* function from the *Methylkit* package (66). Tile coverage was defined as the coverage sum of all CpGs inside the tile, and methylation was defined as the ratio between the total number of methylated CpGs and tile coverage. On average ~ 1.3 million tiles were covered in each sample. Due to the different read lengths in PJ17 and PJ20, some tiles displayed heterogeneous coverage leading to systematic biases between the two projects. We thus compared the non-tumor (NT) liver samples from PJ20 (n=4) and PJ17 (n=9), and we removed 58,179 tiles (~3%) with a methylation difference ≥ 0.05 or ≤ -0.05 between all PJ17 and all PJ20 NT samples. To determine the methylation status of 11p15.5 locus, we computed for each sample the mean methylation over imprinted regions IC1 (chr11:1998745-2003509, hg38) and IC2 (chr11:2697587-2700983, hg38). We then used a K-means clustering to identify samples with gain of methylation (GOM) of IC1 and/or loss of methylation (LOM) of IC2.

DNA methylation changes in hepatoblastoma

We generated a DNA methylation-based classification of 84 HB and 13 non-tumor liver samples. To do so, we selected the 5,000 tiles with coverage >50X in all samples and the highest standard deviation, and we used the *hclust* R function to perform a hierarchical clustering with Pearson distance and Ward.D linkage method. We used Independent

Component analysis (ICA) as previously described (67) to characterize independent sources of DNA methylation changes in HB, based on ~212,000 tiles with coverage >10X in all samples and standard deviation >0.035. We identified 5 methylation components (MCs) including components related to the 'Liver Progenitor' (MC1) and 'Mesenchymal' (MC2) subgroups, and to patient age (MC3). We explored various (epi)genomic features associated with differentially methylated tiles and MCs, including gene features (promoter / gene body), CpG island features (island / shore / shelf / open sea), chromatin states (68) and DNA methylation domains (69) (highly methylated domains vs. partially methylated domains) in normal liver, and replication timing in HepG2 cell line (70). We used ELMER v2 package (71) to identify transcription factor binding sites (TFBS) enriched within tiles hypermethylated in the 'Mesenchymal' component. We used an in-house adaptation of the GSEA (Gene Set Enrichment Analysis) method to identify overrepresented gene sets (MSigDB v6 database) among genes paired with tiles hypermethylated in 'Mesenchymal' samples. We compared the vst-normalized expression of transcription factors involved in liver differentiation (*HNF1A*, *HNF4A*, *PPARG*, *NR2F6*) with the methylation of tiles containing their respective binding motifs and hypermethylated in the 'Mesenchymal' component. These correlations were validated in Carrillo-Reixach's data set comprising 26 HB analyzed with Illumina 850k methylation arrays (16).

Cell lines

Nine pediatric liver tumor cell lines were collected from collaborations or obtained from commercial sources (**Supplementary Table S8**). Cells were grown either in Dulbecco's modified Eagle's medium (DMEM) or Advanced DMEM F-12 supplemented with 10% fetal bovine serum and using usual conditions (100 U/mL penicillin/streptomycin, 1% Glutamine at 37 °C, 5% CO₂, identity confirmed using whole-exome sequencing, mycoplasma-free verified with MycoAlert Mycoplasma PLUS detection kit (Lonza)).

Determination of drug sensitivity

Doxorubicin (S1208, Selleck chemicals), YM155 (S1130, Selleck chemicals), BI-2536 (S1109, Selleck chemicals), AZD7762 (S1532, Selleck chemicals), trametinib (S2673, Selleck chemicals) were dissolved in DMSO at 10mM final concentration. Cisplatin (S1166, Selleck chemicals) was dissolved in H₂O, 0.9% NaCl, 0.3% Tween20 at a 0.5g/L final concentration, carboplatin (S1215, Selleck chemicals) was dissolved in H₂O, 0.1% Triton. Cells were seeded at 2500-4500 cells/well (**Supplementary Table S10**). After overnight incubation at 37°C and 5% CO₂, cells were treated with 5 different concentrations of drugs (0.001, 0.01, 0.1, 1, and 10µM) using HP D300 digital dispenser (Tecan, Mannedorf, Switzerland). Growth inhibition was measured 72h after treatment with MTS diluted 1:6 in fresh culture medium. Cell viability was assessed by recording absorbance at 490nm using a FLUOstar microplate reader. Dose-response curves were performed using GraphPad Prism 6 Software to determine two parameters reflecting drug sensitivity: GI50 and the area under the curve (AUC). When the GI50 was not reached, the values were set to the highest concentration tested (10µM). Each concentration was tested in duplicate. Polyclonal goat anti-IGF2 antibody (Ref AF-292-NA, Biotechne) and control anti-Igg goat antibody (Ref AB-108-C, Bio-techne) were tested using 5 concentrations (0.1, 0.5, 1, 5 and 10 µg/mL). Monoclonal human anti-IGF1/IGF2 therapeutic antibody (Xentuzumab,

#TAB-475CQ, Creative Biolabs) was resuspended in PBS at 1 mg/mL final concentration and tested at 1 μ M.

In vivo xenografts treatments

Mice were housed in a specific pathogen-free facility and experiments were conducted using protocols and conditions approved by the Institutional animal ethical committee (Authorization n°2015082610113065.01, Ethics Committee Paris-Nord C2EA 121). At day₀, 5.10⁶ HepG2 cells with 1:1 matrigel were inoculated subcutaneously on each flank of female BALB/cAnNRj-Foxn1 5 weeks old mice. At day₂₀, when tumors started growing (volume >200mm³), mice were randomized in 4 groups of 6 mice allocated to the following treatment arms: cisplatin, cisplatin/doxorubicin, BI-2536 and Vehicle. Mice in cisplatin group were injected intraperitoneally at day₂₀ and day₂₇ with 5mg/kg cisplatin. Mice from cisplatin/doxorubicin group received one injection of cisplatin at day₂₀ (5mg/kg) and one intraperitoneal doxorubicin at day₂₇ (2mg/kg). Mice in BI-2536 group were intravenously injected with 40mg/kg BI-2536 at day₂₀ and day₂₇ whereas mice in the vehicle control group received injections of H₂O, 0.9% NaCl. Mice were weighted two times a week and tumor volume was measured at the same time using a caliper and the following formula: (Length*width²)/2 and was expressed as a percentage of initial tumor volume at day₂₀.

Generation of SBS35 signature *in vitro*

To test cisplatin ability to induce SBS35 mutational signature in hepatoblastoma, 6 pediatric liver tumor cell lines were treated with 0.5 μ M cisplatin for 4 weeks. At day₀, 60 000 cells were seeded in a 6-well plate and reseeded each week at the same concentration. When mortality was too high, cells were not split and fresh medium supplemented with 0.5 μ M cisplatin was added. At day₂₈, cells were grown in fresh DMEM and limit dilutions were performed in order to expand clonal cell lines. After amplification, samples were extracted using Allprep DNA/RNA/miRNA universal extraction kit (Qiagen). Finally, mutational signature analysis was performed on cells derived from limit dilution and compared with bulk non-treated baseline mutational profile.

Methylation specific Multiplex ligation-dependent probe amplification (MS-MLPA)

MLPA was used to determine the status of locus 11p15.5 in 27 samples without RRBS data available. MLPA was carried out with 50-100ng of DNA diluted in Tris-HCl, 0.1mM EDTA. DNA samples were screened using ME030-C3 BWS/RSS kit (MRC-Holland, Amsterdam, The Netherlands) containing 42 (MS-) MLPA probes: 26 probes specific of BWS 11p15 region, 2 probes targeting NSD1 gene, 13 reference probes targeting other chromosomes and 1 digestion control probe. Two reference probes targeting respectively 2p25 and 2q24 were excluded because of the presence of very frequent copy number variation in hepatoblastoma tumors. Among the 26 probes targeting 11p15 region, 10 were methylation specific, allowing for the detection of methylation abnormalities. To determine thresholds for IC1 gain of methylation and IC2 loss of methylation, we used K-means clustering method as well. Samples with a gain of methylation in at least 2/3 probes targeting IC1 were annotated as GOM IC1 whereas samples with a loss of methylation in at least 2/4 probes were considered LOM IC2. One probe covering IC1

(H19.11.001.976583) was excluded of analysis because its distribution among samples was not discriminating samples with and without GOM IC1.

***In situ* hybridization IGF2 in FFPE slides**

In situ hybridization experiments were performed using RNAscope® 2.5 HD Detection Reagent BROWN kit (Cat #322300) and *IGF2* probe (Cat. #594361) according to manufacturer's protocol. FFPE slides underwent target retrieval under standard conditions (15 minutes in target retrieval reagent >98°C) and a 20-minutes protease digestion at 40°C in the HybEZ oven.

Immunohistochemistry

Immunohistochemical analyses anti- β -catenin (BD Transduction, clone 14, 610154, 1/300) and anti-GS (Bioscience, 1/500) were performed as in (34).

TERT promoter screening

TERT promoter mutations were identified with WGS data when available and completed with Sanger sequencing as previously described (72) for other samples.

Data availability

The sequencing data reported in this paper have been deposited to the EGA (European Genome-phenome Archive) database (European Genome-phenome Archive) database (accession numbers [EGAS EGAS00001005108] and [EGAS00001003536]).

Acknowledgments: We thank all the clinicians, radiologists, pathologists, laboratory technicians and scientists who participated to this work: Jayendra Shinde, Sandrine Imbeaud, Samantha Schaeffer, Barkha Gupta (Inserm UMR 1138) ; Yves Lebouc (Inserm UMR 938), Isabelle Janoueix (Inserm UMR 830) ; Anne Boland, Robert Olaso (CEA, CNRGH) ; Pr. Frédéric Gauthier, Pr. Hélène Martelli, Cécile Jeannot, Dr. Danièle Pariente, Dr. Stéphanie Franchi-Abella, Isabelle Friteau, Dr. Charlotte Mussini, Katia Posseme, Olivier Trassard, Martine Prsle, Véronique Bruna, Marie-Josée Redon, Véronique Mamesier, Aurélie Jove-Garcia, Damien Schmidt, Ludivine Meni, Violette Gendre, Delphine Colmant, Christophe Avignon, Julie Tisserand, (CHU Bicêtre, APHP, Le Kremlin Bicêtre) ; Dr. Dominique Debray, Gisèle Le Gall (CHU Necker, APHP, Paris) ; Pr. Michel Peuchmaur, Pr. Dominique Berrebi, Pascal Blain, Muriel Mateo-Gajardo, Elodie Dauvergne, Maxette Pierin, Louise Le Guen (CHU Robert Debré, APHP, Paris) ; Pr. Aurore Coulomb L'hermine, Dr. Sabah Boudjemaa, Dr. Hélène Dariane (CHU Trousseau, APHP, Paris) ; Dr. Julien Calderaro (CHU Henri Mondor, APHP, Créteil) ; Pr. Jean Yves Scoazec, Brenda Mallone (Institut Gustave Roussy, Villejuif) ; Dr. Paul Freneaux, Dr. Daniel Orbach (Institut Curie, Paris) ; Dr. Carole Coze, Pr. Dominique Figarella Branger, Dr. Andre Maues de Paula (CHU de la Timone, Marseille) ; Pr. Valérie Rigau, Dr. Yuri Musizzano (CHU Gui de Chauiac, Montpellier) ; Pr. Jannick Selves (CHU Toulouse) ; Dr. Estelle Thebaud, Dr. Louise Galmiche, Pr. Jean-François Mosnier (CHU Nantes) ; Dr. Cécile Dumesnil, Dr. Arnaud François (CHU Charles Nicolle, Rouen) ; Dr. Anne Lutin, Dr. Antoine Gourmel, Dr. Gilles Morin (CHU Amiens) ; Dr. Claire Berger (CHU Saint-Etienne) ; Dr. Frédérique Dijoud, Dr. Cécile Picard (CHU Lyon) ; Dr Clémence Yguel (CHU Nancy) ; Charles Marcaillou (Integrigen) ; Josh Waterfall (Institut Curie, INSERM U830), Andrei Zinovyev (Institut

Curie, INSERM U900) ; HEPATOBIO database. We warmly thank Xin Chen, USCF – USA and Karl-Dimiter Bissig, Baylor College – USA for their kind gift of Hep293TT and B6-2 cell lines.

REFERENCES

1. Darbari A, Sabin KM, Shapiro CN, Schwarz KB. Epidemiology of primary hepatic malignancies in U.S. children. *Hepatology*. 2003;38:560–6.
2. Hadzic N, Finegold MJ. Liver neoplasia in children. *Clin Liver Dis*. 2011;15:443–62, vii–x.
3. Arai Y, Honda S, Haruta M, Kasai F, Fujiwara Y, Ohshima J, et al. Genome-wide analysis of allelic imbalances reveals 4q deletions as a poor prognostic factor and MDM4 amplification at 1q32.1 in hepatoblastoma. *Genes Chromosomes Cancer*. 2010;49:596–609.
4. Eichenmüller M, Trippel F, Kreuder M, Beck A, Schwarzmayer T, Häberle B, et al. The genomic landscape of hepatoblastoma and their progenies with HCC-like features. *J Hepatol*. 2014;61:1312–20.
5. Jia D, Dong R, Jing Y, Xu D, Wang Q, Chen L, et al. Exome sequencing of hepatoblastoma reveals novel mutations and cancer genes in the Wnt pathway and ubiquitin ligase complex. *Hepatology*. 2014;60:1686–96.
6. Sumazin P, Chen Y, Treviño LR, Sarabia SF, Hampton OA, Patel K, et al. Genomic analysis of hepatoblastoma identifies distinct molecular and prognostic subgroups. *Hepatology*. 2017;65:104–21.
7. Sekiguchi M, Seki M, Kawai T, Yoshida K, Yoshida M, Isobe T, et al. Integrated multiomics analysis of hepatoblastoma unravels its heterogeneity and provides novel druggable targets. *npj Precision Oncology*. Nature Publishing Group; 2020;4:1–12.
8. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555:321–7.
9. Khanna R, Verma SK. Pediatric hepatocellular carcinoma. *World J Gastroenterol*. 2018;24:3980–99.
10. Iannelli F, Collino A, Sinha S, Radaelli E, Nicoli P, D’Antiga L, et al. Massive gene amplification drives paediatric hepatocellular carcinoma caused by bile salt export pump deficiency. *Nature Communications*. Nature Publishing Group; 2014;5:3850.
11. Haines K, Sarabia SF, Alvarez KR, Tomlinson G, Vasudevan SA, Heczey AA, et al. Characterization of pediatric hepatocellular carcinoma reveals genomic heterogeneity and diverse signaling pathway activation. *Pediatr Blood Cancer*. 2019;66:e27745.
12. Honeyman JN, Simon EP, Robine N, Chiaroni-Clarke R, Darcy DG, Lim IIP, et al. Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. *Science*. 2014;343:1010–4.
13. Franchi-Abella S, Branchereau S. Benign hepatocellular tumors in children: focal nodular hyperplasia and hepatocellular adenoma. *Int J Hepatol*. 2013;2013:215064.
14. López-Terrada D, Alaggio R, de Dávila MT, Czauderna P, Hiyama E, Katzenstein H, et al. Towards an international pediatric liver tumor consensus classification: proceedings of the Los Angeles COG liver tumors symposium. *Mod Pathol*. 2014;27:472–91.
15. Feng T-C, Zai H-Y, Jiang W, Zhu Q, Jiang B, Yao L, et al. Survival and analysis of prognostic factors for hepatoblastoma: based on SEER database. *Ann Transl Med*. 2019;7:555.
16. Carrillo-Reixach J, Torrens L, Simon-Coma M, Royo L, Domingo-Sàbat M, Abril-Fornaguera J, et al. Epigenetic footprint enables molecular risk stratification of hepatoblastoma with clinical implications. *J Hepatol*. 2020;73:328–41.
17. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500:415–21.
18. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The

repertoire of mutational signatures in human cancer. *Nature*. Nature Publishing Group; 2020;578:94–101.

19. Pilati C, Shinde J, Alexandrov LB, Assié G, André T, Hélias-Rodzewicz Z, et al. Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J Pathol*. 2017;242:10–5.

20. Boot A, Huang MN, Ng AWT, Ho S-C, Lim JQ, Kawakami Y, et al. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res*. 2018;28:654–65.

21. Galluzzi L, Senovilla L, Vitale I, Michels J, Martins I, Kepp O, et al. Molecular mechanisms of cisplatin resistance. *Oncogene*. 2012;31:1869–83.

22. Martinez-Quetglas I, Pinyol R, Dauch D, Torrecilla S, Tovar V, Moeini A, et al. IGF2 Is Up-regulated by Epigenetic Mechanisms in Hepatocellular Carcinomas and Is an Actionable Oncogene Product in Experimental Models. *Gastroenterology*. 2016;151:1192–205.

23. Caruso S, Calatayud A-L, Pilet J, La Bella T, Rekik S, Imbeaud S, et al. Analysis of Liver Cancer Cell Lines Identifies Agents With Likely Efficacy Against Hepatocellular Carcinoma and Markers of Response. *Gastroenterology*. 2019;157:760–76.

24. Chang MT, Bhattarai TS, Schram AM, Bielski CM, Donoghue MTA, Jonsson P, et al. Accelerating Discovery of Functional Mutant Alleles in Cancer. *Cancer Discov*. American Association for Cancer Research; 2018;8:174–83.

25. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555:321–7.

26. Coorens THH, Treger TD, Al-Saadi R, Moore L, Tran MGB, Mitchell TJ, et al. Embryonal precursors of Wilms tumor. *Science*. 2019;366:1247–51.

27. Cairo S, Armengol C, De Reyniès A, Wei Y, Thomas E, Renard C-A, et al. Hepatic stem-like phenotype and interplay of Wnt/beta-catenin and Myc signaling in aggressive childhood liver cancer. *Cancer Cell*. 2008;14:471–84.

28. Hooks KB, Audoux J, Fazli H, Lesjean S, Ernault T, Dugot-Senant N, et al. New insights into diagnosis and therapeutic options for proliferative hepatoblastoma. *Hepatology*. 2018;68:89–102.

29. Northcott PA, Buchhalter I, Morrissy AS, Hovestadt V, Weischenfeldt J, Ehrenberger T, et al. The whole-genome landscape of medulloblastoma subtypes. *Nature*. Nature Publishing Group; 2017;547:311–7.

30. van Groningen T, Koster J, Valentijn LJ, Zwijnenburg DA, Akogul N, Hasselt NE, et al. Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat Genet*. 2017;49:1261–6.

31. Boeva V, Louis-Brennetot C, Peltier A, Durand S, Pierre-Eugène C, Raynal V, et al. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat Genet*. 2017;49:1408–13.

32. Letouzé E, Shinde J, Renault V, Couchy G, Blanc J-F, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun*. 2017;8:1315.

33. Pich O, Muiños F, Lolkema MP, Steeghs N, Gonzalez-Perez A, Lopez-Bigas N. The mutational footprints of cancer therapies. *Nat Genet*. 2019;51:1732–40.

34. Morcrette G, Hirsch TZ, Badour E, Pilet J, Caruso S, Calderaro J, et al. APC germline hepatoblastomas demonstrate cisplatin-induced intratumor tertiary lymphoid structures. *Oncoimmunology*. 2019;8:e1583547.

35. Tesniere A, Schlemmer F, Boige V, Kepp O, Martins I, Ghiringhelli F, et al.

- Immunogenic death of colon cancer cells treated with oxaliplatin. *Oncogene*. 2010;29:482–91.
36. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
37. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–303.
38. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–8.
39. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11.10.1–11.10.33.
40. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biology*. 2016;17:122.
41. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–43.
42. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–6.
43. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32:1220–2.
44. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. *Cell*. 2012;149:994–1007.
45. Popova T, Manié E, Stoppa-Lyonnet D, Rigaiil G, Barillot E, Stern MH. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome Biol*. 2009;10:R128.
46. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004;5:557–72.
47. 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.
48. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499:214–8.
49. Mularoni L, Sabarinathan R, Deu-Pons J, Gonzalez-Perez A, López-Bigas N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biology*. 2016;17:128.
50. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*. 2018;555:371–6.
51. Shinde J, Bayard Q, Imbeaud S, Hirsch TZ, Liu F, Renault V, et al. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics*. 2018;34:3380–1.
52. Letouzé E, Shinde J, Renault V, Couchy G, Blanc J-F, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun*. 2017;8:1315.

53. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med*. 2015;21:751–9.
54. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*. 2013;14:R36.
55. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–9.
56. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
57. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biology*. 2019;20:213.
58. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
59. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17:218.
60. Slyper M, Porter CBM, Ashenberg O, Waldman J, Drokhlyansky E, Wakiro I, et al. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med*. 2020;26:792–802.
61. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*. Nature Publishing Group; 2017;8:14049.
62. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. *Cell*. Elsevier; 2019;177:1888-1902.e21.
63. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols*. Nature Publishing Group; 2011;6:468–81.
64. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*. 2013;14:774.
65. Guo W, Zhu P, Pellegrini M, Zhang MQ, Wang X, Ni Z. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics*. 2018;34:381–7.
66. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*. 2012;13:R87.
67. Meunier L, Hirsch TZ, Caruso S, Imbeaud S, Bayard Q, Roehrig A, et al. DNA Methylation Signatures Reveal the Diversity of Processes Remodeling Hepatocellular Carcinoma Methylomes. *Hepatology*. 2021; doi: 10.1002/hep.31796.
68. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–30.
69. Salhab A, Nordström K, Gasparoni G, Kattler K, Ebert P, Ramirez F, et al. A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains. *Genome Biology*. 2018;19:150.
70. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the

human genome. *Nature*. 2012;489:57–74.

71. Silva TC, Coetzee SG, Gull N, Yao L, Hazelett DJ, Noushmehr H, et al. ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics*. 2019;35:1974–7.

72. Nault JC, Mallet M, Pilati C, Calderaro J, Bioulac-Sage P, Laurent C, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat Commun*. 2013;4:2218.

Figures

Figure 1

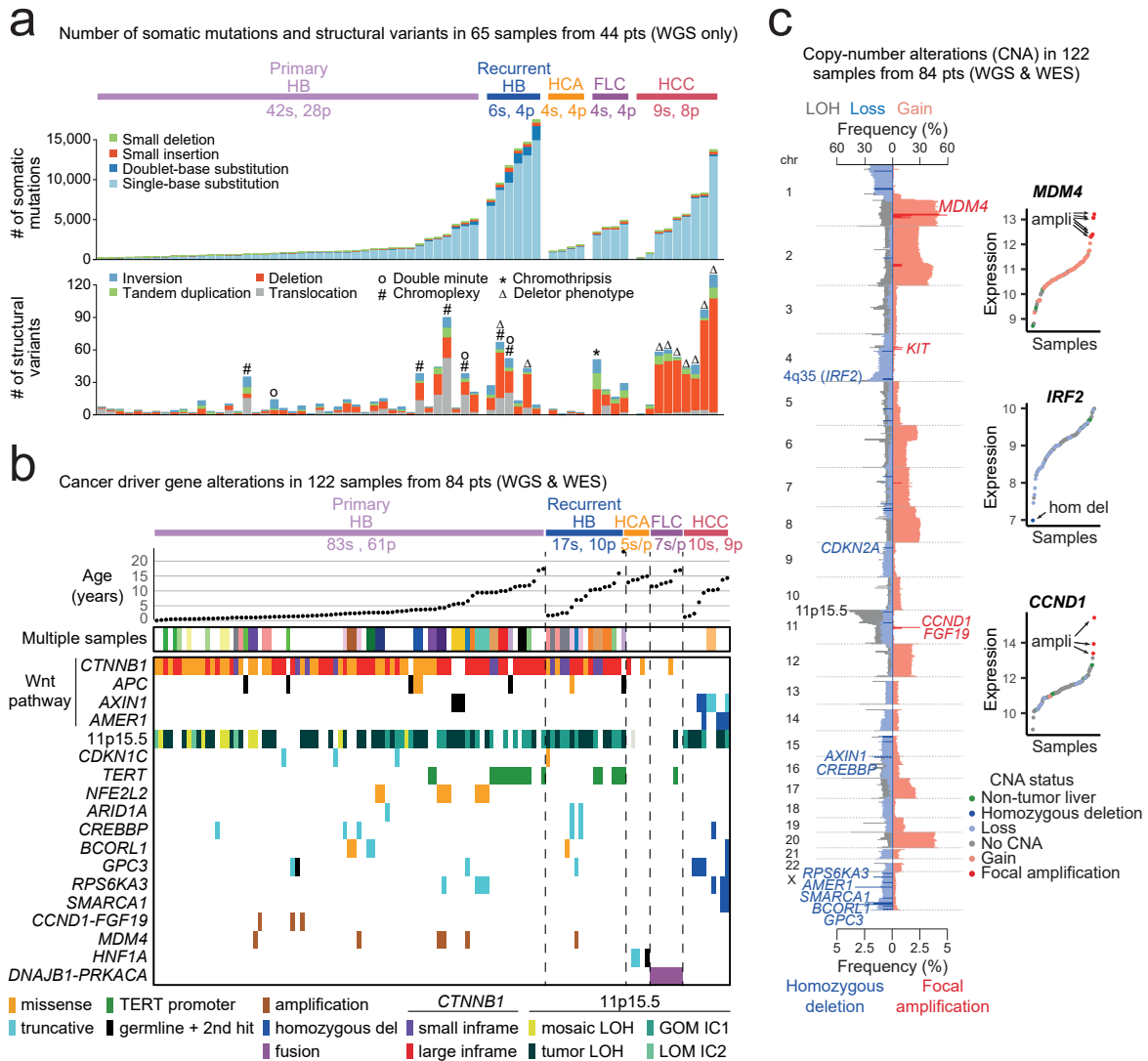


Figure 1. Genomic landscape of pediatric liver cancers. **a** Number of somatic mutations and structural variants identified in 65 pediatric liver cancers (PLC) by whole genome sequencing (WGS). Samples are ordered by diagnosis and mutation burden. Alteration types are indicated with a color code, and specific structural variant phenotypes are highlighted. **b** Heatmap representation of driver alterations across 122 PLC analyzed by WGS or whole exome sequencing (WES). Samples are ordered by diagnosis and patient age at sampling. **c** Frequency of copy-number alterations along the genome for 122 PLC analyzed by WGS or WES. The top axis indicates the frequency of low-amplitude changes (gains, losses and losses of heterozygosity (LOH)); the bottom axis indicates the frequency of high-amplitude changes (focal amplifications and homozygous deletions). Target genes of amplifications and homozygous deletions are indicated. Correlation between gene expression (variance-stabilized) and copy-number is displayed for 3 selected genes.

Figure 2

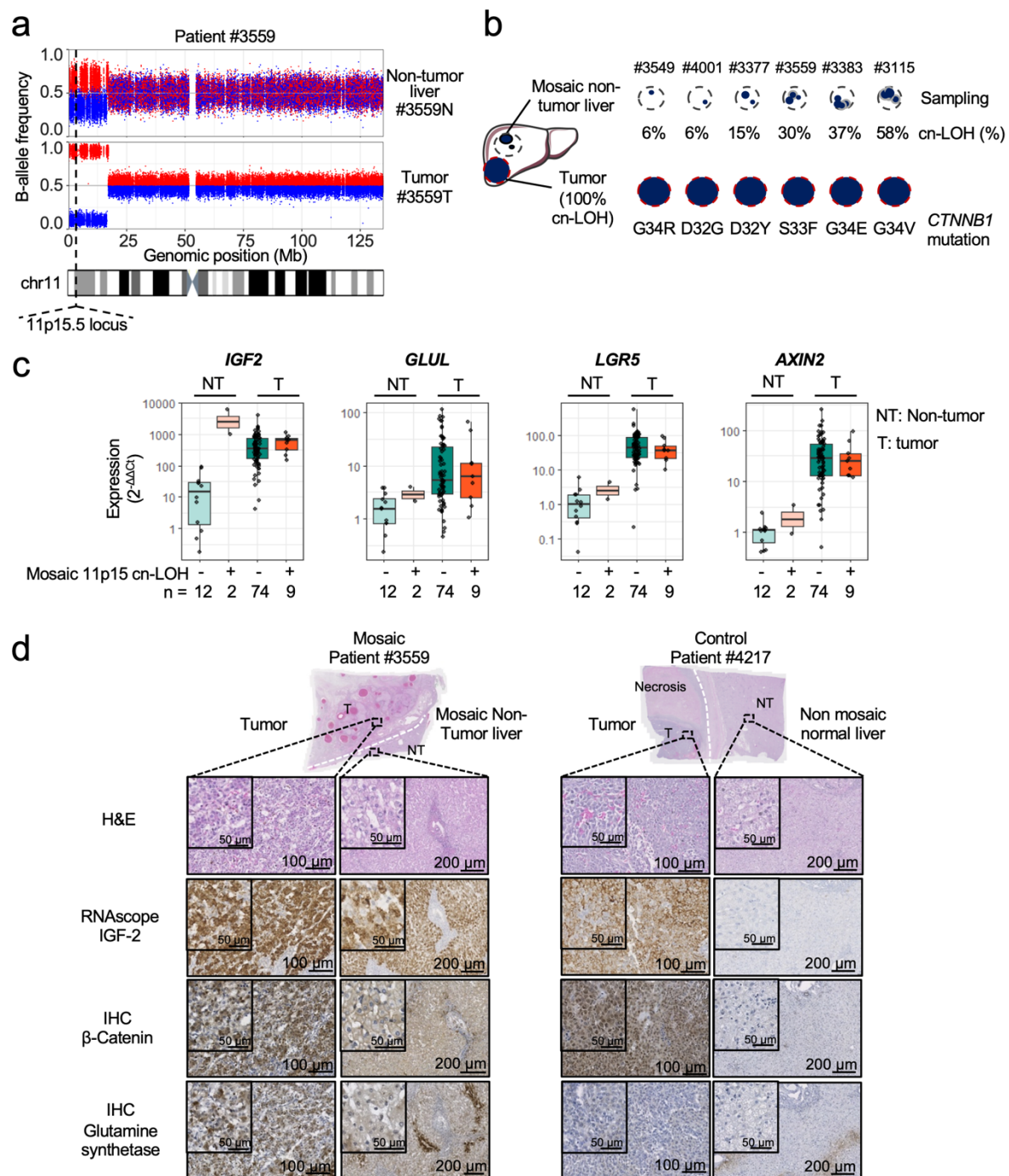


Figure 2. Pre-malignant clonal expansions with 11p15 alteration in hepatoblastoma patients. **a** Identification of a copy-neutral LOH (cn-LOH) at 11p15 locus in the non-tumor liver of patient #3559. **b** Allele frequencies (BAF) of heterozygous single-nucleotide polymorphisms (SNPs) are represented along chromosome 11. SNPs with a BAF greater (resp. lower) than 0.5 in the tumor are colored in red (resp. blue). In the cn-LOH region, red (resp. blue) SNPs correspond to those for which the B allele was retained (resp. lost). The same BAF imbalance is identified in the non-tumor sample, with the same boundaries, demonstrating the presence of the cn-LOH. The amplitude of BAF changes indicate that the cn-LOH is present in 30% of cells in the non-tumor sample. **b**

Pre-malignant expansions with cn-LOH at 11p15 locus were identified in 6 hepatoblastoma patients. The proportion of non-tumor cells harboring the alteration is indicated below. Copy-neutral LOH became clonal in matched hepatoblastoma that had acquired in addition activating *CTNNB1* mutations. **c** Expression levels of *IGF2* and the β -catenin targets *GLUL*, *LGR5*, *AXIN2* ($2^{-\Delta\Delta C_t}$) in tumor and non-tumor samples from patients with and without 11p15.5 cn-LOH premalignant expansions. **d** Representative areas of FFPE slides from 2 HB patients: patient #3559 with mosaic 11p15.5 cn-LOH and control patient #4217. Four types of staining were performed: hematoxylin and eosin staining (H&E), IGF2 RNAscope *in situ* hybridization, and β -catenin and Glutamine synthetase immunostainings.

Figure 3

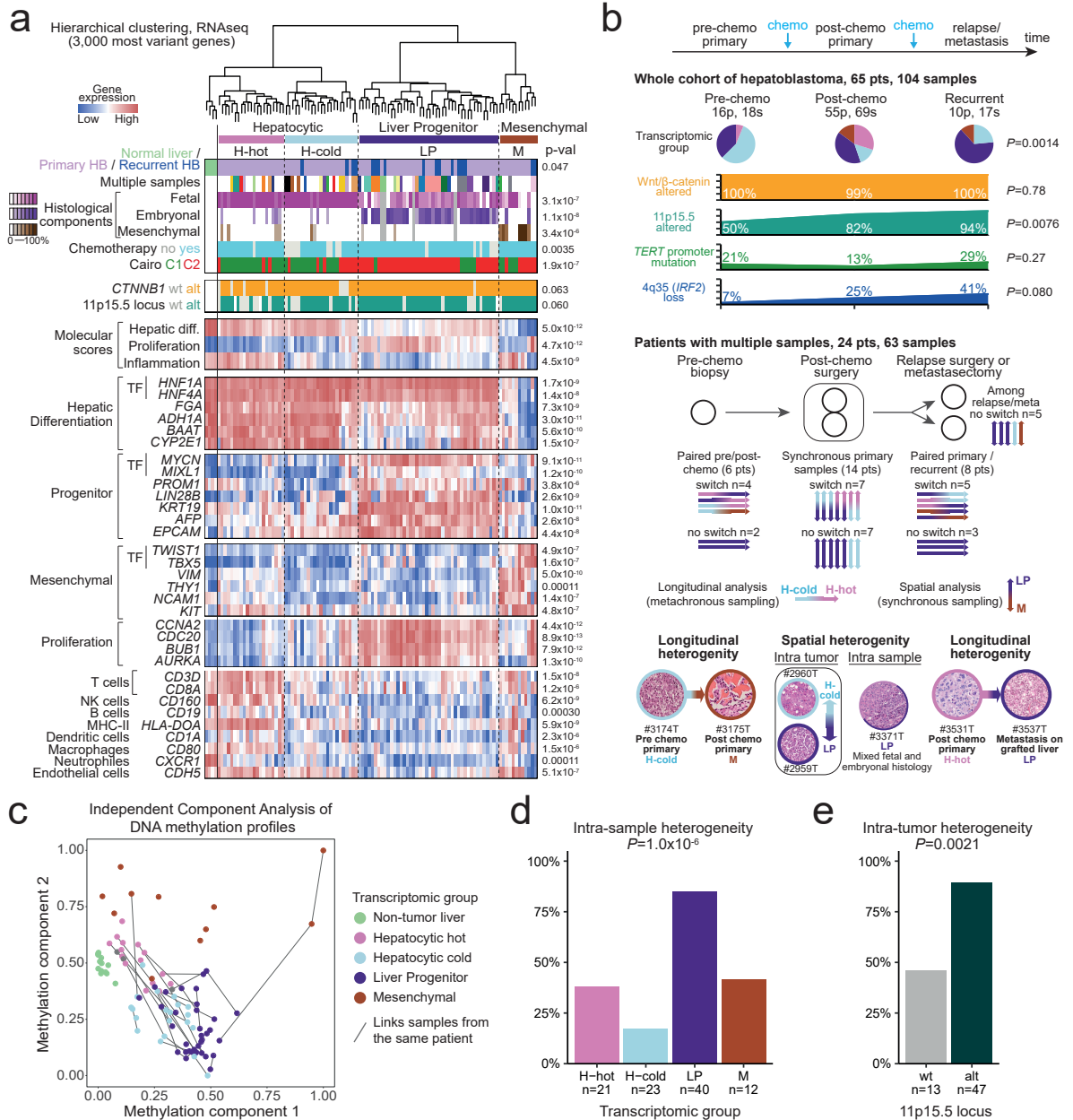


Figure 3. Molecular plasticity of hepatoblastoma across three differentiation states.

a Gene expression-based classification of hepatoblastoma. Unsupervised hierarchical clustering of 100 HB samples from 64 patients and 4 non-tumor liver samples revealed 4 molecular groups. Clinical and molecular annotations are depicted below the dendrogram, with p-values indicating their association with molecular groups. A heatmap shows the expression of key transcription factors (TF) and marker genes representative of each group, as well as molecular scores of hepatic differentiation, cell proliferation and immune infiltration. **b** Top: Proportion of transcriptomic groups and a selection of driver alteration frequencies at different steps of hepatoblastoma progression. Middle: Transcriptomic group switches identified in 24 patients with multiple sampling, including patients with pre/post-chemotherapy samples, synchronous samples at distinct locations in the primary tumor, and/or paired primary and relapse/metastasis. The number of patients with/without molecular switch is indicated for each type of multiple sampling,

and the transcriptomic switch is represented by a color code on the arrows. Bottom: Examples of histological heterogeneity matching transcriptomic group switches. **c** Projection of hepatoblastomas and non-tumor liver samples over two independent methylation components. Hepatoblastomas are colored by their transcriptomic group, and samples from a same patient are linked with black lines. **d** Association of intra-sample histological heterogeneity with transcriptomic groups. **e** Association of intra-tumor histological heterogeneity with 11p15.5 locus alteration.

Figure 4

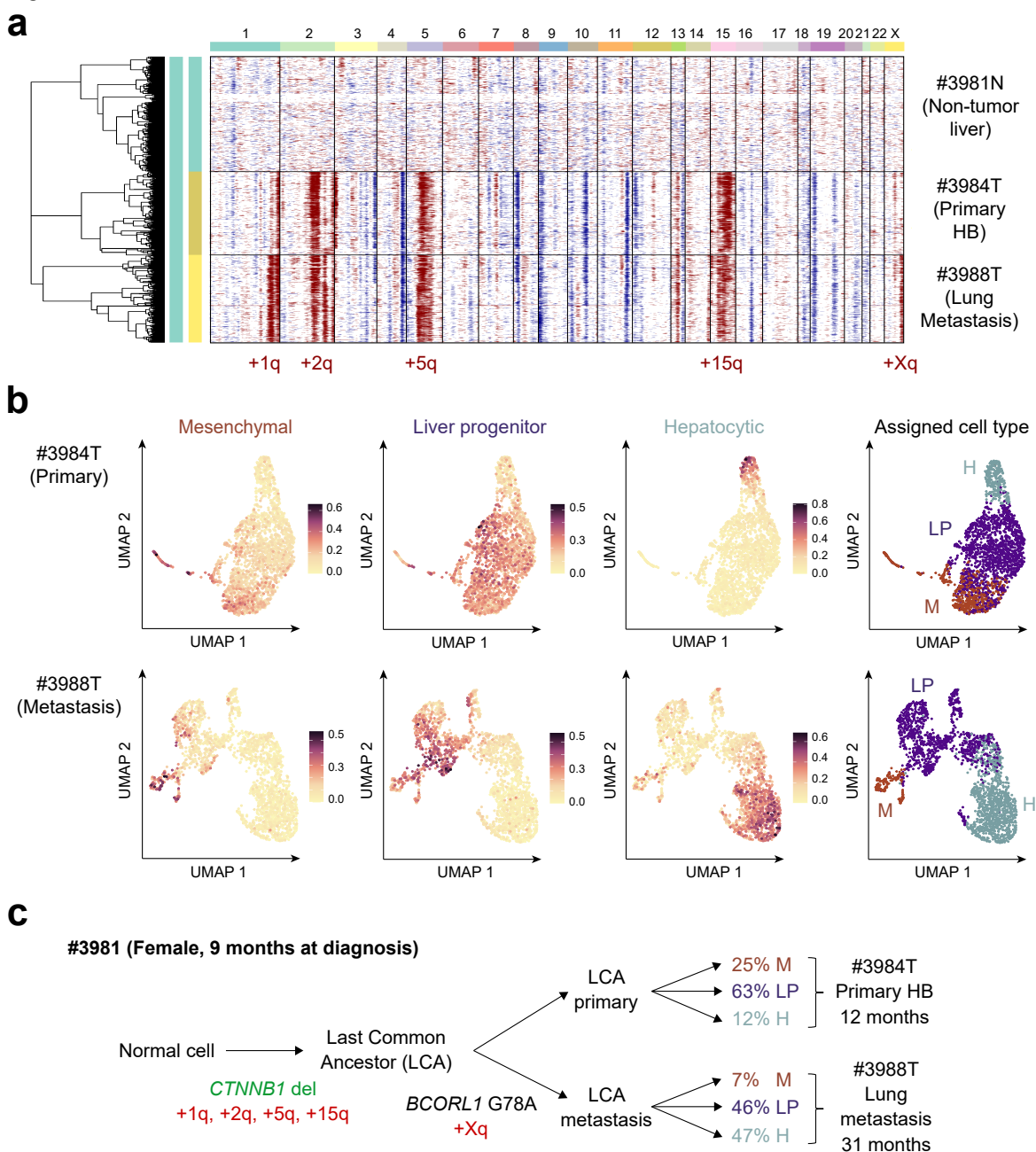


Figure 4. Single-nucleus RNA-seq reveals molecular plasticity along tumor progression in one patient. **a** Virtual copy-number profiles discriminate tumor and normal cells. In agreement with WES data, all tumor cells display gains at 1q, 2q, 5q and 15q whereas the Xq gain is specific to metastatic cells. **b** Uniform manifold approximation and projection (UMAP) of tumor nuclei in the primary tumor (top) and metastasis (bottom). Mean expression of marker genes for each differentiation state are displayed with a color code. The last UMAP on the right shows the assignment of each cell cluster to 'Mesenchymal', 'Liver Progenitor' or 'Hepatocytic' cell type based on the expression of marker genes. **c** Schematic representation of genetic and non-genetic evolution in patient #3981. LCA: Last Common Ancestor; M: Mesenchymal; LP: Liver Progenitor; H: Hepatocytic.

Figure 5

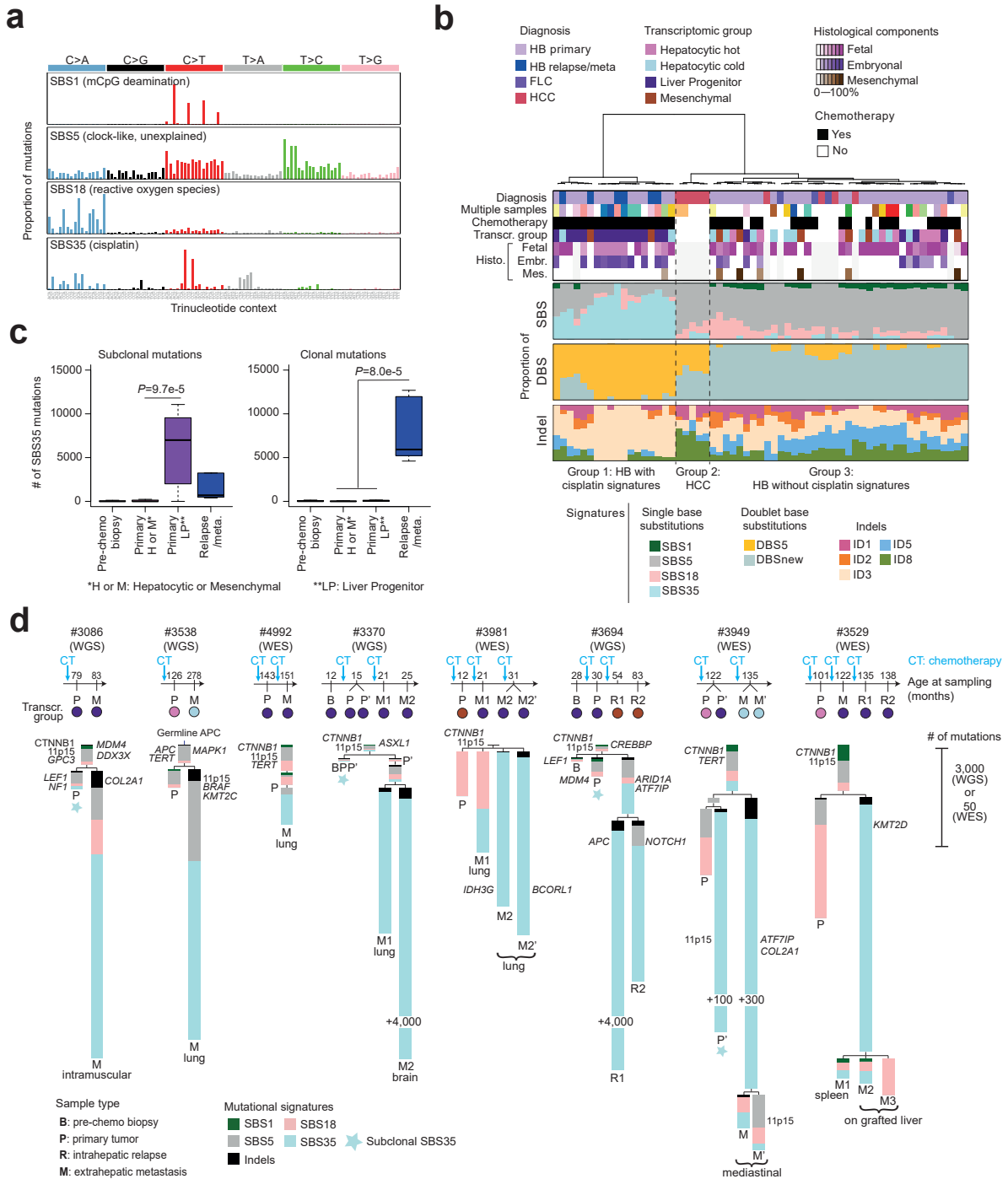


Figure 5. Massive load of cisplatin-induced mutations in chemoresistant hepatoblastomas. **a** Single base substitution (SBS) signatures identified in pediatric liver cancers (PLC). Each signature is displayed according to the 96-substitution classification defined by substitution type and sequence context immediately 5' and 3' to the mutated base. **b** Unsupervised classification of 65 PLC genomes based on their mutational signature exposures. Clinical and molecular annotations are depicted below the dendrogram. Bar graphs indicate the proportion of the 4 single base substitution (SBS), 2 doublet base substitution (DBS) and 5 indel (ID) signatures in each sample. **c** Number of subclonal (left) and clonal (right) mutations attributed to signature SBS35 (cisplatin) in

hepatoblastoma samples according to sample type (pre-chemotherapy biopsy, primary tumor or relapse / metastasis) and molecular group (Hepatocytic or Mesenchymal vs. Liver Progenitor). **d** Phylogenetic trees reconstructed for 8 HB patients with primary tumors and relapses / metastases analyzed by WGS or WES. The time and molecular group of each sample is shown above the trees, together with chemotherapy treatment. Driver alterations are indicated. Branch lengths are proportional to the number of mutations acquired, with a color code indicating the contribution of each mutational signature.

Figure 6

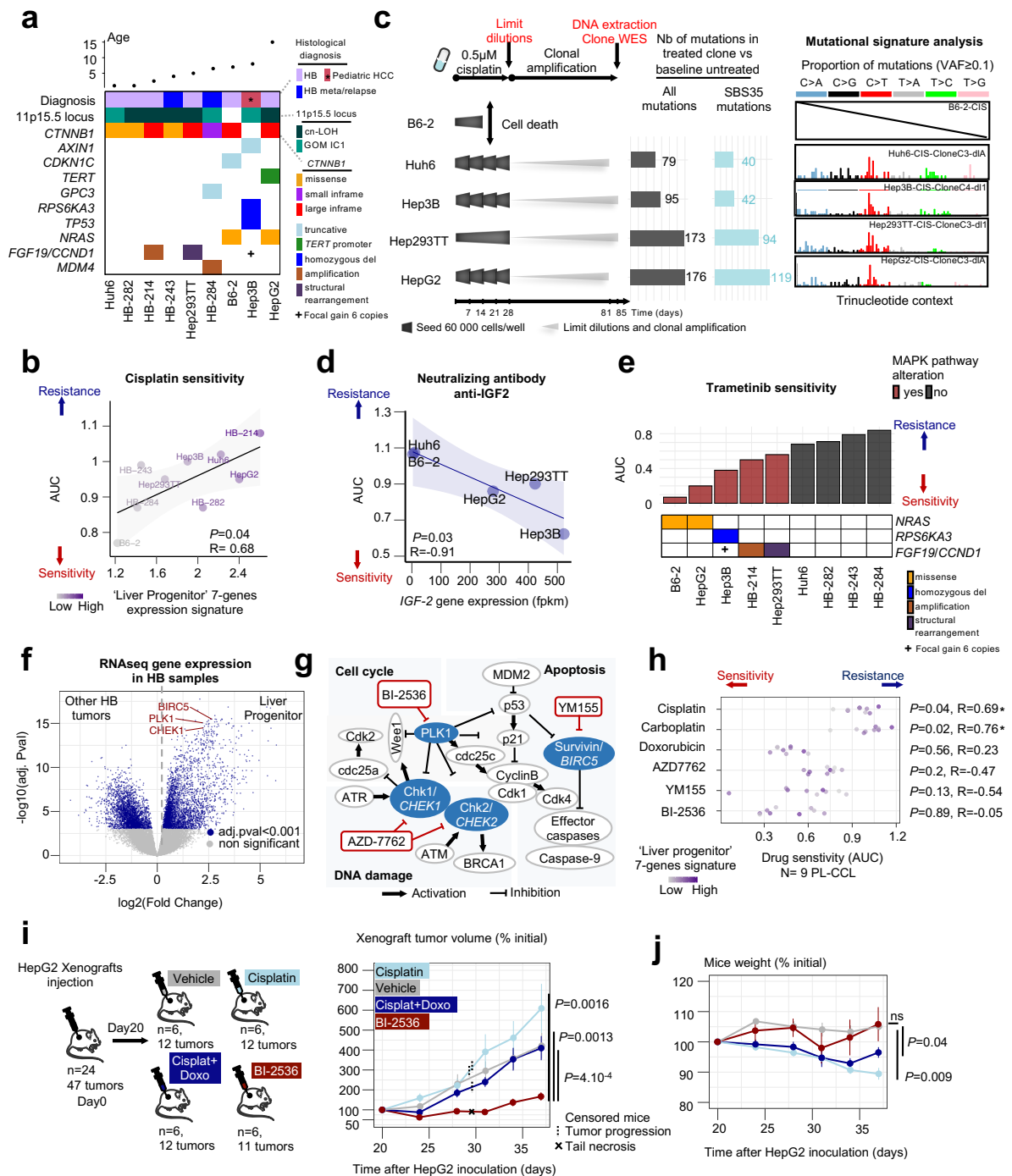


Figure 6. New therapeutic strategies targeting hepatoblastoma based on molecular features. **a** Heatmap representation of somatic alterations in 9 pediatric liver cancer cell lines (PL-CCL) and corresponding patient age at diagnosis (Top). **b** Correlation between sensitivity to cisplatin assessed with area under the curve (AUC) and 'liver progenitor' 7-genes expression signature. Color gradient intensity reflects levels of 'Liver Progenitor' signature. Pearson correlation test was performed. **c** Accumulation of SBS35 signature in 4/5 PL-CCL. Trapezium indicates cell passaging and triangle symbolized cell line limit dilution and clonal expansion. **d** Correlation between sensitivity to neutralizing antibody anti-IGF2 (AUC) and *IGF2* expression in 5 PL-CCL (Pearson test). **e** Trametinib sensitivity

(AUC) in 9 PL-CCL (top) and corresponding genetic alterations in Ras MAPK pathway (bottom). **f** Volcano plot representing differentially expressed genes in HB belonging to Hepatocytic (H-hot and H-cold) and 'Liver Progenitor' transcriptomic subgroups. **g** PLK1, BIRC5 and CHEK1 interactions and roles in cell cycle, DNA damage response and apoptosis. **h** Drug response assessed with AUC in 9 PL-CCL. Color gradient intensity reflects levels of 'Liver Progenitor' signature. Pearson correlations are performed between AUC and levels of 'Liver progenitor' signature. **i** Sensitivity to Cisplatin, Doxorubicin and BI-2536 *in vivo* in HepG2 xenografts from 24 nude mice and **j** mice weight follow-up. Mann-Whitney- Wilcoxon tests were performed at day 37.