
Supplementary information

**Deep whole-genome analysis of 494
hepatocellular carcinomas**

In the format provided by the
authors and unedited

Supplementary Information for

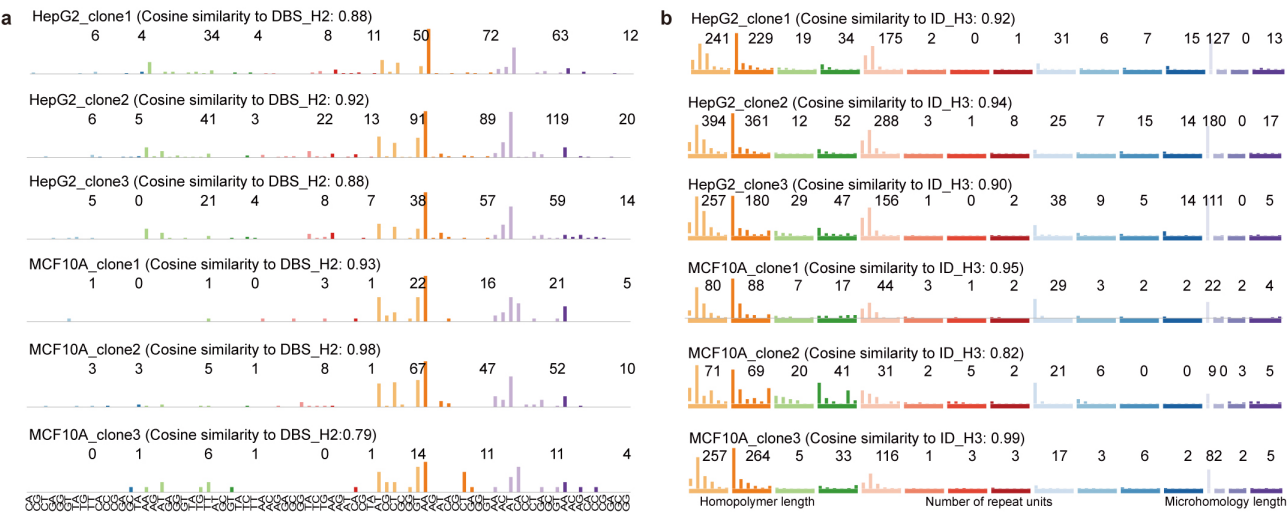
Deep Whole Genome Analysis of 494 Hepatocellular Carcinomas

Lei Chen^{1,13,*}, Chong Zhang^{2,13}, Ruidong Xue^{3,13}, Mo Liu^{4,13}, Jian Bai^{5,13}, Jinxia Bao^{6,13}, Yin Wang^{5,13}, Nanhai Jiang⁴, Zhixuan Li¹, Wenwen Wang⁷, Ruiru Wang⁵, Bo Zheng^{1,7}, Airong Yang⁵, Ji Hu^{1,7}, Ke Liu⁵, Siyun Shen^{1,7}, Yangqianwen Zhang¹, Mixue Bai¹, Yan Wang⁵, Yanjing Zhu^{1,7}, Shuai Yang^{1,7}, Qiang Gao⁸, Jin Gu⁹, Dong Gao¹⁰, Xin Wei Wang¹¹, Hidewaki Nakagawa¹², Ning Zhang³, Lin Wu^{5,*}, Steven G. Rozen^{4,*}, Fan Bai^{2,*}, and Hongyang Wang^{1,*}

Table of contents

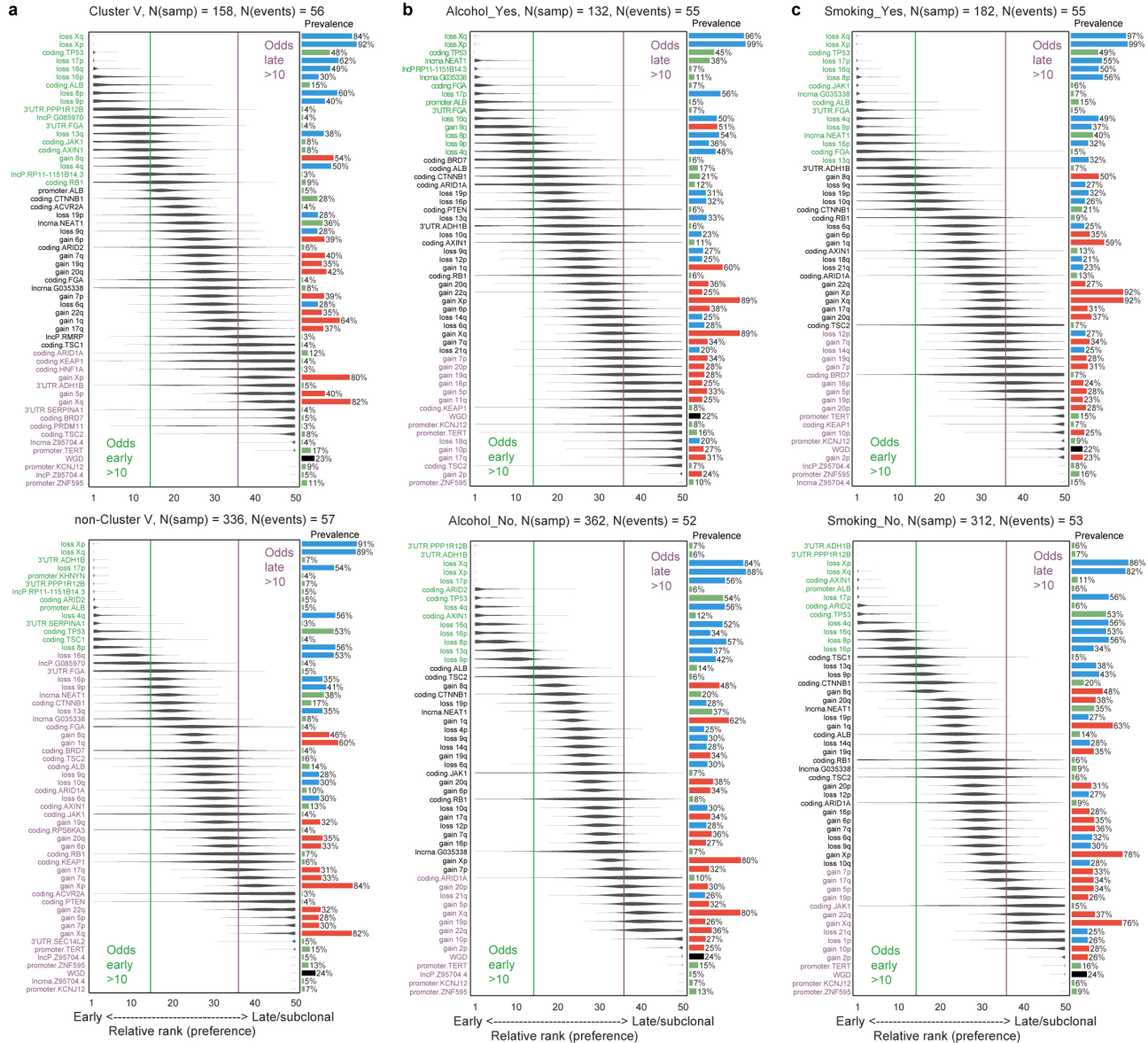
Supplementary Figures.....	2
Supplementary Fig. 1 Signature analysis of cancer cell lines after AA exposure.	2
Supplementary Fig. 2 Evolutionary history of tumors in different groups.	3
Supplementary Fig. 3 Source gel data 1	4
Supplementary Fig. 4 Source gel data 2	5
Supplementary Fig. 5 Source gel data 3	6
Supplementary Notes.....	7
Supplementary Note 1 Study background and Patient cohort	7
Supplementary Note 2 Candidate coding and noncoding drivers.....	8
Supplementary Note 3 Mutational signatures.....	9
Supplementary Note 4 SBS_H8 is linked with poor prognosis	10
Supplementary Note 5 Genomic rearrangements	11
Supplementary Note 6 Evolutionary history.....	12
Supplementary Note 7 Noncoding mutations are linked to dysregulation of liver metabolic pathway.	14
Supplementary Note 8 Functional validation of candidate noncoding drivers	15
Supplementary Discussion	18
Supplementary References	20

Supplementary Figures



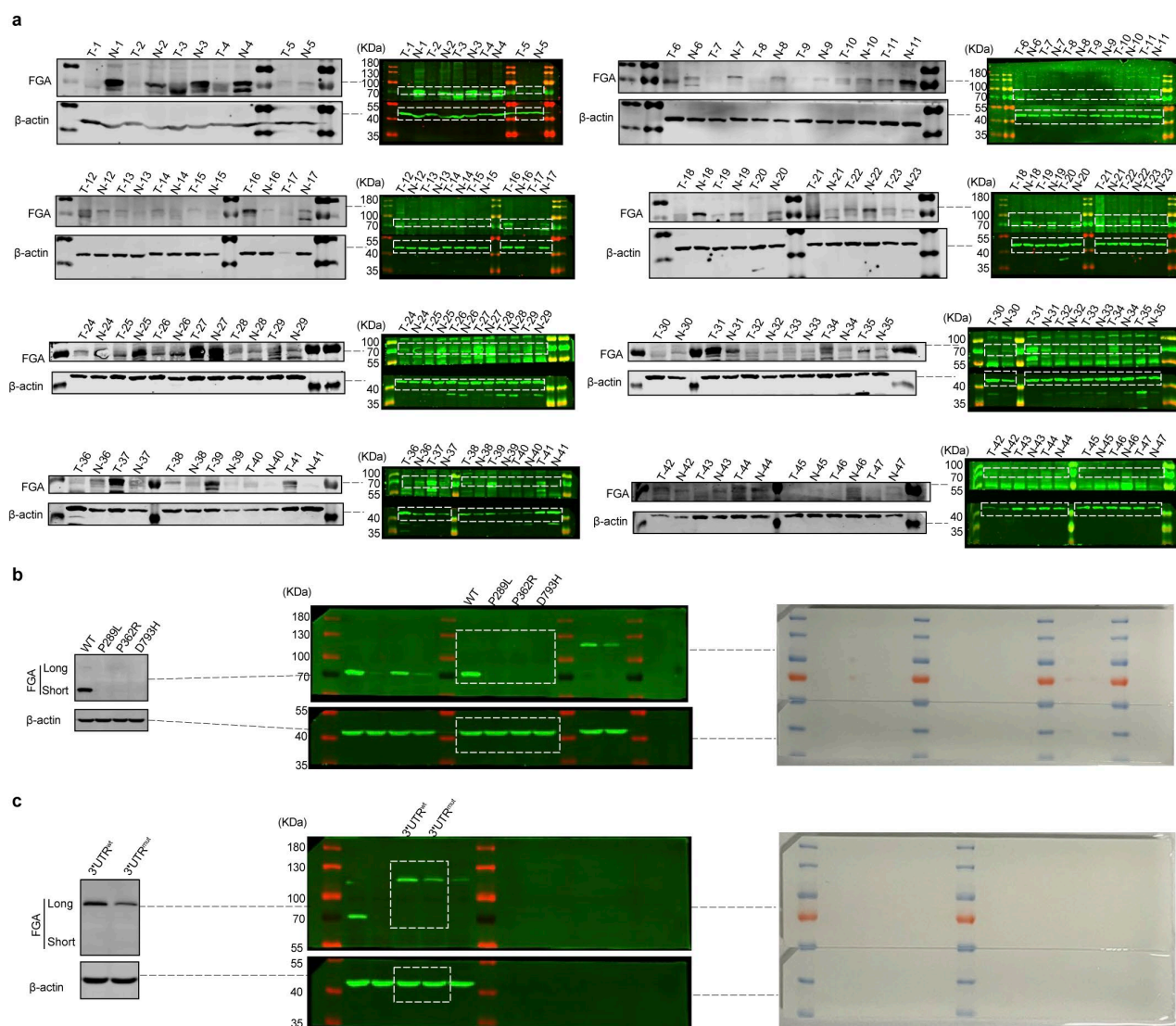
Supplementary Fig. 1 | Signature analysis of cancer cell lines after AA exposure.

a-b, DBS and ID mutation spectrum of six single cell clones from 2 cancer cell lines treated with AA exposure. The cosine similarity between each spectrum and DBS_H2 (**a**) or ID_H3 (**b**) is indicated.



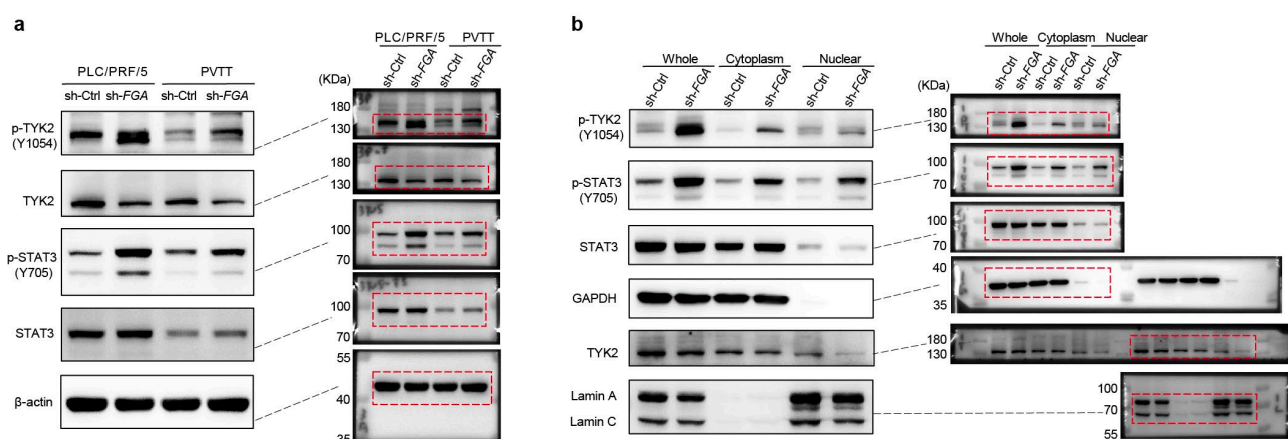
Supplementary Fig. 2 | Evolutionary history of tumors in different groups.

a-c, Preferential ordering diagrams for patients stratified based on Cluster V (**a**), alcohol (**b**) and smoking (**c**). The relative ordering of candidate drivers was compared (**Extended Data Fig. 7f**).



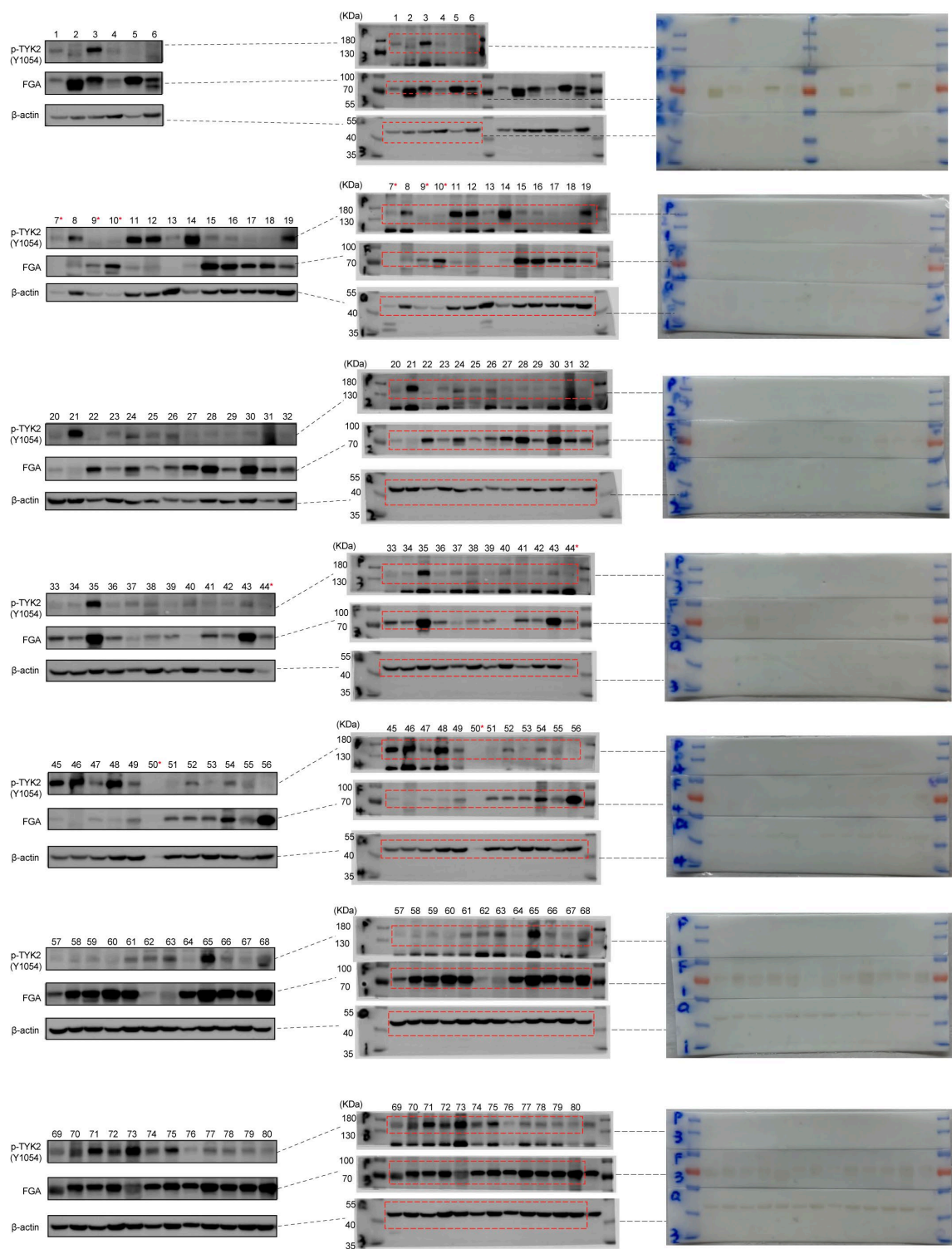
Supplementary Fig. 3 | Source gel data 1

a, Source gel images for **Fig. 5c** showing FGA protein levels in paired tumour and normal samples ($n = 47$). Samples derive from the same experiment and gels were processed in parallel. **b**, Source gel images for **Fig. 5g** showing FGA protein levels across wild-type (WT) and point-mutated (P289L, P362R and D793H) HepG2 cell lines in the *FGA* coding regions. β -actin was run on the same gel. **c**, Source gel images for **Extended Data Fig. 10f** showing FGA protein levels across wild-type (3'UTR^{wt}) and 3'UTR-mutated (3'UTR^{mut}) HepG2 cell lines in the *FGA* noncoding regions. β -actin was run on the same gel. Proteins were detected by western blotting, using the indicated antibodies and the white rectangles show the cropping location.



Supplementary Fig. 4 | Source gel data 2

a, Source gel images for **Extended Data Fig. 10n** showing p-TYK2 (Y1054) and p-STAT3 (Y705) protein levels in PLC/PRF/5 and PVTT cell lines following *FGA* knockdown. p-, phosphorylated. **b**, Source gel images for **Fig. 5l** showing subcellular localization of p-TYK2 (Y1054) and p-STAT3 (Y705) in PLC/PRF/5 cells after *FGA* knockdown. GAPDH and Lamin A/C were utilized as the cytoplasmic and nuclear reference proteins, respectively. Proteins were detected by the indicated antibodies and the red rectangles show the cropping location.



Supplementary Fig. 5 | Source gel data 3

Source gel images for **Fig. 5n** and **Extended Data Fig. 10r** showing FGA protein and TYK2 phosphorylation levels in an independent HCC patient cohort ($n = 75$). “*” denotes five samples removed due to low β -actin signals. Proteins were detected by western blotting, using the indicated antibodies and the red rectangles show the cropping location. Samples derive from the same experiment and gels were processed in parallel.

Supplementary Notes

Supplementary Note 1 | Study background and Patient cohort

Primary liver cancer (PLC) is the third leading cause of cancer-related mortality worldwide¹. Hepatocellular carcinoma (HCC) accounts for over 80% of PLC incidence, half of which is in China². The predisposing factors for HCC exhibit extensive regional differences. Hepatitis B virus (HBV) and aflatoxin are the major risk factors for Chinese HCC, while hepatitis C virus (HCV) is more prevalent in Japanese HCC. For Western populations, alcohol abuse, non-alcoholic steatohepatitis (NASH), and metabolic disorder are the major HCC risk factors³. Recently, the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium of the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) have analyzed the genomic complexity of cancer at unprecedented scale⁴, covering noncoding drivers⁵, mutational signatures⁶, structural variations⁷, and evolutionary histories⁸. Nevertheless, the HCC data of PCAWG (PCAWG-HCC) were mainly from Japanese⁹ and some from European populations¹⁰. Given the great disparity of etiological factors and related differences in carcinogenesis across different populations, a comprehensive whole genome investigation of HBV-enriched Chinese HCC is needed.

Our cohort represent the epidemiology of the Chinese liver cancer population, for instance, the high gender bias¹¹⁻¹³. According to Global Cancer Statistics 2020, rates of both incidence and mortality are 2 to 3 times higher among men than among women in most regions¹. This bias is even more dramatic in China. According to a large-scale multicenter retrospective investigation of 18,275 HCC cases of hepatectomy in China, the male accounts for 84% of all patients¹¹. Our data reporting 85% of men in CLCA is highly consistent with this report. Recently, it is reported that loss of chromosome Y (LOY) also plays an important role in the high risk of cancer for men^{14,15}. In our CLCA cohort, 59.2% of cases (253/427) showed a LOY event, which may also explain this high gender disparity of liver cancer.

Supplementary Note 2 | Candidate coding and noncoding drivers

We explored potential coding drivers by combining three independent methods, MutSigCV¹⁶, dndsCV¹⁷, and OncodriveFML¹⁸. We identified 23 candidate coding drivers (Fig. 1b), including *TP53*, *CTNNB1*, and *ALB*. Compared with the American-enriched TCGA-HCC and Japanese-enriched PCAWG-HCC cohort⁵, only 10 genes were called as coding drivers in all three cohorts (Extended Data Fig. 1d-g). Excluding *TSC2* and *JAK1* that had been reported in previous exome studies of Chinese HCC^{19,20}, six novel candidate coding drivers were identified in CLCA, including *FGA* (4.5%), *HNFI1A* (2.8%), *PRDM11* (2.8%), *CDKN1B* (2.0%), *BMP5* (1.6%), and *ECHS1* (0.4%). The mutational frequency of *TP53* is significantly higher in CLCA compared with either PCAWG-HCC (51% vs. 31%, $P = 2.4\text{e-}08$, $Q < 0.1$) or TCGA-HCC (51% vs. 31%, $P = 1.6\text{e-}04$, $Q < 0.1$).

We explored potential noncoding drivers by combining four methods: MutSigCV-NC²¹, NBR⁵, ActiveDriverWGS²², and OncodriveFML¹⁸. A total of 31 candidate noncoding drivers were identified, including six promoters, eight lncRNAs, four lncRNA promoters, five 5'-untranslated regions (UTRs), and eight 3'UTRs (Fig. 1b). Five genes were determined as driver events in different forms, highlighting the important roles of these genes in HCC development and indicating convergent evolution. *ALB*, encoding albumin, was determined as both coding and noncoding (promoter) drivers. Mutations in both coding regions and 3'UTRs of *ALB* have been reported in HCC, and our study identified promoter mutation as a new form of driver for *ALB*. *FGA*, encoding fibrinogen alpha chain, was also determined as both candidate coding and noncoding (3'UTR) drivers. Further, mutations in the 3'UTR of *FGA* were not enriched for 2-5bp indels ($Q = 0.73$), suggesting that these mutations were not related to the transcription-associated indel signature enriched in 3'UTR and downstream of lineage-related genes²³. *Z95704.4*, *G085970*, and *RMRP* were determined as drivers in both categories of lncRNA and lncRNA promoter.

Surprisingly, only three (9.7%) noncoding drivers were reported by PCAWG-HCC, including *TERT* promoter, lncRNA *NEAT1*, and lncRNA promoter of *RMRP*. As a well-established noncoding

driver of HCC, *TERT* promoter accounted for 35% of CLCA patients, significantly lower than that of PCAWG-HCC cohort (54.3%, $P = 7.9\text{e-}07$) and TCGA-HCC cohort (44%, $P = 0.0237$) ([Extended Data Fig. 1e-g](#)). LncRNA *NEAT1* mutations were first identified in the HCV-enriched Japanese population⁹, but later were considered as artefacts by the re-analysis of PCAWG⁵. Conversely, several reports supported the oncogenic role of lncRNA *NEAT1* mutations²⁴. LncRNA *NEAT1* was mutated in 37% of CLCA cases, comparable to that of 30% in PCAWG-HCC. Whether lncRNA *NEAT1* mutations contribute to HCC development merits further investigations. Considering that three noncoding drivers reported by PCAWG-HCC were successfully recapitulated by our cohort, particularly the low frequency driver of *RMRP* lncRNA promoter, these results suggested the robustness of our identification of noncoding drivers. All other 28 events (90.3%), including all the eight 3'UTRs and five 5'UTRs, were new candidate noncoding drivers not reported in HCC before.

Supplementary Note 3 | Mutational signatures

To explore the mutational processes in CLCA, we extracted mutational signatures of single base substitutions (SBS), doublet base substitutions (DBS), and small insertions and deletions (ID). In total, 17 SBS (SBS_H1-H17), 3 DBS (DBS_H1-H3) and 8 ID (ID_H1-H8) signatures supported by both the SigProfiler and mSigHdp algorithms were detected ([Extended Data Fig. 1j](#), [Extended Data Fig. 2](#), and [Extended Data Fig. 3a](#)). Comparing with the COSMICv3.2 signature set, 23 signatures were known and five signatures were novel ([Supplementary Table 3](#)). Among the known signatures, SBS_H1, H4, H14, H16 and H17 were found to be splits of COSMIC SBS5 and SBS40 (two aging related mutational processes), SBS_H2 to be COSMIC SBS22 (related to aristolochic acid (AA) exposure), and SBS_H3 to be COSMIC SBS24 (related to aflatoxin exposure) ([Extended Data Fig. 2](#)). A full description of the mapping between CLCA signatures and COSMIC signatures was provided in [Supplementary Table 3](#).

The five novel signatures contained one SBS signature, SBS_H8, two DBS signatures, DBS_H1 and DBS_H2, and two ID signatures, ID_H3 and ID_H8. DBS_H1 signature consisted mainly of

CC>NN, GC>NN and TC>NN mutations ([Extended Data Fig. 3d](#)). This signature was found in most tumors and correlated with age as well as other age-related signatures. ID_H8 showed mostly 1bp cytosine deletions and thymine insertions. It was exclusively found in SBS_H3 (COSMIC SBS24) positive tumors and correlated with SBS24 (Spearman $\rho = 0.90$, $P = 1.36 \times 10^{-22}$), showing that this signature is related to aflatoxin exposure.

Supplementary Note 4 | SBS_H8 is linked with poor prognosis

To estimate the number of mutations each signature accounted for in each tumor, we analyzed signature attributions with mSigAct on CLCA ([Extended Data Fig. 3h](#); [Supplementary Table 3](#)). For simplification, we combined SBS_H1, SBS_H4, SBS_H14, SBS_H16, and SBS_H17, which were similar to or splits of aging-related COSMIC signatures SBS1, SBS5, and SBS40, as the “Aging” SBS signature ([Extended Data Fig. 4a-d](#)).

Unsupervised hierarchical clustering based on mutational signatures classified 494 tumors into 5 clusters ([Extended Data Fig. 3h](#)). Cluster I was characterized by high levels of AA-related signatures, SBS_H2, DBS_H2, and ID_H3, suggesting that these tumors were associated with AA mutagenesis. Interestingly, tumors from cluster I were enriched for non-smokers, indicating that the two exogenous mutagenic factors, AA and smoking, might be mutually exclusive in our cohort. Cluster II tumors were dominated by aging-related signatures. Cluster III was characterized by the presence of the aflatoxin-related signatures, SBS_H3 and ID_H8. The enrichment for *TP53* mutations in this group was consistent with a previous report²⁵. Cluster IV’s ID signature was dominated by ID_H1, which is related to slippage during DNA replication. This group featured a high recurrence rate and *AXINI* mutations. Cluster V had strong contributions from SBS_H8 and featured *CTNNB1* mutations ([Extended Data Fig. 3h-j](#)). Of note, higher percentages of SBS_H8 were significantly associated with poorer prognosis across CLCA ([Extended Data Fig. 3k](#), [Extended Data Fig. 5a](#)), demonstrating that the underlying risk factor related to SBS_H8 was closely associated with worse survival for HCC patients. Collectively, these results showed that mutational signatures can stratify CLCA patients into distinct subgroups associated with different

genomic features and prognosis.

We summarized the enrichment of mutational signatures in clonal versus subclonal mutations ([Extended Data Fig. 4h](#)). Strikingly, most mutational signatures related to exogenous factors, such as SBS_H2 (AA), SBS_H3 (aflatoxin), SBS_H8, DBS_H2 (AA), ID_H3 (AA), SBS_H10 (tobacco) and ID_H8 (aflatoxin), were enriched in clonal mutations, suggesting that they occurred at earlier stages of tumorigenesis. In contrast, subclonal mutations were more likely driven by mutational signatures related to endogenous factors, including SBS_H1 (aging), SBS_H17 (aging), ID_H7 (defective DNA damage repair), SBS_H13 (aging), SBS_H9 (APOBEC), and SBS_H4 (aging), suggesting that they occurred in later stages of tumorigenesis.

Supplementary Note 5 | Genomic rearrangements

Our deep WGS data enabled a comprehensive profiling of genomic rearrangements in CLCA, including copy number alterations (CNAs), structural variations (SVs), HBV integrations, ecDNAs, and three forms of clustered alterations, kataegis, chromothripsis, and chromoplexy ([Extended Data Fig. 5b-d](#), [Extended Data Fig. 6](#)). GISTIC analysis was performed to identify recurrent CNAs in our cohort ([Extended Data Fig. 5b](#); [Supplementary Table 4](#)). We found 36 focally amplified regions, including 5p15.33, 11q13.3, 17q25.3, and 8q24.3, and 53 focally deleted regions, including 3p12.3, 4q22.1, 11q23.2, and 12p13.2. The broad CNA profile of the CLCA is largely consistent with previous Chinese HCC cohorts^{20,26}, but quite different from that of PCAWG-HCC cohort, sharing only 9 amplified regions and 6 deleted regions ([Supplementary Table 4](#)).

Next, we explored HBV integrations in our cohort. A total of 1,719 HBV integrations were identified in 339 of 467 (72.6%) HBV-associated HCC cases. 42 of these HBV integrations were also identified by the RNA-seq data ([Extended Data Fig. 5c](#), [Supplementary Table 4](#)). We identified 13 recurrent genes integrated by HBV, including *TERT* (21.4% of 467 cases, n = 100), *KMT2B* (5.4%, n = 25), and *CCNA2* (1.5%, n = 7)^{27,28}. Three miRNA genes, *MIR4457* (n = 6), *MIR4435-2* (n = 3) and *MIR4315-1* (n = 3), were found to be recurrently inserted by HBV, consistent with the previous report that HBV can integrate microRNA genes and regulate their expression²⁹. Five genes

with recurrent HBV integration, *AUTS2*, *NAV2*, *LOC100288778*, *MYOCD*, and *ZBTB32*, were newly found in the CLCA cohort. Intriguingly, most integration breakpoints were located near the 3' end of the *HBx* gene (position: 1,374-1,838), suggesting that this gene plays an important role in HBV-related hepatocarcinogenesis (Extended Data Fig. 5c).

ecDNA has been reported to contribute to virus-related gene amplification³⁰. Whether such circular structure can be exploited by HBV remains poorly defined. We first identified ecDNAs incorporating HBV segments in seven patients (Fig. 3d-f). Interestingly, HBV segments in ecDNA showed an elevated number of copies, as well as increased expression levels, in comparison with those not in ecDNA (Fig. 3d-e). These results showed that HBV integrations in HCC may reach a high copy number via ecDNA amplification. Many well-known oncogenes were also identified in HBV-ecDNAs, including *TERT* and *GATA3* (Fig. 3f). Despite that HBV-*TERT* integration has been identified in HCC in previous studies, our results demonstrated, for the first time, that these integrations can exploit the circular structure of ecDNA and thus amplify to hundreds of copies. The existence of ecDNA were successfully validated by CIRCLE-seq. For example, sequencing reads indicative of ecDNA containing *TERT* and multiple chromosome segments were detected in CLCA_0203 (Fig. 3g).

Supplementary Note 6 | Evolutionary history

Reconstruction of the evolutionary history of CLCA categorized 19.48% point mutations being early clonal, 3.64% late clonal, 31.90% unspecified clonal, and 44.98% subclonal (Extended Data Fig. 7b; Supplementary Table 5). In PCAWG-HCC, cancer drivers were mostly clonal with only 8% being subclonal. By contrast, we found that more drivers were subclonal, potentially owing to the high sequencing depth of CLCA. Among the 23 candidate coding drivers, 926 oncogenic point mutations were identified, of which 33.69% were early clonal, 2.8% late clonal, 51.73% unspecified clonal and 11.23% subclonal (Extended Data Fig. 7c). As to the 28 candidate noncoding driver genes, 955 oncogenic point mutations were identified, of which 27.23% were early clonal, 3.35% late clonal, 44.82% unspecified clonal and 24.61% subclonal. In comparison

with candidate coding drivers, candidate noncoding drivers were more enriched in the subclonal category ($P = 1.4 \times 10^{-12}$, Chi-square test), implying that candidate noncoding drivers may contribute more to the subclonal diversification ([Extended Data Fig. 7c](#)).

To estimate the relative timing of driver events, including mutations and CNAs, we aggregated them across samples and defined a probabilistic ranking ([Extended Data Fig. 7d](#)), reflecting whether these events occur preferentially early or late during tumour evolution. In general, losses typically preceded gains, whole-genome duplications (WGD) occurred relatively late, and noncoding drivers could arise at various stages of the evolutionary history of HCC. The earliest events were *PPP1R12B* 3'UTR mutation and 17p loss, followed by mutations in *TP53*, *ARID2* and the *ADH1B* 3'UTR. In contrast, *TP53* mutation was found to be the earliest mutational event in the PCAWG-HCC and PCAWG pan-cancer cohorts⁸. Our results showed that noncoding mutations may be the earliest mutational events during cancer progression. Surprisingly, *TERT* promoter mutations were among the latest events, including mutations in lncRNA promoter *Z95704.4*, promoter *KCNJ12* and *ZNF595*. This result was distinct to the observation that *TERT* promoter was an early event in European HCC¹⁰. Different forms of alterations in the same genes, including *ALB*, *FGA*, and *Z95704.4*, were present at different evolutionary stages. These results revealed the distinct evolutionary history of the Chinese CLCA HCC cohort in comparison with the PCAWG-HCC cohort, and highlighted the pervasive contributions of noncoding mutations during HCC progression. In addition, the SBS signatures related to tobacco, aflatoxin and AA exposure (SBS_H10, SBS_H3, and SBS_H2), as well as the novel signature SBS_H8, tended to occur early across all 494 cases ([Extended Data Fig. 7e](#)), consistent with the results shown in [Extended Data Fig. 4h](#).

To explore the potential association between evolutionary history and etiology, we stratified CLCA patients based on Cluster V enriched with SBS_H8, alcohol, and smoking ([Extended Data Fig. 7f](#), [Supplementary Fig. 2](#), and [Supplementary Table 5](#)). The comparisons between each group showed distinct evolutionary histories, as exemplified by the different rankings of potential drivers. For instance, 14 candidate drivers were assigned to different evolutionary stages when comparing

the evolutionary history of Cluster V with that of the non-Cluster V. Interestingly, *FGA* mutations were among the earliest drivers in Cluster V patients, alcohol drinking patients and smoking patients.

Supplementary Note 7 | Noncoding mutations are linked to dysregulation of liver metabolic pathway

To systematically evaluate the contributions of different types of somatic alterations across CLCA, we grouped 84 potential driver genes into eight signaling pathways, including cell cycle, RTK-RAS-MAPK, PI3K-PTEN-mTOR, Wnt/ β -catenin, telomere maintenance, JAK-STAT, epigenetic regulation and metabolism (Extended Data Fig. 8; Supplementary Table 6). Eight forms of somatic alterations were assessed, including coding SNVs, noncoding SNVs (further divided into promoters, lncRNAs and UTRs), CNAs, ecDNAs, SVs and HBV integrations. Most of the well-known driver genes in HCC were disrupted by alterations of multiple forms, highlighting the diversity of the dysregulation mechanisms exploited by cancer cells. Notably, six types of alterations were identified for both *TERT* and *PBX1*. Various types of *TERT* alterations including point mutations and HBV integrations in its promoter region have been reported in HCC³¹. *PBX1*, encoding pre-B-cell leukemia transcription factor 1, is a proto-oncogene acting as a *NOTCH* target and a transcriptional repressor of *CDKN2B*³². *PBX1* has been reported as a driver gene in multiple cancer types, but it has not yet been identified in HCC.

To further weigh the relative contributions of coding SNVs and noncoding SNVs across these signaling pathways, we calculated the percentages of cases altered by coding SNVs only, by noncoding SNVs only, and by both categories for each pathway. Intriguingly, we identified three signaling pathways that were disrupted by noncoding SNVs more frequently than by coding SNVs, including RTK-RAS-MAPK (22.1% vs. 6.5%), telomere maintenance (34% vs. 1.4%) and liver metabolism (23.1% vs. 18.2%). Since most targeted drugs for HCC, like Sorafenib³³ and Lenvatinib³⁴, are multi-target tyrosine kinase inhibitors, which target RTKs in addition to VEGFR2, the prevalence of noncoding mutations in the RTK-RAS-MAPK pathway implied that these

previously unrecognized noncoding mutations might be associated with therapeutic responses. In the tumour metabolism pathway, a total of 15 potential driver genes were identified. These alterations affected various metabolic programs, including hepatic metabolism (*APOB*, *ALB* and *HNF1A*), oxidative stress (*KEAP1* and *NFE2L2*), urea metabolism (*CPS1*), alcohol metabolism (*ADH1B* and *ADH4*), fatty acid metabolism (*SERPINA1* and *SERBP1*), and hypoxia (*ARNT*). *FGA* in the JAK-STAT pathway also plays a role in hepatic metabolism. Many of these genes have not been identified as drivers in previous exome sequencing studies of HCC. Given that the liver is a key metabolic organ and metabolism dysregulation is an important feature of liver cancer³⁵, this result suggested that noncoding alterations may play an important role in HCC development, and underlined the necessity of adopting WGS to investigate the metabolic status of HCC. This result is in agreement with a recent WGS analysis of chronic liver disease, which showed that metabolic pathways are a frequent target of convergent somatic mutation³⁶.

A total of 14 potential driver genes were not classified into any of the above signaling pathways. A very high percentage (52.2%) of CLCA HCC patients were affected by noncoding mutations in genes of this unclassified category. This result implies that further efforts are needed to better understand the important roles of noncoding SNVs in HCCs.

Supplementary Note 8 | Functional validation of candidate noncoding drivers

To investigate whether the candidate noncoding drivers we identified have tumorigenic functions, we selected three representative candidate genes to perform functional assays, including *KCNJ12*, *PPP1R12B* (Extended Data Fig. 9), and *FGA* (Fig. 5 and Extended Data Fig. 10). Knockdown by short hairpin RNA (shRNA) or knockout by short guide RNA (sgRNA) for the three drivers were constructed on a total of seven human cell lines, including six liver cancer cell lines (PLC/PRF/5, PVTT, SNU387, SNU182, Huh7, and HepG2) and one normal liver cell line (HHL5).

For the potential tumour suppressor *PPP1R12B*, disrupted cell lines were successfully constructed both by knockdown (PLC/PRF/5, PVTT, SNU387, and HHL5) and by knockout

(HepG2 and Huh7). Consistent results observed between each two independent disrupted cell lines showed that low expression of *PPP1R12B* significantly enhanced tumour migration, invasion, self-renewal, and cell proliferation ([Extended Data Fig. 9a](#)). To further explore whether point mutations of *PPP1R12B* identified in our CLCA cohort are able to cause phenotypic changes, we knocked in the mutations into the HepG2 cell line by the CRISPR based Prime Editing technology. Six point mutations in the noncoding region, including four in 3'UTR and two in 5'UTR, were successfully constructed ([Extended Data Fig. 9b](#)). We first confirmed that these point mutations lead to lower mRNA expression ([Extended Data Fig. 9c](#)). Functional assays showed that these mutations significantly enhanced cell proliferation, migration, self-renewal, and invasion capacities, consistent with that observed with the disrupted cell lines ([Extended Data Fig. 9d-e](#)). Taken together, these data support that *PPP1R12B* is a potential tumour suppressor and candidate noncoding driver for HCC.

For the potential oncogene *KCNJ12*, the disruption cell lines were successfully constructed by knockdown (SNU182 and HepG2) and by knockout (PLC/PRF/5, PVTT, SNU387 and HHL5). Consistent results were observed between each two independent disrupted cell lines. Compared with the normal cell line HHL5, results from five tumour cell lines together showed that *KCNJ12* disruption significantly impaired tumour migration, invasion, self-renewal, and cell proliferation ([Extended Data Fig. 9f](#)). To further explore whether point mutations identified in *KCNJ12* is able to cause phenotypic changes, we knocked in the mutations into the HepG2 cell line by Prime Editing. Three point mutations, including one in the coding region and two in the noncoding regions, were successfully constructed ([Extended Data Fig. 9g](#)). We first confirmed that these mutations lead to a higher level of mRNA expression ([Extended Data Fig. 9h](#)). Functional assays showed that these point mutations significantly enhanced cell proliferation (mut1 and mut2), migration (mut2), invasion (mut1 and mut2), and self-renewal capacities (mut1, mut2, and mut3), consistent with that observed with disrupted cell lines ([Extended Data Fig. 9i-j](#)). These data collectively validate that *KCNJ12* is indeed an oncogenic candidate noncoding driver for HCC.

For the potential tumour suppressor *FGA*, the disruption cell lines were successfully

constructed both by knockdown (PLC/PRF/5, PVTT, SNU387, and SNU182) and by knockout (Huh7). Consistent results observed between each two independent disrupted cell lines showed that low expression of *FGA* significantly enhanced tumour migration, invasion, self-renewal, and cell proliferation ([Extended Data Fig. 10i-j](#)).

To evaluate whether point mutations of *FGA* are enough to lead to phenotypic changes, we exploited the CRISPR/Cas9-based Prime Editing technology to induce point mutations in *FGA* ([Extended Data Fig. 10a](#)). Three point mutations in the coding region, including P289L, P362R and D793H, were successfully constructed ([Fig. 5e](#)). These mutations lead to lower mRNA and protein expression ([Fig. 5f-g](#)). Functional experiments showed that the three point mutations significantly enhanced cell proliferation, migration, invasion and self-renewal capacities, respectively ([Fig. 5h-i](#)). Meanwhile, we failed to knock in the detected noncoding 3'UTR mutations of *FGA* by either Prime Editing technology or other base editors, like cytosine base editor (CBE) or the adenine base editor (ABE). To confirm the functional role of *FGA* 3'UTR mutations, we first knocked out endogenous *FGA* by sgRNA in the HepG2 cell line and then induced ectopic stable expression of mutant 3'UTR with wild-type as control ([Extended Data Fig. 10e](#)). We confirmed the lower protein level of *FGA* in the mutant cell line compared with the wild-type group by Western blot ([Extended Data Fig. 10f](#)). We found that the capabilities of cell proliferation, migration, invasion, and self-renewal were significantly increased in 3'UTR mutant HepG2 cells ([Extended Data Fig. 10g-h](#)). These data are consistent with that observed with the coding mutations, together supporting that *FGA* is a potential tumour suppressor gene and candidate driver gene for HCC.

To further explore the downstream signals of *FGA*, we performed phospho-specific protein microarray containing 584 site-specific phosphorylated proteins and their unphosphorylated counterparts. The sh-*FGA* cells were found to possess 62 sites with elevated phosphorylation levels in 58 genes, as well as 49 sites with decreased phosphorylation levels in 45 genes. Among the up-regulated sites, phosphorylated tyrosine kinase 2 (p-TYK2) was ranked first, and its phosphorylation level in sh-*FGA* cells was three times higher than that of sh-Ctrl cells ([Extended Data Fig. 10l-m](#)). Western blot assays confirmed the enhanced phosphorylation level of TYK2 and

its target protein signal transducer and activator of transcription 3 (STAT3, Tyr705) in sh-*FGA* cells in comparison with sh-Ctrl cells (Extended Data Fig. 10n). To further substantiate the role of *FGA* in the JAK-STAT3 pathway, we analyzed the subcellular distribution of p-TYK2. Accumulation of p-TYK2 was observed in the cytoplasm rather than the nucleus, suggesting that TYK2 phosphorylation in the cytoplasm is required for *FGA*-regulated JAK-STAT signaling (Fig. 5l). Next, we applied a specific inhibitor of p-TYK2 (BMS-986165) to sh-*FGA* cells, which attenuated their migration capacity (Extended Data Fig. 10o). By contrast, the administration of two AKT inhibitors (MK-2206 and AZD5363) had no effect on the migration capacity of sh-*FGA* cells. These results suggested that *FGA* knockdown might not activate *AKT* signaling in HCC, distinct to the proposed role of *FGA* loss facilitated lung cancer tumour metastasis by activating *AKT* signaling³⁷. *STAT3* is a transcription factor critical for interleukin-6 (IL-6) production. The levels of *IL6* mRNA and cellular supernatant IL-6 protein were significantly higher in sh-*FGA* cells compared to sh-Ctrl cells (Extended Data Fig. 10p-q and Fig. 5m). Western blot assays across an additional cohort of 75 HCC patients showed significant negative correlations between *FGA* protein level and TYK2 phosphorylation, and between *FGA* protein level and IL-6 concentration (Extended Data Fig. 10r and Fig. 5n).

Supplementary Discussion

Previous large-scale genomic analysis of HCC mainly focused on exome regions^{19,20,38-42}. Although some recent whole genome analyses of Chinese HCC revealed recurrent HBV integrations^{19,27,43}, both the cohort size and sequencing depth are very limited and preclude further analyses. In this study, we performed an in-depth, integrative, whole-genome analysis of 494 HCC tumors from China, covering coding and noncoding drivers, mutational signatures, genomic rearrangements, clustered alterations, ecDNAs and evolutionary history. These results together depict a comprehensive whole genome landscape of Chinese HCC.

We identified a total of 54 significantly mutated genes for HCC, larger than that reported in any previous HCC study. Notably, 28 of the 31 noncoding drivers identified in our cohort were

firstly reported for HCC, suggesting that our understanding of HCC genome is still very limited. These candidate drivers can only be determined as true drivers by detailed functional validation. We selected three representative candidate genes to perform functional assays, including *KCNJ12*, *PPP1R12B*, and *FGA*. Next, we focused on one candidate driver gene determined as both coding and noncoding driver, *FGA*. *FGA* is a hepatic lineage-related gene. Whether the increased mutational rates adjacent to lineage-related genes contribute to tumorigenesis remains debatable^{23,44}. We showed that *FGA* functioned as a potential tumour suppressor by *in vitro* assays with cell lines and *in vivo* assays with the subcutaneous xenograft mouse model, and these experiments also revealed that *FGA* dysfunction leads to elevated tumour proliferation and invasion. Further analysis showed that *FGA* mutation may activate the JAK-STAT pathway, resulting in improved transcription of *IL6*, which contributes to a chronic inflammatory state that might accelerate hepatocarcinogenesis. Previous studies have reported the pivotal role of JAK-STAT signaling during hepatocarcinogenesis⁴⁵, in which both STAT family transcription factors and cytokines induce malignant hepatocyte transformation and aggressive tumour cell behavior by activating downstream target genes⁴⁶ and reprogramming the tumour environment⁴⁷. Specifically, IL-6 is regarded as the key link between chronic inflammation and tumour development. In addition to activation of JAK signaling, HBV-encoded genes (*HBx* and *HBs*)⁴⁸, alcohol consumption⁴⁹ and other stresses⁵⁰ could enhance the expression or secretion of IL-6 and initiate the cancer-promoting positive feedback circuit.

Collectively, our whole genome analysis of all classes of somatic alterations in CLCA portrayed a comprehensive genomic landscape of Chinese HCC. Our CLCA study serves as a valuable resource that provides important biological insights into HCC carcinogenesis and clinical implications to HCC diagnosis and treatment.

Supplementary References

- 1 Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249, doi:10.3322/caac.21660 (2021).
- 2 Llovet, J. M. *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers* **7**, 6, doi:10.1038/s41572-020-00240-3 (2021).
- 3 Villanueva, A. Hepatocellular Carcinoma. *N Engl J Med* **380**, 1450-1462, doi:10.1056/NEJMr1713263 (2019).
- 4 The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82-93, doi:10.1038/s41586-020-1969-6 (2020).
- 5 Rheinbay, E. *et al.* Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**, 102-111, doi:10.1038/s41586-020-1965-x (2020).
- 6 Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101, doi:10.1038/s41586-020-1943-3 (2020).
- 7 Li, Y. *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112-121, doi:10.1038/s41586-019-1913-9 (2020).
- 8 Gerstung, M. *et al.* The evolutionary history of 2,658 cancers. *Nature* **578**, 122-128, doi:10.1038/s41586-019-1907-7 (2020).
- 9 Fujimoto, A. *et al.* Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet* **48**, 500-509, doi:10.1038/ng.3547 (2016).
- 10 Letouze, E. *et al.* Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun* **8**, 1315, doi:10.1038/s41467-017-01358-x (2017).
- 11 Zhang, B. *et al.* 42,573 cases of hepatectomy in China: a multicenter retrospective investigation. *Science China Life Sciences* **61**, 660-670, doi:10.1007/s11427-017-9259-9 (2018).
- 12 Xue, R. *et al.* Liver tumour immune microenvironment subtypes and neutrophil heterogeneity. *Nature* **612**, 141-147, doi:10.1038/s41586-022-05400-x (2022).
- 13 Xue, R. *et al.* Genomic and Transcriptomic Profiling of Combined Hepatocellular and Intrahepatic Cholangiocarcinoma Reveals Distinct Molecular Subtypes. *Cancer Cell* **35**, 932-947 e938, doi:10.1016/j.ccell.2019.04.007 (2019).
- 14 Qi, M., Pang, J., Mitsiades, I., Lane, A. A. & Rheinbay, E. Loss of chromosome Y in primary tumors. *Cell* **186**, 3125-3136.e3111, doi:10.1016/j.cell.2023.06.006 (2023).
- 15 Abdel-Hafiz, H. A. *et al.* Y chromosome loss in cancer drives growth by evasion of adaptive immunity. *Nature*, doi:10.1038/s41586-023-06234-x (2023).

- 16 Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218, doi:10.1038/nature12213 (2013).
- 17 Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029-1041 e1021, doi:10.1016/j.cell.2017.09.042 (2017).
- 18 Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol* **17**, 128, doi:10.1186/s13059-016-0994-0 (2016).
- 19 Kan, Z. *et al.* Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res* **23**, 1422-1433, doi:10.1101/gr.154492.113 (2013).
- 20 Gao, Q. *et al.* Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* **179**, 561-577 e522, doi:10.1016/j.cell.2019.08.052 (2019).
- 21 Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495-501, doi:10.1038/nature12912 (2014).
- 22 Zhu, H. *et al.* Candidate Cancer Driver Mutations in Distal Regulatory Elements and Long-Range Chromatin Interaction Networks. *Mol Cell* **77**, 1307-1321 e1310, doi:10.1016/j.molcel.2019.12.027 (2020).
- 23 Imielinski, M., Guo, G. & Meyerson, M. Insertions and Deletions Target Lineage-Defining Genes in Human Cancers. *Cell* **168**, 460-472.e414, doi:<https://doi.org/10.1016/j.cell.2016.12.025> (2017).
- 24 Kou, J. T. *et al.* LncRNA NEAT1 regulates proliferation, apoptosis and invasion of liver cancer. *Eur Rev Med Pharmacol Sci* **24**, 4152-4160, doi:10.26355/eurrev_202004_20995 (2020).
- 25 Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science* **253**, 49-53 (1991).
- 26 Xue, R. *et al.* Variable Intra-Tumor Genomic Heterogeneity of Multiple Lesions in Patients With Hepatocellular Carcinoma. *Gastroenterology* **150**, 998-1008, doi:10.1053/j.gastro.2015.12.033 (2016).
- 27 Sung, W. K. *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* **44**, 765-769, doi:10.1038/ng.2295 (2012).
- 28 Zhao, L. H. *et al.* Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun* **7**, 12992, doi:10.1038/ncomms12992 (2016).
- 29 Sartorius, K. *et al.* The Regulatory Role of MicroRNA in Hepatitis-B Virus-Associated Hepatocellular Carcinoma (HBV-HCC) Pathogenesis. *Cells* **8**, 1504, doi:10.3390/cells8121504 (2019).
- 30 Deshpande, V. *et al.* Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat Commun* **10**, 392, doi:10.1038/s41467-018-08200-y (2019).

- 31 Nault, J.-C., Ningarhari, M., Rebouissou, S. & Zucman-Rossi, J. The role of telomeres and telomerase in cirrhosis and liver cancer. *Nat Rev Gastroenterol Hepatol*, doi:10.1038/s41575-019-0165-3 (2019).
- 32 Veiga, R. N., de Oliveira, J. C. & Gradia, D. F. PBX1: a key character of the hallmarks of cancer. *J Mol Med (Berl)* **99**, 1667-1680, doi:10.1007/s00109-021-02139-2 (2021).
- 33 Llovet, J. M. *et al.* Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med* **359**, 378-390, doi:10.1056/NEJMoa0708857 (2008).
- 34 Kudo, M. *et al.* Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. *Lancet* **391**, 1163-1173, doi:10.1016/S0140-6736(18)30207-1 (2018).
- 35 Satriano, L., Lewinska, M., Rodrigues, P. M., Banales, J. M. & Andersen, J. B. Metabolic rearrangements in primary liver cancers: cause and consequences. *Nat Rev Gastroenterol Hepatol* **16**, 748-766, doi:10.1038/s41575-019-0217-8 (2019).
- 36 Ng, S. W. K. *et al.* Convergent somatic mutations in metabolism genes in chronic liver disease. *Nature* **598**, 473-478, doi:10.1038/s41586-021-03974-6 (2021).
- 37 Wang, M. *et al.* Fibrinogen Alpha Chain Knockout Promotes Tumor Growth and Metastasis through Integrin-AKT Signaling Pathway in Lung Cancer. *Mol Cancer Res* **18**, 943-954, doi:10.1158/1541-7786.Mcr-19-1033 (2020).
- 38 Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* **47**, 505-511, doi:10.1038/ng.3252 (2015).
- 39 Totoki, Y. *et al.* Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat Genet* **46**, 1267-1273, doi:10.1038/ng.3126 (2014).
- 40 Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327-1341 e1323, doi:10.1016/j.cell.2017.05.046 (2017).
- 41 Chaisaingmongkol, J. *et al.* Common Molecular Subtypes Among Asian Hepatocellular Carcinoma and Cholangiocarcinoma. *Cancer Cell* **32**, 57-70 e53, doi:10.1016/j.ccell.2017.05.009 (2017).
- 42 Candia, J. *et al.* The genomic landscape of Mongolian hepatocellular carcinoma. *Nat Commun* **11**, 4383, doi:10.1038/s41467-020-18186-1 (2020).
- 43 Jhunjhunwala, S. *et al.* Diverse modes of genomic alteration in hepatocellular carcinoma. *Genome Biol* **15**, 436, doi:10.1186/s13059-014-0436-9 (2014).
- 44 Dietlein, F. *et al.* Genome-wide analysis of somatic noncoding mutation patterns in cancer. *Science* **376**, eabg5601, doi:10.1126/science.abg5601 (2022).
- 45 Calvisi, D. F. *et al.* Ubiquitous Activation of Ras and Jak/Stat Pathways in Human HCC.

Gastroenterology **130**, 1117-1128, (2006).

46 Ding, X. *et al.* Genomic and Epigenomic Features of Primary and Recurrent Hepatocellular Carcinomas. *Gastroenterology* **157**, 1630-1645.e1636, (2019).

47 Makino, Y. *et al.* CTGF Mediates Tumor-Stroma Interactions between Hepatoma Cells and Hepatic Stellate Cells to Accelerate HCC Progression. *Cancer Res* **78**, 4902-4914, doi:10.1158/0008-5472.CAN-17-3844 (2018).

48 Xiang, W.-Q. *et al.* Hepatitis B virus X protein stimulates IL-6 expression in hepatocytes via a MyD88-dependent pathway. *J Hepatol* **54**, 26-33, doi:<https://doi.org/10.1016/j.jhep.2010.08.006> (2011).

49 Johnson, C. *et al.* Interleukin-6 and its receptor, key players in hepatobiliary inflammation and cancer. *Transl Gastrointest Cancer* **1**, 58-70, doi:10.3978/j.issn.2224-4778.2011.11.02 (2012).

50 He, G. & Karin, M. NF- κ B and STAT3 – key players in liver inflammation and cancer. *Cell Res* **21**, 159-168, doi:10.1038/cr.2010.183 (2011).