



OPEN

# Identification of serum metabolite biomarkers and metabolic reprogramming mechanisms to predict recurrence in cholangiocarcinoma

Piya Prajumwongs<sup>1</sup>, Attapol Titapun<sup>1,2</sup>, Vasin Thanasukarn<sup>1,2</sup>, Apiwat Jareanrat<sup>1,2</sup>, Natcha Khuntikeo<sup>1,2</sup>, Nisana Namwat<sup>1,3</sup>, Poramate Klanrit<sup>1,3</sup>, Arporn Wangwiwatsin<sup>1,3</sup>, Jarin Chindaprasirt<sup>1,4</sup>, Supinda Koonmee<sup>1,5</sup>, Prakasit Sa-Ngiamwibool<sup>1,5</sup>, Nattha Muangritdech<sup>1</sup>, Sittiruk Roytrakul<sup>6</sup> & Watcharin Loilome<sup>1,3,7</sup>✉

Cholangiocarcinoma (CCA) has high recurrence rates that severely limit long-term survival. Effective tools for accurate recurrence monitoring and diagnosis remain lacking. Metabolic reprogramming, a key driver of CCA growth and recurrence, is underutilized in cancer screening and management. This study aimed to identify metabolite-based biomarkers to evaluate recurrence severity, enhance disease management, and elucidate the molecular mechanisms underlying CCA recurrence. A comprehensive, non-targeted serum metabolomics analysis using ultra-high-performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry was conducted. Support Vector Machine (SVM) modeling was employed to develop a predictive framework based on metabolite biomarkers. The analysis revealed significant alterations in metabolomics and lipidomics across CCA recurrence subtypes. Notably, changes in metabolites such as amino acids, lipid-derived carnitines, and glycerophospholipids were associated with cancer progression through enhanced energy production and lipid remodeling. The SVM-constructed metabolite-based predictive model demonstrated predictive accuracy comparable to current clinical diagnostic standards. These findings provide novel insights into the metabolic mechanisms underlying CCA recurrence, addressing critical clinical challenges. By advancing early diagnostic approaches, particularly for preoperative detection, this study offers a reliable method for predicting recurrence in CCA patients. This enables effective treatment planning and supports the development of personalized therapeutic strategies, ultimately improving patient outcomes.

**Keywords** Cholangiocarcinoma, Recurrence, Metabolite biomarkers, Support vector machine, Metabolic reprogramming

Cholangiocarcinoma (CCA) is an aggressive biliary tract malignancy, often diagnosed at advanced stages due to its asymptomatic early phase. Surgical resection followed by adjuvant chemotherapy is the primary treatment, but survival outcomes remain poor due to high recurrence rates even after complete resection<sup>1</sup>. Recurrence, whether early or late, is a critical determinant of prognosis, with early recurrence linked to aggressive tumor traits such as poor differentiation, high malignancy grade, and lymphovascular invasion. Whereas, late recurrence is associated with slow tumor growth or micrometastases<sup>2-5</sup>. Therefore, accurately predicting early

<sup>1</sup>Cholangiocarcinoma Research Institute, Khon Kaen University, Khon Kaen, Thailand. <sup>2</sup>Department of Surgery, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. <sup>3</sup>Department of Systems Biosciences and Computational Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. <sup>4</sup>Medical Oncology Unit, Department of Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. <sup>5</sup>Department of Pathology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. <sup>6</sup>Proteomics Research Laboratory, Genome Institute, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathum Thani, Thailand. <sup>7</sup>Cholangiocarcinoma Research Institute, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand. ✉email: watclo@kku.ac.th

recurrence for each regimen in individual patients may guide the selection or modification of adjuvant treatment plans. Although Carcinoembryonic Antigen (CEA) and Cancer Antigen 19–9 (CA 19–9) have been utilized in screening, diagnosis, treatment monitoring, recurrence detection, and disease progression for CCA, they also have several limitations including specific to cancer types, overlap with benign conditions, limited diagnostic values, inconsistent levels of biomarker, lack of established cut-off values and limited role in early detection<sup>6,7</sup>. The ideal biomarker or set of biomarkers for accurately predicting recurrence remains challenging to identify. Thus, there is a pressing need to develop a rapid and highly efficient method to enhance diagnostic accuracy.

Biomarkers are crucial for improving the diagnosis and treatment of cholangiocarcinoma (CCA). The strong link between cancer and metabolic changes makes metabolomics a promising tool for discovering new biomarkers. Recent studies show that metabolomics can provide valuable molecular biomarkers, offering insights into the full range of metabolites in a biological system. By analyzing a wide array of small molecules in a single sample, metabolomics expands beyond clinical diagnostics to identify disease-specific markers<sup>8,9</sup>.

Metabolomics captures the physiological state of a biological system, where deviations from the normal metabolome can indicate disease, especially in cancer patients with recurrence or poor survival outcomes after surgery. Alterations in metabolite profiles have the potential to reveal novel diagnostic biomarkers. Additionally, serum and plasma metabolite analysis has been validated in multiple studies, demonstrating its ability to identify disease-specific metabolic signatures, distinguish between healthy and diseased states, and assess disease stage or severity<sup>10–12</sup>. Furthermore, applications in metabolomics can also be utilized to predict the recurrence of various types of cancer<sup>13–17</sup>. A previous study by Padthaisong et al. revealed that the metabolites in CCA patients after surgery differ significantly between those with no recurrence and those with recurrence. High-rate energy metabolism in recurrent cancer is recognized as a hallmark of cancer development, as it is essential for maintaining energy balance to support cancer cell survival and growth. In addition to this, alterations in other metabolic pathways, such as lipid and amino acid metabolism, have also been reported to play a role in tumor progression<sup>17</sup>. However, this study did not perform metabolomics analysis by dividing patients into early and late recurrence groups. Previous reports have suggested that patients with early recurrence and those with late recurrence should be analyzed separately, as they exhibit significantly different survival outcomes.

This study aimed to identify potential metabolic biomarkers and evaluate the effectiveness of serum metabolomics combined with machine learning in predicting early and late recurrence in CCA patients. Serum samples were analyzed using ultrahigh-performance liquid chromatography-mass spectrometry (UPLC-MS) to characterize metabolite profiles, while Support Vector Machine (SVM) was employed to develop a predictive model for recurrence stratification. The identified metabolites not only serve as promising biomarkers for clinical application but also provide valuable insights into the biochemical pathways underlying CCA recurrence, advancing our understanding of disease progression and supporting precision medicine approaches for improved patient management.

## Results

### Patient characteristics and patient outcomes

In this study, 88 CCA patients who underwent curative surgery and experienced disease recurrence during follow-up, as per clinical guidelines, were included. The patients were divided into a training cohort ( $n=60$ ) and a testing cohort ( $n=28$ ). Based on a previous study<sup>2</sup>, recurrent CCA cases were classified by recurrence-free survival (RFS) into early recurrence (RFS < 365 days) and late recurrence (RFS  $\geq$  365 days). The training cohort included 28 early recurrence cases and 32 late recurrence cases, while the testing cohort comprised 10 early recurrence cases and 18 late recurrence cases. Overview of clinicopathological characteristics—such as age, sex, tumor site, histological type, and TNM stage (based on the 8th edition of the American Joint Committee on Cancer [AJCC] Staging Manual)—and preoperative laboratory (liver function test and tumor biomarkers) was shown in Table 1.

### Global metabolomics analysis of recurrence in CCA patients

The untargeted metabolomics analysis identified 2,369 metabolites in positive mode and 1,872 in negative mode using accurate mass and MS/MS fragmentation. Metabolites were filtered via MetaboAnalyst 6.0, and OPLS-DA effectively distinguished early and late recurrence groups in both modes. A permutation test (100 iterations) confirmed model significance, with permR<sup>2</sup> = 0.98, permQ<sup>2</sup> = 0.92 (positive mode), and permR<sup>2</sup> = 0.929, permQ<sup>2</sup> = 0.986 (negative mode), all with  $p < 0.01$ , indicating no overfitting (Fig. 1A–B).

The variable importance in projection (VIP) plot identified the top 15 metabolites contributing to group separation (Fig. 1C–D). A volcano plot highlighted significant metabolites based on fold change (FC > 1.2 or < 0.83) and FDR-adjusted  $p < 0.05$  (Fig. 1E–F). Metabolites meeting VIP > 1.2, FC > 1.2 or < 0.83, and FDR-adjusted  $p < 0.05$  were retained, with duplicates removed. Those with lower AUC values or complex profiles were excluded, resulting in 90 significant metabolites, including amino acids, fatty acids, and lipids.

The identified candidate metabolites were categorized into common metabolites (predominantly detected in positive mode; see Supplementary Table S1) and lipid metabolites (primarily detected in negative mode; see Supplementary Table S2). This categorization enabled a more focused analysis of distinct metabolic pathways. This separation was based on the distinct biological roles and metabolic pathways associated with each group: common metabolites are involved in broad metabolic processes such as energy production (e.g., glycolysis, TCA cycle), biosynthesis of amino acids, nucleotides, and other essential molecules, while lipid metabolites are specifically associated with lipid metabolism, including energy storage, signaling, and membrane formation. This categorization allowed us to examine the unique contributions of lipid and non-lipid metabolic pathways to the observed metabolic alterations, providing clearer insights into the underlying biology of early and late recurrence in CCA.

	Training set (n = 60)		Testing set (n = 28)	
	Early recurrence (n = 28; 100%)	Late recurrence (n = 32; 100%)	Early recurrence (n = 10)	Late recurrence (n = 18)
Age; median (range)	63 (44–73)	62 (49–77)	59 (41–73)	62 (34–72)
Gender (n (%))				
Male	15 (54%)	21 (66%)	8 (80%)	12 (67%)
Female	13 (46%)	11 (34%)	2 (20%)	6 (33%)
Tumor location				
iCCA	17 (61%)	19 (59%)	6 (60%)	10 (56%)
eCCA	11 (39%)	13 (41%)	4 (40%)	8 (44%)
Tumor morphology				
MF/PI	14 (50%)	12 (38%)	5 (50%)	9 (50%)
ID/mixed type	14 (50%)	20 (62%)	5	9 (50%)
Surgical margin (R)				
R0	15 (54%)	21 (66%)	4 (40%)	11 (61%)
R1	13 (46%)	11 (34%)	6 (60%)	7 (39%)
Histological type				
Well	20 (71%)	26 (81%)	8 (80%)	14 (78%)
Moderately/poorly	8 (29%)	6 (19%)	2 (20%)	4 (22%)
Lymph node metastasis (N)				
N0	9 (32%)	21 (66%)	8 (80%)	12 (67%)
N1	19 (68%)	11 (34%)	2 (20%)	6 (33%)
TNM <sup>s</sup> staging				
I-II	5 (18%)	10 (31%)	2 (20%)	5 (28%)
III-IV	23 (82%)	22 (69%)	8 (80%)	13 (72%)
Preoperative laboratory data <sup>s</sup>				
Cholesterol (mg/dL)	174.5 (135–290)	180 (93–266)	197 (103–251)	190.5 (107–257)
Albumin (g/dL)	4.2 (2.9–4.8)	4.2 (2.9–5.3)	3.7 (2.6–4.4)	4.3 (2.8–4.7)
Globulin (g/dL)	3.6 (2.5–4.8)	3.3 (1.7–4.8)	3.3 (2.6–5.1)	3.1 (2.5–4.9)
Total protein (g/dL)	7.8 (5.4–8.3)	7.6 (4.6–8.8)	7.1 (6.6–7.7)	7.5 (5.4–9.4)
Direct bilirubin (mg/dL)	0.2 (0.1–1.3)	0.2 (0.1–1)	1.4 (0.2–5.1)	0.3 (0.1–2.9)
Total bilirubin (mg/dL)	0.4 (0.2–1.8)	0.5 (0.3–5.2)	1.9 (0.5–6.2)	0.5 (0.3–3.1)
AST (U/L)	29 (10–57)	29 (17–193)	56 (25–95)	44 (18–195)
ALT (U/L)	25 (10–99)	31 (2–88)	51 (30–78)	31 (17–228)
ALP (U/L)	177.5 (65–1068)	125 (61–716)	264.5 (92–899)	133.5 (67–397)
Tumor biomarkers <sup>s</sup>				
CA19-9 (ng/mL)	279.4 (0.6–1000)	112.2 (0.8–1000)	275.8 (1.5–1000)	195 (1.07–1000)
CEA (U/mL)	9.72 (1.03–1000)	7.2 (1.1–109.1)	8.8 (2.6–128.5)	7.3 (1.94–402)

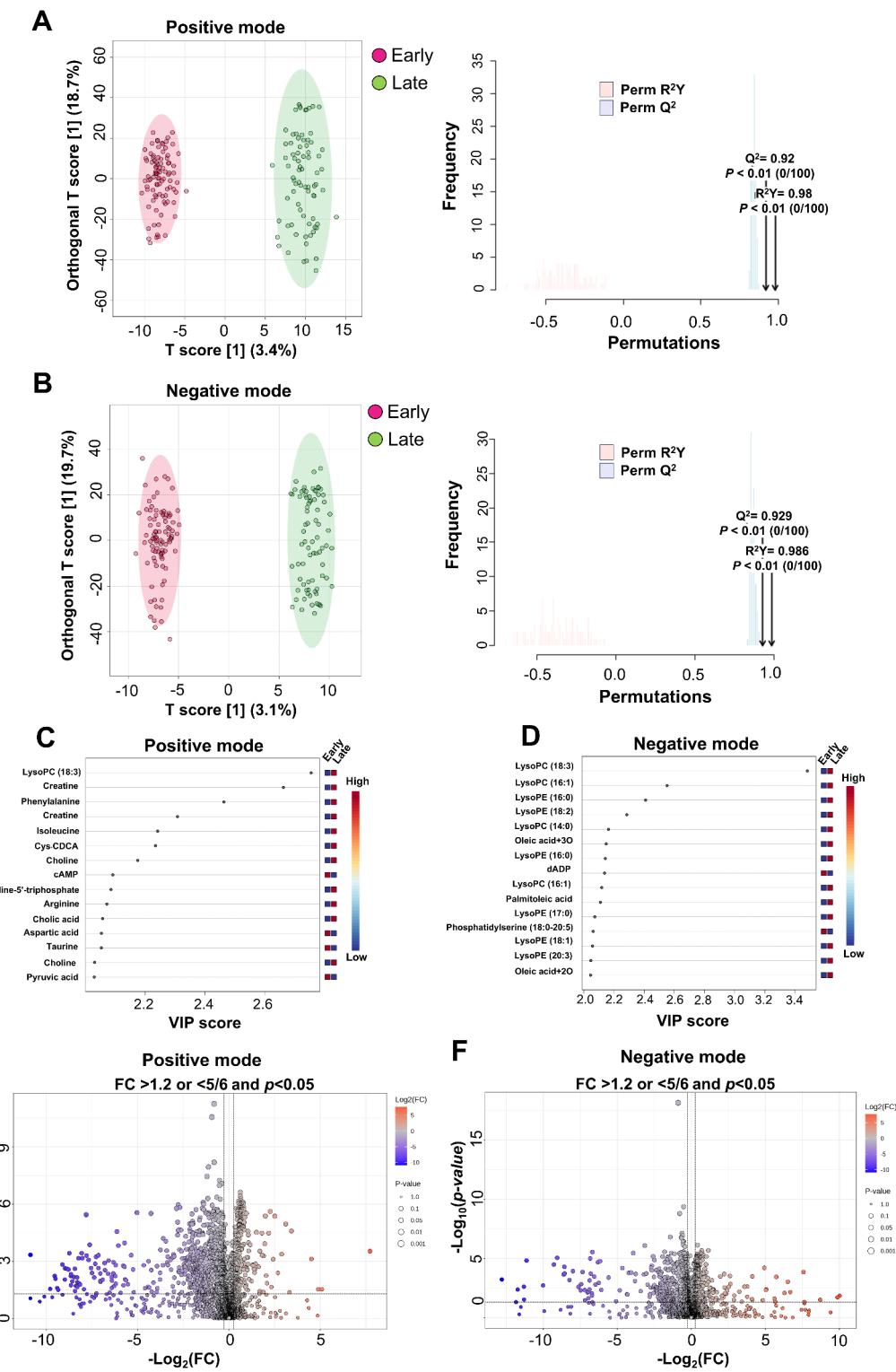
**Table 1.** Clinicopathological characteristics. <sup>s</sup>There is missing data in some cases.

The top 10 candidate metabolites with significant AUROC values included LysoPC(18:3/0:0), LysoPC(16:1/0:0), LysoPI(16:0/0:0), kynurenine, LysoPE(18:3/0:0), LysoPE(16:0/0:0), LysoPE(17:0/0:0), thymidine 5'-triphosphate, creatinine, and L-cysteine (Fig. 2A). Hierarchical clustering analysis (HCA) clearly separated early and late recurrence groups based on 90 differentially expressed metabolites—46 common metabolites (Fig. 2B) and 44 lipid metabolites (Fig. 2C). Among them, 20 metabolites and 5 lipids were upregulated in early recurrence, while 26 metabolites and 39 lipids were elevated in late recurrence, indicating distinct metabolic profiles between recurrence subtypes.

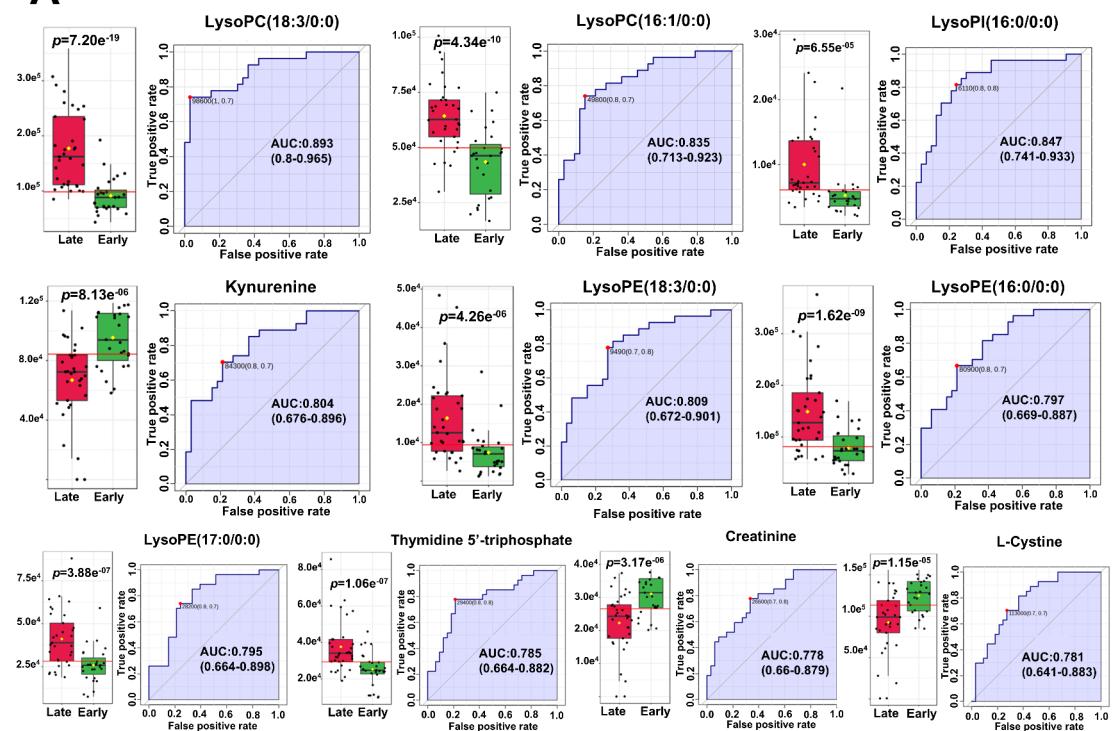
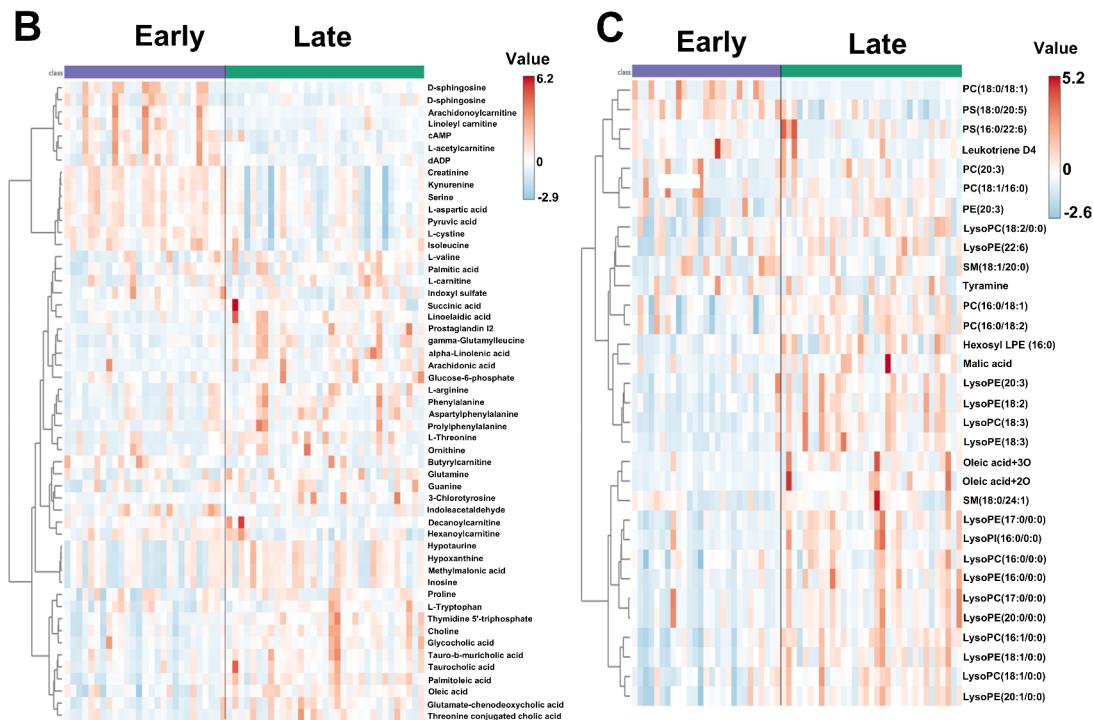
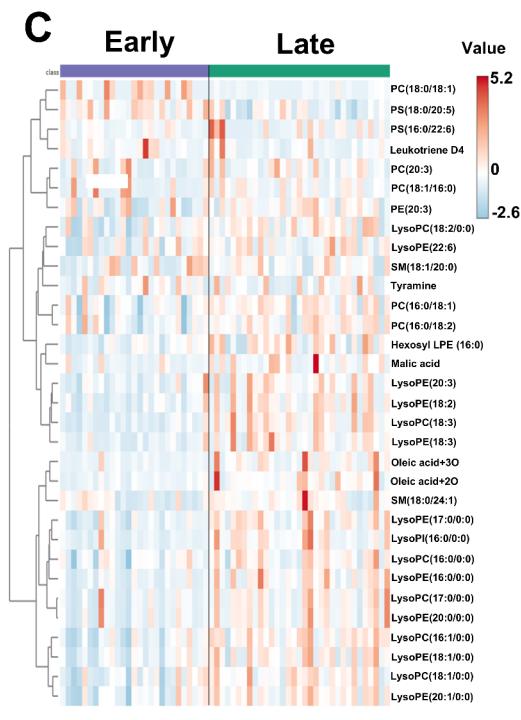
### Efficiency of candidate metabolites in differentiating early and late recurrence in CCA patients

To evaluate the ability of the 90 candidate metabolites to distinguish between recurrence groups, we performed OPLS-DA analysis in two steps. In the first step, the analysis was conducted using a training set, which included samples used for identifying the 90 metabolites and for training the model. The OPLS-DA analysis clearly separated the early and late recurrence groups, with an  $R^2$  of 0.573 (explanatory ability) and  $Q^2$  of 0.557 (predictive ability), indicating moderate to good performance. Cross-validation through a permutation test (2000 iterations) confirmed the model's robustness, yielding  $R^2 = 0.921$  and  $Q^2 = 0.868$ , with  $p < 5e^{-4}$ , indicating no overfitting (Fig. 3A).

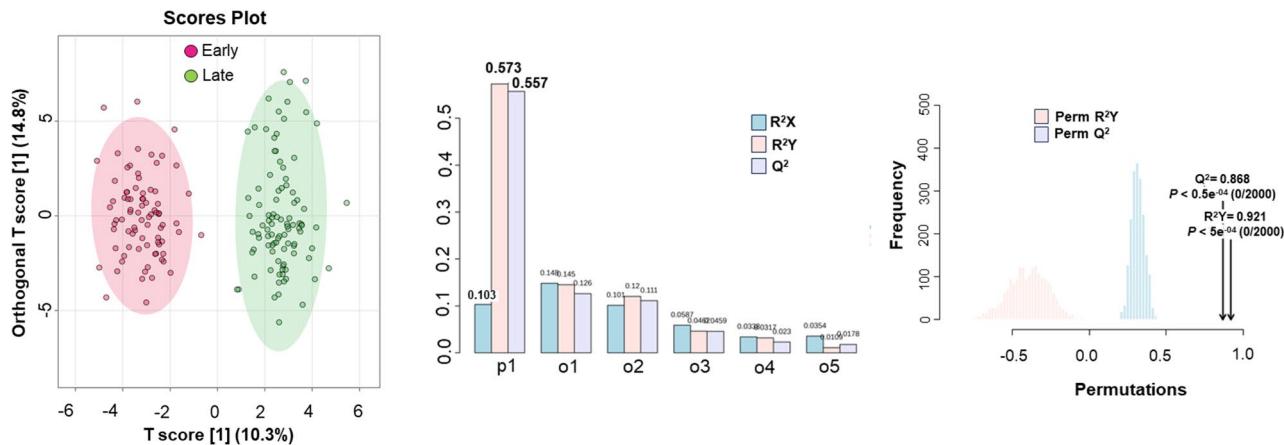
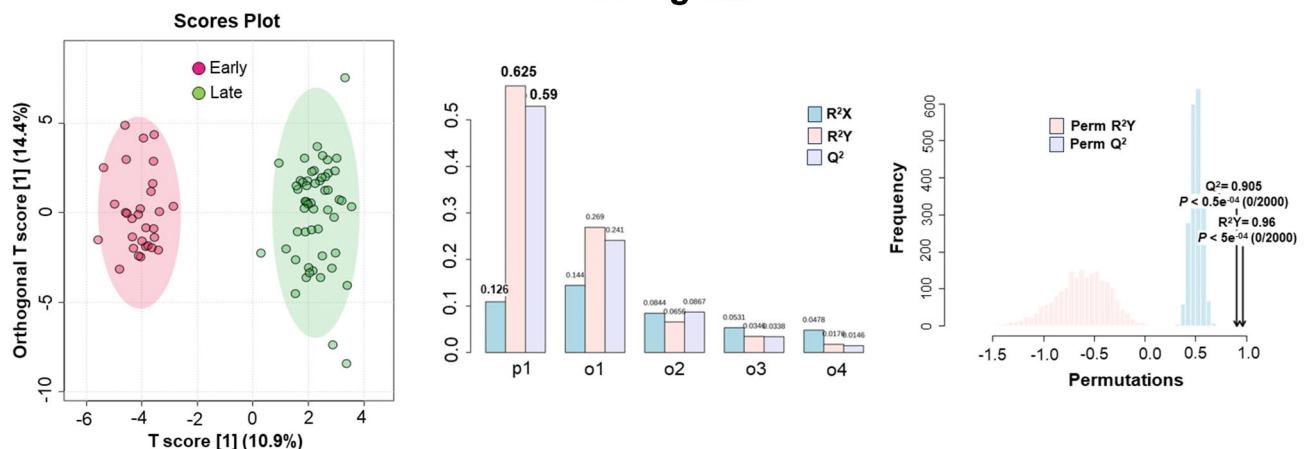
In the second step, the same 90 candidate metabolites were used to evaluate the model performance in a testing set. The OPLS-DA analysis successfully differentiated the two patient groups, with an  $R^2$  of 0.625 and  $Q^2$  of 0.59, indicating moderate to good performance. A permutation test with 2,000 iterations validated the



**Fig. 1.** Chemometrics analysis of metabolomics data. The OPLS-DA score plots for metabolites in positive (A) and negative (B) modes, with permutation tests to assess model robustness and stability. VIP plots of the top 15 metabolites in positive (C) and negative (D) modes, highlighting key metabolites contributing to group separation. Volcano plots of metabolites in positive (E) and negative (F) modes, displaying significant metabolites between the groups.

**A****B****C**

**Fig. 2.** Differential expression of candidate metabolites between early and late recurrence. Boxplots of the top 10 metabolites with the highest AUROC values, illustrating differential expression between early (green) and late (red) recurrence groups, alongside their corresponding ROC curves and AUC values (A). Heatmap of common metabolites, highlighting expression differences between the early and late recurrence groups (B). Heatmap of lipid metabolites, emphasizing significant lipid profile changes associated with the early and late recurrence groups (C).

**A****Training set****B****Testing set**

**Fig. 3.** Chemometric analysis of candidate metabolites for distinguishing early and late recurrence. OPLS-DA score plot for the training set showing the separation of 90 candidate metabolites between early and late recurrence groups. Model performance was evaluated by cross-validation (middle panel), with  $R^2$  indicating the proportion of variance explained and  $Q^2$  representing predictive accuracy. A permutation test (2000 permutations) yielded  $p < 0.001$  (right panel), confirming the model's robustness and statistical significance (A). OPLS-DA score plot for the testing set, illustrating separation of early and late recurrence based on the 90 candidate metabolites. Model validation was conducted with  $R^2$  and  $Q^2$  values from cross-validation (middle panel), and the permutation test (2000 permutations) showed  $p < 0.001$  (right panel), validating the model reliability and discriminative power (B).

results, yielding  $R^2 = 0.96$  and  $Q^2 = 0.905$ , with  $p < 5e^{-4}$ , further confirming the model's reliability and absence of overfitting (Fig. 3B). The results from both the training and testing sets demonstrate the effectiveness of the 90 candidate metabolites in distinguishing between recurrence groups in this study.

#### Support vector machine-based predictive modeling of early and late recurrence in CCA using candidate metabolites

This study identified serum metabolites that distinguished early and late recurrence groups, suggesting their potential as biomarkers for preoperative screening in CCA patients. These biomarkers might serve to predict recurrence patterns, guide treatment decisions, and improve patient monitoring. To construct predictive models, we employed a support vector machine (SVM) approach based on candidate metabolites, using a training set to develop the classification model and a testing set to validate model performance. The SVM model was constructed using the training set, which included 60 recurrent CCA patients (27 early and 33 late recurrence cases), through an initial feature selection process that identified 90 candidate metabolites, prioritizing their relevance to the classification task. The selection of metabolites for model construction was guided by the SVM algorithm, which focused on their ability to effectively differentiate between classes, emphasizing their contribution to model

performance rather than relying solely on statistical significance. Using these 90 metabolites, six distinct SVM models were developed, ensuring the identification of the most informative features for accurate classification (Fig. 4A and Supplementary Fig. S1A). The model 4, which incorporated 20 metabolites (Supplementary Table S3), demonstrated the highest predictive accuracy. The predicted class probabilities from the best-performing classifier, based on AUC, were presented in Fig. 4A. Confusion matrix analysis showed Model 4 achieved a true positive rate of 92.9%, a true negative rate of 94.3%, a false positive rate of 3.0%, and a false negative rate of 7.1%. The performance metrics were as follows: accuracy = 95%, precision = 96.29%, recall (sensitivity) = 92.86%, specificity = 96.88%, and F1-score = 94.4%. These findings indicate Model 4 as the most stable and accurate classifier. Feature selection frequency and mean importance measures highlighted the key metabolites contributing to model performance (Supplementary Table S3). The predictive outcomes of Model 4 closely aligned with diagnostic results. The Kaplan-Meier (KM) analysis of diagnosis showed significantly shorter recurrence-free survival (RFS) and overall survival (OS) in early vs. late recurrence patients (RFS = 200 vs. 786 days,  $p < 0.001$ ; OS = 264 vs. 960 days,  $p < 0.001$ ) (Table 2; Fig. 4B and Supplementary Fig. S1B). SVM-based predictions reflected a similar trend of RFS and OS as diagnostic results (RFS = 200 vs. 829 days,  $p < 0.001$ ; OS = 264 vs. 960 days,  $p < 0.001$ ) (Table 2; Fig. 4C and Supplementary Fig. S1C). Interestingly, model 3, which utilized only 10 metabolites, demonstrated predictive performance comparable to Model 4 while significantly reducing model complexity. In the training set, model 3 achieved similar AUC values and overall classification metrics to model 4 (the best model). Additionally, the prediction accuracy closely aligned with diagnosis (Table 2; Fig. 4D and Supplementary Fig. S1D).

The testing set included 28 recurrent CCA patients (10 early and 18 late recurrence cases). The same 90 metabolites from the training set were used to construct six SVM models. Model 4, incorporating the same 20 metabolites, was selected for evaluation. In the testing set, model 4 maintained strong predictive performance with an AUC of 0.898. Confusion matrix analysis yielded an accuracy of 85.71%, precision of 90%, recall of 75%, specificity of 93.75%, and an F1-score of 81.82% (Table 2; Fig. 4E and Supplementary Fig. S1E). Predictive accuracies across all six models ranged from 78.9 to 84%. The KM analysis confirmed that early recurrence patients exhibited significantly shorter survival outcomes in both diagnostic and SVM-based predictions. The diagnostic results showed RFS = 208 vs. 787 days ( $p < 0.001$ ) and OS = 345 vs. 984 days ( $p < 0.001$ ) (Table 2; Fig. 4F, and Supplementary Fig. S1F). Similarly, SVM-based prediction of model 4 produced RFS = 233 vs. 787 days ( $p < 0.001$ ) and OS = 381 vs. 984 days ( $p < 0.001$ ), further demonstrating the model's robustness and consistency across datasets (Table 2; Fig. 4G, Supplementary Fig. S1G). Moreover, to validate performance of model 3 of the training set, the same 10-metabolite model was validated in the testing set, where its predictions closely matched clinical outcomes (diagnosis) and model 4 (20-metabolite model). This highlights the potential for a more streamlined model with fewer metabolites while maintaining strong predictive accuracy (Table 2; Fig. 4H and Supplementary Fig. S1H).

Additionally, a comparative analysis between metabolite-based models and tumor biomarkers (CA19-9 and CEA) was conducted. The metabolite-based models demonstrated superior predictive performance, while tumor biomarkers showed lower accuracy in distinguishing early and late recurrence groups (Table 2). These results reinforce the clinical relevance of serum metabolites as recurrence biomarkers in CCA patients.

### Pathway analysis highlights global metabolic changes in recurrence of CCA patients

A metabolic pathway impact analysis using MetaboAnalyst 6.0 identified key pathways associated with serum metabolite changes. To improve accuracy, metabolites and lipid metabolites were analyzed separately. The analysis integrated metabolite set enrichment and pathway topology to extract biological insights for early and late recurrence. Pathway impact and enrichment were assessed using a global test and relative betweenness centrality, excluding unidentified metabolites or those lacking HMDB IDs. Figure 5 illustrates the results, where circle color and size indicate  $p$ -values and pathway impact values, respectively.

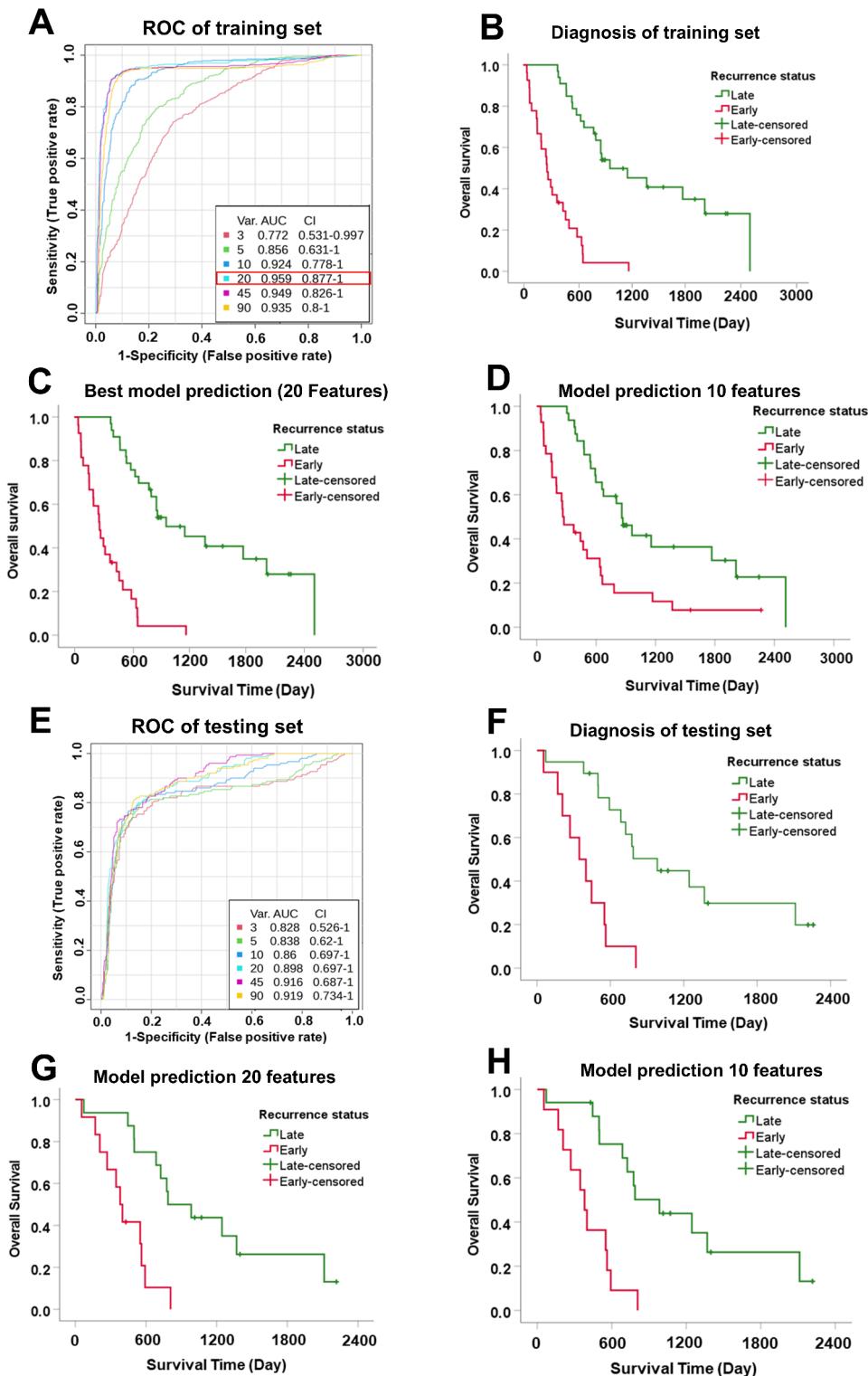
For common metabolites, nine significant pathways were identified for metabolites, including six amino acid metabolism pathways (arginine biosynthesis, alanine/aspartate/glutamate metabolism, glycine-serine/threonine metabolism, valine/leucine/isoleucine biosynthesis, arginine/proline metabolism, and tryptophan metabolism), along with the TCA cycle, glyoxylate/dicarboxylate metabolism, and sphingolipid metabolism (Fig. 5A). Pathway impact analysis confirmed significant involvement of amino acid metabolism, fatty acid oxidation, and the TCA cycle (Fig. 5C and Supplementary Table S4).

For lipid metabolites, key pathways included unsaturated fatty acid biosynthesis, glycerophospholipid metabolism, sphingolipid metabolism, alpha-linolenic acid metabolism, and arachidonic acid metabolism. Enrichment analysis highlighted significant pathways such as alpha-linolenic acid metabolism, arachidonic acid metabolism, phospholipid biosynthesis, beta-oxidation of long-chain fatty acids, glycolipid metabolism, and oxidation of branched-chain fatty acids (Fig. 5B). Fatty acid metabolism showed strong links to metabolic alterations (Fig. 5D and Supplementary Table S5).

These results suggested that the identified pathways, related to both metabolites and lipid metabolites, are closely linked to the observed alterations in amino acid metabolism, lipid metabolism, and energy metabolism (TCA cycle), which may contribute to disease progression.

### Discussion

Cholangiocarcinoma (CCA) has a high recurrence rate after surgical resection, with the recurrence-free interval being a key factor in disease severity and survival. Early recurrence, occurring within 6 months to 1 year, is associated with significantly lower survival rates and serves as an independent prognostic factor. In this study, we categorized recurrence into early and late groups using a 365-day cut-off. Understanding recurrence biology is critical for developing diagnostic tools to detect relapse. Cancer recurrence involves proliferation, inflammation, migration, invasion, immune evasion, and cell membrane remodeling, all influenced by metabolic alterations.



These metabolic changes, particularly in nutrient uptake, provide energy for growth and development. Identifying these alterations may reveal biomarkers for CCA, offering new strategies for treatment and reducing early recurrence risks.

Currently, no reliable biomarkers exist for preoperative screening to predict recurrence risk or subtype. Common markers like CEA and CA19-9 have limitations, including low specificity and inability to distinguish between early and late recurrence<sup>6,7</sup>. Consistent with previous studies, our findings show that CEA and CA19-9 have low accuracy, high false-positive rates, and poor classification of recurrence subtypes. Additionally, their suboptimal performance in predicting DFS and OS might be attributed to the wide variation in biomarker levels, complicating the determination of appropriate cut-off values (Table 1). This raises the question of whether a biomarker panel or multiple biomarkers in combination could enhance diagnostic accuracy. These limitations

**Fig. 4.** Support Vector Machine (SVM)-based predictive modeling of early and late recurrence using candidate metabolites. SVM classification of early and late recurrence using 90 candidate metabolites in the training set. Model 4 (20 metabolites) achieved the highest accuracy (A). Kaplan-Meier (KM) curves based on diagnostic outcomes, showing shorter overall survival (OS) in early recurrence than late recurrence (B). KM curves based on SVM Model 4 predictions, confirming shorter OS in early recurrence than late recurrence (C). KM curves based on SVM model 3 (10 metabolites) predictions, confirming shorter OS in early recurrence than late recurrence (D). The SVM classification in the testing set using the same 90 metabolites. Model 4 achieved an AUC of 0.898, with performance metrics assessed similarly to the training set (E). KM curves based on diagnostic outcomes in the testing set, showing survival differences between early and late recurrence (F). KM curves based on SVM model 4 predictions in the testing set, mirroring diagnostic trends in OS (G). Model 3 (10 metabolites) predictions in the testing set, demonstrating clear separation of OS between early recurrence and late recurrence as similar trend in model 4 and diagnostic outcomes in the testing set (H).

	Training set		Testing set		Tumor biomarkers (training set)			Diagnosis result <sup>a</sup>	
	20 Features	10 Features	20 Features	10 Features	CA19-9 (ng/mL)	CEA (U/mL)	Combined	Training set	Testing set
Model performance									
ROC	0.959	0.924	0.898	0.860	0.654	0.521	0.589	–	–
Accuracy	95%	94.92%	85.71%	79.31%	63.33%	53.33%	60%	–	–
Precision	96.29%	92.86%	90%	66.67%	42.86%	39.29%	39.29%	–	–
Recall (sensitivity)	92.86%	96.77%	75%	88.24%	66.67%	50%	61.11%	–	–
Specificity	96.88%	96.30%	93.75%	80.00%	52.17%	44%	47.78%	–	–
F1-score	94.4%	94.52%	81.82%	72.73%	81.25%	65.63%	78.13%	–	–
Clinical outcome									
RFS early vs. late (day); <i>p</i> -value	200 vs. 829 days <i>p</i> <0.001	201 vs. 655 days <i>p</i> <0.001	233 vs. 787 days <i>p</i> <0.001	233 vs. 787 days <i>p</i> <0.001	281 vs. 431 days <i>p</i> =0.109	382 vs. 436 days <i>p</i> =0.234	281 vs. 409 days <i>p</i> =0.158	200 vs. 786 days <i>p</i> <0.001	208 vs 787 days <i>p</i> <0.001
OS early vs. late (day); <i>p</i> -value	264 vs. 960 days <i>p</i> <0.001	264 vs. 861 days <i>p</i> <0.001	381 vs. 984 days <i>p</i> <0.001	381 vs. 984 days <i>p</i> <0.001	408 vs. 663 days <i>p</i> =0.069	375 vs. 663 days <i>p</i> =0.043	477 vs. 663 days <i>p</i> =0.058	264 vs. 960 days <i>p</i> <0.001	345 vs. 984 days <i>p</i> <0.001

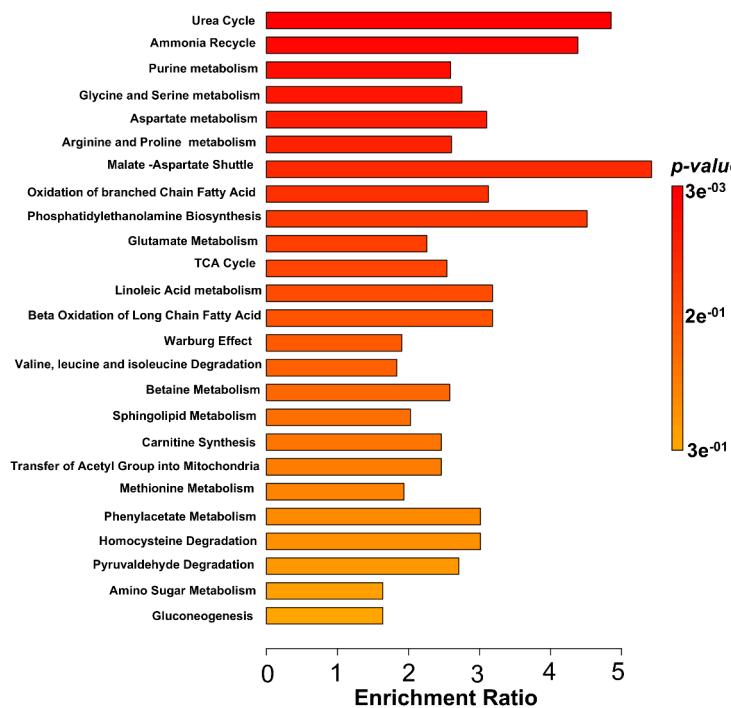
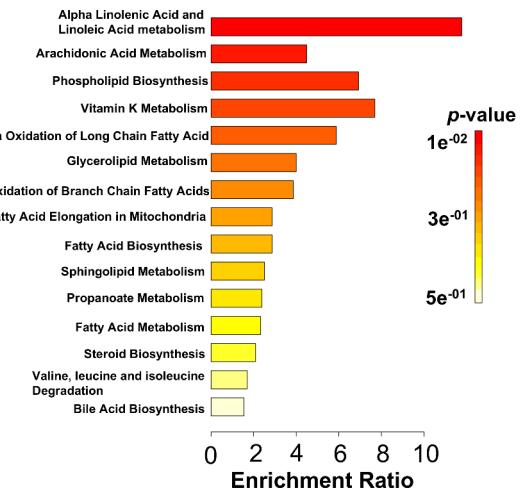
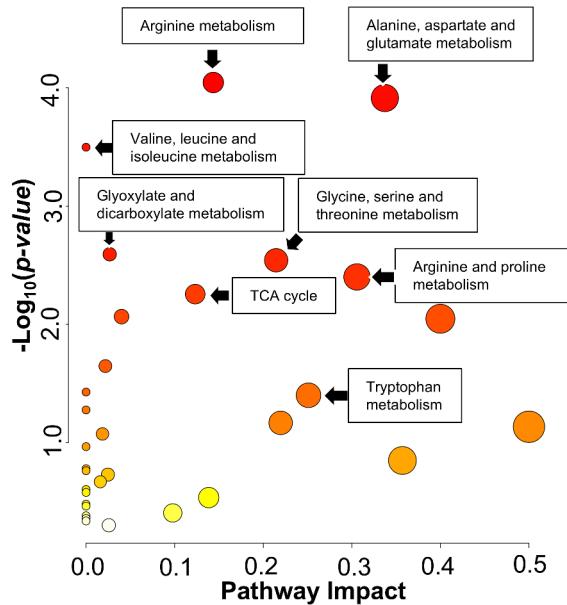
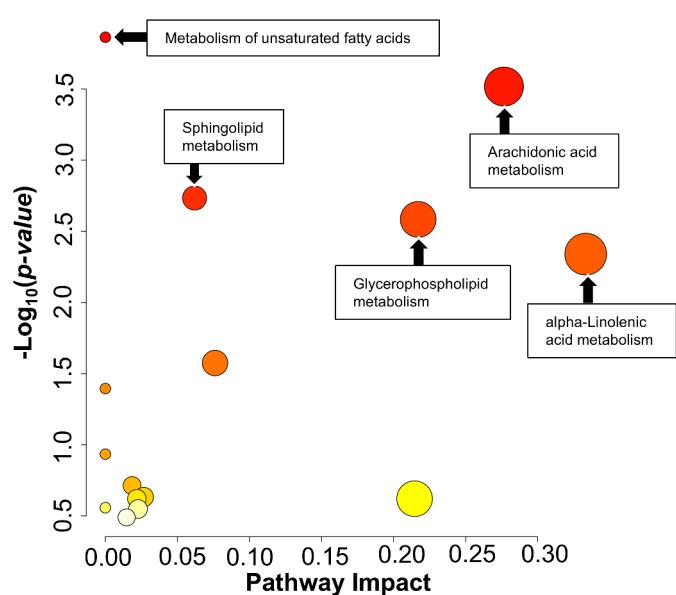
**Table 2.** Model performance and prediction of metabolite-base biomarker using SVM and tumor biomarkers.

<sup>a</sup>The diagnosis result served as a reference for evaluating the model performance in both the training set and testing set.

highlight the need for novel, highly specific biomarkers that can accurately predict recurrence patterns and improve clinical management strategies.

In this study, we analyzed metabolites from preoperative serum samples of CCA patients, revealing distinct metabolic profiles between early and late recurrence groups. Our findings align with prior research in breast, lung, and liver cancers. Using chemometric tools, specifically OPLS-DA, along with statistical analysis via Volcano Plot, we established criteria for identifying metabolites differentiating between early and late recurrence. We identified 90 metabolites (Supplementary Table S1-2), categorized into general metabolites and lipid metabolites (Fig. 2B-C). OPLS-DA effectively distinguished recurrence subtypes, demonstrating the strong discriminatory power of these metabolites and their potential as biomarkers for preoperative recurrence classification in CCA.

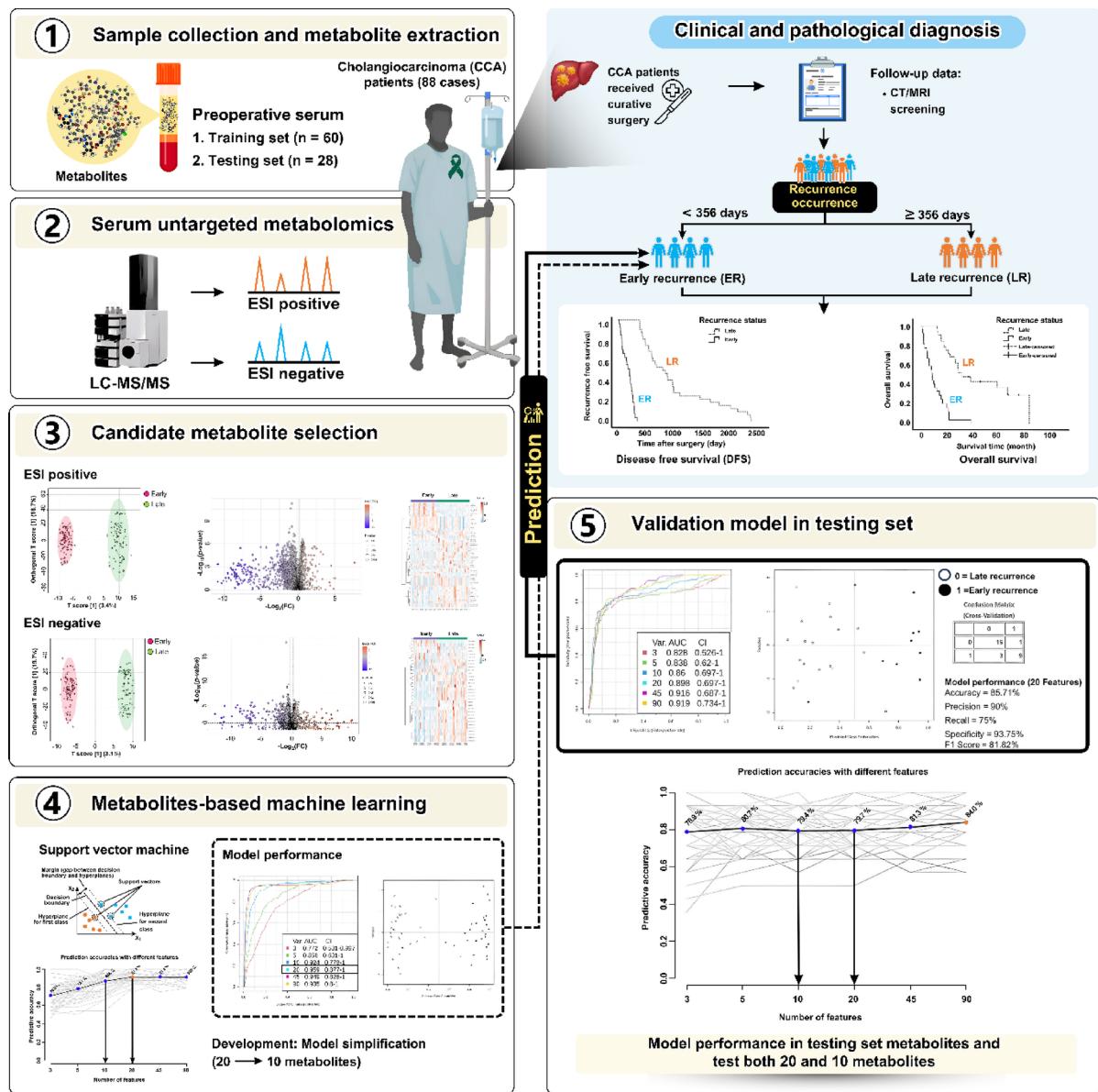
To enhance predictive capability, we employed Support Vector Machine (SVM), a robust machine learning algorithm optimized for classification tasks. SVM identifies an optimal hyperplane to maximize separation between distinct data groups, ensuring high accuracy and reliability<sup>18,19</sup>. It is important to note that although the top candidate metabolites identified in Figs. 1 and 2 from the initial analysis were statistically significant, not all of them were incorporated into the final metabolite signatures in SVM model construction (Fig. 4A). This discrepancy stems from the differing selection criteria employed in these analyses. Specifically, the top 10/15 metabolites were selected based on their individual statistical significance, whereas the final metabolite signatures were constructed by prioritizing metabolites with the highest collective discriminatory power within the SVM model from candidate metabolites. This methodological distinction is consistent with previous studies<sup>20,21</sup>, wherein metabolites that demonstrate statistical significance in univariate analyses do not necessarily contribute to the predictive capacity of multivariate models. By emphasizing metabolites with superior classification performance, this approach enhances the robustness, generalizability, and interpretability of the predictive model, ensuring optimal differentiation between study groups. In our study, SVM significantly improved classification accuracy between early and late recurrence groups. Among the 90 metabolites, SVM constructed six models with varying metabolite numbers, achieving AUROC values between 0.772 and 0.959. Model 4, containing 20 metabolites, demonstrated the highest predictive performance. This model provided DFS and OS predictions comparable to physician diagnoses post-surgery and follow-up, highlighting its potential as a clinical tool for

**A****B****C****D**

**Fig. 5.** Pathway analysis of serum metabolic alteration in CCA recurrence. Pathway enrichment analysis for common metabolites (A). Pathway enrichment analysis for lipid metabolites (B). Pathway impact analysis for common metabolites (C). Pathway impact analysis for lipid metabolites (D), highlighting significant pathways. Circle size and color represent *p*-value and pathway impact, with larger circles indicating greater pathway impact.

recurrence classification. The model's performance was validated in an independent testing group, where results closely aligned with observed clinical outcomes as shown in Table 2; Fig. 4F-G and Supplementary Fig. S1F-G.

To facilitate clinical application, using 20 metabolites for prediction could be considered complex and challenging to implement in practice. Therefore, simplifying the model by reducing the number of metabolites is crucial for clinical translation. However, this simplification often comes at the cost of decreased model



**Fig. 6.** Summary of key findings and concepts discussed in the study. Illustration created using Canva software (<https://www.canva.com/>).

performance. In this study, we found that model 3, constructed with 10 metabolites, exhibited predictive performance comparable to model 4, which utilized the full set of 20 metabolites and represented the best-performing model. When applying model 3 to predict recurrence in CCA patients, the predictions for RFS and OS were closely aligned with clinical diagnoses made by physicians. Furthermore, the predictive accuracy of model 3 was comparable to that of model 4, demonstrating its robustness despite the reduced number of metabolites. Notably, the 10-metabolite set of model 3 was validated effectively in the testing set, reinforcing its potential reliability in clinical settings. These findings suggested that predictive models leveraging a reduced metabolite set could offer a practical alternative for preoperative recurrence prediction in CCA (Table 2). Such models provide a balance between maintaining sufficient predictive power and reducing complexity, paving the way for their potential integration into routine clinical workflows. By adopting this approach, clinicians might gain a valuable tool for identifying patients at risk of recurrence, enabling timely and personalized interventions to improve patient outcomes (Fig. 6). Our findings are consistent with previous research, where the SVM algorithm has been widely applied in metabolomics, particularly for biomarker identification, patient classification, and disease prediction models. Notable applications include research on obesity<sup>22</sup> and various cancers, such as esophageal squamous cell carcinoma<sup>23</sup>, breast cancer<sup>11</sup>, colon cancer<sup>12</sup>, and ovarian cancer<sup>24</sup>. These studies could identify abnormalities using SVM, which is constructed from metabolites, and could be further developed for future management and treatment planning.

To investigate the molecular mechanisms underlying the 90 metabolites identified in this study, we aimed to elucidate the biological pathways contributing to recurrence. Our findings provide deeper insights into the metabolic processes driving CCA recurrence, highlighting the involvement of metabolites in various pathways.

Cancer progression, especially in recurrent cancer, cells have an elevated ATP demand to support rapid proliferation, survival under stress, and resistance to therapy. To meet these demands, cancer cells undergo significant metabolic changes, including enhanced amino acid degradation and phospholipid breakdown for fatty acid oxidation. Amino acid catabolism fuels the tricarboxylic acid (TCA) cycle, providing ATP and key intermediates, while fatty acids undergo beta-oxidation to generate large amounts of ATP. This metabolic flexibility allows recurrent cancers to thrive in nutrient-limited, hypoxic microenvironments, supporting continued growth and metastasis. Enhanced ATP production also supports survival pathways, DNA repair, and drug resistance mechanisms, contributing to the resilience of recurrent cancers against conventional therapies. This study revealed that the metabolic pathways associated with CCA recurrence primarily involved amino acid metabolism, lipid metabolism (including beta-oxidation, glycolipid, phospholipid, and arachidonic acid metabolism), and energy production (TCA cycle) (Fig. 5). These metabolic alterations reflected the adaptive changes of cancer cells during both early and late recurrences.

The study identified a set of amino acids, detected by LC-MS/MS, associated with CCA recurrence, including both essential amino acids like phenylalanine, arginine, tryptophan, threonine, valine, isoleucine, and non-essential amino acids like aspartic acid, serine, cysteine, asparagine, glutamine, and proline. Notably, branched-chain amino acids, such as valine and isoleucine, were emphasized for their roles in cancer progression. These amino acids are involved in key metabolic pathways, including arginine, alanine, glutamate, and tryptophan metabolism, which converge on energy production through the TCA cycle, fueling cancer cell growth and survival. The study observed reduced levels of most amino acids in the serum of early recurrence, highly aggressive CCA subtypes compared to late recurrence, suggesting that cancer cells consume amino acids more rapidly to support growth, protein synthesis, and energy production through the malate-aspartate shuttle and TCA cycle. This enhanced amino acid utilization for energy production promotes uncontrolled proliferation, metastasis, and adaptation to microenvironmental stress, driving cancer progression. Previous studies corroborate this finding, demonstrating that amino acids are vital for maintaining redox balance, regulating epigenetics, and modulating immune responses linked to tumorigenesis<sup>25,26</sup>. Additionally, Padthaisong et al. showed that recurrent (aggressive) CCA requires increased energy production, reflected by a significant reduction in amino acid levels compared to non-recurrent (less aggressive) CCA<sup>17</sup>.

Our study also revealed the pivotal roles of specific amino acids in cancer progression, where they modulate signaling pathways, promote angiogenesis, and influence immune responses. Notably, tryptophan plays a crucial role in enabling tumors to adapt to fluctuating environments, evade immune surveillance, and establish a microenvironment that supports growth and metastasis, primarily through its metabolic products, especially kynurenone<sup>27</sup>. Additionally, during cancer progression, the synthesis of glutathione from glutamate, cysteine<sup>28</sup>, and serine-derived one-carbon units via the folate cycle supports NADPH production<sup>29</sup>, which is vital for maintaining redox balance, especially given the high energy demands of cancer cells. This metabolic process generates reactive oxygen species (ROS), and specific amino acids help regulate redox balance, as observed in pancreatic cancer growth<sup>30</sup> and breast cancer recurrence<sup>31</sup>.

Tryptophan metabolism is critical in cancer progression beyond energy production. In early recurrences, we observed upregulation of tryptophan metabolites, including kynurenone, indole-3-acetaldehyde (IAAld), and indoxyl sulfate. The kynurenone pathway, the primary metabolic route for tryptophan, produces metabolites that promote tumor progression by suppressing immune responses, promoting regulatory T-cell differentiation, and inducing inflammation. Jia Y et al. showed that high kynurenone levels in non-small cell lung cancer correlate with higher cancer stages and lower overall survival<sup>32</sup>. Similarly, elevated kynurenone metabolites are associated with increased mortality in stage I–III colorectal cancer patients<sup>33</sup>. Additionally, tryptophan is metabolized by gut microbiota into indole derivatives like indole-3-acetaldehyde (3-IAA) and indoxyl sulfate. Tintelnot J et al. found that 3-IAA in the serum of pancreatic cancer patients and chemotherapy-sensitive mice correlates with improved progression-free and overall survival<sup>34</sup>. Meanwhile, indoxyl sulfate has been shown to promote colorectal cancer cell proliferation by activating the aryl hydrocarbon receptor and Akt signaling pathways, inducing EGFR expression<sup>35</sup>.

Moreover, we also found that lipid metabolism pathways—such as beta-oxidation, glycolipid metabolism, phospholipid metabolism, unsaturated fatty acid metabolism, and arachidonic acid metabolism—in CCA recurrence. In particular, lipid-derived carnitines, including the acetylated forms of L-carnitine like L-acetylcarnitine (LAC), butyrylcarnitine, hexanoylcarnitine, and others, were upregulated in early recurrences of CCA. This underscores the critical role of energy metabolism in the aggressive nature of CCA, which is a rapidly progressing cancer. Lipid-derived carnitines facilitate the transport of fatty acids into mitochondria, where they are oxidized via  $\beta$ -oxidation, activating the TCA cycle and contributing to energy production (Fig. 5B). This supports tumorigenesis and cancer progression, alongside glucose and amino acid metabolism. Such metabolic reprogramming allows cancer cells to survive under nutrient-limiting conditions, supporting processes like membrane biosynthesis, energy production, and the generation of signaling intermediates<sup>36</sup>. Previous studies have also shown that lipid metabolism alterations are linked to poor survival outcomes and cancer recurrence in cancers such as breast<sup>37</sup>, liver<sup>38</sup>, and cholangiocarcinoma (CCA)<sup>17,39</sup>.

Our study identified significant changes in glycerophospholipid and sphingolipid metabolism during early recurrence of CCA. Phospholipids such as phosphatidylcholines (PC), phosphatidylinositols (PI), phosphatidylethanolamines (PE), phosphatidylserines (PS), and sphingomyelin (SM) are crucial for cell membrane structure, fluidity, and permeability. They facilitate substance exchange between cells and their environment, playing key roles in signaling, immune responses, and apoptosis. Lysophospholipids (e.g., LysoPC, LysoPE, LysoPI) derived from phospholipid degradation regulate inflammation, immune responses,

and cellular signaling. Altered lysophospholipid levels can indicate cancer. Metabolic reprogramming of phospholipids, marked by enhanced turnover, supports energy production, membrane biosynthesis, and tumor growth, contributing to malignant transformation and metastasis. They could serve as valuable biomarkers for assessing cancer progression and recurrence risk, providing a means of early diagnosis and prognosis in oncology<sup>40,41</sup>. Notably, we observed significant differences in the serum levels of specific phospholipids that could effectively distinguish early from late recurrence in CCA patients. These included LysoPC(18:3/0:0), LysoPC(16:1/0:0), LysoPE(0:0/18:3), LysoPC(14:0/0:0), LysoPE(16:0/0:0), LysoPI(16:0/0:0), LysoPE(20:1/0:0), LysoPE(18:1/0:0), LysoPC(18:1/0:0), and LysoPE(20:3/0:0), which are consistent with previous studies that highlight the relationship between lipid profile alterations and cancer pathogenesis<sup>42</sup> such as bladder<sup>43</sup>, ovarian<sup>44</sup>, liver cancer<sup>45</sup>, squamous carcinoma and adenocarcinoma lung cancers<sup>46</sup>. Supporting our findings, a previous report by Padthaisong et al. demonstrated distinct lipid metabolism profiles—particularly in triglycerides (TG) and diglycerides (DG)—between recurrent and non-recurrent CCA. They also identified certain phospholipids, such as PC (18:0/22:6) and PE (16:0, 18:0, 18:1, and 20:3), with no detectable levels of lysophospholipids. Despite these variations, lipid metabolism profiles appear to be sufficiently robust to serve as potential biomarkers for predicting recurrence and non-recurrence (NR) in CCA<sup>17</sup>. Furthermore, in 2024, Bi et al. examined alterations in phospholipid metabolites, particularly PC(32:0), LysoPA(16:0/0:0), and LysoPC(16:0/0:0), in CCA patients compared to healthy controls. They identified significant differences in lipid metabolite profiles, particularly in several phospholipids and lysophospholipids, which can be used to discriminate between healthy individuals and CCA patients. Moreover, they demonstrated that specific phospholipids and lysophospholipids, notably LysoPA(16:0/0:0) and LysoPC(16:0/0:0), effectively differentiate CCA from HCC<sup>10</sup>. This underscores the critical role of alterations in lipid metabolites, particularly phospholipids and lysophospholipids, in reflecting disease-specific abnormalities and indicating the severity of the condition. Building on previous studies in CCA and our findings, these results strengthen the hypothesis that lipid metabolite patterns—especially those involving lysophospholipids (LPL) and phospholipids (PL)—could distinguish distinct profiles among cancer subtypes, including non-recurrent (NR) CCA, recurrent (R) CCA; early recurrence and late recurrence, and healthy tissues.

Finally, our study revealed significant changes in unsaturated fatty acid metabolism, with reduced levels of oleic acid, alpha-linolenic acid (ALA), and linoleic acid (TFA) in the early recurrence group compared to the late recurrence group. These essential fatty acids are crucial for membrane fluidity, cellular signaling, and lipid storage. Their decrease suggests altered fatty acid metabolism, potentially promoting inflammation, modulating the tumor microenvironment, and inducing immunosuppression, which contribute to cancer progression, especially in early recurrence<sup>47</sup>. Monounsaturated (MUFA) and polyunsaturated fatty acids (PUFA) play a dual role in cancer, supporting energy production, membrane formation, and pro-inflammatory processes that promote tumor-supportive environment and cancer survival under nutrient-limiting conditions<sup>48,49</sup>. Our findings align with studies showing that aggressive cancers, like ovarian and colorectal cancer, increase unsaturated fatty acid uptake to meet metabolic demands. Additionally, we observed an interaction between unsaturated fatty acid metabolism and arachidonic acid metabolism. Decreased arachidonic acid levels and increased leukotriene A4 suggested that arachidonic acid may convert into leukotrienes, promoting inflammation and supporting cancer progression<sup>50</sup>.

Our results were consistent with previous studies on metabolomics-based prediction of cancer recurrence. In CCA, significant metabolic differences were observed between post-surgical patients who experienced recurrence and those who remained recurrence-free. Integrated global metabolomics and lipidomics analyses revealed that key energy metabolism pathways, such as pyruvate metabolism and the tricarboxylic acid (TCA) cycle, were downregulated in patients with recurrence. In contrast, most lipid species—including triglycerides, phosphatidylcholines, phosphatidylethanolamines, and phosphatidic acids—were upregulated in these patients. These findings suggest that dysregulation of energy metabolism and lipid homeostasis may play a critical role in CCA recurrence, offering potential biomarkers for recurrence prediction and therapeutic targeting<sup>17</sup>. Similarly, in pancreatic ductal adenocarcinoma (PDAC), metabolomics studies have highlighted key metabolic alterations and potential therapeutic targets following neoadjuvant chemoradiation therapy. Choline metabolism emerged as a critical pathway associated with recurrence in PDAC patients. Furthermore, levels of phosphocholine, carnitine, and glutathione were strongly correlated with recurrence-free survival, particularly in patients undergoing this treatment regimen<sup>15</sup>. In gastric cancer, metabolomic profiling has enabled the stratification of patients into high- and low-risk recurrence groups. Four metabolites—asppartate, beta-alanine, guanosine diphosphate, and glycine—were identified as robust predictive biomarkers for recurrence risk. Lower concentrations of these metabolites were associated with an increased risk of recurrence and poorer survival outcomes, establishing their utility as potential cutoff values for risk assessment<sup>14</sup>. Likewise, in ovarian cancer, specific metabolites have been identified as indicators of cancer risk, poor survival outcomes, and increased recurrence likelihood. Phospholipids and their derivatives, along with alterations in amino acids and lipid profiles, were significantly associated with the risk of ovarian cancer recurrence<sup>16</sup>. These results highlight the potential of metabolic biomarkers in predicting cancer recurrence, a process driven by significant metabolic reprogramming in cancer cells, particularly emphasizing the critical roles of energy production, amino acid and lipid metabolism, and glycophospholipid metabolism (cell membranes) in supporting cancer progression and resilience.

In conclusion, our study highlights the potential of metabolic profiling to identify biomarkers for predicting recurrent status in CCA. Distinct metabolic signatures were detected between early and late recurrence, with key pathways such as amino acid metabolism, lipid metabolism, and energy production playing pivotal roles. The application of SVM models, demonstrated effective classification of recurrence subtypes, offering promise for preoperative screening. The simplification of the metabolite panel to 10 metabolites ensures high predictive

accuracy and feasibility for clinical use. These findings pave the way for utilizing metabolic signatures in clinical decision-making, enhancing early detection, and improving patient outcomes in CCA.

## Materials and methods

### Ethics approval

This study was conducted in accordance with Good Clinical Practice guidelines, the Declaration of Helsinki, and relevant national laws and regulations governing clinical research. Informed consent was obtained from all participants, and all study procedures were reviewed and approved by the Khon Kaen University Ethics Committee for Human Research (reference number HE661318).

### Population and sample group

This research involved serum samples from 88 patients diagnosed with cholangiocarcinoma. Data were retrospectively collected from medical records at Srinagarind Hospital, Faculty of Medicine, Khon Kaen University, Thailand, covering clinical information from January 1, 2017, to December 31, 2021. Serum samples were sourced from the biobank at the Cholangiocarcinoma Research Institute (CARI), Khon Kaen University. Prognostic factors were gathered using a retrospective data collection form and the ISAN Cohort database from CARI. The collected data included the age at diagnosis, gender, histological confirmation, tumor size, cancer grade, cancer staging, surgical margins, lymph node metastasis, lymphovascular invasion, histological grade, and details of chemotherapy received. Finally, tumor morphology (gross examination) and pathological findings were correlated with previous report and the 8th AJCC Staging Manual<sup>51</sup>, respectively.

### Clinical outcome follow-up

The follow-up period for patients with cholangiocarcinoma extended from the date of surgery for a minimum of five years, commencing on January 1, 2017. All causes of death were monitored through life status verification from the Ministry of Interior's database, supplemented by additional data from medical records documented by physicians. Typically, patients were scheduled for follow-up visits at Srinagarind Hospital, Faculty of Medicine, Khon Kaen University, every six months for at least five years after treatment. The variables studied included overall survival (OS), and disease-free survival (DFS).

### Sample collection and serum preservation

For serum sample collection from patients with cholangiocarcinoma prior to surgical treatment, blood was drawn from a vein (venipuncture) with a volume of 5 milliliters into a clot blood tube. It was ensured that clot formation was complete before centrifugation. Serum was then separated from red blood cells using a centrifuge at 3,000–3,500 RPM at 4 °C for 10 min. The serum was aspirated and aliquoted into 1 µl portions in Eppendorf tubes to avoid repeated thawing of samples, and then stored at -80 °C in the biobank of the Cholangiocarcinoma Research Institute, Khon Kaen University, until further analysis.

### Sample preparation

Serum was centrifuged at 14,000 rpm for 10 min. Twenty µl of supernatant was aliquoted and mixed with 80 µl of methanol, containing 25 ng/ml sulfa mix standards (sulfamethizole, sulfamethazine, sulfachloropyridazine and sulfadimethoxine). Fifteen µl of each sample was mixed to create a pooled QC sample. Seven dilution QCs were made by diluting pooled QC sample to 0%, 1%, 10%, 20%, 50%, 80% and 100% in concentration<sup>52</sup>. All samples were centrifuged at 14,000 rpm for 10 min. Supernatant was transferred to LC-MS vial and subjected to LC-MS analysis.

### LC-MS/MS acquisition

Dilution QCs were acquired at the beginning from low to high concentration. Serum sample was run in triplicates, where pooled QC sample was inserted every 10 runs. All samples were analyzed using an Agilent 1290 Infinity II LC system connected with a 6545XT Q-TOF mass spectrometer (Agilent Technologies, USA). Electrospray ionization (ESI) was used as an ionization source. LC separation was conducted on Agilent Poroshell 120 EC-C18 column (2.1 × 100 mm, 2.7 µm) at 50 °C. Injection volume was 10 µl for both positive and negative ionization modes. Mobile phases A and B were a 0.1% formic acid (FA) in water and acetonitrile (ACN), respectively. LC gradient was set as follows; 0% B for 0.5 min, 0–55% B in 10 min, 55–75% B in 2 min, 75–100% B in 1.5 min, 100% B for 3 min, 100–0% B in 0.5 min, 0% B for 2.5 min, with constant flow rate of 0.4 ml/min. A 10% isopropanol in water was used as a needle wash. MS analysis was conducted with MS1 mass range of 100–1700 m/z and MS2 range of 25–1000 m/z. MS parameters were set as follows; gas temperature at 325 °C, nebulizer at 45 psi, drying gas at 13 L/min, sheath gas temperature at 275 °C, sheath gas flow at 12 L/min, nozzle voltage at 500 V, fragmentor voltage at 175 V and skimmer voltage at 65 V. Capillary voltage was 4000 and 3000 V for positive and negative modes, respectively. Acquisition rate was 3.35 spectrum per s. Maximum 10 precursor ions per cycle with precursor threshold at 5,000 counts were selected for MS/MS fragmentation. Collision energy (CE) was 20 and 10 eV for positive and negative modes, respectively. Purine, trifluoroacetic acid ammonium salt and hexakis (1 H, 1 H, 3 H-tetrafluoropropoxy) phosphazine were used as reference masses. Data was collected in centroid mode<sup>53</sup>.

### Metabolite identification, data filtering, preprocessing, and statistical analysis

Raw data in Agilent.d format were converted to .abf files using the Reifycs Abf Converter and subsequently uploaded to MS-DIAL software (version 5.3)<sup>54</sup> for further processing. Peak detection, sample alignment, and compound identification were performed using default settings. Alignment was conducted against QC samples to ensure data consistency. Normalization was carried out using the LOWESS method, with sulfadimethoxine

(311.0809 m/z in positive mode and 309.0658 m/z in negative mode) applied as the internal standard for quality control. Mass exclusion was applied to remove features with m/z values of 121.0509 and 922.0098 in positive mode, and 112.9856, 119.0363, 966.0007, and 1033.9881 in negative mode. Reference masses were excluded from the analysis to ensure accuracy. Metabolite identification was conducted using the online ESI (+/-) MS/MS database derived from authentic standards, supplemented by searches against the Human Metabolome Database (HMDB). Detected adducts included M+H in positive ionization mode and M-H in negative ionization mode, with a retention time window of 0.2 to 18 min. Data were pre-filtered based on the following criteria: Pearson correlation coefficient  $\geq 0.70$  against 7 dilution QC samples, coefficient of variation (%CV) among QC samples  $\leq 50\%$ , and an identification score  $\geq 0.70^{55}$ . Features that did not meet these criteria were excluded from further analysis.

The raw metabolite profile data were normalized using log10 transformation and Pareto scaling prior to statistical analysis. The data were then split into training and test sets. Feature selection was exclusively performed on the training set to avoid data leakage and to ensure that the testing set remained unseen until the final evaluation. This process ensured the integrity of the model by preventing overfitting and maintaining an unbiased assessment. Multivariate analysis was performed using MetaboAnalyst 6.0<sup>56</sup> (<https://www.metaboanalyst.ca/>), including principal component analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA). Clustering heatmaps were generated to visualize differentially abundant metabolites. A volcano plot was constructed to illustrate fold changes and statistical significance, with *p*-values adjusted using the Benjamini–Hochberg false discovery rate (FDR) method (threshold: FDR  $< 0.05$ ). In addition, receiver operating characteristic (ROC) curves were plotted, and the area under curve (AUC) was calculated to evaluate the diagnostic performance of the selected metabolites. Feature selection for candidate metabolites were selected based on their statistical significance according to the following criteria: variable importance in projection (VIP)  $> 1.2$  from OPLS-DA analysis, fold change (FC)  $> 1.2$  or  $< 0.83$ , and FDR-adjusted *p*  $< 0.05$ , which were considered significant for the discriminatory model. Duplicate metabolites were removed, and only metabolites with higher AUC values from the ROC analysis were retained. Subsequently, the candidate metabolites that passed the feature selection criteria were further analyzed in the Support Vector Machine (SVM) model and pathway analysis.

### Support vector machine model for candidate metabolites

The Support Vector Machine (SVM) model was used to classify and predict potential biomarkers from the candidate metabolites identified in serum samples. The candidate metabolites were selected based on feature selection criteria according to VIP  $> 1.2$  from OPLS-DA analysis, FC  $> 1.2$  or  $< 0.83$ , and FDR-adjusted *p*  $< 0.05$ , which were considered significant for the discriminatory model. Duplicates were removed, and only metabolites with higher AUC values in ROC analysis. Using the MetaboAnalyst 6.0 platform, candidate metabolite data from the training set were first normalized to ensure uniform scaling. The SVM model was applied using a Radial Basis Function (RBF) kernel (default kernel in MetaboAnalyst), and cross-validation techniques such as 10-fold cross-validation were used to assess the model generalizability and to avoid overfitting. The default parameters for the SVM model included a cost value of 1 and gamma value of 1/n, where n is the number of features (default settings in MetaboAnalyst). The model performance was evaluated using classification metrics such as accuracy, sensitivity, specificity, and area under the curve (AUC) based on the training set<sup>18</sup>. The SVM model developed from the training set was then validated using an independent testing set, and model performance was further assessed through the same classification metrics (accuracy, sensitivity, specificity, and AUC) derived from the testing set.

### Pathway analysis

Pathway enrichment and pathway impact analysis were performed using MetaboAnalyst 6.0<sup>56</sup> (<https://www.metaboanalyst.ca/>), with KEGG pathways used as a reference for the analysis<sup>57</sup>. To ensure a focused and accurate interpretation, significantly altered candidate metabolites were categorized into two groups: common (non-lipid) metabolites and lipid metabolites, which were analyzed separately. The candidate metabolite lists were uploaded to the platform, and their corresponding Human Metabolome Database (HMDB) IDs were mapped to reference pathways. The pathway analysis module in MetaboAnalyst 6.0 utilizes a hypergeometric test to evaluate the overrepresentation of metabolites within specific pathways and a relative-betweenness centrality measure to assess the topological importance of pathways. To address multiple testing, False Discovery Rate (FDR) correction was applied. We established research-specific criteria to effectively elucidate the biological phenomena associated with cancer progression and recurrence for this study. Pathways were deemed statistically significant if they met either of the following thresholds: (1) *p*  $< 0.05$  and FDR  $< 0.2$ , or (2) *p*  $< 0.05$ , FDR  $< 0.3$ , and an impact score  $> 0.2$ . The results were visualized using interactive pathway maps, highlighting key metabolic alterations in both common and lipid metabolites.

### Bioinformatics analysis

The overall survival (OS) was calculated using the Kaplan-Meier (KM) method, where disease-free survival (DFS) was defined as the time from surgery to recurrence, and overall survival was defined as the time from surgery to death. Patients who survived beyond the study period had their median DFS and OS calculated, and comparisons between groups were analyzed using the Log-rank test. A *p*-value of  $< 0.05$  was considered statistically significant. All analyses were conducted using IBM SPSS Statistics version 26.

## Data availability

The datasets generated during this study are available in the ProteomeXchange repository under the accession IDs JPST003566 and PXD060214 (via <https://repository.jpostdb.org/entry/JPST003566.0>). Clinical data supporting the findings of this study are available from the corresponding author upon reasonable request.

Received: 4 February 2025; Accepted: 7 April 2025

Published online: 14 April 2025

## References

1. Banales, J. M. et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 557–588. <https://doi.org/10.1038/s41575-020-0310-z> (2020).
2. Nassar, A. et al. Factors of early recurrence after resection for intrahepatic cholangiocarcinoma. *World J. Surg.* **46**, 2459–2467. <https://doi.org/10.1007/s00268-022-06655-1> (2022).
3. Yamauchi, N. et al. Clinical significance of early recurrence after curative resection of colorectal cancer. *Anticancer Res.* **42**, 5553–5559. (2022).
4. Groot, V. P. et al. Defining and predicting early recurrence in 957 patients with resected pancreatic ductal adenocarcinoma. *Ann. Surg.* **269**, 1154–1162. <https://doi.org/10.1097/sla.0000000000002734> (2019).
5. Yamauchi, N. et al. Clinical significance of early recurrence after curative resection of colorectal cancer. *Anticancer Res.* **42**, 5553–5559. <https://doi.org/10.21873/anticanres.16061> (2022).
6. Kim, H. S. et al. Serum carcinoembryonic antigen and carbohydrate antigen 19-9 as preoperative diagnostic biomarkers of extrahepatic bile duct cancer. *BJS Open.* **5** <https://doi.org/10.1093/bjsopen/zrab127> (2021).
7. Kang, J. S. et al. Limits of serum carcinoembryonic antigen and carbohydrate antigen 19-9 as the diagnosis of gallbladder cancer. *Ann. Surg. Treat. Res.* **101**, 266–273. <https://doi.org/10.4174/astr.2021.101.5.266> (2021).
8. Zhang, A., Sun, H., Yan, G., Wang, P. & Wang, X. Mass spectrometry-based metabolomics: applications to biomarker and metabolic pathway research. *Biomed. Chromatogr.* **30**, 7–12 (2016).
9. Jacob, M., Lopata, A. L., Dasouki, M. & Abdel Rahman, A. M. Metabolomics toward personalized medicine. *Mass Spectrom. Rev.* **38**, 221–238 (2019).
10. Bi, Y. et al. Glycerophospholipid-driven lipid metabolic reprogramming as a common key mechanism in the progression of human primary hepatocellular carcinoma and cholangiocarcinoma. **23**, 326 (2024).
11. Silva, A. A. R. et al. Plasma metabolome signatures to predict responsiveness to neoadjuvant chemotherapy in breast cancer. **16**, 2473 (2024).
12. Xie, T. et al. Blood metabolomic profiling predicts postoperative gastrointestinal function of colorectal surgical patients under the guidance of goal-directed fluid therapy. **13**, 8929 (2021).
13. Qu, N. et al. Integrated proteogenomic and metabolomic characterization of papillary thyroid cancer with different recurrence risks. *Nat. Commun.* **15**, 3175. <https://doi.org/10.1038/s41467-024-47581-1> (2024).
14. Kaji, S. et al. Metabolomic profiling of gastric cancer tissues identified potential biomarkers for predicting peritoneal recurrence. *Gastric Cancer* **23**, 874–883. <https://doi.org/10.1007/s10120-020-01065-5> (2020).
15. Wada, Y. et al. Tumor metabolic alterations after neoadjuvant chemoradiotherapy predict postoperative recurrence in patients with pancreatic cancer. *Jpn J. Clin. Oncol.* **52**, 887–895. <https://doi.org/10.1093/jco/hyac074> (2022).
16. Ahmed-Salim, Y. et al. The application of metabolomics in ovarian cancer management: a systematic review. *Int. J. Gynecol. Cancer.* **31**, 754–774. <https://doi.org/10.1136/ijgc-2020-001862> (2021).
17. Padthaisong, S. et al. Integration of global metabolomics and lipidomics approaches reveals the molecular mechanisms and the potential biomarkers for postoperative recurrence in early-stage cholangiocarcinoma. *Cancer Metab.* **9** <https://doi.org/10.1186/s40170-021-00266-5> (2021).
18. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297. <https://doi.org/10.1007/BF00994018> (1995).
19. Mahadevan, S., Shah, S. L., Marrie, T. J. & Slupsky, C. M. Analysis of metabolomic data using support vector machines. *Anal. Chem.* **80**, 7562–7570. <https://doi.org/10.1021/ac800954c> (2008).
20. Xie, T. et al. Blood metabolomic profiling predicts postoperative gastrointestinal function of colorectal surgical patients under the guidance of goal-directed fluid therapy. *Aging* **13** <https://doi.org/10.18632/aging.202711> (2021).
21. Szczerbinski, L. et al. Untargeted metabolomics analysis of the serum metabolic signature of childhood obesity. *Nutrients* **14** <https://doi.org/10.3390/nu14010214> (2022).
22. Szczerbinski, L. et al. Untargeted metabolomics analysis of the serum metabolic signature of childhood obesity. **14**, 214 (2022).
23. Chen, Z., Huang, X., Gao, Y., Zeng, S. & Mao, W. Plasma-metabolite-based machine learning is A.Pomising diagnostic A.proach for esophageal squamous cell carcinoma investigation. *J. Pharm. Anal.* **11**, 505–514 (2021).
24. Gaul, D. A. et al. Highly-accurate metabolomic detection of early-stage ovarian cancer. *Sci. Rep.* **5**, 16351. <https://doi.org/10.1038/srep16351> (2015).
25. Lieu, E. L., Nguyen, T., Rhyne, S. & Kim, J. Amino acids in cancer. *Exp. Mol. Med.* **52**, 15–30. <https://doi.org/10.1038/s12276-020-0375-3> (2020).
26. Chen, J., Cui, L., Lu, S. & Xu, S. Amino acid metabolism in tumor biology and therapy. *Cell Death Dis.* **15** <https://doi.org/10.1038/s41419-024-06435-w> (2024).
27. Abd El-Fattah, E. E. IDO/kynurenone pathway in cancer: possible therapeutic approaches. *J. Translational Med.* **20**, 347. <https://doi.org/10.1186/s12967-022-03554-w> (2022).
28. Lo, M., Ling, V., Wang, Y. Z. & Gout, P. W. The xc- cystine/glutamate antiporter: a mediator of pancreatic cancer growth with a role in drug resistance. *Br. J. Cancer.* **99**, 464–472. <https://doi.org/10.1038/sj.bjc.6604485> (2008).
29. Fan, J. et al. Quantitative flux analysis reveals folate-dependent NADPH production. *Nature* **510**, 298–302. <https://doi.org/10.1038/nature13236> (2014).
30. Son, J. et al. Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature* **496**, 101–105. <https://doi.org/10.1038/nature12040> (2013).
31. De Marchi, T. et al. Phosphoserine aminotransferase 1 is associated to poor outcome on Tamoxifen therapy in recurrent breast cancer. *Sci. Rep.* **7**, 2099. <https://doi.org/10.1038/s41598-017-02296-w> (2017).
32. Jia, Y. et al. Low expression of Bin1, along with high expression of IDO in tumor tissue and draining lymph nodes, are predictors of poor prognosis for esophageal squamous cell cancer patients. *Int. J. Cancer.* **137**, 1095–1106. <https://doi.org/10.1002/ijc.29481> (2015).
33. Damerell, V. et al. Circulating tryptophan–kynurenone pathway metabolites are associated with all-cause mortality among patients with stage I–III colorectal cancer. **156**, 552–565 <https://doi.org/10.1002/ijc.35183> (2025).
34. Tintelnot, J. et al. Microbiota-derived 3-IAA influences chemotherapy efficacy in pancreatic cancer. *Nature* **615**, 168–174. <https://doi.org/10.1038/s41586-023-05728-y> (2023).
35. Ichisaka, Y., Yano, S., Nishimura, K., Niwa, T. & Shimizu, H. Indoxyl sulfate contributes to colorectal cancer cell proliferation and increased EGFR expression by activating AhR and Akt. *Biomedical Res. (Tokyo Japan)* **45**, 57–66. <https://doi.org/10.2220/biomed.res.45.57> (2024).

36. Koundouros, N. & Poulogiannis, G. Reprogramming of fatty acid metabolism in cancer. *Br. J. Cancer.* **122**, 4–22 (2020).
37. Abdelzaher, E. & Mostafa, M. F. Lysophosphatidylcholine acyltransferase 1 (LPCAT1) upregulation in breast carcinoma contributes to tumor progression and predicts early tumor recurrence. *Tumour Biology: J. Int. Soc. Oncodevelopmental Biology Med.* **36**, 5473–5483. <https://doi.org/10.1007/s13277-015-3214-8> (2015).
38. Morita, Y. et al. Lysophosphatidylcholine acyltransferase 1 altered phospholipid composition and regulated hepatoma progression. *J. Hepatol.* **59**, 292–299. <https://doi.org/10.1016/j.jhep.2013.02.030> (2013).
39. Tomacha, J. et al. Targeting fatty acid synthase modulates metabolic pathways and inhibits cholangiocarcinoma cell progression. *Front. Pharmacol.* **12**, 696961. <https://doi.org/10.3389/fphar.2021.696961> (2021).
40. Sunshine, H. & Iruela-Arispe, M. L. Membrane lipids and cell signaling. *Curr. Opin. Lipidol.* **28**, 408–413 (2017).
41. Molendijk, J., Robinson, H., Djuric, Z. & Hill, M. M. Lipid mechanisms in hallmarks of cancer. *Mol. Omics.* **16**, 6–18. <https://doi.org/10.1039/c9mo00128j> (2020).
42. Wolrab, D., Jirásko, R., Chocholoušková, M., Peterka, O. & Holčapek, M. Oncolipidomics: Mass spectrometric quantitation of lipids in cancer research. *TrAC Trends Anal. Chem.* **120**, 115480 (2019).
43. Nizioł, J. et al. Untargeted ultra-high-resolution mass spectrometry metabolomic profiling of blood serum in bladder cancer. *Sci. Rep.* **12**, 15156. <https://doi.org/10.1038/s41598-022-19576-9> (2022).
44. Yagi, T. et al. Challenges and inconsistencies in using lysophosphatidic acid as a biomarker for ovarian cancer. *Cancers* **11** <https://doi.org/10.3390/cancers11040520> (2019).
45. Lu, Y. et al. Comparison of hepatic and serum lipid signatures in hepatocellular carcinoma patients leads to the discovery of diagnostic and prognostic biomarkers. *Oncotarget* **9**, 5032–5043. <https://doi.org/10.18632/oncotarget.23494> (2018).
46. Ravipati, S., Baldwin, D. R., Barr, H. L., Fogarty, A. W. & Barrett, D. A. Plasma lipid biomarker signatures in squamous carcinoma and adenocarcinoma lung cancer patients. *Metabolomics* **11**, 1600–1611 (2015).
47. Martin-Perez, M., Urdiroz-Urricelqui, U., Bigas, C. & Benitah, S. A. The role of lipids in cancer progression and metastasis. *Cell Metab.* **34**, 1675–1699 (2022).
48. Westheim, A. J. F. et al. The modulatory effects of fatty acids on cancer progression. *Biomedicines* **11** <https://doi.org/10.3390/biomedicines11020280> (2023).
49. Zhao, G. et al. Ovarian cancer cell fate regulation by the dynamics between saturated and unsaturated fatty acids. *119*, e2203480119 (2022).
50. Zhao, H. et al. Inflammation and tumor progression: signaling pathways and targeted intervention. *Signal. Transduct. Target. Therapy.* **6**, 263. <https://doi.org/10.1038/s41392-021-00658-5> (2021).
51. Amin, M. B. et al. The eighth edition AJCC cancer staging manual: continuing to build a Bridge from a population-based to a more personalized approach to cancer staging. *Cancer J. Clin.* **67**, 93–99. <https://doi.org/10.3322/caac.21388> (2017).
52. Lewis, M. R. et al. Development and application of ultra-performance liquid chromatography-TOF MS for precision large scale urinary metabolic phenotyping. *Anal. Chem.* **88**, 9004–9013. <https://doi.org/10.1021/acs.analchem.6b01481> (2016).
53. Wu, Q. et al. UPLC-Q-TOF/MS based metabolomic profiling of serum and urine of hyperlipidemic rats induced by high fat diet. *J. Pharm. Anal.* **4**, 360–367. <https://doi.org/10.1016/j.jpha.2014.04.002> (2014).
54. Tsugawa, H. et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods.* **12**, 523–526. <https://doi.org/10.1038/nmeth.3393> (2015).
55. Saw, N. et al. Influence of extraction solvent on nontargeted metabolomics analysis of enrichment reactor cultures performing enhanced biological phosphorus removal (EBPR). *Metabolites* **11** <https://doi.org/10.3390/metabol11050269> (2021).
56. Pang, Z. et al. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucl. Acids Res.* **52**, W398–W406. (2024).
57. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucl. Acids Res.* **51**, D587–D592. (2022).

## Acknowledgements

All authors are truly thankful for helpful discussions and guidance from Prof. Narong Khuntikeo at Department of Surgery, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand, Cholangiocarcinoma Research Institute (CARI), Khon Kaen University, Khon Kaen, Thailand and Cholangiocarcinoma Screening and Care Program (KKU), Khon Kaen University, Khon Kaen, Thailand. We are also indebted to all members of CASCAP, particularly the cohort members, and researchers at CARI, Faculty of Medicine, Khon Kaen University for collecting and proofing of CCA patient data. In addition, we also thank Professor Ross H. Andrew for editing the MS via the Publication Clinic KKU, Thailand.

## Author contributions

Conceptualization—P.P., A.T., V.T., and W.L.; Methodology— P.P., A.T., V.T., A.J., N.K., N.N., P.K., A.W., J.C., S.K., P.S., N.M. and W.L.; Formal analysis— P.P., A.T., V.T. and W.L.; Investigation— P.P., A.T., V.T. and W.L.; Resources— A.T., S.R. and W.L.; Data curation—P.P., A.T., V.T. and W.L.; Visualization—P.P., V.T. and N.M.; Supervision— A.T., S.R. and W.L.; Project administration— W.L.; Funding acquisition— W.L.; Original draft preparation— P.P. and W.L.; Draft review and editing—all authors.

## Funding

This work was supported by the National Science Research and Innovation Fund (NSRF) through Khon Kaen University, the National Research Council of Thailand through the Hub of Knowledge Grant and Srinagarind Diamond Research Fund (grant no. DR63201) to WL.

## Declarations

### 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/s41598-025-97641-9>.

**Correspondence** and requests for materials should be addressed to W.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

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

© The Author(s) 2025