

# Whole-exome sequencing reveals genomic landscape of intrahepatic cholangiocarcinoma and identifies *SAV1* as a potential driver

Received: 27 January 2022

Accepted: 7 November 2024

Published online: 17 November 2024

 Check for updates

Zheng-Jun Zhou<sup>1,2,4</sup>, Yu-Hang Ye<sup>1,2,4</sup>, Zhi-Qiang Hu<sup>1,2,4</sup>, Yue-Ru Hou<sup>3,4</sup>, Kai-Xuan Liu<sup>1,2</sup>, Rong-Qi Sun<sup>1,2</sup>, Peng-Cheng Wang<sup>1,2</sup>, Chu-Bin Luo<sup>1,2</sup>, Jia Li<sup>1,2</sup>, Ji-Xue Zou<sup>2</sup>, Jian Zhou<sup>1,2</sup>, Jia Fan<sup>1,2</sup>, Cheng-Li Song<sup>3</sup> ✉ & Shao-Lai Zhou<sup>1,2</sup> ✉

Intrahepatic cholangiocarcinoma (ICC) is the second most common primary hepatic malignancy after hepatocellular carcinoma, with poor prognosis and limited treatment options. The genomic features of ICC in Chinese patients remain largely unknown. In this study, we perform deep whole-exome sequencing of 204 Chinese primary ICCs and characterize genomic alterations and clonal evolution, and reveal their associations with patient outcomes. We identify six mutational signatures, including Signatures A and F, which are highly similar to previously described signatures linked to aristolochic acid and aflatoxin exposures, respectively. We also identify 13 significantly mutated genes in the ICC samples, including *SAV1*. We find that *SAV1* was mutated in 2.9% (20/672) of 672 ICC samples. *SAV1* mutation is associated with lower *SAV1* protein levels, higher rates of tumor recurrence, and shorter overall patient survival. Biofunctional investigations reveal a tumor-suppressor role of *SAV1*: its inactivation suppresses Hippo signaling, leading to YAP activation, thereby promoting tumor growth and metastasis. Collectively, our results delineate the genomic landscape of Chinese ICCs and identify *SAV1* as a potential driver of ICC.

Intrahepatic cholangiocarcinoma (ICC) is the second most common primary hepatic malignancy after hepatocellular carcinoma (HCC) and has an increasing incidence worldwide<sup>1,2</sup>. Advances in diagnostic modalities and clinical screening have made early detection and curative resection of ICC possible; however, long-term patient survival is still poor because of high relapse rates<sup>3,4</sup>. Recently, several pioneering studies expanded our understanding of ICC pathogenesis at the molecular level by delineating the landscape of genetic alterations that underlie ICC carcinogenesis<sup>5–11</sup>. However, most of these studies contained limited numbers of ICC samples, and the patients were from

multiple countries with different epidemiological and ancestral backgrounds. Furthermore, most of the studies used targeted sequencing, and whole-exome sequencing (WES) or whole-genome sequencing was only performed with low coverage. More importantly, exhaustive follow-up was performed on very few of the patients. These shortcomings may impede the discovery of potential drivers of ICC, as well as hinder the exploration of potential etiological factors contribute to genetic alterations and their association with patient outcomes.

In this work, we aim to systematically define genomic alterations in Chinese patients with ICC. We perform deep WES on 204 Chinese

<sup>1</sup>Department of Liver Surgery and Transplantation, Zhongshan Hospital, Fudan University, Shanghai, China. <sup>2</sup>Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai, China. <sup>3</sup>Institute of Cancer Stem Cell, Dalian Medical University, Dalian, Liaoning, China. <sup>4</sup>These authors contributed equally: Zheng-Jun Zhou, Yu-Hang Ye, Zhi-Qiang Hu, Yue-Ru Hou. ✉ e-mail: [songchengli@dmu.edu.cn](mailto:songchengli@dmu.edu.cn); [zhoushaolai99@sina.com](mailto:zhoushaolai99@sina.com)

primary ICCs and analyze the associations of genomic alterations with tumor progression and patient outcomes. Focusing on a potential driver of ICC, *SAVI*, we perform Sanger sequencing to evaluate all the *SAVI* coding exons in an additional 468 ICC specimens. We also investigate the functional effects and associated mechanisms of *SAVI* on tumor growth and metastasis. Collectively, our results delineate the genomic landscape of Chinese ICCs and identify *SAVI* as a potential driver of ICC.

## Results

### Overview of genomic alterations in patients with primary ICC

To discover the genomic alterations in Chinese patients with ICC, we performed WES of tumor and matched non-cancerous liver-tissue samples from 204 patients with ICC. The average sequencing depth was 445.9-fold for tumors and 138.7-fold for normal tissues (Supplementary Data 1). We identified a total of 21,772 somatic coding SNVs and 1302 indels, with 2–1628 mutations per tumor genome (Supplementary Data 2). We performed Sanger sequencing to validate 1756 randomly selected somatic non-synonymous mutations, which revealed a high true-discovery rate (95.9%). There was an average of 3.77 (range = 0.07–54.27) somatic coding mutations per Mb (Fig. 1A).

The WES results indicated that C > A transversions and C > T and T > C transitions were ubiquitous in all of the ICCs (Fig. 1A), a feature shared by other ICC cohorts<sup>8,9,11</sup>. In addition, we identified 17 amplified segments, which harbored several known oncogenes, including *MYC*, *ERBB2*, and *CCND1*. We also identified 24 lost segments, which harbored tumor suppressors such as *TP53* (17p13) and *CDKN2A/B* (9q21) (Fig. 1A; Supplementary Fig. 1A, B; Supplementary Data 3). In addition, we identified whole-genome doubling in 44 (21.6%) of the 204 tumors (Supplementary Data 4).

We correlated the tumor mutation burden (TMB) and copy number variation (CNV) burden with patient prognosis. The results showed that high TMB was predictive of low recurrence-free survival (RFS) but had no ability to predict OS. In addition, CNV burden showed no associations with patient prognosis, either for OS or RFS (Supplementary Fig. 1C).

### Mutational signatures

We applied NMF and identified six mutation signatures in the 204 ICC samples (Signatures A–F; Fig. 1B, C; Supplementary Fig. 2A; Supplementary Data 5). Signature A was characterized by dominant T > A mutations and was highly similar to the previously described Signature 22 (cosine correlation similarity = 0.99), which is known to result from exposure to aristolochic acid and to be associated with a high mutational burden<sup>12</sup>. Consistent with that, Signature A was a predominant signature in the hypermutated ICC samples. Signature B, characterized by C > G mutations, was similar to the previously described Signature 13, which has been proposed as an APOBEC-driven hypermutated phenotype. Signatures C and E were highly similar to the previously described signatures 1 and 5, respectively (correlation coefficient = 0.96 and 0.93, respectively), which have been associated with aging. Signature D showed similarity to the previously described Signature 20 (correlation similarity = 0.85), which was previously confirmed to be driven by defective DNA mismatch repair and associated with high numbers of indels at mono/polynucleotide repeats. Consistent with that, Signature D was mainly identified in four ICCs with high numbers of indels. Signature F, characterized by C > A mutations, showed strong similarity to the previously described Signature 24 (correlation similarity = 0.94), which is associated with exposure to aflatoxin in cancers.

Hierarchical clustering, based on the prevalence of the mutation signatures, identified nine groups of tumors (Msig-clusters 1–9; Fig. 1C) characterized by enrichment of specific mutation signatures. For example, Msig-clusters 1 and 6 were characterized by Signature A, as they contained more mutations that fit Signature A than other clusters. Furthermore, Msig-cluster 3 was characterized by Signature D, Msig-

cluster 4 was characterized by signature F, and Msig-cluster 5 was characterized by Signature B. Associations between the signature-based clusters and clinical background and genetic alterations were also revealed (Supplementary Fig. 2B). Msig-clusters 1 and 6 exhibited significant correlations with higher TMB ( $P = 0.029$  and  $P = 0.003$ , respectively), while Msig-clusters 8 and 9 were correlated with lower TMB ( $P = 9.59 \times 10^{-5}$  and  $P = 0.021$ , respectively). Msig-clusters 4 and 5 demonstrated higher TMB ( $P = 8.97 \times 10^{-5}$  and  $P = 0.003$ , respectively) and higher CNV burden ( $P = 0.038$  and  $P = 0.035$ , respectively). Especially, we confirmed correlations between Msig-cluster 5 and the presence of hepatolithiasis ( $P = 0.017$ ), implicating AID/APOBEC enzymes in cholangitis-induced tumorigenesis. In addition, Msig-clusters 3 and 9 respectively exhibited correlations with the presence of *ARID1A* mutations or *IDH1* mutations ( $P = 0.005$  and  $P = 0.027$ , respectively).

### Significantly mutated protein-coding genes and associations with patient outcomes

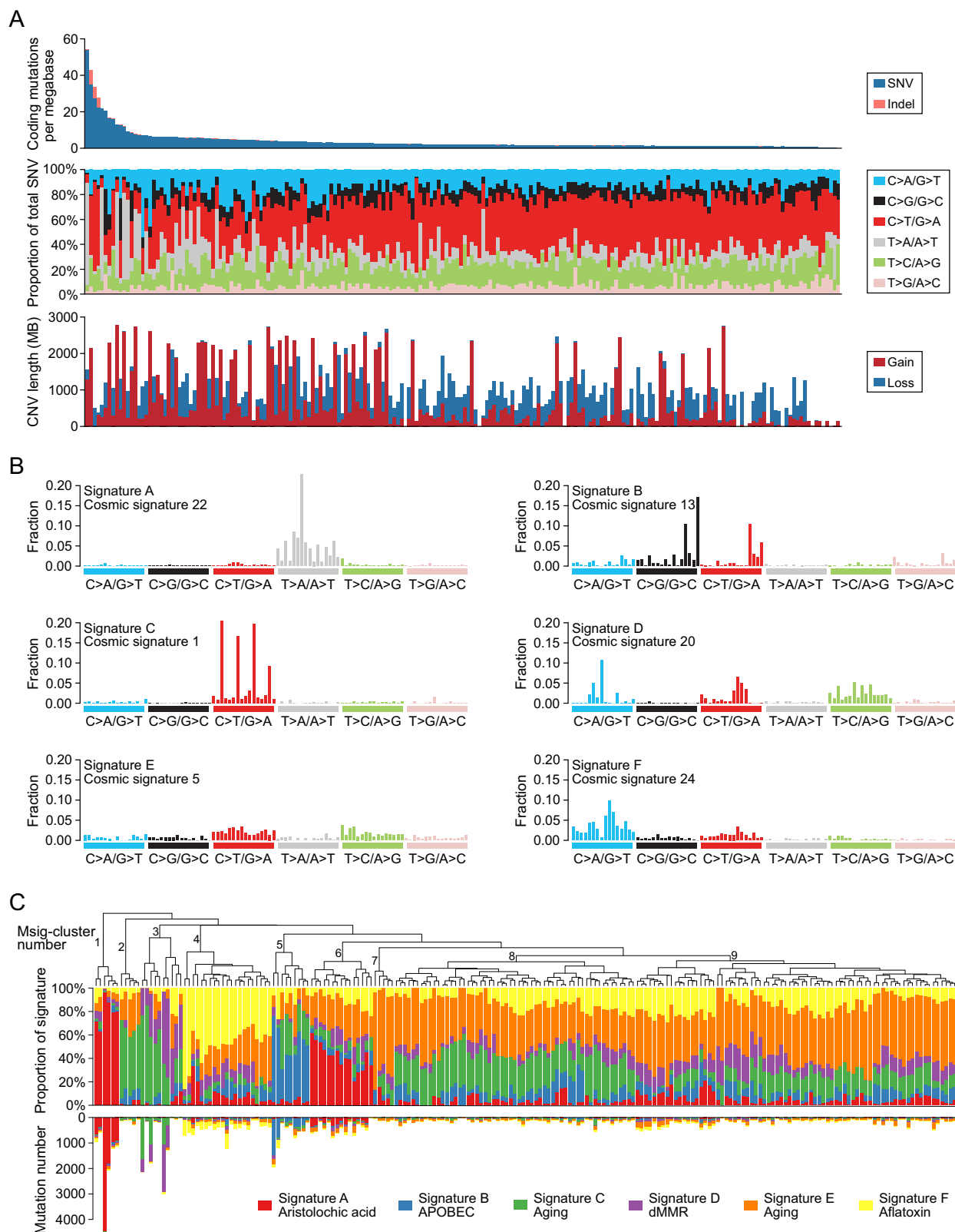
We used MutSigCV to identify cancer driver genes<sup>13</sup>. Thirteen genes were significantly enriched with damaging mutations ( $q < 0.1$ ; Fig. 2A; Supplementary Data 6): *TP53*, *KRAS*, *IDH1*, *ELF3*, *SAVI*, *BAP1*, *ARID1A*, *PBRM1*, *PTEN*, *ARID2*, *BRD7*, *CDKN2A*, and *ACVR2A*. We also identified oncogenes (*MYC*, *ERBB2*, and *CCND1*) with copy-number amplification and tumor-suppressor genes (*CDKN2A* and *CDKN2B*) with copy-number deletion. We then compared the mutation frequencies in the 13 genes to those in a combined TCGA ( $n = 30$ ) and MSK ( $n = 158$ ) cohort. We found that the mutation frequencies in several of the 13 genes were higher in our ICC cohort than in the TCGA and MSK cohort, including *TP53* (30% versus 16%), *KRAS* (14% versus 6%), *ELF3* (5% versus 1%), and *SAVI* (3% versus 0%), whereas the mutation frequencies of *IDH1* (10% versus 26%) and *PBRM1* (7% versus 13%) were slightly lower in our ICC cohort than in the TCGA and MSK cohort (Fig. 2B).

We next correlated somatic mutations in specific genes with patient outcomes. We found that somatic mutations in *TP53*, *KRAS*, *ELF3*, and *SAVI* were correlated with patient OS and/or RFS (Supplementary Fig. 3). When we looked at the combination of these four genes (*TP53*, *KRAS*, *ELF3*, and *SAVI*; Fig. 2C), the cumulative recurrence rate was highest (92.9%) in patients with mutations in two or more of the four genes, which was significantly higher than the cumulative recurrence rate in patients with mutation in only one of the four genes (76.3%) or in none of the four genes (60.9%; Fig. 2D). Similarly, the 5-year OS among patients with mutations in two or more of the four genes was 7.7%, which was significantly lower than that among patients with mutation in only one of the four genes (27.2%) or in none of the four genes (45.3%; Fig. 2E).

### Clonal diversity and evolution of recurrent somatic mutations

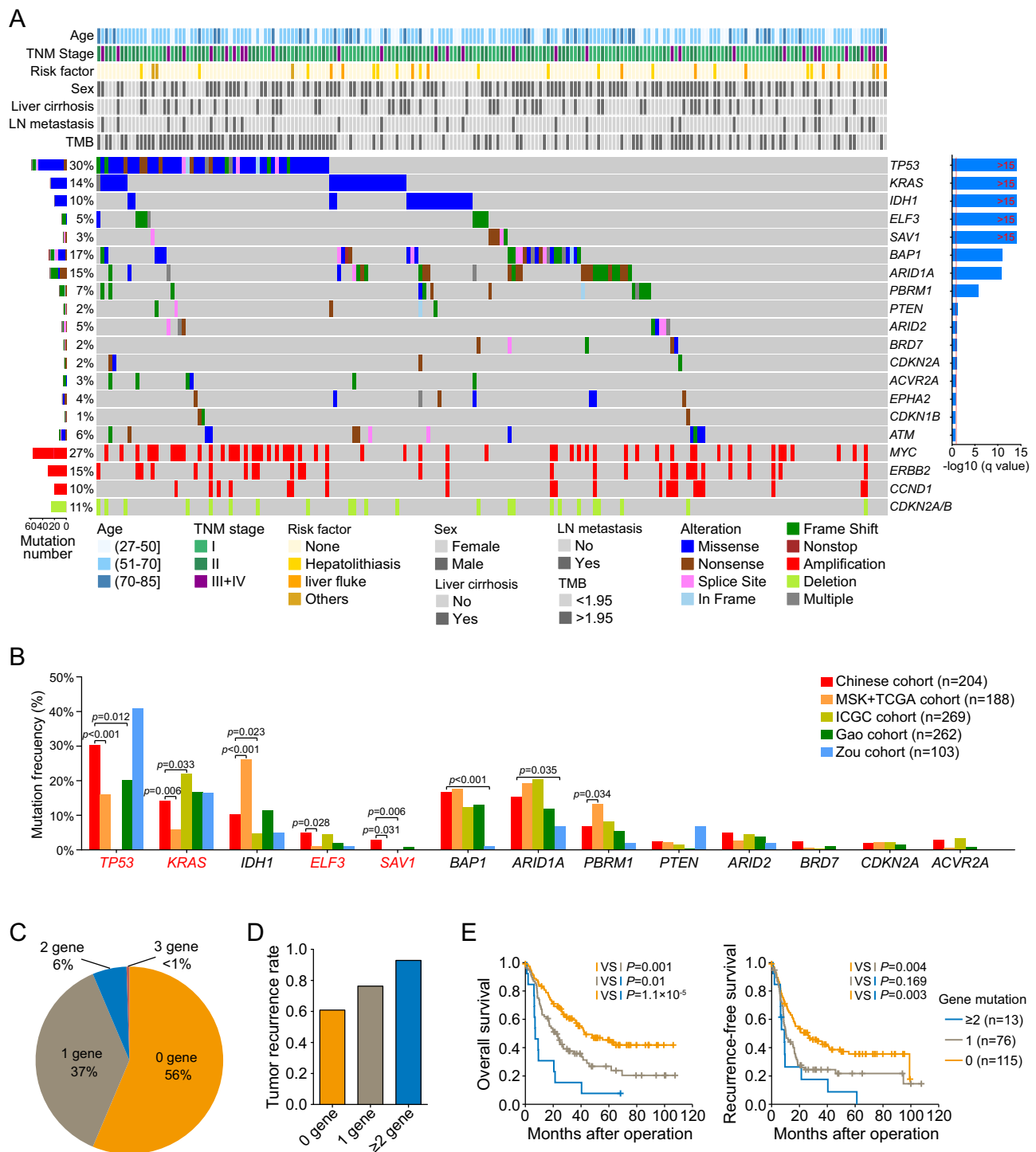
To estimate genetic diversity based on WES of all the ICC samples, we used an analysis pipeline described in a previous study<sup>14</sup> to compute a Shannon index for each patient<sup>15,16</sup>, which estimates not only the number of clones, but also their relative abundance. A patient with a single clone would have a Shannon index of 0, with increasing Shannon index corresponding to a greater number of clones and a smaller dispersal of CCFs (Fig. 3A). We divided the patients into a “clonal equilibrium” group, defined by a Shannon index higher than the median value among all patients, and a “clonal dominance” group, defined by a Shannon index below the median value (Fig. 3A). Kaplan–Meier survival analysis showed that clonal dominance was associated with significantly shorter OS and RFS compared with clonal equilibrium. Univariate and multivariate analyses revealed that the Shannon index was an independent prognostic factor for both OS and RFS (Fig. 3B; Supplementary Data 7).

We next deduced the approximate sequential order in which the somatic alterations occurred during ICC development by calculating how often a recurrent non-synonymous mutation resided on a clonal



**Fig. 1 | Global genomic alterations in 204 Chinese primary ICCs. A** Mutational burden in coding regions across 204 ICCs (top panel). Distribution of six substitution patterns sorted by the number of somatic coding mutations (middle panel). Somatic copy number alterations, including copy number gains and losses in each patient (bottom panel). **B** Patterns of six mutation signatures (Signatures A,

B, C, D, E, and F) identified in 204 ICCs; the most similar COSMIC signature to each is also indicated. **C** Unsupervised hierarchical clustering of 204 primary ICCs based on the intensity of the signatures in each sample. SNV single-nucleotide variants, Indel insertions or deletions. Source data are provided as a Source Data file.



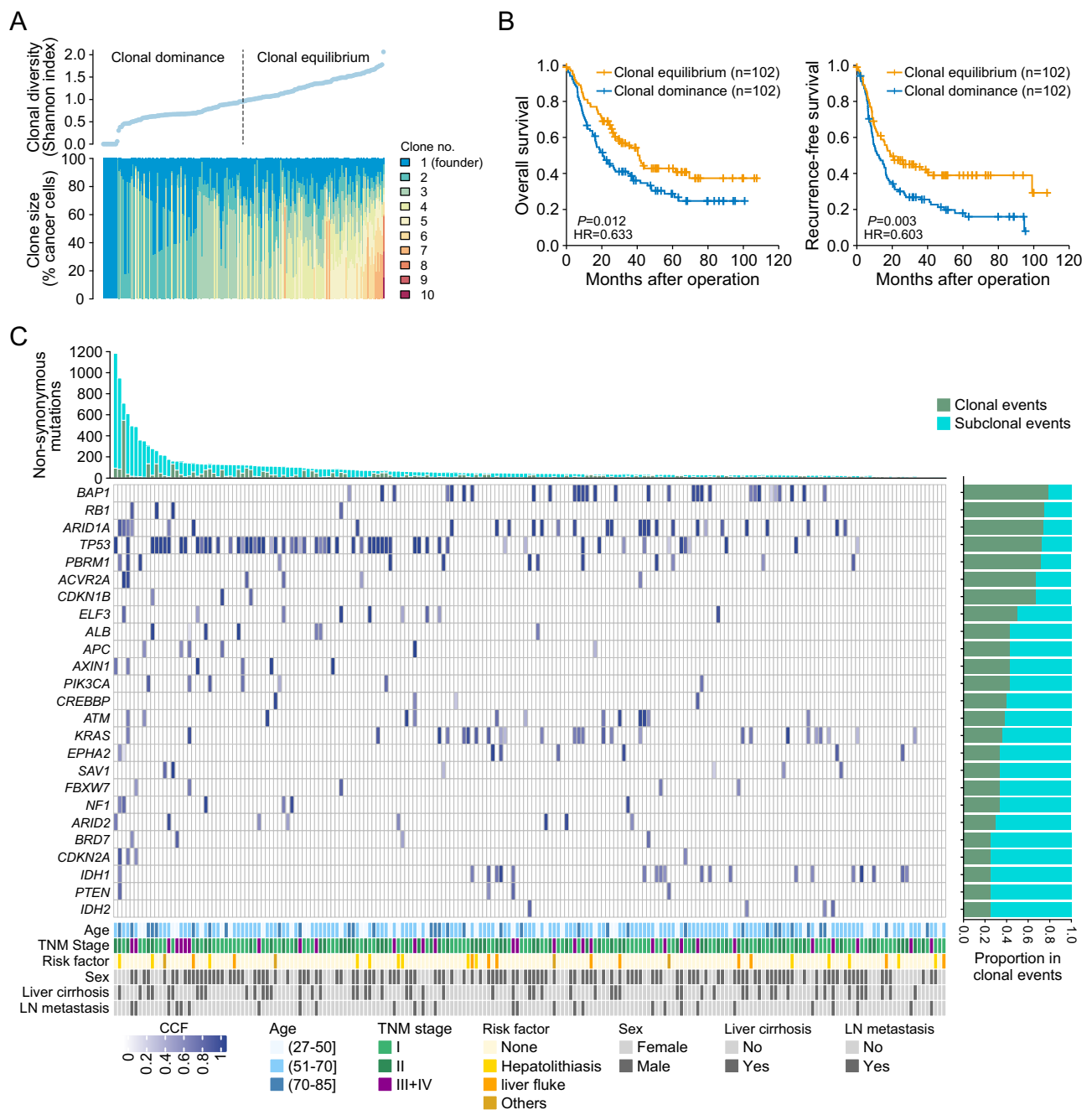
**Fig. 2 | Significantly mutated protein-coding genes and associations with patient outcome.** **A** Thirteen genes with statistically significant levels of mutation (MutSigCV, FDR < 0.1), together with *EPHA2*, *CDKN1B*, and *ATM* with FDR < 0.2, and significant copy number alterations in likely cancer driver genes are shown. **B** Comparison of mutation frequencies for 13 significantly mutated genes between the 204 ICCs in our cohort and ICCs from other cohorts, Chi-square test.

**C** Proportions of 204 ICCs carrying *TP53*, *KRAS*, *ELF3*, and *SAVI* mutations. **D** Tumor recurrence rates in patients with HCC carrying different numbers of mutated genes (*TP53*, *KRAS*, *ELF3*, and *SAVI*). **E** Kaplan-Meier survival analysis showing overall survival and recurrence-free survival based on *TP53*, *KRAS*, *ELF3*, and *SAVI* mutations. P: log-rank test, two-sided. TNM tumor node metastasis, LN lymph node, TMB tumor mutational burden. Source data are provided as a Source Data file.

event versus a subclonal event. This analysis confirmed that *BAP1*, *ARID1A*, and *TP53* mutations were the earliest mutations to evolve, followed by *PBRM1*, *ELF3*, and *KRAS* mutations. By contrast, *ARID2* and *IDH1/2* mutations were mostly restricted to subclonal events, which occurred at later time points (Fig. 3C).

**Somatic mutation of *SAVI* was associated with reduced *SAVI* expression and predictive of patient outcomes after curative resection**

We evaluated all coding exons of *SAVI* in an additional 468 pairs of ICCs and normal tissue samples by Sanger sequencing. We identified



**Fig. 3 | Clonal diversity and evolution of recurrent somatic mutations. A** Clonal diversity quantified and sorted by Shannon index in 204 ICCs. **B** Kaplan–Meier survival analysis showing overall survival and recurrence-free survival based on the Shannon index (clonal equilibrium and clonal dominance). P: log-rank test, two-sided. **C** Evolution of recurrent mutations in likely cancer driver genes. Patients are

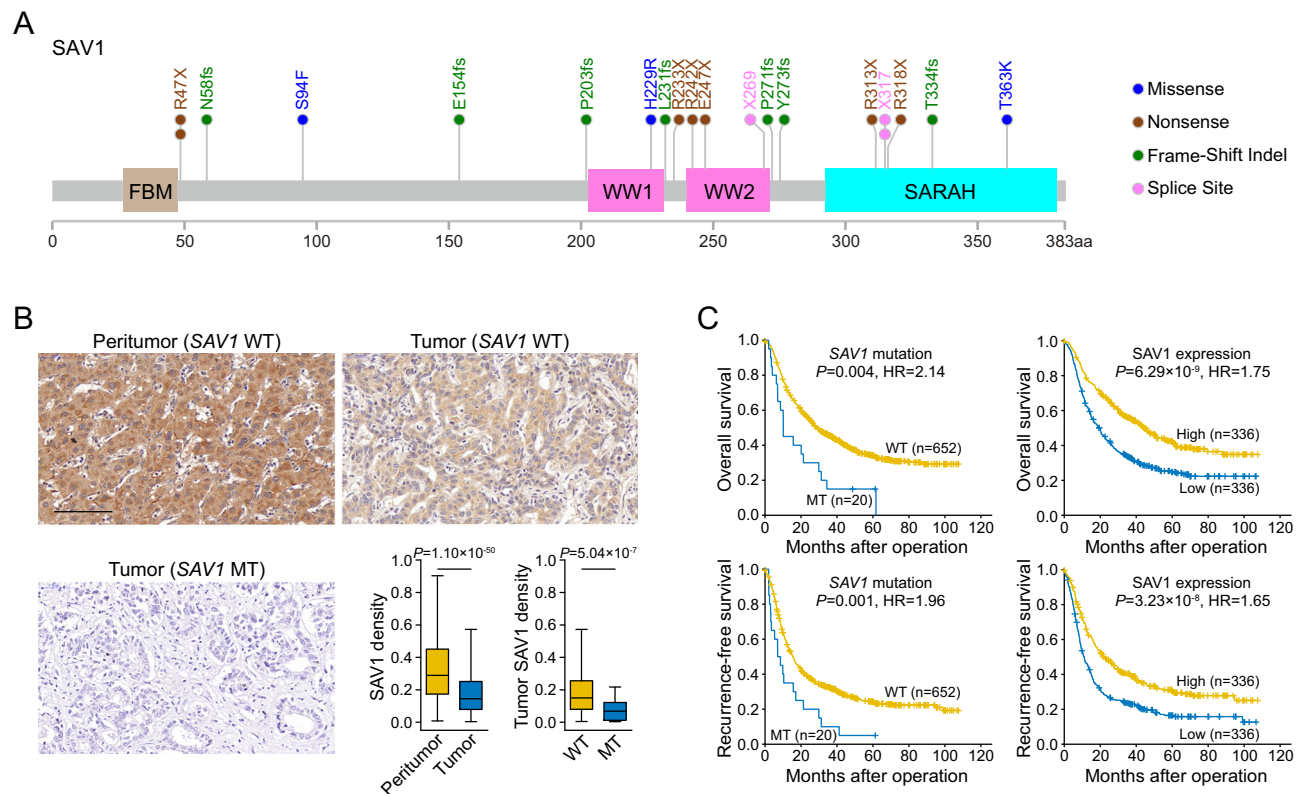
ordered by the total number of non-synonymous mutations; recurrent mutations are sorted by proportion defined as trunk and overall frequency. The cancer cell fraction (CCF) measures the proportion of the tumor population that harbors the mutation. TNM tumor node metastasis, LN lymph node, CCF cancer cell fraction.

*SAVI* somatic mutations in 14 of the 468 ICCs. Thus, in the overall total of 672 ICCs included in our study, we identified 20 *SAVI* somatic mutations in 20 different patients (Fig. 4A; Supplementary Fig. 4; Supplementary Data 8). Specifically, 14 (70%) of the 20 *SAVI* mutations were truncating mutations, which suggested a possible tumor-suppressor role of *SAVI* in ICC.

We further evaluated *SAVI* expression by immunohistochemistry in all 672 ICCs. The results showed that *SAVI* expression was down-regulated among all the tumor samples compared with that in the matched non-tumor liver samples (Fig. 4B). Moreover, the patients

with *SAVI* somatic mutation showed a further decrease in tumor *SAVI* expression (Fig. 4B).

We then correlated *SAVI* somatic mutation and expression with the patients' clinical characteristics and outcomes. *SAVI* somatic mutation was correlated with increased tumor size and poor tumor differentiation. The *SAVI* expression level was correlated with higher CA19-9 and GGT levels, increased tumor size, and lymphatic metastasis (Supplementary Data 9). Among the 672 patients with ICC, those with *SAVI* somatic mutation or reduced *SAVI* expression exhibited shorter OS and RFS compared with patients without those features (Fig. 4C).



**Fig. 4 | Clinical significance of *SAV1* mutation in 672 ICCs. A** Schematic diagram of the location of the mutation sites in *SAV1* identified in this study. **B** Representative *SAV1* staining in peritumor tissues and tumor tissues with wild-type *SAV1* or mutant *SAV1* (R47X). The statistics of *SAV1* staining density in different groups are shown. Scale bars = 100  $\mu$ m. In the boxplots, the central line represents median, the bounds

of boxes represent the first and third quartiles, and the upper and lower whiskers extend to the highest or the smallest value. Student's *t*-test, two-sided. **C** Kaplan–Meier survival analysis showing overall survival and recurrence-free survival based on *SAV1* mutation and expression level. P: log-rank test, two-sided. WT wild-type, MT mutation-type. Source data are provided as a Source Data file.

Univariate and multivariate analyses confirmed that *SAV1* somatic mutation and low expression were independent prognostic factors for OS and RFS (Supplementary Data 10).

### Structural analysis of the effects of *SAV1* mutations on protein structure

We obtained the full-length structure of human *SAV1* based on the prediction provided in the AlphaFold predicted-structure database (<https://alphafold.ebi.ac.uk/>), and showed the structure of 198–383 residues region in this study. We also built a structural model of the *SAV1*–MST2 heterotetramer to more intuitively understand the significance of dimerization of the WW12 domain to the function of *SAV1* (Supplementary Methods).

The structural and biochemical results of a recent study confirmed that *SAV1* can interact with MST2 and form a heterodimer through the SARAH domain<sup>17</sup>, which is essential for the interaction between *SAV1* and MST kinase<sup>18</sup>. Two *SAV1* molecules interact through their WW12 domains, composed of WW1 and WW2, to form a homodimer; WW2 is mostly responsible for the formation of the dimer interaction interface<sup>19</sup>. Furthermore, two *SAV1*–MST2 heterodimers further dimerize through the *SAV1* WW12 domain to form a heterotetramer, which phosphorylates the downstream LATS/MOB complex and promotes Hippo activation<sup>19</sup>.

The R47X, N58fs, E154fs, and P203fs mutations in our ICC samples caused deletion of the WW12 and SARAH domains. The deletion of WW12 is predicted to hinder the formation of the *SAV1* homodimer, and the absence of the SARAH domain is likely to impede the interaction between *SAV1* and MST kinase to form a heterodimer (Supplementary Fig. 5B). In addition, the L231fs, R233X, R242X, and E247X mutations in our ICC samples is predicted to destroy the integrity of the WW2

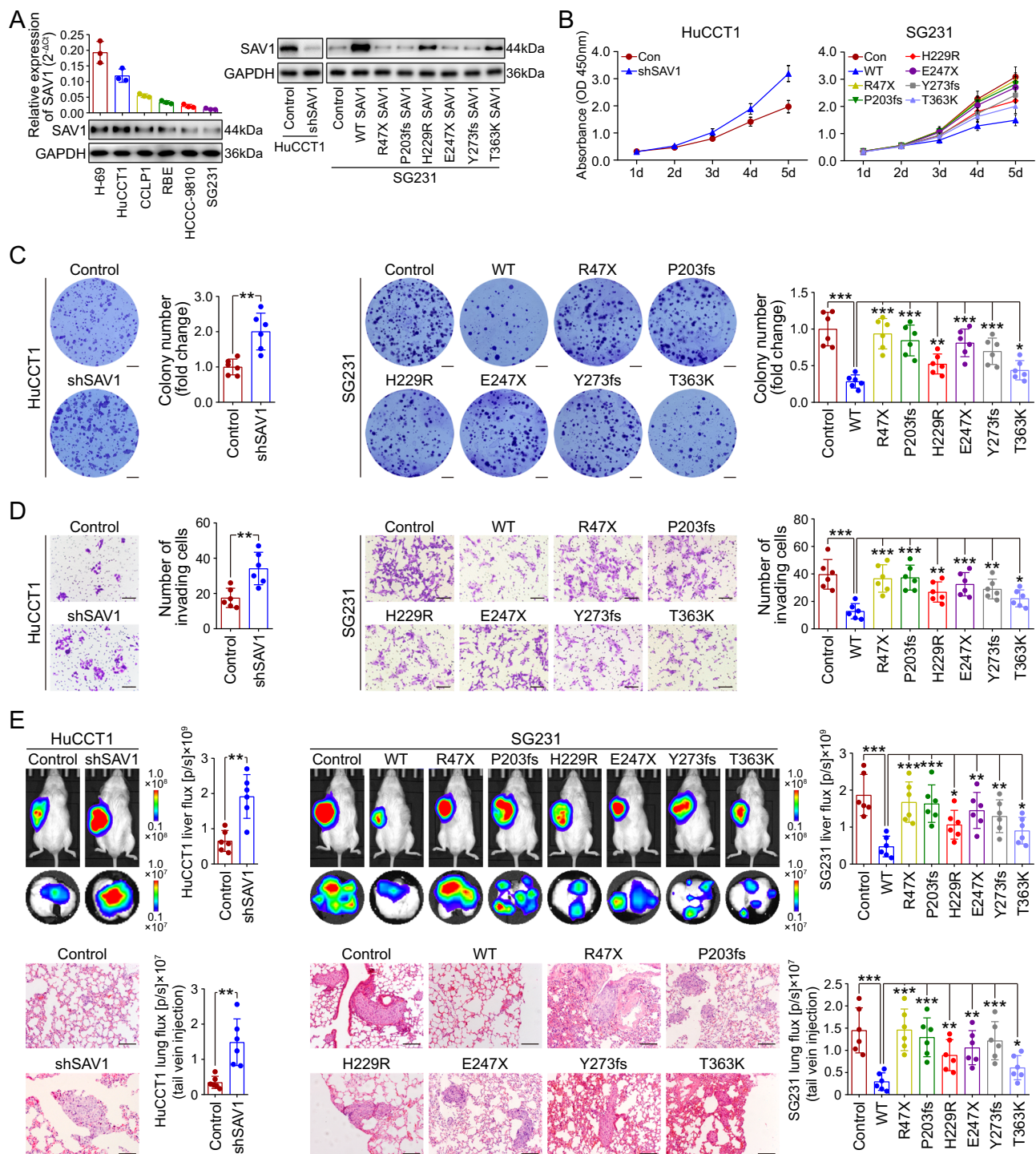
domain and caused deletion of the SARAH domain. Deletion of WW2 is expected to cause *SAV1* to fail to form a homodimer (Supplementary Fig. 5B). The absence of the SARAH domain is predicted to impede the interaction between *SAV1* and MST kinase. Similarly, the P271fs, Y273fs, R313X, R318X, and T334fs mutations caused the deletion of the SARAH domain, which is essential for the interaction between *SAV1* and MST kinase (Supplementary Fig. 5B). Therefore, all of the above mutations could affect *SAV1*-mediated downstream signal transduction.

### *SAV1* functions as a tumor-suppressor gene in ICC

Our results suggested a possible tumor-suppressor role for *SAV1* in ICC. To test that hypothesis, we performed WES on five ICC cell lines (HuCCT1, CCLP1, RBE, HCCC-9810, and SG231). All of the ICC cell lines were identified as having WT *SAV1* using the same filter criteria used for the ICC samples. The WES results for the cell lines were validated by Sanger sequencing. Western blots and qRT-PCR confirmed that *SAV1* mRNA and protein levels in the five ICC cell lines, especially HCCC-9810 and SG231, were lower than those in H-69 cells (Fig. 5A).

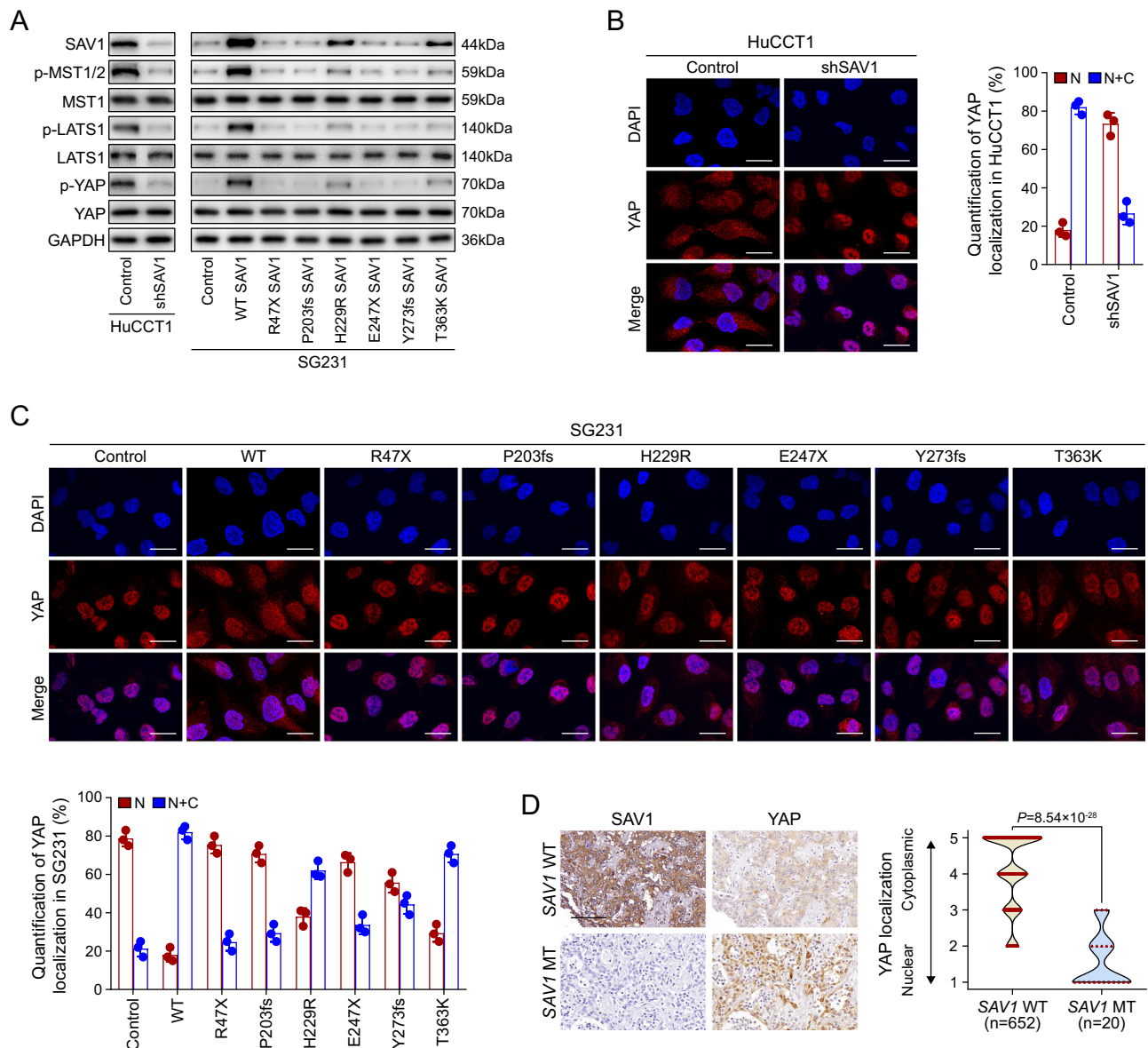
Next, we knocked down *SAV1* in HuCCT1 cells (Fig. 5A). Biofunctional investigations revealed that the knockdown of *SAV1* resulted in an increase in ICC cell proliferation, colony formation, and invasive ability (Fig. 5B–D). In vivo ICC mouse models showed that *SAV1* knockdown accelerated tumor growth and metastasis (Fig. 5E).

We generated lentiviral constructs to re-express WT and selected mutant *SAV1* variants in SG231 cells (Fig. 5A). The results showed that overexpression of WT *SAV1* substantially suppressed ICC cell proliferation, colony formation, and invasion ability. By contrast, expression of the *SAV1* mutants failed to cause these effects in whole or in part, suggesting that these *SAV1* mutations were inactivating mutations (Fig. 5B–D).



**Fig. 5 | Identification of SAV1 as a tumor-suppressor gene in ICC.** **A** SAV1 expression examined by qRT-PCR and western blot in five ICC cell lines (HuCCT1, CCLP1, RBE, HCCC-9810, and SG231), one human immortalized nonmalignant cholangiocyte cell line (H-69), and stably transfected cells,  $n = 3$ , errors are in  $\pm$ SD. **(B)** Proliferation of HuCCT1 cells after SAV1 knockdown and SG231 cells expressing wild-type or mutant SAV1 compared with that of controls,  $n = 4$ , errors are in  $\pm$ SD. **C** Colony formation activity of HuCCT1 cells after SAV1 knockdown and SG231 cells expressing wild-type or mutant SAV1 compared with that of controls. The bar graphs illustrate quantification of the colony formation assay. Student's  $t$ -test, two-sided,  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $n = 6$ , errors are in  $\pm$ SD. Scale bar: 400  $\mu$ m.

**D** Invasion of HuCCT1 cells after SAV1 knockdown and SG231 cells expressing wild-type or mutant SAV1 compared with that of controls. The graphs depict the number of invasive cells after 48 h. Student's  $t$ -test, two-sided,  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $n = 6$ , errors are in  $\pm$ SD. Scale bar: 100  $\mu$ m. **E** Representative bioluminescence images of mouse liver tumors and pulmonary metastasis, and H&E stained images of metastatic nodules in lungs. The color-scale bar depicts the photon flux emitted from the mice. Student's  $t$ -test, two-sided,  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $n = 6$ , errors are in  $\pm$ SD. Scale bar: 100  $\mu$ m. Con control, WT wild-type, sh short hairpin RNA. Source data are provided as a Source Data file.



**Fig. 6 | SAV1 inactivation suppresses Hippo signaling, leading to YAP activation.** **A** Western blot showed the expression of p-YAP, YAP, and the YAP upstream kinases pMST1/2, MST1, pLATS1, and LATS1 in ICC cells upon alteration of SAV1 expression. **B, C** Immunofluorescence staining showing subcellular YAP localization in the indicated cells. The graphs depict the percentages of cells with exclusively nuclear (N) YAP and with both nuclear and cytoplasmic (N + C) YAP. No cells showed YAP that was localized exclusively in the cytoplasm,  $n = 3$ , errors are in  $\pm$ SD.

Scale bar: 20  $\mu$ m. **D** Representative SAV1 and YAP staining in tumor tissues with wild-type or mutant SAV1. Scale bars = 50  $\mu$ m. The graphs depict the correlation between YAP nuclear localization and SAV1 mutation in ICC specimens. The subcellular localization of YAP was scored from 1 to 5, with 1 representing exclusive nuclear localization and 5 representing exclusive cytoplasmic localization. Student's  $t$ -test, two-sided. sh short hairpin RNA, WT wild-type, MT mutation-type. Source data are provided as a Source Data file.

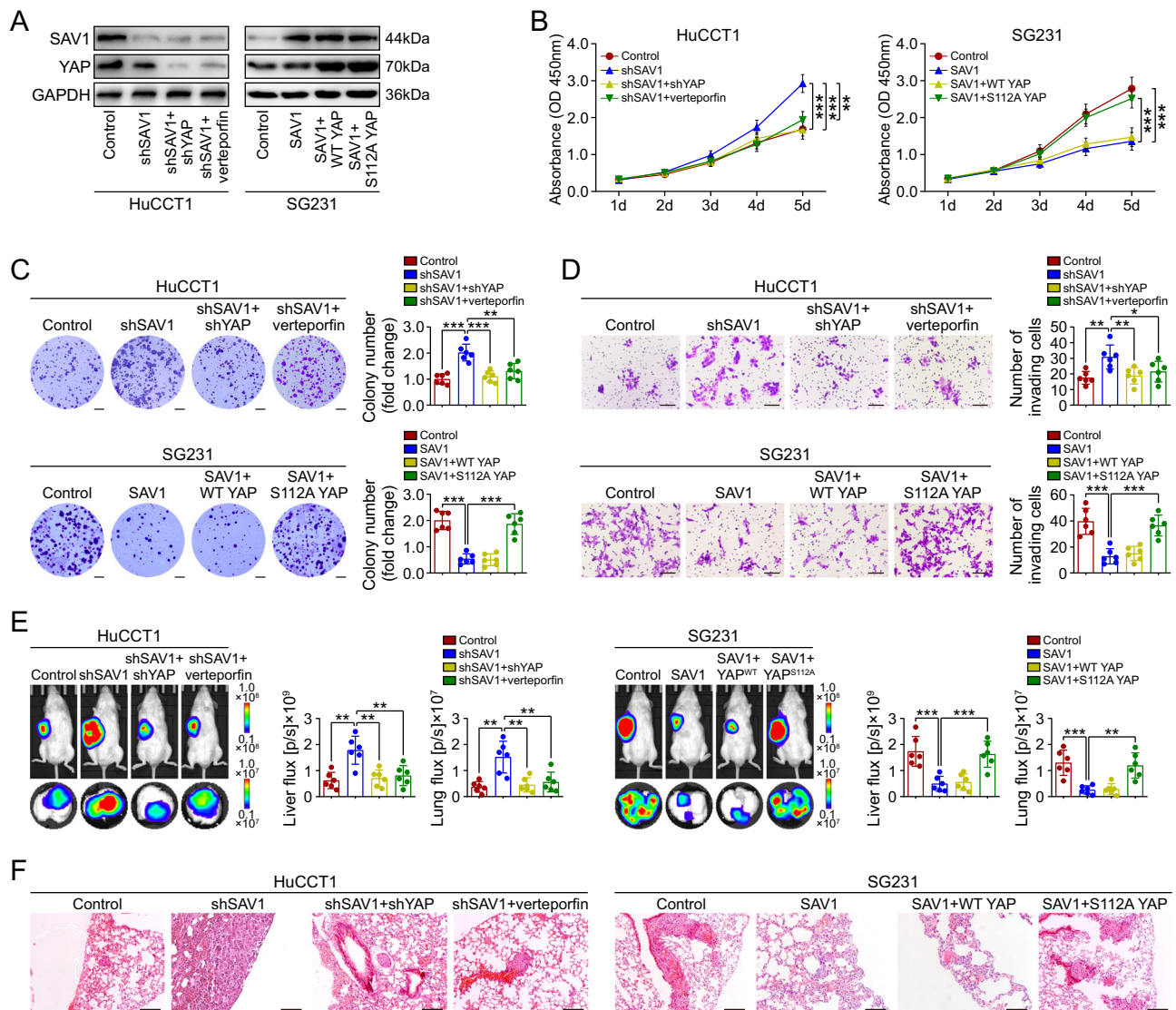
Next, we used the SAV1 mutants together with WT SAV1 for in vivo study. In agreement with the in vitro results, WT SAV1 significantly suppressed tumor growth and metastasis, whereas the SAV1 mutants yielded a larger tumor volume and increased metastasis compared with the WT SAV1 (Fig. 5E). These results further suggested that the SAV1 mutations were inactivating types, supporting the notion that SAV1 is a tumor-suppressor gene in ICC and that certain somatic mutations abolish its function and its tumor-inhibitory effect.

#### SAV1 inactivation suppresses Hippo signaling, leading to YAP activation

In mammals, the scaffold protein SAV1 promotes activation of the Hippo kinase cascade<sup>20</sup>. Therefore, we tested the effect of SAV1 on Hippo signaling in ICC cells. Western blot analysis showed that

knockdown of SAV1 in HuCCT1 cells reduced YAP Ser127 phosphorylation. Conversely, ectopic expression of SAV1 increased YAP Ser127 phosphorylation in SG231 cells (Fig. 6A). Consistent with the YAP phosphorylation, the phosphorylation of the cytosolic kinases upstream of YAP (MST1, MST2, and LATS1) was also decreased in HuCCT1 cells after knockdown of SAV1, whereas it was increased in SG231 cells after restoration of SAV1 expression (Fig. 6A). Immunofluorescence staining showed that knockdown of SAV1 resulted in nuclear accumulation of YAP in HuCCT1 cells, whereas overexpression of SAV1 in SG231 cells led to increased levels of cytoplasmic YAP and reduced levels of nuclear YAP (Fig. 6B, C). These results suggested that SAV1 triggers a Hippo kinase cascade that leads to phosphorylation, cytoplasmic retention, and inactivation of YAP. We also confirmed that compared with WT SAV1, mutated SAV1 failed to trigger this kinase





**Fig. 7 | SAV1 inactivation promotes ICC growth and metastasis by activating YAP.** **A** Western blot showed the expression of SAV1 and YAP in stably transfected ICC cells. **B** Proliferation of YAP knockdown or verteporfin (1  $\mu$ g/ml)-treated HuCCT1 cells following SAV1 knockdown and of SG231 cells expressing wild-type YAP or S112A YAP following SAV1 overexpression compared with that of controls. Student's *t*-test, two-sided,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$ ,  $n = 4$ , errors are in  $\pm$ SD. **C** Colony formation activity of YAP-knockdown or verteporfin-treated HuCCT1 cells following SAV1 knockdown and of SG231 cells expressing wild-type YAP or S112A YAP following SAV1 overexpression compared with that of controls. The bar graphs illustrate quantification of the colony formation assay. Student's *t*-test, two-sided,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$ ,  $n = 6$ , errors are in  $\pm$ SD. Scale bar: 400  $\mu$ m. **D** Invasion

activity of YAP-knockdown or verteporfin-treated HuCCT1 cells following SAV1 knockdown and of SG231 cells expressing wild-type YAP or S112A YAP following SAV1 overexpression compared with that of controls. The graphs depict the number of invasive cells after 48 h. Student's *t*-test, two-sided,  $^{*}P < 0.05$ ,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$ ,  $n = 6$ , errors are in  $\pm$ SD. Scale bar: 100  $\mu$ m. **E**, **F** Representative bioluminescence images of mouse liver tumors and pulmonary metastasis, and H&E stained images of metastatic nodules in lungs. The color-scale bar depicts the photon flux emitted from the mice. Student's *t*-test, two-sided,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$ ,  $n = 6$ , errors are in  $\pm$ SD. Scale bar: 100  $\mu$ m. sh short hairpin RNA, WT wild-type, MT mutation-type. Source data are provided as a Source Data file.

cascade, which suggests that SAV1 positively modulates Hippo signaling and that its inactivation due to mutation suppresses Hippo signaling, leading to YAP activation in ICC cells (Fig. 6A–C). Clinical ICC samples with SAV1 somatic mutation tended to display nuclear accumulation of YAP, as revealed by immunohistochemistry staining (Fig. 6D), which further indicated that SAV1 inactivation suppresses Hippo signaling and leads to YAP activation in ICC cells. Furthermore, YAP knockdown or verteporfin (a YAP inhibitor) treatment alleviated the SAV1 knockdown-induced increases in in vitro cell proliferation, colony formation, and invasion (Fig. 7A–D) and in vivo tumor growth and metastasis in HuCCT1 cells (Fig. 7E, F). Moreover, in SG231 cells, overexpression of S112A mutant YAP (a nonphosphorylatable YAP mutant), but not that of WT YAP, reversed the inhibitory effect of SAV1

on in vitro cell proliferation, colony formation, and invasion and in vivo tumor growth and metastasis (Fig. 7B–F). These results suggest that SAV1 inactivation promotes ICC growth and metastasis by activating YAP.

## Discussion

In this study, we delineated the genomic landscape, including somatic SNVs/indels and CNV, in 204 patients with ICC. Through NMF and hierarchical clustering, we identified the Msig-cluster 1 and 6 groups, which were characterized by Signature A, and the Msig-cluster 4 group, which was characterized by Signature F. Signatures A and F were highly similar to the previously described Signature 22 (aristolochic acid) and 24 (aflatoxin), respectively, which is different than the results of

previous ICC cohort studies conducted in other countries and regions<sup>9–11</sup>. Exposures to aristolochic acid or aflatoxin have been revealed as important environmental risk factors for HCC development, especially in China, and are known to contribute to hypermutation in cancer<sup>11,12,21</sup>. Our results confirm that exposures to aristolochic acid or aflatoxin are also important environmental risk factors for ICC development in China, and that the SNV burden is higher in ICCs potentially linked to aristolochic acid or aflatoxin exposure than in other ICCs. In addition, we identified Signature D mainly in four ICC samples with higher numbers of indels, indicating similarity to the previously described Signature 20, which was confirmed to be driven by defective DNA mismatch repair. Hence, we propose that the four patients with Signature D would likely benefit from immunotherapy.

We identified 13 genes that were significantly mutated in ICC. About half of these genes, including *TP53*, *KRAS*, and *IDH1*, presented a different mutation frequency in our ICC cohort than in a combined TCGA and MSK cohort. We propose that this might be due to differences in racial or etiological factors. We also identified mutations in *TP53*, *KRAS*, *ELF3*, and *SAVI* that were correlated with patient OS and/or RFS. The combined mutational profile of these four genes showed strong predictive value and correlated with OS and RFS. Although some other genes such as *IDH1*, *BAP1* and *PBRM1* have been confirmed as drivers of ICC, they have no ability to predict ICC prognosis. Our results suggest that *TP53*, *KRAS*, *ELF3*, and *SAVI* may play roles in ICC growth and metastasis.

We also demonstrated the role of clonal diversity in predicting ICC outcomes. We showed that clonal dominance could independently confer a higher risk of recurrence and shorter OS, which is consistent with the results of previous studies<sup>14,22</sup>. Although some other studies found that higher numbers of clones were associated with worse patient prognosis<sup>23,24</sup>, they did not investigate the contribution of relative clone size to prognosis, which might explain the contradiction between previous findings and our results. We propose that the presence of a dominant clone at the time of diagnosis could reflect the capacity of ICC to undergo rapid clonal turnover upon acquisition of a newly acquired adaptive mutation, leading to tumor recurrence<sup>14</sup>.

Among the genes whose mutational frequency correlated with patient OS and/or RFS, *SAVI*, which was mutated in about 3% of the 672 ICC samples in our cohort, was not previously known as a driver of ICC. We did not find *SAVI* to be significantly mutated in the TCGA cohort, possibly because of limited sample size, low sequencing coverage, or differences in racial or etiological factors. A search of cancer genomics datasets available at [www.cbioportal.org](http://www.cbioportal.org) showed that apart from two nonsense mutations (C697T and C724T), the other *SAVI* somatic mutations were identified in our ICC cohort. Furthermore, we found that the majority of *SAVI* mutations in our cohort were truncating or splice-site mutations. Although lack of biochemical experiments to support our structure model, through structural analysis and inference, we speculatively predicted that these mutations can block the normal downstream signal transduction process.

In addition to these truncating or splice-site mutations, we also identified 3 missense mutations: S94F, H229, and T363. H229 is located near the junction between the WW1 and WW2 domains. In wild-type (WT) *SAVI*, the side chain of H229 interacts with L199 and L231 through two stable hydrogen bonds (Supplementary Fig. 5A). L199 is located at the beginning of the WW1 domain, and the interaction between L199 and H229 contributes to the stability of WW1 conformation. L231, along with H229, is located in a loop region that also contains the R233 residue, which is important for the dimerization function of WW2 (Supplementary Fig. 5A, B)<sup>19</sup>. An R233Q mutation was previously shown to weaken the dimerization of WW1<sup>19</sup>. Similarly, in the H229R mutant that we identified in ICC, the flexibility of the side chain of the substituted R229 residue is predicted to weaken the ability of the residue to form stable hydrogen bonds with L199 and L231. The resulting lack of hydrogen bond interaction between R229 and L199 probably

weakens the conformational stability of the WW1 domain. Moreover, the lack of interaction between R229 and L231 is predicted to affect the stability of the loop where R233 is located (Supplementary Fig. 5A, B), thus affecting the ability of R233 to stabilize the dimerization of WW2 and preventing *SAVI* from forming a stable heterotetramer with *MST2*. The reduced phosphorylation function of the unstable heterotetramer results in disruption of the normal downstream signal transduction process.

T363 is located in the middle of the SARA domain. The side chain of T363 interacts with the Q359 residue through a stable hydrogen bond (Supplementary Fig. 5A). Although T363 is not a part of the interface where *SAVI* interacts with *MST2*, the hydrogen bond between T363 and Q359 is important for the secondary structure stability of the  $\alpha$ -helix of the SARA domain, which supports the hydrophobic interaction between L361 and L365 of *SAVI* and L448 and L452 of *MST2*<sup>17</sup>. Moreover, previous studies showed that the stability of the SARA structure is essential for the interaction between *SAVI* and *MST2* kinase<sup>18</sup>. In the T363K mutant, K363 is predicted not to form a stable hydrogen bond with Q359, which is expected to weaken the secondary structure stability of the SARA domain, further weaken the stability of the interaction between *SAVI* and *MST2* to form a heterotetramer, and thus influence the activation of the Hippo pathway.

*SAVI* is a human homolog of *Salvador* that acts as a scaffolding protein and contains two protein–protein interaction modules known as WW domains<sup>19</sup>. Down-regulation of *SAVI* in several types of cancer including HCC<sup>25</sup>, clear cell renal cell carcinoma<sup>26</sup>, and colorectal cancer<sup>27</sup> can promote cell proliferation, anti-apoptosis, migration, and invasion. Furthermore, previous studies in fish and flies have shown a cholangio-suppressive role of *SAVI* orthologs<sup>28,29</sup>. The functional impact of genetic mechanisms affecting *SAVI* is still unknown, however, especially in cancers such as ICC. Although a previous TCGA publication identified *SAVI* deletions in 5% of ICC patients, the sample size was small, with only two *SAVI* deletions identified in two patients<sup>30</sup>. Another previous study revealed focal deletions at 14q22.1, which contain coding regions of *SAVI* and occur in 12% of ICCs<sup>31</sup>, our results did not identify recurrent genomic deletions involving *SAVI* (Supplementary Fig. 1A, B; Supplementary Data 3), which suggests that CNV might not be the main factor that leads to *SAVI* inactivation. Nevertheless, we showed that *SAVI* was significantly mutated in our ICC cohort, resulting reduced *SAVI* expression. Moreover, we demonstrated that *SAVI* somatic mutation and down-regulation were both predictive of tumor recurrence and shorter OS. Importantly, we revealed through gain-of-function and loss-of-function studies that *SAVI* plays a tumor-suppressor role in ICC. We also confirmed that *SAVI* inactivation by somatic mutation promotes ICC growth and metastasis.

*SAVI* is one of the core kinase components of the Hippo signaling pathway in mammals<sup>20</sup>. Our results further confirm its role in modulating the Hippo pathway in ICC: *SAVI* inactivation suppresses Hippo signaling, leading to YAP activation. Through in vitro and in vivo functional analyses, we confirmed that *SAVI* inactivation promotes ICC growth and metastasis by activating YAP. Inactivation of the Hippo pathway has been shown to be associated with poor prognosis in HCC<sup>25</sup>, yet its role in ICC was largely unknown. Our results revealed that *SAVI* mutation leads to inactivation of the Hippo pathway, which in turn promotes ICC growth and metastasis. In addition, our results revealed a tumor-suppressor role for verteporfin, which exhibited an effect similar to that of YAP knockdown in *SAVI*-inactivated ICC cells. Verteporfin is already approved by the FDA for eye disease indications such as macular degeneration, so it would be relatively easy to adopt for ICC treatment. Hence, our finding of ICC inhibition by verteporfin offers exciting possibilities for the treatment of *SAVI*-mutated ICC. Notably, the FDA-approved application of verteporfin is based on photodynamic therapy to eliminate neovascularization of blood vessels. Light activation is required for verteporfin to inhibit neovascular

angiogenesis, but it is not required for verteporfin to disrupt the interaction between YAP and TEAD<sup>32</sup>. Therefore, verteporfin might have a dual function inhibiting both angiogenesis and YAP activity, both of which can contribute to inhibiting the progression of SAV1-mutated ICC.

Taken together, our results delineate the genomic landscape that characterizes Chinese ICCs and identify SAV1 as a potential driver.

## Methods

### Patients and follow-up

The Research Ethics Committee of Zhongshan Hospital granted ethical approval for the use of human subjects (Y2018-060). Each participant gave informed consent in signed form for inclusion in the study. We enrolled and continuously collected tumor and matched non-cancerous liver-tissue samples from 672 Chinese patients with primary ICC who received curative resection from 2010 to 2016 in the Department of Liver Surgical Oncology of Zhongshan Hospital, Fudan University, Shanghai, China. Patients that received palliative surgeries or prior interventions (such as trans-hepatic artery embolization, chemotherapy, or radiotherapy) or that developed other primary malignancies or inflammatory diseases during follow-up were excluded from the study. A total of 204 ICCs were subjected to WES, and the remaining 468 ICCs were subjected to Sanger sequencing. Curative resection was defined as complete resection of tumor nodules, with cancer-free tumor margins shown by histologic examination, and resection of regional lymph nodes, including the hilar, hepatoduodenal-ligament, and caval lymph nodes, with no cancerous thrombus in the portal vein (main trunk or major branches), hepatic veins, or bile duct<sup>33</sup>. Patients with further lymph node involvement were considered to have distant metastasis and were excluded from the study<sup>34</sup>. Tumor differentiation was graded histologically according to the Edmondson–Steiner criteria<sup>35</sup>. Liver function was graded according to the Child–Pugh system. Tumor stage was determined according to the 2017 International Union against Cancer TNM system. The clinicopathologic characteristics of the patients are listed in Table 1.

The present study includes follow-up data collected through December 2018. The follow-up procedures are described in detail elsewhere<sup>36,37</sup>. We diagnosed tumor recurrence on the basis of computed tomography scans, magnetic resonance imaging, digital subtraction angiography, and elevated serum carbohydrate antigen 19-9 (CA19-9) level, with or without histological confirmation<sup>38</sup>. We defined disease-free survival as the interval between curative surgery and any diagnosis of recurrence (intrahepatic or extrahepatic)<sup>39</sup>. We defined overall survival (OS) as the time from the date of curative surgery until death or the end of follow-up. The surviving patients were censored at the end of follow-up.

### DNA preparation, DNA capture, and sequencing

Snap frozen tissue samples from tumor and matched non-cancerous liver were obtained and embedded in OCT compound, sectioned by a cryostat, and stained by hematoxylin and eosin. We performed macrodissection to enrich the tumor fraction relative to the dominant stromal component and other normal cells. DNA was extracted using a general protocol for genome sequencing. Preparation of sequencing libraries and DNA capture methods were carried out according to the manufacturer's protocols.

### Whole-exome sequencing

Snap-frozen tissue samples from 204 primary tumors and matched non-cancerous liver from 204 ICC patients were subjected to whole exome sequencing. Sequencing libraries were constructed using a modification of the KAPA Library Preparation Kit. Briefly, 1 µg genomic DNA was sheared to an average fragment size of 200 bp using a Biorupter (Diagenode, Belgium). Fragments were purified using AMPureXP beads (Beckman Coulter Inc, Brea, CA, USA) to remove small products.

Then, the DNA was subjected to three enzymatic steps: end repair, A-tailing, and ligation to Illumina paired-end indexed adapters, as outlined in the KAPA Library Preparation Kit protocol (Kapa, Inc.). The libraries were subjected to minimal PCR cycling and were quantified using the Qubit 2.0 Fluorometer (Thermo Fisher, Germany). Libraries were combined into pools for solution phase hybridization using the Roche NimbleGen SeqCap EZ exome enrichment kits V3 (Roche NimbleGen Inc). The captured libraries were analyzed on Agilent's 2100 Bioanalyzer, and the DNA concentrations were measured using the Qubit 2.0 Fluorometer (Thermo Fisher, Germany) and then sent for sequencing to generate 2 × 150 bp paired-end reads using a HiSeq or Novaseq platform (Illumina, Inc.).

### Data quality control

Sequence artifacts; including reads containing adapter contamination, low-quality nucleotides, and unrecognizable nucleotides (N); undoubtedly set a barrier to subsequent reliable bioinformatics analysis. Hence, quality control is an essential step and is applied to guarantee a meaningful downstream analysis. The steps of data processing were as follows:

- (1) Discard a pair of reads if either read contains adapter contamination;
- (2) Discard a pair of reads if it contains poly-N;
- (3) Discard a pair of reads if the proportion of low-quality (Phred quality <5) bases is over 50% in either read.

### Reads mapping and detection of somatic genetic alterations

Valid sequencing data was mapped to the reference human genome (UCSC hg19) using the Burrows-Wheeler Aligner (BWA)<sup>40</sup> software to get the original mapping results stored in BAM format. We performed local realignment of the original BAM alignment using the GATK2<sup>41</sup> and then marked duplicate reads using Sambamba<sup>42</sup>.

Somatic SNVs were detected by muTect1 and muTect2<sup>43</sup>, and somatic indels were detected by Strelka<sup>44</sup>. High-confidence somatic mutations were called if the following criteria were met: (1) the tumor samples were covered sufficiently ( $\geq 10\times$ ) and the normal samples were covered  $\geq 8\times$  at the genomic level; (2) the variants were supported by at least 4% of the total reads in the tumor and less than 1% of the total reads in the normal tissue; (3) the variants were supported by at least three reads in the tumor. (4) Only SNVs agreed on by both tools (muTect1 and muTect2) were retained. ANNOVAR<sup>4</sup> was performed to do annotation for the Variant Call Format obtained in the previous effort.

### Tumor copy number variation detection, ploidy, and purity prediction

We used FACETS<sup>45</sup>, an allele-specific copy number analysis tool to detect genome-wide total, allele-specific, and integer DNA copy number, and predict tumor ploidy. The GISTIC (V2.0) algorithm was used to infer recurrently amplified or deleted genomic regions<sup>46</sup>. Genes with a total copy number greater than the gene-level median ploidy were considered gains. Genes with more than twice the median ploidy were considered amplifications. Genes with less than the median ploidy were considered losses. Genes with a total copy number of 0 were considered deletions.

### Spectrum and signatures of somatic mutations

We grouped all identified somatic single-nucleotide variants (SNVs) according to the 16 possible flanking nucleotide contexts into 96 trinucleotides (mutated base plus its sequence context)<sup>47,48</sup>. We then used a non-negative matrix factorization (NMF) module to decipher the underlying mutation signatures<sup>49</sup>. The overall number of mutation signatures was selected based on the reproducibility of the signatures and low overall reconstruction error. We then compared the inferred signatures to the COSMIC signatures by cosine similarity. We

**Table 1 | Clinicopathologic characteristics of patients with intrahepatic cholangiocarcinoma**

Variable		WES (n = 204)		Sanger sequencing (n = 468)	
		N	%	N	%
Age (year)	≤50	41	20.1	84	17.9
	>50	163	79.9	384	82.1
Sex	Female	82	40.2	200	42.7
	Male	122	59.8	268	57.3
HBsAg	Negative	137	67.2	324	69.2
	Positive	67	32.8	144	30.8
Liver fluke	Negative	193	94.6	437	93.4
	Positive	11	5.4	31	6.6
CA19-9	≤36	94	46.1	198	42.3
	>36	110	53.9	270	57.7
GGT(U/L)	≤54	110	53.9	232	49.6
	>54	94	46.1	236	50.4
Liver cirrhosis	No	148	72.5	353	75.4
	yes	56	27.5	115	24.6
Tumor size(cm)	≤5	100	49.0	216	46.2
	>5	104	51.0	252	53.8
Tumor number	Single	157	77.0	339	72.4
	Multiple	47	23.0	129	27.6
Microvascular/Bile duct invasion	No	156	76.5	372	79.5
	Yes	48	23.5	96	20.5
Lymphatic metastasis	No	176	86.3	399	85.3
	Yes	28	13.7	69	14.7
Tumor encapsulation	complete	32	15.7	69	14.7
	none	172	84.3	399	85.3
Tumor differentiation	I + II	86	42.2	211	45.1
	III + IV	118	57.8	257	54.9
TNM stage	I	109	53.4	246	52.6
	II	65	31.9	144	30.8
	III	29	14.2	73	15.6
	IV	1	0.5	5	1.1

performed hierarchical clustering of the tumors based on the contributions of the mutational signatures in each patient<sup>50</sup>.

### Significantly mutated genes

For significantly mutated genes analysis, all identified somatic SNVs and indels were subjected MutSigCV<sup>13</sup> to identify significantly mutated genes (false discovery rate [FDR] <0.1).

### Cancer cell fraction estimation and mutation cluster analysis

For each somatic mutation, the VAF was calculated using the number of reads supporting the variant allele (Rmut) and the number of reads supporting the reference allele (Rnorm; namely,  $VAF = Rmut / (Rmut + Rnorm)$ ). Then, we calculated the value of cancer cell fractions (CCF) as follows:  $VAF = p * CCF / (CPNnorm * (1-p) + p * CPNmut)$ , where CPNmut indicates the local copy number in the tumor, CPNnorm indicates the local copy number in the normal controls (usually assumed to be 2 except for sex chromosomes), and p indicates the tumor purity in each sequenced sample. The VAF was defined as the VAF of each somatic mutation. CCF is represented as a distribution between 0 and 1. Then all somatic mutations (SNVs and indels) identified were applied to infer the mutation cluster according to PyClone (a Bayesian clustering method)<sup>51</sup>.

### Estimation of Shannon index

To assess the clonal diversity in each patient, we used the Shannon index of ecological diversity<sup>52</sup>. After we inferred mutation clusters using PyClone, we filtered out clusters with a single mutation and those involving only silent mutations. We then used the remaining mutation clusters to quantitate the Shannon diversity index in each patient<sup>14</sup>.

In this analysis, mutations were assumed to accumulate in a linear way, with each mutation giving rise to a new cell population<sup>14</sup>. We ordered the mutation clusters by increasing cancer cell fraction (CCF) in each patient. The raw size  $s_i$  of each of the  $N$  mutations in each patient was estimated as follows:

$$s_1 = CCF_1 \quad (1)$$

$$s_i = CCF_i - CCF_{i-1} \quad (2)$$

The relative abundance of each population  $p_i$  was then computed as the size of the mutation  $s_i$  divided by the sum of the sizes of all  $N$  mutations in the patient.

$$p_i = \frac{s_i}{\sum_i s_i} \quad (3)$$

The resulting relative abundances  $p_i$  were then used to derive the Shannon diversity index for the patient.

$$Shannon\ Index = -\sum_i p_i \log(p_i) \quad (4)$$

### Sanger sequencing

We randomly selected 1600 somatic non-synonymous SNVs and 156 indels identified by WES. We used Sanger sequencing to validate the selected mutations. All mutations presented in Fig. 2A were also validated by Sanger sequencing. In addition, all coding exons of SAV1 were further screened in an additional 468 ICC samples. Sanger sequencing primers were designed using the Primer3 software (<http://frodo.wi.mit.edu/>). All mutations identified in tumors were confirmed by independent PCR and Sanger sequencing in the specific tumors and paired normal tissues to determine their somatic nature.

### Modeling of the SAV1 and SAV1–MST2 complex structures

Because there was no previously determined full-length SAV1 structure, we obtained the full-length structure of SAV1-HUMAN based on the prediction provided in the AlphaFold predicted-structure database (<https://alphafold.ebi.ac.uk/>). There was no stable secondary structure for residues 1–197 in the predicted structure, so only the structure of residues 198–383 was shown in the present study. In addition, because of the flexibility of the connection-region loop between the WW1 and WW2 domains, we adjusted the deflection angle of the connection region between the WW1 and WW2 domains with reference to the highly homologous crystal structure of the WW12 domain of mouse SAV1 (PDB code: 7BQG). Based on the crystal structure of the dimer in the WW12 domain (PDB Code: 7BQG) and the crystal structure of the SAV1–MST2 heterodimer complex (PDB Code: 6A05), we built a structural model of the SAV1–MST2 heterotetramer to more intuitively understand the significance of WW12 dimerization to the function of SAV1. The structural model of the SAV1–MST2 heterotetramer was generated in PyMOL (<https://pymol.org/2/>). Specifically, to obtain the WW12 domain dimer as described in previous work<sup>19</sup>, a WW12 domain dimer was generated using the generate symmetry mates module in PyMOL. Then, the SAV1 structure modeled by AlphaFold was aligned with the WW12 domain dimer in PyMOL to obtain the SAV1 dimer. Subsequently, after

aligning the SAV1-MST2 heterodimer complex (PDB Code: 6AO5) with the SAV1 dimer, the angle of the flexible loop between the WW12 domain and the SARAH domain of SAV1 was manually adjusted in PyMOL to ensure that the distance between the two MST2 molecules was reasonable. Finally, a schematic diagram of the structure of the SAV1-MST2 heterotetramer was presented.

### Cell lines, animals, and lentiviral vector

One human immortalized, nonmalignant cholangiocyte cell line: H-69, and five ICC cell lines were used in this study: H-69, HuCCT1, SG231, and CCLP1 were kindly provided from Dr. Robert Anders at Johns Hopkins, while HCCC-9810 and RBE cell lines were purchased from the Institute of Biochemistry and Cell Biology (Chinese Academy of Sciences). Four-to-six-week-old male NOD-*Prkdc<sup>scid</sup> IL2rg<sup>tm1</sup>/Bcgen* mice were obtained from the Beijing Biocytogen Co., Ltd and maintained under specific pathogen-free conditions. Humane care was provided for all animals in concordance with the criteria described in the “Guide for the Care and Use of Laboratory Animals” (National Institutes of Health publication 86-23, revised 1985). The maximal tumor size permitted by The Research Ethics Committee of Zhongshan Hospital is 2 cm in length for subcutaneous tumor and the tumors of all animals in this study didn't exceed the criteria.

The following lentiviral vectors were purchased from Shanghai GeneChem Co: wild-type and mutant SAV1 expression vectors and corresponding control lentiviral vector (Ubi-MCS-SV40-firefly\_Luciferase-IRES-Puromycin), and the shRNA-SAV1 and its negative control (hU6-MCS-Ubiquitin-firefly\_Luciferase-IRES-puromycin). The Ubi-MCS-SV40-firefly\_Luciferase-IRES-Puromycin-SAV1 wild-type and mutant vectors were transfected into SG231 cells with low intrinsic SAV1 levels. The hU6-MCS-Ubiquitin-firefly\_Luciferase-IRES-puromycin-shRNA-SAV1 was transfected into HuCCT1 cells with high intrinsic SAV1 expression. The Ubi-MCS-SV40-firefly\_Luciferase-IRES-Puromycin and hU6-MCS-Ubiquitin-firefly\_Luciferase-IRES-puromycin lentiviral vectors were used as controls. Stably transfected clones were validated by immunoblotting.

### Cell proliferation, colony formation, and matrigel invasion assays

Cells (2000 cells/well) were seeded in 100  $\mu$ L of media in a 96-well plate. Then, 10  $\mu$ L CCK-8 solution (Dojindo) was added to the cells at the indicated time points. The cells were then incubated for an additional 2 h. The numbers of viable cells were determined by measurement of absorbance at 450 nm.

To assess the colony formation abilities of the cells, 500–1000 cells were seeded into each well of six-well plates and incubated at 37 °C for 12–16 days. Then, the cells were fixed with 100% methanol and stained with 0.1% crystal violet. Image-Pro Plus v6.2 (Media Cybernetics) was used to count the megascopic cell colonies.

To assay cell invasion, 24-well Transwell plates with an 8  $\mu$ m pore size (Minipore) were precoated with Matrigel (BD Biosciences). Then, 100  $\mu$ L Dulbecco's modified Eagle medium (DMEM) with 1% fetal bovine serum (FBS) containing  $1 \times 10^5$  cells was added to the upper chamber. The lower chamber contained 600  $\mu$ L DMEM with 10% FBS. After 48 h, both the Matrigel and the remaining cells in the upper chamber were removed. Cells that had invaded the lower surface of the membrane were fixed using 4% paraformaldehyde and then stained with Giemsa. Cells from five microscopic (200 $\times$ ) fields were counted.

### In vivo assays for tumor growth and metastasis

For tumor growth assay, 100  $\mu$ L serum-free DMEM and Matrigel (BD Biosciences; 1:1) containing  $1 \times 10^7$  ICC cells were injected subcutaneously into the upper left flank region of mice. For tumor metastasis assay, mice were placed in a restrainer, and tumor cells ( $1 \times 10^6$  cells in 200  $\mu$ L serum-

free DMEM) were injected through the tail vein using an insulin needle. In some experiment groups for verteporfin treatment, intraperitoneal injection of verteporfin (100 mg/kg) thrice weekly was performed starting 7 days after inoculation. The mice were monitored every 5 days and were sacrificed after 5 weeks. Bioluminescence imaging was performed using an IVIS Lumina K Series III, and image radiance values were normalized using Living Image (Perkinelmer).

### RNA isolation and qRT-PCR

Total RNA was extracted from cells using Trizol reagent (Invitrogen) according to the manufacturer's instructions. The mRNA expression in the cells was assessed with qRT-PCR using an ABI7900HT instrument (Applied Biosystems). Quantitative RT-PCR (qRT-PCR) was performed using the SYBR PrimeScript RT-PCR Kit (Takara Bio). GAPDH was used as an internal control for qRT-PCR. Relative mRNA levels were estimated on the basis of Ct values and normalized according to GAPDH expression using the following equation:  $2^{-\Delta Ct} (\Delta Ct = Ct [\text{target gene}] - Ct [\text{GAPDH}])$ . All experiments were performed in triplicate.

### Western blot and immunofluorescence analysis

Western blotting was performed as described previously<sup>53</sup>. Briefly, proteins from total cell lysates were separated by 10% SDS-PAGE. The proteins were then transferred to polyvinylidene difluoride (PVDF) membranes. After washing and blocking steps, the membranes were incubated with primary antibodies. After washing, the membranes were incubated with horseradish peroxidase-conjugated secondary antibodies. Antibody binding was detected using enhanced chemiluminescence assays.

For immunofluorescence assays, cells cultured on glass slides were fixed in 4% paraformaldehyde for 15 min. Subsequently, the cells were permeabilized with 0.1% Triton X-100 for 15 min at room temperature, washed with phosphate buffered saline (PBS), and blocked with PBS containing 1% (w/v) bovine serum albumin (BSA) and 0.15% (w/v) glycine (BSA buffer) for 1 h at room temperature. The cells were then treated with primary antibody for 2 h at room temperature. A negative control (primary antibody omitted) was included on every slide. The cells were then washed with BSA buffer and incubated with 2  $\mu$ g/mL Alexa Fluor 488-conjugated goat anti-mouse antibody (Molecular Probes, Eugene, OR) for 1 h at room temperature. After rinsing in PBS, the slices were counter-stained with diamidino phenylindole and examined by fluorescence microscopy (Leica Microsystems Imaging Solutions, Cambridge, UK).

### Immunohistochemistry (IHC) and evaluation of immunohistochemical variables

IHC staining was performed using an avidin-biotin-peroxidase complex as described previously<sup>54</sup>. Briefly, rehydration and microwave antigen retrieval were performed. Then, the slides were incubated with monoclonal antibodies at 4 °C overnight. Next, the slides were incubated with secondary antibody (GK500705, Gene Tech) at 37 °C for 30 min. The slides were stained with 3,3'-diaminobenzidine and then counter-stained with Mayer's hematoxylin. Slides that were treated identically but without primary antibody served as negative controls.

IHC staining was analyzed by three independent investigators who were blinded to patient characteristics. Any discrepancies were resolved by consensus. Five representative microscope fields were photographed under high-power magnification (200 $\times$ ) using the Leica QWin Plus v3 software. Each image was captured using identical settings. Image-Pro Plus v6.2 software (Media Cybernetics, Inc.) was used to determine density. In each photograph, the integrated optical density of positive SAV1 and YAP staining was determined. The ratio of that density to the total area of each photograph was calculated as the density.

### Statistical analysis

Statistical analyses were performed in the R 3.6.2 environment or using SPSS 16.0 for Windows. The data were expressed as the mean  $\pm$  SD of three independent experiments unless otherwise specified. Student's *t*-test was used to compare quantitative data between groups. The chi-square or Fisher's exact test was used to compare categorical data. The Kaplan–Meier method was used to calculate both the OS and the cumulative recurrence rates. Differences were analyzed by the log-rank test. All statistical tests were two-sided, and  $P < 0.05$  were considered statistically significant.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) in National Genomics Data Center (Nucleic Acids Res 2022), China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA008513). The accession number HRA008513 contains all raw sequencing data (204 WES data) with restrictedly controlled access, which is a property of Zhongshan Hospital, Fudan University, provided to the scientists or researchers through a restricted-access agreement that prevents sharing the dataset with a third party or publicly. Researchers could send data requests for scientific purpose via the website [<https://bigd.big.ac.cn/gsa-human/browse/HRA008513>] after registration. The correspondence author will check the requests with caution and make the decision whether to permit or not, in 1 month upon receiving the requests. The data will be available for 1 month once access has been granted by the correspondence author. Source data are provided with this paper.

### References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 7–34 (2019).
2. Bridgewater, J. et al. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J. Hepatol.* **60**, 1268–1289 (2014).
3. Rizvi, S. & Gores, G. J. Pathogenesis, diagnosis, and management of cholangiocarcinoma. *Gastroenterology* **145**, 1215–1229 (2013).
4. Mavros, M. N., Economopoulos, K. P., Alexiou, V. G. & Pawlik, T. M. Treatment and prognosis for patients with intrahepatic cholangiocarcinoma: systematic review and meta-analysis. *JAMA Surg.* **149**, 565–574 (2014).
5. Ong, C. K. et al. Exome sequencing of liver fluke-associated cholangiocarcinoma. *Nat. Genet.* **44**, 690–693 (2012).
6. Jiao, Y. et al. Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas. *Nat. Genet.* **45**, 1470–1473 (2013).
7. Chan-On, W. et al. Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. *Nat. Genet.* **45**, 1474–1478 (2013).
8. Zou, S. et al. Mutational landscape of intrahepatic cholangiocarcinoma. *Nat. Commun.* **5**, 5696 (2014).
9. Nakamura, H. et al. Genomic spectra of biliary tract cancer. *Nat. Genet.* **47**, 1003–1010 (2015).
10. Jusakul, A. et al. Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. *Cancer Discov.* **7**, 1116–1135 (2017).
11. Wardell, C. P. et al. Genomic characterization of biliary tract cancers identifies driver genes and predisposing mutations. *J. Hepatol.* **68**, 959–969 (2018).
12. Ng, A. W. T. et al. Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. *Sci. Transl. Med.* **9**, eaan6446 (2017).
13. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
14. Cerrano, M. et al. Clonal dominance is an adverse prognostic factor in acute myeloid leukemia treated with intensive chemotherapy. *Leukemia* **35**, 712–723 (2021).
15. Maley, C. C. et al. Classifying the evolutionary and ecological features of neoplasms. *Nat. Rev. Cancer* **17**, 605–619 (2017).
16. Bertucci, F. et al. Genomic characterization of metastatic breast cancers. *Nature* **569**, 560–564 (2019).
17. Bae, S. J. et al. SAV1 promotes Hippo kinase activation through antagonizing the PP2A phosphatase STRIPAK. *eLife* **6**, e30278 (2017).
18. Callus, B. A., Verhagen, A. M. & Vaux, D. L. Association of mammalian sterile twenty kinases, Mst1 and Mst2, with hSalvador via C-terminal coiled-coil domains, leads to its stabilization and phosphorylation. *FEBS J.* **273**, 4264–4276 (2006).
19. Lin, Z., Xie, R., Guan, K. & Zhang, M. A WW tandem-mediated dimerization mode of SAV1 essential for Hippo signaling. *Cell Rep.* **32**, 108118 (2020).
20. Yu, F. X., Zhao, B. & Guan, K. L. Hippo pathway in organ size control, tissue homeostasis, and cancer. *Cell* **163**, 811–828 (2015).
21. Zhang, W. et al. Genetic features of aflatoxin-associated hepatocellular carcinoma. *Gastroenterology* **153**, 249–262.e242 (2017).
22. Turajlic, S. et al. Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. *Cell* **173**, 581–594.e512 (2018).
23. Andor, N. et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* **22**, 105–113 (2016).
24. Bochtler, T. et al. Clonal heterogeneity as detected by metaphase karyotyping is an indicator of poor prognosis in acute myeloid leukemia. *J. Clin. Oncol.* **31**, 3898–3905 (2013).
25. Sohn, B. H. et al. Inactivation of Hippo pathway is significantly associated with poor prognosis in hepatocellular carcinoma. *Clin. Cancer Res.* **22**, 1256–1264 (2016).
26. Matsuura, K. et al. Downregulation of SAV1 plays a role in pathogenesis of high-grade clear cell renal cell carcinoma. *BMC Cancer* **11**, 523 (2011).
27. Jiang, J. et al. SAV1 represses the development of human colorectal cancer by regulating the Akt-mTOR pathway in a YAP-dependent manner. *Cell Prolif.* **50**, e12351 (2017).
28. Brandt, Z. J., Echert, A. E., Bostrom, J. R., North, P. N. & Link, B. A. Core Hippo pathway components act as a brake on Yap and Taz in the development and maintenance of the biliary network. *Development* **147**, dev184242 (2020).
29. Lee, K. P. et al. The Hippo-Salvador pathway restrains hepatic oval cell proliferation, liver size, and liver tumorigenesis. *Proc. Natl Acad. Sci. USA* **107**, 8248–8253 (2010).
30. Farshidfar, F. et al. Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. *Cell Rep.* **18**, 2780–2794 (2017).
31. Sia, D. et al. Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. *Gastroenterology* **144**, 829–840 (2013).
32. Liu-Chittenden, Y. et al. Genetic and pharmacological disruption of the TEAD-YAP complex suppresses the oncogenic activity of YAP. *Genes Dev.* **26**, 1300–1305 (2012).
33. Zhou, S. L. et al. CXCL5 contributes to tumor metastasis and recurrence of intrahepatic cholangiocarcinoma by recruiting infiltrative intratumoral neutrophils. *Carcinogenesis* **35**, 597–605 (2014).

34. Zhou, X. D. et al. Intrahepatic cholangiocarcinoma: report of 272 patients compared with 5,829 patients with hepatocellular carcinoma. *J. Cancer Res. Clin. Oncol.* **135**, 1073–1080 (2009).
35. Wittekind, C. [Pitfalls in the classification of liver tumors]. *Pathologe* **27**, 289–293 (2006).
36. Zhou, S. L. et al. Genomic sequencing identifies WNK2 as a driver in hepatocellular carcinoma and a risk factor for early recurrence. *J. Hepatol.* **71**, 1152–1163 (2019).
37. Zhou, S. L. et al. Tumor-associated neutrophils recruit macrophages and T-regulatory cells to promote progression of hepatocellular carcinoma and resistance to Sorafenib. *Gastroenterology* **150**, 1646–1658.e17 (2016).
38. Zhou, Z. et al. Tumor-associated neutrophils and macrophages interaction contributes to intrahepatic cholangiocarcinoma progression by activating STAT3. *J. Immunother. Cancer* **9**, e001946 (2021).
39. Zhou, S. L. et al. A positive feedback loop between cancer stem-like cells and tumor-associated neutrophils controls hepatocellular carcinoma progression. *Hepatology* **70**, 1214–1230 (2019).
40. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
41. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
42. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
43. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
44. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
45. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
46. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
47. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
48. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15**, 585–598 (2014).
49. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
50. Fujimoto, A. et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500–509 (2016).
51. Roth, A. et al. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).
52. Lee, J. Y. et al. Lobular carcinomas in situ display intralésion genetic heterogeneity and clonal evolution in the progression to invasive lobular carcinoma. *Clin. Cancer Res.* **25**, 674–686 (2019).
53. Zhou, S. et al. Tacrolimus enhances the invasion potential of hepatocellular carcinoma cells and promotes lymphatic metastasis in a rat model of hepatocellular carcinoma: involvement of vascular endothelial growth factor-C. *Transpl. Proc.* **43**, 2747–2754 (2011).
54. Zhou, Z. J. et al. HNRNPAB induces epithelial-mesenchymal transition and promotes metastasis of hepatocellular carcinoma by transcriptionally activating SNAIL. *Cancer Res.* **74**, 2750–2762 (2014).

## Acknowledgements

This study was jointly supported by the National Natural Science Foundation of China (No. 82372985 to S.L.Z., No. 82373418 to Z.J.Z., No. 82273247 to C.L.S., No. 82173260 to S.L.Z., No. 82072681 to Z.J.Z., No. 82003082 to Z.Q.H.), Shanghai Technical Standard Program (21DZ2201100 to S.L.Z.) and Shanghai Medical Innovation Research Project (22Y11907300 to S.L.Z.).

## Author contributions

Z.J.Z., Y.H.Y., Z.Q.H. and S.L.Z. performed the experiments; Y.R.H., S.L.Z., and C.L.S. analyzed the data; K.X.L., R.Q.S., P.C.W., C.B.L., J.L., and J.X.Z. provided the samples; Z.J.Z., C.L.S. and S.L.Z. wrote the paper; J.Z. and J.F. commented on the study and revised the paper; C.L.S. and S.L.Z. obtained funding and designed the research.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54387-8>.

**Correspondence** and requests for materials should be addressed to Cheng-Li Song or Shao-Lai Zhou.

**Peer review information** *Nature Communications* thanks Rory Smoot, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024