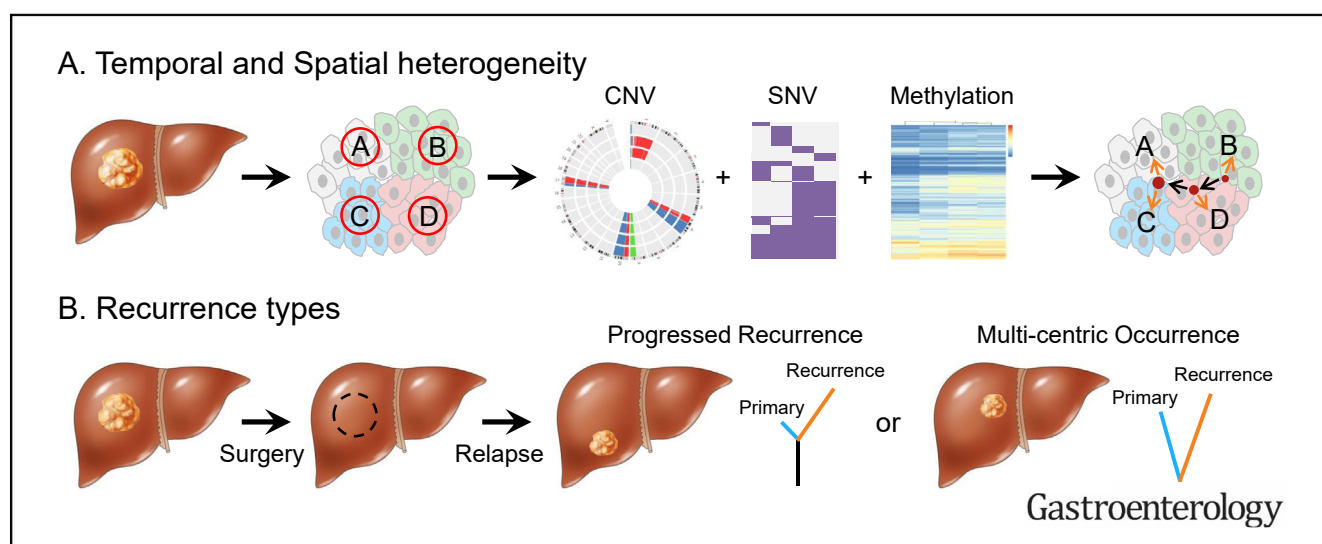


Genomic and Epigenomic Features of Primary and Recurrent Hepatocellular Carcinomas



Xiaofan Ding,^{1,*} Mian He,^{1,*} Anthony W. H. Chan,¹ Qi Xiu Song,¹ Siu Ching Sze,¹ Hui Chen,¹ Matthew K. H. Man,¹ Kwan Man,² Stephen L. Chan,³ Paul B. S. Lai,⁴ Xin Wang,⁵ and Nathalie Wong^{1,4,6,7}

¹Department of Anatomical and Cellular Pathology at Sir Y.K. Pao Centre for Cancer, The Chinese University of Hong Kong, Shatin, Hong Kong, China; ²Department of Surgery, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong, China; ³Department of Clinical Oncology, The Chinese University of Hong Kong, Shatin, Hong Kong, China; ⁴Department of Surgery, The Chinese University of Hong Kong, Shatin, Hong Kong, China; ⁵Department of Biomedical Sciences, City University of Hong Kong, Kowloon Tong, Hong Kong, China; ⁶State Key Laboratory of Translational Oncology, The Chinese University of Hong Kong, Shatin, Hong Kong, China; and ⁷State Key Laboratory of Digestive Disease, The Chinese University of Hong Kong, Shatin, Hong Kong, China



See Covering the Cover synopsis on page 1447; see editorial on page 1477.

BACKGROUND & AIMS: Intratumor heterogeneity and divergent clonal lineages within and among primary and recurrent hepatocellular carcinomas (HCCs) produce challenges to patient management. We investigated genetic and epigenetic variations within liver tumors, among hepatic lesions, and between primary and relapsing tumors. **METHODS:** Tumor and matched nontumor liver specimens were collected from 113 patients who underwent partial hepatectomy for primary or recurrent HCC at 2 hospitals in Hong Kong. We performed whole-genome, whole-exome, or targeted capture sequencing analyses of 356 HCC specimens collected from multiple tumor regions and matched initial and recurrent tumors. We performed parallel DNA methylation profiling analyses of 95 specimens. Genomes and epigenomes of nontumor tissues that contained areas of cirrhosis or fibrosis were analyzed. We developed liver cancer cell lines that endogenously expressed a mutant form of TP53 (R249S) or overexpressed mutant forms

of STAT3 (D170Y, K348E, and Y640F) or JAK1 (S703I and L910P) and tested the abilities of pharmacologic agents to reduce activity. Cells were analyzed by immunoblotting and chromatin immunoprecipitation with quantitative polymerase chain reaction. **RESULTS:** We determined the monoclonal origins of individual tumors using a single sample collection approach that captured more than 90% of mutations that are detected in all regions of tumors. Phylogenetic and phyloepigenetic analyses revealed interactions and codependence between the genomic and epigenomic features of HCCs. Methylation analysis revealed a field effect in cirrhotic liver tissues that predisposes them to tumor development. Comparisons of genetic features revealed that 52% of recurrent HCCs derive from the clonal lineage of the initial tumor. The clonal origin of recurrent HCCs allowed construction of a temporal map of genetic alterations that associated with tumor recurrence. Activation of JAK signaling to STAT was a characteristic of HCC progression via mutations that associate with response to drug sensitivity. The combination of a mutation that increases the function of TP53 and the 17p chromosome deletion might provide liver cancer cells with a replicative advantage. Chromatin immunoprecipitation analysis of TP53

with the R249S substitution revealed its interaction with genes that encode chromatin regulators (*MLL1* and *MLL2*). We validated *MLL1* and *MLL2* as direct targets of TP53^{R249S} and affirmed their association in the Cancer Genome Atlas dataset. The MLL-complex antagonists MI-2-2 (inhibitor of protein interaction) and OICR-9492 (inhibitor of activity) specifically inhibited proliferation of HCC cells that express TP53^{R249S} at nanomolar concentrations. **CONCLUSIONS:** We performed a systematic evaluation of intra- and intertumor genetic heterogeneity in HCC samples and identified genetic and epigenetic changes that associate with tumor progression and recurrence. We identified chromatin regulators that are upregulated by mutant TP53 in HCC cells and inhibitors that reduce proliferation of these cells. DNA methylation patterns in cirrhotic or fibrotic liver tissues might be used to identify those at risk of HCC development.

Keywords: Hepatic Carcinogenesis; Tumorigenesis; Tumor Progression; Treatment.

Hepatocellular carcinoma (HCC) is a major cause of cancer-related deaths worldwide.¹ Surgical resection remains the mainstay of treatment for early localized HCC, but hepatic recurrence occurs in >70% of patients within 5-years postsurgery that reduces survival.² Early recurrent tumors, typically within 2-years postsurgery, are believed to share the same clonal origin as initial HCC, whereas late recurrences (>2 years) are considered to represent de novo development of second primary tumors bearing a different clonal lineage.^{3,4} Although these hypotheses have yet to be fully validated, because the clinical management of progressive or secondary tumor differs substantially, it is important to build a practical approach to discriminate genetic clonality between initial and recurrent HCCs to distinguish between these 2 possibilities.

For advanced stage patients, sorafenib is the first-line therapy but offers minimal survival benefit.⁵ It is believed that intratumor heterogeneity has limited the development of targeted therapies in HCC and determination of the spatial distribution of topographically distinct areas of genetic alterations remains a challenge. In addition, the frequent presence of multinodular HCC at diagnosis poses the question of whether these lesions represent synchronous multifocal tumors of different clonality, or intrahepatic metastases with a shared clonal relationship, and has further complicated development of systemic targeted therapies. The characteristics of tumor heterogeneity and the extent of genomic diversity in HCC tumors are therefore of potential importance for the clinical management of patients, underscoring the need to better understand the genomic architecture within a bulk tumor, clonal expansion and subclonal diversification in tumor dissemination, and temporal dynamics during recurrence.

Like many cancer types, HCC evolves through a sequential order of clonal expansion and selection of genomic and epigenomic alterations.⁶ However, unlike most cancers, HCC usually arises from an organ that has been damaged by chronic inflammation and persistent hepatic

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

We analyzed hepatocellular carcinomas (HCCs) and surrounding non-tumor tissues to identify genetic and epigenetic variations within liver tumors, among hepatic lesions, and between primary and relapsing tumors.

NEW FINDINGS

We identified DNA methylation patterns associated with risk of HCC recurrence after surgery. We identified chromatin regulators that are upregulated by mutant TP53 in HCC cells and inhibitors that reduce proliferation of these cells.

LIMITATIONS

This was a retrospective analysis of liver tissues obtained from patients undergoing surgery at 2 hospitals. Analyses of larger numbers of samples and prospective studies are needed.

IMPACT


Genetic and DNA methylation patterns of HCCs can be used to identify patients at high risk for recurrence after surgery. DNA methylation patterns in cirrhotic or fibrotic liver tissues might be used to identify those at risk of HCC development and recurrence after surgery.

injury from etiologies such as hepatitis viruses, alcohol, and fatty liver.^{7,8} More than 90% of HCCs develop on a background of liver cirrhosis or fibrosis,⁷ which has long been regarded to prime cancer development, but there is a lack of concrete evidence to support their premalignant state. It is plausible that initial “gatekeeper” alterations in premalignant stages provide a selective growth advantage to normal epithelial cells that are insufficient for tumor onset.⁹ Later deleterious “driver” events unleash malignant transformation and allow clonal expansion of tumor cell growth.⁹ We hypothesize that “gatekeeper” and “driver” events in HCC, whether genetic or epigenetic, are trunk alterations that are clonally dominant, given that tumors arise from a single cell and trunk alterations are initiating pro-oncogenic events in cancer evolution.

To characterize spatial and temporal tumor heterogeneity, and determine the truncal “gatekeepers” and “drivers” in the clonal evolution leading to HCC, we investigated heterogeneity at 3 levels: (1) analysis of intratumor heterogeneity by whole-exome sequencing (WES) and DNA methylation to quantify the variability among different sectors from multiregion sampling within

* Authors share co-first authorship.

Abbreviations used in this paper: CGI, CpG island; CNA, copy number alteration; GOF, gain-of-function; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HGV, high genetic variation; NLs, normal livers; WES, whole-exome sequencing; WGS, whole-genome sequencing.

 Most current article

© 2019 by the AGA Institute. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

0016-5085

<https://doi.org/10.1053/j.gastro.2019.09.005>

the same tumor; (2) analysis of intertumor heterogeneity among hepatic lesions to accurately distinguish the genomic and epigenomic lineages of HCC tumors; and (3) analysis of the heterogeneity between initial and relapse tumors. We also scored the tumor-neighboring liver cirrhosis and fibrosis nodules for genomic and epigenomic changes, and identified HCC-specific “gatekeeper” events. In addition, focusing on clinically and therapeutically relevant aspects of tumor heterogeneity, we explored gain-of-function (GOF) *TP53* mutation and *STAT3* for possible therapeutic interventions.

Methods

Patient Samples

Tumors and matching adjacent nontumoral liver samples were collected from 113 patients who underwent partial hepatectomy for primary or recurrent HCC at Prince of Wales Hospital and Queen Mary Hospital, Hong Kong. Six tumors were subjected to extensive multiple region sampling as demonstrated in [Supplementary Figure 1](#). Informed consent was obtained from the recruited patients and the study protocol was approved by the Clinical Research Ethics Committee of the Chinese University of Hong Kong and Human Research Ethics Committee of the University of Hong Kong. Clinical information of the patient cohort is provided in [Supplementary Table 1](#). Tumor samples were either snap frozen or formalin fixed and paraffin embedded. The diagnosis of HCC was confirmed in all cases by 2 experienced pathologists through histology examination.

Whole-Genome Sequencing, WES, Targeted Capture Sequencing, and DNA Methylation Profiling

Thirty-six samples from 9 patients were analyzed by whole-genome sequencing (WGS); 86 samples from 6 patients by WES, and 234 samples from 98 patients by targeted next generation sequencing that covered 824 HCC-associated vital genes ([Supplementary Tables 1–4](#)). For WGS, Illumina sequencing libraries were prepared from genomic DNA passing quality control, and paired-end 100 base pair sequencing was performed on HiSeq 2500. Raw genome coverage of 100x for tumor samples at first surgery and 60x at relapse within the initial-relapse dataset were obtained, giving a high potential for the detection of events that could persist in relapse tumors. For samples studied for intratumor heterogeneity, Agilent (Santa Clara, CA) SureSelect Human All Exons v6 kit was used for whole-exome capture of genomic DNA and the captured DNA was sequenced by HiSeq 2500. One case ITH-15, both the T1 and T2 tumors were subjected to WGS. Genomic DNA from 95 samples was analyzed on Infinium HumanMethylationEPIC BeadChip (Illumina, San Diego, CA) arrays to obtain genome-wide DNA methylation profiles. We also included one reported case, HCC-15,¹⁰ which is the sole published case with extensive multiregion sampling, for intratumor heterogeneity analysis. To explore an association between public and nonpublic mutations with tumor size, we also included mutations called in 9 patients with HCC from the study by Zhai et al.¹¹

Accession Code

Sequencing and methylation data are accessible in the EBI database, no. EGAS00001002094.

Results

Spatial Genomic Heterogeneity

In 96 multiregion samples from 6 HCC tumors, we performed WGS on 12 samples from a case of bifocal HCC (ITH-15) and the remaining by WES ([Supplementary Figure 1A–G](#) and [Supplementary Tables 1](#) and [2](#)). According to their spatial organization, somatic aberrations including mutations, indels, and copy number alterations (CNAs) in different sectors were classified into 3 groups: public (ubiquitous in all regions), regional (found in 2 or more regions), and private (found in only 1 region). The phylogenetic trees varied between cases but marked public mutations were prominent in all tumor sectors (average 59.9%) highlighting the monoclonal origin of individual HCCs ([Figure 1Ai](#)). On average, we found 110 nonsilent public mutations per tumor sector ([Supplementary Table 3A](#)). Cancer-related drivers, including *TP53*, *CTNNB1*, and *TERT*, shared many public mutations ([Figure 1Aii](#)). Their presence in a clonal manner reinforces their early carcinogenic role. Notably, the number of public mutations did not correlate with abundance of nonpublic mutations, nor with tumor size ([Supplementary Figure 1H](#)). We next analyzed CNA profiles for patterns of intratumor heterogeneity ([Figure 1B](#) and [C](#), and [Supplementary Figure 1I–L](#)). Besides public CNAs in all tumor sectors, we also found a progressive increase in CNA level from one tumor sector to another (eg, increasing allelic magnitude of gain[7] in ITH-15T1). Also, copy neutral loss of heterozygosity was more common than expected and could be pervasive in all sectors (eg, copy neutral loss of heterozygosity [3p] in ITH-15T2) ([Figure 1C](#)). The only driver CNA found to be highly ubiquitous was del(17p), which encompasses *TP53*.

To address the extent of genetic variation, we generated 2 indices to estimate distribution of genetic variability within a tumor: the number of unique private mutations and Nei's score.^{12,13} Both methods concurred in highlighting high genetic variation (HGV) of substantial divergence between regions in ITH-625T1, HCC-15, and ITH-15T1, whereas the rest showed low genetic variation, especially ITH-15T2 was uniformly conferred by common drivers ([Figure 1D](#) and [E](#) and [Supplementary Figure 1M](#)). To explore factors associated with HGV, we formulated a spatial model to simulate 3-dimensional tumor growth.¹⁴ Based on observed data, we simulated multiregion sequencing data on virtual tumors and used Approximate Bayesian Computation to infer tumor characteristics, including mutation rate (μ) and selection (s). Interestingly, instead of being driven by selection, genetic variation strongly correlated with inferred mutation rates ([Figure 1F](#) and [Supplementary Figure 2A–D](#)). Meanwhile, these results were supported by 2 intrinsic public mutations in the DNA mismatch repair pathway in HGV-HCCs, ITH-625 (*ATRX*), and HCC-15 (*MSH2*), which likely elevated the acquired mutations along with tumor growth. The development of new

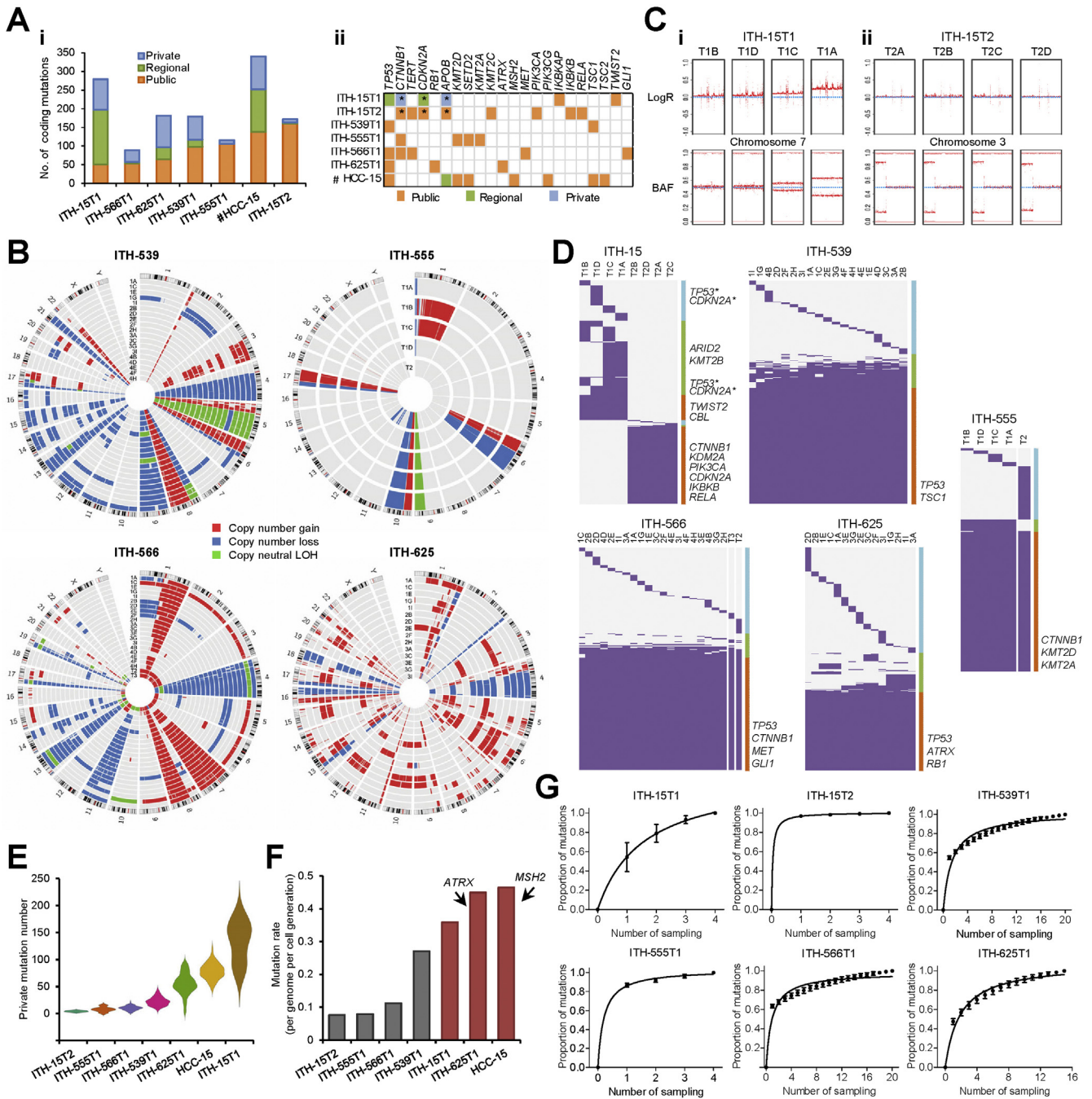
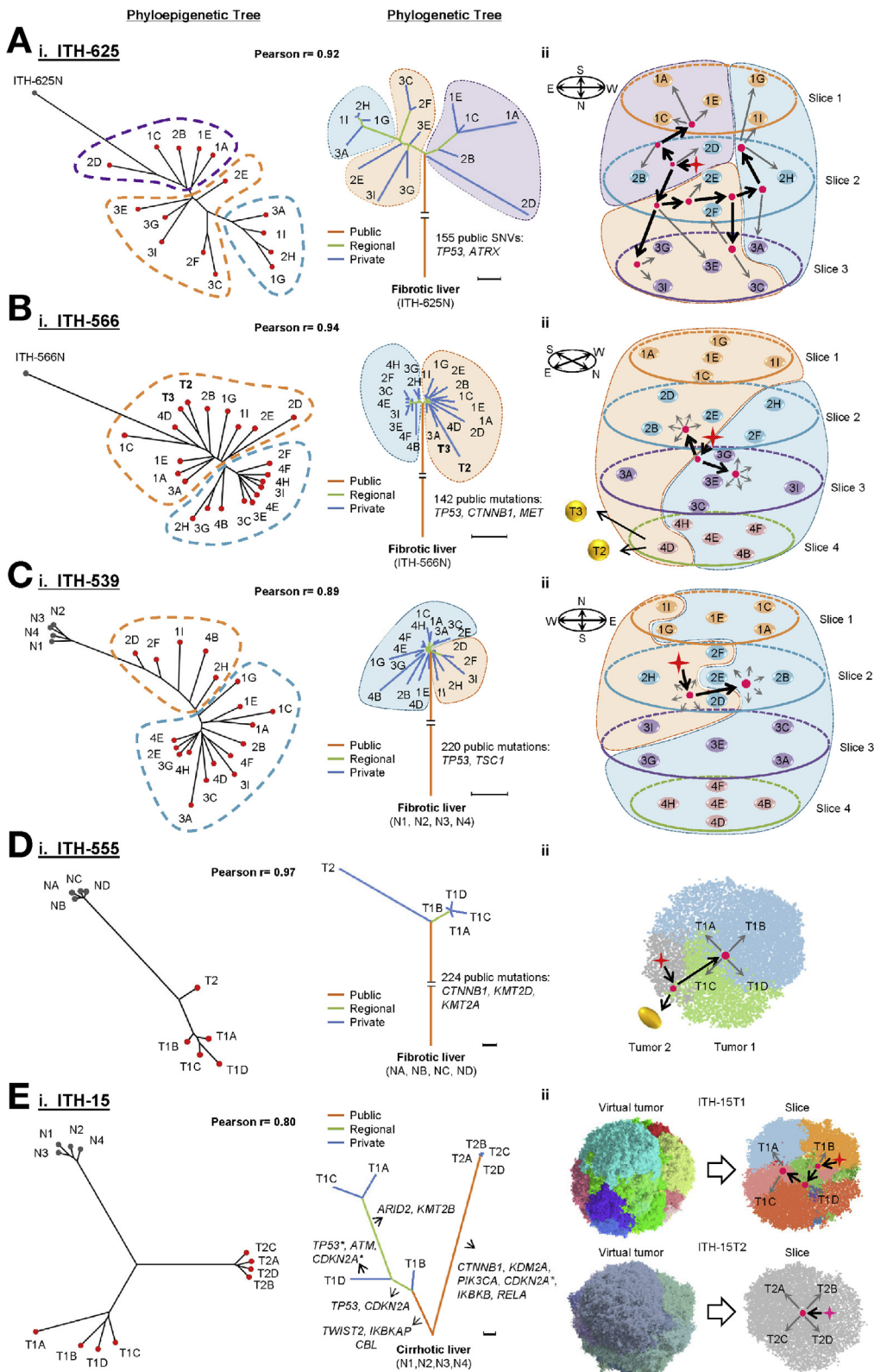


Figure 1. Patterns of intratumor heterogeneity in HCC. (A) Distribution of somatic events. (i) Left panel displays public (orange), regional (green), and private (blue) mutations in each tumor. (ii) Right panel shows driver genes identified in each tumor. Symbol (*) represents differing mutations in the same gene identified in the same patient ITH15. Symbol (#) HCC-15 represents a published case (Supplementary Methods). (B) Circos plots showing CNA profiles from ITH-539, -555, -566, and -625. Each track represents CNA determined from 1 tumor sample. (C) Representative CNAs from ITH-15T1 and ITH-15T2. Progressive increment of BAF magnitude from sector T1B>T1D>T1C>T1A suggested an increasing level of gain chr.7 in clonal trajectory between tumor sectors. Pervasive copy neutral loss of heterozygosity of chr.3p in T2A, T2B, T2C, and T2D was suggested from the simultaneous balanced logR ratios and imbalanced BAF values. (D) Mutational profiles of ITH-15, -539, -555, -566, and -625. Grid colors indicate the presence (blue) or absence (gray) of aberrations in each sample. Mutations were grouped into 3 classes: public (orange bar), regional (green bar), and private (blue bar). Columns of each mutation profiles were ranked in orders obtained from hierarchical clustering. (E) Violin plot illustrating the distribution of ITH levels, with y-axis representing the distribution of private mutation number after pairwise comparison across all regions in each tumor. (F) Inferred mutation rates based on Approximate Bayesian Computation (ABC) and spatial simulation. The y-axis stands for inferred mutation rates (mutations per cell cycle per site) within coding regions. (G) Mutation saturation curves. Relationship between tumor sampling number and proportion of mutations detected, with the x-axis representing sampling number and y-axis representing number of mutations.

simulation models has allowed identification of pertinent factors associated with HCC.

To evaluate links between sampling number and the number of mutations uncovered, we calculated the number

of identifiable nonsilent mutations with stepwise increase in profiled sampling count (Figure 1G and Supplementary Figure 2E). The saturation curves of all cases showed variances in shape but, on average, a single sample could



achieve a 0.67 detection of all somatic mutations. For the single-sampling approach, our results suggested that the distribution of clonal fraction was always higher for public mutations than nonpublic mutations (Supplementary Figure 2F). To increase the relative proportion of public mutations in the single-sample study, we grouped mutations into clonal (≥ 0.9) and subclonal (< 0.9) fractions and found average 90.65% clonal mutations to be public (Supplementary Figure 2G).

Genome and Epigenome Codependency in Evolutionary Trajectory

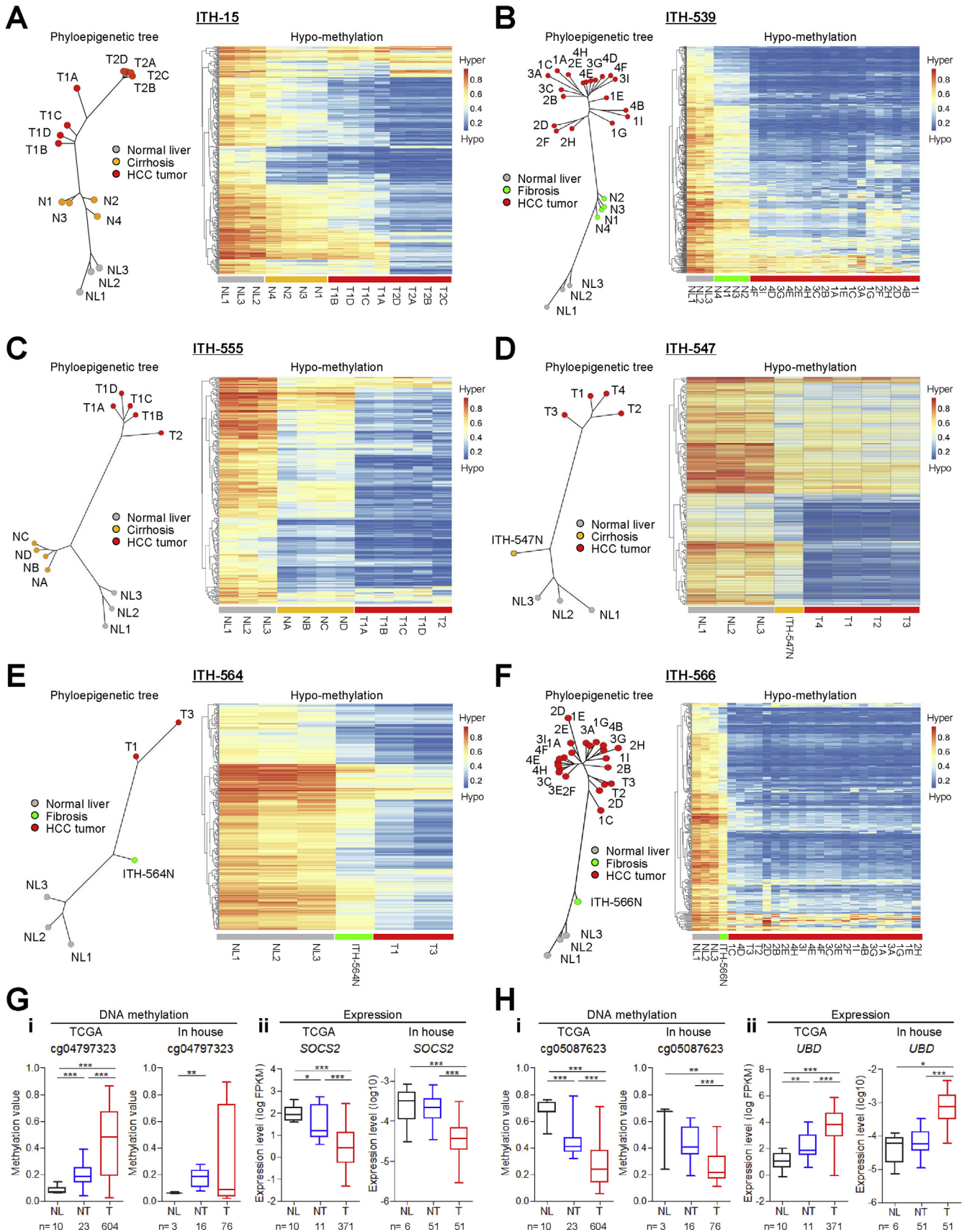
To decipher intratumor heterogeneity at the epigenome level and the potential relationship with genomic alterations, we profiled the DNA methylome of all tumor sectors that corresponded to samples that were sequenced by WGS and WES. Together with 3 normal livers (NLs), these multiregional tumor sectors were array profiled using Illumina Infinium MethylationEPIC (Supplementary Table 1A). We identified CpG probes that showed significant methylation differences from NLs, and accordingly classified them into public and nonpublic changes. Differential CpG probes suggestive of hyper- or hypomethylated loci were assigned to their regulatory categories, including CpG islands (CGIs), CGI shores, promoters, enhancers and gene body (Supplementary Figure 3A). On average, both public and nonpublic CpG probes showed enrichment of hypermethylated probes in CGI and promoter regions. This would imply that nonpublic CpG probes could also have a role in subclonal diversification. Gene ontology analysis supported our view, where nonpublic hypermethylated genes showed significant enrichment in cancer-related processes, including regulation of cell proliferation, adhesion, and transcription (Supplementary Figure 3B). Our observations underlined the likely involvement of methylation in the intratumor heterogeneity of HCC evolutionary biology.

We next sought to determine the spatial distribution of genomic and epigenomic alterations in multisector regions and map their trajectory path in tumor expansion. The extent of trunk and branch variables from WGS and WES allowed determination of the clonal relationship between individual lesions (Figure 2A–E). For DNA methylation, we first defined those epigenetic driver events (Epidrivers) that were

pertinent to HCC using 2 independent HCC methylation studies as training cohorts, The Cancer Genome Atlas¹⁵ ($n = 380$) and the study by Villanueva et al¹⁶ ($n = 243$) (Supplementary Figure 3C and D). Similar to genetic aberrations, both hyper- and hypo methylation Epidrivers exhibited extensive intratumoral heterogeneity, with an average 40.95% and 29.09% changes defined as nonpublic, respectively. As expected, the levels of intratumor heterogeneity reflected through methylation changes were also significantly associated with those of genetic changes, including somatic mutations and CNAs (Supplementary Figure 3E and F).

We next constructed a phyloepigenetic tree for each patient based on the level of variability across CpG sites. Meanwhile, we also inferred phylogenetic sample trees and clone trees based on somatic mutations of each case (Figure 2 and Supplementary Figures 4–6). Interestingly, both phylogenetic and phyloepigenetic trees displayed strong spatial overlap in all cases (Supplementary Figure 5, Pearson's r ranging from 0.80 to 0.92, permutation test $P < .01$). We next attempted to reconstruct the trajectory path of tumor development. The phylogenetic tree of all cases revealed a progressive pattern of mutational accumulation that could be mapped according to our spatial sampling locations (Figure 2), depicting a picture of evolutionary history in both spatial and temporal dimensions. In most cases, the initiating clone originated from the spatial center, except for ITH-555 and ITH-15T1 from the edge, but all showed continual evolution into distant branches containing divergent subclones. The phyloepigenetic trees also displayed an evolutionary pattern highly similar to phylogenetic trees in relative distance among samples (Pearson's r ranging from 0.80 to 0.92). Contrary to the spatial variation of the Big Bang model in colorectal cancer growth,¹⁷ our results suggested that HCC followed an evolutionary path in which tumor grew as a spatially continuous single clone expansion with sequential accumulation of mutations. Taking ITH-625 as an example, the phylogenetic tree suggested highly localized confinement of subclones and most regions in fact carried subclones from a single phylogenetic branch. The combined phylogenetic and phyloepigenetic analyses further suggested that the initiator cells for ITH-625 were from the spatial center, and the ancestral cells further divided into 2 spatially separated clusters of accumulated genetic and epigenetic aberrations.

Figure 2. Evolutionary history of tumor growth based on genome and epigenome profiles. (A–E) (i) Left panel shows the phyloepigenetic trees constructed on hypermethylated CpG sites, and middle panel shows the phylogenetic trees constructed based on nonsilent mutations. The branch length correlated with identified mutation number. (ii) Right panel presents the inferred growth model of each tumor. *Red star* in the center represents initiating cells. *Bold arrows* show the major evolutionary trajectories and *gray arrows* link the inferred branch and sampling locations. ITH-625T1 initiated from the spatial center with public mutations in *TP53*, *RB1*, and *ATRX* that propagated into 2 directions: one close to 2B and the other between slice 2 and slice 3. The 2B branched into subclones on slice 1. The branch on slices 2 and 3 further spread into one cluster of 3E, 3G, and 3I, and another cluster of 2E, 2F, and 3C that finally evolved into the most distant branch containing divergent subclones: 1G, 1I, 2H, and 3A. In contrast, both ITH-566T1 and ITH-539T1 were spatially separated into 2 big branches with each cluster including a set of tumor sectors with few genetic and epigenetic heterogeneity. ITH-566T2 and T3 originated from a common ancestor that already existed in the region of 4D. Both genetic and epigenetic data supported the divergent time for ITH-555T1 and T2 was quite early before the clonal expansion of ITH-555T1. ITH-15T1 and T2, however, were seeded from different ancestral tumor cells and displayed distinct ITH patterns, with high heterogeneity found in ITH-15T1 and high genetic homogeneity in ITH-15T2. Nonetheless, phyloepigenetic tree of ITH-15 showed T1 and T2 shared a truncal branch that arose from a common set of CpG probes.



Clonal Status of Multifocal HCC and Field Effect

Multiple hepatic lesions at presentation are common in patients with HCC. To address the clonal relationship among multifocal tumors, we assessed the intertumor heterogeneity in 5 patients with synchronous tumors; of these, 3 patients had also been analyzed for intratumor heterogeneity. In this way, we were able to map the route of tumor dissemination as well as clonal assessment. Both phylogenetic and phyloepigenetic trees concurred in highlighting common clonal ancestry among intrahepatic tumors ITH-566, -555, -547, and -564 (Figure 2 and Supplementary Figure 5). The genetic similarity of ITH-566T2 and T3 satellites with T1 could be directly traced to one of T1's specific sectors, 4D, where the route of spread likely initiated (Figure 2B). Similarly, ITH-555T2 likely diverged from sector T1C and before growth expansion of T1 (Figure 2D). In addition to the monoclonal origin of multifocal tumors arising from a single ancestor, we also observed tumors of independent origins in patient ITH-15. Despite ITH-15T1 being localized in close proximity to ITH-15T2 (3 cm apart), WGS showed their clonal origins to differ distinctively from each other. There was complete absence of shared somatic variants and CNAs (Figure 1D and Supplementary Figure 1J). Their pattern of hepatitis B virus (HBV) integration also showed no common viral insertion sites (Supplementary Figure 5D and Supplementary Table 3B). Together, WGS fully illustrated and strongly suggested that ITH-15T1 and T2 originated from 2 different clones. Surprisingly, contrary to the genomic findings, the phyloepigenetic tree of ITH-15 showed branches of T1 and T2 arising from a common set of tumor-specific probes that arose from their neighboring cirrhotic liver (Figure 2E). The aberrant hyper- and hypomethylation patterns maintained from cirrhosis to both T1 and T2 tumors suggested that epigenetic alterations already existed in the early preneoplastic liver tissue of this patient, and convergent evolution underlined the development of these tumors.

Little is known about the extent of DNA methylation changes in premalignant cirrhotic/fibrotic lesions, which has limited our knowledge on how they participate in HCC development. To address these questions, we performed WGS, WES, and methylation profiling on single or multiple regenerative nodules of neighboring cirrhotic or fibrotic liver from all cases studied (Supplementary Table 1A). We also filtered out the organ-related methylation pattern using 3 healthy NLS as reference. Analysis of cirrhotic or fibrotic nodules showed consistent absence of Indel or CNA abnormalities, except for occasional single nucleotide variants (SNVs) that were undetectable in the corresponding tumor. We next explored field effect at the epigenetic level. Focusing on epigenetic probes that could demonstrate a

progressive change in sequential order from normal to cirrhosis/fibrosis to HCC liver states, we defined a subset of the most variable CpGs (Figure 3A–F, Supplementary Figures 6 and 7). We found all liver cirrhosis/fibrosis sectors were epigenetically divergent from NLS, except for ITH-625, which was a case of mild fibrosis and closely resembled healthy controls. Notably, individual cirrhotic/fibrotic livers showed distinct clusters of aberrant methylation characteristic of their corresponding HCC. A linear progression of specific cluster nodes linking liver disease to HCC underscored epigenetic changes within cirrhotic/fibrotic nodules as truncal gatekeepers of tumorigenic potential. We found many of the early hyper- and hypomethylation changes targeted genes involved in cancer pathways, implying that they are functionally relevant.^{18–21} For instance, *SOCS2* hypermethylation correlated with repressed expression, which has been implicated in augmenting liver inflammation and fibrosis leading to HCC,¹⁹ whereas *UBD* hypomethylation (also *FAT10*) is an important mediator that promotes chronic inflammation-associated tumorigenesis when upregulated²⁰ (Figure 3G and H and Supplementary Figure 6D–F). Our results illustrated a prominent field effect from methylation abnormalities in the early preneoplastic phases of HCC even when the genome is stable.

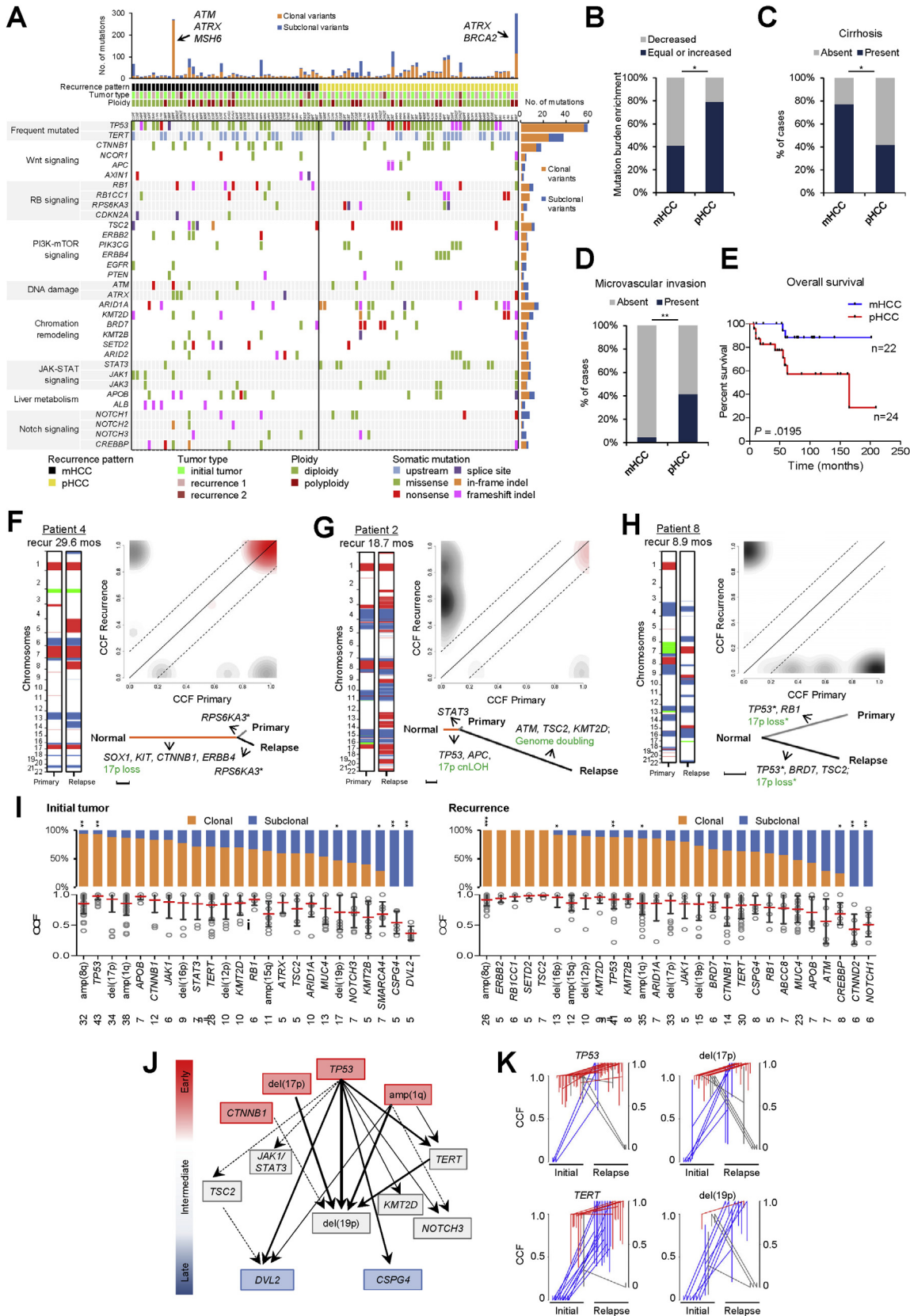
Temporal Mutational Profiles and Mutational Signatures in Tumor Progression

The supposition that early <2-year recurrences share clonal lineage with initial HCC has gained attention as it may allow the development of predictive biomarkers. To verify this, we performed targeted resequencing of 824 key HCC-associated genes (including the *TERT* promoter) in 154 tumors and matched nontumoral liver from 106 patients, including 41 matched initial-recurrent tumor pairs, 5 cases with 3 longitudinal tumors from more than 1 relapse, 29 primary HCCs with no recurrence in a median follow-up of 12.4 years, and 26 relapse tumors alone (Supplementary Tables 1–4 and Figure 4A). Direct genetic comparisons showed 2 idiosyncratic types of recurrence: progressive-HCC (52.2%) that shared extensive mutational changes with initial tumor, and multicentric-HCC (47.8%) that arose from an independent tumor lineage (Supplementary Figures 8A and 9). Intriguingly, mutation burden in progressive-HCCs was significantly higher than multicentric-HCCs ($P < .05$) (Figure 4B). Although a general trend for multicentric-HCC to relapse after >2 years was observed (Supplementary Figure 8B), 36.4% of these tumors (mean 44.1 months) shared the same clonal origin as their initial HCC. Noteworthy, 33.3% of early recurrent tumors (mean 10.3 months) were in fact multicentric-HCC. Evidently, these

Figure 3. Precancerous epigenetic changes. (A–F) Precancerous DNA methylation changes are identified in nontumoral sectors of tumor-neighboring liver cirrhosis/fibrosis from ITH-15, ITH-539, ITH-555, ITH-547, ITH-564, and ITH-566. Phyloepigenetic trees (left) and heatmap (right) illustrate precancerous epigenetic changes that diverged from NLS but are maintained in tumor lesions. Phyloepigenetic trees are constructed based on hypermethylation changes. Heatmaps are constructed based on hypomethylation changes. (G, H) Validation of precancerous methylation changes using independent dataset, The Cancer Genome Atlas (left) and in-house data (right) on both methylation (i) and expression level (ii). NT, tumor adjacent nonmalignant tissues; T, tumor tissues.

multicentric-HCC tumors commonly arose from liver cirrhosis, which again emphasizes a “field effect” in promoting multiple de novo tumors (Figure 4C). Conversely, the

genetic resemblance of progressive-HCC showed strong association with microvascular invasion and overall less favorable survival prognosis (Figure 4D and E).



To evaluate the timing of evolutionary branching in progressive-HCC and determine whether early subclonal diversification occurred in multicentric-HCC, a full genomic landscape was determined on 8 HBV-associated cases by WGS (patients 1–8) (Supplementary Tables 1B and 3D). Overall, the distribution of mutational changes, and the spectrum of CNA and HBV integration sites, all concurred in supporting the same ancestral origins in all initial-relapse pairs (progressive-HCC), except patient 5 and patient 8 (multicentric-HCC), which showed complete genetic dissimilarities (Figure 4F–H, Supplementary Figures 10 and 11A and B, Supplementary Table 3E). Two-dimensional clustering of estimated clonal fraction further suggested that progressive-HCCs exhibited 2 types of evolutionary branching: early (patients 1–3, patient 6) and late (patient 4, patient 7). Their inferred clonal fractions also revealed the majority of shared mutations were clonal in both initial-relapse tumors of progressive-HCC, although subclonal mutations in the initial tumors (mean 24.95%) were mostly lost at recurrence. Our WGS data showed that most progressive-HCC recurrences arose mainly from the divergence of a dominant driver clone rather than by subclonal evolution.

We next examined the temporal order of driver acquisitions in the evolution of HCC. Based on computed clonal fraction, we classified driver events as clonal or subclonal in 154 initial and relapse tumors. A high incidence of clonal dominance of *TP53* mutations, *del(17p)*, *amp(1q)*, and *amp(8q)* in both initial and relapse tumors underlined their importance in tumor initiation (Figure 4I). To further explore the relationship between drivers, we constructed a temporal trajectory of genetic variations by integrating both clonal fraction and co-occurrence information. For each driver, we calculated their relative enrichment to infer their emergence timepoint as early, intermediate or late. The temporal network showed that the ancestor clone was initialized by few events, mainly mutations in *TP53*, *CTNNB1*, and CNAs of *del(17p)* and *amp(1q)*, which were ubiquitously shared across time (Figure 4J and K and Supplementary Figure 11D and E). Other intermediate-stage drivers include *TERT* promoter mutations, chromatin

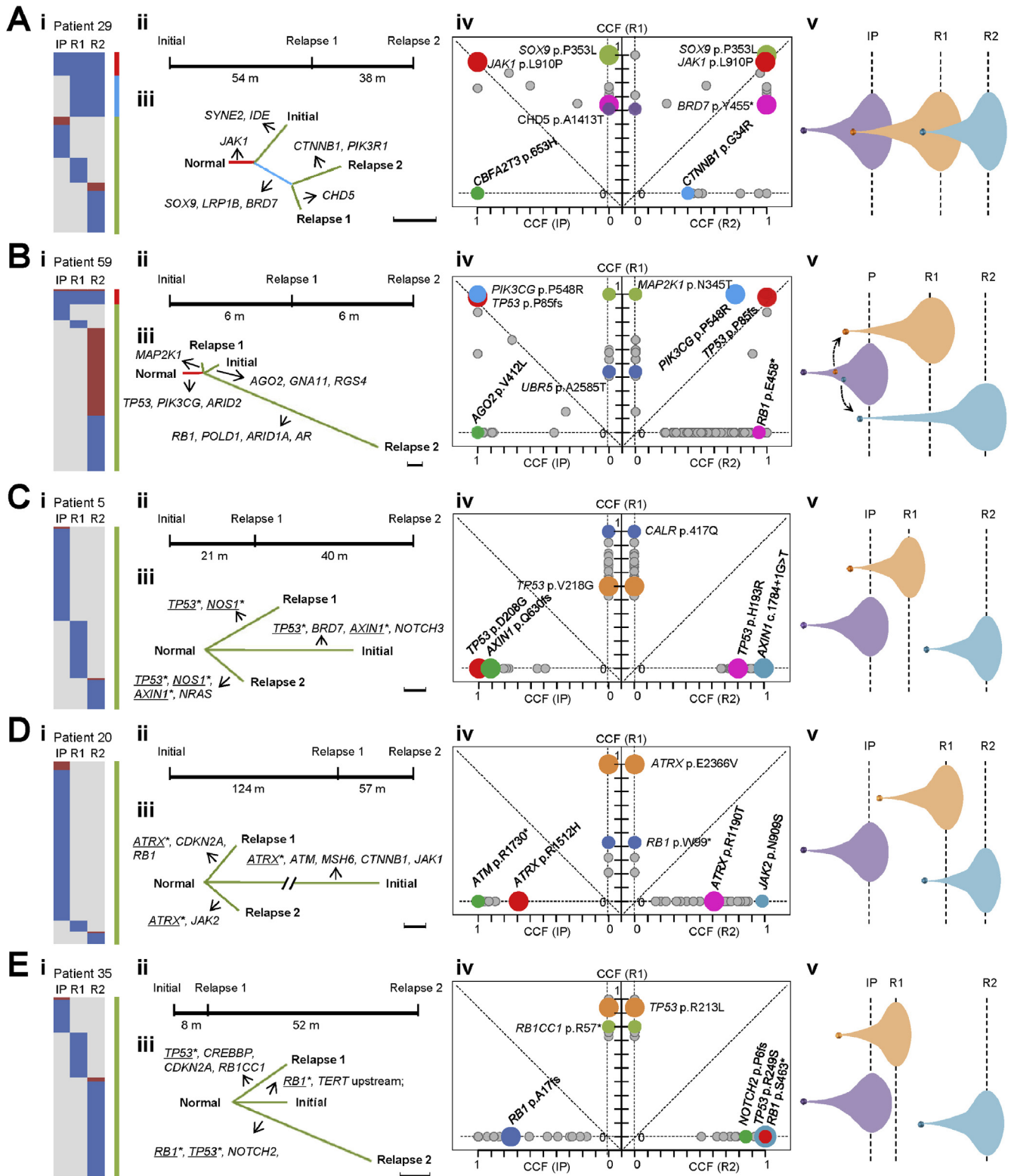
remodeling genes, and mammalian target of rapamycin, NOTCH and JAK/STAT signaling pathways.

To elucidate mutational signature in HCC progression, we analyzed clonal and subclonal mutational spectra in WGS primary-relapse pairs and ITH-15 (Supplementary Figures 11C, 12, and 13). In general, we found 2 features of maintained and evolved mutational signatures. We note maintenance of age-related signature 5 and transcriptional coupled damage-related mutational signature 16 in both clonal and subclonal mutational spectra (Supplementary Figure 12). Conversely, signatures that evolved substantially between clonal and subclonal mutations include a decrease of tobacco-related signature 4 and aflatoxin B1-related signature 24 (Supplementary Figure 12), enforcing activities of these 2 carcinogens in the early stages of tumor evolution. More excitingly, we found cases of multicentric-HCC occurrences. Although they showed divergent genetic makeup between lesions, the distribution of mutational signatures was similar, that is, between initial and relapse tumors (patients 5 and 8) and between T1 and T2 of ITH-15 (Supplementary Figure 13). Compared with previous WES study,²² our WGS analysis provided a comprehensive overview on both genome-wide mutational spectra and signatures, and direct observation of progression-related pattern of mutational signatures within each patient.

Convergent Evolution Sourced Parallel Seeding

We next analyzed patients with 2 sequential recurrences (R1 and R2) to determine the longitudinal progression of relapse tumors from initial HCC (Figure 5A–E). The phylogenetic trees indicated that 2 patients (patients 29 and 59) showed progressive-HCC recurrences where both R1 and R2 shared clonal similarities to the first tumor. Of interest, patient 59's R1 and R2 tumors branched off early from the initial HCC, probably existing as micrometastases before first surgery. In patient 5, patient 20, and patient 35, multicentric-HCC arose from independent clones in each initial, R1, and R2 tumor. We carefully examined the mutational profiles of these multicentric-HCCs and, strikingly, found many shared driver genes but mutated at

Figure 4. Temporal genetic heterogeneity. (A) Mutational profiles in targeted resequencing data. Stacked bar chart (top) shows the number of clonal and subclonal mutations for each sample. Grid plot represents clinical elements, mutations and CNAs. Right stacked bar chart shows the distribution of clonal and subclonal status for mutations identified in representative drivers. (B) Mutation burden in progressive-HCC (pHCC) and multicentric-HCC (mHCC) recurrences. Cases are separated into 2 groups based on the comparison of mutation burden between recurrent tumor and matched primary tumor (Fisher's exact test: $P < .05$). (C, D) The distribution of cirrhosis (C), microvascular invasion (D) in initial tumors with recurrence of multicentric-HCC (mHCC; $n = 22$) and initial tumors with recurrence of progressive-HCC (pHCC; $n = 24$) (Fisher's exact test: $*P < .05$, $**P < .01$). (E) Kaplan-Meier curves (Log-rank test: $P = .0195$). (F–H) Genetic divergence between initial and recurrent tumors. Two-dimensional density plot (upper-right panel) based on clustering of CCFs of all mutations identified at initial (x -axis) and at relapse (y -axis). Clouds with shared mutations are in red; the others in gray. CNAs were illustrated on the left panel (red, CN gain; blue, CN loss; green, copy neutral loss of heterozygosity). Phylogenetic trees (bottom) are inferred from somatic mutations. One scale unit represents 20 mutations. Genes labeled with an asterisk (*) represent different mutations identified from the same individual. Mos, months. (I) Cancer Cell Fraction (CCF) distributions. The CCF distribution of each driver tested using Fisher's exact test, and compared with the rest of the drivers ($*P < .05$, $**P < .01$, $***P < .001$). (J) Inferred temporal order. Dashed line, evidenced by 2 samples; black line, evidenced by more than 2 samples, with line thickness proportionally related to the number of supporting samples. (K) Clonal fraction changes as determined by CCF between initial-relapse pairs. Red, CCF maintained; blue, CCF increased; gray, CCF decreased.



BASIC AND TRANSLATIONAL LIVER

F

| | With convergent mutation | Without convergent mutation |
|------|--------------------------|-----------------------------|
| mHCC | 14 | 8 |
| pHCC | 6 | 18 |

Fisher's exact test $P = .0163$

G

| | HCC drivers | Others |
|-----------------------------|-------------|--------|
| With convergent mutation | 16 | 26 |
| Without convergent mutation | 27 | 364 |

Fisher's exact test $P < .001$

different positions. For instance, different *TP53* GOF mutations within the DNA-binding domain were found in patient 5's initial, R1, and R2 tumors, and between R1 and R2 tumors of patient 35 (Figure 5C and E). Likewise, 3 distinct inactivating mutations of *ATRX* were found in the initial, R1, and R2 tumors of patient 20 (Figure 5D). We further asserted such mutation switching was prevalent between initial-relapse pairs (Supplementary Table 5), although more common in multicentric-HCC (63.64%) than progressive-HCC (25.0%) (Fisher's exact test $P = .016$; Figure 5F). More so, such convergent mutations were frequent among HCC driver genes (Fisher's exact test $P < .0001$, Figure 5G). Notably, in addition to *TP53*, other convergent drivers included activating mutations of *CTNNB1* and *TERT* promoter (also convergent HBV integration sites in the *TERT* promoter seen in patient 8) (Supplementary Figure 11A) and inactivating frameshift, splicing or stop gain mutations in *RB1*, *AXIN1* and *RPS6KA3* (Supplementary Table 5). Overall, the dysfunction of genes from convergent evolution reflects their importance in carcinogenesis,^{23,24} and infers an environmental selection that allows parallel seeding of ancestor clones before first surgery.

Exploring Truncal Mutations for Potentials as Actionable Targets

GOF *TP53* mutations were detected in 41.3% of cases, in which co-occurring del(17p) underscored a dominant effect in >25%. Missense mutations of *TP53* clustered within the core DNA-binding domain, where R249S was most common (Figure 6A). To explore the transcriptional effect of mutant p53^{R249S}, we performed chromatin immunoprecipitation-sequencing in the HCC cell line HKCI-11, which has a p53^{R249S} mutation, and the human hepatocyte line MIHA, which is p53^{WT} (Supplementary Table 2D). Our analysis showed high disparity in overall number and distribution of transcription start site-proximal peaks between R249S and WT (Figure 6B–D and Supplementary Table 6). Notably, R249S peaks were enriched in the promoter proximal regions that colocalized with active histone mark H3K27ac signals, which supports their direct participation in active transcription. Based on gene ontology analysis, we were intrigued to find significant enrichment of genes related to histone methylation activities (Figure 6E). Strong promoter occupancy for selected chromatin regulators was confirmed

for R249S and another GOF mutant Y220C by chromatin immunoprecipitation-quantitative polymerase chain reaction. In contrast, MIHA TP53^{WT} did not appear to bind any of these genes, although as expected it bound to the promoter region of canonical target *CDKN1A* (Figure 6F–I and Supplementary Figure 14A and B). We validated *MLL1* and *MLL2* as direct functional targets of p53^{R249S} and affirmed their association in the Cancer Genome Atlas dataset (Figure 6J and Supplementary Figure 14C). Because targeting chromatin regulators by small molecules has emerged as a promising avenue for cancer therapy, we proceeded to test the pharmacological effects of 2 MLL-complex antagonists, MI-2-2 (inhibitor of protein interaction) and OICR-9492 (inhibitor of activity). Both MI-2-2 and OICR-9492 showed potent inhibition of R249S cell growth at nanomolar concentrations, but had a negligible effect on p53^{WT} cells (Figure 6K and L). These results provide the first evidence in HCC of the growth inhibitory effect of pharmacological compounds that intervene MLL functionality downstream of GOF p53.

Given that activating *STAT3* mutations are found in 20% of patients with progressive-HCC, we next attempted to target GOF mutants using a *STAT3* inhibitor, S3I-201. Stably expressed mutants in 3 wild-type cell lines consistently showed elevated pSTAT3-Y705 and potent reporter activity in the absence of interleukin-6 induction, suggesting a strong oncogenic effect of common somatic variants (D170Y, K348E, and Y640F) (Supplementary Figure 14D and E). When treated with S3I-201, constitutive activation caused by *STAT3* oncogenic mutants showed marked diminution in all cell lines (Figure 6M and N). Beyond *STAT3* mutants, persistent pSTAT3-Y705 also could be induced by *JAK1* activating mutations (S703I and L910P) (Supplementary Figure 14F and G). Our results provide an approach to interfere with progressive-HCC recurrence.

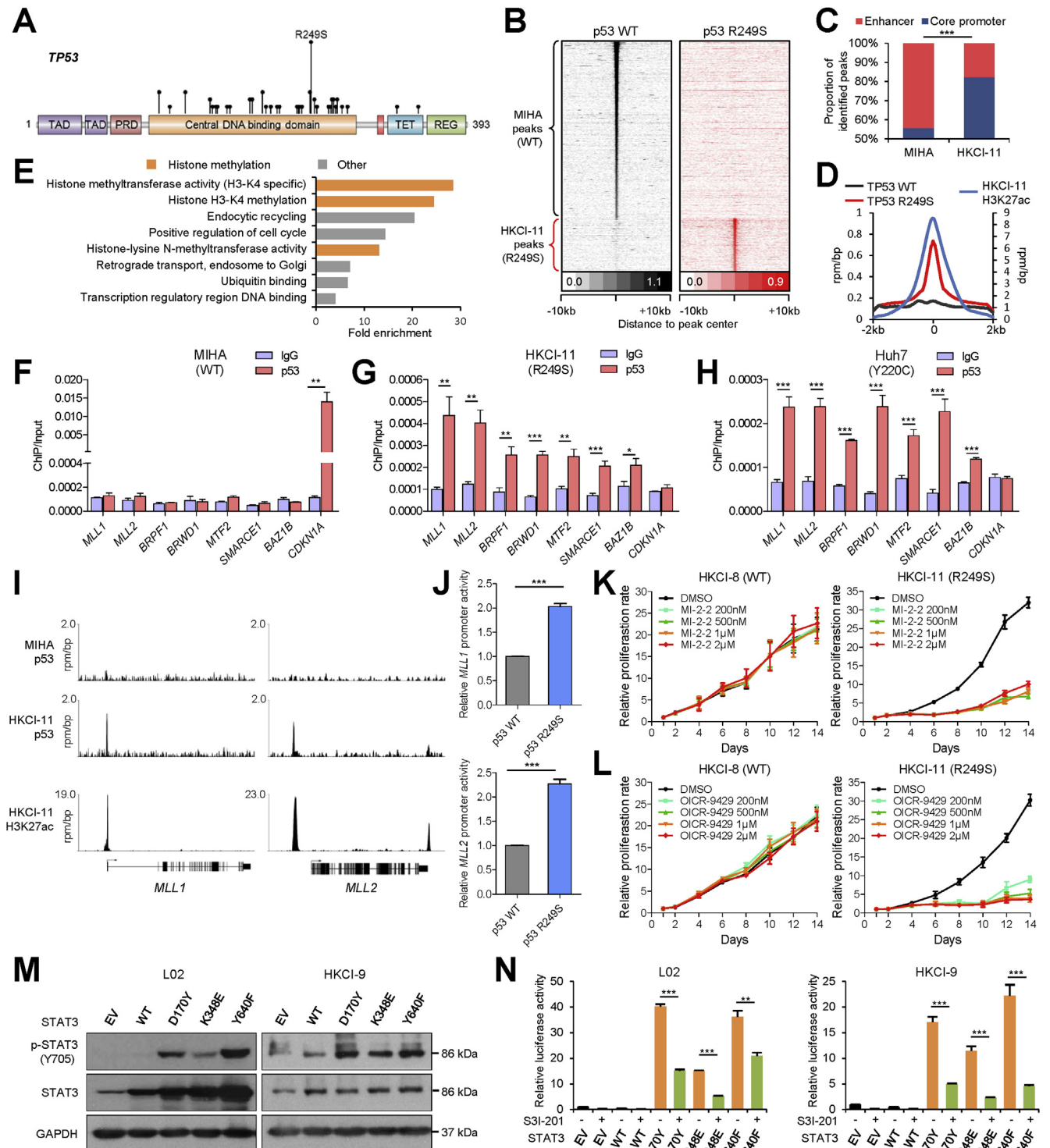
Discussion

Here, we systemically examined tumor heterogeneity in HCC, focusing on the clonal aspects of intra- and intertumor heterogeneities, and tumor recurrence. Based on trajectories derived from genetic and epigenetic aberrations, we reconstructed the evolutionary history and course of clonal development of HCC (Figure 7). We identified founder drivers, including double hits on *TP53*, and GOF mutations in *CTNNB1* and the *TERT* promoter, and more importantly a

Figure 5. Evolution patterns in sequential recurrences. (A–E) Genetic profiles on patients with 2 recurrences. (i) Panel shows the proportion of ubiquitous mutations (blue) and indels (dark red) across initial primary (IP) and their matched 2 recurrences (R1 and R2). The right bar represents aberrations private to IP or shared by 2 recurrences (light blue), or ubiquitously identified. (ii) The timeline of surgical histories for each patient. (iii) Phylogenetic trees constructed based on the number of identified mutations. One scale unit represents 5 nonsynonymous mutations. Convergent mutations are highlighted with both asterisk (*) and underlined. (iv) Panel shows the distribution of Cancer Cell Fraction (CCF) values for each mutation in drivers (left x-axis, initial primary; y-axis, first recurrence; right x-axis, second recurrence). (v) Schematic representation of progressive pattern in each patient. Patient 29 displays a typical progressive pattern with R2 derived from ancestor cells that existed in its R1 counterpart. In contrast, phylogenetic tree of patient 59 indicates the 2 recurrences branched off at similar time point, highlighting an early divergence of initiators from initial tumor. Multicentric-HCCs are suggested in patients 5, 20, and 35. (F, G) Distribution of convergent mutations between recurrent types and between passenger and driver genes. Recurrent key drivers for HCC include *TP53*, *TERT*, *CTNNB1*, *TSC2*, *JAK1*, *NOTCH1*, *FGFR3*, *ATRX*, and *RPS6KA3*.

common scenario of convergent mutations that favor each ancestor clone to experience a parallel series of expansions. While dysfunctional genes involved in convergent evolution have been described for other cancer types,^{23,24} they have been minimally discussed in HCC. We defined in this study convergent key drivers of HCC, including *TP53*, *TERT*, *CTNNB1*, *TSC2*, *JAK1*, *NOTCH1*, *FGFR3*, *ATRX*, and *RPS6KA3*.

Furthermore, our integrated phylogenetic and phyloepigenetic analysis provide new insight into a co-evolutionary relationship between the genome and epigenome in HCC. Notably, such genomic-epigenomic codependency has also been reported in other cancer types, such as brain tumors,²⁵ prostate,²⁶ and esophageal cancers.²⁷ Together, this highlights that close interplay between the genome and



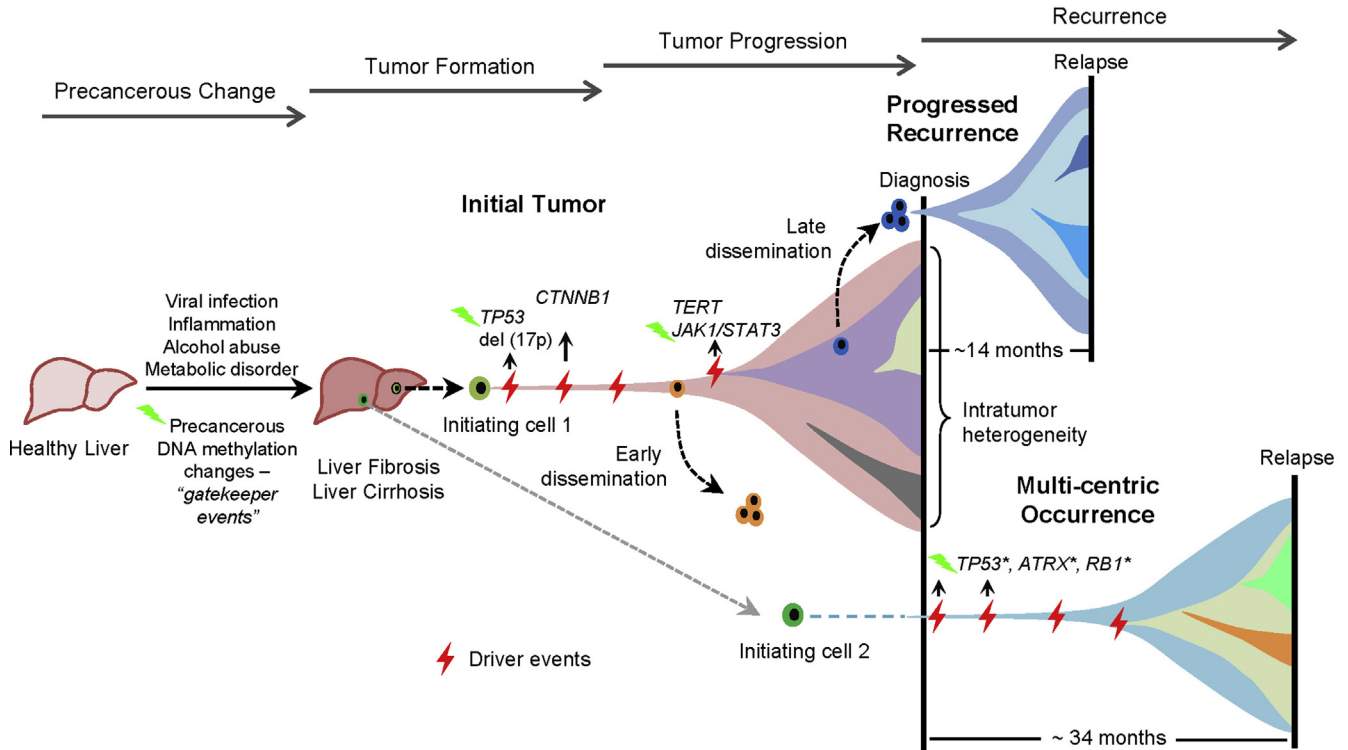


Figure 7. An example of tumor evolutionary trajectories and types of recurrence in HCC.

epigenome is a general phenomenon during malignant progressions of human cancers.

The inference that early recurrences within 2 years are attributed to intrahepatic micrometastasis has been a conventional classification. We found a concordant trend for progressed recurrence to relapse at mean 14 months but it is noteworthy that a third of these early recurrences were in fact second primary tumors. Conversely, 36.4% of late relapses were actually progressed recurrences suggesting dormancy of disseminated tumor cells is also common in HCC. From a clinical standpoint in patient management, it may be necessary to re-biopsy recurrences regardless of relapse time to establish clonal origin; although a single-sampling approach may suffice capturing the clonality because truncal events are ubiquitously distributed.

Our study also highlighted multicentric-HCC could arise as early as 4 months following first surgery. The apparent synchronous development of de novo clones exemplifies a “field cancerization” effect from liver cirrhosis/fibrosis. The presence of convergent mutations despite the lack of clonal resemblance in more than half of multicentric-HCC cases suggests these tumors underwent selection from the same microenvironment. Despite tumor adjacent liver cirrhosis and fibrosis has long been considered the putative pre-malignant lesion of HCC, existing literature does not suggest presence of recurrent somatic mutations in these alleged precancerous states.^{11,28,29} Our data show that individual cirrhotic/fibrotic nodules had acquired DNA methylation changes characteristic of HCC, a mechanism that appeared to predominate over genome alterations. These epigenetic

BASIC AND TRANSLATIONAL LIVER

Figure 6. GOF mutations in *TP53* and *STAT3* can be actionable targets. (A) Schematic diagram of identified missense mutations in *TP53*. (B) Heatmap of p53 binding peaks in MIHA cells and HKCI-10 cells. Each row shows binding intensity centered on called peaks with a flanking region of 10 kb. (C) The proportion of regions defined as enhancers or core promoters bound by wild-type p53 and mutant p53 (R249S) (Fisher’s exact test: ****P* < .001). (D) Average binding intensity for wild-type p53, mutant p53 (R249S) and H3K27ac in HKCI-11 across the identified binding region of R249S. Left y-axis scaled for binding peaks of wild-type p53 and R249S p53; whereas, right y-axis was scaled for binding peaks of H3K27ac in HKCI-11. (E) Gene ontology analysis of genes with transcription start site–proximal peaks in R249S p53 chromatin immunoprecipitation (ChIP)-sequencing data. (F–H) ChIP-quantitative polymerase chain reaction of selected chromatin regulators in HKCI-11 (R249S), Huh7 (Y220C) and MIHA (WT). (I) ChIP-Sequencing signals for *MLL1* and *MLL2*. (J) Dual-luciferase assay showing promoter activities of *MLL1* and *MLL2* in HEK293FT cells overexpressed with wild-type p53 and R249S p53 (Student *t* test: ****P* < .001). (K, L) Growth curve of HKCI-8 and HKCI-11 on treatment with varying dosages of MI-2-2 (K) and OICR-9492 (L). (Error bars represent mean ± SEM from 3 independent experiments; Student *t* test: **P* < .05; ***P* < .01; ****P* < .001). (M) Western blot measuring the phosphorylation levels of *STAT3* in L02 cells and HKCI-9 cells stably expressing vector control, wild-type *STAT3*, *STAT3* mutants. All the tested mutants displayed markedly elevated phosphorylation level at Y705 of *STAT3* protein. (N) The increased *STAT3* transcription activity by *STAT3* mutants is readily inhibited by *STAT3* inhibitor S3I-201 (Student *t* test: ****P* < .001).

changes include truncal aberrations such as *UBD* and *SOCS2* that have reported functions in linking inflammation to cancer.^{19,20} It is plausible that initial “gatekeeper” changes in DNA methylation provide a selective growth advantage to a normal hepatocyte, allowing it to outgrow surrounding cells and become a microscopic clone when combined with additional genetic “drivers.” Similar to molecular subtyping in cancers, we hypothesize that assigning liver cirrhosis and fibrosis to groups according to their DNA methylation patterns could serve as a predictive marker to reflect the potential risk of HCC development and de novo recurrence in high-risk patients after surgery.

Therapeutic targeting of truncal events in HCC remains a challenge. Our data show that GOF p53 mutation leads to independent transcription activation in HCC that can directly regulate a distinct set of chromatin regulators, such as histone methylation (*MLL1*, *MLL2*), histone acetylation (*BRPF1*), chromatin remodeling complex (*SMARCE1*, *BRWD1*, *BAZ1B*) and component of the Polycomb repressive complex (*MTF2*). We provide first evidence on the feasibility to target chromatin regulators activated by mutant p53 in HCC using small molecule inhibitors, and described a new area of genetic dependencies related to mutant p53 that could be tested for effects of chromatin drugs. In fact, 2 MI-2-2-derived lead compounds, MI-503 and MI-463, are currently being developed for human trials. In summary, our work demonstrates the genomic and epigenomic architectures of HCC tumors can inform targeted therapeutic interventions and identify high-risk patients.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at <https://doi.org/10.1053/j.gastro.2019.09.005>.

References

1. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;385:117–171.
2. Portolani N, Coniglio A, Ghidoni S, et al. Early and late recurrence after liver resection for hepatocellular carcinoma: prognostic and therapeutic implications. *Ann Surg* 2006;243:229–235.
3. Imamura H, Matsuyama Y, Tanaka E, et al. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. *J Hepatol* 2003;38:200–207.
4. Wu JC, Huang YH, Chau GY, et al. Risk factors for early and late recurrence in hepatitis B-related hepatocellular carcinoma. *J Hepatol* 2009;51:890–897.
5. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* 2008;359:378–390.
6. Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012;481:306–313.
7. El-Serag HB. Hepatocellular carcinoma. *N Engl J Med* 2011;365:1118–1127.
8. Llovet JM, Zucman-Rossi J, Pikarsky E, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers* 2016;2:16018.
9. Vogelstein B, Papadopoulos N, Velculescu VE, et al. Cancer genome landscapes. *Science* 2013;339:1546–1558.
10. Ling S, Hu Z, Yang Z, Yang F, et al. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci U S A* 2015;112:E6496–E6505.
11. Zhai W, Lim TK, Zhang T, et al. The spatial organization of intra-tumour heterogeneity and evolutionary trajectories of metastases in hepatocellular carcinoma. *Nat Commun* 2017;8:4565.
12. Lee JK, Wang J, Sa JK, Ladewig E, et al. Spatiotemporal genomic architecture informs precision oncology in glioblastoma. *Nat Genet* 2017;49:594–599.
13. Sun R, Hu Z, Sottoriva A, et al. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat Genet* 2017;49:1015–1024.
14. Waclaw B, Bozic I, Pittman ME, et al. A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* 2015;525:261–264.
15. Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* 2017;169:1327–1341.e23.
16. Villanueva A, Portela A, Sayols S, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology* 2015;61:1945–1956.
17. Sottoriva A, Kang H, Ma Z, et al. A Big Bang model of human colorectal tumor growth. *Nat Genet* 2015;47:209–216.
18. Chin SP, Marthick JR, West AC, et al. Regulation of the *ITGA2* gene by epigenetic mechanisms in prostate cancer. *Prostate* 2015;75:723–734.
19. Fan Y, Fang X, Tajima A, et al. Evolution of hepatic steatosis to fibrosis and adenoma formation in liver-specific growth hormone receptor knockout mice. *Front Endocrinol (Lausanne)* 2014;5:218.
20. Ren J, Wang Y, Gao Y, et al. FAT10 mediates the effect of TNF- α in inducing chromosomal instability. *J Cell Sci* 2011;124:3665–3675.
21. Tao R, Wang ZF, Qiu W, et al. Role of S100A3 in human hepatocellular carcinoma and the anticancer effect of sodium cantharidinate. *Exp Ther Med* 2017;13:2812–2818.
22. Chen G, Cai Z, Li Z, et al. Clonal evolution in long-term follow-up patients with hepatocellular carcinoma. *Int J Cancer* 2018;143:2862–2870.
23. Heaphy CM, de Wilde RF, Jiao Y, et al. Altered telomeres in tumors with ATRX and DAXX mutations. *Science* 2011;333:425.
24. Liu XY, Gerges N, Korshunov A, et al. Frequent ATRX mutations and loss of expression in adult diffuse astrocytic tumors carrying IDH1/IDH2 and TP53 mutations. *Acta Neuropathol* 2012;124:615–625.

25. **Mazor T, Pankov A**, Johnson BE, et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell* 2015;28:307–317.
26. Brocks D, Assenov Y, Minner S, et al. Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell Rep* 2014;8:798–806.
27. **Hao JJ, Lin DC, Dinh HQ, Mayakonda A, Jiang YY**, et al. Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma. *Nat Genet* 2016;48:1500–1507.
28. **Lin DC, Mayakonda A, Dinh HQ, Huang PB, Lin L**, et al. Genomic and epigenomic heterogeneity of hepatocellular carcinoma. *Cancer Res* 2017;77:2255–2265.
29. **Nault JC, Calderaro J**, Di Tommaso L, et al. Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of pre-malignant nodules in hepatocellular carcinoma on cirrhosis. *Hepatology* 2014;60:1983–1992.

Author names in bold designate shared co-first authorship.

Reprint requests

Address requests for reprints to: Nathalie Wong, Department of Surgery, The Chinese University of Hong Kong, Prince of Wales Hospital, N.T., Hong Kong. e-mail: natwong@cuhk.edu.hk.

Acknowledgments

The authors are indebted to the Silent Teachers Program of CUHK for the HCC materials from deceased patient. We are grateful to Prof. Simon Herrington from the University of Edinburgh for his critical reviewing and scientific editing. We also thank the Core Utilities of Cancer Genomics and Pathobiology (CUHK) for providing the facilities and assistance in support of this research.

Author contributions: XD and MH analyzed and interpreted data, prepared, and wrote the manuscript. XD performed the bioinformatics analysis of WGS, methylation, and targeted deep sequencing. XD and XW performed the bioinformatics analysis of chromatin immunoprecipitation-sequencing. MH and QXS prepared samples for WGS and target capture screening. MH, SCS, HC, and MKHM performed biological experiments, chromatin immunoprecipitation-sequencing validation, luciferase, quantitative polymerase chain reaction, WB, and Sanger sequencing. AWHC, SLC, PBSL, and KM provided liver cancer specimens and the associated clinical annotations. NW conceived and supervised the study, and prepared and wrote the manuscript.

Conflicts of interest

The authors disclose no conflicts.

Funding

This work was supported by the Theme-based Research Scheme from the Hong Kong Research Grants Council (RGC) (Ref. T12-403/11), in part by a Collaborative Research Fund from the Hong Kong RGC (Ref. C4041-17) and the VC's discretionary fund from the Chinese University of Hong Kong (CUHK).

Supplementary Materials and Methods

Targeted Resequencing: Custom Panel Probe Design

Custom capture probes targeting a 3.2 Mb genomic region was designed using SeqCap EZ design service (Roche, Basel, Switzerland). Target regions were selected to include the promoter region of TERT and coding exons of 823 selected genes, which included (1) mutated genes that might play roles in carcinogenesis, (2) genes that were recurrently mutated in our WGS cohort, and (3) genes that were reported to be frequently mutated in HCC. These genes were confirmed to have substantial detection power in 1129 HCCs from the ICGC project, 12 among which 97.87% showed at least one mutation within our selected regions.

Targeted Capture Sequencing

Extracted DNA samples were constructed into Illumina pre-capture libraries using KAPA Hyper Prep Kit (KAPA Biosystems, Roche Sequencing and Life Science, Wilmington, MA). The pre-capture libraries passed QC were pooled and enriched for target regions using NimbleGen SeqCap EZ Choice Kit (Roche). Briefly, pooled pre-capture libraries were hybridized in reaction mixture containing COT Human DNA, Multiplex Hybridization Enhancing Oligo Pool, Hybridization Buffer, Hybridization Component A, and SeqCap EZ custom probes using SeqCap EZ Accessory Kit (Roche), SeqCap HE-Oligo Kit A and B (Roche), and SeqCap EZ Hybridization and Wash Kit (Roche). The hybridization mixture was then incubated at 47°C for 72 hours. Enriched libraries were captured by SeqCap EZ Pure Capture Beads (Roche) and nonspecific sequences were washed away by Stringent Wash Buffer, Wash Buffer I, II, and III using SeqCap EZ Hybridization and Wash Kit (Roche). The post-capture libraries were then amplified by ligation-mediated polymerase chain reaction (PCR) using SeqCap Accessory Kit (Roche). The efficiencies of the capture reactions were evaluated by performing quantitative PCR (qPCR)-based quality check on built-in Nimblegen Sequence Capture controls using Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA). Post-capture libraries were quantified by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA) and sequenced by Illumina (San Diego, CA) HiSeq 2000. Library fragment sizes were resolved by BioAnalyzer High Sensitivity DNA Analysis Kit (Agilent, Santa Clara, CA).

Sequence Alignment and Variant Calling

Genomic sequence data were initially processed with the following steps:

- (1) Raw reads were quality checked with Fastqc (v0.11.4) and filtered by fastq-mcf (v1.04.636). The remaining high-quality reads were aligned to a combined reference genome, including human genome hg19 and consensus HBV genome, using BWA-MEM algorithm with default parameters (version 0.7.12)^{1,2} and sorted by samtools (version 0.1.19).³ Aligned reads were tagged by Markduplicate in Picard (v1.115), locally

realigned (IndelRealigner) and recalibrated (Base-Recalibrator) with GATK Analysis Toolkit (v3.1.1) (Broad institute, Cambridge, MA).⁴

- (2) Confident single nucleotide polymorphism (SNP) callings were generated by Unifiedgenotyper. Somatic SNVs were called by MuTect (v1.1.7) (Broad institute, Cambridge, MA).⁵ Somatic Indels were detected by Platypus (v1.0) (The Wellcome Centre for Human Genetics, Roosevelt Drive, Oxford, UK),⁶ Strelka (v1.0) (Illumina, San Diego, CA),⁷ and Scalpel (v0.5.2) (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).⁸
- (3) Somatic SNV candidates were filtered when detected in normal sample panel and dbSNP 135⁹ except when they were recorded in COSMIC database (v71).¹⁰ Somatic Indel candidates were only retained with one of the following conditions satisfied: (1) Indels at least called by 2 tools; (2) manually checked.
- (4) The final variants were then combined and annotated by SnpEff and SnpSift (v3.6a).^{11,12} Most non-silent mutations were confirmed by extensive manual inspection.

HBV Integration Detection in Whole-Genome Sequence Data

HBV integration in human genome would cause reads spanned the integration site (human-virus chimeric reads). Therefore, we focused on reads mapped to HBV genome, chimeric reads as well as unmapped reads and realigned them by BWA-SW algorithm (v0.7.12).² Next, we extracted all the unique soft-clipped reads with one end on the virus genome and the other on the human genome. Based on these integration sites, we defined a set of supporting reads by including paired end of human-virus chimeric reads and read pairs close to these sites. We selected only confident HBV integration events with more than 3 supporting read pairs and at least 1 chimeric read for analysis.

Copy Number Variation Analysis

For WGS data, we first obtained allele frequencies of all the germline SNPs with sequence depth more than 30, which were then used to yield phased haplotype with Impute2^{13,14} in 5M segments. We further divided the haplotype blocks into non-overlapped neighboring windows and calculated their representative B-allele frequency (BAFs) by averaging phased allele frequencies. Meanwhile, Control-Freec (v6.8)^{15,16} was exploited to get the log₂ read ratios between samples and control with 500 base pair (bp) window size, which were then transferred into intensity values array platform. On the basis of the normalized ratios and phased BAFs, we used ASCAT2¹⁷ to generate ploidy, purity and CNAs within each sample.

For captured sequence data, we divided human genome into 200k nonoverlapped consecutive windows and counted the reads within each region by the coverage module in Bedtools (v2.17) (Quinlan Lab, University of Utah, Salt Lake City, UT).¹⁸ We then selected regions with positive read count and built a linear model to remove background noise.

The corrected read counts in tumor samples and their matched normal samples were next integrated to get the log2 ratios, which were segmented by circular binary segmentation algorithm.¹⁹ Copy number analysis was then performed by ABSOLUTE (v1.0.6).²⁰

To define CNA in intratumor heterogeneity, we summarized the percentage of genomic regions with public CNA and the percentage of genomic regions subjected to nonpublic CNA.²¹ The proportion of public CNA was defined as the percentage of genome subjected to public CNA divided by the percentage of genome subjected to any CNA in any region from the same tumor.

Estimation of Clonal Fraction by Cancer Cell Fraction (CCF) and 2-Dimensional Clustering of CCF

We used the framework in ABSOLUTE (v1.0.6)²⁰ to infer the posterior probability distribution of CCF for each mutation. Briefly, it consisted of 2 parts: inferring mutation multiplicity (m) and calculating posterior probability of CCF with estimated multiplicity. We used the estimated values of purity (p), CNA (q), and CNA CCF (f_{cnv}) from CNA analysis to build a probabilistic model, where mutant allele frequency was represented as f . We first calculate the expected allele frequency for clonal mutations with $m = 1$:

$$f_s = \frac{p}{2(1-p) + pqf_{cnv}}$$

When both clonal and subclonal CNAs were identified within the same genomic region, we consider only clonal CNA here for simplicity. The possible multiplicities for mutations was then:

$$m_q = \{1, \dots, q\}$$

The inferred allele frequency with different multiplicity was as follows:

$$f_m = mf_s$$

The total likelihood of observed allele frequency was then the following:

$$Lik(\hat{f}|p, q, n, w_c, w_s) = \sum_{m \in m_q} [w_c \text{Beta}(f_m | n\hat{f} + 1, n(1 - \hat{f}) + 1)] + w_s \int_0^{\hat{f}_s} \text{Beta}(\hat{f} | n\hat{f} + 1, n(1 - \hat{f}) + 1) \frac{1}{f_s} df$$

where \hat{f} was the observed allele frequency. The weights for clonal events (w_c) and subclonal events (w_s) were both set to 0.5. The probability of mutation with different multiplicities was then assessed and we select multiplicity with highest probability for the next step. The posterior distribution of CCF for mutations was then obtained by the following:

$$\Pr(f|f_m, n, \hat{f}) = \frac{\text{Beta}(ff_m | n\hat{f} + 1, n(1 - \hat{f}) + 1)}{\int_0^1 \text{Beta}(ff_m | n\hat{f} + 1, n(1 - \hat{f}) + 1) df}$$

Here we used an approximation approach by calculating such distribution based on 100 equally spanned discrete points from 0.01 to 1. We also tried another way to define the probability distribution by integrating over all possible multiplicities,²² and the results showed high concordance with this simplified method. Clonal events were then defined as aberrations with modal CCF greater than 0.9.

We also used a parametric Bayesian Dirichlet process to model the mutation clusters in initial-relapse pairs as described in Bolli's study.^{22,23} Gibbs sampling was used to estimate the posterior distribution of clonal fraction of each cluster. The Markov chain was run for 2000 iterations, with the first 1000 discarded.

Phylogenetic Tree and Phyloepigenetic Tree

We constructed a phylogenetic tree on the basis of shared and private nonsilent aberrations. In each phylogenetic tree, branch length was linearly related to the mutations it covered. To find the most representative subclones from each region, we used only mutations with CCF > 0.2 to construct the phylogenetic tree for tumors with multiregional sampling.

Methylation data was processed by minfi package.²⁴ Somatic methylation changes within individual samples were defined as probes with methylation changes large than 0.2 when compared with those in control samples. To construct phyloepigenetic trees, we first defined epigenetic drivers in HCC showing consistent changes in 2 independent cohorts, The Cancer Genome Atlas (TCGA) ($n = 380$)²⁵ and the study of Villanueva et al²⁶ ($n = 243$). Only hypermethylated Epdrivers were selected to build phyloepigenetic trees. Genetic and epigenetic distance matrices were calculated using the Euclidean distance metric. Phylogenetic and phyloepigenetic trees were then inferred by the neighbor-joining method in ape package.²⁷ Bootstrapping with 1000 bootstrap replicates was applied

for each tree to assess strength of support for each clade. Similarity was assessed by calculating the Pearson's correlation coefficient between the genetic and epigenetic distance matrices. Permutation test was applied to generate empirical P values by comparing correlation values between these 2 kinds of distance matrices to a null distribution generated through randomly permuting sample labels for 100,000 times. Hierarchical clustering also

was performed accordingly on the basis of hypomethylation Epidrivers with the greatest variations. To confirm association between phyloepigenetic and phylogenetic trees, we then generated a cophylogenetic plot for each case by the “cophylo” function from the phytools package in R (<http://www.phytools.org>).

Inference of Mutation Order

Based on the assumption that clonal mutations were earlier than subclonal mutations, we used the same approach in Landau’s study to infer the relative temporal order of driver mutations.²² First, we collected all the edges that start with one clonal driver event d1 and end with one subclonal driver event from the same sample. We defined early, late, and intermediate events as events with significantly higher number of out-degrees compared with that of in-degrees, extremely lower number of out-degrees compared with that of in-degrees, and balanced out-degrees and in-degrees, respectively. The relative order of drivers was then drawn on selected drivers with minimum of 2 supporting evidence.

GOF p53 Data Analysis

All the chromatin immunoprecipitation sequencing (ChIP-seq) data were aligned to human genome hg19 using Bowtie2.²⁸ HOMER (v4.8.3)²⁹ was exploited here to generate and annotate regions with significant enrichment of binding signals. Enhancer regions were defined as regions with peaks identified within certain distance (5 kb to 50 kb) from transcription start sites of protein-coding transcripts and colocalized with H3K27ac binding peaks in HKCI-11. All of binding intensities were normalized by bamliquidator (v1.0) (<https://github.com/BradnerLab/pipeline/wiki/bamliquidator>) to values in unit of reads per million mapped reads per bp (rpm/bp). Average binding profiles centered on GOF p53 binding regions (± 2 kb distance) were generated based on the average binding values from neighboring windows surrounding p53 binding centers. Heatmap of ChIP-seq occupancy for wild-type and GOF p53 were created for the ± 10 kb region flanking all identified peaks with each row representing a specific binding region. Rows were first ranked by average binding intensity in wild-type p53 and then ranked by average binding intensity in GOF p53. For gene ontology analysis, we selected transcription start site proximal peaks (within 5 kb of the transcription start site) of GOF p53 with a focus density ratio greater than 0.73 and processed their associated genes in DAVID.^{30,31} Gene ontology terms with fewer than 2 genes also were discarded.

Mutation data and RNA expression data for public HCC samples were directly downloaded from TCGA. We defined GOF mutations in *TP53* as those reported and missense mutations found within the DNA binding domains of p53; whereas, GOF mutations in *STAT3* and *JAK1* were those known missense mutations. The original RNA expression values (normalized read counts) from TCGA datasets were used in our analysis.

Simulation Model

Similar to the described computational frameworks,³² we built a 3-dimensional tumor growth model with tumor cells occupying sites of a regular 3-dimensional square lattice and proliferating according to certain rules.³³ Each cell in this model owned a set of genetic aberrations (passenger or driver), which determined both its growth rate and survival rate. The mutation rate for passengers, drivers were initially set per each simulated tumor together with cell size and growth advantage. After 5000 simulations, we sliced each virtual tumor according to our in-house sampling strategy. We next exploited Approximated Bayesian computation to obtain posterior distributions of each parameter by fitting properties of simulated tumors to those of our ITH samples. Here, these summary properties including statistics of subclonal mutations, such as number and frequency distribution, and intratumoral genetic indices.³⁴ The ITH indices were calculated based on average values of all pairwise comparison values for private mutation numbers and Nei’s genetic distances within a single tumor. Nei’s genetic distance was defined as follows³⁵:

$$D_{nei} = -\log \frac{\sum (x_i y_i + (1 - x_i)(1 - y_i))}{\sqrt{(\sum x_i^2 + (1 - x_i)^2)(\sum y_i^2 + (1 - y_i)^2)}}$$

where x_i and y_i represented the CCF values in sample x and y for mutation i and all mutations from sample x and y were calculated in this genetic distance. By including these 2 metrics here, we were able to perform inference analysis to obtain estimated parameters for each tumor. Meanwhile, despite a varied number of regions in our HCC cases, distributions of these 2 metrics also could be used directly to compare relatively intratumoral genetic heterogeneity for cases with multiregional sampling.

Clone Tree Construction

To construct clone tree for cases with progressive recurrences, we applied both CITUP³⁶ and PyClone (v 0.13.1).³⁷ We first used PyClone to perform clustering analysis of cancer cell fractions from each mutation with default settings. The mutation clusters were then used to construct phylogenetic trees through CITUP, the outputs of which were also manually checked to ensure no violation of evolutionary conflicts mentioned previously.

Mutational Signature Analysis

Mutational signatures analysis was performed by the MutationalPatterns package³⁸ in R. Mutational signatures described in liver cancers^{39–41} and other signatures with marked contributions were considered, including signatures 3, 4, 5, 6, 8, 12, 16, 17, 18, 22, 24, and 29. Mutational signatures analysis was applied for samples with whole-genome sequence data.

Cell Culture

Huh7, PLC/RPF/5, and HEK293FT were obtained from the American Type Culture Collection (Manassas, VA) and

maintained in Dulbecco's modified Eagle's complete medium. The in-house established human HCC cell lines HKCI-8, HKCI-9, HKCI-10, and HKCI-11 were cultured in AIM-V medium (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum. Immortalized hepatocyte cell lines, MIHA and L02, were propagated in Dulbecco's modified Eagle's complete medium. Stable cell lines, stably expressing wild-type (WT) and mutant forms of STAT3 and JAK1, were created⁴² and maintained in complete medium containing 1 $\mu\text{g}/\text{mL}$ puromycin. In brief, cells were transduced with lentivirus, which was generated by transfection of HEK293FT cells with pLenti-STAT3 (WT or mutants), pLenti-JAK1 (WT or mutants) or pLenti-puro empty vector. After incubation for 24 hours, cells were selected using 2 $\mu\text{g}/\text{mL}$ puromycin (Sigma Merck KGaA, Darmstadt, Hesse, Germany) for at least 10 days. Transduction efficiency was checked by Western and quantitative reverse-transcribed PCR (RT-PCR).

RT-PCR, ChIP-qPCR, and ChIP-sequencing

RNA was isolated by RNeasy Mini Kit (Qiagen, Hilden, Germany). Purified RNA was subjected to DNase I treatment and reverse transcribed into complementary DNA (Thermo Fisher Scientific). Real-time PCR was conducted using standard procedures on StepOnePlus Real-Time PCR System (ABI). ChIP was performed following the study of Zhu et al⁴³ with adjustments. Briefly, cells were crosslinked in 1% formaldehyde (Thermo Fisher Scientific) for 10 minutes at room temperature. After neutralizing by glycine, cells were harvested and lysed in Farnham lysis buffer (5 mM HEPES, pH 8.0, 85 mM KCl, 0.5% NP-40) with 1X Protease Inhibitor (Thermo), and then sheared by sonication. The supernatant from lysates was subjected to immunoprecipitation with corresponding antibodies at 4 °C overnight and pulled down using magnetic beads. Immunoprecipitates were washed and reverse-crosslinked by incubation at 65 °C overnight. DNA was further purified by QIAquick PCR Purification Kit (Qiagen). Following ChIP, qPCR was conducted by SYBR green-based detection method using equal amounts of ChIP DNA. For ChIP-sequencing, libraries were prepared with KAPA Hyper Library Prep Kit (Kapa Biosystems), and then sequenced on Illumina HiSeq 2000.

BrdU Cell Proliferation Assay

Cell replication was measured using a Cell Proliferation ELISA kit (Roche). Briefly, 1000 cells were seeded in a 96-well plate with presence or absence of MI-2-2 or OICR-9429 drug treatment. For every 2 days, cells were incubated with 10 μM BrdU for 2 hours at 37 °C and fixed by FixDenat buffer. Newly synthesized cellular DNA was detected using an anti-BrdU peroxidase-conjugated monoclonal antibody and visualized by the subsequent substrate reaction. Optical density was quantified by measuring the absorbance at 450 nm using a scanning multiwell spectrophotometer.

Cloning

Full-length *TP53*, *STAT3*, and *JAK1* ORFs were amplified from MIHA cells and cloned into pcDNA3.1 or pLenti-puro vectors, respectively. For promoter luciferase constructs, up to 3 kb upstream of *MLL1* and *MLL2* were cloned into pGL3-basic vector. Mutants of *TP53*, *STAT3*, and *JAK1* were generated from WT constructs by QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies). All sequences were verified by Sanger sequencing. Primers are listed in [Supplementary Table 7](#).

Western Blot

After protein quantitation, 20- μg protein lysates were resolved on 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis and electrophoretically transferred onto nitrocellulose membranes. Specific primary antibodies used in this study were as follows: p53 (DO-1, Santa Cruz Biotechnology, Dallas, TX), H3K27ac (Abcam, Cambridge, MA), JAK1 (Cell Signaling, Danvers, MA), STAT3 (Cell Signaling), p-JAK1 (Tyr1022/1023) (Cell Signaling), p-STAT3 (Tyr705) (Cell Signaling), and GAPDH (Millipore, Bedford, MA).

Luciferase

STAT3 luciferase reporter, a luciferase construct under the control of multiple STAT3 consensus binding sites, was purchased from Qiagen. For luciferase assay, L02 and HKCI-9 were co-transfected with STAT3 luciferase reporter and STAT3 or JAK1 (WT or mutants) expression vectors. Two days after transfection, cells were treated or untreated with interleukin-6 for 3 hours. Then cells were lysed and the activities of firefly and Renilla luciferases were determined using Dual-Luciferase Reporter Assay System (Promega, Madison, WI). For STAT3 inhibitor, cells were treated with 250 μM S31-201 or dimethyl sulfoxide for 24 hours before measuring luciferase activities. To determine the transcription activity of p53, *MLL1* or *MLL2* promoter constructs were co-transfected with expression plasmid for WT or mutant p53 in HEK293FT cells. Relative luciferase activities were calculated by normalizing firefly luciferase activities to the Renilla luciferase activities in each cell lysate.

Statistical Analysis

Statistical analysis was performed in R (v3.0.2) and GraphPad (La Jolla, CA) Prism (v6.05). Categorical data were compared with Fisher Exact test or χ^2 test and continuous data were compared with Student *t* test or Mann-Whitney *U* test when appropriate. Survival analysis was performed by the Kaplan-Meier method and compared by the log-rank test. Cox regression model was then applied to obtain hazard ratios together with their corresponding 95% confidence intervals. Odds ratio for recurrence was calculated based on the occurrence of mutated drivers across initial tumors with progressive-HCC (pHCC) recurrence and their occurrence across initial tumors without recurrence.

Supplementary References

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25:1754–1760.
2. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010; 26:589–595.
3. **Li H, Handsaker B, Wysoker A, et al.** The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
4. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43:491–498.
5. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; 31:213–219.
6. **Rimmer A, Phan H, Mathieson I, et al.** Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet* 2014;46:912–918.
7. Saunders CT, Wong WS, Swamy S, et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 2012; 28:1811–1817.
8. Narzisi G, O’Rawe JA, Iossifov I, et al. Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nat Methods* 2014;11:1033–1036.
9. 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
10. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* 2011;39:D945–D950.
11. Cingolani P, Patel VM, Coon M, et al. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program. *Snpsift*. *Front Genet* 2012;3:35.
12. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6:80–92.
13. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; 5:e1000529.
14. **Nik-Zainal S, Van Loo P, Wedge DC, et al.** The life history of 21 breast cancers. *Cell* 2012;149:994–1007.
15. Boeva V, Zinovyev A, Bleakley K, et al. Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics* 2011;27:268–269.
16. Boeva V, Popova T, Bleakley K, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 2012; 28:423–425.
17. Rasmussen M, Sundstrom M, Goransson Kultima H, et al. Allele-specific copy number analysis of tumor samples with aneuploidy and tumor heterogeneity. *Genome Biol* 2011;12:R108.
18. Quinlan AR. BEDTools: The Swiss-Army tool for genome feature analysis. *Curr Protoc Bioinformatics* 2014;47. 11.12.1–34.
19. Olshen AB, Venkatraman ES, Lucito R, et al. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004;5:557–572.
20. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 2012;30:413–421.
21. **Jamal-Hanjani M, Wilson GA, McGranahan N, et al.** Tracking the evolution of non-small-cell lung cancer. *N Engl J Med* 2017;376:2109–2121.
22. **Landau DA, Carter SL, Stojanov P, et al.** Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 2013;152:714–726.
23. Bolli N, Avet-Loiseau H, Wedge DC, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun* 2014;5:2997.
24. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;30:1363–1369.
25. Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* 2017;169:1327–1341.e23.
26. **Villanueva A, Portela A, Sayols S, et al.** DNA methylation-based prognosis and epdrivers in hepatocellular carcinoma. *Hepatology* 2015;61:1945–1956.
27. Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 2004;20:289–290.
28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.
29. **Heinz S, Benner C, Spann N, et al.** Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–589.
30. **Huang da W, Sherman BT, Lempicki RA.** Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.
31. **Huang da W, Sherman BT, Lempicki RA.** Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37:1–13.
32. Sottoriva A, Kang H, Ma Z, et al. A Big Bang model of human colorectal tumor growth. *Nat Genet* 2015; 47:209–216.
33. Waclaw B, Bozic I, Pittman ME, et al. A spatial model predicts that dispersal and cell turnover limit intra-tumour heterogeneity. *Nature* 2015;525:261–264.
34. **Sun R, Hu Z, Sottoriva A, et al.** Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat Genet* 2017;49:1015–1024.
35. **Lee JK, Wang J, Sa JK, et al.** Spatiotemporal genomic architecture informs precision oncology in glioblastoma. *Nat Genet* 2017;49:594–599.

36. **Malikic S, McPherson AW, Donmez N**, et al. Clonality inference in multiple tumor samples using phylogeny. *Bioinformatics* 2015;31:1349–1356.
 37. Roth A, Khattra J, Yap D, et al. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* 2014;11:396–398.
 38. Blokzijl F, Janssen R, van Boxtel R, et al. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* 2018;10:33.
 39. **Letouze E, Shinde J**, Renault V, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun* 2017;8:1315.
 40. **Schulze K, Imbeaud S, Letouze E**, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* 2015;47:505–511.
 41. **Alexandrov LB**, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415–421.
 42. Horn PA, Keyser KA, Peterson LJ, et al. Efficient lentiviral gene transfer to canine repopulating cells using an overnight transduction protocol. *Blood* 2004;103:3710–3716.
 43. Zhu J, Sammons MA, Donahue G, et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature* 2015;525:206–211.
-

Author names in bold designate shared co-first authorship