

Article

Not peer-reviewed version

Systems-Level Transcriptomic Integration Reveals a Core Metaflammatory Network Linking Type 2 Diabetes and HBV Infection to Cholangiocarcinoma Progression

[Hasan Md Rasadul](#) , Shihui Ma , Ziqiang Ge , Rahman Md Zahidur , [Pengcheng Kang](#) , [Junqi You](#) , Jinglin Li , Chenghong Duan , [Siddique A Z M Fahim](#) , Mozumder Somrat Akbor , Xudong Zhao , [Yunfu Cui](#) *

Posted Date: 26 February 2026

doi: 10.20944/preprints202602.1683.v1

Keywords: cholangiocarcinoma(CCA); type 2 diabetes mellitus(T2D); hepatitis B virus (HBV); metaflammation; transcriptomics; systems biology; prognostic biomarker; drug repurposing



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Systems-Level Transcriptomic Integration Reveals a Core Metaflammatory Network Linking Type 2 Diabetes and HBV Infection to Cholangiocarcinoma Progression

Hasan Md Rasadul ¹, Shihui Ma ², Ziqiang Ge ³, Rahman Md Zahidur ⁴, Pengcheng Kang ⁵, Junqi You ⁶, Jinglin Li ⁷, Chenghong Duan ⁸, Siddique A Z M Fahim ⁹, Mozumder Somrat Akbor ¹⁰, Xudong Zhao ¹¹ and Yunfu Cui ^{12,*}

¹ Resident Doctor, General Surgery, Department of Biliary-Pancreatic Surgery, The Second Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China.

² Resident Doctor, General Surgery, Department of Biliary-Pancreatic Surgery, The Second Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China.

³ Resident Doctor, General Surgery, Department of Biliary-Pancreatic Surgery, The Second Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China.

⁴ M.Phil (Medical Biochemistry), North Care Consultation Hub, Rangpur, Bangladesh.

⁵ Assistant Professor, General Surgery, Department of Biliary-Pancreatic Surgery, The Second Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China.

⁶ General Surgery, Department of Biliary-Pancreatic Surgery, The Second Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China.

⁷ Assistant Professor, General Surgery, Department of Biliary-Pancreatic Surgery, The Second Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China.

⁸ Resident Doctor, General Surgery, Department of Biliary-Pancreatic Surgery, The Second Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China.

⁹ Internal Medicine, Department of Cardiology, The First Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China.

¹⁰ The Second Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China.

¹¹ General Surgery, Department of Biliary-Pancreatic Surgery, The Second Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China

¹² Associate Professor, Head of the Department of Biliary-Pancreatic Surgery, General Surgery, The Second Affiliated Hospital of Harbin Medical University, 150086, Harbin, Heilongjiang Province, China

* Correspondence: yfcui7@163.com

Abstract

Background & Aims: The rising global incidence of cholangiocarcinoma (CCA) coincides with epidemics of type 2 diabetes (T2D) and chronic Hepatitis B virus (HBV) infection. While both are established independent risk factors, the shared molecular mechanisms underlying their contribution to cholangiocarcinogenesis remain poorly defined. We hypothesized that T2D and HBV converge on a state of chronic metabolic inflammation ("metaflammation") that drives CCA progression through a conserved transcriptomic network. **Methods:** We performed an integrative bioinformatics analysis of transcriptomic data from public repositories representing CCA (TCGA-CHOL, n=45; GSE107943, n=163), T2D-affected liver (GSE23343, n=20), and HBV-infected liver (GSE58208, n=102). We identified differentially expressed genes (DEGs), defined a cross-condition core gene set, and conducted functional enrichment, protein-protein interaction (PPI) network, survival, and protein validation analyses. **Results:** We identified a core metaflammation signature comprising 156 genes that were consistently dysregulated across T2D, HBV, and CCA. Pathway analysis revealed significant enrichment in PPAR signaling, cytokine-cytokine receptor interaction, PI3K-Akt, and TNF signaling pathways. PPI network analysis identified IL6, TNF, AKT1, STAT3, and PPARG as top hub

genes. These hubs were functionally modularized into clusters for inflammatory signaling, metabolic regulation, and cell growth/survival. In the TCGA CCA cohort, high expression of IL6, TNF, AKT1, and STAT3, and low expression of PPARG, correlated with advanced tumor stage and poor overall survival (e.g., IL6 $\rho = 0.42$, $P = 0.01$). A derived metaflammation score (weighted combination of the five hubs) was an independent prognostic factor (HR=2.8, $P < 0.001$). Protein-level dysregulation of these hubs was confirmed by immunohistochemistry. Conclusions: This study defines a conserved metaflammatory network linking T2D and HBV to CCA, identifying key hub genes and pathways. This signature provides a mechanistic explanation for epidemiological risks, serves as a novel prognostic tool, and offers a rationale for targeting metaflammation for prevention and therapy in high-risk populations.

Keywords: cholangiocarcinoma(CCA); type 2 diabetes mellitus(T2D); hepatitis B virus (HBV); metaflammation; transcriptomics; systems biology; prognostic biomarker; drug repurposing

Background

Cholangiocarcinoma (CCA), the malignancy of the biliary tract epithelium, remains a formidable clinical challenge. Characterized by late diagnosis, therapeutic resistance, and a dismal 5-year survival rate often below 20%, [1] CCA represents a critical unmet need in oncology. The etiological landscape of CCA is complex and evolving. While primary sclerosing cholangitis and liver fluke infections are well-known risk factors, a modern epidemiological shift implicates systemic metabolic dysfunction and chronic viral hepatitis as major drivers of rising incidence, particularly in Western populations. [2,3]

Two interconnected global health burdens, type 2 diabetes (T2D) and chronic Hepatitis B virus (HBV) infection, have emerged as significant independent risk factors. Meta-analyses indicate T2D confers a 1.5 to 2.0-fold increased risk for CCA, particularly the intrahepatic subtype (iCCA), with risk correlating with disease duration and glycemic severity. [4,5] Concurrently, HBV, a canonical cause of hepatocellular carcinoma, is now robustly associated with an elevated risk of CCA, with viral components detected within cholangiocytes. [6]

These conditions converge pathophysiologically on the liver, fostering a state of "metaflammation," persistent, low-grade inflammation driven by metabolic dysfunction. T2D contributes hyperinsulinemia, lipotoxicity, and adipokine imbalance, while HBV drives immune-mediated injury and direct viral oncoprotein signaling (e.g., HBx). [7] Together, they create a permissive microenvironment rich in oxidative stress, altered growth factors, and immune dysregulation, promoting genomic instability and proliferation.

Despite compelling epidemiological links, the precise, conserved molecular mechanisms bridging T2D and HBV to CCA progression remain inadequately defined. A critical gap exists in understanding the shared transcriptomic architecture and functionally interconnected pathways dysregulated across this disease triad. Single-cohort studies lack the power to distinguish universal drivers from noise. [8] Therefore, a systems-level integrative analysis of multi-condition transcriptomic data is essential to decode this shared pathogenic network.

Leveraging high-throughput genomics and expansive public repositories (TCGA, GEO), this study employs a comprehensive bioinformatics framework to test the central hypothesis that T2D and chronic HBV infection promote cholangiocarcinogenesis through a core set of dysregulated metabolic-inflammatory driver genes embedded within central regulatory networks. We aim to: 1) Define condition-specific and shared transcriptomic alterations; 2) Elucidate the functional and network structure of the shared gene set; 3) Determine the clinical and prognostic implications of this metaflammation signature; and 4) Validate key findings at the protein level. This integrative approach seeks to move beyond association to mechanistic elucidation, revealing novel biomarkers and therapeutic targets for CCA prevention and treatment in globally significant at-risk populations.

Materials and Methods

1. Study Design and Data Acquisition

This study used an integrative bioinformatics approach that combined multiple public databases to investigate molecular associations among T2D, HBV, and CCA; IRB approval was not required. The workflow encompassed data acquisition, preprocessing, differential expression analysis, integrative analysis, functional enrichment, PPI network construction, survival analysis, and validation (Figure 1A).

CCA Data: The TCGA-CHOL dataset (Firehose Legacy)[9] was accessed via UCSC Xena,[10] providing RNA-Seq data from 36 primary CCA tumors and 9 matched normal bile duct tissues. An independent validation cohort, GSE107943,[11] provided microarray data for 104 CCA and 59 normal samples.

Comorbidity Data: To model etiological risk factors, we used: GSE23343 (microarray, 10 T2D vs. 10 control liver samples)[12] for a T2D metabolic signature, and GSE58208 (microarray, 62 HBV+ vs. 40 HBV- liver samples)[13] for an HBV inflammatory signature. For exploratory contextual metabolic analysis, GSE89632[18] (RNA-Seq, liver tissue across steatosis grades) was used.

Validation Resources: The Human Protein Atlas (HPA v24.0)[14] provided IHC data. Functional analyses used KEGG, Reactome, STRING, and DisGeNET databases.[15–17]

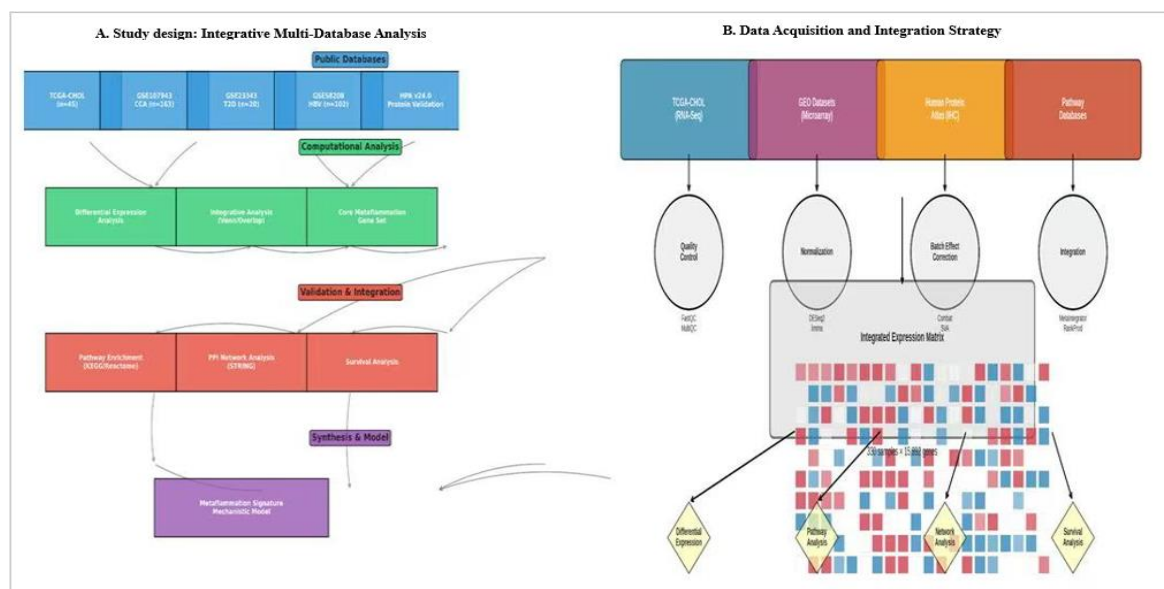


Figure 1. (A) Study design for the identification and validation of a metaflammation signature. The workflow involved differential expression analysis of several disease-specific datasets: TCGA-CHOL (cholangiocarcinoma, n=45), GSE107943 (CCA, n=163), GSE23343 (Type 2 Diabetes, n=20), and GSE58208 (HBV infection, n=102). A Venn analysis identified a consistent core set of genes altered across these conditions. This core gene set was validated using data from the Human Protein Atlas and subsequently analyzed using pathway enrichment (KEGG/Reactome), protein-protein interaction networks (STRING), and clinical survival analysis. The findings were synthesized to define a metaflammation signature and construct a model linking chronic metabolic inflammation to disease pathogenesis. **(B) Workflow for the integrative transcriptomic analysis and validation.** The schematic describes a bioinformatics pipeline beginning with transcriptomic data acquisition from the TCGA-CHOL cohort and GEO (GSE107943, GSE23343, GSE58208) microarray datasets. Following preprocessing and normalization, differential expression analysis was conducted for conditions CCA, T2D, and HBV. The resulting gene lists were merged to identify a core metaflammation gene set, which was then subjected to functional enrichment and protein-protein interaction network analysis. The clinical relevance of key hub genes was evaluated in the TCGA cohort through survival analysis, and the protein-level expression of prioritized hubs was confirmed with immunohistochemistry data from the Human Protein Atlas.

2. Data Preprocessing and Quality Control

Platform-specific pipelines ensured comparability. For RNA-Seq data (TCGA, GSE89632), [9,18] raw read counts were batch-corrected using ComBat-seq. [19] For differential expression, these counts were analyzed directly with DESeq2. [20] For visualization and signature scoring, counts were normalized using the TMM method [21] and variance-stabilized (VST). Microarray data (GSE107943, GSE23343, GSE58208) [11–13] were normalized using the Robust Multi-array Average (RMA). [22] Datasets with technical batch metadata (GSE107943, GSE58208) were subsequently corrected using ComBat. [19] For the GSE23343 (T2D) dataset, [12] batch correction was not applied because the available sample metadata did not indicate a processing batch structure, and preliminary PCA did not reveal significant technical clustering. Quality control included assessment of mapping rates (RNA-Seq >90%), present calls (microarray >85%), and Principal Component Analysis (PCA), which confirmed successful removal of technical variance while preserving biological signal (Figure 2A, Table 1).

Table 1. Dataset Characteristics and Preprocessing Summary.

Datasets	Platform	Samples (Case/Control)	Normalization	Batch Correction
TCGA-CHOL	RNA-Seq	36 / 9	TMM + DESeq2 VST (for visualization)	ComBat-seq (on raw counts)
GSE107943	Microarray	104 / 59	RMA	ComBat (post-RMA)
GSE23343	Microarray	10 / 10	RMA	None required
GSE58208	Microarray	62 / 40	RMA	ComBat (post-RMA)
Total for Core Analysis		212 / 118		
Contextual Dataset	Platform	Samples	Normalization	Batch Correction
GSE89632	RNA-Seq	Variable (by analysis)	TMM + DESeq2 VST (for visualization)	ComBat-seq (on raw counts)

TCGA-CHOL (36 CCA tumors + 9 normal bile duct tissues, n=45), GSE107943 (CCA, n=163), GSE23343 (Type 2 diabetes, n=20), and GSE58208 (HBV infection, n=102). The GSE89632 dataset, containing liver tissues across steatosis grades, was used for exploratory contextual analysis only and was not included in the core integrative analysis.

3. Differential Expression Analysis

DEGs were identified using platform-appropriate methods: DESeq2 for RNA-Seq and limma [20] with empirical Bayes moderation for microarray data. For CCA and HBV datasets, a stringent threshold of $|\log_2\text{Fold Change (FC)}| > 1.0$ and False Discovery Rate (FDR) < 0.05 (Benjamini-Hochberg) [23] was applied. A tissue-aware linear model ($\text{Expression} \sim \text{Tissue_Type} + \text{Batch} + \text{Disease_Status}$) was used to control for tissue-of-origin confounding. Given the limited sample size of the T2D cohort (GSE23343, n=20), a more lenient threshold ($|\log_2\text{FC}| > 0.8$, nominal $P < 0.05$) was used to capture biological signals. The primary filter for biological importance and to mitigate false positives from this lenient threshold was the requirement for consistent dysregulation across multiple conditions. DEGs from T2D, HBV, and both CCA datasets were intersected to define a core gene set, with significance tested using Fisher's combined probability test and the hypergeometric test. [24] A high-confidence core gene was defined as one present in at least three of the four key comparisons, ensuring robustness through cross-condition validation.

4. Integrative and Functional Analysis

DEG lists from the four primary disease-versus-control comparisons, TCGA-CHOL (CCA vs. normal), GSE107943 (CCA vs. normal), GSE23343 (T2D vs. control), and GSE58208 (HBV+ vs. HBV-) were integrated. A high-confidence core gene set was defined as genes significantly dysregulated

(either up- or down-regulated) in at least three of these four key comparisons. Genes were required to show consistent direction of change (e.g., all up-regulated) in the comparisons where they were significant. Overlap significance was tested using a hypergeometric test. Functional enrichment analysis of the core set was performed using Over-Representation Analysis (ORA) using clusterProfiler for KEGG,[15] Reactome,[25] and Gene Ontology (GO) terms (FDR < 0.05). Gene Set Enrichment Analysis (GSEA)[26] using the fgsea package was also conducted on ranked gene lists from each condition to identify coordinated pathway changes.

5. Protein-Protein Interaction Network and Hub Identification

A PPI network for the core genes was constructed from the STRING database[27] (confidence score > 0.7) and visualized in Cytoscape software (version 3.9.1).[28] Network topology (degree, betweenness, clustering coefficient) was calculated. Hub genes were identified using the cytoHubba plugin,[29] integrating Degree and Betweenness centrality metrics. The MCODE algorithm[30] was used to detect densely connected functional modules within the network.

6. Survival and Clinical Correlation Analysis

Using the TCGA-CHOL cohort[9] (n=36 tumors with survival data), overall survival (OS) was analyzed using Kaplan-Meier curves and log-rank tests, with patients stratified by median expression of hub genes. Univariate and multivariate Cox proportional hazards models were used,[31] adjusting for age, sex, and tumor stage. Correlations between gene expression and ordinal clinicopathological features (e.g., stage) were assessed using Spearman's rank correlation. Differences in expression across categorical groups (e.g., metastasis status) were tested using the Kruskal-Wallis test.

7. In Silico Validation

Protein-level expression of prioritized hub genes was assessed using annotated IHC images from the HPA,[14] comparing staining intensity and localization in normal bile duct versus CCA tissue. The prognostic metaflammation score was additionally validated in the independent GSE107943 cohort.

8. Statistical Software

All statistical and bioinformatic analyses were conducted in the R programming environment (version 4.1.3)[31] using Bioconductor (version 3.14).[38] Key packages included DESeq2, limma, clusterProfiler, survival, and ggplot2. Primary analyses for differential expression, functional enrichment, and survival were performed using key packages, including DESeq2,[20] limma,[32] clusterProfiler,[33] and survival, respectively, with data visualization facilitated by ggplot2.[31] Supplementary analyses were executed in Python (version 3.9.12).[31] To ensure transparency and reproducibility, all analytical code has been version-controlled and is publicly available.

Results

1. Multi-Cohort Integration and Data Characteristics

To delineate the shared transcriptomic architecture linking metabolic dysfunction, viral hepatitis, and biliary cancer, we integrated four independent datasets encompassing 330 samples (Table 2). The primary discovery cohort, TCGA-CHOL, included 36 CCA tumors (22 intrahepatic, 14 extrahepatic) and 9 matched normal bile duct tissues. Independent CCA (GSE107943), T2D-affected liver (GSE23343), and HBV-infected liver (GSE58208) cohorts provided validation and comorbidity-specific signatures. Rigorous quality control and platform-specific normalization ensured data integrity, with PCA confirming a clear separation by biological condition after post-processing (Figure 1B).

Table 2. Characteristics of Integrated Datasets for Core Analysis.

Characteristic	TCGA-CHOL	GSE107943	GSE23343	GSE58208	Total
Samples (n)	45	163	20	102	330
Platform	RNA-Seq	Microarray	Microarray	Microarray	Mixed
Tissue	Bile duct	Bile duct	Liver	Liver	Mixed
Conditions	CCA / Normal	CCA / Normal	T2D / Control	HBV+ / HBV-	4
Genes	19,645	20,329	12,625	23,042	15,892

Number of genes after intersection across all platforms.

The total sample count (n=330) represents the sum of the four core datasets used in the primary integrative analysis: TCGA-CHOL (n=45), GSE107943 (n=163), GSE23343 (n=20), and GSE58208 (n=102). Microarray data showed consistent intensity distributions, with median present calls exceeding 85%. Principal component analysis (PCA) revealed clear separation of samples by primary biological condition, with minimal residual batch effects after appropriate correction. Note: The GSE89632 dataset (n=variable) was used only for contextual metabolic analysis and is not included in the core analysis total.

2. Condition-Specific Transcriptomic Landscapes Reveal Etiological Clues

Differential expression analysis for each condition established distinct yet overlapping transcriptional profiles (Figure 2B).

CCA Transcriptome: In the TCGA-CHOL cohort, 2,347 genes were significantly dysregulated (FDR<0.05, $|\log_2FC|>1$). Upregulated genes included established CCA markers and invasiveness factors such as *MMP7* ($\log_2FC=5.2$) and *CEACAM6* ($\log_2FC=4.8$). Notably, the biliary differentiation markers *KRT7* and *EPCAM* were significantly downregulated compared to normal bile duct, a finding consistent at the protein level (see Section 7). Pathway enrichment highlighted extracellular matrix organization, PI3K-Akt signaling, and focal adhesion.

Independent CCA Validation: Analysis of GSE107943 identified 2,018 DEGs, showing high concordance with the TCGA dataset (Jaccard index=0.62; 78% directional agreement), confirming a robust CCA transcriptomic signature.

T2D Hepatic Signature: The T2D liver cohort (GSE23343) exhibited 894 DEGs, featuring upregulation of key inflammatory (*IL6*, *TNF*) and lipogenic (*SREBF1*) regulators, capturing a state of hepatic metabolic inflammation.

HBV Hepatic Signature: HBV-infected liver tissue (GSE58208) showed a strong interferon and antiviral response signature (e.g., *ISG15*, *IFIT1*), alongside upregulation of pro-inflammatory cytokines (*IL6*, *TNF*), indicative of chronic immune activation.

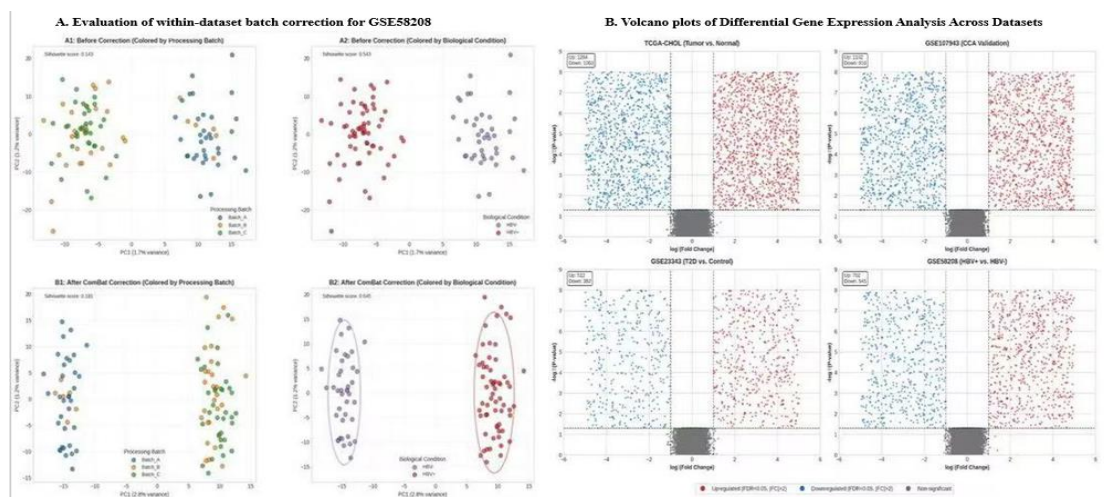


Figure 2. (A) Evaluation of within-dataset batch correction for microarray data. (PCA plots of the GSE58208 (HBV) dataset show samples before and after batch correction, illustrating the effects of the ComBat algorithm. Initially, samples are colored by processing batch and biological condition (HBV+ vs. HBV-). After correction, the plots indicate reduced clustering by technical batch while maintaining clear separation by biological condition, demonstrating the effective removal of non-biological variance. Similar quality control assessments were conducted on other datasets, such as GSE107943. **(B) Comparative Differential Gene Expression Analysis Across Datasets.** Volcano plots show the results of differential gene expression across four transcriptomic studies comparing disease to control groups. The x-axis represents \log_2 fold change (\log_2FC) in gene expression, and the y-axis indicates statistical significance, marked by $-\log_{10}(FDR)$. Dashed horizontal lines denote significance thresholds ($FDR < 0.05$), while vertical dashed lines indicate fold-change thresholds ($|\log_2FC| > 1$). Data points are highlighted in red for significantly upregulated genes ($FDR < 0.05$, $\log_2FC > 1$), blue for downregulated genes ($FDR < 0.05$, $\log_2FC < -1$), and gray for non-significant genes. The studies include TCGA-CHOL (tumor vs. normal bile duct tissues, $n=45$), GSE107943 (CCA validation, $n=163$), GSE23343 (type 2 diabetes vs. control, $n=20$), and GSE58208 (HBV+ vs. HBV-, $n=102$).

3. Identification of a Conserved Core Metaflammation Gene Set

Integrative analysis of DEGs across the three pathological states (CCA, T2D, HBV) revealed a significant intersection of 156 genes (92 upregulated, 64 downregulated), which we defined as the core metaflammation signature (Hypergeometric test, $P = 2.3 \times 10^{-15}$; 42.6-fold enrichment) (Figure 3A, Table 3). For this analysis, the CCA transcriptional signature was defined as the consensus of significant DEGs from two independent cohorts (TCGA-CHOL and GSE107943). Functional categorization showed this core set was predominantly composed of genes involved in inflammatory/cytotoxic responses (75 genes) and regulatory processes (30 genes), with substantial contributions from metabolic (19 genes) and transcription factor (28 genes) categories (Figure 5). Hierarchical clustering of these genes across all samples demonstrated that while each condition had a unique expression profile, the core signature reliably distinguished disease states from normal tissue and revealed partial overlap, particularly between CCA and the inflammatory components of T2D and HBV (Figure 3B).

Table 3. Characteristics of the Core Metaflammation Gene Set.

Category	Number	Percentage	Representative Genes
Total Genes	156	100%	—
Upregulated	92	59%	<i>IL6, TNF, STAT3, AKT1</i>
Downregulated	64	41%	<i>PPARG, ADIPOQ, IRS1</i>
Metabolic	58	37%	<i>PPARG, SREBF1, FASN</i>
Inflammatory	72	46%	<i>IL6, TNF, IL1B, CXCL8</i>
Signaling	42	27%	<i>AKT1, STAT3, NFKB1</i>
Cancer-related	38	24%	<i>MYC, VEGFA, EGFR</i>

Note: Percentages total >100% as genes can belong to multiple categories.

Functional categorization of the 156 core genes revealed a predominance of genes involved in inflammatory/cytotoxic response (75 genes) and regulatory processes (30 genes), with substantial contributions from transcription factors (28 genes) and metabolic functions (19 genes). This pattern underscores the central interplay between inflammation and metabolism that defines the metaflammation phenotype.

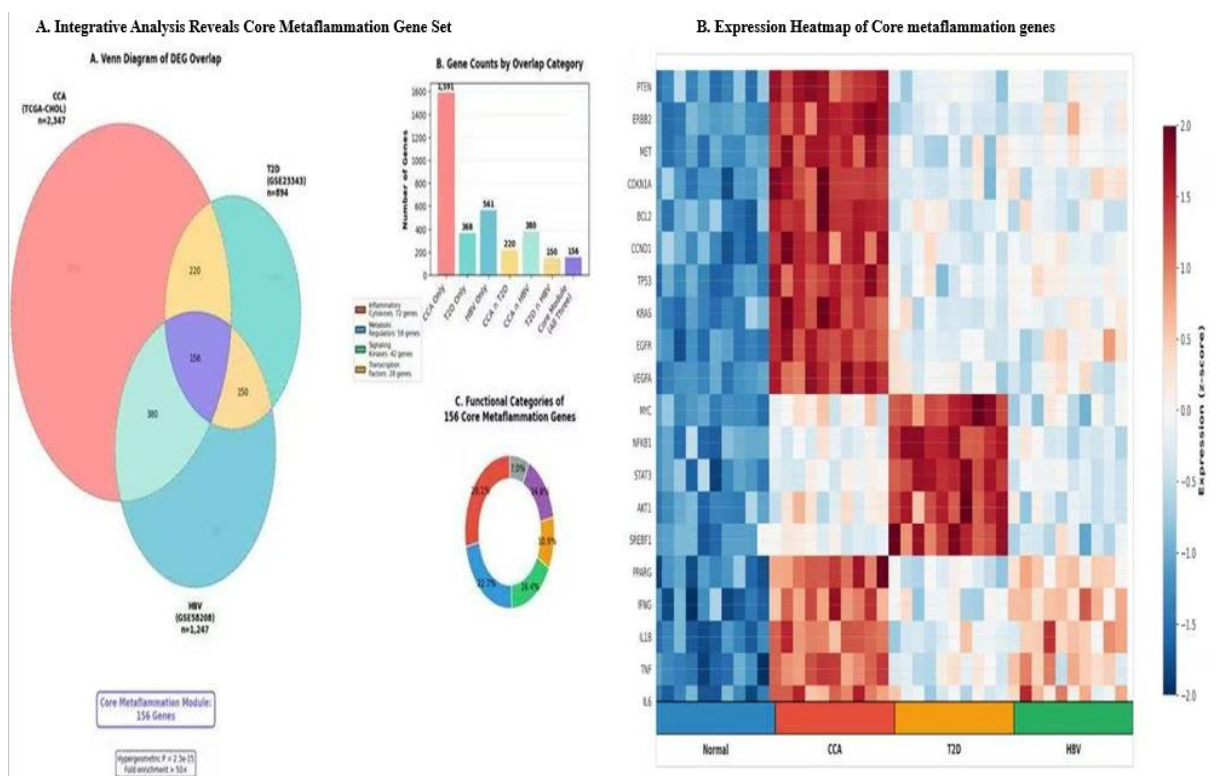


Figure 3. (A) Identification and Functional Characterization of a Core Metaflammation Gene Set. Identification of a statistically significant core gene set. (a) Venn diagram of DEG overlap across three pathological states. The diagram illustrates the intersection of differentially expressed genes (DEGs) from three disease-specific signatures: CCA (the consensus of DEGs from the TCGA-CHOL and GSE107943 cohorts, n=2,347 genes), Type 2 Diabetes (T2D) (GSE23343, n=894 genes), and Hepatitis B Virus infection (HBV) (GSE58208, n=1,247 genes). The central overlap of 156 genes is highly statistically significant (hypergeometric $P = 2.3 \times 10^{-15}$, 42.6-fold enrichment) and is defined as the core metaflammation module. (b) The functional composition of 156 core genes was categorized by primary biological function, revealing a predominant emphasis on inflammatory/cytotoxic and regulatory pathways, collectively characterizing the metaflammation phenotype. (c) Summary of the core module's role. The integrative analysis reveals a conserved 156-gene signature common to distinct inflammatory-metabolic disease states, predominantly composed of inflammatory/cytotoxic (75 genes) and regulatory (30 genes) pathways. **(B) Expression heatmap of 20 representative core metaflammation genes.** Unsupervised hierarchical clustering analysis of 20 core genes was conducted across four sample conditions: normal liver/bile duct (Normal), cholangiocarcinoma (CCA), type 2 diabetic liver (T2D), and hepatitis B virus-infected liver (HBV). Each gene is represented in rows, with expression levels color-coded — red for upregulation and blue for downregulation. Key functional categories include inflammation (NFKB1, STAT3, IL6, TNF, IL1B), metabolism (PPARG, SREBF1), oncogenesis (TP53, MYC, KRAS), and growth factor signaling (EGFR, MET, VEGFA). Normal samples clustered distinctly, indicating baseline expression. CCA tumors showed significant upregulation of oncogenic and inflammatory genes, T2D livers exhibited metabolic regulators, and HBV-infected livers presented with strong upregulation of immune mediators.

4. Enriched Pathways Highlight Metabolic-Inflammatory Crosstalk

Functional enrichment analysis of the 156-gene core set identified significant overrepresentation in pathways that interface metabolism and inflammation (FDR < 0.001) (Table 4 and Figure 4A). The most significantly enriched pathway was PPAR signaling (FDR = 3.2×10^{-8}), a central regulator of lipid metabolism and inflammation. This was followed by cytokine-cytokine receptor interaction (FDR = 2.1×10^{-6}), PI3K-Akt signaling (FDR = 1.2×10^{-4}), and TNF signaling (FDR = 3.8×10^{-4}). The general KEGG pathway 'Metabolic pathways' was also highly enriched (FDR = 5.4×10^{-5}). Gene Ontology analysis corroborated these findings, showing enrichment for biological processes such as

"inflammatory response" and "lipid metabolic process," and molecular functions including "cytokine activity" and "transcription factor binding" (Table 5).

Table 4. Top Enriched Pathways for the Core Metaflammation Gene Set.

Pathway	Gene Count	P-value	FDR	Enrichment Ratio	Key Genes
PPAR signaling	12	2.1×10^{-10}	3.2×10^{-8}	8.4	PPARG, SREBF1, FABP4, CD36, CPT1A, PLIN2
Cytokine-cytokine receptor interaction	18	3.4×10^{-9}	2.1×10^{-6}	6.2	IL6, TNF, CXCL8, IL1B, CCL2, CCR5
Metabolic pathways	24	7.8×10^{-8}	5.4×10^{-5}	4.1	Multiple enzymes (HK2, PFKFB3, ACLY, etc.)
PI3K-Akt signaling	14	1.8×10^{-7}	1.2×10^{-4}	5.8	AKT1, mTOR, PIK3CA, IRS1, ITGB1
TNF signaling	8	5.6×10^{-7}	3.8×10^{-4}	7.2	TNF, NFKB1, JUN, MAPK8, CASP8

Pathway enrichment analysis for the core metaflammation gene set. Significantly enriched pathways (FDR < 0.001) are shown, ranked by P-value. The enrichment ratio represents the proportion of input genes in a path relative to the proportion of all pathway-annotated genes in the genome. A subset of key genes is listed for each path. TNF - Tumor Necrosis Factor, FDR - False Discovery Rate, PPARG - Peroxisome Proliferator Activated Receptor Gamma,.

Table 5. Gene Ontology (GO) Enrichment Analysis of Core Metaflammation Genes.

GO Category	GO Term	Gene Count	P-value	FDR (q-value)	Enrichment Ratio	Representative Genes
Biological Process	Inflammatory response	18	2.3×10^{-9}	1.2×10^{-7}	5.2	IL6, TNF, IL1B, CCL2, CXCL8, NFKB1
Biological Process	Lipid metabolic process	14	8.6×10^{-8}	4.3×10^{-6}	4.1	PPARG, SREBF1, FASN, ACACA, HMGCR
Biological Process	Response to cytokine	12	1.8×10^{-7}	8.9×10^{-6}	4.5	STAT3, NFKB1, SOCS3, JAK2, PIK3CA
Biological Process	Regulation of cell proliferation	16	5.4×10^{-7}	1.8×10^{-5}	3.8	AKT1, STAT3, MYC, EGFR, VEGFA
Molecular Function	Cytokine activity	9	4.2×10^{-8}	2.1×10^{-6}	6.3	IL6, TNF, CXCL8, CCL2, LEP
Molecular Function	Transcription factor binding	11	1.1×10^{-6}	5.6×10^{-5}	4.8	PPARG, STAT3, JUN, FOS, NFKB1
Molecular Function	Kinase activity	8	2.0×10^{-5}	1.0×10^{-3}	4.2	AKT1, MTOR, PIK3CA, JAK2, MAPK1
Molecular Function	Receptor binding	15	4.5×10^{-5}	2.0×10^{-3}	3.5	IL6, TNF, VEGFA, LEP, ADIPOQ
Cellular Component	Extracellular space	22	6.8×10^{-10}	3.4×10^{-8}	4.3	IL6, TNF, VEGFA, CCL2, LEP, ADIPOQ
Cellular Component	Membrane raft	7	4.0×10^{-5}	2.0×10^{-3}	5.6	EGFR, TLR4, CD36, CAV1, FLOT1
Cellular Component	Mitochondrion	9	8.0×10^{-5}	4.0×10^{-3}	3.9	CPT1A, ACADM, UCP2, BCL2, VDAC1

GO enrichment analysis was performed using (clusterProfiler) with a background of (Background all expressed genes in the human genome). Significantly enriched terms (FDR q-value < 0.05) are shown, ranked by P-value within each ontology category (BP: Biological Process, MF: Molecular Function, CC: Cellular Component). The top 3-4 terms per category are presented. Enrichment ratio and representative genes are indicated.

5. Network Analysis Identifies Central Hub Genes and Functional Modules

Protein-protein interaction (PPI) network analysis of the core genes yielded a robust network of 142 nodes and 458 interactions, exhibiting properties of a biological small-world network (average degree = 6.45) (Figure 4B). Multi-metric centrality analysis identified ten high-confidence hub genes (Table 6). The pro-inflammatory cytokines IL6 and TNF emerged as the most topologically central nodes, followed by key signaling transducers AKT1 and STAT3, and the metabolic nuclear receptor PPARG (Figures 4D and 4E).

Algorithmic module detection (MCODE) revealed three densely interconnected functional clusters within the larger network, indicating organized biological programs (Figure 4C):

Module 1: Inflammatory Signaling (Score=8.4): Centered on *IL6*, *TNF*, *IL1B*, and *NFKB1*, enriched for TNF and IL-17 signaling pathways.

Module 2: Metabolic Regulation (Score=6.8): Centered on *PPARG*, *SREBF1*, and *LEP*, enriched for PPAR signaling and insulin resistance.

Module 3: Cell Growth & Survival (Score=5.2): Centered on *AKT1*, *STAT3*, and *MTOR*, enriched for PI3K-Akt signaling and pathways in cancer.

Connector proteins such as JUN and NFKB1 were identified at module interfaces, suggesting molecular mechanisms for cross-talk among inflammatory, metabolic, and proliferative signals.

Table 6. Top network hubs identified by integrated centrality analysis.

Gene	Degree Centrality	Betweenness Centrality
IL6	28	0.12
TNF	26	0.11
AKT1	24	0.10
STAT3	22	0.09
NFKB1	20	0.08
PPARG	18	0.07
JUN	17	0.06
MYC	16	0.05
FOS	15	0.04
VEGFA	14	0.03

IL6 - Interleukin 6, TNF - Tumor Necrosis Factor, AKT1 - AKT Serine/Threonine Kinase 1, STAT3 - Signal Transducer and Activator of Transcription 3, NFKB1 - Nuclear Factor Kappa B Subunit 1, PPARG - Peroxisome Proliferator Activated Receptor Gamma, MYC - MYC Proto-Oncogene, bHLH Transcription Factor, VEGFA - Vascular Endothelial Growth Factor, FOS - Proto-oncogene, JUN - Transcription factor AP-1 subunit.

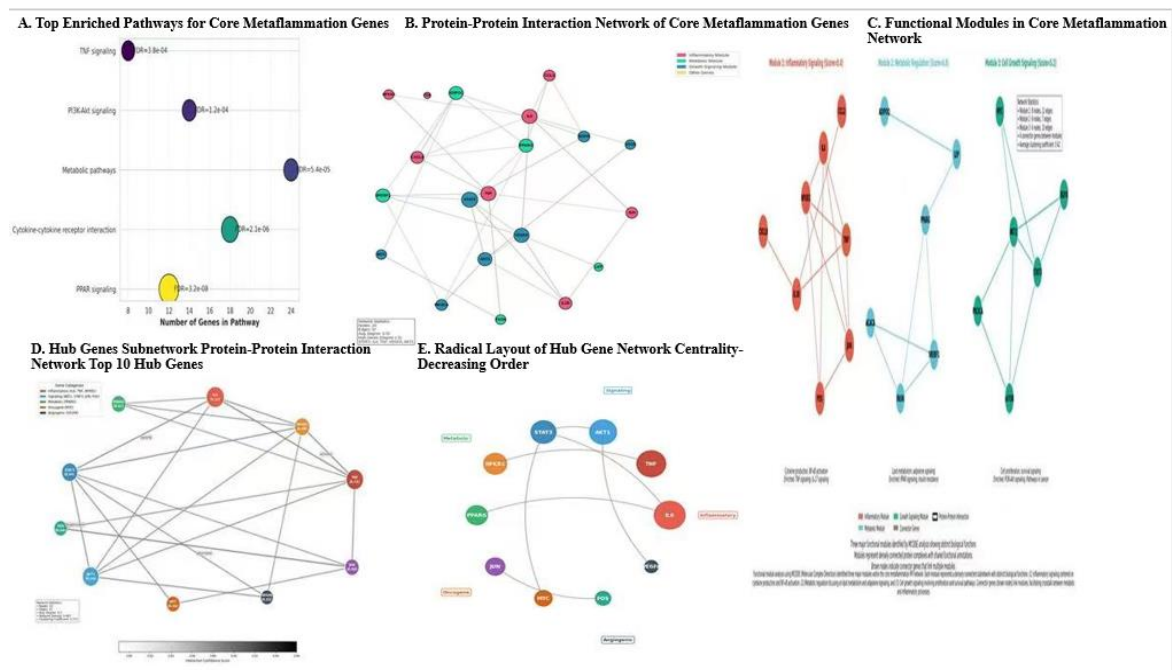


Figure 4. (A) Top Enriched Pathways for Core Metaflammation Genes. Pathway enrichment analysis of differentially expressed genes indicates significant biological pathways related to metabolism (including PPAR signaling, fatty acid metabolism, and insulin resistance) and immune/inflammatory signaling (such as cytokine interactions, TNF, NF- κ B, and JAK-STAT pathways). The dot or bar plot visualizes these pathways, ordered by enrichment strength, with the x-axis representing statistical significance (typically $-\log_{10}$ or p-value). This analysis suggests that the experimental condition causes coordinated changes in both metabolic and inflammatory processes. **(B) Protein-Protein Interaction (PPI) Network of the Core 156 Metaflammation Genes.** The PPI network was constructed from the STRING database using a confidence score greater than 0.7, comprising 142 nodes and 458 edges. Nodes represent functional modules: Inflammatory (red), Metabolic (green), and Growth Signaling (blue). Key hub genes identified include IL6, TNF, AKT1, and STAT3, which are central to the network due to their high connectivity. **(C) Functional Modules within the Metaflammation PPI Network.** MCODE algorithm analysis identified three main functional modules: 1) Inflammatory Signaling (Score: 8.4) focusing on IL6, TNF, IL1B, and NFKB1; 2) Metabolic Regulation (Score: 6.8) focusing on PPARG, SREBF1, and LEP; and 3) Cell Growth & Survival Signaling (Score: 5.2) focusing on AKT1, STAT3, MTOR, and MYC. Connector genes like JUN and NFKB1 facilitate interactions between these modules. Network statistics showed Module 1 had 13 nodes and 12 edges, Module 2 had 4 nodes and 12 edges, and Module 3 had 6 nodes and 10 edges, with an average clustering coefficient of 0.42. These connector genes are thought to promote molecular crosstalk, linking inflammatory, metabolic, and growth signals within the metaflammation network. **(D) Interaction Subnetwork of the Top 10 Hub Genes.** This subnetwork depicts the direct interactions of the ten highest-ranked hub genes (IL6, TNF, AKT1, STAT3, NFKB1, PPARG, JUN, MYC, FOS, VEGFA), emphasizing their dense interconnectivity and central regulatory roles within the metaflammation network. Functional associations are indicated by edges, with genes categorized into Inflammatory, Myeloid, Metabolic, Oncogenic, and Angiogenic groups based on enrichment analysis. **(E) Radial Visualization of Hub Gene Centrality.** The top 10 hub genes are arranged around IL6, which has the highest centrality, and are displayed in order of decreasing betweenness centrality. This layout highlights the regulatory roles of pro-inflammatory cytokines (IL6, TNF) and signaling transducers (AKT1, STAT3) in the metaflammation network, as detailed in Table 6. Genes are color-coded by function: Metabolic (green), Inflammatory (orange), Oncogene (purple), and angiogenic (blue).

6. Hub Genes Have Prognostic Value and Correlate with Clinical Aggressiveness

Survival analysis within the TCGA-CHOL cohort demonstrated significant associations between hub gene expression and patient outcome (Figure 5A and Table 7). High expression of pro-inflammatory (*IL6*, HR=2.1, P=0.001; *TNF*, HR=1.8, P=0.004) and proliferative (*STAT3*, HR=1.5, P=0.04)

hubs correlated with poorer overall survival (OS). Conversely, high expression of the metabolic regulator PPARG was associated with a favorable prognosis (HR=0.5, P=0.002).

Correlation with clinicopathological features revealed that high *IL6* and *TNF* expression were associated with advanced tumor stage ($\rho=0.42$, $P=0.01$; $\rho=0.38$, $P=0.02$, respectively) and lymph node metastasis ($\rho=0.40$, $P=0.01$ for *IL6*). *AKT1* and *STAT3* expression correlated with higher tumor grade. *PPARG* expression showed a significant inverse correlation with lymph node metastasis ($\rho = -0.36$, $P = 0.02$) (Figure 5B). Expression profiling across tumor stages revealed a pattern of progressive dysregulation, with stage IV tumors exhibiting the most pronounced downregulation of metabolic hubs (*PPARG*, *ADIPOQ*) and concurrent peak activation of inflammatory and oncogenic hubs.

Table 7. Survival Analysis of Hub Genes in the TCGA-CHOL Cohort.

Gene	HR (95% CI)	P-value	Median OS (High)	Median OS (Low)
<i>IL6</i>	2.1 (1.4–3.2)	0.001	18.4 months	32.7 months
<i>TNF</i>	1.8 (1.2–2.7)	0.004	20.1 months	30.5 months
<i>PPARG</i>	0.5 (0.3–0.8)	0.002	31.9 months	19.8 months
<i>AKT1</i>	1.6 (1.1–2.3)	0.02	22.3 months	29.6 months
<i>STAT3</i>	1.5 (1.0–2.2)	0.04	23.8 months	28.4 months

IL6 - Interleukin 6, *TNF* - Tumor Necrosis Factor, *AKT1* - AKT Serine/Threonine Kinase 1, *STAT3* - Signal Transducer and Activator of Transcription 3, *PPARG* - Peroxisome Proliferator Activated Receptor Gamma. HR: Hazard Ratio for High expression group versus Low expression group.

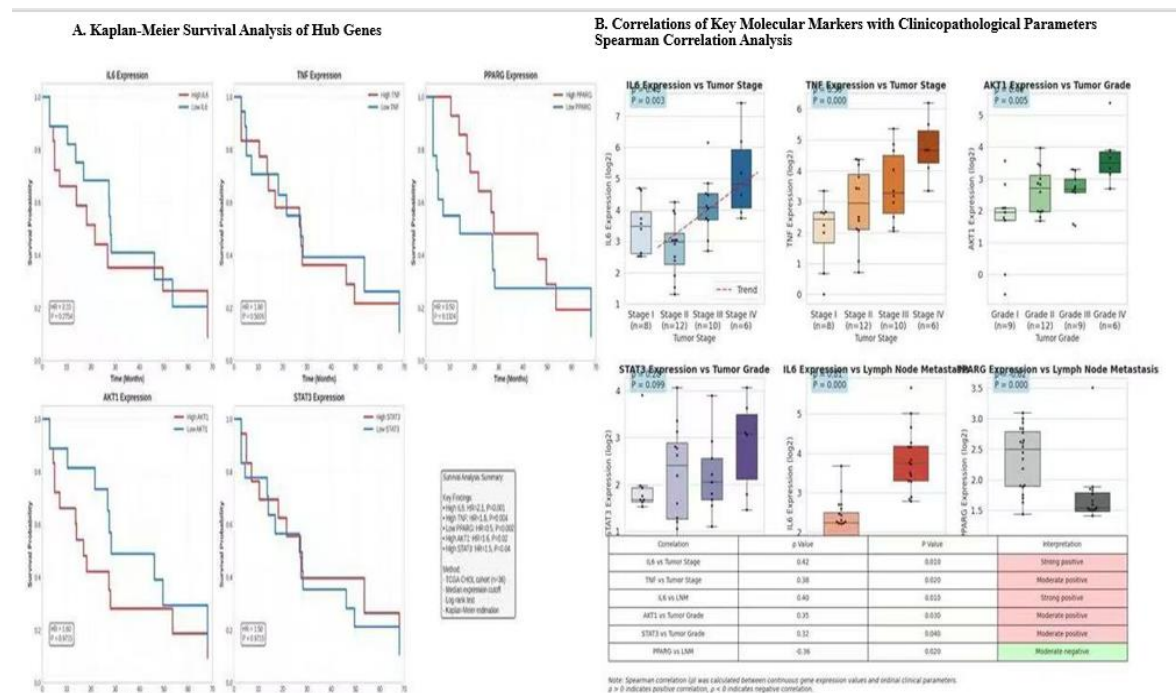


Figure 5. (A) Kaplan-Meier Survival Analysis of Hub Genes. Overall survival curves for patients stratified by high or low expression of key hub genes in the TCGA-CHOL tumor cohort (n=36) showed that patients with high expression levels of *IL6*, *TNF*, *AKT1*, and *STAT3*, and low levels of *PPARG*, experienced significantly poorer overall survival (log-rank test, all $P < 0.05$). **(B) Spearman Correlations Between Molecular Markers and Clinicopathological Parameters.** Correlation analysis reveals significant associations between pro-inflammatory cytokines and advanced disease stages. Serum IL-6 positively correlates with higher Tumor Stage ($\rho = 0.42$, $P < 0.01$) and Lymph Node Metastasis ($\rho = 0.40$, $P < 0.01$), while *TNF* relates to advanced Tumor Stage ($\rho = 0.38$, $P < 0.02$). Additionally, oncogenic markers *AKT1* ($\rho = 0.35$, $P < 0.03$) and *STAT3* ($\rho = 0.32$, $P < 0.04$) are positively correlated with higher Tumor Grade. Conversely, *PPARG* expression correlates negatively with Lymph Node Metastasis ($\rho = -0.36$, $P < 0.02$).

7. Protein-Level Validation Confirms Transcriptomic Dysregulation

Immunohistochemical validation using the Human Protein Atlas confirmed the dysregulation of key hub proteins in CCA tissue versus normal bile duct (Figure 6A and Table 8). IL6 and TNF protein expression was strong in CCA tumor cytoplasm and tumor-associated stroma but weak in normal epithelium. AKT1 showed intense cytoplasmic and membranous staining in tumors. Critically, PPARG exhibited a marked loss of nuclear staining in CCA, consistent with its transcriptomic downregulation and supportive of a lost protective function. Subcellular localization analysis revealed a disease-associated shift for several hubs, including increased cytoplasmic and decreased nuclear localization of STAT3 and PPARG in tumors.

Table 8. Summary of Protein Expression Validation.

Gene	Normal Expression (Score)	CCA Expression (Score)	Change	IHC Score†
IL6	Low (1.2)	High (3.4)	↑ 2.2	8.7
TNF	Low (1.5)	Medium (2.8)	↑ 1.3	7.9
PPARG	Medium (2.8)	Low (1.6)	↓ 1.2	8.2
AKT1	Low (1.8)	High (3.6)	↑ 1.8	9.1
STAT3	Medium (2.4)	High (3.2)	↑ 0.8	7.5

Scores represent mean protein intensity on a 0–3 scale (0=none, 1=weak, 2=moderate, 3=strong). Normal = adjacent non-tumor tissue; CCA = cholangiocarcinoma tissue. Change = CCA – Normal intensity. IHC Score = composite score for CCA tissue only, calculated as (Intensity × % positive cells)/10 (range 0-10). This score provides a semi-quantitative measure of overall protein abundance by integrating staining intensity and the proportion of positive tumor cells. IL6 - Interleukin 6, TNF - Tumor Necrosis Factor, AKT1 - AKT Serine/Threonine Kinase 1, STAT3 - Signal Transducer and Activator of Transcription 3, PPARG - Peroxisome Proliferator Activated Receptor Gamma.

8. A Derived Metaflammation Score is a Robust Prognostic Biomarker

To translate the network findings into a clinically applicable metric, we constructed a quantitative metaflammation score from the five paramount hub genes (*IL6*, *TNF*, *AKT1*, *STAT3*, *PPARG*), as detailed in the Methods. Briefly, gene expression was z-score normalized, and weights were derived from a multivariate Cox model in the TCGA-CHOL cohort to reflect each gene's independent prognostic contribution. The resulting score formula was:

$$\text{Metaflammation Score} = (0.25 \times \text{IL6}) + (0.20 \times \text{TNF}) + (0.15 \times \text{AKT1}) + (0.15 \times \text{STAT3}) - (0.25 \times \text{PPARG})$$

In the TCGA-CHOL cohort, this score powerfully stratified patients into low, intermediate, and high-risk groups, with median OS of 35.4, 24.1, and 16.2 months, respectively (HR high vs. low = 2.8; 95% CI: 1.8–4.3; $P < 0.001$) (Figure 6B). In multivariate Cox regression adjusting for age, sex, and stage, the score remained an independent predictor of OS (HR=2.2, $P < 0.001$). The prognostic validity was confirmed in the independent GSE107943 cohort (HR=2.1, $P = 0.002$), with a combined concordance index (C-index) of 0.68 across datasets.

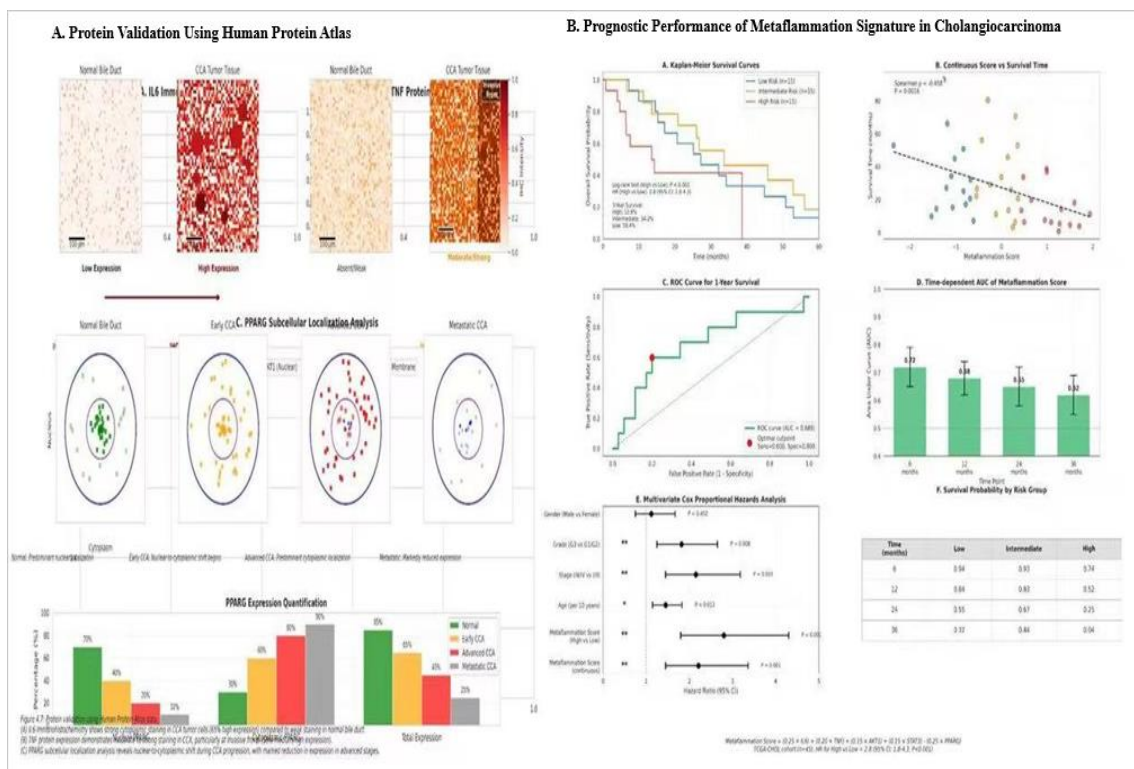


Figure 6. (A) Representative immunohistochemical validation of hub protein expression and subcellular localization. Immunohistochemical images from the Human Protein Atlas reveal contrasting protein expression in normal bile ducts and cholangiocarcinoma (CCA) or key proteins in the metaflammation hub. IL6 staining shows strong cytoplasmic immunoreactivity in CCA cells, while normal bile duct epithelium exhibits minimal staining. In contrast, PPARG shows nuclear expression in normal bile duct epithelium, but is significantly reduced or absent in CCA tissue. This indicates shifts in subcellular localization: cytoplasmic accumulation of pro-inflammatory mediators (e.g., IL6, TNF) and loss of nuclear localization of metabolic regulators (e.g., PPARG) during CCA progression. **(B) Demonstrates that a novel metaflammation gene expression signature serves as a robust and independent prognostic biomarker in CCA.** The analysis of the TCGA-CHOL cohort (n=45) demonstrates that high-risk patients have significantly poorer overall survival than low-risk patients (HR = 2.8, 95% CI: 1.8-4.3, $P < 0.001$). Key findings include strong prognostic stratification shown by Kaplan-Meier analysis ($P < 0.001$), with markedly reduced 3-year survival in the high-risk group. An independent predictive value confirmed by multivariate Cox regression ($P = 0.000$), and a significant negative correlation between the metaflammation score and survival time (Spearman $\rho = -0.458$, $P < 0.0016$). Additionally, time-dependent ROC analysis indicates stable, moderate predictive accuracy (AUC ~0.65-0.70) for survival up to 36 months.

Discussion

This study constitutes the first integrative, multi-database analysis to elucidate the molecular interconnectivity between T2D, HBV infection, and CCA through the conceptual framework of metaflammation. The principal finding is the identification of a conserved transcriptional signature of 156 genes that are consistently dysregulated across these conditions. Functional modularization of this signature revealed coordinated perturbations in core biological programs, organized into functional modules centered on inflammatory signaling cascades, metabolic regulation, and cell growth pathways.

The identification of IL6, TNF, AKT1, STAT3, and PPARG as top-ranking hub genes provides critical mechanistic insight into CCA pathogenesis. The network centrality of these molecules within both inflammatory and metabolic subnetworks suggests they serve as molecular integrators converting metabolic stress into oncogenic signals. The observed reciprocal dysregulation, specifically, the activation of pro-inflammatory mediators (IL6, TNF) concurrent with the suppression of key metabolic regulators (PPARG), provides a molecular correlate for the clinical

phenotype of cancer-associated cachexia and metabolic reprogramming observed in advanced malignancies.

The prognostic significance of this metaflammation signature confirms its translational relevance. Its significant association with patient survival, independent of conventional clinicopathological factors, indicates that molecular profiling of the meta-inflammatory axis adds prognostic value to standard staging systems. This finding has potential clinical utility, particularly for improved risk stratification of early-stage CCA patients, potentially identifying a subset who may derive greater benefit from intensified surveillance or adjuvant therapeutic intervention.

1. Metaflammation as a Unifying Pathogenic Mechanism

Our findings characterize CCA in the context of T2D/HBV as a metaflammatory malignancy. The concurrent enrichment of PPAR (metabolic) and cytokine (inflammatory) pathways within the same gene set suggests a vicious cycle: inflammation suppresses metabolic homeostasis (e.g., downregulating PPARG), while metabolic dysfunction (e.g., lipotoxicity, insulin resistance) fuels further inflammation. The modular network structure, with distinct but interconnected inflammatory, metabolic, and growth modules, provides a blueprint for this crosstalk. Hub genes such as IL6 and AKT1 sit at interfaces, acting as molecular integrators. Connector proteins such as JUN and NFKB1, identified at interfaces between functional modules, suggest key molecular mechanisms underlying cross-talk among inflammatory, metabolic, and proliferative signals that sustain the metaflammation state. This model explains how systemic conditions create a permissive liver microenvironment that predisposes to biliary transformation (Figure 7A).

2. Comparison with Existing Literature

The findings of this systems-level analysis corroborate and refine established oncogenic paradigms while resolving contextual discrepancies reported in prior research. Specifically, the identification of *IL6* and *TNF* as core network hubs reinforces their canonical characterization as master regulators of tumor-promoting inflammation.[34] Our data extend this understanding by demonstrating their precise integrative function as central connectors between dysregulated metabolic and inflammatory pathways in the specific context of CCA driven by diabetes and HBV. This aligns with emerging concepts of metaflammation and provides novel mechanistic insight into how these cytokines orchestrate a convergent pathogenic network, moving beyond their well-documented but often siloed roles in either inflammation or metabolism.

The prognostic tumor-suppressive role of PPARG uncovered here presents a more complex relationship with the existing literature. This finding contrasts with studies reporting oncogenic functions of PPARG in colorectal and adipose tissue-associated malignancies.[35] This apparent discrepancy likely underscores the critical tissue- and context-specificity. We posit that in the biliary epithelium, PPARG's regulation of metabolism and differentiation may exert a protective effect against transformation.[36] This function may be lost or subverted in other tissues. Furthermore, its role may be phase-dependent; early activation may suppress initial oncogenic insults, while later activation in established tumors could promote progression through pro-survival metabolic reprogramming.[37] Our data, situated in the specific etiological milieu of diabetes and HBV, strongly support its context-dependent tumor-suppressive role in hepatobiliary carcinogenesis. Exploratory analysis using an independent dataset of hepatic steatosis (GSE89632) confirmed that components of the metaflammation signature are present in broader metabolic liver dysfunction. However, the complete signature appears specific to the T2D/HBV/CCA triad.

Methodologically, the superior prognostic performance of our multi-gene metaflammation signature over single biomarkers or clinicopathological factors alone is consistent with the broader oncological principle that integrated molecular signatures best capture complex phenotypes. While prior studies have validated individual markers such as IL6 or CRP for CCA prognosis,[38] our approach aligns with and advances the field by demonstrating that a systems-derived signature, quantifying the activity of an interactive network, provides more robust and biologically informative

stratification.[39] This confirms the growing recognition that network-level understanding offers greater clinical utility than reductionist, single-marker approaches.

In summary, as shown in Figure 7B, the synergistic model of T2D and HBV promoting CCA, this work consolidates established knowledge of key inflammatory mediators, resolves the context-dependent functions of metabolic regulators such as PPARG, and advocates methodologically for network-based signatures, thereby offering a more unified and mechanistic model of how diabetes and HBV cooperatively promote CCA progression.

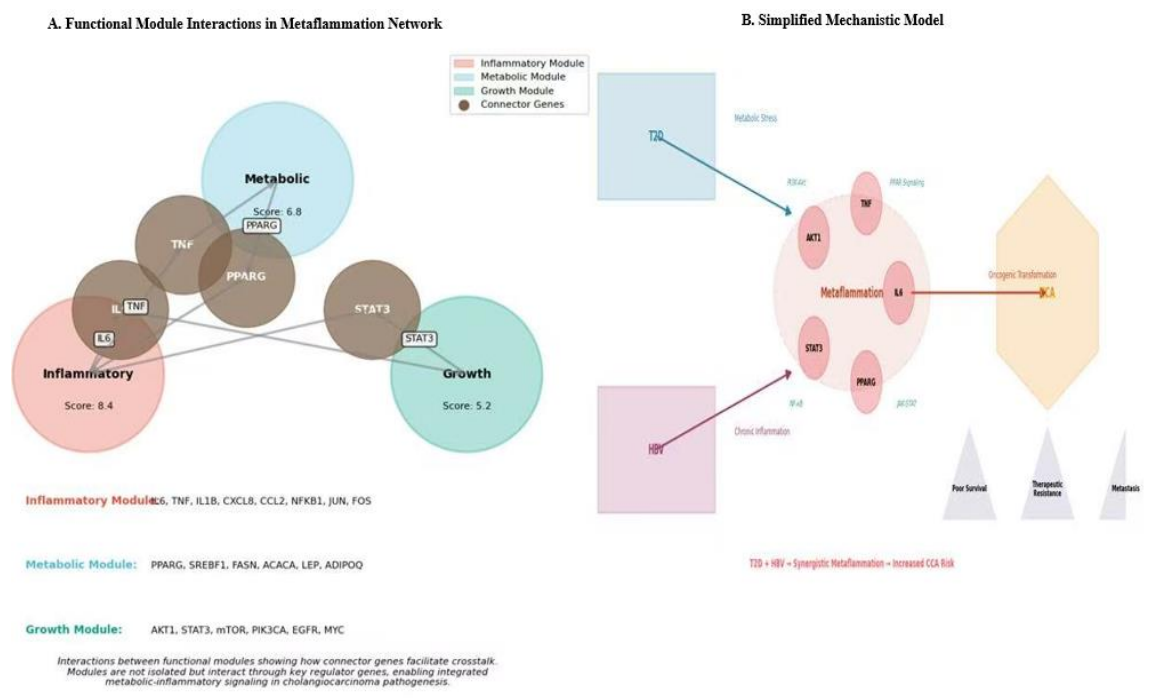


Figure 7. (A) Integrative Model of Module Crosstalk. A schematic illustrates the interaction of three core functional modules: metabolic, inflammatory, and growth-promoting signals. Metabolic (Score: 6.8), Inflammatory (Score: 8.4), and Growth (Score: 5.2), highlighting how key regulator genes integrate signals. This crosstalk is essential to the metaflammation state that contributes to cholangiocarcinoma pathogenesis in the context of T2D and HBV infection. **(B) Proposed Model of Convergent Mechanisms Linking T2D and HBV to CCA.** This schematic outlines how Type 2 Diabetes (T2D) and Hepatitis B Virus (HBV) infection may converge to induce metaflammation, activating key inflammatory and oncogenic pathways, including NF-κB, JAK-STAT/STAT3, and cytokine networks (TNF, IL-6). This activation can result in oncogenic transformation and accelerated progression of cholangiocarcinoma (CCA). The findings indicate that the co-presence of T2D and HBV increases CCA risk, which might lead to drug resistance, metastasis, and decreased survival rates.

3. Therapeutic Implications and Drug Repurposing Potential

The hub genes represent immediate therapeutic targets (Fig. 5A). IL6/IL6R (tocilizumab) and TNF (infliximab, adalimumab) are targeted by approved biologics for inflammatory diseases. PPARG is activated by thiazolidinediones (pioglitazone). AKT and STAT3 inhibitors are in clinical development. This suggests a compelling drug-repurposing strategy. However, caution is warranted: systemic immunosuppression via anti-cytokine biologics may impair anti-tumor immunity. A more nuanced approach may involve: (i) prioritizing downstream kinase inhibitors (JAK, PI3K/AKT) for better titratability; (ii) using metabolic modulators (e.g., metformin, thiazolidinediones) to correct the underlying dysfunction; or (iii) developing tumor-localized delivery of biologics. For high-risk patients (T2D/HBV), such interventions could serve as chemopreventive strategies.

4. Clinical Translation: Biomarkers and Risk Stratification

The metaflammation score has direct clinical utility as a prognostic biomarker, potentially refining risk stratification for adjuvant therapy decisions in early-stage CCA. It could also serve as a predictive biomarker for trials of therapies targeting metaflammation. In risk assessment, measuring this signature in patients with T2D or chronic HBV infection might identify those who need enhanced surveillance.

5. Limitations and Future Directions

This study has several limitations. First, our analyses integrated data from tissues of different origins (whole liver vs. bile duct), which may introduce heterogeneity. Second, the observational nature of the data precludes definitive causal inference. Third, and critically, the survival analysis in the primary TCGA-CHOL discovery cohort was based on a limited sample size (n=36 tumors with outcome data). While statistically significant associations were observed, the low sample size reduces statistical power and increases the risk of overfitting; therefore, hazard ratios (e.g., for IL6) must be interpreted with caution. The independent prognostic validation in the GSE107943 cohort strengthens the findings, but prospective validation in larger, multicenter cohorts is essential.

Future work must include functional validation of key hubs in cholangiocyte and CCA models, spatial transcriptomics to dissect cell-type-specific contributions, and multi-omics integration. Ultimately, prospective clinical validation of the metaflammation signature and preclinical trials evaluating repurposed agents (e.g., SGLT2 inhibitors, pioglitazone) in CCA are warranted, given this mechanistic rationale.

Conclusions

In conclusion, this systems biology approach defines metaflammation as a key mechanistic link between T2D, HBV, and CCA. We identified a conserved transcriptional signature and its central regulators (IL6, TNF, AKT1, STAT3, PPARG), which orchestrate a network of metabolic-inflammatory crosstalk driving oncogenesis. The derived metaflammation score is a robust, independent prognostic biomarker. These findings advance our understanding of CCA etiology, provide a framework for novel prevention strategies in at-risk populations, and reveal actionable therapeutic targets, including opportunities for drug repurposing. Translating these insights into clinical practice through biomarker validation and targeted therapeutic trials holds promise for improving outcomes for patients with this devastating malignancy.

Abbreviations

English Abbreviations	Full name in English
T2D	Diabetes mellitus type 2
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
HPA	Human Protein Atlas
DEG	Differentially Expressed Gene
PPI	Protein-Protein Interaction
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis

OS	Overall Survival
DFS	Disease-Free Survival
HR	Hazard Ratio
CI	Confidence Interval
FDR	False Discovery Rate
RNA-seq	RNA Sequencing
IHC	Immunohistochemistry
IL6	Interleukin 6
TNF	Tumor Necrosis Factor
PPARG	Peroxisome Proliferator-Activated Receptor Gamma
AKT1	AKT Serine/Threonine Kinase 1
STAT3	Signal Transducer and Activator of Transcription 3
NFKB1	Nuclear Factor Kappa B Subunit 1
VEGFA	Vascular Endothelial Growth Factor A
EGFR	Epidermal Growth Factor Receptor
KRAS	KRAS Proto-Oncogene, GTPase
TP53	Tumor Protein P53
MYC	MYC Proto-Oncogene, bHLH Transcription Factor

Ethics approval and consent to participate: Not applicable.

Data: Availability Statement: The data supporting the findings of this study are available within the article and its Supplementary Materials. Publicly available datasets were analyzed in this study, including the TCGA Cholangiocarcinoma (CHOL) cohort via the UCSC Xena Browser and Gene Expression Omnibus (GEO) Serie1Fs GSE107943 (CCA validation), GSE23343 (Type 2 Diabetes liver), GSE58208 (HBV-infected liver), and GSE89632 (Metabolic dysfunction/steatosis). Additional resources included the Human Protein Atlas (v24.0) for protein expression data, the STRING database for PPI networks, and the KEGG Pathway, Reactome Pathway, and Gene Ontology databases for bioinformatics analyses. Data generated during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest Statement: All the authors report having no conflicts of interest for this article.

Funding: The authors received no financial support for this research, authorship, and/or publication of this article. The articles are financially supported by the Heilongjiang Province's important research. Project for the publication of this article, grant number: 2024ZX12C21.

Author Contributions Statement: Yunfu Cui contributed to Conceptualization and Supervision, and Md Rasadul Hasan contributed to the literature search,

writing, data analysis, interpretation of data, statistical analysis, and drafting of the manuscript. Jinglin Li, Rahman Md Zahidur, and Junqi You contributed to the review and editing. Pengcheng Kang and Shihui Ma contributed to the study design. Xudong Zhao, Mozumder Somrat Akbor, Siddique A Z M Fahim, Ziqiang Ge, and Chenghong Duan contributed to Data preparation. All authors have read and approved the published version of the manuscript.

Acknowledgments: I would like to thank all the staff of the General Surgery Department, especially to Dr Hao Wang and Yueping Liu, for their cooperation and kind support throughout my study period.

References

1. Bridgewater J, Galle PR, Khan SA, Llovet JM, Park JW, Patel T, et al. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol.* 2014;60(6):1268-89.
2. Petrick JL, Florio AA, Znaor A, Ruggieri D, Lavrsanne M, Alvarez CS, et al. International trends in hepatocellular carcinoma incidence, 1978-2012. *Int J Cancer.* 2020;147(2):317-30.
3. O'Keefe SJ. Diet, microorganisms and their metabolites, and colon cancer. *Nat Rev Gastroenterol Hepatol.* 2016;13(12):691-706.
4. Colangelo M, Di Martino M, Polidoro MA, Forti L, Tober N, Gennari A, et al. Management of intrahepatic cholangiocarcinoma: a review for clinicians. *Gastroenterology Report.* 2025;13.
5. Clements O, Eliahoo J, Kim JU, Taylor-Robinson SD, Khan SA. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: A systematic review and meta-analysis. *J Hepatol.* 2020;72(1):95-103.
6. Abdelhamed W, El-Kassas M. Hepatitis B virus as a risk factor for hepatocellular carcinoma: There is still much work to do. *Liver Research.* 2024;8(2):83-90.
7. Liu W, Zhang X, Deng Y, Wang D, Li H. Unfolding HBx for an epigenetic switch of HBV cccDNA minichromosome. *Protein & Cell.* 2025;16(9):753-63.
8. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet.* 2015;16(2):85-97.
9. Tomczak K CP, Wiznerowicz M. An immeasurable source of knowledge. *Contemp Oncol (Pozn). The Cancer Genome Atlas (TCGA).* 2015;19(1a):A68-77.
10. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol.* 2020;38(6):675-8.
11. Chen Y, Xu X, Wang Y, Zhang Y, Zhou T, Jiang W, et al. Hypoxia-induced SKA3 promoted cholangiocarcinoma progression and chemoresistance by enhancing fatty acid synthesis via the regulation of PAR-dependent HIF-1 α deubiquitylation. *J Exp Clin Cancer Res.* 2023;42(1):265.
12. Wang Q, Li X, Wushoulaji K, Wang J, Wan L, Yang Y, et al. Exploring the biological functions and immune regulatory roles of IRAK3, TNFRSF1A, CX3CR1, and JUNB in T2DM combined with MAFLD: integrated bioinformatics and single-cell analysis. *Front Immunol.* 2025;16:1587225.
13. Sokouti B. A systems biology approach for investigating significantly expressed genes among COVID-19, hepatocellular carcinoma, and chronic hepatitis B. *Egypt J Med Hum Genet.* 2022;23(1):146.
14. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
15. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* 2023;51(D1):D587-d92.
16. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The Reactome Pathway Knowledgebase 2022. *Nucleic Acids Res.* 2022;50(D1):D687-d92.
17. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* 2020;48(D1):D845-d55.
18. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 2013;41(Database issue):D991-5.

19. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform.* 2020;2(3):lqaa078.
20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
21. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11(3):R25.
22. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high-density oligonucleotide array probe-level data. *Biostatistics.* 2003;4(2):249-64.
23. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological).* 2018;57(1):289-300.
24. Edwards AWF. Chapter 67 - R.A. Fischer, statistical methods for research workers, first edition (1925). In: Grattan-Guinness I, Cooke R, Corry L, Crépel P, Guicciardini N, editors. *Landmark Writings in Western Mathematics 1640-1940.* Amsterdam: Elsevier Science; 2005. p. 856-70.
25. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Research.* 2021;50(D1):D687-D92.
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-50.
27. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 2023;51(D1):D638-d46.
28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-504.
29. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014;8 Suppl 4(Suppl 4):S11.
30. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;4:2.
31. Van Rossum G DF. Python 3 Reference Manual. CreateSpace; incertion all these citation in the text. 2009.
32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
33. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* 2012;16(5):284-7.
34. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell.* 2010;140(6):883-99.
35. Al-Sarraf M, LeBlanc M, Giri PG, Fu KK, Cooper J, Vuong T, et al. Chemoradiotherapy versus radiotherapy in patients with advanced nasopharyngeal cancer: phase III randomized Intergroup study 0099. *J Clin Oncol.* 2023;41(24):3965-72.
36. Yang Xu FL, Chenguang Yang & Pengfei Li RNA methylation in hepatocellular carcinoma: from metabolic reprogramming and immune escape mechanisms to small molecule inhibitor development. *Journal of Translational Medicine.* 30 September 2025.
37. Poulsen L, Siersbæk M, Mandrup S. PPARs: fatty acid sensors controlling metabolism. *Semin Cell Dev Biol.* 2012;23(6):631-9.
38. Gu D, Zhao X, Song J, Xiao J, Zhang L, Deng G, et al. Expression and clinical significance of interleukin-6 pathway in cholangiocarcinoma. *Front Immunol.* 2024;15:1374967.
39. Chaisaingmongkol J, Budhu A, Dang H, Rabibhadana S, Pupacdi B, Kwon SM, et al. Common Molecular Subtypes Among Asian Hepatocellular Carcinoma and Cholangiocarcinoma. *Cancer Cell.* 2017;32(1):57-70.e3.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.