EDITORIAL

A race to uncover a panoramic view of primary liver cancer

Ruidong Xue[1,2,3*], Jing Li[4,5*], Fan Bai[3], Xinwei Wang[6], Junfang Ji[7], Yinying Lu[1]

[1]Comprehensive Liver Cancer Center, Beijing 302 Hospital of PLA, Beijing 100039, China; [2]Translational Cancer Research Center, Peking University First Hospital, Beijing 100034, China; [3]Biodynamic Optical Imaging Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China; [4]Department of Gastroenterology, Peking University People's Hospital, Beijing 100044, China; [5]Clinical Center of Immune-Mediated Digestive Diseases, Peking University People's Hospital, Beijing 100044, China; [6]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA; [7]Life Sciences Institute, Zhejiang University, Hangzhou 310058, China

Introduction

Primary liver cancer, the second most common cause of cancer related death worldwide[1], presents ethnic, etiological, sex, and geographical diversity[2] (**Figure 1A**). At the histological level, liver cancer includes two major types: hepatocellular carcinoma (HCC, about 80%) and cholangiocarcinoma (CCA, about 15%). Many etiological factors contribute to HCC development, such as hepatitis B virus (HBV), hepatitis C virus (HCV), aflatoxin B_1 (AFB1), alcohol, and metabolic diseases[3]. By contrast, the major risk factors for CCA are liver flukes (*Opisthorchis viverrini* and *Clonorchis sinensis*) and primary sclerosing cholangitis[4], although HBV, HCV, and metabolic diseases may also be linked to CCA (**Figure 1A**). According to their anatomical locations, CCA can also be classified as intrahepatic (ICC), perihilar (PCC), and extrahepatic (ECC) tumors (**Figure 1B**).

Currently, HCC and CCA are regarded as distinct entities that differ in their clinical features, guidelines for diagnosis and treatment, as well as their cellular phenotypes and molecular profiles. Marked inter-tumor heterogeneity has been described in patients with liver cancer[2,3] (**Figure 1C**). Recent genomic studies also identified intra-tumor heterogeneity in both HCC and CCA[4,5]. All these speak for the lack of targeted therapy in liver cancer, underlining its genetic complexity and requiring international efforts to portray a more comprehensive genetic landscape of liver cancer.

Improvements in next-generation sequencing technology in recent years have allowed researchers to study the liver cancer genome more effectively and rapidly[2,3]. Dozens of studies have investigated the genome of liver cancer, identifying numerous somatic driver events during hepatocarcinogenesis, including point mutations, small insertions and deletions, copy number aberrations (CNAs), virus integrations, and gene fusions. Integrated analysis of these driver events reveals that telomere maintenance, the p53-cell cycle, WNT/β-catenin, kinase-RAS, and epigenetic regulation are the major signaling pathways that are disrupted in liver cancer[2]. More recently, several international studies spearheaded the analysis by integrating multiple "omics" data, including genome, epigenome, transcriptome and metabolome, to provide a better molecular classification and key genetic features of liver cancer. These efforts have been led by The Cancer Genome Atlas (TCGA) network, the International Cancer Genome Consortium (ICGC), and Thailand Initiative in Genomics and Expression Research for Liver Cancer (TIGER-LC) Consortium to better define liver cancer heterogeneity. The profound implications of these international efforts are clear that there is an urgent need to develop a well-annotated tumor biobank to control the clinical, biological, etiological, and genomic heterogeneity of liver cancer, and to study the landscape at each omic-level, and integrate the multiple-omics data to uncover the true genetic panorama of liver cancer.

TCGA HCC multi-platform analysis

The TCGA HCC team[6] sequenced a full set of 363 HCC cases at the genomic level and selected a core set of 196 cases to profile their methylomes, transcriptomes, and proteomes. Integrating these data from distinct platforms via

---

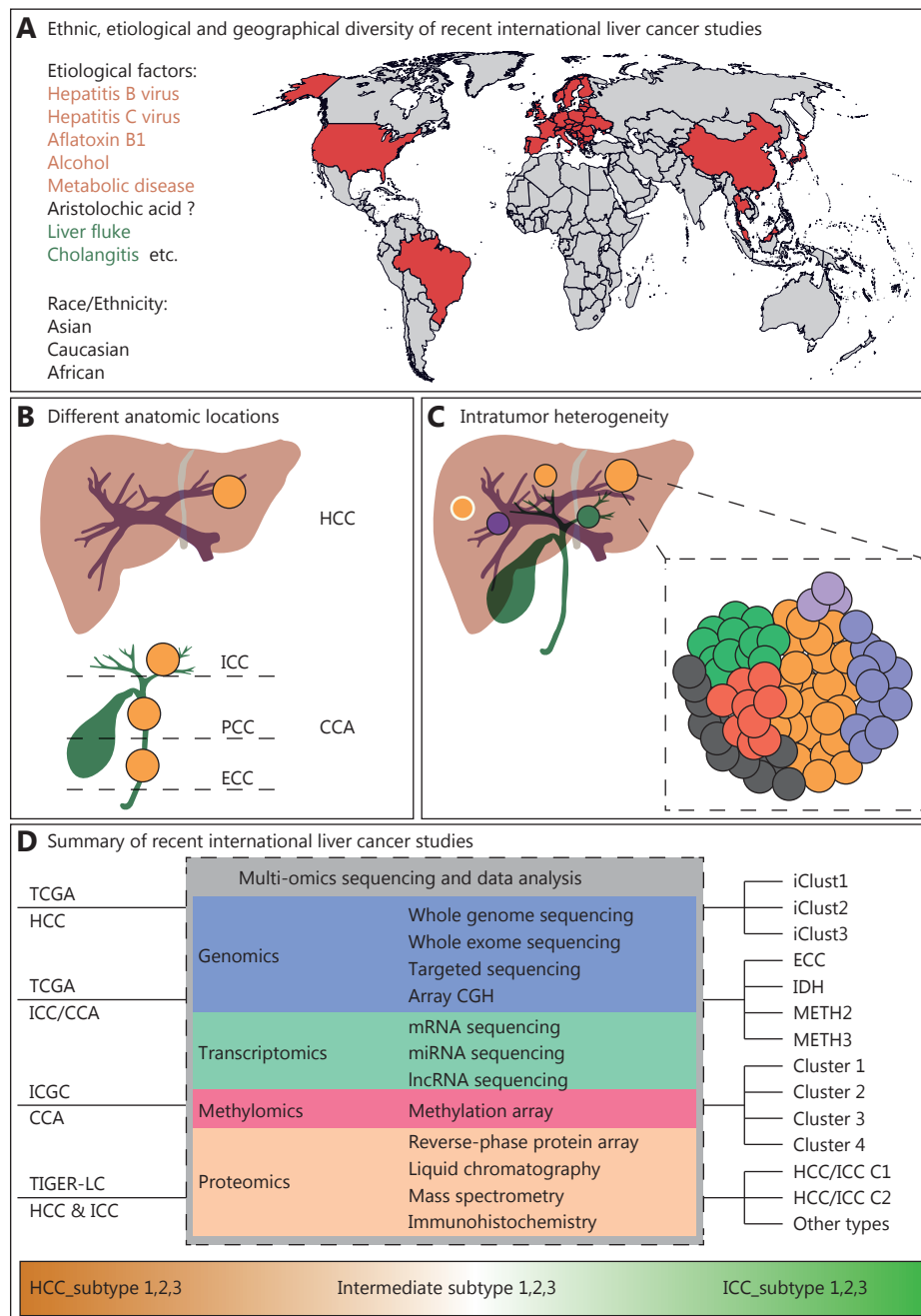*These authors have contributed equally to this work
Correspondence to: Junfang Ji and Yinying Lu
E-mail: junfangji@zju.edu.cn and luyinying1973@163.com

Figure 1 Summary of recent international multi-omics liver cancer studies. (A) Ethnic, etiological and geographical diversity of recent international liver cancer studies. (B) The anatomic locations of primary liver cancer. The portal vein is in purple and the bile duct is in green. Orange circles denote tumor. HCC, hepatocellular carcinoma. ICC, PCC and ECC stand for intrahepatic, perihilar, and extrahepatic cholangiocarcinoma (CCA), respectively. (C) Intra-tumor heterogeneity in liver cancer. The big orange circle denotes the primary tumor, small orange circle denotes intrahepatic metastasis, the small orange circle with white outline denotes a multiple occurrence, the purple and green circles stand for portal vein and bile duct tumor thrombosis. The enlarged box shows the complex subclonal structure of the primary tumor. Different circles mean different subclones. (D) Summary of the molecular subtypes identified in recent international multi-omics studies. The grey box lists the multi-omics sequencing methods involved in these studies. The lower panel illustrates the existence of intermediate molecular subtypes between HCC subtypes and ICC/CCA subtypes.

unsupervised clustering allowed them to identify three different molecular subtypes of HCC, named iClust1, iClust2, and iClust3. These molecular subtypes were each characterized by different clinical and genetic variables. For instance, iCust1 was associated with younger age, Asian ethnicity, female gender, normal body weight, higher tumor grade, macrovascular invasion, lower frequency of *CDKN2A* silencing, *CTNNB1* mutation, and *TERT* promoter mutation. They also built a prognosis prediction model that could be successfully validated in three independent cohorts. Moreover, a combined analysis of significantly mutated genes (*ALB* and *APOB*) and hypermethylated genes (*CPS1*) unveiled the potential driver role of metabolic reprogramming in the progression of hepatocytes to malignancy. By analyzing the expression of p53-targeted genes, they identified a p53 signature and revealed that certain HCC cases without TP53 mutations had inactive p53 signaling. This p53 signature would be clinically useful to identify patients who have an inactivated p53 status, rather than just a p53 mutation. Then the team also characterized the immune phenotypes of HCC via histopathology and gene expression analyses of immune cell markers. The results show that certain HCC cases have high levels of immune infiltration and the shift of the immune microenvironment from activating effector cells to resting suppressive immune cells implies the potential clinical benefit of immunotherapy. This study is the first comprehensive multi-omics analysis of HCC and reveals its different molecular subtypes. Potential molecular targets identified in each subtype await for further validation and hopefully lead to improvements in targeted therapy and patient prognosis.

TCGA CCA multi-platform analysis

The TCGA CCA team[7] analyzed a smaller set of 53 CCA cases, including 42 ICC, using a similar strategy to that described in the HCC study. It should be noted that all the CCA cases analyzed in the study were not associated with liver fluke and hepatitis infection, the major etiological factors for this tumor type. The TCGA team again integrated various -omics data from a core set of 38 cases and found that they contained four subgroups (termed COCA1–4). COCA1 was named as an ECC subtype because it contained three of four extrahepatic CCA. COCA2 was named as IDH because it was enriched for all the *IDH* hotspot mutations. COCA3 and COCA4 were named as METH2 and METH3, respectively, because of their enrichment for specific DNA methylation patterns. Possibly due to the small sample size, patients from these clusters did not show significant

difference in survival. In addition, the *IDH* mutant-enriched subtype could also be identified by mRNA expression clustering. Further integrated -omics analysis of the IDH subtype confirmed its distinct molecular features, including low expression of chromatin modifiers, high expression of mitochondrial genes, and amplifications in the mitochondrial DNA.

Furthermore, the TCGA team integrated CCA data with HCC and pancreatic ductal adenocarcinomas from TCGA and performed clustering analysis. Interestingly, a small percentage (4%, 7/179) of HCC were clustered into the CCA group, and four out of these seven cases harbored *IDH* hotspot mutations. Meanwhile, 100% (4/4) of *IDH* mutated HCC cases were clustered into the CCA group. Both histological examination and mRNA expression analysis also showed that these *IDH*-mutant HCC cases have certain similarities to CCA. The existence of these intermediate cases supports the view that liver tumors comprise a continuous spectrum.

ICGC CCA multi-platform analysis

The ICGC CCA team collected a large cohort of 489 CCA cases from multiple countries and regions, among which 133 were fluke-associated[8]. Data from single genomic platforms revealed new driver alterations in CCA, comprising coding mutations in *RASA1*, structural variations (such as an *FGFR* 3' untranslated region deletion), and noncoding mutations in H3K27me3 promoters. Integrative genomic, epigenomic, and transcriptomic analysis clustered these cases into four subtypes, termed Cluster 1–4, which exhibited distinct clinical and genetic features. Etiologically, Cluster 1 comprised mainly fluke-positive tumors, Cluster 3 and Cluster 4 contained mainly fluke-negative tumors, and Cluster 2 had both types. *TP53* mutations, *ERBB2* amplifications, and elevated *ERBB2* expression were significantly enriched in Clusters 1 and 2. In addition, Cluster 1 showed a higher mutation burden, enrichment of *ARID1A* and *BRCA1/2* mutations, and higher mutation levels in histone-3-lysine-27-trimethylated gene promoters. Cluster 3 showed a higher level of CNAs and exclusive upregulation of immune checkpoint genes, which indicated potential cluster-specific immunotherapies. In Cluster 4, BAP1 and IDH1/2 mutations, FGFR alterations, and upregulated *FGFR* family and *PI3K* pathway signatures were enriched. Although Clusters 1 and 4 both showed DNA hypermethylation, Cluster 1 contained hypermethylation in CpG islands, while Cluster 4 had hypermethylation in CpG promoter shores. Clinically, Cluster 4 showed better prognosis compared with

338

Xue et al. Molecular classification of liver cancer

the other clusters, which was validated in an expanded cohort and in data from other published studies. Notably, patients with CCA that were classified by anatomical sites did not present a difference in terms of their survival. Collectively, this study showed that molecular subtypes based on multi-omics data offer a new patient stratification criterion with a better understanding of CCA, a prognostic prediction power, and a potential value in guiding patients for targeted therapy.

TIGER-LC HCC and ICC multi-platform analysis

As an Asian country, Thailand has a high incidence of liver cancer, with a higher proportion of CCA. The TIGER-LC Consortium has dedicated their efforts to delineating liver cancer at multiple omic levels. The team analyzed the first sequential 199 Thai cases, which included 69 HCC and 130 ICC cases[9]. Unsupervised hierarchical clustering and principal-component analyses with transcriptome data of all cases revealed that some of the HCC and ICC cases shared similar molecular features, which was consistent with the two most recent reports[10,11]. Consensus clustering further classified HCC and ICC into three and four major subtypes, respectively. Comparison among these subtypes showed that there were similar features between HCC-C1 and ICC-C1, as well as between HCC-C2 and ICC-C2. HCC/ICC-C1 subtypes were characterized by high chromosomal instability, while HCC/ICC-C2 subtypes were linked to inflammatory responses. Absence of significant correlations with cirrhosis, etiology, TNM staging, and diagnostic factors suggested that these molecular subtypes were unique. Significant differences in prognosis between them indicated that this classification had clinical relevance. Such common subtypes were further investigated in a larger cohort, including 582 Asian and 265 Caucasian patients.

In the in-depth analysis, the team found that the HCC/ICC-C2 subtype had a better prognosis in Asian than in Caucasian patients, while the HCC/ICC-C1 subtype showed the reverse results. Notably, the HCC/ICC-C2 subtypes were not found in the Caucasian cohorts, possibly because of the relative small sample size. Consistent with the TCGA study, it seemed that common molecular subtypes are also race/ethnicity-dependent. Moreover, combining CNA and expression data, they revealed that *PLK1/ECT2* signaling was enriched in the HCC/ICC-C1 subtype. Immunohistochemical analysis further validated these results and demonstrated that *PLK1/ECT2* could be used as both classification and prognostic markers. Collectively, this study showed the existence of an HCC/ICC intermediate subgroup, indicating

the importance of molecular subtyping and paving the way to establish new classification criteria with the hope of developing novel targeted therapy across HCC and ICC.

Intermediate subtypes of HCC and CCA

The findings of the TIGER-LC and TCGA studies reveal that certain HCC cases are molecularly and histologically similar to CCA[7,9]. These studies both shed light on the existence of the intermediate subtype of HCC/CCA. Further efforts are needed to dissect the intermediate subtype of HCC/CCA (such as the C1 and C2 groups from TIGER-LC) and to explore its clinical relevance, e.g., to test *PLK1/ECT2* as classification and prognostic markers for intermediate HCC/CCA subgroups in multiple independent cohorts.

It should be noted that in the TCGA study, both HCC and CCA contained unique *IDH*-mutant-like subtypes. Interestingly, *IDH* mutations were enriched in non-Asian HCC cases. Considering the proposed function of *IDH* mutations in shifting liver progenitor cells from a hepatocellular towards a cholangiocellular fate[12], *IDH* mutations might play critical roles in the CCA-like-phenotype presentation of these HCC cases. More interestingly, *IDH*-mutations were enriched in an unmatched group rather than in the C1/C2 subtype of TIGER-LC[9]. In this vein, it is likely that the *IDH*-mutant cases may also represent a group of intermediate liver cancer, especially in Caucasian populations. Further validation is needed in a broader patient cohort with both HCC and ICC across different races and ethnicities.

Collectively, the above large-scale multi-omics studies suggest a possible hepatocarcinogenesis model in which liver cancer might comprise a continuous molecular/biological spectrum from distinct subtypes of HCC and CCA with hepatocellular and cholangiocellular features, respectively, to intermediate subtypes that exhibit overlapping phenotypes of HCC and CCA (**Figure 1D**).

Aristolochic acid and liver cancer

It is well known that multiple etiological factors are involved in hepatocarcinogenesis (**Figure 1A**). These factors have distinct carcinogenic mechanisms and can generate a unique molecular stamp in the tumor genome, which is termed the mutational signature[13]. One of the best examples is a unique p53-Arg249Ser mutation in HCC induced by AFB1[14]. A recent study indicates that aristolochic acid and their derivatives, collectively called AA, are strong mutagens that can bind to DNA and cause T:A > A:T transversions[15,16]. AA

has long been known to cause kidney failure and may lead to bladder and upper tract urothelial carcinomas. A recent study analyzing HCC genome data from multiple countries and regions revealed the prevalence of AA-associated mutational signature in HCC throughout Asia, especially in Taiwan[17], China. Data from their previous study[16] and other studies[4,18] showed that AA signatures can be found in Chinese patients with liver cancer. To comprehensively explore the potential influence of AA on liver cancer, they first collected 98 HCC cases from Taiwan that had a wide exposure to AA. Analyzed by their newly designed algorithm mSigAct, exome sequencing data showed that AA signature can be found in 78% (76/98) of HCC. They further examined this signature in sequencing data from 1,400 HCC cases from diverse geographical regions. AA signatures were found in 47% (42/89) of HCC from mainland China, in 29% (10/35) from Southeast Asian HCC, in 13% (29/231) from South Korean HCC, in 2.7% (13/477) from Japanese HCC, in 4.8% (10/209) from North American HCC, and in 1.7% (4/230) from European HCC. Collectively, these data showed that the AA signature is closely associated with HCC worldwide.

The high prevalence of AA signature in HCC from East and Southeast Asia raises the existing carcinogenic alert concerning AA-containing herbal medicines, which may be still consumed by local residents. Acknowledging the carcinogenic effect of AA was already validated long before, these results reminded us that local governments and research institutes should carefully re-evaluate the usage and circulation of AA-containing herbs and then establish up-to-date regulations and solutions accordingly. Nevertheless, there are still several points that are not clear concerning the causal role of AA in hepatocarcinogenesis. First, the analyses were merely based on the detection of T:A > A:T. It is possible that certain, as yet unidentified, chemicals may have similar effects. To rule out this possibility, independent analysis based on other techniques should be used; for example, detecting the AA-adducts in the tumor or circulating DNA. Second, the new detection algorithm may overestimate the percentage of AA signatures. In some AA-positive calling, the AA signature may only be one of the signatures in the tumor, suggesting that AA may act as a passenger, rather than a driver, of tumorigenesis. The AA percentage in the Chinese cohort varies dramatically from 11/93 in the previous study[16] to 42/89 in the current study[17]. This suggests that the percentage of tumors that are driven by AA may be overestimated. To perform mutational signature analysis, the immediate flanking 5' and 3' bases of the mutated base are incorporated, as a result, there are 96 substitutions in total[13]. Considering the, on average,

hundreds of mutations detected in the exome data, only tens of mutations segregated into the 96 bins, which further weakens the power of this detection. Third, the published data for the 1,400 HCC cases used in this study are not ideal. Interestingly, the data with a high percentage of AA signature are all in small cohorts (~100 cases), while the data with a low percentage are all large cohorts (200–500 cases). The percentage of AA signature in Chinese HCC is questionable given the enormous population and vast regional diversity in China because it was derived from a relatively small cohort from Hong Kong, China (88 cases).

To determine the relationship between AA and Asian-HCC, well-designed cell line and animal model studies, together with well-annotated large biobanks, are required in the future. Regarding patients with liver cancer, a thorough review of their medical history focusing on AA and related guidelines is needed. Acknowledging the existing complexities and difficulties, epidemiological investigations for AA are required. There is also a need for more studies on the AA-containing herbs, and more broadly, Chinese medicine.

Chinese medicines frequently comprise combinations of many ingredients, which are specific parts of herbs and animals. Not all parts of AA-containing herbs have AAs. The "cooking" process follows documented protocols, during which these ingredients are added at different times, followed by specific cooking steps. Dozens of steps are involved in their production. Thus, how much AA is left in the final form remains unknown when it comes to certain drugs, which requires standard composition and toxicological analysis in the future. Enormous efforts are urgently needed to circumvent the potential toxicity of certain Chinese medicines.

Perspective

Taking the TCGA, ICGC, and TIGER-LC studies together, these pioneering integrated analyses of multi-omics data from two different histological cancer types reveal novel molecular taxonomy of liver cancer. Among these, each subtype can be characterized by different genomic, epigenomic, RNA levels, and metabolomic alterations, which may serve as biomarkers and potential druggable targets. Further efforts are required to validate the molecular subgroups, to develop companion diagnostic tests for precision therapy, and to establish clinical trials based on the identified mutations, signaling pathways, and clinically available inhibitors or compounds.

The fact that certain subtypes are associated with specific

underlining etiological factors indicates that distinct liver cancer subtypes are the result of different extrinsic and intrinsic carcinogenic processes. Beyond the subtypes of HCC and CCA, one very intriguing possibility is that, if put together, the HCC and CCA subtypes may not be mutually exclusive. As a result, liver cancer might comprise a spectrum of HCC-specific subtypes, HCC/CCA intermediate subtypes, and CCA specific subtypes (**Figure 1D**). In this scenario, global efforts with more large-scale, multi-omics studies across different etiologies and geographies are needed in the future to explore all the existing subtypes and to finally complete the painting of the genetic panorama of liver cancer.

Acknowledgements

Conflict of interest statement

No potential conflicts of interest are disclosed.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015; 136: E359-86.

2. Marquardt JU, Andersen JB, Thorgeirsson SS. Functional and genetic deconstruction of the cellular origin in liver cancer. Nat Rev Cancer. 2015; 15: 653-67.

3. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. Genetic landscape and biomarkers of hepatocellular carcinoma. Gastroenterology. 2015; 149: 1226-39.

4. Xue RD, Li RY, Guo H, Guo L, Su Z, Ni XH, et al. Variable intra-tumor genomic heterogeneity of multiple lesions in patients with hepatocellular carcinoma. Gastroenterology. 2016; 150: 998-1008.

5. Walter D, Doring C, Feldhahn M, Battke F, Hartmann S, Winkelmann R, et al. Intratumoral heterogeneity of intrahepatic cholangiocarcinoma. Oncotarget. 2017; 8: 14957-68.

6. The Cancer Genome Atlas Research Network. Cancer Genome Atlas Research N. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell. 2017; 169: 1327-41.e23

7. Farshidfar F, Zheng SY, Gingras M-C, Newton Y, Shih J, Robertson AG, et al. Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. Cell Rep. 2017; 18: 2780-94.

8. Jusakul A, Cutcutache I, Yong CH, Lim JQ, Huang MN, Padmanabhan N, et al. Whole-genome and epigenomic landscapes of etiologically distinct subtypes of cholangiocarcinoma. Cancer Discov. 2017; 7: 1116-35.

9. Chaisaingmongkol J, Budhu A, Dang H, Rabibhadana S, Pupacdi B, Kwon SM, et al. Common molecular subtypes among asian hepatocellular carcinoma and cholangiocarcinoma. Cancer Cell. 2017; 32: 57-70.

10. Oishi N, Kumar MR, Roessler S, Ji JF, Forgues M, Budhu A, et al. Transcriptomic profiling reveals hepatic stem-like gene signatures and interplay of miR-200c and epithelial-mesenchymal transition in intrahepatic cholangiocarcinoma. Hepatology. 2012; 56: 1792-803.

11. Woo HG, Lee JH, Yoon JH, Kim CY, Lee HS, Jang JJ, et al. Identification of a cholangiocarcinoma-like gene expression trait in hepatocellular carcinoma. Cancer Res. 2010; 70: 3034-41.

12. Saha SK, Parachoniak CA, Ghanta KS, Fitamant J, Ross KN, Najem MS, et al. Mutant IDH inhibits HNF-4α to block hepatocyte differentiation and promote biliary cancer. Nature. 2014; 513: 110-4.

13. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500: 415-21.

14. Hsu IC, Metcalf RA, Sun T, Welsh JA, Wang NJ, Harris CC. Mutational hotspot in the p53 gene in human hepatocellular carcinomas. Nature. 1991; 350: 427-8.

15. Hoang ML, Chen CH, Sidorenko VS, He J, Dickman KG, Yun BH, et al. Mutational signature of aristolochic acid exposure as revealed by whole-exome sequencing. Sci Transl Med. 2013; 5: 197ra102

16. Poon SL, Pang ST, McPherson JR, Yu W, Huang KK, Guan P, et al. Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. Sci Transl Med. 2013; 5: 197ra101

17. Ng AWT, Poon SL, Huang MN, Lim JQ, Boot A, Yu W, et al. Aristolochic acids and their derivatives are widely implicated in liver cancers in Taiwan and throughout Asia. Sci Transl Med. 2017; 9. pii: eaan6446.

18. Zou S, Li J, Zhou H, Frech C, Jiang X, Chu JS, et al. Mutational landscape of intrahepatic cholangiocarcinoma. Nat Commun. 2014; 5: 5696.