Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets

Kornelius Schulze[#,1,2,3,4], **Sandrine Imbeaud**[#,1,2,3,4], **Eric Letouzé**[#,1,2,3,4], **Ludmil B Alexandrov**[5,6], **Julien Calderaro**[1,2,3,4,7], **Sandra Rebouissou**[1,2,3,4], **Gabrielle Couchy**[1,2,3,4], **Clément Meiller**[1,2,3,4], **Jayendra Shinde**[1,2,3,4], **Frederic Soysouvanh**[1,2,3,4], **Anna-Line Calatayud**[1,2,3,4], **Roser Pinyol**[8], **Laura Pelletier**[1,2,3,4], **Charles Balabaud**[9], **Alexis Laurent**[10,11], **Jean-Frederic Blanc**[9,12], **Vincenzo Mazzaferro**[13], **Fabien Calvo**[1,2,3,4], **Augusto Villanueva**[8,16], **Jean-Charles Nault**[1,2,3,4,14], **Paulette Bioulac-Sage**[9,15], **Michael R Stratton**[5], **Josep M Llovet**[8,16,17], and **Jessica Zucman-Rossi**[1,2,3,4,18]

[1]Inserm, UMR-1162, Génomique fonctionnelle des Tumeurs solides, Equipe Labellisée Ligue Contre le Cancer, Institut Universitaire d'Hematologie, Paris, F-75010 France

[2]Université Paris Descartes, Labex Immuno-Oncology, Sorbonne Paris Cité, Faculté de Médecine, Paris, France

[3]Université Paris 13, Sorbonne Paris Cité, Unité de Formation et de Recherche Santé, Medecine, Biologie humaine, F-93000 Bobigny, France

[4]Université Paris Diderot, F-75013 Paris

[5]Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK

[6]Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America

[7]Assistance Publique-Hôpitaux de Paris, Department of Pathology, Centre hospitalier universitaire Henri Mondor, F-94000 Créteil, France

[8]Hepatocellular Carcinoma Translational Research Laboratory, Barcelona-Clínic Liver Cancer Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Liver Unit. Centro de Investigación Biomédica en Red: Enfermedades Hépaticas y Digestivas; Hospital Clínic, Barcelona, Catalonia, Spain

[9]Inserm, UMR-1053; Université de Bordeaux, Bordeaux, F-33076, France

[10]Assistance Publique-Hôpitaux de Paris, Department of Digestive and Hepatobiliary Surgery, Centre hospitalier universitaire Henri Mondor, F-94000 Créteil, France

[11]Inserm, UMR-955, F-94000 Créteil, France

[12]Centre hospitalier universitaire de Bordeaux, Department of Hepatology, Hôpital Saint-André, Bordeaux, F-33076, France

[13]Department of Liver Surgery and Transplant, Fondazione Istituto Tumori, via Venezian 1, 20133 Milan, Italy

[14]Assistance Publique Hôpitaux de Paris, Hôpitaux Universitaires Paris – Seine Saint-Denis, Site Jean Verdier, Pôle d'Activité Cancérologique Spécialisée, Service d'Hépatologie, F-93143 Bondy, France

[15]Centre hospitalier universitaire de Bordeaux, Pellegrin Hospital, Department of Pathology, Bordeaux, F-33076, France

[16]Mount Sinai Liver Cancer Program (Division of Liver Diseases), Mount Sinai School of Medicine, New York, USA

[17]Institució Catalana de Recerca i Estudis Avançats, Barcelona. Catalonia, Spain

[18]Assistance Publique-Hôpitaux de Paris, Hopital Europeen Georges Pompidou, F-75015 Paris, France

[#] These authors contributed equally to this work.

Abstract

Genomic analyses promise to improve tumor characterization in order to optimize personalized treatment for patients with hepatocellular carcinoma (HCC). Exome sequencing analysis of 243 liver tumors revealed mutational signatures associated with specific risk factors, mainly combined alcohol/tobacco consumption, and aflatoxin B1. We identified 161 putative driver genes associated with 11 recurrent pathways. Associations of mutations defined 3 groups of genes related to risk factors and centered on *CTNNB1* (alcohol), *TP53* (HBV), and *AXIN1*. Analyses according to tumor stage progression revealed *TERT* promoter mutation as an early event whereas *FGF3, FGF4, FGF19/CCND1* amplification, *TP53* and *CDKN2A* alterations, appeared at more advanced stages in aggressive tumors. In 28% of the tumors we identified genetic alterations potentially targetable by FDA-approved drugs. In conclusion, we identified risk factor-specific mutational signatures and defined the extensive landscape of altered genes and pathways in HCC which will be useful to design clinical trials for targeted therapy.

Hepatocellular carcinoma (HCC) is a heterogeneous disease which usually develops within liver cirrhosis related to various etiologies. Hepatitis B virus (HBV) infection, with or without aflatoxin B1 (AFB1) exposure, is the most frequent etiology in Asia and Africa, whereas hepatitis C virus (HCV), chronic alcohol abuse, or metabolic syndrome are frequently related to HCC in Western countries[1,2]. In cirrhosis, hepatocarcinogenesis is a multi-step process where pre-cancerous dysplatic macronodules (DMN) transform into early HCC that progress into "small and progressed HCC" then leading to advanced HCC[3,4]. In rare cases, HCC develops in normal liver with some of these tumors potentially resulting from malignant transformation of hepatocellular adenomas (HCA)[5-7]. Unraveling the patterns of genomic alterations in these heterogeneous tumors is pivotal towards identifying targeted therapies that could improve patient care[8,9].

To understand HCC diversity we analyzed the whole coding sequences of 243 liver tumors surgically resected in Europe (France, Italy and Spain) associated with cirrhosis (METAVIR F4, n=118), fibrosis (F2-F3, n=46) or non-fibrotic livers (F0-F1, n=79)[10]. Various risk factors were identified, including alcohol intake (41%), HCV (26%), nonalcoholic steatohepatitis as co-occurrence of metabolic syndrome (18%), HBV (14%), hemochromatosis (7%), or no known etiology (11%). The 118 tumors associated with cirrhosis represented different stages along HCC progression: 7 DMN, 7 early, 17 small and progressed, 58 classic, and 29 poor-prognosis HCC (Supplementary Tables 1, 2 and **Online Methods** for clinical definitions). Exome sequencing was performed in tumors and matched non-tumor liver samples to a mean depth of 72-fold (Supplementary Fig. 1). We identified 28,478 somatic mutations, 6,184 of which occurred in a single tumor with a hypermutation phenotype (Supplementary Table 3). Excluding this sample, we identified a median of 21 silent and 64 non-silent mutations per tumor (ranging from 1 to 706) corresponding to a mean somatic mutation rate in coding sequences of 1.3 mutations per megabase, consistent with previous reports[11-13].

We analyzed the mutational spectrum of the 243 liver tumors to identify mutagenic processes operative in HCC. *De novo* signature analysis using the Wellcome Trust Sanger Institute mutational signatures framework revealed 4 signatures (Supplementary Fig. 2, 3). Two signatures were previously identified in a pan-cancer analysis (i.e., signatures 6 and 16)[14,15] and two were novel (signatures 23 and 24). Mutational catalogues derived from exome data may underestimate the number of operative signatures[16], and 8 signatures were previously identified by analyzing 88 whole liver cancer genomes (signatures 1A, 1B, 4, 5, 6, 12, 16, and 17)[14]. To determine the complete set of signatures in our series, we evaluated the contribution of these 8 previously identified signatures and the 2 novel signatures identified *de novo* (see **Online Methods**). Altogether, 8 signatures (all except signatures 12 and 17) were found at different strengths in the 243 samples (Fig. 1a, Supplementary Table 4).

Hierarchical clustering of the samples, based on the contributions of the mutational signatures in each sample, revealed 6 groups of tumors (MSig1 to 6) and 4 singletons which were significantly associated with demographic, etiologic and molecular features (Fig. 1b-d, Supplementary Tables 4, 5). Msig1 included 9 HCC enriched in signature 4 (frequent C>A and dinucleotide mutations) which has been previously shown to be associated with

tobacco[14,17]. The prevalence of smokers in Msig1 was high (57%) but not significantly different from the other groups. Msig2 group (5 HCC) was characterized by the novel "signature 24" showing a high rate of C>A mutations. All of the patients in the Msig2 group were migrants born in subtropical African countries and infected by HBV. Three of these tumors displayed a somatic R249S *TP53* mutation typical of AFB1-exposed HCC[18,19] suggesting that signature 24 reflects the mutational pattern induced by AFB1. There was an equal prevalence of mutations for signatures 5 and 16 (49% each) in MSig3. This particular group included 19 highly mutated HCC associated with alcohol and tobacco exposure (67%, $P = 6\times10^{-5}$, Fisher's exact test), *CTNNB1* mutation (74%, $P = 8\times10^{-4}$), and non-cirrhotic livers (79%, $P = 0.02$). These results suggest a genotoxic synergistic effect of alcohol and tobacco exposures in 8% of the patients that developed a homogeneous molecular subtype of tumors. MSig4 tumors were characterized by frequent *TP53*-mutations and signature 5, enriched in T>C transitions. Mutations in the 46 tumors classified in Msig5 were related to signatures 1A and 1B resulting of spontaneous deamination at NpCpG trinucleotides acquired through the patient's lifetime[14]. This group of tumors demonstrated a low mutation rate and an early histological stage, but these tumors also had a lower tumor cell content, so this group may partly be related to normal cell contamination (see **Online Methods**). Signature 16 accounted for 79% of mutations in MSig6 tumors and was associated with older patients, *TERT* and *CTNNB1* mutations. Among the tumors classified alone, one (BCM723T) showed signature 6 which is characteristic of cancers with defective DNA mismatch repair, displaying numerous indels. The newly identified signature 23 was only encountered in the hypermutated tumor CHC892T, occuring in a 71 year old female presenting with a non-fibrotic liver having black anthracotic pigment deposition, predominantly in macrophages and vessels (Supplementary Fig. 4). This signature was characterized by a predominance of C>T mutations with a strong strand bias that may result from the interplay between an unknown mutagenic process affecting predominantly guanine residues and transcription coupled nucleotide excision repair[15]. To validate the mutational signatures identified in our series, we analyzed the mutation patterns identified by exome sequencing in the ICGC-Japan (452 tumors)[20] and the TCGA cohorts of 198 tumors. Consistent with the findings by Totoki *et al.*, we found an enrichment of T>C mutations in an ApTpN context in the Japanese data. We also found two tumors with mutational patterns corresponding to our signatures 6 and 23, validating the occurrence of these signatures in rare HCC cases (Supplementary Fig. 5). Six tumors of the TCGA data clustered with our MSig2 cases and displayed patterns similar to the AFB1-related signature 24. These cases were all African or Asian ($P = 0.002$, Fisher's exact test) and 3 of them displayed the characteristic R249S *TP53* mutation ($P = 8\times10^{-4}$, Fisher's exact test) (Supplementary Fig. 6).

We used MutSigCV to identify cancer driver genes[21]. Fourteen genes were significantly enriched for damaging mutations (q < 0.05, Fig. 2a): *TP53*, *CTNNB1*, *AXIN1*, *ALB*, *ARID1A*, *ARID2*, *ACVR2A*, *NFE2L2*, *RPS6KA3*, *KEAP1*, *RPL22*, *CDKN2A*, *CDKN1A* and *RB1* (Supplementary Table 6). We also analyzed copy-number alterations (CNAs) by comparing the sequence coverage in 243 tumors and matched non-tumor liver samples. The pattern of broad gains and losses was consistent with previous reports in HCC (Fig. 2b)[11,22,23]. We identified recurrent homozygous deletions of the *CFH* locus, *IRF2*,

CDKN2A, *PTPN3*, *PTEN*, *AXIN1* and *RPS6KA3*, and recurrent focal amplifications of *TERT*, *VEGFA*, *MET*, *MYC*, the *FGF3, FGF4, FGF19/CCND1* locus, *JAK3* and *CCNE1* (Supplementary Tables 7a,b). Next, we developed a pipeline integrating focal CNAs and mutations identifying 161 putative driver genes in liver cancer (see **Online Method**, Fig. 2c, Supplementary Table 8).

To identify cellular pathways associated with HCC we annotated the 161 candidate driver genes using the Gene Ontology database and manually curated the precise role of each gene in pathways (see **Online Methods**). We found 11 pathways altered in 5% of HCC (Fig. 3 and Supplementary Fig. 7,8,9): *TERT* promoter mutations activating telomerase expression (60%)[24], WNT/ß-Catenin (54%), PI3K/AKT/mTOR (51%), TP53/cell cycle (49%), MAP kinase (43%), hepatic differentiation (34%), epigenetic regulation (32%) chromatin remodeling (28%), oxidative stress (12%), Il6/JAK/STAT (9%), and TGFß (5%). This analysis revealed new genes recurrently mutated in HCC: ß-catenin inhibitors (*ZNRF3, USP34* and *MACF1*), hepatocyte secreted proteins(*APOB* and FGA), and the TGFß-receptor *ACVR2A*, recently associated with chondrosarcoma[25]. We then identified significant associations between mutations and risk factors (Supplementary Table 9). Alcohol-related HCC were significantly enriched in *CTNNB1*, *TERT*, *CDKN2A*, *SMARCA2*, and *HGF* alterations ($P < 0.05$, chi-square tests for trend in proportions). HBV-related HCC were frequently mutated for *TP53*, and *IL6ST* mutations were exclusively identified in HCC with no known etiology. In contrast, HCV infection, metabolic syndrome, and hemochromatosis did not show significant associations.

Because each tumor accumulates numerous damaging mutations, we identified three major clusters of associated alterations, centered on *CTNNB1, AXIN1* and *TP53* (Fig. 4, Supplementary Fig. 10, and Supplementary Table 10). Interestingly, alterations of genes belonging to a same pathway were frequently distributed in different clusters. These results may reflect cooperation, functional redundancy, or lethality of gene combinations and may contribute to better predict efficacy of targeted therapies.

Next, we defined the catalog of actionable genomic alterations among the 11 major pathways. Altogether, 28% of patients harbored at least one damaging alteration potentially targetable by an FDA-approved drug (Fig. 3, Supplementary Fig. 11 and Supplementary Table 11) and 86% by a drug studied in phase I to phase III clinical trials (Supplementary Fig. 11 and Supplementary Table 12). Targetable alterations by FDA-approved drugs comprised focal amplifications or mutations of *FLTs* (6%), *FGF3/4/19* (4%), *PDGFRs* (3%), *EPHA4* (3%), *JAK3* (3%), *VEGFA* (1%), *HGF* (3%), *MTOR (2%), EGFR* (1%), *FGFRs* (1%), *IL6R* (1%), *KIT* (1%), *MET* (1%), *TEK* (1%), *BRAF* (<1%), *ERBB2* (<1%), *JAK1* (<1%), *KDR* (<1%). We also showed *in vitro* that inactivating mutations of *RPS6KA3* (7%) induced an activation of the RAS/MAPK pathway with an over-expression of phosphorylated ERK1/2 (Supplementary Fig. 12) suggesting that RPS6KA3-mutated HCC could be targeted by ERK or MEK inhibitors. Notably, advanced-stage tumors harbored more potentially targetable alterations, including, in particular, *FGF/CCND1* amplifications (Supplementary Fig. 13).

Additionally to driver genes that can be directly inhibited by targeted drugs, other alterations may potentiate drug sensitivity of cancer cells. This includes NQO1, which markedly increases sensitivity to the HSP90 inhibitor, 17-AAG, by reducing this compound to a more potent inhibitor[26,27]. NQO1 expression is induced by the oxidative stress pathway, which is activated in 12% of HCC, primarily by mutations of *KEAP1* or *NFE2L2*. To test whether these tumors could be more sensitive to HSP90 inhibitors, we assessed the response of 29 liver cancer cell lines to 17-AAG and 17-DMAG (Fig. 5a). The growth inhibition of 50% (GI50) was significantly inversely correlated with the expression of NQO1 (Pearson's r = −0.56, *P* = 0.0015, Fig. 5b). Two of the three cell lines harboring *KEAP1* inactivating mutations were highly sensitive to HSP90 inhibitors, whereas the third mutated cell line, which was less sensitive, was homozygous for the *NQO1* P187S *(NQO1*2 allele, rs1800566)* missense variant causing NQO1 deficiency (Fig. 5c)[28]. These findings suggest that tumors with high NQO1 expression may be more sensitive to HSP90 inhibitors except in patients with homozygous P187S genotype.

Finally, we explored the progression of HCC in cirrhotic and non-cirrhotic livers. We identified an increased number of gene mutations ($P = 1.2 \times 10^{-3}$, Jonckheere-Terpstra test) and chromosome aberrations ($P = 1.3 \times 10^{-5}$, Jonckheere-Terpstra test) along DMN malignant transformation to poor prognosis HCC (Fig. 6a). Although *TERT* promoter mutations were already frequent at early stages, *CTNNB1* and *TP53* mutation frequencies increased significantly with progression, and focal amplifications at the *CCND1/FGF* locus were mostly encountered in poor prognosis HCC (*P* < 0.01, Chi-square test for trend in proportions, Fig. 6b). Interestingly, chromosome aberrations appearred later than gene mutations during progression. While similar findings were observed in tumors which developed in non-fibrotic liver relative to the progressive accumulation of mutations and CNAs, *TERT* promoter mutations were later events during malignant transformation. Moreover, *HNF1A* and *IL6ST* mutations were restricted mostly to HCA suggesting that most of the HCC in non-fibrotic liver did not derive from the transformation of an adenoma (Fig. 6b). A link between *FGF19* expression and overall survival has been described in the literature[22,24,29-32]. Using a multivariate survival analysis, we identified that *CDKN2A* inactivation and *FGF/CCND1* amplification were associated with poor prognosis in our cohort of resected HCC, independently of classical prognostic clinical and histological features (Fig. 6c and Supplementary Table 13).

In conclusion, our study identified relationships between environmental exposures and mutational patterns in HCC as well as the landscape of driver genes and pathways altered in different clinical stages and etiological backgrounds. For patient care, genomic alterations identified in targetable genes will be useful to determine HCC patients that could potentially benefit from targeted treatment in future clinical trials.

Online Methods

Liver Samples

A series of 243 liver tumor samples and their non-tumor counterparts were collected from patients surgically treated in Europe: 193 cases from France (Créteil, Bordeaux), 9 from Spain (Barcelona), and 41 from Italy (Milan). The study was approved by institutional

review board committees (CCPRB Paris Saint-Louis, 1997, 2004, and 2010, approval number 01-037; Bordeaux 2010-A00498-31). Written informed consent was obtained in accordance with french legislation. All samples were immediately frozen in liquid nitrogen and stored at −80 °C.

Clinical Data

Clinicopathological data were available in all cases. Risk factors were defined by significant alcohol intake, HCV, HBV, hemochromatosis, with no known etiology, and non-alcoholic steatohepatitis (NASH) as co-occurrence of metabolic syndrome. Metabolic syndrome consists of a combination of disorders including central obesity [waist circumference >102 cm (M), >88 cm (F)], hypertriglyceridemia (triglycerides >150 mg/dL), low high-density lipoprotein serum levels (<40 mg/dL), arterial hypertension (>130 mmHg systolic or >85 mmHg diastolic), and raised fasting plasma glucose (FPG) 1.1 mg/dL or previously diagnosed type 2 diabetes. At least three of the latter criteria have to be fulfilled for diagnosis. NASH can be the chronic consequence of a non-alcoholic fatty liver disease (NAFLD), which frequently co-occurs in patients with metabolic syndrome and is characterized by hepatocellular accumulation of triglycerides in the absence of significant alcohol consumption. By contrast, NASH additionally includes the presence of inflammation and can display different degrees of fibrosis. Patients without known etiology are those that do not display the above frequent etiologies or rare etiologies (such as primary biliary cirrhosis, autoimmune hepatitis and primary sclerosing cholangitis). Samples were classified according to the clinical, pathological and genetic features as previously described by Guichard *et al.* (Supplementary Table 1, 2)[11]. In all HCC samples, the ratio of tumor cells/non-tumor cells was evaluated >50%; the PurBayes method[33] estimated an average 70% (range 39-100%) of tumor purity based upon sequencing data (Supplementary Table 14). Definition of DMN, early, small and progressed, classic, and poor prognosis HCC is based on histopathological criteria in HCC proposed by the International Consensus Group for Hepatocellular Neoplasia[3,4]. DMN: macronodules containing low (a cell population lacking architectural atypia with mild increase in cellularity as compared to surroundings; portal tracts detectable) or high (frank cytological and architectural atypia as compared to surroundings but insufficient for a diagnosis of malignancy; portal tracts detectable) grade dysplastic nodules. Early HCC: diameter 2 cm, vaguely nodular lesion with indistinct margins with a well differentiated histology which may require careful distinction from high grade dysplastic nodules; few portal tracts detectable. Small and progressed HCC: diameter 2 cm, distinctly nodular lesion with well (G1) to moderately (G2) differentiated histology in which malignancy is recognized at first glance; no portal tracts detectable. Poor prognosis HCC: HCC cases displaying recurrence within 2 years. Classic HCC: non-early, non-small, non-poor prognosis HCC.

Genomic DNA extraction

We extracted DNA using a salting-out procedure[34]. Genomic DNA was loaded on a 0.8% agarose gel for quality control, only DNA >10Kb were selected. DNA quantification was performed using Hoechst 33258 from Sigma Chemical Co. (St. Louis, MO, USA).

Exome capture, library construction and sequencing

Sequence capture, enrichment and elution from 243 pairs of genomic DNA were performed by IntegraGen (Evry, France) as previously described in Guichard et al. with some modifications[11]. Agilent in-solution enrichment was used with their biotinylated oligonucleotides probes library (SureSelect Human All-Exon kit v2-46Mb (n=36 pairs); v3-52Mb (n=7 pairs); v4-70Mb (n=56 pairs); v5+UTRs-75Mb (n=144 pairs), Agilent technologies) according to manufacturer's instruction. Eluted-enriched DNA sample was sequenced on an Illumina HiSeq 2000 sequencer as paired-end 75b reads as previously described[35]. Image analysis and base-calling was performed using Illumina Real Time Analysis (RTA) Pipeline v1.12 with default parameters. Whole-exome sequencing pre-analysis was based on the Illumina pipeline (CASAVA1.8.2). Only the positions included in the bait coordinates were conserved. Each sample was sequenced to an average depth of 72.0X, with ~96.9% of the targeted regions covered 1×, ~92.6% 10× and ~82.9% 25× (Supplementary Fig. 1).

Identification of somatic variants

A list of variants was generated (Supplementary Table 3) considering only somatic mutations in coding regions plus consensus intronic bases (missense/nonsense/splice-site/ indels/synonymous mutations) as previously described in Guichard *et al.* with some modifications[11]. Polymorphisms referenced in dbSNP135 or 1000Genomes with minor allele frequency over 2% were removed. Functional evidence of predictive drastic consequence of the variants was investigated using Polyphen-2 v2.2.2[36]. A total of 11,823 (41%) putative somatic mutations were validated manually using the Integrated Genomics Viewer (IGV) and 3,126 (11%) using Sanger sequencing (Supplementary Table 3). A systematic Sanger sequencing was performed in a subset of 155 samples on a list of 11 genes (*CTNNB1*, *TP53*, *ARID1A*, *AXIN1*, *RPS6KA3*, *CDKN2A*, *NFE2L2*, *ARID2*, *PI3KCA*, *KRAS* and *KEAP1*) and we used these data to benchmark our exome mutation calling pipeline. A total of 151 somatic variants were called by both methods, whereas 21 variants were only identified by Sanger sequencing and 10 by exome sequencing. Variants not identified by exome sequencing were mostly mutations in the poorly covered GC-rich exon 1 of *ARID1A* (n=10 variants) and large deletions of *CTNNB1* exon 3 (n=3). This results in a sensitivity of our somatic calling pipeline of 88% [82-92%, IC95] that reaches 95% [90-98%, IC95] when excluding those 2 specific regions, and a specificity of 99% [98-100%, IC95].

Mutations were annotated using the Alamut Batch, Alamut Visual v2.4 (Interactive biosoftware, France) and Oncotator (http://www.broadinstitute.org/cancer/cga/Oncotator/). All sequences has been deposited in the EGA (European genome-phenome archive - http:// www.ebi.ac.uk/ega/) database (accessions EGAS00001000217, EGAS00001000679 and EGAS00001001002) and ICGC data portal (http://dcc.icgc.org/, release 18, Dec10th, 2014).

De novo mutational signature analysis

The mutational catalogues of the 243 liver tumors were analyzed using the Wellcome Trust Sanger Institute mutational signatures framework[16]. This algorithm makes use of a well-known blind source separation technique, termed nonnegative matrix factorization (NMF).

NMF identifies the matrix of *N* mutational signature, *P*, and the matrix of the exposures of these signatures, *E*, by minimizing a Frobenius norm while maintaining non-negativity:

$$\min_{\substack{P \in \mathbb{M}_{\mathbb{R}_+}^{(\dot{K},N)} \quad E \in \mathbb{M}_{\mathbb{R}_+}^{(N,G)}}} \parallel \mathrm{M} - P \times E \parallel_F^2$$

The method for deciphering mutational signatures, including evaluation with simulated data and list of limitations, can be found in ref [16].

First, all mutation data were converted into a matrix, *M*, that is made up of 96 features comprising mutations counts for each mutation type (C>A, C>G, C>T, T>A, T>C, and T>G; somatic mutations presented in a pyrimidine context) using each possible 5′ (C, A, G, T) and 3′ (C, A, G, T) context for all samples. Then, the algorithm was applied to the matrix that contains *K* mutation types and *G* samples, deciphering the minimal set of mutational signatures that optimally explains the proportion of each mutation type, thus estimating the contribution of each signature to each sample.

After extraction, 4 stable and reproducible mutational signatures were deciphered (see stability and error plot in Supplementary Fig. 2) and termed Signatures A, B, C, and D. Those signatures were compared to the catalogue of 27 consensus signatures that were previously identified by a pan-cancer analysis encompassing 7,042 samples and 30 cancer types[14]. The comparison was performed by cosine similarity as described[16] as well as based on other biological features exhibited by the mutational signatures (e.g., transcriptional strand bias, presence of small insertions and/or deletions at specific context, *etc.*).

The comparison revealed that 2 of the 4 signatures are novel, while the patterns of the remaining 2 were previously identified through the pan-cancer analysis. Notably, Signatures C exhibited a strong transcriptional strand bias of T>C mutations especially at TpA dinucleotides (70% vs. 30%) a behaviour consistent with the one of Signature 16. The pattern of mutations of Signature C showed a cosine similarity of 0.95 with the pattern of mutations of Signature 16. Similarly, Signature D's pattern is extremely similar to the one of Signature 6 (cosine similarity of 0.90), both exhibiting prevalence for high numbers of indels at repetitive elements. In contrast, Signatures A and B had cosine similarity <0.90 with any of the previously identified mutational signatures. Please note that we previously used a cut-off of ~0.90 to cluster mutational signatures into consensus mutational signatures[14].

Reintroduction of consensus mutational signature analysis and sample clustering

Previous analysis of 88 whole-genomes of liver cancers has revealed 6 mutational signatures: Signatures 1B, 4, 6, 12, 16, and 17[14]. Further, signature 1B can be decomposed into Signatures 1A and 5[14]. The *de novo* extraction of mutational signatures of our 243 liver exomes revealed two additional mutational signatures, termed signatures 23 and 24. Thus, the complete compendium of mutational signature that can be present in a liver cancer is: Signatures 1A, 1B, 4, 5, 6, 12, 16, 17, 23, and 24. To evaluate the presence of all these

signatures in the 243 liver exomes, we used a previously described approach to find the *Exposure* matrix minimizing the following constrained linear function for each sample[37]:

$$\min_{Exposures_i \geq 0} \| SampleMutations - \sum_{i=1}^{N} \left(\overrightarrow{Signature_\iota} * Exposure_i \right) \|_F^2$$

Here, $\overrightarrow{Signature_\iota}$ represents a vector with 96 components (corresponding to the six somatic substitutions and their immediate sequencing context) and *Exposure_i* is a nonnegative scalar reflecting the number of mutations contributed by this signature. *N* is equal to 10 and it reflects the number of all possible signatures that can be found in a liver sample. Any mutational signature contributing less than 1% of the somatic mutations in a sample was removed and the sample was re-analyzed with the remaining signatures. Any signature that did not improve the cosine similarity between the original sample and the sample reconstructed using the consensus mutational signatures and their respective exposures with more than 0.02 was removed and the sample was reanalyzed with the remaining signatures. The analysis revealed that signatures 1A, 1B, 4, 5, 6, 16, 23, and 24 are present in these liver cancer exomes but not signatures 12 and 17. The re-introduction allowed better evaluating the presence of mutational signatures in each sample by leveraging the set of consensus mutational signatures previously deciphered from a larger dataset of 88 whole-genomes.

All samples were clustered, based on the number of somatic mutations contributed by each signature in each sample, using unsupervised hierarchical clustering with cosine distance and Ward linkage.

Copy-number analysis

To identify copy-number alterations, we calculated the log ratio of the coverage in each tumor and its matched non-tumor liver sample for each bait of the exome capture kit. Log ratio profiles were then smoothed using the circular binary segmentation algorithm as implemented in the Bioconductor package *DNAcopy*[38]. The most frequent smoothed value was considered to be the *zero* level of each sample. Segments with a smoothed log ratio above (*zero*+0.3) or below (*zero*−0.3) were considered to be gained and deleted, respectively. High-level amplifications and homozygous deletions thresholds were defined as the (mean+ 5 standard deviations) of smoothed log ratios in gained and deleted regions, respectively. Chromosome instability was quantified as the frequency of aberrant arms (FAA), i.e. the proportion of chromosome arms with an aberrant copy-number status on >60% of their length[11].

Identification of putative driver genes

We first used the MutSigCV[21] algorithm to identify genes harboring significantly more mutations than expected by chance. This approach takes into account the nucleotide context, gene-expression, replication time, the observed silent mutations and the presence of mutations in the surrounding regions. It estimates the background mutation rate for each gene–patient–category combination and tests the null hypothesis that all the observed mutations in each gene are a consequence of random background mutation. Genes for which

this hypothesis is rejected based on the Benjamini-Hochberg false discovery rate-corrected q-value are considered significantly mutated. Analysis was done using default settings together with a liver-oriented covariates table comprising HCC-derived gene expression levels (GSE62232).

We then combined mutations and focal CNAs to define an enlarged list of putative drivers. We considered all genes with 6 alterations (non-silent mutations, high-level amplifications or homozygous deletions) among the 235 HCC (excluding DMN samples and hypermutated HCC sample CHC892T), corresponding to a frequency of 2.5%. We then remove large genes (cds length 10.000 amino acid), genes not expressed in HCC considering our Affymetrix microarray GSE62232 dataset collected on a series of 81 HCCs (RMA normalized intensities 20 units or <20 units., with 20 standard deviation) and genes displaying ratio of silent versus non-silent mutations 0.15. 161 genes passed more than two of these three filters and were considered to be putative driver genes (Supplementary Table 8).

Identification of recurrently altered pathways and targetable genes

Gene annotations, including unique gene and transcript identifiers, description and functional reports, were retrieved from Ensembl release 75 given the variant genomic location using the Bioconductor package *biomaRt.* The definition of the cellular pathways found to be associated with HCC was established by annotating the 161 candidate driver genes using the Gene Ontology database (http://www.geneontology.org/), then stating their precise roles through an expert reviewing of the literature and gene annotation or pathways databases including GeneCards (http://www.genecards.org/), KEGG (http://www.genome.jp/kegg/pathway.html), and PubMed (http://www.ncbi.nlm.nih.gov/pubmed/). Using this approach, we identified 11 major pathways altered in more than 5% of HCC. Some putative driver genes or genes already reported in HCC were added to the general scheme (Fig. 3a), as well as genes with only minor or no alterations in our cohort but playing key functional roles in the identified pathways. Interactions between pathways and repartition of the genes in the cellular compartments were highlighted as reported in KEGG pathway database. The FDA-approved drugs or drugs screened in different phases of clinical trials that were found related to one of the genes or pathways were reviewed in ClinicalTrails.gov (http://www.clinicaltrials.gov) and NCI Drug Dictionary (http://www.cancer.gov/drugdictionary) databases.

Sanger sequencing

Eleven percent of the mutations identified by exome sequencing were confirmed by independent PCR and Sanger sequencing. All HCC and liver cancer cell lines were systematically screened for *CTNNB1*, *TP53*, *ARID1A*, *AXIN1*, *RPS6KA3*, *KEAP1*, *CDKN2A*, *NFE2L2*, *ARID2*, *PIK3CA* and *KRAS* mutations (Supplementary Table 15) as described in Guichard *et al.* and the promoter region of *TERT* was sequenced as described in Nault *et al.*[11,24]. In all cases the somatic origin of the mutation found in tumor was verified by sequencing the corresponding adjacent, normal liver sample[11,24].

Cell lines

The 29 liver cancer cell lines were obtained from commercial sources. Cells were grown in Dulbecco's modified Eagle's medium (DMEM) or William's E medium supplemented with 10% fetal bovine serum and 100 U/mL penicillin/streptomycin and maintained at 37 °C in a humidified incubator in 5% CO2. Cell line identity was confirmed by exploring known gene mutations in each cell line. All the cells were mycoplasma-free, as tested by a PCR assay (Sigma).

Drugs and cell viability assay

17-AAG and 17-DMAG, two benzoquinone ansamycin HSP90 inhibitors were purchased from Sigma-Aldrich and dissolved in DMSO. Cells were seeded into 96-well plates at a density of 1500-3000 cells per well. After overnight incubation, cells were treated for 48h with vehicle alone (0.1% DMSO) or with various concentrations of 17-AAG or 17-DMAG (0.001, 0.01, 0.1, 1 and 10 uM, in 0.1% DMSO) in 100 ul of culture medium supplemented with 10% fetal bovine serum and 100 U/ml penicillin/streptomycin. Each concentration was tested in duplicate; experiments were repeated two to three times for each cell. Cell viability was measured by MTS assays (Promega) according to the manufacturer's recommendations. The concentration of drug inhibiting cells growth by 50% relative to the untreated control (GI50) was calculated after curve fitting with GraphPad Prism 5.0 software.

Cell lines transfection and western-blotting

Cells were transfected with 2 nM of siRNA using the lipofectamine RNAiMAX reagent (Invitrogen) in 6-well plates, according to the manufacturer's protocol. Three different siRNA duplexes targeting *RPS6KA3* (coding RSK2) (s12279, s12280 and s12280, Life Technologies) were tested. Block-iT Alexa Fluor Red Fluorescent Oligo siRNA (Life Technologies) was used as a double-stranded RNA negative control. The effect of the gene knockdown was verified on the protein level by western blotting. The total protein extracts were obtained by lysis in the RIPA buffer supplemented with protease and phosphatase inhibitors. Western-blot analyses were performed using primary antibodies specific for: RSK2 (sc-1430 Santa Cruz, dilution 1:2000), ERK1/2 (#9102 Cell Signaling Technology, dilution 1:500) and phospho-ERK1/2 (#9101 Thr202/Tyr204) (Cell Signaling Technology, dilution 1:300). Polyclonal rabbit anti-β-actin (A5060 Sigma, 1:3000) was used as loading control.

Quantitative RT-PCR

Quantification of *NQO1* mRNA level was performed by quantitative RT-PCR on the BioMark HD™ platform (Fluidigm) using a set of pre-designed primers and probe from Life technologies (Hs00168547_m1). Ribosomal 18S (R18S) was used for the normalization of expression data.

Reverse phase protein array

RPPA technology was used to quantify NQO1 protein level in cells as previously described[39]. Briefly, equal amounts of protein lysates were printed onto nitrocellulose covered slides. Four serial dilutions and two technical replicates per dilution were deposited

for each sample. Arrays were revealed with an anti-NQO1 antibody (HPA007308 from Sigma). Quantification and normalization of RPPA data was performed using the NormaCurve method[39].

NQO1 P187S SNP analysis

NQO1 P187S *(NQO1*2 allele, rs1800566)* SNP genotyping was performed on genomic DNA extracted from cells using a pre-designed TaqMan assay (C_2091255_30, Life Technologies), on an ABI *7900HT* instrument (Applied Biosystems), according to the manufacturer's instructions.

Statistical Analysis

R software v2.15.0 (http://www.R-project.org) and Bioconductor packages were used for statistical analysis and data visualization. Tests of independence were performed using Chi-square and Fisher's exact tests. *P* values were adjusted by Monte Carlo simulation according to Hope *et al.*[40]. The strength of association among gene mutation events was modeled using a binomial logistic regression. We used chi-square tests for trend in proportions to identify genes associated with HCC progression and the Jonckheere-Terpstra test to assess the increase of mutation and CNA numbers along tumor stages. Only genes mutated in 3% of cases were included.

Variables associated with overall survival at 60 months were identified using univariate and multivariate Cox proportional hazards regression models (Wald test), using the *survival* package. Only patients with curative (R0) resection were included in survival analysis (n=216, exclusion of non-curative resections and liver transplantations). Kaplan-Meier plots were used to describe survival rates among all cases.

All reported *P* values were two-tailed and differences were considered significant when the *P* value was under 0.05.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. Lancet. 2012; 379:1245–55. [PubMed: 22353262]

2. El-Serag HB. Hepatocellular carcinoma. N Engl J Med. 2011; 365:1118–27. [PubMed: 21992124]

3. International Consensus Group for Hepatocellular NeoplasiaThe International Consensus Group for Hepatocellular, N. Pathologic diagnosis of early hepatocellular carcinoma: a report of the international consensus group for hepatocellular neoplasia. Hepatology. 2009; 49:658–64. [PubMed: 19177576]

4. Roncalli M, et al. Liver precancerous lesions and hepatocellular carcinoma: the histology report. Dig Liver Dis. 2011; 43(Suppl 4):S361–72. [PubMed: 21459342]

5. Zucman-Rossi J, et al. Genotype-phenotype correlation in hepatocellular adenoma: new classification and relationship with HCC. Hepatology. 2006; 43:515–24. [PubMed: 16496320]

6. Nault JC, Bioulac-Sage P, Zucman-Rossi J. Hepatocellular benign tumors-from molecular classification to personalized clinical care. Gastroenterology. 2013; 144:888–902. [PubMed: 23485860]

7. Pilati C, et al. Genomic profiling of hepatocellular adenomas reveals recurrent FRK-activating mutations and the mechanisms of malignant transformation. Cancer Cell. 2014; 25:428–41. [PubMed: 24735922]

8. Bruix J, Gores GJ, Mazzaferro V. Hepatocellular carcinoma: clinical frontiers and perspectives. Gut. 2014; 63:844–55. [PubMed: 24531850]

9. Llovet JM, Hernandez-Gea V. Hepatocellular carcinoma: reasons for phase III failure and novel perspectives on trial design. Clin Cancer Res. 2014; 20:2072–9. [PubMed: 24589894]

10. The French METAVIR Cooperative Study Group. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. Hepatology. 1994; 20:15–20. [PubMed: 8020885]

11. Guichard C, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet. 2012; 44:694–8. [PubMed: 22561517]

12. Fujimoto A, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet. 2012; 44:760–4. [PubMed: 22634756]

13. Kan Z, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. Genome Res. 2013; 23:1422–33. [PubMed: 23788652]

14. Alexandrov LB, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–21. [PubMed: 23945592]

15. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet. 2014; 15:585–98. [PubMed: 24981601]

16. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013; 3:246–59. [PubMed: 23318258]

17. Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Curr Opin Genet Dev. 2014; 24:52–60. [PubMed: 24657537]

18. Hsu IC, et al. Mutational hotspot in the p53 gene in human hepatocellular carcinomas. Nature. 1991; 350:427–8. [PubMed: 1849234]

19. Bressac B, Kew M, Wands J, Ozturk M. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. Nature. 1991; 350:429–31. [PubMed: 1672732]

20. Totoki Y, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet. 2014; 46:1267–73. [PubMed: 25362482]

21. Lawrence MS, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214–8. [PubMed: 23770567]

22. Sawey ET, et al. Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. Cancer Cell. 2011; 19:347–58. [PubMed: 21397858]

23. Chiang DY, et al. Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. Cancer Res. 2008; 68:6779–88. [PubMed: 18701503]

24. Nault JC, et al. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. Nat Commun. 2013; 4:2218. [PubMed: 23887712]

25. Totoki Y, et al. Unique mutation portraits and frequent COL2A1 gene alteration in chondrosarcoma. Genome Res. 2014; 24:1411–20. [PubMed: 25024164]

26. Barretina J, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012; 483:603–7. [PubMed: 22460905]

27. Garnett MJ, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012; 483:570–5. [PubMed: 22460902]

28. Siegel D, et al. Rapid polyubiquitination and proteasomal degradation of a mutant form of NAD(P)H:quinone oxidoreductase 1. Mol Pharmacol. 2001; 59:263–8. [PubMed: 11160862]

29. Ahn SM, et al. Genomic portrait of resectable hepatocellular carcinomas: Implications of RB1 and FGF19 aberrations for patient stratification. Hepatology. 2014; 60:1972–82. [PubMed: 24798001]

30. Wang K, et al. Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. Hepatology. 2013; 58:706–17. [PubMed: 23505090]

31. Hyeon J, Ahn S, Lee JJ, Song DH, Park CK. Expression of fibroblast growth factor 19 is associated with recurrence and poor prognosis of hepatocellular carcinoma. Dig Dis Sci. 2013; 58:1916–22. [PubMed: 23456506]

32. Miura S, et al. Fibroblast growth factor 19 expression correlates with tumor progression and poorer prognosis of hepatocellular carcinoma. BMC Cancer. 2012; 12:56. [PubMed: 22309595]

References Methods

33. Song S, et al. qpure: A tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. PLoS One. 2012; 7:e45835. [PubMed: 23049875]

34. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988; 16:1215. [PubMed: 3344216]

35. Gnirke A, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009; 27:182–9. [PubMed: 19182786]

36. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–9. [PubMed: 20354512]

37. Waltz RA, Morales JL, Nocedal J, Orban D. An interior algorithm for nonlinear optimization that combines line search and trust region steps. Mathematical Programming. 2006; 107:391–408.

38. 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. [PubMed: 15475419]

39. Troncale S, et al. NormaCurve: a SuperCurve-based method that simultaneously quantifies and normalizes reverse phase protein array data. PLoS One. 2012; 7:e38686. [PubMed: 22761696]

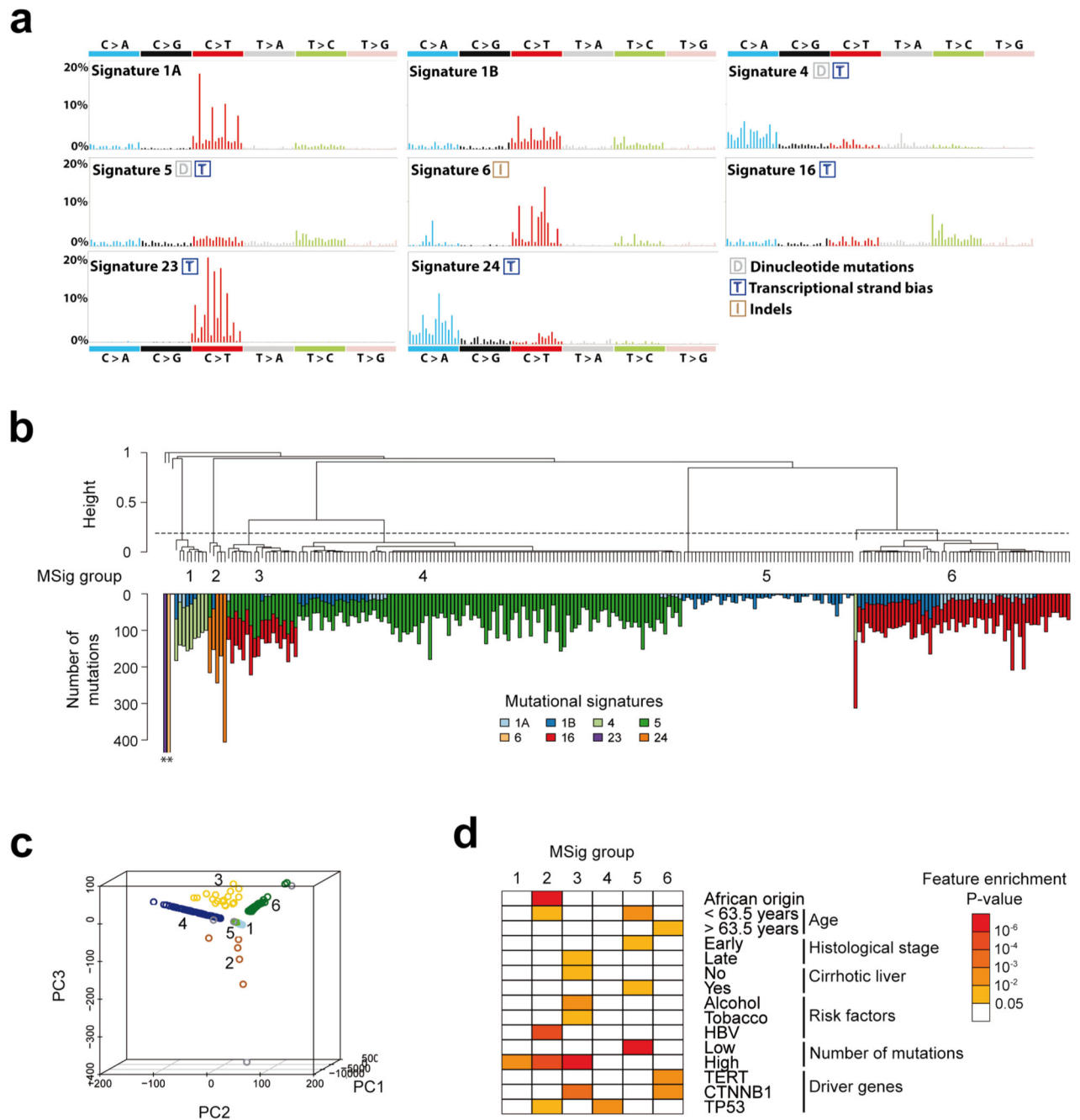40. Hope ACA. A simplified Monte Carlo significance test procedure. JSTOR. 1968; 30:582598.

Figure 1.

Consensus signatures of mutational processes in hepatocellular carcinoma. (**a**) Patterns of the signatures of the mutational processes operative in 243 liver exomes. Signatures 23 and 24 were identified using *de novo* WTSI mutational signatures framework, while the presence of Signatures 1A, 1B, 4, 5, 6, and 16 were identified via re-introduction of consensus mutational signatures previously identified in liver cancer by a pan-cancer analysis[14]. Each signature is displayed according to the 96 substitution classification defined by the substitution class and sequence context immediately 3′ and 5′ to the mutated base.

Signatures 1A, 1B, 4, 5, 6 and 16 match signatures previously described in a pan-cancer study[14] and the plotted patterns correspond to the updated consensus signatures. Signatures 23 and 24 are new.(**b**) Unsupervised hierarchical clustering of 243 liver tumors based on the intensity of signatures operative in each sample. Tumors were classified into 6 mutational signature (MSig) groups and 4 singletons. The number of mutations attributed to each signature in each tumor is represented by color bars below the dendogram. *Tumors harboring more than 500 mutations. (**c**) Principal component analysis of mutational signature intensities in 243 liver tumors. Tumor samples are plotted in three dimensions using their projections onto the first three principal components (PC). MSig group membership is represented by a color code and labels. (**d**) Clinical and molecular features associated with each MSig group. Associations were assessed using Chi-square tests for categorical variables and ANOVA for quantitative features.
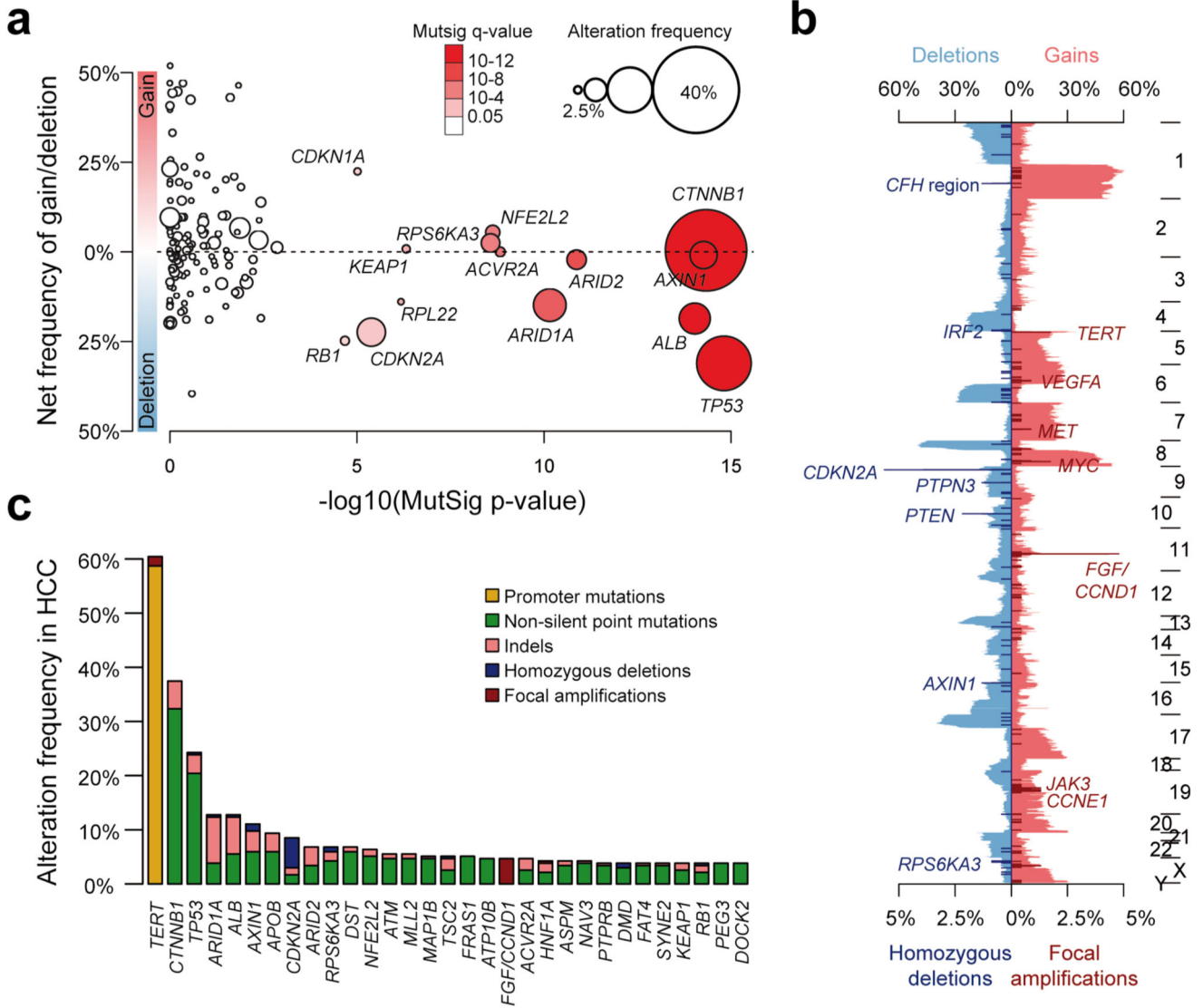
Figure 2.

Integration of mutations, focal amplifications and homozygous deletions identifies putative driver genes in hepatocellular carcinoma. (**a**) The 161 putative driver genes identified by integrating mutations and focal CNAs are represented, with the log-transformed mutation significance on the x-axis and the net frequency of gains and deletions on the y-axis. The size and color of each sample represent the alteration frequency and MutSig q-values, respectively. Significantly mutated genes ($q < 0.05$) are indicated. (**b**) Frequency of CNAs along the genome. The top axis indicates the frequency of low-amplitude changes (gains and losses); the bottom axis indicates the frequency of high-amplitude changes (focal amplifications and homozygous deletions). Target genes of recurrent amplifications and homozygous deletions are indicated. (**c**) Bar plot indicating the number and type of events for the most frequently altered genes (4%).
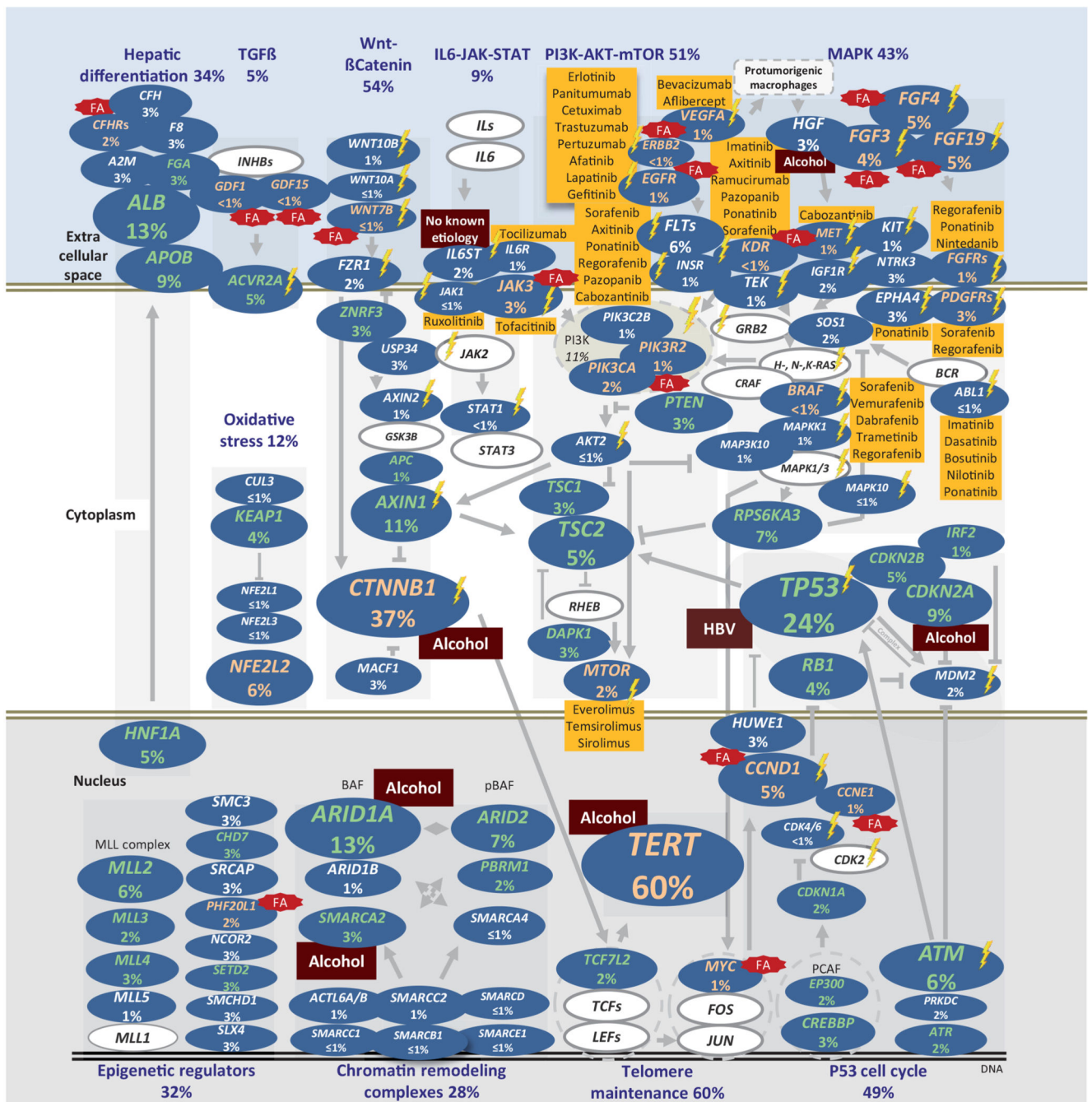
Figure 3.
The landscape of altered genes and pathways in HCC. Eleven pathways altered in 5% of HCC. Genes belonging to each pathway are represented with their alteration frequencies (green font: inactivation; red font: activation; white font: unknown) and activating or inhibiting interactions between pathways are highlighted. Major genes lacking alterations are highlighted in white. Focal amplifications (FA) and significant associations with etiologies are indicated. Potential druggable genes are assigned (yellow: FDA-approved drugs; lighting: drugs screened in phase I-III clinical trials).
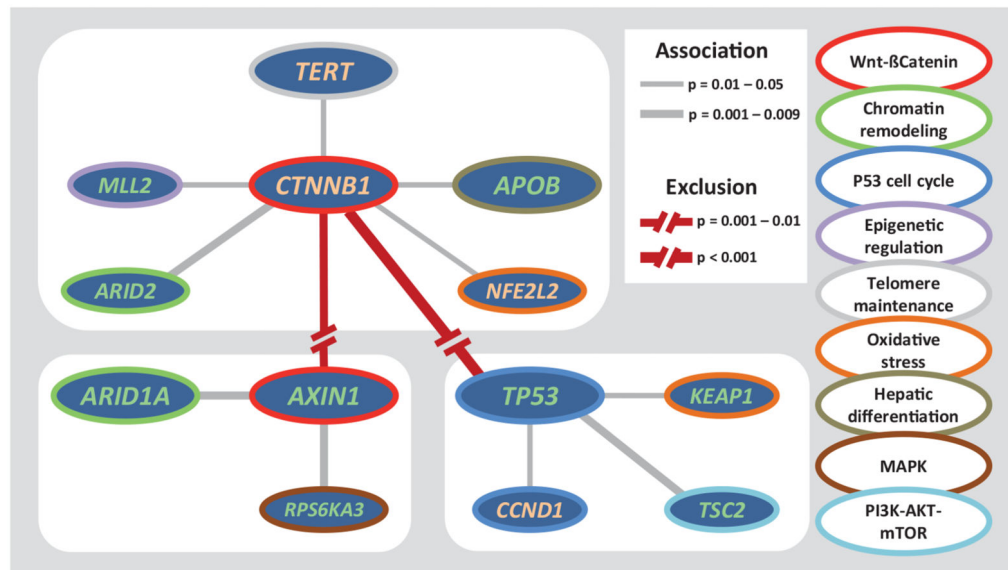
Figure 4.
Major clusters of associated alterations. Three major groups of associated alterations: *CTNNB1*-cluster, *AXIN1*-cluster, and *TP53*-cluster. Significant associations and exclusions between genes are represented, with line widths proportional to significance, as indicated in the legend. Surrounding colors of each gene indicate its participation to a particular pathway.
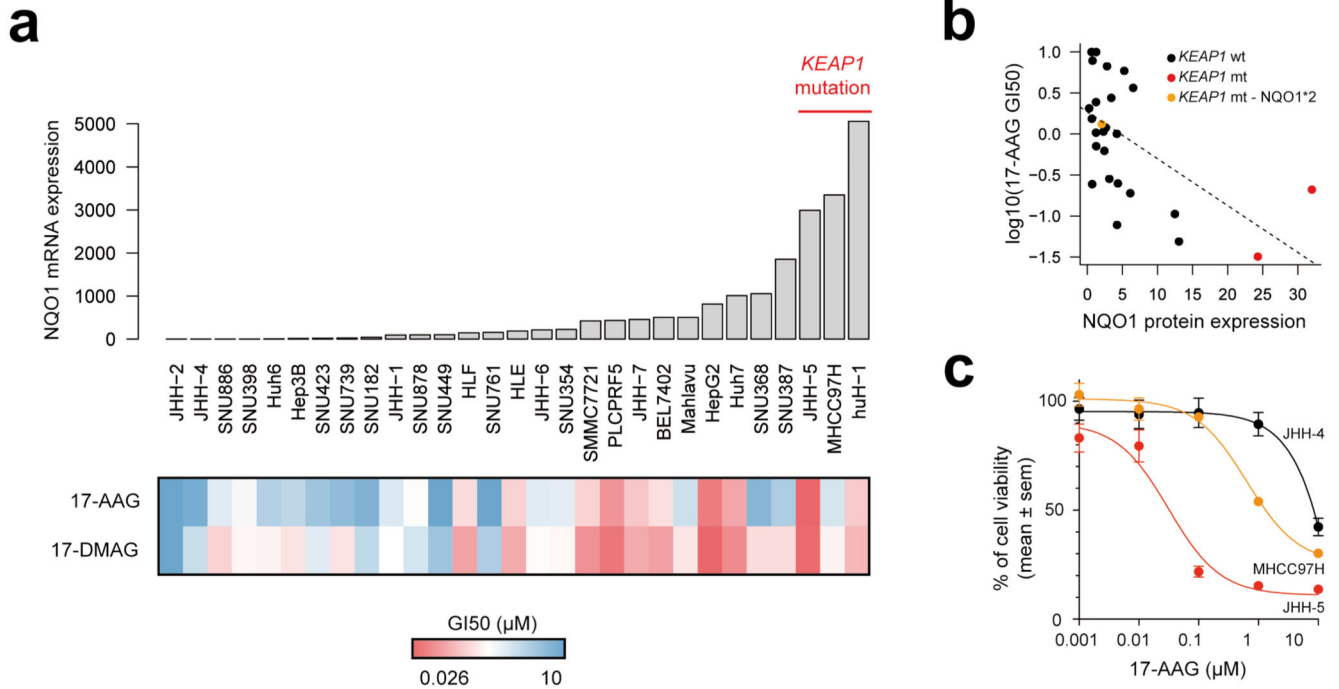
Figure 5.

Sensitivity of liver cancer cell lines to HSP90 inhibitors is associated with *NQO1* expression. (**a**) Sensitivity of 29 liver cancer cell lines to 17-AAG and 17-DMAG was assessed by a cell viability assay. The heatmap represents the GI50 of the two drugs in each cell line, and mRNA expression levels of *NQO1* are represented atop the heatmap. Three cell lines harboring *KEAP1* mutations are indicated. (**b**) Correlation between NQO1 protein expression and sensitivity to 17-AAG. A color code indicates cell lines with *KEAP1* mutation, and the MHCC97H cell line that harbors a *KEAP1* mutation but is homozygous for the *NQO1*2* variant causing NQO1 enzymatic deficiency. (**c**) Cell viability curves for cell lines JHH-4 (*KEAP1*-wild-type), JHH-5 (*KEAP1*-mutated) and MHCC97H *(KEAP1*-mutated and *NQO1*2* homozygous) in the presence of the HSP90 inhibitor 17-AAG. Results are shown as mean + sem (standard error of the mean). Each experiment was repeated two to three times for each cell line.
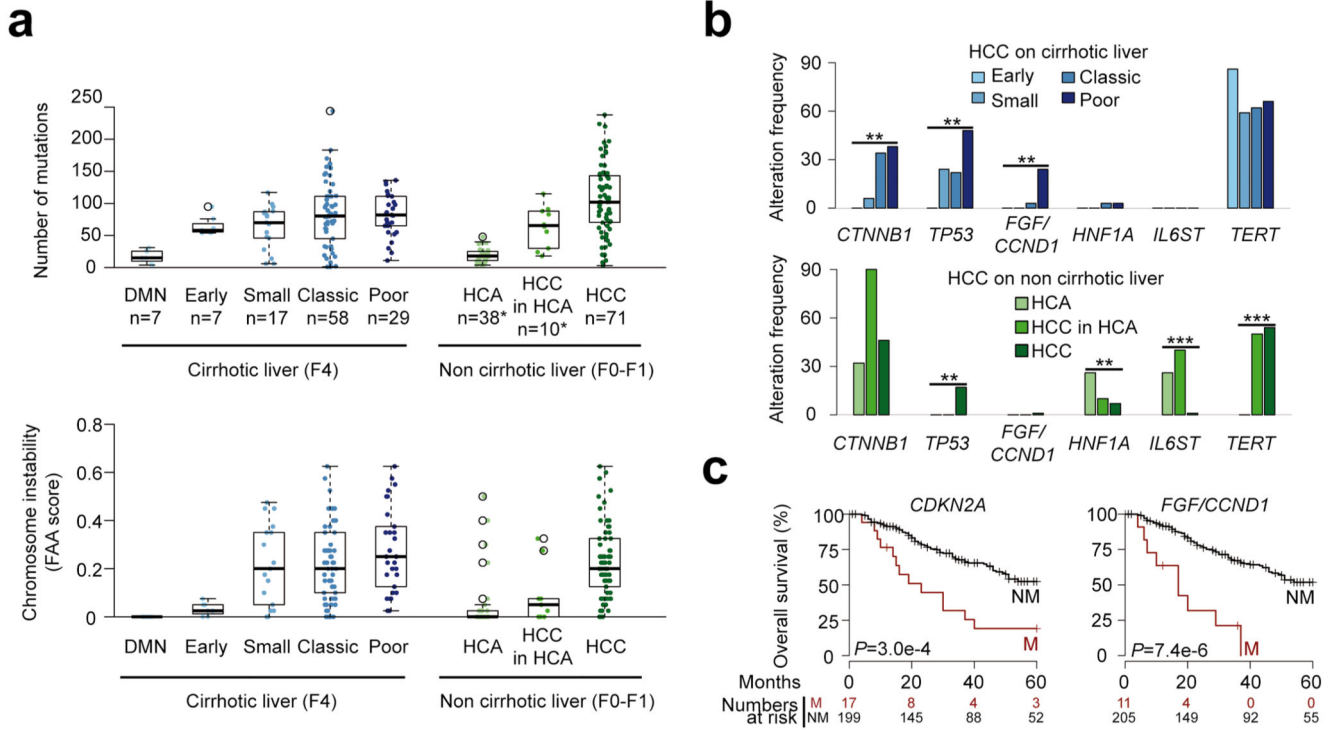
Figure 6.
Molecular features of HCC progression in cirrhotic and non-cirrhotic liver. (**a**) Box-and-whisker plots show the distribution of mutation numbers (top) and chromosome instability scores (bottom) relative to tumor stage in cirrhotic (left) and non-cirrhotic liver (right). FAA, fraction of aberrant arms; DMN, dysplastic macronodule; HCA, hepatocellular adenoma; HCC, hepatocellular carcinoma. Middle bar, median; box, interquartile range; bars extend to 1.5 times the interquartile range. *Most HCA and HCC in HCA samples are from a previous study[7]. (**b**) Genetic alterations with significantly different frequencies across tumor stages in cirrhotic (top) and non-cirrhotic liver (bottom). **, $P < 0.01$, ***, $P < 0.001$, Chi-square test for trends in proportion. (**c**) Kaplan-Meier curves for overall survival in the presence or absence of *CDKN2A* alteration (left) or *FGF3, FGF4, FGF19/CCND1* amplification (right), two features found significantly associated with survival in a multivariate Cox model. *P*-values were obtained using Wald tests.