Molecular Therapy
Oncology
Original Article

# Bile molecular landscape provides pathological insight and classifies signatures predictive of carcinoma of the gall bladder

Nupur Sharma,[1] Sadam H. Bhat,[1] Babu Mathew,[1] Manisha Yadav,[1] Gaurav Tripathi,[1] Vasundhra Bindal,[1] Sanju Yadav,[1] Neha Sharma,[1] Sushmita Pandey,[1] Hami Hemati,[4] Deepika Bohra,[3] Rashmi Rana,[3] Narender K. Sharma,[5] Sanyam Falari,[2] Viniyendra Pamecha,[2] and Jaswinder Singh Maras[1]

[1]Department of Molecular and Cellular Medicine, Institute of Liver and Biliary Sciences, New Delhi 110070, India; [2]Department of Liver Transplant and HepatoPancreato Biliary Surgery, Institute of Liver and Biliary Sciences, New Delhi 110070, India; [3]Department of Research, GRIPMER, New Delhi 110060, India; [4]Department of Toxicology and Cancer Biology, University of Kentucky, Lexington, KY 40506, USA; [5]Department of Bioscience and Biotechnology, Banasthali Vidyapith, Rajasthan 304022, India

**Carcinoma of the gall bladder (CAGB) has a poor prognosis. Molecular analysis of bile could classify indicators of CAGB. Bile samples ($n$ = 87; training cohort) were screened for proteomics and metabolomics signatures of cancer detection. In bile, CAGB showed distinct proteomic (217 upregulated, 258 downregulated) and metabolomic phenotypes (111 upregulated, 505 downregulated, $p < 0.05$, fold change > 1.5, false discovery rate <0.01) linked to significantly increased inflammation (coagulation, arachidonic acid, bile acid) and alternate energy pathways (pentose-phosphate pathway, amino acids, lipid metabolism); and decreased glycolysis, cholesterol metabolism, PPAR, RAS, and RAP1 signaling, oxidative phosphorylation, and others compared to gallstone or healthy controls ($p < 0.05$). Bile proteins/metabolites signatures showed significant correlation ($r^2 > 0.5$, $p < 0.05$) with clinical parameters. Metabolite/protein signature-based probability of detection for CAGB (cancer) was >90% ($p < 0.05$), with area under the receiver operating characteristic curve >0.94. Validation of the top four metabolites—toluene, 5,6-DHET, creatine, and phenylacetaldehyde—in separate cohorts ($n$ = 80; bile [T1] and paired plasma [T2]) showed accuracy (99%) and sensitivity/specificity (>98%) for CAGB detection. Tissue validation showed bile 5,6-DHET positively correlated with tissue PCNA (proliferation), and caspase-3 linked to cancer development ($r^2$ >0.5, $p < 0.05$). In conclusion, the bile molecular landscape provides critical molecular understanding and outlines metabolomic indicator panels for early CAGB detection.**

## INTRODUCTION

Carcinoma of the gallbladder (CAGB) has a poor prognosis, often due to slow/vague symptoms, aggressive disease progression, and consequently poor outcomes.[1,2] Early detection and timely surgery can lead to long-term survival.[3] An increase in the incidence rate is closely related to elevated prevalence of gallstones.[4] Therefore, it is critical to provide early diagnosis and to decipher the dysregulated proteins and metabolites (bio-molecules) sub-network, which could provide insight into the pathobiology and help in the classification of putative indicators used for the screening and stratification of CAGB.

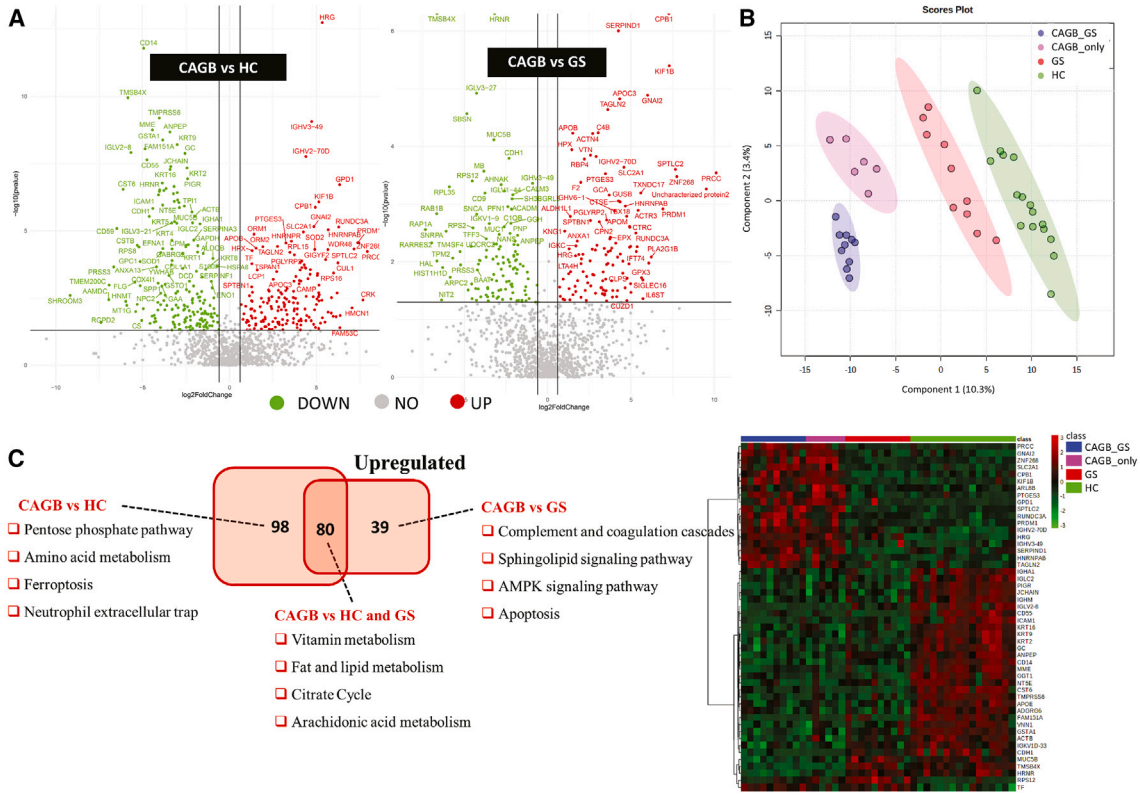
Bile (concentrated in GB) helps in fat digestion, acts as excretory medium for endogenous and exogenous compounds, and is known for the assimilation of proteins and metabolites linked to the development of GB cancer.[5] Change in the biliary proteome or metabolome is concordant with biliary pathology and contributes to the progression of biliary diseases.[6] Many published series showed the utility of bile as a choice of fluid for gallstone (GS) identification or GB-associated disease indicators,[7] although the study of biliary proteins and metabolites in the carcinogenesis of the GB is obscure and there are no defined indicators that could help in early stratification of CAGB.

Literature has demonstrated the Royal Marsden Hospital score as a significant prognostic marker across various cancers, particularly in predicting survival outcomes in patients undergoing immunotherapy.[8] Additionally, DNA damage repair alterations have emerged as predictive biomarkers, especially in biliary tract cancers (BTCs), where they can inform response to therapies such as immune checkpoint inhibitors (ICIs).[9] However, the role of these biomarkers in CAGB remains uncertain.

Over the past decade, immunotherapy has revolutionized the treatment landscape for solid malignancies, offering new hope through therapies targeting immune pathways. ICIs, in particular, have become prominent for advanced biliary cancers, although the clinical response remains modest.[10,11] Research suggests that ICIs can also alleviate symptoms like peripheral neuropathy and headaches in cancer patients but also lead to ICI-related adverse events such as hearing

Received 26 June 2024; accepted 4 November 2024;
https://doi.org/10.1016/j.omton.2024.200904.

**Correspondence:** Jaswinder Singh Maras, Associate professor, Department of Molecular and Cellular Medicine, Institute of Liver and Biliary Sciences, New Delhi 110070, India.
**E-mail:** jassi2param@gmail.com




*(legend on next page)*

loss.[12] Further research is crucial to fully establish the efficacy of ICIs in CAGB or identify novel therapeutic targets. This study aims to bridge that gap and may help in identifying new treatment strategies for CAGB patients.

Bile collection is a challenging task due to the surgical procedure involved. However, bile often spills from enterohepatic circulation to systemic circulation, suggesting that plasma samples could also be used as a substitute for CAGB detection and warrants validation.[13]

In the present study, we investigated the bile samples from CAGB and non-cancer (GS or healthy control [HC]) groups. Proteomics and metabolomics analyses outlined systemic change in the bile linked with CAGB development. A panel of the top four potential protein and/or metabolite indicators of CAGB was identified and subjected to validation in bile (test cohort: T1) and paired plasma (test cohort: T2) samples of CAGB ($n$ = 40) and GS ($n$ = 40) patients using five different machine learning (ML) models. Our result demonstrates the efficacy of the selected panel of metabolites, and the optimal ML model was used for reliable and infallible detection of CAGB.

## RESULTS

### Demographic profile of the study cohort

The study consisted of 167 patients who were divided into the training cohort ($n$ = 87 patients) and the test cohort (T1 = bile, and T2 cohort = plasma; $n$ = 80 patients; Figure S1). Clinical profiles of the training and test cohorts were similar and showed bilirubin, aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, and γ-glutamyl transferase levels increased whereas total bile acids, triglycerides, and high-density lipoprotein levels were decreased in CAGB patients (Figure S2; Table S1).

### Bile proteome profile could stratify patients with CAGB

Bile functionality relies on the lipids, metabolites, and proteins present in it, and analysis of bile proteome could provide critical insight into the development of CAGB. In the training cohort, bile proteomic analysis identified 366 (178 upregulated and 188 downregulated) and 235 (119 upregulated and 116 downregulated) differentially expressed proteins (DEPs) in CAGB compared to GS and HC (Figure 1A; Table S2; fold change [FC] > 1.5, $p$ < 0.05, false discovery rate [FDR] < 0.01). Partial least-squares discriminant analysis (PLS-DA) and hierarchical clustering analysis segregated CAGB from GS and HC (Figures 1B and S3–S5). A total of 137 proteins (unique to CAGB; Figure 1C) were linked to increase in pentose phosphate pathway, amino acid metabolism, ferroptosis, and immunological pathways (e.g., complement and neutrophil extracellular trap [NET]), sphingolipid signaling, AMPK pathway, apoptosis and others (Figure 1C; Tables S3 and S4; $p$ < 0.05). Furthermore, 80 proteins (common to CAGB vs. HC or GS) were associated with vitamin, fat, lipid, arachidonic acid metabolism, and citrate cycle (Figure 1C; Table S5; $p$ < 0.05). Similarly, proteins significantly downregulated in bile samples of CAGB (Figure 1D) showed a decrease in glycolysis/gluconeogenesis, cholesterol metabolism, platelet activation and ATP-binding cassette (ABC) transporters, glutathione metabolism, pyruvate metabolism, oxidative phosphorylation, and others (Figure 1D; Tables S6–S8; $p$ < 0.05). The protein panel (PRCC, PRDM1, RUNDC3A, and CPB1) showed significantly low mean decrease in accuracy (Figure 1E) and were the most important proteins (FC, area under the receiver operating characteristic curve [AUC], and $p$ value) with a probability of cancer detection (POD) >85% (AUC = 0.99, 95%-CI = 0.973–1) in CAGB patients (Figure 1F). Overall, CAGB showed distinct protein profiles in which the majority of proteins were downregulated and the identified protein panel could be used for stratification of CAGB patients.

### Bile proteome could stratify patients with GS among CAGB patients

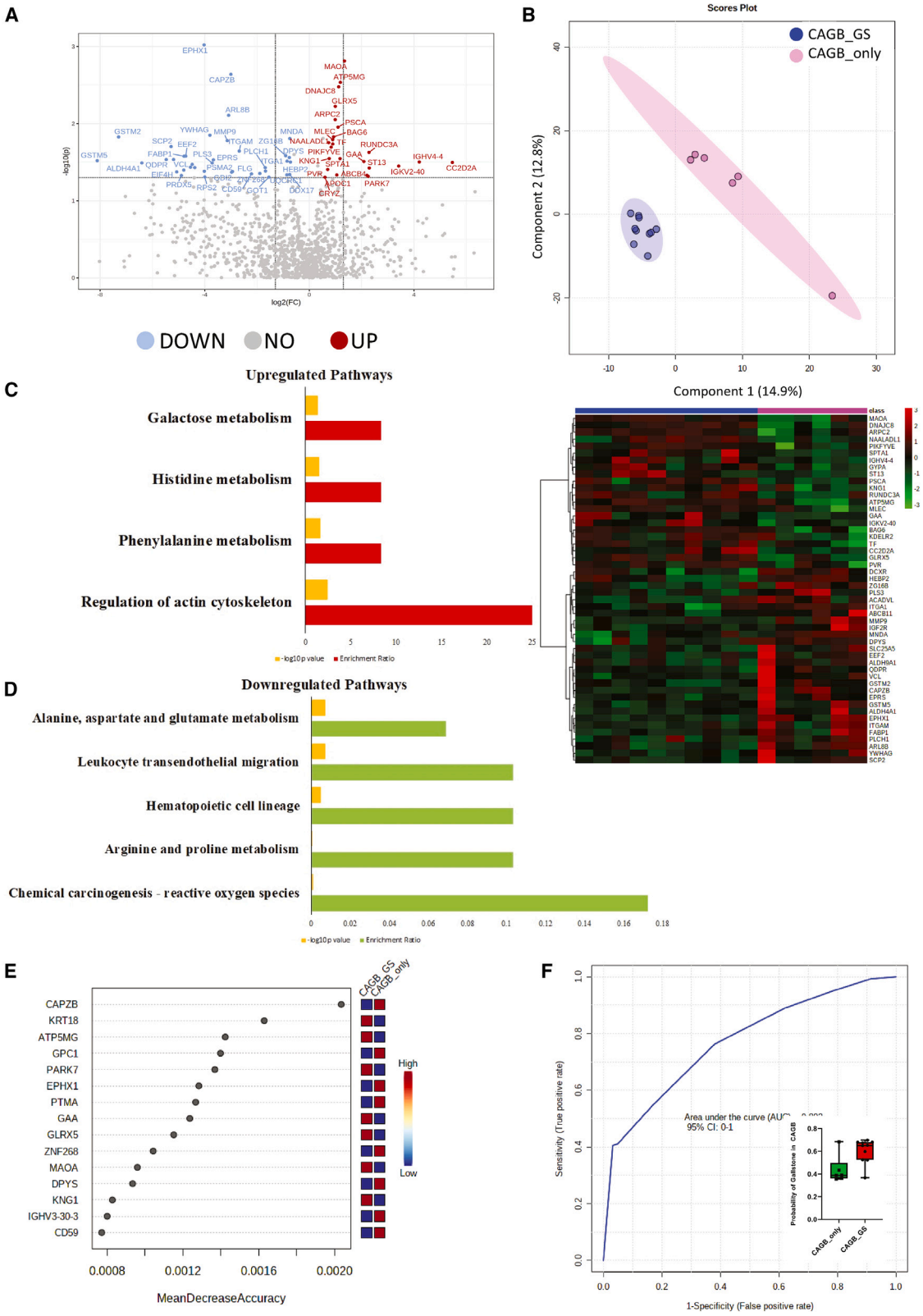
Next, we compared the proteome profiling of CAGB patients with GS to those without GS (CAGB only). Results showed 24 upregulated and 38 downregulated proteins in CAGB with GS (Figure 2A; Table S2; FC > 1.5, $p$ < 0.05). PLS-DA, along with hierarchical clustering, segregated the study groups (Figure 2B). Proteins significantly upregulated in CAGB with GS were linked to regulation of actin cytoskeleton, phenylalanine, and galactose metabolism (Figure 2C), and proteins downregulated were linked to arginine, aspartate, and glutamate metabolism and to leukocyte migration (Figure 2D). The protein panel of PARK7, ST13, GSTM2, and PAICS showed the lowest mean decrease in accuracy (Figure 2E) and were the most important proteins (FC, AUC, and $p$ value), with a POD of gallstone in CAGB >70% (AUC = 0.80, confidence interval [95%-CI] = 0.73–1) in CAGB patients (Figure 2F). Overall, our results suggest that CAGB with GS is associated with an increase in phenylalanine and galactose metabolism. Furthermore, the identified proteomic indicator panel could be used for stratification of CAGB patients with GS.

### Bile proteome profile identifies signatures associated with GS

Untargeted proteomic analysis identified 216 DEPs (87 upregulated and 129 downregulated) in GS compared to HC (Figure S6;

**Figure 1. Bile proteome profile among CAGB, GS, and HC**

(A) Volcano plot showing differentially expressed proteins (DEPs) in the bile of CAGB patients when compared to GS or HC. The red dots represent upregulated, while the green dot represents downregulated proteins (FC > 1.5, $p$ < 0.05 considered significant). (B) Partial least-squares discriminant analysis (PLS-DA) score plot and hierarchical clustering show clear segregation of CAGB patients compared to GS or HC. In PLS-DA, the percentage of explained variability for component 1 and component 2 is 10.3% and 3.4%, respectively. (C) Venn diagram showing enrichment of upregulated pathways from CAGB patients when compared to GS or HC. (D) Venn diagram showing enrichment of downregulated pathways from CAGB patients when compared to GS or HC. (E) Random forest (RF) score plot representing the mean decrease in accuracy (MDA) plot of 15 different proteins and their expression status (red = upregulated and blue = downregulated). (F) AUC value for prediction of CAGB on the training set cohort is shown based on top four protein panel. AUC = 0.99. Estimation of probability of cancer detection (POD) with >85% is also shown for these proteins (CAGB is indicated as cancer, while GS or HC is indicated as non-cancer).




*(legend on next page)*

Table S2; FC > 1.5, $p < 0.05$). PLS-DA and heatmap analysis showed clear segregation of GS (Figures S7 and S8). The mean decrease in accuracy analysis identified GS-specific protein signatures (Figure S9). Among the signatures, the panel of four proteins showed a diagnostic efficacy of 98% (AUC = 0.98, 95%-CI = 0.926–1) and POD >80% for gallstone detection (Figure S10). We found that the GS proteomic profile is distinct and that the identified proteomic signature panel could be used for stratification of patients with GS, which warrants further validation.

### Blood transcription module space analysis of DEPs

Next, DEPs from cancer and GS were searched on the blood transcription module (BTM) space. More than 20 BTM modules were enriched in cancer and GS. The majority of modules was downregulated in cancer and GS (Figures S11 and S12). Venn analysis showed that platelet activation, blood coagulation, monocyte activation, and the KLF12 targets network specifically increased and associated with cancer, whereas antigen presentation, cell movement, and adhesion pathways were specific to GS (Figure S13).

### Bile proteomic profiling of patients with cholecystitis is distinct

In our study, seven patients had cholestasis in addition to CAGB. A comparative analysis of patients with and without cholestasis revealed 83 DEPs (68 upregulated and 15 downregulated; FC > 1.5, $p < 0.05$; Figure S14). The expression profile segregated CAGB patients with cholestasis in PLS-DA and hierarchical clustering (Figures S15 and S16). Proteins significantly upregulated in CAGB patients with cholestasis were linked to glycolysis, chemical carcinogenesis-like DNA adducts, and reactive oxygen species (ROS) pathways, and those significantly downregulated in CAGB with cholestasis were linked to selenocompound metabolism, pantothenate biosynthesis, and β-alanine metabolism pathways, respectively (Figures S17 and S18). Furthermore, random forest (RF) analysis identified significant protein signatures associated with cholestasis (Figure S19). The identified signature could be explored for their diagnostic capability for the prediction of cholestasis in CAGB.

### Bile metabolome profile could stratify patients with CAGB

Bile metabolomics analysis of CAGB patients, when compared to GS or HC, showed 585 (102 upregulated and 483 downregulated) and 466 metabolites (45 upregulated and 421 downregulated) metabolites differentially expressed (Figure 3A; Table S9; FC > 1.5, $p < 0.05$, FDR < 0.01). PLS-DA and hierarchical clustering analysis segregated CAGB from other groups (Figure 3B). Metabolites significantly increased in CAGB were associated with primary and secondary bile acid, ABC transporters, pentose-phosphate pathway, arachidonic acid aspartate and glutathione, and tryptophan metabolism (Figure 3C; Tables S10–S12; $p < 0.05$). Metabolites downregulated in CAGB were associated with purine, folate biosynthesis, histidine, tyrosine, nicotinate and nicotinamide, riboflavin metabolism, and others (Figure 3D; Tables S13–S15; $p < 0.05$). The mean decrease in accuracy was lowest for metabolite panels C00601, C00300, C14772, and C01455 (Figure 3E) and were the most significant metabolites (FC, AUC, and $p$ value), with a POD of >70% (AUC = 0.94, 95%-CI = 0.88–0.98) (Figure 3F). Overall, these results suggest that CAGB has a distinct profile, and the identified metabolite panel could be used for the stratification of patients with CAGB.

### Bile metabolome could stratify patients with GS among CAGB patients

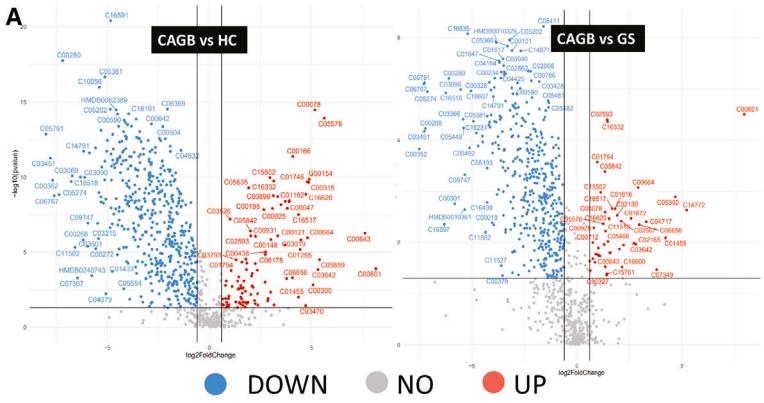
The bile metabolomic profile of CAGB with GS when compared to CAGB without GS outlined that the metabolomic profile of CAGB with GS was distinct and showed 12 metabolites upregulated and 20 metabolites downregulated (Figure 4A; Table S9; FC > 1.5, $p < 0.05$). PLS-DA showed clear segregation of CAGB with GS and CAGB only (Figure 4B). Metabolites significantly increased in CAGB with GS were linked to mitochondrial fatty acid β-oxidation and steroidogenesis (Figure 4C), and those downregulated were linked to biotin, alanine, and glutamate metabolism (Figure 4D). The mean decrease in accuracy was lowest (Figure 4E) for the metabolite panel C16565, C06188, C14874, and C04167 and was the most significant metabolites (FC, AUC, malondialdehyde [MDA], and $p$ value) with a POD for GS in CAGB of >70% (AUC = 0.84, 95%-CI = 0.81–0.95) (Figure 4F). Overall, these results suggest that metabolome profiling of CAGB with GS is distinct, and the identified metabolite panel could be used for the stratification of patients with GS in CAGB; this observation warrants further validation.

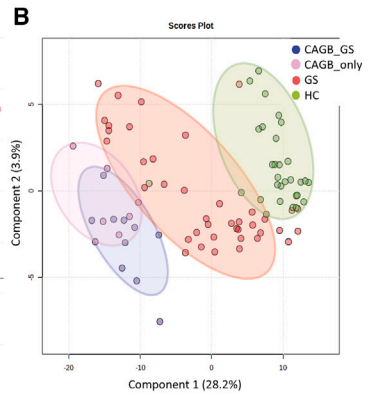
### Bile metabolome profile identifies signatures linked with GS

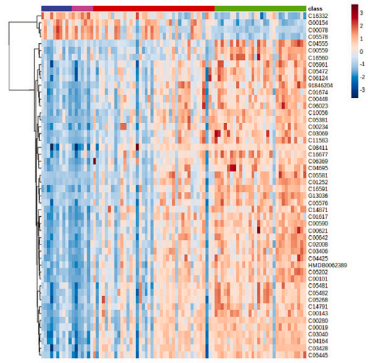
Bile metabolomics analysis of GS identified 358 metabolites (95 upregulated and 263 downregulated; Figure S20; Table S9; FC > 1.5, $p < 0.05$). PLS-DA and a heatmap showed clear segregation of GS and HC (Figures S21 and S22). MDA identified metabolite panels C00315, C01746, G00154, and C00077 as the most significant (FC, AUC, MDA, and $p$ value) and showed a POD of GS of >85% (Figures S23 and S24; AUC = 0.90, 95%-CI = 0.83–0.99). Altogether, patients with GS represent a distinct metabolome profile, and the

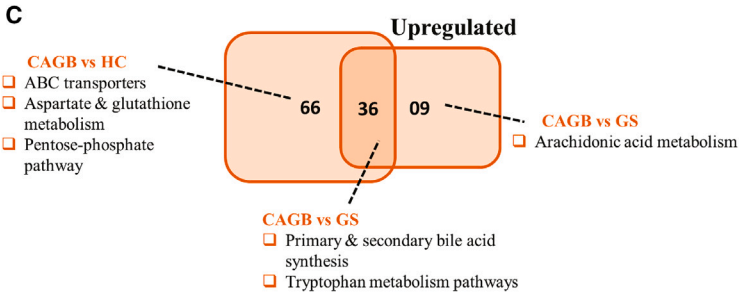
**Figure 2. Comparative proteome of CAGB with GS and CAGB without GS**

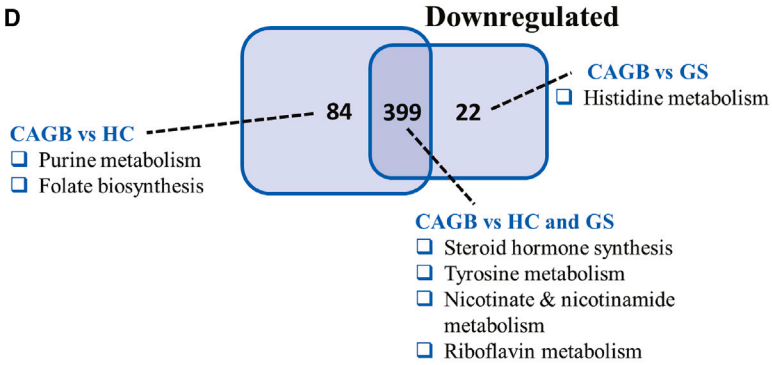
(A) Volcano plot of proteins in CAGB with GS when compared to CAGB without GS. The maroon dots represent upregulated, while the blue dots represent downregulated proteins (FC > 1.5, $p < 0.05$ considered significant). (B) PLS-DA score plot and hierarchical clustering show segregation of CAGB_GS (CAGB with GS) and CAGB_only (CAGB without GS) based on proteome profile. The percentage of explained variability for component 1 and component 2 is 14.9% and 12.8%, respectively. (C) Bar plot shows pathway enriched with upregulated DEPs in CAGB with GS ($p$ value is represented as −log10 $p$ value; red bars show enrichment ratio, while yellow bars show $p$ value). (D) Bar plot showing pathway enriched with downregulated DEPs in CAGB with GS ($p$ value is represented as −log10 $p$ value; green bars show enrichment ratio, while yellow bars show $p$ value). (E) RF score plot representing the MDA plot of 15 different proteins and their expression status (red = upregulated and blue = downregulated). (F) AUC value for prediction of CAGB with GS is shown based on top four protein panel. AUC = 0.80. Estimation of probability of CAGB with GS detection with >70% is also shown for proteins.

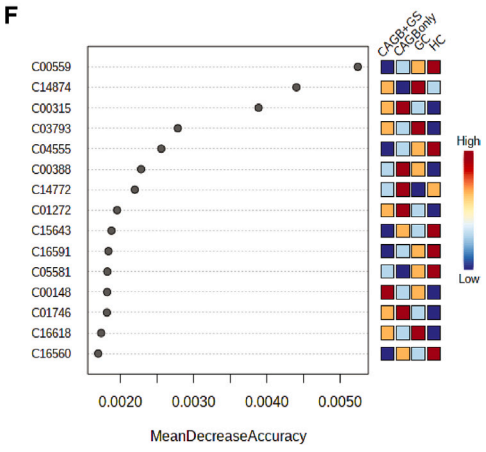


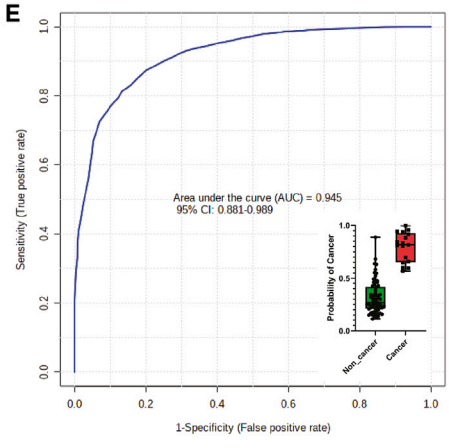






*(legend on next page)*

identified metabolites panel represents putative indicators for the stratification of patients with predisposition to GS.

### Bile metabolomics profiling of patients with cholecystitis is distinct

Analysis of patients with and without cholestasis revealed 57 differentially expressed metabolites (DEMs) (24 upregulated and 33 downregulated; Figures S25, S6, and S27; FC > 1.5, $p < 0.05$). Upregulated metabolites in CAGB patients with cholestasis were linked to arachidonic acid and folate metabolism pathways and those downregulated were linked to steroidogenesis and bile acid biosynthesis pathways, respectively (Figures S28 and S29). Random forest analysis identified significant protein signatures associated with cholestasis (Figure S30). These signatures could be explored for their ability to be putative indicators of cholestasis in CAGB patients.

### Integrome analysis of bile proteome and metabolome functionality

Bile proteomic and metabolomic integration analysis identified four unique positively co-related clusters (Figure 5A). The mean cluster intensity was higher in clusters 1 and 2 and lower in clusters 3 and 4 for cancer development (Figure 5B). Cluster-wise pathway analysis showed that in cluster 1, the increase in phenylacetaldehyde is associated with membrane trafficking and carbohydrate metabolism (Figure 5C). Proteins of cluster 2 were mainly linked to metabolism of lipids and lipoproteins that established positive correlation with metabolites associated with ABC transporters, tryptophan, and glutathione metabolism (Figure 5D). Next, cluster 3 (reduced in cancer) revealed that deregulation of CD14 and TMSB-4 (Thymosin Beta 4 X-Linked) proteins correlate with several metabolites of steroid and folate metabolism, others (Figure 5E). Furthermore, cluster 4 (reduced in cancer) is associated with a decrease in inosinic acid and 1-amino-2-propanol and several proteins that belong to complement cascade, platelet activation and signaling, glycosylation, and others (Figure 5F). As a result, the analysis of integration revealed a remarkable correlation between the functionality of proteins and metabolites in the development of CAGB.

### Metabolomic and proteomic signatures as putative indicators for CAGB stratification and validation of the metabolite panel for CAGB detection using ML

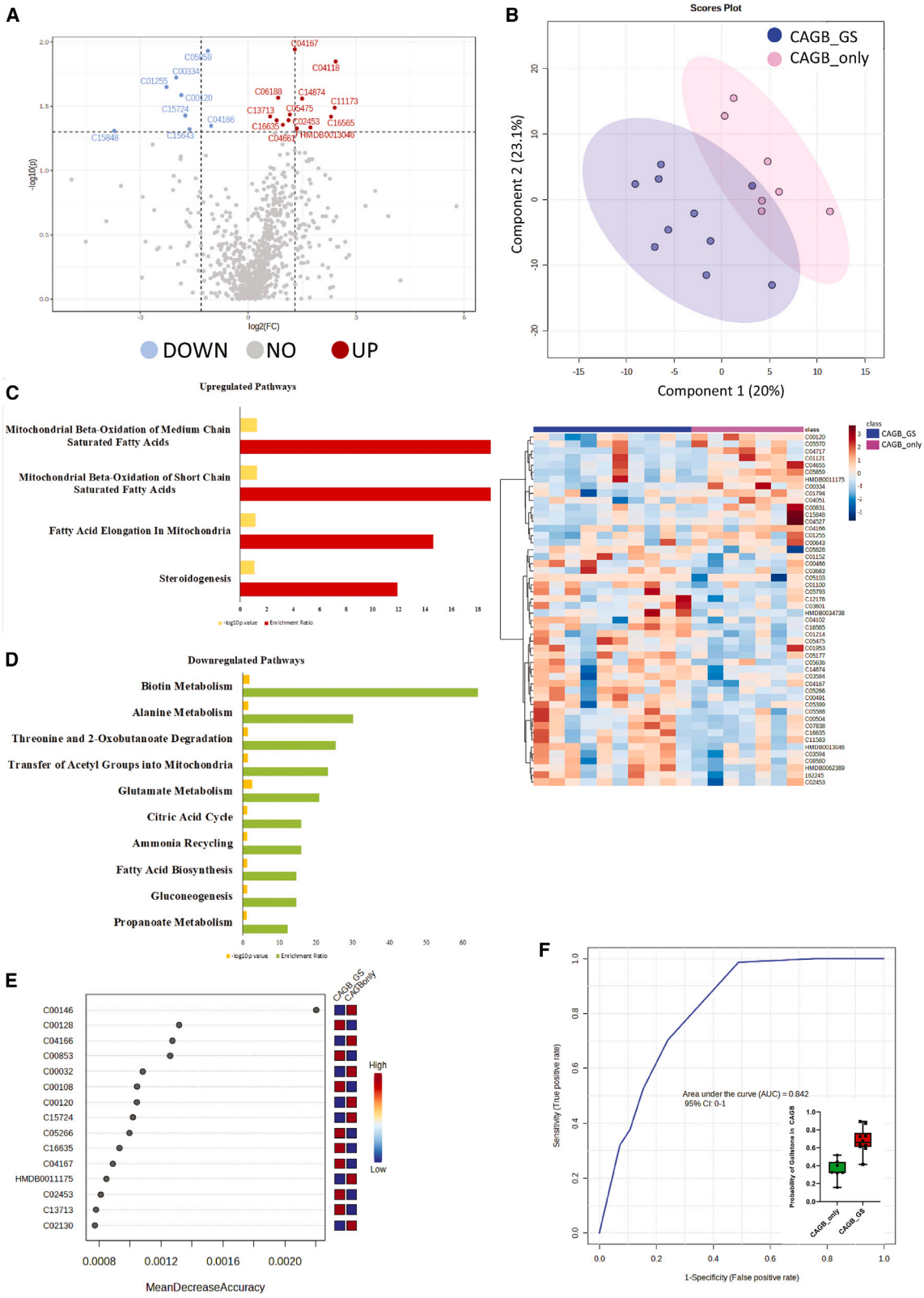
The proteomic and metabolic analyses led to the identification of putative signatures that were selected based on their AUC, FC, $p$ value, and POD (Figure 6A). Cross-correlation analysis (individual metabolites/proteins, POD panels, clinical profile, bile acid profile) showed significant and direct correlation of POD metabolites and POD proteins with the clinical and bile acid profiles in CAGB patients (Figure S31). When compared, the POD of metabolites showed higher diagnostic efficacy, with an AUC of 0.99, in comparison to the POD of proteins for determination of CAGB (Figure S32). Furthermore, the POD of metabolites correlated directly with the American Joint Committee on Cancer (AJCC) grade of differentiation in CAGB patients ($8^{th}$ edition; Figure 6B). The metabolite panel selected for validation showed significant FC, AUC, and relative abundance in the training cohort, test cohort 1 (bile; T1), and test cohort 2 (plasma; T2) (Figure 6C). The workflow using five different ML algorithms (linear discriminant analysis [LDA], K-nearest neighbor [KNN], support vector machine [SVM], classification and regression trees [CART], and RF) resulted in 20 trained and tested ML models. In the ML analysis, both test cohorts (T1 and T2) showed significant accuracy and kappa values for all four metabolites (Figure 6D). The metabolite panel altogether showed the highest accuracy, sensitivity, specificity, and significant $p$ value in both bile and plasma cohorts when compared to individual metabolites. RF represented the best ML model for CAGB detection in bile as well as plasma samples (Figures 6E and 6F). We also compared the results of metabolites panel with our previous study showing that the lipidomics panel[7] could segregate CAGB. A close similarity was observed in results in terms of accuracy, sensitivity, specificity, and $p$ value (Figure S33). Overall, our results suggest that the panel of metabolites selected for POD can be used for early demarcation and stratification of CAGB patients. Additionally, the four-metabolite panel could be added to the previous lipid panel or could be individually used for early detection of CAGB patients.

### Tissue level validation of the identified indicator of CAGB

Next, probable mechanisms of CAGB development and their association with the identified metabolite panel were highlighted (Figure 7A). To validate the link between 5,6-dihydroxyeicosatrienoic acid (5,6-DHET) and cancer, all arachidonic acid metabolites were identified, and it showed 5,6-DHET and leukotriene B4 significantly upregulated in CAGB vs. GS (Figure 7B). The literature suggests that epoxyeicosatrienoic acids (EETs) help in cancer development via accumulation of caspase-3.[14] Concordantly, the expression of proliferating cell nuclear antigen (PCNA) and caspase-3 was significantly increased in CAGB (Figures 7C and 7D). Moreover, PCNA and caspase-3 showed a direct correlation with 5,6-DHET in CAGB patients (Figure 7E). Overall, our results suggest that the increase in bile 5,6-DHET level corroborates the increase in PCNA (proliferation)

**Figure 3. Bile metabolome profile among CAGB, GS, and HC**

(A) Volcano plot showing differentially expressed metabolites (DEMs) in bile of CAGB patients compared to GS or HC. The orange dots represent upregulated, while the blue dots represent downregulated proteins (FC > 1.5, $p < 0.05$ considered significant). (B) PLS-DA score plot and hierarchical clustering show clear segregation of CAGB patients when compared to GS or HC. In PLS-DA, the percentage of explained variability for component 1 and component 2 is 10.3% and 3.4%, respectively. (C) Venn diagram shows enrichment of upregulated pathways from CAGB patients when compared to GS or HC. (D) Venn diagram shows enrichment of downregulated pathways from CAGB patients when compared to GS or HC. (E) RF score plot representing the MDA plot of 15 different metabolites and their expression status (red = upregulated and blue = downregulated). (F) AUC value for prediction of CAGB on the training set cohort is shown based on top four metabolome panel. AUC = 0.94. Estimation of POD with >85% is also shown for these metabolites (CAGB is indicated as cancer, while GS or HC is indicated as non-cancer).




*(legend on next page)*

and accumulation of caspase-3 (anti-apoptosis), which further supports cancer development.

## DISCUSSION

Bile has multifaceted metabolic roles and is composed of proteins and metabolites that are secreted or shed by the hepatobiliary system.[15] Bile is a complex mixture, and variations in individual bile composition complicate biomarker identification. Researchers are tackling this by refining analytical methods and improving cross-validation approaches.[16] Although many published studies outline proteome and metabolome analysis in hepatobiliary diseases,[17,18] detailed evaluation of the biliary proteome and metabolome and their association with CAGB is obscure. Furthermore, putative indicators for early prediction of gall bladder carcinoma should be delineated.

This study investigated bile protein and metabolite profiles in patients with CAGB, GS, and HC. The primary objective of the study was to identify putative indicators (proteins and metabolites) that have diagnostic capability for CAGB detection. We further aimed to elucidate the interplay between the bile proteome and the metabolic network to understand how proteome-metabolome interactions contribute to CAGB development. Finally, we validated the identified metabolites using a ML approach in both bile and plasma samples from a paired cohort. This validation aimed to confirm the potential of these metabolites as a panel of biomarkers for CAGB detection. Bile analysis offers a promising way to identify the molecular changes associated with developing GS or cancer.[16] Patients recruited as training and test cohorts were clinically similar. Patients recruited with CAGB (training or test cohort) showed increased severity and liver function parameters and had variable bile acid profiles.

The bile proteome of CAGB patients showed a significant increase in proteins linked to inflammatory pathways, such as NET formation, ferroptosis, sphingolipid signaling, coagulation cascade activation, arachidonic acid metabolism, alternate energy metabolism, pentose phosphate pathway, and fat and lipid metabolism. Notably, the increase in NETosis-related proteins suggests enhanced neutrophil activation and infiltration[19] and is also reported in BTC and extrahepatic cholangiocarcinoma tissues.[20] The increase in NETosis also promotes cancer-linked thrombosis and tumor growth and is known to protect cancer cells in breast and lung cancer.[21] Our findings also revealed a significant increase in complement protein C5 and proteins associated with arachidonic acid metabolism. An increase in C5 proteins contributes to cancer growth and metastasis by increasing angiogenesis specifically shown in biliary cancer.[22,23] Increases in arachidonic acid metabolism proteins (prostaglandin-E synthase, leukotriene-A4 hydrolase, and others) found in our study are responsible for increasing downstream prostaglandins and leukotrienes, which play a key role in cancer-associated immune mediators and enhance angiogenesis.[24] Further increases in sphingolipid metabolism,[25] ferroptosis,[26] alternate energy activation (e.g., pentose phosphate pathway),[27] and lipid and fat metabolism are classical features of cancer development.[28] Concordantly, BTM analysis showed that platelet activation, blood coagulation, monocyte activation, and KLF12 target network were specifically increased in CAGB. The proteomic analysis of the bile sample in CAGB patients provides us with an insight that tumor development in CAGB is primarily linked to increases in inflammatory, angiogenesis, and tumor expansion-associated proteins that may lead to highly aggressive cancers.

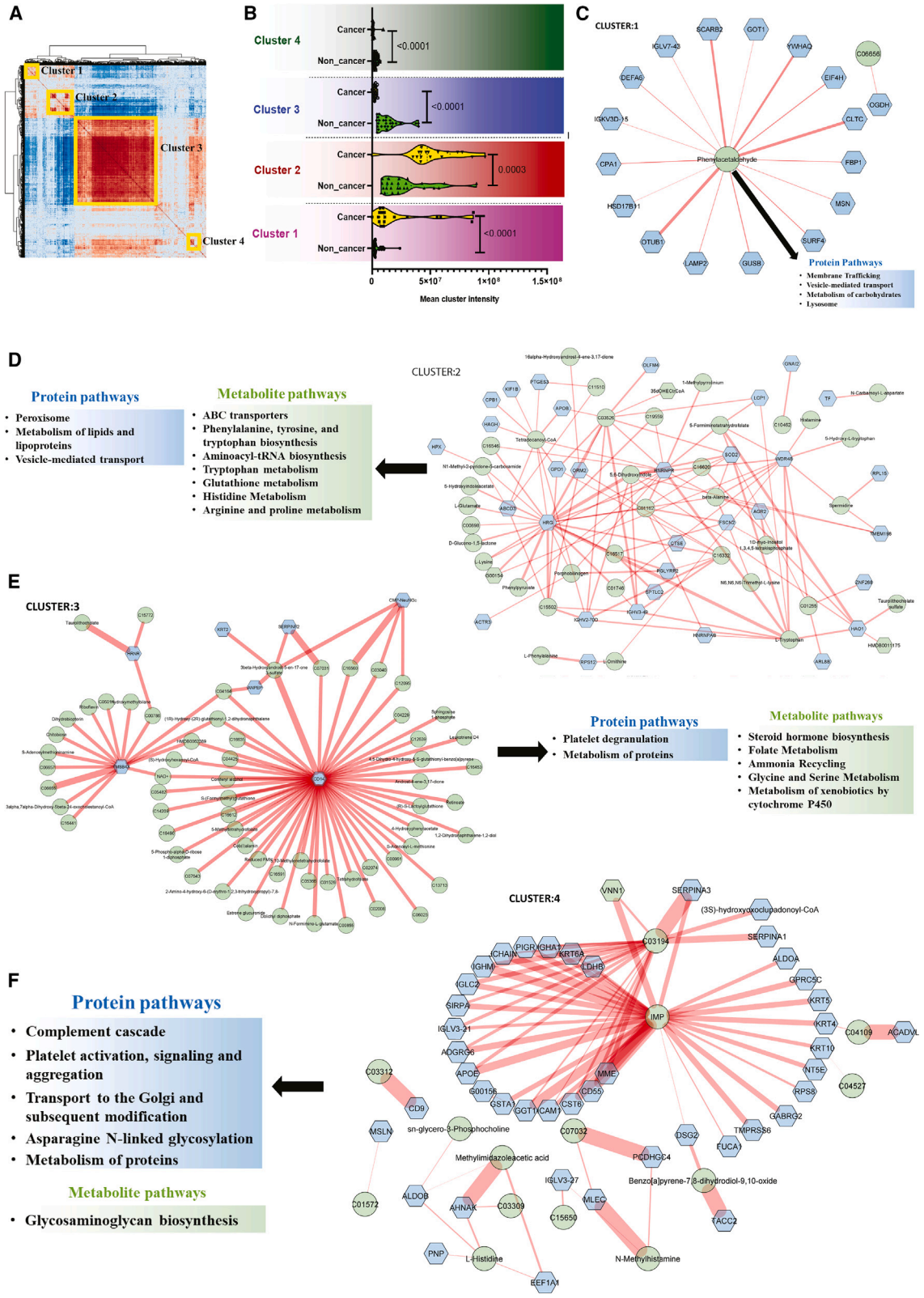
Differential regulations of metabolites are the readout of various metabolic pathways that are altered during the process of cancer development, and thus the identification of such metabolites could prove advantageous in the early diagnosis of cancer. Interestingly, the metabolome profile of CAGB patients was distinct. Bile samples of CAGB patients showed persistent increases in primary and secondary bile acid synthesis and tryptophan metabolism known to be inflammatory.[29,30] Elevated levels of chenodeoxycholic acid or glycochenodeoxycholic acid are known to induce cancer cell invasion and provide stemness to cells via the STAT3 pathway,[31] while the increase in tryptophan metabolism supports cancer progression by suppressing anti-tumor immunity.[32] We also observed a unique increase in arachidonic metabolism, pentose phosphate pathways, and others in CAGB bile samples, suggestive of pertinent inflammation and shift to alternate energy metabolism in CAGB patients.

Bile proteomic analysis of CAGB with and without GS showed significant increases in proteins associated with the regulation of actin cytoskeleton, histidine metabolism, phenylalanine metabolism, and galactose metabolism in CAGB with GS. Increases in PARK7, ST13, and GSTM2 relate to excessive increases in oxidative stress-linked cancer progression due to the presence of GS.[33–35]

Again, bile analysis in CAGB patients with GS identified increased fatty acid metabolism (generation, breakdown, and steroid production) compared to those without GS. Additionally, specific metabolites, including aminopropylcadaverine, salicin 6-phosphate, glutathione episulfonium ion, and *trans*-acenaphthene-1,2-dioland, emerged as

**Figure 4. Comparative metabolome of CAGB with GS and CAGB without GS**

(A) Volcano plot of metabolites in CAGB with GS when compared to CAGB without GS. The maroon dots represent upregulated, while the blue dots represent downregulated metabolites (FC > 1.5, $p < 0.05$ considered significant). (B) PLS-DA score plot and hierarchical clustering show segregation of CAGB_GS (CAGB with GS) and CAGB_only (CAGB without GS) based on metabolite profile. The percentage of explained variability for component 1 and component 2 is 14.9% and 12.8%, respectively. (C) Bar plot shows pathway enriched with upregulated DEMs in CAGB with GS ($p$ value is represented as $-\log10\,p$ value; red bars show enrichment ratio, while yellow bars show $p$ value). (D) Bar plot shows pathway enriched with downregulated DEMs in CAGB with GS ($p$ value is represented as $-\log10\,p$ value; green bars show enrichment ratio, while yellow bars show $p$ value). (E) RF score plot representing the MDA plot of 15 different metabolites and their expression status (red = upregulated and blue = downregulated). (F) AUC value for prediction of CAGB with GS is shown based on top four metabolites panel. AUC = 0.80. Estimation of probability of CAGB with GS detection with >70% is also shown for metabolites.




*(legend on next page)*

potential markers to differentiate CAGB with and without GS. Our results suggest the involvement of dysregulated fat metabolism in the development of GS in CAGB patients.

Interestingly, analysis of bile samples in GS patients showed significant increases in protein RAP1A (Ras-related protein 1), known to promote cancer via mitogen-activated protein kinase and Notch signaling pathway.[36] An increase in small nuclear ribonucleoprotein polypeptide A is known to modulate the nerve growth factor expression in cancer.[37] Glypican 1 is well known to promote the growth and migration of cancer cells through regulating the transforming growth factor-β1/SMAD2 signaling pathway, whereas an increase in Cystatin 6 promotes angiogenesis and cancer development.[38,39] Therefore, the association of signature proteins with cancer signaling pathways dictates the predisposition of GS toward cancer development.

Similar to the proteomic analysis of bile samples of GS as compared to HC the metabolite analysis showed like the proteomic analysis and highlighted significant increases in metabolites, which suggest the dysregulation of fat metabolism, change in β-oxidation status, and inflammation.[40,41] This dysregulation of lipid metabolism seen in the bile sample of GS patients is concordant with the development of GS in such patients.

Interestingly, CAGB patients with cholestasis showed significant increases in the protein expression profile of bile. This may be because of the retention of bile constituents in cholestasis. CAGB patients who had cholestasis showed increases inUDP-glucose pyrophosphorylase 2, which is compensatory to maintain glycogen synthesis and glycoprotein production; however, this increase also aids in cancer production.[42] Increases in Small Heat Shock Protein β-1 and DEAD-box helicase 1 seen in CAGB patients with cholestasis is known to activate intracellular pathways for inflammatory processes linked to the induction of cholestasis.[43,44]

Metabolite analysis of CAGB patients with cholestasis showed significant increases in arachidonic acid and folate metabolism with decreases in steroidogenesis and polyamine levels. Decreased steroidogenesis suggests the accumulation of cholesterol and thus decreases in the metastability of the bile samples in CAGB patients.[45] Our metabolic results are concordant with the proteomic findings and state that bile samples of CAGB with cholestasis are more inflammatory and have dysregulated lipid metabolism, which could be one of the reasons for the observed phenotype.

Next, for integrating both proteomic and metabolomic analyses of CAGB bile, differentially expressed proteome and metabolome in CAGB were subjected to cross-correlation network analysis. We identified four unique positively co-related clusters documenting clear associations among DEPs and metabolites in CAGB patients. Interestingly, phenylacetaldehyde (PAA), found in cluster 1, showed significant correlation with proteins linked with membrane trafficking and carbohydrate metabolism. PAA is known to increase ROS, thereby altering carbohydrate metabolism in cancer cells.[46,47] These findings suggest that PAA and alterations in carbohydrate metabolism may contribute to CAGB development and progression. Cluster 2 was significantly associated with lipoproteins, ABC transporters, and tryptophan metabolism. Increased tryptophan metabolites suggest its association with other cancers, potentially indicating its contribution to CAGB.[48] Cluster 3, which was downregulated, was associated with proteins of immune cells (CD14) or those involved in the organization of the cytoskeleton (TMSB-4), suggesting the dysregulated immune defense and increased chances of metastasis and development of CAGB.[49] In cluster 4, we observed concordant results, with significant downregulation of immune pathways like platelet activation and complement activation, suggesting that dysregulation of immune activation contributes to the development of CAGB.[50] The proteome and metabolome integration highlighted key metabolic disturbances that play a vital role in the development of CAGB.

Next, the clinical correlation showed that metabolites and proteins signatures were directly associated with liver injury parameters and the bile acid profile. Interestingly, serum albumin and cholesterol showed an inverse correlation to these markers. Additionally, positive correlation emerged between POD metabolites and CAGB grades of differentiation, suggesting a potential link between POD and CAGB progression.

The POD proteins, four key protein signatures selected for CAGB detection, include PRCC, known to protect cancer cells from DNA damage, leading to cell survival and cancer progression[51]; PRDM1, a tumor suppressor gene, may have tumor regulatory properties[52]; RUNDC3A, which regulates SNAP25 expression and is linked with cancer progression[53]; and CPB1, which is found to be increased in cancer, but the mechanistic role is uncertain.[54] In addition, POD metabolites, four signature metabolites selected for CAGB detection, include phenylacetaldehyde, known to increase ROS and STAT3-mediated decrease in antitumor activity[46]; 5,6-DHET, an EET that promotes angiogenesis, cell proliferation, and anti-apoptosis activity in the tumor[55]; creatine, which induces SMAD3 signaling for metastasis[56]; and toluene, which is produced by the microbiome, is a carcinogen derived from phenylacetate.[57] Moreover, POD of metabolites exhibited higher AUCs than that of proteins, which was further employed for validation.

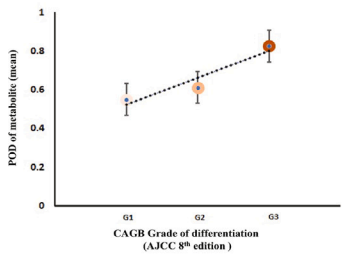
**Figure 5. Integrome analysis of DEPs and DEMs in bile of CAGB**

(A) Correlation plot showing correlation clusters between DEPs and DEMs in bile of CAGB patients (highlighted in yellow). (B) Violin plot shows mean cluster intensity in each cluster for cancer and non-cancer. (C) Expression correlation network of DEPs (in blue background) and DEMs (in green background) in cluster 1 and their associated pathways ($r^2$ = 0.7; edge width represents the degree of correlation). (D) Expression correlation network of DEPs and DEMs in cluster 2 and their associated pathways ($r^2$ = 0.7). (E) Expression correlation network of DEPs and DEMs in cluster 3 and their associated pathways ($r^2$ = 0.7). (F) Expression correlation network of DEPs and DEMs in cluster 4 and their associated pathways ($r^2$ = 0.7).

**A**

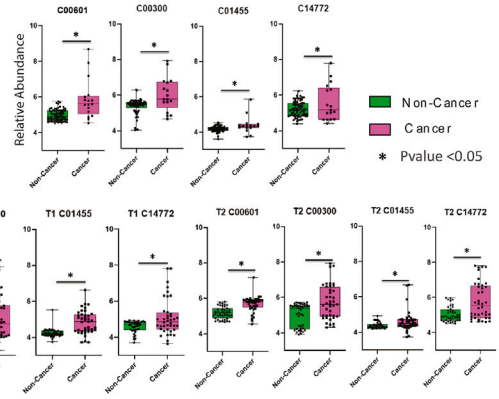
| | Proteins | AUC | Pval | FC | | Metabolite | AUC | Pval | FC |
|---|---|---|---|---|---|---|---|---|---|
| Cancer vs Non-Cancer | PRCC | 0.84 | 0.00 | 8.44 | Cancer vs Non-Cancer | C00601 | 0.92 | 0.00 | 8.04 |
| | PRDM1 | 0.82 | 0.00 | 7.18 | | C00300 | 0.84 | 0.00 | 4.84 |
| | RUNDC3A | 0.84 | 0.00 | 5.82 | | C14772 | 0.74 | 0.00 | 4.33 |
| | CPB1 | 0.97 | 0.00 | 5.47 | | C01455 | 0.91 | 0.00 | 4.28 |
| CAGB_GS vs CAGB_only | PARK7 | 0.85 | 0.05 | 2.25 | CAGB_GS vs CAGB_only | C16565 | 0.80 | 0.03 | 2.04 |
| | ST13 | 0.85 | 0.04 | 2.29 | | C06188 | 0.81 | 0.03 | 1.78 |
| | GSTM2 | 0.87 | 0.01 | -7.29 | | C14874 | 0.81 | 0.03 | 2.83 |
| | PAICS | 0.80 | 0.06 | -5.97 | | C04167 | 0.85 | 0.01 | 2.46 |
| Gallstone vs Healthy | SNRPA | 0.81 | 0.01 | 5.74 | Gallstone vs Healthy | C00315 | 0.88 | 0.00 | 4.97 |
| | RAP1A | 0.86 | 0.00 | 7.28 | | C01746 | 0.86 | 0.00 | 4.22 |
| | GPC1 | 0.82 | 0.00 | -5.47 | | G00154 | 0.87 | 0.00 | 4.20 |
| | CST6 | 0.94 | 0.00 | -8.46 | | C00077 | 0.87 | 0.00 | 4.02 |

**B**

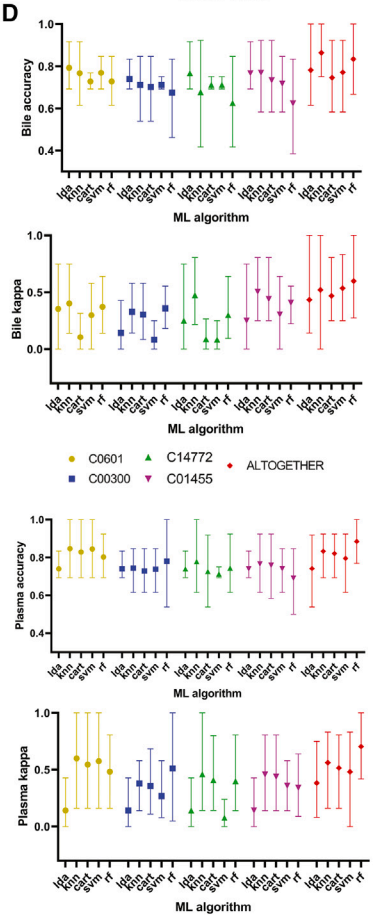



**C**

| Metabolite | Training Cohort | | | Test Cohort 1 (Bile) | | Test Cohort 2 (Plasma) | |
|---|---|---|---|---|---|---|---|
| | AUC | FC | P-value | FC | P-value | FC | P-value |
| C00601 | 0.92 | 8.04 | 0.00 | 1.99 | 0.05 | 3.27 | 0.00 |
| C00300 | 0.84 | 4.84 | 0.00 | 1.58 | 0.03 | 1.86 | 0.00 |
| C14772 | 0.74 | 4.33 | 0.00 | 2.77 | 0.05 | 1.28 | 0.00 |
| C01455 | 0.91 | 4.28 | 0.00 | 2.71 | 0.01 | 1.61 | 0.05 |




**D**




**E**

## Bile

**Accuracy**

| | lda | knn | cart | svm | rf |
|---|---|---|---|---|---|
| C00601 | 0.76 | 0.80 | 0.71 | 0.80 | 0.98 |
| C00300 | 0.72 | 0.75 | 0.76 | 0.71 | 0.98 |
| C14772 | 0.72 | 0.79 | 0.71 | 0.71 | 1.00 |
| C01455 | 0.72 | 0.80 | 0.81 | 0.87 | 0.97 |
| Altogether | 0.77 | 0.90 | 0.81 | 0.85 | 1.00 |

**p-value**

| | lda | knn | cart | svm | rf |
|---|---|---|---|---|---|
| C00601 | 0.10 | 0.01 | 0.54 | 0.02 | 0.00 |
| C00300 | 0.39 | 0.19 | 0.10 | 0.54 | 0.00 |
| C14772 | 0.47 | 0.03 | 0.54 | 0.54 | 0.00 |
| C01455 | 0.47 | 0.02 | 0.01 | 0.00 | 0.00 |
| Altogether | 0.07 | 0.00 | 0.01 | 0.00 | 0.00 |

**Sensitivity**

| | lda | knn | cart | svm | rf |
|---|---|---|---|---|---|
| C00601 | 0.19 | 0.35 | 0.00 | 0.30 | 0.95 |
| C00300 | 0.05 | 0.38 | 0.27 | 0.00 | 1.00 |
| C14772 | 0.03 | 0.38 | 0.00 | 0.00 | 1.00 |
| C01455 | 0.03 | 0.54 | 0.51 | 0.62 | 0.89 |
| Altogether | 0.32 | 0.70 | 0.51 | 0.73 | 1.00 |

**Specificity**

| | lda | knn | cart | svm | rf |
|---|---|---|---|---|---|
| C00601 | 1.00 | 0.99 | 1.00 | 1.00 | 0.99 |
| C00300 | 1.00 | 0.90 | 0.97 | 1.00 | 0.98 |
| C14772 | 1.00 | 0.96 | 1.00 | 1.00 | 1.00 |
| C01455 | 1.00 | 0.90 | 0.93 | 0.97 | 1.00 |
| Altogether | 0.96 | 0.98 | 0.93 | 0.90 | 1.00 |

**F**

## Plasma

**Accuracy**

| | lda | knn | cart | svm | rf |
|---|---|---|---|---|---|
| C00601 | 0.72 | 0.87 | 0.88 | 0.88 | 0.99 |
| C00300 | 0.72 | 0.80 | 0.83 | 0.75 | 1.00 |
| C14772 | 0.72 | 0.82 | 0.80 | 0.71 | 0.98 |
| C01455 | 0.72 | 0.83 | 0.85 | 0.84 | 0.99 |
| Altogether | 0.77 | 0.87 | 0.87 | 0.87 | 1.00 |

**p-value**

| | lda | knn | cart | svm | rf |
|---|---|---|---|---|---|
| C00601 | 0.47 | 0.00 | 0.00 | 0.00 | 0.00 |
| C00300 | 0.39 | 0.01 | 0.00 | 0.19 | 0.00 |
| C14772 | 0.47 | 0.00 | 0.02 | 0.54 | 0.00 |
| C01455 | 0.47 | 0.00 | 0.00 | 0.00 | 0.00 |
| Altogether | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 |

**Sensitivity**

| | lda | knn | cart | svm | rf |
|---|---|---|---|---|---|
| C00601 | 0.03 | 0.68 | 0.76 | 0.68 | 0.97 |
| C00300 | 0.05 | 0.62 | 0.73 | 0.24 | 1.00 |
| C14772 | 0.03 | 0.59 | 0.76 | 0.00 | 0.97 |
| C01455 | 0.03 | 0.65 | 0.68 | 0.62 | 0.89 |
| Altogether | 0.38 | 0.73 | 0.57 | 0.78 | 1.00 |

**Specificity**

| | lda | knn | cart | svm | rf |
|---|---|---|---|---|---|
| C00601 | 1.00 | 0.94 | 0.93 | 0.97 | 1.00 |
| C00300 | 1.00 | 0.88 | 0.87 | 0.96 | 1.00 |
| C14772 | 1.00 | 0.91 | 0.81 | 1.00 | 0.99 |
| C01455 | 1.00 | 0.90 | 0.92 | 0.93 | 0.99 |
| Altogether | 0.93 | 0.92 | 0.99 | 0.91 | 1.00 |

*(legend on next page)*

Our previous research demonstrated the potential of using plasma samples as surrogates for bile in stratification CAGB. The present study strengthens this finding by showing that bile metabolite variations are also validated in plasma samples. This is likely because bile components can leak from the gut into the bloodstream, as reported in other studies. Therefore, plasma analysis holds promise as a non-invasive approach for CAGB detection.[7]

Finally, we identified a panel of four key metabolites using *p* value, FC, area under the receiver operating characteristic curve, RF, and POD calculation, thereby validating them in the bile cohort T1 and plasma cohort T2 using an ML approach. The combined panel showed the highest accuracy for CAGB detection, and the LDA model serves as the preferred ML algorithm for CAGB detection in both bile and plasma samples. Furthermore, we propose literature-driven mechanisms linked to cancer development for each metabolite, and GB tissue analysis supported our findings. The elevated 5,6-DHET level in bile and its association with cancer was validated by correlating it with caspase-3 accumulation, which helps in the development of cancer by increasing PCNA (proliferation marker).[14]

Interestingly, the downregulation of four proposed CAGB biomarkers (4,5-DHET, creatine, toluene, phenylacetaldehyde) in patients of CAGB with GS suggests that these markers not only provide insights about cancer development but significant increases in such markers could also be taken as a risk factor for the development of cancer specifically in patients who do not have an evident GS. In addition, regardless of the presence or absence of GS, a significant increase in the plasma level of these metabolites indicates the development of CAGB.

Many US Food and Drug Administration-approved mass spectrometry (MS)-based *in vitro* diagnostic methods for the identification of microbes, newborn screening, and quantification of therapeutic drugs are now part of routine diagnostics.[58] Our method identified a panel of metabolites within 30 min, suggesting a rapid and potentially transformative approach for patient screening. Further validation in a larger CAGB cohort is warranted.

While bile collection limits large-scale studies due to its invasive nature, it could not be considered for large population studies, which is the major lacuna of the study. Furthermore, apart from cancer, the bile components may also come from the hepatocytes, biliary tree, or GB epithelium. This might limit the interpretation of results. Also, this study did not thoroughly explore the specific biological functions of these biomarkers in CAGB, leaving this aspect for future investigations. However, our robust findings and validation using ML demonstrate the potential of the identified metabolite panel that accurately segregates patients with CAGB.

In conclusion, we present a pilot study showcasing the bile proteome and the bile metabolome in patients with CAGB. We employed a robust integrative analysis of proteomic and metabolomic investigation of bile samples in patients with CAGB. Our validation of metabolites in the bile and paired plasma cohorts (ML approach) documented the highest accuracy, sensitivity, and specificity for CAGB detection. Thus, the identified metabolite panel is proposed as a putative candidate for CAGB detection. Moreover, the present study holds significant potential for improving early diagnosis and treatment. Validation of these biomarkers in plasma offers an even more accessible and less invasive diagnostic approach, potentially transforming clinical practice. In the coming years, we can expect advancements in the development of plasma-based biomarker panels for clinical use. Along with improvements in analytical validation and AI-driven data analysis, this could revolutionize the early detection and management of GB diseases. All this could also aid in investigating the role of these biomarkers in CAGB.

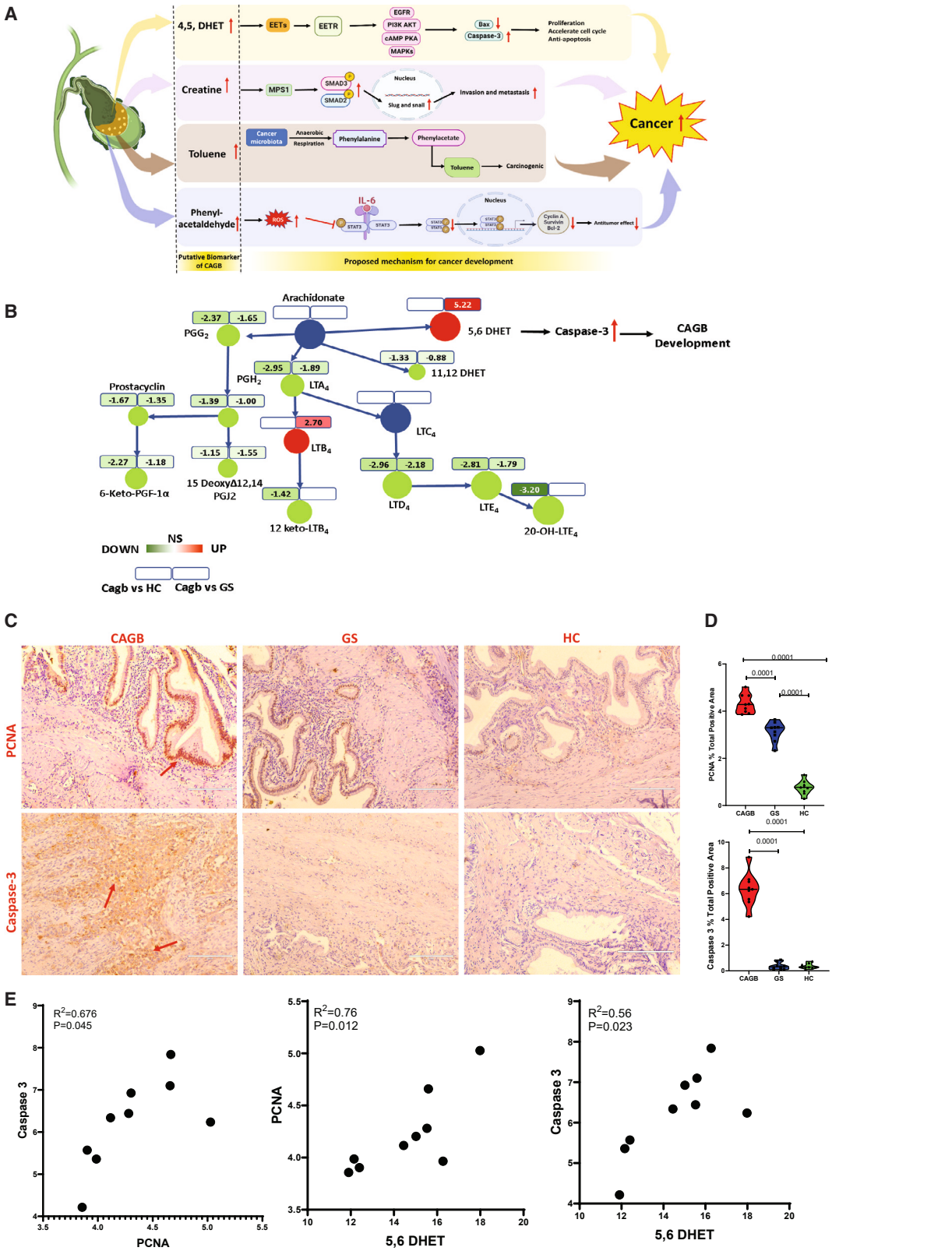
## MATERIALS AND METHODS

### Patient selection

A total of 167 participants were recruited at the Institute of Liver and Biliary Sciences and Ganga Ram Institute for Postgraduate Medical Education & Research, New Delhi, India, from 2018 to 2021 for this cross-sectional study. Samples were divided into training and test cohorts. In the training cohort (samples recruited between 2018 and 2020), bile and plasma samples were collected from patients undergoing cholecystectomy for GS disease (GS, $n = 40$) and patients undergoing surgery for CAGB were grouped as CAGB ($n = 17$); we also collected HC samples (HC, $n = 30$). The patients enrolled as healthy in our study were live liver donors for living-donor liver transplant surgery. During the surgical intervention, ~5 mL bile was also collected intraoperatively by aspirating from the GB. In the test cohort, baseline plasma and bile of 80 participants (samples recruited between 2020 and 2021) were also collected (Figure S1). Samples were stored at −80°C until analyzed. All procedures involved in the study were conducted per the institutional ethical committee, and written informed consent was obtained from all subjects enrolled in this study.

**Figure 6. Metabolomic and proteomic signatures as putative indicators for CAGB stratification and validation of metabolite panel for CAGB detection using machine learning**

(A) Panel of proteins and metabolites selected based on their AUC, p value, and FC for POD calculation in training cohort. (B) POD of metabolite panel (represented in mean) showing positive correlation with CAGB grade of differentiation (AJCC 8th edition). (C) Panel of four metabolites selected from POD along with their AUC, FC, and relative abundance in training and test cohorts (bile and their paired plasma). (D) Accuracy and kappa value of the different ML models used for characterization of principal biomarker associated with CAGB as compared to HC or GS Patients in test cohort T1 (in bile) and T2 (in plasma). (E) Accuracy, specificity, sensitivity, and *p* value of four different metabolites individually and combined in T1 for different ML models. (F) Accuracy, specificity, sensitivity, and *p* value of four different metabolites individually and combined in T2 for different ML models.




*(legend on next page)*

### Bile and plasma metabolomics

We used the organic phase extraction method for metabolite extraction for bile and plasma. In brief, 400 μL cold methanol (100% v/v) was added to 100 μL bile, and plasma samples were kept overnight at −20°C for protein precipitation, followed by 10-min centrifugation at 13,000 rpm. Then, the supernatant was transferred, vacuum dried, and reconstituted in (5:95:5) 5% acetonitrile: 95% water: 5% internal standards + external standards at known concentrations. The samples were run onto reverse-phase chromatography using C18 column (Thermo Fisher Scientific [Waltham, MA] 25003102130: 3 μm, 2.1 mm, 100 mm) using an ultra-high performance liquid chromatographic system followed by high-resolution MS (Figure S1). Compound Discoverer 3.0 was used for metabolite features identification (Thermo Fisher Scientific). Feature annotation was performed using mzVault, mzCloud (www.mzcloud.org), and mass list searches. Each metabolite was annotated with its Kyoto Encyclopedia of Genes and Genomes or Human Metabolome Database ID. After filtration, the data were saved in .csv file format and subjected to statistical analysis.[59]

### Bile proteomics

A total of 100 μL bile was diluted with 400 μL 0.1% formic acid and sonicated. These samples were centrifuged at 13,000 rpm for 10 min, and the collected supernatant was albumin depleted in Akta Pure. Equivalent proteins, 100 μg, were subjected to reduction with DTT, alkylation with indole-3-acetic acid, and digestion for 24 h at 37°C using trypsin (sequencing-grade modified). Post-desalting, the samples were dried and reconstituted in 0.1% formic acid. A total of 3 μL of each sample was subjected to nano-electrospray ionization and MS/MS using Q-Exactive Plus (Thermo Fisher Scientific) at the collision-induced dissociation mode, with the electrospray voltage at 2.3 kV. Further data were analyzed using Proteome Discoverer (version 2.0, Thermo Fisher Scientific). Uniprot *Homo sapiens* (human) database (UP000005640) with Mascot algorithm (Mascot 2.4, Matrix Science, London, UK) was used. Identified proteins were subjected to conventional statistical analysis and network and pathway analysis.

### Statistical analysis

Statistical analysis was performed using GraphPad Prism version 6 and SPSS version 20; $p < 0.05$ using Benjamini-Hochberg correction were considered significant. Results are shown as mean and SD unless indicated otherwise. Unpaired (two-tailed) Student's t test and the Mann-Whitney *U* test were performed for two-groups comparison. For comparison among more than two groups, one-way ANOVA and the Kruskal-Wallis test were performed. All correlations were performed using Spearman correlation analysis, and $r^2 > 0.5$ and $p < 0.05$ were considered significant. For proteomics and metabolomics, Perseus and Metaboanalyst 5.0 were used; data were log normalized and subjected to Pareto scaling and further subjected to multivariate projection analyses including principal-component analysis, PLS-DA, RFanalysis. Pathway enrichment patterns were identified for metabolites and proteins individually using Metabolanalyst and EnrichR.[60] Significant proteins were enriched on 346 BTMs to identify the different pathway modules.[61] Pathway integration analysis was done using both Impala and Cytoscape 3.9.1.[62] R version 4.0.1 was used to make the correlation plot and validation using ML algorithm.

The ML approach was employed using five different MLs (LDA, RF, SVM, CART, and KNN) algorithms to validate the plausible biomarker lipids species. In total, we implemented 25 ML models comprising 5 ML algorithms along with 5 metabolites individually and combined. Fourfold (outer) nested repeated (five times) 10-fold (inner) cross-validation (with randomized stratified splitting) was done on training and test cohorts in R with the Caret package. In this way, repeated 10-fold cross-validation was performed 20 times, and the models were obtained with the best results. In addition, the accuracy, sensitivity, and specificity performance measures summarized the overall cross-validation prediction performance. The equations used to quantify these performance measures are presented below (in which TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives):

$$\text{Accuracy} = TP + TNTP + TN + FP + FN,$$

$$\text{Sensitivity} = TPTP + FN,$$

$$\text{Specificity} = TNTN + FP.$$

In the present study, POD for cancer development, POD for presence of GS in cancer patients, and POD for GS development were calculated using the top four proteins or metabolites identified based on p value, AUC, and mean decrease in accuracy in each comparison (cancer vs. no cancer, GS in cancer vs. no GS, and GS vs. healthy) using SPSS version 20.

## DATA AND CODE AVAILABILITY

The raw data for proteomics and metabolomics performed in the present study are available from the corresponding author upon reasonable request. All analyses done in this study are included in the published article and its supplemental information files.

## ACKNOWLEDGMENTS

The work was supported by the Indian Council of Medical Research (5/4/8-3/CD/JS/2021-NCD-II, project ID: 2020–4958).

**Figure 7. Tissue-level validation showing association of 5,6-DHET with CAGB**

(A) The proposed mechanism of CAGB development and its association with putative indicators. (B) Arachidonic acid metabolites identified in the study (red = upregulated, green = downregulated, white = non-significant). (C) Immunohistochemistry (IHC) shows expression of PCNA and caspase-3 in CAGB, GS, and HC ($n = 9$). (D) IHC analyses. Relative quantization of positively stained cells is expressed as mean number of positive cells/5 high-power field (20×) and $p < 0.0001$. (E) Expression correlation PCNA and caspase-3, PCNA, and 5,6-DHET, and caspase-3 and 5,6-DHET in CAGB ($n = 9$, $r^2 > 0.5$, $p < 0.05$).

## AUTHOR CONTRIBUTIONS

J.S.M. conceptualized the work. Samples were provided by V.P., R.R., S.F., and D.B.. N.S., S.H.B., and B.M. were responsible for sample processing and experimental work and were helped by V.B., M.Y., G.T., S.Y., N.S., S.P., N.K.S., and H.H. Data analysis was performed by N.S. under the guidance of J.S.M. The manuscript was drafted by N.S. and J.S.M. This manuscript has been seen and approved by all authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omton.2024.200904.

## REFERENCES


1. Goetze, T.O. (2015). Gallbladder carcinoma: Prognostic factors and therapeutic options. World J. Gastroenterol. *21*, 12211–12217. https://doi.org/10.3748/wjg.v21.i43.12211.
2. Hundal, R., and Shaffer, E.A. (2014). Gallbladder cancer: epidemiology and outcome. Clin. Epidemiol. *6*, 99–109. https://doi.org/10.2147/CLEP.S37357.
3. Roa, J.C., García, P., Kapoor, V.K., Maithel, S.K., Javle, M., and Koshiol, J. (2022). Gallbladder cancer. Nat. Rev. Dis. Primers *8*, 69. https://doi.org/10.1038/s41572-022-00398-y.
4. Rawla, P., Sunkara, T., Thandra, K.C., and Barsouk, A. (2019). Epidemiology of gallbladder cancer. Clin. Exp. Hepatol. *5*, 93–102. https://doi.org/10.5114/ceh.2019.85166.
5. Reshetnyak, V.I. (2013). Physiological and molecular biochemical mechanisms of bile formation. World J. Gastroenterol. *19*, 7341–7360. https://doi.org/10.3748/wjg.v19.i42.7341.
6. Holm, M., Joenväärä, S., Saraswat, M., Tohmola, T., Saarela, T., Tenca, A., Arola, J., Renkonen, R., and Färkkilä, M. (2022). Quantitative bile and serum proteomics for the screening and differential diagnosis of primary sclerosing cholangitis. PLoS One *17*, e0272810. https://doi.org/10.1371/journal.pone.0272810.
7. Sharma, N., Yadav, M., Tripathi, G., Mathew, B., Bindal, V., Falari, S., Pamecha, V., and Maras, J.S. (2022). Bile multi-omics analysis classifies lipid species and microbial peptides predictive of carcinoma of gallbladder. Hepatology *76*, 920–935. https://doi.org/10.1002/hep.32496.
8. Sahin, T.K., Rizzo, A., Aksoy, S., and Guven, D.C. (2024). Prognostic Significance of the Royal Marsden Hospital (RMH) Score in Patients with Cancer: A Systematic Review and Meta-Analysis. Cancers *16*, 1835. https://doi.org/10.3390/cancers16101835.
9. Ricci, A.D., Rizzo, A., and Brandi, G. (2020). The DNA damage repair (DDR) pathway in biliary tract cancer (BTC): a new Pandora's box? ESMO Open *5*, e001042. https://doi.org/10.1136/esmoopen-2020-001042.
10. Uson Junior, P.L.S., and Araujo, R.L. (2022). Immunotherapy in biliary tract cancers: Current evidence and future perspectives. World J. Gastrointest. Oncol. *14*, 1446–1455. https://doi.org/10.4251/wjgo.v14.i8.1446.
11. Ricci, A.D., Rizzo, A., and Brandi, G. (2020). Immunotherapy in Biliary Tract Cancer: Worthy of a Second Look. Cancer Control *27*, 1073274820948047. https://doi.org/10.1177/1073274820948047.
12. Guven, D.C., Erul, E., Kaygusuz, Y., Akagunduz, B., Kilickap, S., De Luca, R., and Rizzo, A. (2023). Immune checkpoint inhibitor-related hearing loss: a systematic review and analysis of individual patient data. Support. Care Cancer *31*, 624. https://doi.org/10.1007/s00520-023-08083-w.
13. Di Ciaula, A., Baj, J., Garruti, G., Celano, G., De Angelis, M., Wang, H.H., Di Palo, D.M., Bonfrate, L., Wang, D.Q.-H., and Portincasa, P. (2020). Liver Steatosis, Gut-Liver Axis, Microbiome and Environmental Factors. A Never-Ending Bidirectional Cross-Talk. J. Clin. Med. *9*, 2648.
14. Chen, C., Wei, X., Rao, X., Wu, J., Yang, S., Chen, F., Ma, D., Zhou, J., Dackor, R.T., Zeldin, D.C., and Wang, D.W. (2011). Cytochrome P450 2J2 is highly expressed in hematologic malignant diseases and promotes tumor cell growth. J. Pharmacol. Exp. Ther. *336*, 344–355. https://doi.org/10.1124/jpet.110.174805.
15. Boyer, J.L. (2013). Bile formation and secretion. Compr. Physiol. *3*, 1035–1078. https://doi.org/10.1002/cphy.c120027.
16. Li, Y.C., Li, K.S., Liu, Z.L., Tang, Y.C., Hu, X.Q., Li, X.Y., Shi, A.D., Zhao, L.M., Shu, L.Z., Lian, S., et al. (2022). Research progress of bile biomarkers and their immunoregulatory role in biliary tract cancers. Front. Immunol. *13*, 1049812. https://doi.org/10.3389/fimmu.2022.1049812.
17. Zhao, C., Wang, M., Huang, J., Jia, Z., Zhao, X., Li, E., Wei, Z., Dong, Y., Liu, W., Han, T., et al. (2020). Integrative analysis of proteomic and metabonomics data for identification of pathways related to Rhizoma Paridis-induced hepatotoxicity. Sci. Rep. *10*, 6540. https://doi.org/10.1038/s41598-020-63632-1.
18. Urman, J.M., Herranz, J.M., Uriarte, I., Rullán, M., Oyón, D., González, B., Fernandez-Urién, I., Carrascosa, J., Bolado, F., Zabalza, L., et al. (2020). Pilot Multi-Omic Analysis of Human Bile from Benign and Malignant Biliary Strictures: A Machine-Learning Approach. Cancers *12*, 1644.
19. Vorobjeva, N.V., and Chernyak, B.V. (2020). NETosis: Molecular Mechanisms, Role in Physiology and Pathology. Biochemistry. *85*, 1178–1190. https://doi.org/10.1134/S0006297920100065.
20. Branchi, V., Jurgensen, B., Esser, L., Gonzalez-Carmona, M., Weismuller, T.J., Strassburg, C.P., Henn, J., Semaan, A., Lingohr, P., Manekeller, S., et al. (2021). Tumor Infiltrating Neutrophils Are Frequently Found in Adenocarcinomas of the Biliary Tract and Their Precursor Lesions with Possible Impact on Prognosis. J Pers Med *11*, 233. https://doi.org/10.3390/jpm11030233.
21. Masucci, M.T., Minopoli, M., Del Vecchio, S., and Carriero, M.V. (2020). The Emerging Role of Neutrophil Extracellular Traps (NETs) in Tumor Progression and Metastasis. Front. Immunol. *11*, 1749. https://doi.org/10.3389/fimmu.2020.01749.
22. Afshar-Kharghan, V. (2017). The role of the complement system in cancer. J. Clin. Invest. *127*, 780–789. https://doi.org/10.1172/JCI90962.
23. Son, K.H., Ahn, C.B., Kim, H.J., and Kim, J.S. (2020). Quantitative proteomic analysis of bile in extrahepatic cholangiocarcinoma patients. J. Cancer *11*, 4073–4080. https://doi.org/10.7150/jca.40964.
24. Tian, W., Jiang, X., Kim, D., Guan, T., Nicolls, M.R., and Rockson, S.G. (2020). Leukotrienes in Tumor-Associated Inflammation. Front. Pharmacol. *11*, 1289. https://doi.org/10.3389/fphar.2020.01289.
25. Ogretmen, B. (2018). Sphingolipid metabolism in cancer signalling and therapy. Nat. Rev. Cancer *18*, 33–50. https://doi.org/10.1038/nrc.2017.96.
26. Lu, B., Chen, X.B., Ying, M.D., He, Q.J., Cao, J., and Yang, B. (2017). The Role of Ferroptosis in Cancer Development and Treatment Response. Front. Pharmacol. *8*, 992. https://doi.org/10.3389/fphar.2017.00992.
27. Patra, K.C., and Hay, N. (2014). The pentose phosphate pathway and cancer. Trends Biochem. Sci. *39*, 347–354. https://doi.org/10.1016/j.tibs.2014.06.005.
28. Bian, X., Liu, R., Meng, Y., Xing, D., Xu, D., and Lu, Z. (2021). Lipid metabolism and cancer. J. Exp. Med. *218*, e20201606. https://doi.org/10.1084/jem.20201606.
29. Li, T., and Apte, U. (2015). Bile Acid Metabolism and Signaling in Cholestasis, Inflammation, and Cancer. Adv. Pharmacol. *74*, 263–302. https://doi.org/10.1016/bs.apha.2015.04.003.
30. Platten, M., Nollen, E.A.A., Röhrig, U.F., Fallarino, F., and Opitz, C.A. (2019). Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. Nat. Rev. Drug Discov. *18*, 379–401. https://doi.org/10.1038/s41573-019-0016-5.
31. Shi, C., Yang, J., Hu, L., Liao, B., Qiao, L., Shen, W., Xie, F., and Zhu, G. (2020). Glycochenodeoxycholic acid induces stemness and chemoresistance via the STAT3 signaling pathway in hepatocellular carcinoma cells. Aging (Albany NY) *12*, 15546–15555. https://doi.org/10.18632/aging.103751.
32. Peyraud, F., Guegan, J.P., Bodet, D., Cousin, S., Bessede, A., and Italiano, A. (2022). Targeting Tryptophan Catabolism in Cancer Immunotherapy Era: Challenges and Perspectives. Front. Immunol. *13*, 807271. https://doi.org/10.3389/fimmu.2022.807271.
33. Liang, Z.L., Kim, M., Huang, S.M., Lee, H.J., and Kim, J.M. (2013). Expression of carboxyl terminus of Hsp70-interacting protein (CHIP) indicates poor prognosis in human gallbladder carcinoma. Oncol. Lett. *5*, 813–818. https://doi.org/10.3892/ol.2013.1138.

34. Jin, W. (2020). Novel Insights into PARK7 (DJ-1), a Potential Anti-Cancer Therapeutic Target, and Implications for Cancer Progression. J. Clin. Med. *9*, 1256. https://doi.org/10.3390/jcm9051256.

35. Guo, E., Wei, H., Liao, X., Wu, L., and Zeng, X. (2020). Clinical significance and biological mechanisms of glutathione S-transferase mu gene family in colon adenocarcinoma. BMC Med. Genet. *21*, 130. https://doi.org/10.1186/s12881-020-01066-2.

36. Lu, L., Wang, J., Wu, Y., Wan, P., and Yang, G. (2016). Rap1A promotes ovarian cancer metastasis via activation of ERK/p38 and notch signaling. Cancer Med. *5*, 3544–3554. https://doi.org/10.1002/cam4.946.

37. Dou, N., Yang, D., Yu, S., Wu, B., Gao, Y., and Li, Y. (2018). SNRPA enhances tumour cell growth in gastric cancer through modulating NGF expression. Cell Prolif. *51*, e12484. https://doi.org/10.1111/cpr.12484.

38. Lu, F., Chen, S., Shi, W., Su, X., Wu, H., and Liu, M. (2022). GPC1 promotes the growth and migration of colorectal cancer cells through regulating the TGF-beta1/SMAD2 signaling pathway. PLoS One *17*, e0269094. https://doi.org/10.1371/journal.pone.0269094.

39. Lalmanach, G., Kasabova-Arjomand, M., Lecaille, F., and Saidi, A. (2021). Cystatin M/E (Cystatin 6): A Janus-Faced Cysteine Protease Inhibitor with Both Tumor-Suppressing and Tumor-Promoting Functions. Cancers *13*, 1877. https://doi.org/10.3390/cancers13081877.

40. Xiang, Y., Kong, X., Zhang, C., He, C., Cai, J., Lu, R., Zhang, B., Lu, L., and Yang, Y. (2021). Free fatty acids and triglyceride change in the gallbladder bile of gallstone patients with pancreaticobiliary reflux. Lipids Health Dis. *20*, 97. https://doi.org/10.1186/s12944-021-01527-4.

41. Maurer, K.J., Carey, M.C., and Fox, J.G. (2009). Roles of infection, inflammation, and the immune system in cholesterol gallstone formation. Gastroenterology *136*, 425–440. https://doi.org/10.1053/j.gastro.2008.12.031.

42. Wolfe, A.L., Zhou, Q., Toska, E., Galeas, J., Ku, A.A., Koche, R.P., Bandyopadhyay, S., Scaltriti, M., Lebrilla, C.B., McCormick, F., and Kim, S.E. (2021). UDP-glucose pyrophosphorylase 2, a regulator of glycogen synthesis and glycosylation, is critical for pancreatic cancer growth. Proc. Natl. Acad. Sci. USA *118*, e2103592118. https://doi.org/10.1073/pnas.2103592118.

43. Oh, A., Jeon, S., Jeong, M.G., Kim, H.K., Kang, J., Lee, Y.S., and Hwang, E.S. (2021). HSPB1 inhibitor J2 attenuates lung inflammation through direct modulation of Ym1 production and paracrine signaling. Biomed. Pharmacother. *143*, 112225. https://doi.org/10.1016/j.biopha.2021.112225.

44. Ishaq, M., Ma, L., Wu, X., Mu, Y., Pan, J., Hu, J., Hu, T., Fu, Q., and Guo, D. (2009). The DEAD-box RNA helicase DDX1 interacts with RelA and enhances nuclear factor kappaB-mediated transcription. J. Cell. Biochem. *106*, 296–305. https://doi.org/10.1002/jcb.22004.

45. Chen, Y., Kong, J., and Wu, S. (2015). Cholesterol gallstone disease: focusing on the role of gallbladder. Lab. Invest. *95*, 124–131. https://doi.org/10.1038/labinvest.2014.140.

46. Choi, H.S., Kim, S.L., Kim, J.H., Ko, Y.C., and Lee, D.S. (2020). Plant Volatile, Phenylacetaldehyde Targets Breast Cancer Stem Cell by Induction of ROS and Regulation of Stat3 Signal. Antioxidants *9*, 1119. https://doi.org/10.3390/antiox9111119.

47. Liou, G.Y., and Storz, P. (2010). Reactive oxygen species in cancer. Free Radic. Res. *44*, 479–496. https://doi.org/10.3109/10715761003667554.

48. Li, C., and Zhao, H. (2021). Tryptophan and Its Metabolites in Lung Cancer: Basic Functions and Clinical Significance. Front. Oncol. *11*, 707277. https://doi.org/10.3389/fonc.2021.707277.

49. Janssen, L.M.E., Ramsay, E.E., Logsdon, C.D., and Overwijk, W.W. (2017). The immune system in cancer metastasis: friend or foe? J. Immunother. Cancer *5*, 79. https://doi.org/10.1186/s40425-017-0283-9.

50. Schlesinger, M. (2018). Role of platelets and platelet receptors in cancer metastasis. J. Hematol. Oncol. *11*, 125. https://doi.org/10.1186/s13045-018-0669-2.

51. Liu, C., Lin, X., Sun, B., Mao, Z., Chen, L., Qian, H., and Su, C. (2021). PRCC reduces the sensitivity of cancer cells to DNA damage by inhibiting JNK and ATM/ATR pathways and results in a poor prognosis in hepatocellular carcinoma. Cell Biosci. *11*, 185. https://doi.org/10.1186/s13578-021-00699-x.

52. Casamassimi, A., Rienzo, M., Di Zazzo, E., Sorrentino, A., Fiore, D., Proto, M.C., Moncharmont, B., Gazzerro, P., Bifulco, M., and Abbondanza, C. (2020). Multifaceted Role of PRDM Proteins in Human. Cancer *21*, 2648.

53. Lin, Z., and Kwok, H.F. (2022). RUNDC3A/SNAP25/Akt signaling mediates tumor progression and chemoresistance in gastric neuroendocrine carcinoma. Cell Death Dis. *13*, 840. https://doi.org/10.1038/s41419-022-05294-7.

54. Kothari, C., Clemenceau, A., Ouellette, G., Ennour-Idrissi, K., Michaud, A., Diorio, C., and Durocher, F. (2021). Is Carboxypeptidase B1 a Prognostic Marker for Ductal Carcinoma In Situ? Cancers *13*, 1726. https://doi.org/10.3390/cancers13071726.

55. Wang, B., Wu, L., Chen, J., Dong, L., Chen, C., Wen, Z., Hu, J., Fleming, I., and Wang, D.W. (2021). Metabolism pathways of arachidonic acids: mechanisms and potential therapeutic targets. Signal Transduct. Target. Ther. *6*, 94. https://doi.org/10.1038/s41392-020-00443-w.

56. Zhang, L., Zhu, Z., Yan, H., Wang, W., Wu, Z., Zhang, F., Zhang, Q., Shi, G., Du, J., Cai, H., et al. (2021). Creatine promotes cancer metastasis through activation of Smad2/3. Cell Metab. *33*, 1111–1123.e4. https://doi.org/10.1016/j.cmet.2021.03.009.

57. Zhao, H., Chen, J., Li, X., Sun, Q., Qin, P., and Wang, Q. (2019). Compositional and functional features of the female premenopausal and postmenopausal gut microbiota. FEBS Lett. *593*, 2655–2664. https://doi.org/10.1002/1873-3468.13527.

58. Seger, C., and Salzmann, L. (2020). After another decade: LC-MS/MS became routine in clinical diagnostics. Clin. Biochem. *82*, 2–11. https://doi.org/10.1016/j.clinbiochem.2020.03.004.

59. Sharma, N., Bhat, S.H., Tripathi, G., Yadav, M., Mathew, B., Bindal, V., Sharma, S., Gupta, E., Maras, J.S., and Sarin, S.K. (2022). Global metabolome profiling of COVID-19 respiratory specimen using high-resolution mass spectrometry (HRMS). STAR Protoc. *3*, 101051. https://doi.org/10.1016/j.xpro.2021.101051.

60. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97. https://doi.org/10.1093/nar/gkw377.

61. Li, S., Rouphael, N., Duraisingham, S., Romero-Steiner, S., Presnell, S., Davis, C., Schmidt, D.S., Johnson, S.E., Milton, A., Rajam, G., et al. (2014). Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. Nat. Immunol. *15*, 195–204. https://doi.org/10.1038/ni.2789.

62. Kamburov, A., Cavill, R., Ebbels, T.M.D., Herwig, R., and Keun, H.C. (2011). Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. Bioinformatics *27*, 2917–2918. https://doi.org/10.1093/bioinformatics/btr499.