



OPEN

Tumor metabolism and associated serum metabolites define prognostic subtypes of Asian hepatocellular carcinoma

Yotsawat Pomyen^{1,2,14}, Anuradha Budhu^{1,3,14}, Jittiporn Chaisaingmongkol^{1,4,5}, Marshonna Forgues¹, Hien Dang^{1,6}, Mathuros Ruchirawat^{4,5}, Chulabhorn Mahidol⁴, Xin Wei Wang^{1,3}✉ & the TIGER-LC Consortium*

Treatment effectiveness in hepatocellular carcinoma (HCC) depends on early detection and precision-medicine-based patient stratification for targeted therapies. However, the lack of robust biomarkers, particularly a non-invasive diagnostic tool, precludes significant improvement of clinical outcomes for HCC patients. Serum metabolites are one of the best non-invasive means for determining patient prognosis, as they are stable end-products of biochemical processes in human body. In this study, we aimed to identify prognostic serum metabolites in HCC. To determine serum metabolites that were relevant and representative of the tissue status, we performed a two-step correlation analysis to first determine associations between metabolic genes and tissue metabolites, and second, between tissue metabolites and serum metabolites among 49 HCC patients, which were then validated in 408 additional Asian HCC patients with mixed etiologies. We found that certain metabolic genes, tissue metabolites and serum metabolites can independently stratify HCC patients into prognostic subgroups, which are consistent across these different data types and our previous findings. The metabolic subtypes are associated with β -oxidation process in fatty acid metabolism, where patients with worse survival outcome have dysregulated fatty acid metabolism. These serum metabolites may be used as non-invasive biomarkers to define prognostic tumor molecular subtypes for HCC.

Hepatocellular carcinoma (HCC) is one of the leading cancers worldwide in both incidence and mortality. According to the Global Cancer Observatory, the number of new liver cancer cases worldwide is 4.7% of all cancer incidence but accounted for 8.2% of all cancer deaths in 2018¹. It is estimated that there will be a 64% increase in incidence and a 69% increase in mortality of liver cancer by 2040 based on 2018 statistics². Most of these incidence and deaths will occur in East Asia and Southeast Asia, including Thailand.

In the past, Thailand had one of the highest incidence of HCC due to the high prevalence of hepatitis B virus (HBV) infection³ and food contamination by aflatoxin B₁ (AFB₁)^{4,5}. However, lifestyle changes and economic improvement over several decades have shifted the cause of HCC from HBV infection and food contamination to new causes, such as Hepatitis C virus (HCV) infection⁶ and non-alcoholic fatty liver disease (NAFLD)^{7,8}. The shift in the landscape of hepatocarcinogenesis is now a challenge for healthcare professionals. Currently, α -fetoprotein (AFP) is the most convenient and non-invasive serum biomarker for detecting HCC⁹. However, the reliability of AFP compared to ultrasound is still an ongoing debate for early-stage detection of the HCC^{10,11}. Therefore, there is an urgent need for better serum biomarkers to achieve early detection and better patient stratification for precision oncology to improve patient prognosis^{9,10}.

¹Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA. ²Translational Research Unit, Chulabhorn Research Institute, Bangkok 10210, Thailand. ³Liver Cancer Program, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA. ⁴Laboratory of Chemical Carcinogenesis, Chulabhorn Research Institute, Bangkok 10210, Thailand. ⁵Center of Excellence on Environmental Health and Toxicology, Office of Higher Education Commission, Ministry of Higher Education, Science, Research and Innovation, Bangkok 10400, Thailand. ⁶Present address: Division of Surgery, Thomas Jefferson University, Philadelphia, PA 19107, USA. ¹⁴These authors contributed equally: Yotsawat Pomyen and Anuradha Budhu. *A list of authors and their affiliations appears at the end of the paper. ✉email: xw3u@nih.gov

Previously, our group has shown that serum AFP alone is not sufficient as a biomarker for early detection or patient stratification of aggressive HCC subtypes in Asian patient cohorts^{12,13}. With the shortcomings of AFP in Asian populations, we have established a gene-signature panel of molecular subtypes that can consistently stratify aggressive HCC and intrahepatic cholangiocarcinoma (iCCA) subtypes among several Asian populations¹³. However, this and previous signatures, are mainly based on gene expression and miRNA expression signatures derived from bulk tumor tissues, which are difficult to obtain without surgical resection of the tumors¹⁴. In this study, we reasoned that the molecular subtype-related tumor biology may have a systemic impact and, therefore, be reflected through the systemic serum metabolite profiles of the patients. To identify serum metabolites that are surrogates of gene-based molecular subtypes, we employed metabolite-focused, two-step correlation analysis, first with correlation analysis between tissue gene expression of metabolic genes and tissue metabolite profiles, and second, between tissue metabolite profiles and serum metabolite profiles. This strategy allowed us to identify serum metabolite panels that can stratify HCC patients according to their tumor-related prognostic subtypes.

Results

Tissue metabolites can distinguish between tumor and non-tumor tissue. The overall flowchart of the analysis strategy is illustrated in Fig. 1A, with a detailed analysis flowchart shown in Supplementary Fig. S1. To test whether tissue metabolites can distinguish tumor and non-tumor tissues of 56 HCC patients, Principal Component Analysis (PCA) was used to classify tissue samples based on metabolites measured by an untargeted metabolomics approach (Fig. 1B). Based on the 500 metabolites with highest variability among samples, tumor and non-tumor tissues can be separated clearly. To visualize the extent of tissue metabolites and pathways that are dysregulated, differentially altered metabolites between tumor and non-tumor tissues were overlaid onto overall human metabolite map from KEGG¹⁵. The 562 known detected tissue metabolites can be converted to 352 KEGG compound IDs from the total of 1271 metabolites present on the map (accession number hsa01100). The affected metabolites in HCC fall into four main metabolic categories, i.e., nucleotide metabolism, amino acid metabolism, energy metabolism, and lipid metabolism (Fig. 1C).

Metabolic genes, tissue and serum metabolites can identify patients with worse survival outcome. We used metabolism-specific genes to further determine if they can classify HCC into prognostic subtypes. To ensure that the metabolic genes and tissue metabolites are tumor-specific, we incorporated the calculation of a paired t-test between tumor and adjacent non-tumor tissue gene expression. For the first step of the correlation analysis, the numbers of metabolic genes and metabolites that passed correlation thresholds from a permutation test and a paired t-test with an FDR-adjusted p -value of < 0.05 were 491 genes and 78 metabolites, respectively (Supplementary Table S1). Among 78 metabolites that passed the correlation coefficient thresholds (Supplementary Table S2), 40 metabolites passed an additional 1.5 fold-change cutoff between tumor and adjacent non-tumor tissues (Fig. 2A, Supplementary Fig. S2A, and Supplementary Table S3). We performed hierarchical clustering analysis on Thai HCC patient samples by using consensus clustering algorithm¹⁶ based on these 491 metabolic genes (MetGene) or 40 tumor-specific metabolites (TissueMet) which revealed three stable clusters based on metabolic genes or tumor-specific tissue metabolites, named MetGene-C1, MetGene-C2, and MetGene-C3, or TissueMet-C1, TissueMet-C2, and TissueMet-C3, respectively (Fig. 2B—upper panel). The metabolic gene expression based on 491 genes shows distinct expression patterns between the three clusters, which are similar to previously identified common Asian subtypes (Fig. 2B—middle panel)¹³. Similar to global transcriptome profiling analysis, there was no discernible pattern emerging from standard clinical and etiological parameters of HCC (Fig. 2B—lower panel). These results suggest that metabolic genes can capture major molecular features of tumor samples. We also explored whether these subclasses were consistent with prognostic signatures published among HCC cohorts. To do so, we assessed S1-S3 signature¹⁷, stem cell signature¹⁸, EpCAM signature¹⁹, Metastatic signature²⁰, CCA-like signature²¹, Cluster 2 signature²², proliferation signature²³ and found a concordance between the good and poor survival subclasses defined by the metabolic genes and those defined by published signatures. We also performed consensus clustering using the same set of metabolic genes in two other cohorts: Liver Cancer Institute (LCI)²⁰ cohort and The Cancer Genome Atlas (TCGA)²⁴ cohort on Asian patients only. Similar gene expression patterns were observed in both cohorts (Supplementary Fig. S3A, B). The resulting MetGene clusters from LCI and TCGA cohorts were then subjected to Subclass Mapping analysis to identify clusters that matched the original consensus clusters from the TIGER-LC cohort. MetGene-C1 and MetGene-C2 clusters matched the original clusters from TIGER-LC cohort (Supplementary Fig. S3C, D). To identify serum metabolites that are directly associated with tumor tissue metabolites, we performed a second step of correlation analysis between the 40 tumor-specific tissue metabolites and global serum metabolites, which resulted in a set of 75 correlated serum metabolites (Fig. 3A and Supplementary Table S4). Finally, to ascertain that the identified serum metabolites can stratify patients into similar subgroups as those defined by metabolic genes and tumor tissue metabolites, we performed consensus clustering analysis based on the 75 serum metabolites, which similarly yielded three stable clusters of the samples named SerumMet-C1, SerumMet-C2, and SerumMet-C3. Similar to our assessment of metabolic genes, we found a concordance between good and poor survival subclasses defined by the 75 serum metabolites and those defined by published signatures.

We also investigated the mutation status of two genes that often mutated in HCC, namely TP53 and CTNNB1. These two genes are not usually mutated together in the same patient. We found that TP53 mutation are predominantly in MetGene-C1, TissueMet-C1, and SerumMet-C1 clusters, whereas CTNNB1 mutation are predominantly found in MetGene-C3, TissueMet-C3, and SerumMet-C3 clusters (Fig. 2B—lower panel).

To determine whether the resulting patient subgroups are associated with patient outcome, survival analysis was performed. Survival analysis based on consensus clusters revealed that the C1 cluster from metabolic

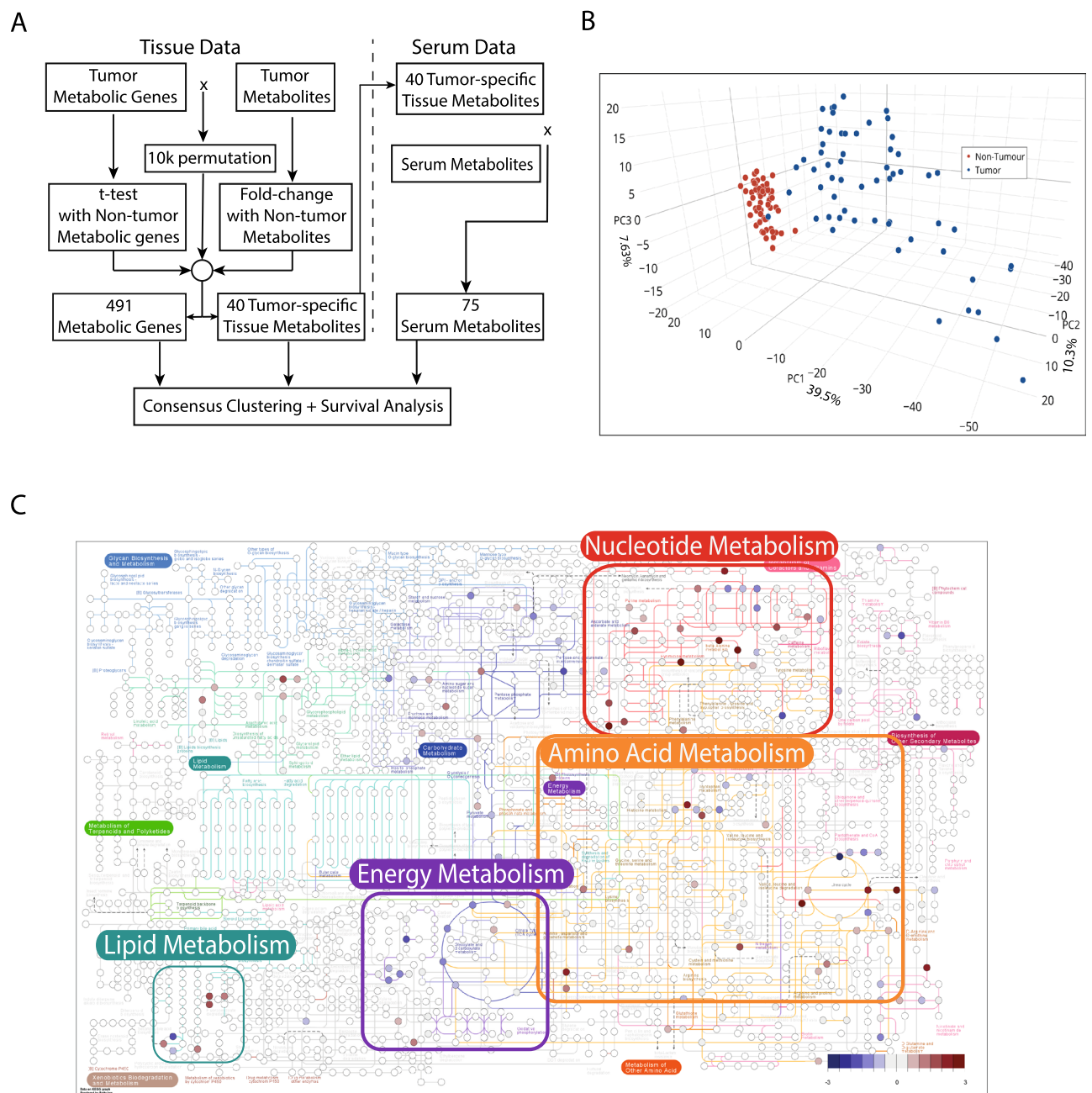
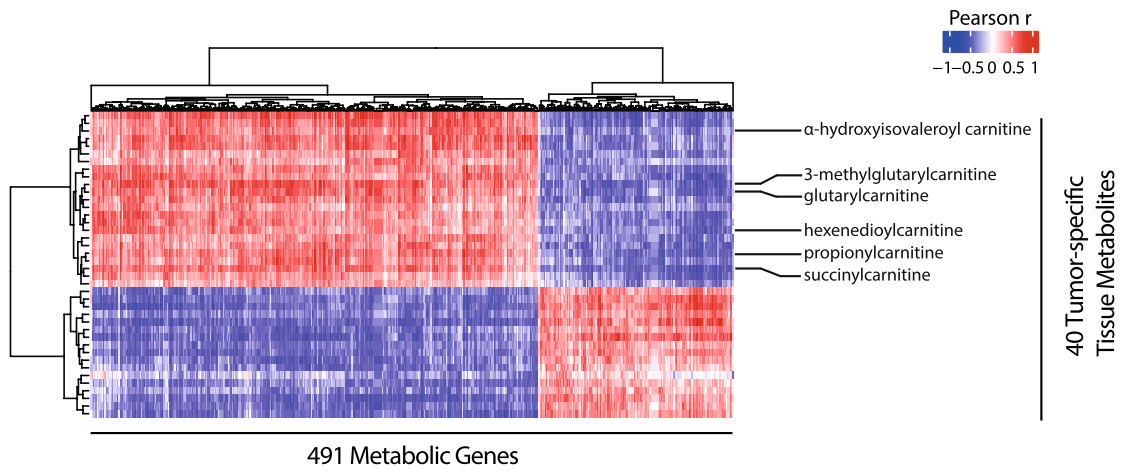


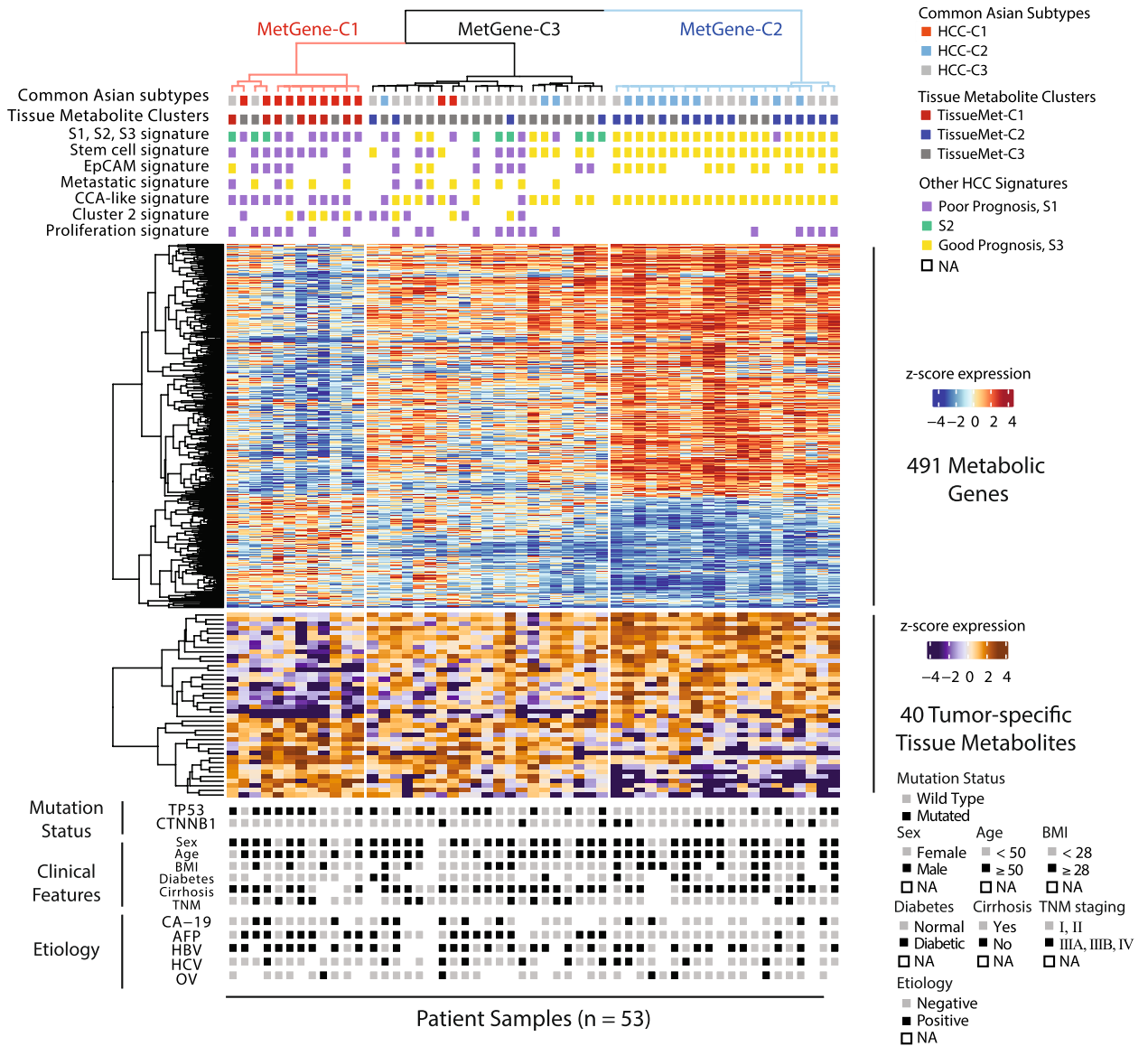
Figure 1. Overall study design, tissue stratification by metabolite abundance and their related pathways. **(A)** Overall flowchart of the analysis. **(B)** 3D PCA plot of tumor and adjacent non-tumor tissue metabolites from Thai HCC patients based on expression of 500 most variable metabolites. The tumor and non-tumor tissues can be clearly separated by tissue metabolites alone. **(C)** KEGG Overall Human Pathways (map accession number hsa01100¹⁵) overlaid with metabolite ratios between tumor and adjacent non-tumor tissues from HCC patients. Each node represents a human metabolite in a metabolic pathway. The colors represent log₂ fold-change of the tissue metabolites. Blue represents increased mean metabolite abundance from adjacent non-tumor tissues, and red represents higher mean metabolite abundance from tumor tissues. White nodes represent metabolites with no data available. Abbreviations: PCA—Principal Component Analysis; HCC—Hepatocellular Carcinoma.

genes (MetGene-C1), tissue metabolites (TissueMet-C1), and serum metabolites (SerumMet-C1) have similarly worse survival outcome compared to other clusters (Fig. 3B–D, respectively). Hazard ratios [with 95% confidence interval (CI)] between the C1 cluster and other clusters defined by metabolic genes, tumor-specific tissue metabolites, and serum metabolites are 5.8 [2.2–15], 3.3 [1–10], and 4.4 [1.6–12] respectively. The same trends were also observed in the LCI and TCGA-Asian cohorts, however the C3 cluster was not present in either cohort (Supplementary Fig. S3E, F). The C2 and C3 clusters from both metabolic genes and metabolites of the TIGER-LC cohort have similar survival trajectories (MetGene-C2 and MetGene-C3 in Fig. 3B, and TissueMet-C2 and TissueMet-C3 in Fig. 3C), which should be expected since these metabolic genes and metabolites are correlated. The trajectory of the C3 cluster in serum metabolites, however, did not resemble those of metabolic

A



B



◀ **Figure 2.** Correlation analysis of tumor metabolic genes and tissue metabolites. (A) Heatmap of correlation coefficients between 491 metabolic genes and 40 tumor-specific tissue metabolites that pass the thresholds of 0.165 and -0.165 defined in by permutation test. (B) Heatmap of gene expression from 491 metabolic genes in the tumor tissues from HCC patients from TIGER-LC cohort. The top panel of the heatmap shows clusters of patients based on three classifications of patients: (1) the dendrogram is based on consensus clustering from 491 metabolic genes identified through permutation test (MetGene clusters); (2) common Asian subtypes identified previously by our research group¹³; and (3) Tissue Metabolite (TissueMet) clusters, which are based on consensus clustering from 40 tumor-specific tissue metabolites. The middle panel shows two heatmaps based on z-score expression of 491 metabolic genes and 40 tumor-specific tissue metabolites. The bottom panel of the heatmap shows mutation status of two genes (TP53 and CTNNB1), clinical features, and etiologies of each tissue sample. Abbreviations: MetGene-C—Metabolic gene-Cluster; TissueMet-C—Tissue metabolite-Cluster; HCC—Hepatocellular Carcinoma; BMI—Body Mass Index; TNM—TNM Classification of Malignant Tumors; CA-19—Cancer Antigen 19-9; AFP—Alpha Fetoprotein; HBV—Hepatitis B virus; HCV—Hepatitis C virus; OV—*Opisthorchis viverrini*. Each colored box above heatmap represents one sample. The color representing “Good” prognosis and S3 signature is yellow, while “Poor” prognosis and S1 signature is represented by purple. S2 signature is represented by green box.

genes or tissue metabolites clusters (Fig. 3D). Traditional prognostic biomarker for HCC, i.e., AFP > 300 ng/ml, cannot separate patients into two subgroups with statistical significance in this cohort (Supplementary Fig. S4). Hazard Ratio [95%CI] for AFP as prognostic biomarker in this cohort is 2 [0.72–5.5]. Interestingly, the samples that were classified as SerumMet-C1 using serum metabolites were predominantly C1 samples identified by metabolic genes, tissue metabolites, and common Asian subtypes in our previous study¹³, and the samples from the C1 cluster classified by tissue metabolites are consistent with the C1 samples from serum metabolites (Fig. 4A—upper panel). Similar to the consensus clusters identified using metabolic genes, there is no discernible pattern associated with any cluster defined by the serum metabolites and standard HCC etiology parameters (Fig. 4A—lower panel). Overall, serum metabolites were representative of metabolic genes and tissue metabolites in their capacity to classify outcome-related molecular subgroups.

Identified serum metabolites are associated with fatty acid metabolism and secondary metabolism of microbial metabolites.

We next ascertained the pathways that were present and/or altered among the metabolite-defined clusters. To do so, we performed pathway analysis based on the metabolic gene expression among the MetGene clusters. When comparing the 491 metabolic gene expression between samples from different MetGene clusters using Ingenuity Pathway Analysis (IPA)²⁵, we found that the top three pathways, namely lipid metabolism, i.e. LPS/IL-1 mediated inhibition of RXR function, fatty acid β -oxidation I, and PXR/RXR activation, were deregulated in all the comparisons between the three MetGene clusters, with $-\log_{10}(p\text{-values})$ of MetGene-C1/C2, MetGene-C1/C3, and MetGene-C2/C3 as 21.7, 24.4, and 21.7, respectively (Supplementary Fig. S5 and Supplementary Table S5). The results concurred with serum metabolite data, where short-chain and long-chain acylcarnitines were differentially abundant among the three serum metabolite clusters (Fig. 4A and Supplementary Fig. S6).

To further confirm that these dysregulated pathways based on gene expression levels are truly associated with lipid metabolism, we determined the composition of the identified 75 serum metabolites by tabulating the metabolites according to their classes and subclasses. Almost half (46%) of the identified serum metabolites were lipids (Supplementary Fig. S2B), which consisting mainly of acylcarnitines (28%), phosphatidylcholine and lysophosphatidylcholine (PC and lysoPC—25%), while the remainder are other assorted lysolipids, steroids and fatty acids (Supplementary Fig. S2C). One class of serum metabolites with highest abundance, namely acylcarnitines, is strongly correlated with tissue metabolites (Fig. 4C, D) and reflect the pathway analysis based on tumor and adjacent non-tumor matched tissue gene expression, with HNF4a as possible transcriptional regulator (Supplementary Figs. S7A, B and S8). Notable metabolites from this class are short-chain acylcarnitines (butyrylcarnitine, 2-methylmalonyl carnitine, and glutarylcarnitine) and long-chain acylcarnitines (laurylcarnitine, myristoleoylcarnitine, and oleoylcarnitine). The median relative abundance levels in log₁₀ of these metabolites for SerumMet-C1, C2, and C3 are as follow: butyrylcarnitine—23.32, 23.44, 23.50; 2-methylmalonyl carnitine—17.9, 18.25, 18.13; glutarylcarnitine—21.40, 21.81, 21.59; laurylcarnitine—21.0, 20.42, 18.23; Myristoleoylcarnitine—21.81, 21.67, 18.23; and Oleoylcarnitine—22.87, 22.92, 21.35. Acylcarnitines are intermediate metabolites in fatty acid metabolism, specifically in process of fatty acid transportation from the cytosol to the mitochondria during fatty acid β -oxidation whereby fatty acids are broken down. The next most abundant class of serum metabolites is lysoPC, which constitutes the majority of biological membranes and is easily obtained through the diet. Other metabolite classes were amino acids, xenobiotics, peptides, carbohydrates, cofactors and vitamins, and energy metabolites. One other notable class of metabolites observed among the 75-serum metabolite set is microbial metabolites, which consisted of *p*-cresol sulfate (PCS), 4-ethylphenyl sulfate (4-EPS), and 4-methylcatechol (Fig. 4A, B, Supplementary Fig. S9 and Supplementary Table S5). The median relative abundance levels in log₁₀ of these metabolites for SerumMet-C1, C2, and C3 are as follow: *p*-cresol sulfate—28.62, 26.95, 27.72; 4-ethylphenyl sulfate—23.84, 18.32, 21.05; and 4-mythylcatechol sulfate—22.64, 21.72, 22.51.

Finally, to determine whether clinical variables were associated with various subtypes identified in this study, we performed a Fisher’s Exact test between these clinical variables and the subtypes. We found that most of the clinical features are independent of these subtypes, with the two exceptions of AFP/MetGene subtypes and BMI/SerumMet subtypes (Supplementary Table S5). The samples with AFP level > 300 ng/ml are clustered in the MetGene-C1 subgroup, which has worse prognosis compared to other subgroups. The group with BMI > 24 (overweight patients) is linked with the SerumMet-C2 subgroup, which has better prognosis compared to other

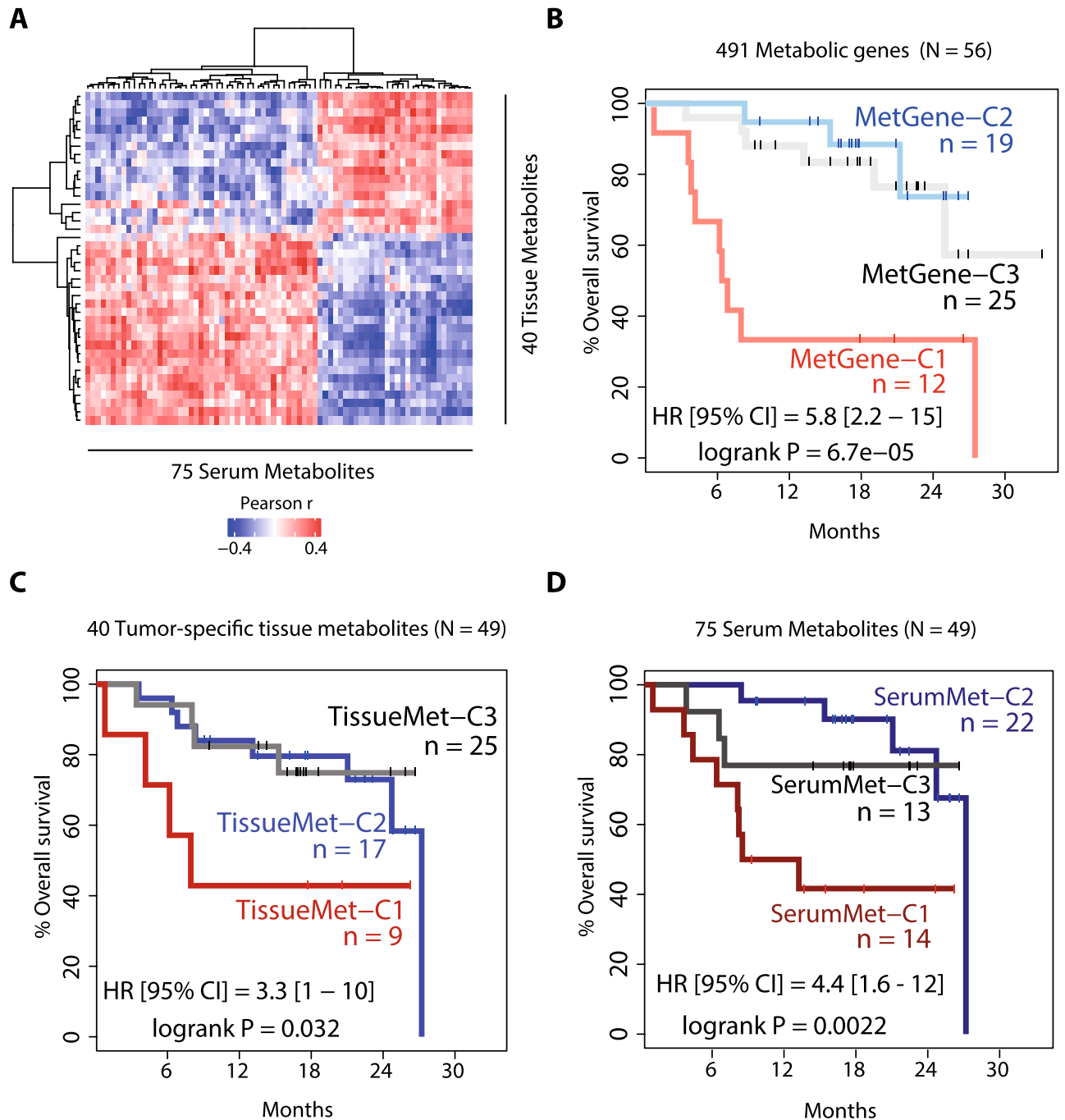


Figure 3. Correlation analysis of tissue and serum metabolites. **(A)** Heatmap of correlation coefficients between 40 tumor-specific tissue metabolites and 75 serum metabolites identified in this study. **(B)** Kaplan–Meier plot showing survival probabilities of HCC patients from TIGER-LC cohort according to consensus clustering based on expression of 491 metabolic genes. **(C)** Kaplan–Meier plot showing survival probabilities of HCC patients from TIGER-LC cohort according to consensus clustering based on abundance of 40 tumor-specific metabolites. **(D)** Kaplan–Meier plot showing survival probabilities of HCC patients from TIGER-LC cohort according to consensus clustering based on abundance of 75 serum metabolites. Abbreviations: HR [95% CI]—Hazard Ratio with 95% confidence interval; MetGene-C—Metabolic gene-Cluster; TissueMet-C—Tissue metabolite-Cluster; SerumMet-C—Serum metabolite-Cluster.

subgroups. This data concurs with the notion that short-chain acylcarnitines in SerumMet-C2 subgroup have the most abundant level among all the subgroups.

Discussion

In this study, we present a set of serum metabolites that can potentially be used as non-invasive prognostic biomarkers that are reflective of tumor-tissue status. The multiple levels of metabolite-related data can independently stratify Asian HCC patients into molecular subtypes that have similar patient outcomes. Compared to our previous study that utilized whole transcriptome analysis to identify HCC molecular subtypes¹³, here, we have considerably decreased the number of markers that are required for patient stratification without compromising the subgroup consistency. Furthermore, the identified metabolic genes and metabolites gave new insights into the molecular underpinnings of Asian HCC.

Serum metabolite measurement is a potential non-invasive tool for early detection and prognosis determination of HCC. Compared to other noninvasive techniques, such as ultrasonography, computed tomography, or magnetic resonance imaging (MRI), serum metabolites are considerably less cumbersome in terms of implementation and results interpretation. Previous publications have identified several fatty acids and acylcarnitines in serum and plasma as early detection and prognostic biomarkers of HCC, indicating that fatty acid metabolism is universally dysregulated in HCC^{12,26–29}. However, to our knowledge, there is no study which directly correlates the chain of events between gene expression and metabolites from tumor tissues to serum. Our study therefore may present a more accurate picture of the molecular underpinnings of HCC.

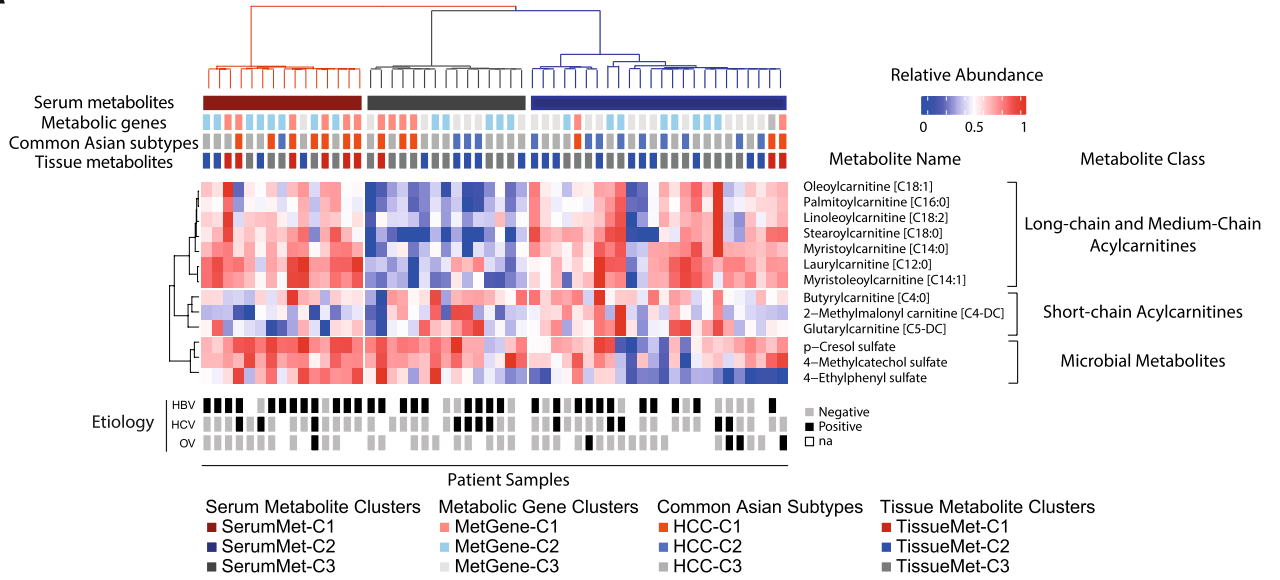
The methodology employed in this study, namely two-step correlation analysis by first correlating tissue gene expression and metabolite and then correlating tissue metabolites and serum metabolites, has unique advantages over single step correlation. This approach combines permutation-based correlation threshold selection with correlation analysis to objectively identify correlated genes and metabolites. The approach also utilized tissue gene expression, tumor tissue and serum metabolites to identify the plausible causal chain of events that should directly reflect molecular processes from tumor tissues to tissue metabolites, and finally to serum metabolites. To our knowledge, this is the first study that utilized these data in this way. This method, combined with gene expression and metabolomics data, allows us to directly observe the chain of associations between gene expression and metabolites in tumor tissue, and between tumor tissue and serum metabolites. It enables us to see a convergence of evidence from multiple levels of data, from which patient subgroups resulting independently from metabolic gene expression, tissue, and serum metabolites had similar patient outcome.

The set of 75 serum metabolites was able to independently stratify patients into subgroups that were also consistent with patient subgroups from metabolic gene expression, tumor tissue metabolites, and whole transcriptome from our previous study¹³. The consistency in these clustering results is another indicator that the identified metabolic genes and metabolites are involved in the disease manifestation. Furthermore, the numbers of genes and metabolites involved in sample stratification were significantly lower compared to our previous study, which will be beneficial for further utilization of these genes and metabolites as prognostic biomarkers. Finally, serum metabolites are more accessible and less invasive than gene expression or metabolites from tumor tissues, therefore they are more suitable as less invasive prognostic biomarkers for Asian HCC patients. This set of serum metabolites can be validated in larger patient cohorts in the future to facilitate HCC patient prognostication in clinical settings.

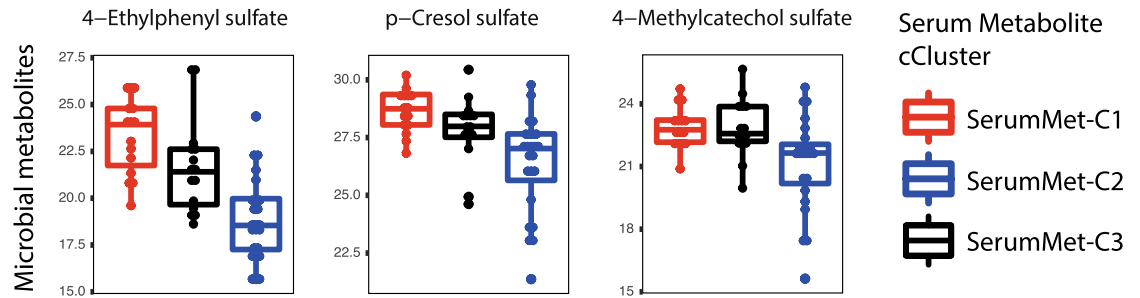
Lipids and fatty acid metabolism play central roles in all cancer types, including HCC³⁰. Pathway analyses based on the abundance of both tumor tissue and serum metabolites, combined with tissue gene expression, indicated that lipid and fatty acid metabolism are among the top dysregulated metabolic pathways in HCC. The roles of one class of fatty acid-related metabolites of interest, i.e., acylcarnitines, are slowly emerging in HCC. Acylcarnitines are known as intermediate metabolites in β -oxidation of fatty acids in mitochondria^{31,32}. Short- and medium-chain acylcarnitines largely reside in peroxisomes and produced mainly by liver and muscle^{33,34}. A recent study showed that acylcarnitine profiles can be used as biomarkers for obesity-driven HCC, pinpointed a serum metabolite, oleoylcarnitine (C18:1 carnitine), which was found in this study, as one of the biomarkers for NAFLD and HCC, and revealed the mechanism of actions of oleoylcarnitine through STAT3 pathway activation³¹. The same study also identified several other acylcarnitines that were also identified in this study, namely laurylcarnitine (C12:0) and myristoleoylcarnitine (C14:1), as possible HCC biomarkers. For short- and medium-chain acylcarnitines, the evidence of the involvement in HCC are limited. Possible master regulators for these metabolites are HNF4 α , which were found in the pathway analysis and interacts with CPT1 and CPT2. HNF4 α was found to be the main regulator for acylcarnitines in brown adipose tissue in mice, in which the brown tissue increases the uptake of acylcarnitines during the cold³⁵. This will need further analyses to confirm the actual link between HNF4 α and acylcarnitines in human HCC.

The gut microbiome has been shown to play several important roles in liver diseases. Several mechanisms involving the gut microbiome have been proposed, along with approaches to mitigate, prevent, or even reverse HCC³⁶. Many essential metabolites and nutrients are produced by gut microbes to provide precursors for more complex molecules in human, such as hormones, vitamins, and other cofactors. Some metabolites identified in this study, i.e., 4-EPS and PCS, are microbial metabolites derived from the diet³⁷. Several studies found that PCS is a highly potent uremic toxin³⁸ and can induce cell proliferation and migration of kidney cancer³⁹. The roles of these compounds in HCC are still obscure. *p*-cresol is produced mainly by *Clostridioides difficile* (also known as *Clostridium difficile* or *C. diff*)⁴⁰. A recent study predicted that there may be other microbes which can also produce *p*-cresol, but this needs to be confirmed⁴¹. *p*-cresol is metabolized in the liver through a conjugation process, i.e. sulfonation, to produce a less toxic form, PCS⁴². There are several possible mechanisms that might be specifically linked to PCS. It was found through molecular docking simulations that PCS can activate epidermal growth factor receptor (EGFR) via interdomain pocket interaction of extracellular EGFR, and finally induce kidney tissue remodeling⁴³. A recent study showed that EGFR signaling promotes survival of HCC cell lines via interaction with E-cadherin and β -catenin, but only in late stages of the disease⁴⁴. Another possible mechanism is that PCS might increase liver detoxification burden, especially sulfonation, affecting the ability of liver to detoxify

A

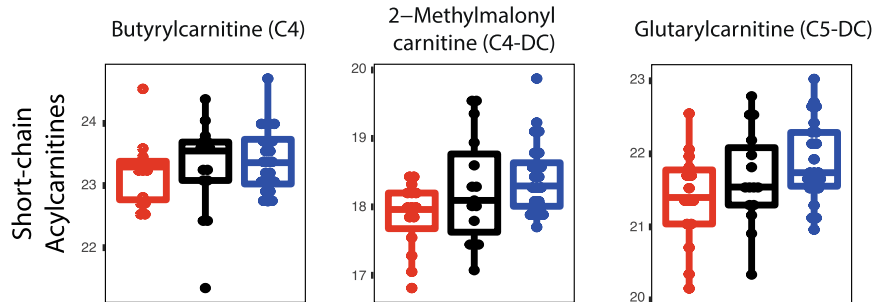


B

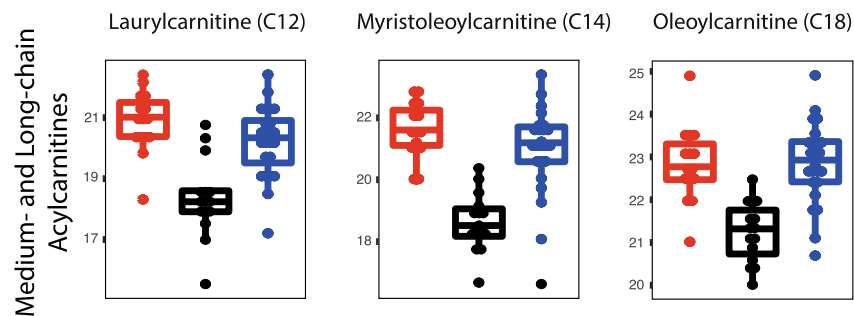


C

Log₁₀ Metabolite Abundance



D



◀ **Figure 4.** Patient clustering by serum metabolites and identification of prognostic serum metabolites. (A) Heatmap of metabolite abundance from three selected groups of metabolites of interest: short-chain acylcarnitines; long-chain acylcarnitines; and microbial metabolites. The top panel is patient clusters based on various consensus clustering results: first row is consensus clustering results based on 75 serum metabolites; second row is consensus clustering results based on 491 metabolic genes; third row is the original consensus clustering of TIGER-LC HCC samples; and fourth row is consensus clustering results based on 40 tumor-tissue metabolites. (B–D) Log-scale metabolite abundance. (B) Microbial metabolites (p-cresol sulfate, 4-ethylphenyl sulfate, and 4-methylcatechol sulfate). (C) Short-chain acylcarnitines (Butyrylcarnitine [C4:1], 2-methylmalonyl carnitine [C4-DC] and Glutarylcarnitine [C5-DC]). (D) Medium- (laurylcarnitine [C12:1]) and long-chain acylcarnitines (Myristoleoylcarnitine, [C14:1] and Oleoylcarnitine [C18:0]). The y-axis is the log₁₀ metabolite abundance. The x-axis is the SerumMet clusters. Abbreviations: MetGene-C—Metabolic gene-Cluster; TissueMet-C—Tissue metabolite-Cluster; SerumMet-C—Serum metabolite-Cluster; HCC—Hepatocellular Carcinoma.

other highly potent liver toxins, such as acetaminophen (paracetamol) or aflatoxin⁴⁵. This study also proposed a less invasive means for measuring PCS through urine in HCC patients⁴⁵. PCS may also be associated with the microbiome in that PCS-producing *C. diff* was found to be dominating in the gut of mice, due to toxicity of PCS to other bacterial species in the gut⁴⁶ and may be the mechanism of action of dysbiosis from *C. diff* infection. Therefore, PCS might be a key link between gut dysbiosis and HCC pathogenesis.

Biomarker discovery for HCC is an active field of research. The metabolites mapped onto the KEGG's Human Metabolism Map in Fig. 1C were previously found to be dysregulated in HCC and have been proposed as either prognostic or diagnostic biomarkers, but independently they were not able to outperform traditional HCC biomarkers. To further explore these dysregulations, we have assessed tumor specific metabolic gene alterations and serum alterations and determined their association with clinical features including patient outcome. The specific alterations and how they apply to HCC tumorigenesis would require detailed molecular and functional studies which could be addressed in future studies. Numerous studies have been conducted to search for suitable serum biomarkers that can predict survival outcome of patients in different populations^{47–50}. Currently, however, clinically accepted prognostic serum biomarkers are limited to only α -fetoprotein (AFP), AFP-L3, and Des- γ -carboxy prothrombin (DCP)^{51,52}. Various acylcarnitines were suggested as serum or plasma biomarkers in several Asian HCC cohorts, but none of these biomarkers have been validated in any clinical trials^{29,53,54}. The main reason for this may result from the inconsistency of the data from these studies. One study of PCS suggests that it could be a urine-based biomarker for differentiating HCC and cirrhotic patients. However, the study found that HCC patients have lower PCS level compared to cirrhotic patients and suggested that sulfonation alteration might be responsible mechanisms⁵⁵. Thus, additional comprehensive studies are needed to determine the consistency and effectiveness of such biomarkers.

In conclusion, we have established stable metabolic subtypes based on metabolic gene expression, tissue metabolites and importantly, serum metabolites, through a two-step correlation analysis. Metabolic gene subtypes can stratify patients into better and worse prognosis based on the metabolic gene expression across different Asian patient cohorts. Tissue and serum metabolites can also consistently stratify patients into subgroups similar to the metabolic gene subtypes and our previously identified common molecular Asian transcriptome-defined subtypes. Some of the serum metabolites that we identified might serve as prognostic biomarkers after future clinical validation. Validation in other cohorts should be a priority to translate this finding into clinical use. Publicly available resource with both gene expression and serum metabolite measurement are not currently available to further validate our findings. If and when such resources are available, we should have clearer picture of the serum metabolites associated with HCC. Finally, future studies that link metabolic alterations with disease etiologies and clinical features would be of future interest to shed light on the underlying mechanisms contributing to HCC which could be utilized in clinical management.

Methods

Cohorts, clinical, and genomic data. Tissue gene expression of the TIGER-LC cohort¹³, LCI cohort²⁰ and TCGA liver cohort²⁴, were used in this analysis. The list of metabolic genes (3765 genes) were retrieved from Human Metabolic Reaction 2.0 (HMR2.0) database⁵⁶. Tissue metabolite data were described previously among the TIGER-LC cohort¹³. In brief, metabolites were measured in tumor tissues and serum using Metabolon's Discover HD4 Platform, which is an untargeted metabolomics approach that combines positive and negative modes of liquid chromatography (LC+/LC-) and gas chromatography/mass spectrometry (GS/MS) to measure metabolite abundance. After data processing, 718 and 990 metabolites were detected from tumor tissues and serum samples, respectively. After excluding unidentified metabolites (metabolites that cannot be matched to standard library compounds), there were 562 and 615 metabolites remaining for further analyses from tumor tissues and serum, respectively. Clinical features data included in this study are

1. Mutation status of TP53 and CTNNB1 genes, determined in our previous study¹³
2. Sex
3. Age, with two categories of ≤ 50 and > 50 years old
4. Body Mass Index (BMI), with two categories of BMI ≤ 24 and BMI > 24
5. Diabetes status, with two categories of no diabetes and with diabetes
6. CA-19-9 level, with two categories of CA-19-9 ≤ 37 U/ml and CA-19-9 > 37 U/ml
7. AFP level, with two categories of AFP ≤ 300 ng/ml and AFP > 300 ng/ml

8. Cirrhosis status, with two categories of Cirrhosis Positive and Cirrhosis Negative
9. TNM classification of malignant tumors, with two categories of early stage I and II, and late stage of IIIA, IIIB, and IV
10. Hepatitis B virus (HBV) infection status, with two categories of HBV Positive and HBV Negative
11. Hepatitis C virus (HCV) infection status, with two categories of HCV Positive and HCV Negative
12. *Orpisthorchis viverrini* (OV) or liver fluke infection status, with two categories of OV Positive and OV Negative

3D PCA of tissue metabolites. Three-dimensional principal component analysis projections of tumor and adjacent non-tumor tissue metabolites were constructed based on the 500 most variable metabolites. The most variable metabolites were selected based on variance of each metabolites among all HCC samples.

KEGG human and microbial metabolic maps. The KEGG map overlay with metabolite ratio was performed using the R package pathview⁵⁷. The metabolite names are converted to KEGG compound IDs, and the maps required are downloaded through KEGG API by KEGGREST package in R⁵⁸. The tissue metabolites from the TIGER-LC cohort were overlaid onto the KEGG human metabolite map¹⁵ (map accession number: hsa01100, retrieved on April 5, 2017) and reconstructed based on the TIGER-LC tissue metabolite ratio between the mean of tumor tissues and adjacent non-tumor tissues. The log₂-ratio of metabolites between tumor and adjacent non-tumor tissues were calculated and used for map overlay. The serum metabolites were overlaid onto the KEGG Microbial Metabolism in Diverse Environments map (map accession number: KO01120, retrieved on April 10, 2021) and reconstructed based on the serum metabolite ratio between samples from SerumMet-C1 and SerumMet-C2 clusters.

Two-step correlation analysis and correlation coefficient threshold selection by permutation analysis between metabolic gene expression, tissue metabolites, and serum metabolites. There are 3765 genes classified as human metabolic genes according to Human Metabolic Atlas database version 2.0⁵⁹, of which approximately 2000 genes overlapped with the gene expression data from our previous study¹³. Metabolic gene expression between tumor and adjacent non-tumor tissue were compared using a paired t-test. The resulting *p*-values were adjusted using Benjamini–Hochberg false discovery rate (FDR) procedure⁶⁰. Metabolic genes with FDR-adjusted *p*-values of less than 0.05 were retained for further correlation analysis. Spearman correlation coefficients between the expression of 2300 metabolic genes from tumor tissues and 562 metabolites from the TIGER-LC cohort were calculated.

For correlation threshold selection, the gene names, metabolite names, and the sample names in both metabolic genes and metabolite matrices were permuted to destroy original relationships of the original data. Then, the correlation coefficient between permuted metabolic gene expression and metabolites were calculated and recorded with 10,000 permutations. The distribution of correlation coefficients from the original data was plotted against the average of distributions of correlation coefficients from the permuted data. Finally, the intersection points of correlation coefficients between the original data and the average of permuted data was determined to be -0.165 and 0.165 (Supplementary Fig. S10) and were used as correlation coefficient thresholds to select the correlation coefficients between metabolic genes and metabolites for further analyses.

The metabolic genes and tissue metabolites were selected by calculating the average value of the sum of absolute correlation coefficients from the following equations:

$$G_i = \frac{\sum_{j=1}^J |c_{ij}|}{J} \quad \text{and} \quad T_j = \frac{\sum_{i=1}^I |c_{ji}|}{I}$$

where $|c_{ij}|$ and $|c_{ji}|$ are absolute correlation coefficient between metabolic gene *i* and tissue metabolite *j*, *I* is total number of metabolic genes, *J* is total number of tissue metabolites, G_i is the average value of sum of absolute correlation coefficients of metabolic gene *i* over tissue metabolites, and T_j is the average value of sum of absolute correlation coefficients of tissue metabolite *j* over all metabolic genes. G and T that are higher than the correlation coefficient threshold identified by permutation test are then selected for further analyses.

For the first step of correlation, the number of metabolic genes and tissue metabolites from correlation coefficient threshold selection is 491 genes and 78 metabolites, respectively. Tissue metabolites that are tumor-specific was determined based on a metabolite fold-change of 1.5 between tumor and non-tumor adjacent tissues, which yielded 40 tumor-specific tissue metabolites that were used in the second step of correlation with serum metabolites.

For the second step of correlation, correlation thresholds used for correlation analysis between tissue and serum metabolites are -0.15 and 0.15. These thresholds correspond to the sums of correlation coefficients between tumor tissue gene expression and tissue metabolites previously identified. Serum metabolite selection, the same strategy of average of the sum of absolute correlation coefficient is used, following the equation:

$$S_k = \frac{\sum_{m=1}^M |c_{km}|}{M}$$

where $|c_{km}|$ is absolute value of correlation coefficient between serum metabolite *k* and tumor-specific tissue metabolite *m*, *M* is total number of tumor-specific tissue metabolites, and S_k is average value of sum of absolute correlation coefficient of serum metabolite *k* over all tumor-specific tissue metabolites. The final number of serum metabolites that passed the threshold is 75 metabolites.

Consensus clustering analysis and subclass mapping analysis. To define consensus clusters for the patient samples among the Thai cohort, the matrix of 491 metabolic genes was used as the input for consensus clustering⁶¹. Two other cohorts were used to validate identified metabolic genes, namely TCGA-LIHC²⁴ and LCI²⁰ cohorts. The TCGA-LIHC cohort is comprised of liver cancer patients from across US cancer centers as a part of The Cancer Genome Atlas project. The LCI cohort is comprised of Chinese liver cancer patients from the Liver Cancer Institute. In our previous study we found that Asian patients had common molecular features, which are different than patients with European or African descent¹³. Therefore, in this study we only used Asian descent samples from TCGA-LIHC. The data matrices from these two cohorts were obtained by using the original 491 genes from the TIGER-LC cohort analysis and identifying the overlapping genes from the TCGA-LIHC or LCI data sets. The LCI cohort has 385 overlapping genes, while TCGA-LIHC cohort has 461 overlapping genes. Consensus clustering was performed by using ConsensusClusterPlus¹⁶ package version 1.38 in R version 3.3.3. The reason for using consensus clustering in this study is to determine whether the sets of metabolic gene expression, tumor tissue, and serum metabolites can independently stratify patients into subgroups that have similar consistency amongst different molecular data types. The parameters used in all the analyses were as follows: maxK = 8, number of bootstrap = 20,000; item subsampling proportion = 0.8; feature subsampling proportion = 1; cluster algorithm = hc [hierarchical clustering]; inner linkage type = ward.D; final linkage type = ward.D; correlation method = spearman; seed = 3044. The final number of clusters (k) selected for further analyses is three.

The clustering results from the LCI and TCGA-Asian cohorts were then subjected to Subclass Mapping (SubMap)⁶² analysis to determine whether the clusters from these two cohorts matched the clusters from the TIGER-LC cohort. SubMap version 7.0 from GenePattern software suite⁶³ was used with default parameters. After two-step correlation analysis, the list of 40 tumor-specific tissue metabolites and 75 serum metabolites were also used for consensus clustering with the same parameters previously described to confirm that the metabolites identified by correlation analysis can independently and consistently stratify patients into similar subgroups. Submap analysis was also used to determine the if the TIGER-LC samples matched with other HCC signatures. The HCC signatures used in this study were S1-S3 signature based on WNT-TGF β , AKT-MYC signaling pathway activation, and hepatocyte differentiation¹⁷, stem cell signature based on expression of stem cell gene set¹⁸, EpCAM signature based on a gene set that is co-expressed with EpCAM¹⁹, metastatic signature based on a gene set related to metastasis of HCC²⁰, CCA-like signature based on a gene set from cholangiocarcinoma-like HCC²¹, Cluster 1 and 2 signatures based on 238-gene signature from CCA patients²², and proliferation signature based on copy number alterations and activation of oncogenic pathways²³.

Statistical and survival analyses. All statistical and survival analyses were performed using R statistical computing software version 3.3.3. Survival analysis on all cohorts was performed using a Kaplan Meier estimator and a Cox proportional hazards model was calculated between the worst survival group and the rest of the groups, accompanied with Log-rank *p*-value for each comparison. The Cox proportional hazard assumption was tested and met for all the tests carried out. The Paired t-test was used to calculate differential gene expression between tumor and adjacent non-tumor tissues. All *p*-values are two-sided *p*-values unless otherwise stated and adjusted for false discovery rate by Benjamini–Hochberg FDR procedure⁶⁴.

Pathway analyses. The final 491 metabolic genes were subjected to pathway analyses using GSEA⁶⁵ version 18 on GenePattern software suite⁶³ for tumor vs non tumor comparison, and Ingenuity Pathway Analysis (IPA)⁶⁶ version 01-16 for tumor clusters comparison.

Received: 18 December 2020; Accepted: 24 May 2021

Published online: 08 June 2021

References

- Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424. <https://doi.org/10.3322/caac.21492> (2018).
- Ferlay, J. *et al.* Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* **144**, 1941–1953. <https://doi.org/10.1002/ijc.31937> (2019).
- Leroi, C. *et al.* Prevalence of chronic hepatitis B virus infection in Thailand: a systematic review and meta-analysis. *Int. J. Infect. Dis.* **51**, 36–43. <https://doi.org/10.1016/j.ijid.2016.08.017> (2016).
- Hollstein, M. C. *et al.* p53 mutations and aflatoxin B1 exposure in hepatocellular carcinoma patients from Thailand. *Int. J. Cancer* **53**, 51–55. <https://doi.org/10.1002/ijc.2910530111> (1993).
- Gouas, D., Shi, H. & Hainaut, P. The aflatoxin-induced TP53 mutation at codon 249 (R249S): biomarker of exposure, early detection and target for therapy. *Cancer Lett.* **286**, 29–37. <https://doi.org/10.1016/j.canlet.2009.02.057> (2009).
- Wasitthanasem, R. *et al.* Liver disease burden and required treatment expenditures for hepatitis C virus (HCV) infection in Thailand: implications for HCV elimination in the new therapeutic era, a population-based study. *PLoS ONE* **13**, e0196301. <https://doi.org/10.1371/journal.pone.0196301> (2018).
- Fan, J. G., Kim, S. U. & Wong, V. W. New trends on obesity and NAFLD in Asia. *J. Hepatol.* **67**, 862–873. <https://doi.org/10.1016/j.jhep.2017.06.003> (2017).
- Anstee, Q. M., Reeves, H. L., Kotsiliti, E., Govaere, O. & Heikenwalder, M. From NASH to HCC: current concepts and future challenges. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 411–428. <https://doi.org/10.1038/s41575-019-0145-7> (2019).
- Yang, J. D. *et al.* A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 589–604. <https://doi.org/10.1038/s41575-019-0186-y> (2019).
- Llovet, J. M. *et al.* Hepatocellular carcinoma. *Nat. Rev. Dis. Primers* **2**, 16018. <https://doi.org/10.1038/nrdp.2016.18> (2016).

11. Chaiteerakij, R., Addissie, B. D. & Roberts, L. R. Update on biomarkers of hepatocellular carcinoma. *Clin. Gastroenterol. Hepatol.* **13**, 237–245. <https://doi.org/10.1016/j.cgh.2013.10.038> (2015).
12. Budhu, A. *et al.* Integrated metabolite and gene expression profiles identify lipid biomarkers associated with progression of hepatocellular carcinoma and patient outcomes. *Gastroenterology* **144**, 1066–1075e1061. <https://doi.org/10.1053/j.gastro.2013.01.054> (2013).
13. Chaisaingmongkol, J. *et al.* Common molecular subtypes among Asian hepatocellular carcinoma and cholangiocarcinoma. *Cancer Cell* **32**, 57–70e53. <https://doi.org/10.1016/j.ccell.2017.05.009> (2017).
14. Llovet, J. M., Montal, R., Sia, D. & Finn, R. S. Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat. Rev. Clin. Oncol.* **15**, 599–616. <https://doi.org/10.1038/s41571-018-0073-4> (2018).
15. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
16. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573. <https://doi.org/10.1093/bioinformatics/btq170> (2010).
17. Hoshida, Y. *et al.* Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res.* **69**, 7385–7392. <https://doi.org/10.1158/0008-5472.CAN-09-1089> (2009).
18. Lee, J. S. *et al.* Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* **40**, 667–676. <https://doi.org/10.1002/hep.20375> (2004).
19. Yamashita, T. *et al.* EpCAM and alpha-fetoprotein expression defines novel prognostic subtypes of hepatocellular carcinoma. *Cancer Res.* **68**, 1451–1461. <https://doi.org/10.1158/0008-5472.CAN-07-6013> (2008).
20. Roessler, S. *et al.* A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res.* **70**, 10202–10212. <https://doi.org/10.1158/0008-5472.CAN-10-2607> (2010).
21. Woo, H. G. *et al.* Identification of a cholangiocarcinoma-like gene expression trait in hepatocellular carcinoma. *Cancer Res.* **70**, 3034–3041. <https://doi.org/10.1158/0008-5472.CAN-09-2823> (2010).
22. Andersen, J. B. *et al.* Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. *Gastroenterology* **142**, 1021–1031e1015. <https://doi.org/10.1053/j.gastro.2011.12.005> (2012).
23. Sia, D. *et al.* Integrative molecular analysis of intrahepatic cholangiocarcinoma reveals 2 classes that have different outcomes. *Gastroenterology* **144**, 829–840. <https://doi.org/10.1053/j.gastro.2013.01.001> (2013).
24. Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* **169**, 1327–1341e1323. <https://doi.org/10.1016/j.cell.2017.05.046> (2017).
25. QIAGEN Inc. *Ingenuity Pathway Analysis (IPA)*. <https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/> (2019).
26. Huang, Q. *et al.* Metabolic characterization of hepatocellular carcinoma using nontargeted tissue metabolomics. *Cancer Res.* **73**, 4992–5002. <https://doi.org/10.1158/0008-5472.CAN-13-0308> (2013).
27. Lu, Y. *et al.* Acetylcarnitine is a candidate diagnostic and prognostic biomarker of hepatocellular carcinoma. *Cancer Res.* **76**, 2912–2920. <https://doi.org/10.1158/0008-5472.CAN-15-3199> (2016).
28. Wang, B. *et al.* Metabonomic profiles discriminate hepatocellular carcinoma from liver cirrhosis by ultraperformance liquid chromatography-mass spectrometry. *J. Proteome Res.* **11**, 1217–1227. <https://doi.org/10.1021/pr2009252> (2012).
29. Liu, Y. *et al.* NMR and LC/MS-based global metabolomics to identify serum biomarkers differentiating hepatocellular carcinoma from liver cirrhosis. *Int. J. Cancer* **135**, 658–668. <https://doi.org/10.1002/ijc.28706> (2014).
30. Peck, B. & Schulze, A. Lipid metabolism at the nexus of diet and tumor microenvironment. *Trends Cancer* **5**, 693–703. <https://doi.org/10.1016/j.trecan.2019.09.007> (2019).
31. Fujiwara, N. *et al.* CPT2 downregulation adapts HCC to lipid-rich environment and promotes carcinogenesis via acylcarnitine accumulation in obesity. *Gut* **67**, 1493–1504. <https://doi.org/10.1136/gutjnl-2017-315193> (2018).
32. Bjorndal, B. *et al.* Associations between fatty acid oxidation, hepatic mitochondrial function, and plasma acylcarnitine levels in mice. *Nutr. Metab.* <https://doi.org/10.1186/s12986-018-0241-7> (2018).
33. Xu, G. *et al.* Liver and muscle contribute differently to the plasma acylcarnitine pool during fasting and exercise in humans. *J. Clin. Endocrinol. Metab.* **101**, 5044–5052. <https://doi.org/10.1210/jc.2016-1859> (2016).
34. Violante, S. *et al.* Peroxisomes contribute to the acylcarnitine production when the carnitine shuttle is deficient. *Biochim. Biophys. Acta* **1467–1474**, 2013. <https://doi.org/10.1016/j.bbali.2013.06.007> (1831).
35. Simcox, J. *et al.* Global analysis of plasma lipids identifies liver-derived acylcarnitines as a fuel source for brown fat thermogenesis. *Cell Metab.* **26**, 509–522e506. <https://doi.org/10.1016/j.cmet.2017.08.006> (2017).
36. Yu, L. X. & Schwabe, R. F. The gut microbiome and liver cancer: mechanisms and clinical translation. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 527–539. <https://doi.org/10.1038/nrgastro.2017.72> (2017).
37. Zhang, L. S. & Davies, S. S. Microbial metabolism of dietary components to bioactive metabolites: opportunities for new therapeutic interventions. *Genom. Med.* **8**, 46. <https://doi.org/10.1186/s13073-016-0296-x> (2016).
38. Vanholder, R., Schepers, E., Pletinck, A., Nagler, E. V. & Glorieux, G. The uremic toxicity of indoxyl sulfate and p-cresyl sulfate: a systematic review. *J. Am. Soc. Nephrol.* **25**, 1897–1907. <https://doi.org/10.1681/ASN.2013101062> (2014).
39. Wu, T. K. *et al.* The uremic toxin p-cresyl sulfate induces proliferation and migration of clear cell renal cell carcinoma via microRNA-21/ HIF-1alpha axis signals. *Sci. Rep.* **9**, 3207. <https://doi.org/10.1038/s41598-019-39646-9> (2019).
40. Dawson, L. F. *et al.* The analysis of para-cresol production and tolerance in *Clostridium difficile* 027 and 012 strains. *BMC Microbiol.* **11**, 86. <https://doi.org/10.1186/1471-2180-11-86> (2011).
41. Saito, Y., Sato, T., Nomoto, K. & Tsuji, H. Identification of phenol- and p-cresol-producing intestinal bacteria by using media supplemented with tyrosine and its metabolites. *FEMS Microbiol. Ecol.* <https://doi.org/10.1093/femsec/fiy125> (2018).
42. Ogata, N., Matsushima, N. & Shibata, T. Pharmacokinetics of wood creosote: glucuronic acid and sulfate conjugation of phenolic compounds. *Pharmacology* **51**, 195–204. <https://doi.org/10.1159/000139335> (1995).
43. Sun, C. Y. *et al.* Protein-bound uremic toxins induce tissue remodeling by targeting the EGF receptor. *J. Am. Soc. Nephrol.* **26**, 281–290. <https://doi.org/10.1681/ASN.2014010021> (2015).
44. Kim, E. *et al.* Promotion of growth factor signaling as a critical function of beta-catenin during HCC progression. *Nat. Commun.* **10**, 1909. <https://doi.org/10.1038/s41467-019-09780-z> (2019).
45. Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R. & Nicholson, J. K. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14728–14733. <https://doi.org/10.1073/pnas.0904489106> (2009).
46. Passmore, I. J. *et al.* Para-cresol production by *Clostridium difficile* affects microbial diversity and membrane integrity of gram-negative bacteria. *PLoS Pathog.* **14**, e1007191. <https://doi.org/10.1371/journal.ppat.1007191> (2018).
47. Chauhan, R. & Lahiri, N. Tissue- and serum-associated biomarkers of hepatocellular carcinoma. *Biomark. Cancer* <https://doi.org/10.4137/bic.S34413> (2016).
48. Lou, J., Zhang, L., Lv, S., Zhang, C. & Jiang, S. Biomarkers for hepatocellular carcinoma. *Biomark. Cancer* <https://doi.org/10.1177/1179299x16684640> (2017).
49. Ocker, M. Biomarkers for hepatocellular carcinoma: what's new on the horizon?. *World J. Gastroenterol.* **24**, 3974–3979. <https://doi.org/10.3748/wjg.v24.i35.3974> (2018).

50. Liu, X. N. *et al.* Multiple “Omics” data-based biomarker screening for hepatocellular carcinoma diagnosis. *World J. Gastroenterol.* **25**, 4199–4212. <https://doi.org/10.3748/wjg.v25.i30.4199> (2019).
51. Marrero, J. A. *et al.* Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. *Clin. Liver Dis.* **13**, 1. <https://doi.org/10.1002/cld.802> (2019).
52. European Association for the Study of the Liver. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J. Hepatol.* **69**, 182–236. <https://doi.org/10.1016/j.jhep.2018.03.019> (2018).
53. Lu, Y. *et al.* Acetylcarnitine is a candidate diagnostic and prognostic biomarker of hepatocellular carcinoma. *Can. Res.* **76**, 2912–2920. <https://doi.org/10.1158/0008-5472.Can-15-3199> (2016).
54. Chen, S. *et al.* Pseudotargeted metabolomics method and its application in serum biomarker discovery for hepatocellular carcinoma based on ultra high-performance liquid chromatography/triple quadrupole mass spectrometry. *Anal. Chem.* **85**, 8326–8333. <https://doi.org/10.1021/ac4016787> (2013).
55. Shariff, M. I. F. *et al.* Urinary metabotyping of hepatocellular carcinoma in a UK cohort using proton nuclear magnetic resonance spectroscopy. *J. Clin. Exp. Hepatol.* **6**, 186–194. <https://doi.org/10.1016/j.jceh.2016.03.003> (2016).
56. Mardinoglu, A. *et al.* Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 3083. <https://doi.org/10.1038/ncomms4083> (2014).
57. Luo, W. & Brouwer, C. Pathview: an R/bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **29**, 1830–1831. <https://doi.org/10.1093/bioinformatics/btt285> (2013).
58. Tenenbaum, D. KEGGREST: Client-Side REST Access to KEGG (2018).
59. Pornputtpong, N., Nookaew, I. & Nielsen, J. Human metabolic atlas: an online resource for human metabolism. *Database (Oxford)* **2015**, bav068. <https://doi.org/10.1093/database/bav068> (2015).
60. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).
61. Monti, S., Tamayo, P., Mesirov, J. & Golub, T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* **52**, 91–118. <https://doi.org/10.1023/A:1023949509487> (2003).
62. Hoshida, Y., Brunet, J. P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS ONE* **2**, e1195. <https://doi.org/10.1371/journal.pone.0001195> (2007).
63. Reich, M. *et al.* GenePattern 2.0. *Nat. Genet.* **38**, 500–501. <https://doi.org/10.1038/ng0506-500> (2006).
64. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
65. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* **23**, 3251–3253. <https://doi.org/10.1093/bioinformatics/btm369> (2007).
66. Kramer, A., Green, J., Pollard, J. Jr. & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523–530. <https://doi.org/10.1093/bioinformatics/btt703> (2014).

Acknowledgements

We thank other members of the TIGER-LC Consortium, and the patients and families who contributed to this study. This work was supported in part by the intramural research program of the Center for Cancer Research, National Cancer Institute of the United States. Funding support for the establishment of the tissue/tumor biobank and collection of specimens and clinical diagnosis were received from the Chulabhorn Research Institute, Bangkok, Thailand.

Author contributions

X.W.W., Y.P., A.B., M.R. conceptualized and designed the experiment. Y.P., A.B., J.C., M.F., H.D. developed the methods. Y.P., A.B., X.W.W. analyzed and interpreted the data. Y.P., A.B., X.W.W. wrote the manuscript. X.W.W., C.M., M.R. supervised the study. All authors reviewed the manuscript.

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/s41598-021-91560-1>.

Correspondence and requests for materials should be addressed to X.W.W.

Reprints and permissions information is available at 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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2021

the TIGER-LC Consortium

Benjarath Pupacdi², Siritida Rabibhadana⁴, Kannikar Phonphutkul⁷, Nirush Lertprasertsuke⁸, Anon Chotirosniramit⁸, Chirayu U. Auewarakul⁹, Teerapat Ungtrakul⁹, Vajarabhongsa Budhisawasdi¹⁰, Chawalit Pairojkul¹⁰, Suleeporn Sangrajang¹¹, Curtis C. Harris¹, Christopher A. Loffredo¹² & Robert Wiltrot¹³

⁷Rajavej Hospital, Chiang Mai 50000, Thailand. ⁸Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand. ⁹Chulabhorn Royal Academy, Bangkok 10210, Thailand. ¹⁰Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand. ¹¹National Cancer Institute, Bangkok 10400, Thailand. ¹²Georgetown University Medical Center, Washington, DC 20057, USA. ¹³Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA.