JOURNAL OF CLINICAL MEDICINE OF KAZAKHSTAN (E-ISSN 2313-1519)
Original Article DOI: https://doi.org/

Identification of Hub Genes Associated with Bile Duct Cancer Using Integrated Gene Expression Data with Protein-Protein Interaction Network

Narmadha Ramasamy[1], Pogala Hema Vardhan[1], Kannan Muthu[1]

[1]Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India

Received: 2025-11-01.
Accepted: 2026-04-16.

J Clin Med Kaz 2026; 23(3): 11-23

Corresponding author:
Kannan Muthu.
E-mail: kannan@bicpu.edu.in.
ORCID: ___.

ABSTRACT

Introduction: Bile duct cancer (cholangiocarcinoma) is an uncommon but highly aggressive malignancy characterized by late diagnosis and poor prognosis. Molecular profiling is essential to understand its pathogenesis and identify biomarkers for early detection and targeted therapy. This study aimed to identify key regulatory genes and molecular pathways associated with bile duct cancer using an integrated bioinformatics approach.

Methods: Gene expression datasets (GSE131027 and GSE107754) were retrieved from the NCBI Gene Expression Omnibus (GEO) database, and samples annotated as cholangiocarcinoma were selected for analysis. Differentially expressed genes were identified using GEO2R. Protein–protein interaction networks were constructed using the STRING database and analyzed in Cytoscape with CytoHubba for hub gene identification.

Results: A total of over 1,000 differentially expressed genes (DEGs) were identified across the analyzed datasets. Network and functional enrichment analyses highlighted seven major hub genes—TP53, HIST1H3F, H2AFZ, FOS, POLR2B, CAV1, and SMAD3 that occupied central positions within the protein–protein interaction networks. These hub genes were associated with pathways related to transcriptional regulation, oxidative and nitrosative stress responses, RNA processing, and post-transcriptional gene regulation. Gene Ontology and KEGG pathway analyses further indicated enrichment of biological processes previously implicated in cancer-related molecular mechanisms, supporting their relevance in cholangiocarcinoma.

Conclusion: This integrative bioinformatics study identified critical hub genes and molecular pathways that may serve as potential biomarkers and therapeutic targets for bile duct cancer. Further validation and molecular docking studies could facilitate the development of targeted drugs and improve treatment outcomes.

Keywords: Cholangiocarcinoma, biomarkers, Hub genes, therapeutics, gene expression.

Introduction

Bile duct cancer (BDC), or cholangiocarcinoma, is a rare but highly aggressive malignancy arising from the epithelial lining of the bile ducts, which transport bile from the liver to the small intestine. Despite its low incidence, BDC presents significant clinical challenges due to late diagnosis, early local invasion, and poor overall prognosis. Anatomically, cholangiocarcinoma is classified into intrahepatic (iCCA) that originates within the liver parenchyma, perihilar (pCCA) arises at or near the hepatic duct bifurcation, and distal extrahepatic (dCCA) is where the tumor occurs closer to the pancreas and ampulla of vater. This classification is critical for determining surgical strategies, prognostic assessment, and adjuvant therapy planning [1–3]. The Japanese Society of Hepato-Biliary-Pancreatic Surgery (JSHBPS)

provides standardized definitions and classifications for hepatobiliary malignancies, including BDC, to guide clinical management and optimize outcomes [4–6]. The pathogenesis of BDC is multifactorial. Chronic inflammation of the biliary epithelium, such as from primary sclerosing cholangitis, bile duct strictures, and liver fluke infestations, predisposes to malignant transformation. Viral hepatitis B and C infections are significant contributors, particularly in East Asia, where hepatitis prevalence is high[7–9]. Environmental toxins and genetic mutations, including alterations in tumor suppressor genes and oncogenes also been involved [10,11]. BDC predominantly affects older adults, with a higher incidence in males and geographic regions such as Southeast Asia exhibiting elevated prevalence rates.

Clinical presentation is often nonspecific that leads to delayed diagnosis in BDC. Patients commonly present with obstructive jaundice, pruritus, weight loss, and cholangitis. Imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), and endoscopic retrograde cholangiopancreatography (ERCP), assist in tumor localization and staging, yet precise assessment of tumor spread along the bile duct remains difficult [12,13]. Cytological evaluation of ERCP brushings is highly specific but has limited sensitivity (23–56%) due to low cellularity and reactive changes and the most of the diagnosis frequently requires surgical exploration [13,14]. Surgical resection remains the primary curative therapy for BDC. The tumor's location and degree of dissemination dictate the kind and scope of the surgery. iCCA is managed with hepatic resection aiming for negative margins. Perihilar tumors, particularly Bismuth type III and IV lesions, often require major hepatectomy with caudate lobectomy, extrahepatic bile duct excision, and regional lymphadenectomy. Distal cholangiocarcinoma is primarily treated with pancreaticoduodenectomy (PD), whereas selected middle-third common bile duct tumors may undergo segmental bile duct resection to avoid pancreatic transection while achieving oncological safe margins. The proximity of tumors to the hepatic artery, portal vein, and adjacent liver parenchyma complicates surgical planning, and extensive resections are associated with increased postoperative morbidity and mortality [15–17]. Long-term results are still limited despite advancements in surgery. In spite of the aggressive tumor and early recurrence with five-year survival rate after extrahepatic cholangiocarcinoma excision seldom exceeds 50%. Prognostic factors include depth of the tumor invasion, lymph node metastases, perineural invasion, histological differentiation, and resection margin status. Serum carbohydrate antigen 19-9 (CA19-9) serves as an adjunct biomarker, although it is insufficient for early detection [18,19]. Preoperative biliary drainage and stenting are frequently required in obstructive cases but are associated with higher postoperative infectious complications, emphasizing careful patient selection.

Recent advances highlight the significance of molecular profiling in understanding BDC pathogenesis and guiding therapeutic strategies. Analyses of gene expression networks reveal dysregulation of genes involved in cell cycle control, apoptosis, and signaling pathways. Bioinformatics tools such as Cytoscape and datasets from NCBI Gene Expression Omnibus (GEO) facilitate visualization of gene interactions, identification of key regulatory genes, and prediction of survival outcomes. Such analyses can inform targeted drug development and stratify patients based on prognostic risk. Identifying consistently dysregulated genes enables precision medicine approaches and offers potential biomarkers for early diagnosis and therapeutic intervention [20–23]. The concept of “field cancerization” is also relevant in BDC. Chronic bile duct injury and exposure to carcinogens may induce widespread epithelial alterations, predisposing patients to multifocal malignancies. This underscores the importance of thorough intraoperative assessment, histological examination of resection margins, and long-term surveillance for recurrence. Population-level studies utilizing administrative databases, such as the Korean National Health Insurance Service, have provided valuable insights into BDC risk factors, long-term prognosis, and treatment outcomes. These data complement molecular studies by correlating biomarker expression with clinical outcomes.

Advances in high-throughput transcriptomic technologies and computational biology have enabled systematic exploration of the molecular landscape of bile duct cancer. Publicly available gene expression datasets deposited in the NCBI Gene Expression Omnibus (GEO) provide valuable, large-scale data that allow unbiased identification of differentially expressed genes across independent patient cohorts, thereby enhancing the reliability and reproducibility of findings [24,25]. Integrative analysis of multiple GEO datasets helps overcome sample heterogeneity and platform-specific bias commonly encountered in individual studies. Furthermore, protein–protein interaction (PPI) network analysis offers a systems-level perspective by elucidating functional relationships among dysregulated genes and identifying coordinated molecular modules involved in tumorigenesis [26–28]. Within these networks, hub genes defined by their high topological connectivity often represent key regulatory elements that control critical biological processes such as cell cycle regulation, apoptosis, and stress responses [29,30]. Therefore, combining GEO-based transcriptomic analysis with PPI network construction and hub gene identification provides a powerful bioinformatics framework for uncovering potential biomarkers and therapeutic targets in cholangiocarcinoma, bridging clinical observations with molecular mechanisms.

Bile duct cancer is a rare, highly aggressive malignancy arising from complex interactions between chronic inflammation, viral infections, environmental factors, and genetic alterations. Anatomical classification into intrahepatic, perihilar, and distal types guides surgical management and prognosis. While curative resection remains the cornerstone of treatment, survival is limited by aggressive tumor biology and anatomical complexity. Integrating molecular profiling, gene network analysis, and targeted therapy strategies offers promising avenues for improving patient outcomes. Multidisciplinary management—including surgery, molecular diagnostics, supportive care, and palliative interventions—is essential for optimizing survival and quality of life in patients with this challenging malignancy.

Despite advances in imaging and surgical management, the molecular mechanisms underlying cholangiocarcinoma remain incompletely understood, and reliable molecular biomarkers for early diagnosis are limited. Publicly available transcriptomic datasets from the Gene Expression Omnibus (GEO) provide valuable resources for systematically investigating gene expression alterations associated with bile duct cancer. Integrative bioinformatics approaches that combine differential gene expression analysis with protein–protein interaction network construction and hub gene identification enable the discovery of key regulatory genes and biological pathways involved in tumor progression. In this study, we applied a systems biology–based workflow to identify hub genes and functionally enriched pathways associated with cholangiocarcinoma, aiming to provide insights into disease mechanisms and potential molecular targets.

Methods

Bibliographic Search

Gene expression datasets GSE131027 and GSE107754 were obtained from the NCBI Gene Expression Omnibus (GEO) database. GSE131027 is a pan-cancer microarray dataset generated using the Affymetrix Human Genome U133 Plus 2.0 Array, whereas GSE107754 represents an independent dataset containing transcriptomic profiles relevant to biliary tract malignancies. Samples annotated as cholangiocarcinoma were selected from both datasets based on GEO metadata for downstream analysis.

Differential Gene Expression Analysis

Differential gene expression analysis was performed using GEO2R, an interactive web-based tool provided by the NCBI Gene Expression Omnibus. GEO2R applies the limma statistical framework to identify differentially expressed genes between defined sample groups. Samples annotated as cholangiocarcinoma / bile duct cancer were selected based on GEO metadata and compared with corresponding reference samples. Genes with an absolute log2 fold change ($|\log2FC| > 1$ and an adjusted p-value < 0.05 (Benjamini–Hochberg correction, as implemented in GEO2R) were considered statistically significant and used for downstream protein–protein interaction and hub gene analyses

Formation of STRING network using the STRING network analysis and Cytoscape

Protein–protein interaction (PPI) networks were constructed using the STRING database to identify functional associations among differentially expressed genes. The interaction networks were visualized using Cytoscape through the STRINGApp plugin. Nodes represent proteins, and edges represent predicted or experimentally validated protein–protein interactions as curated by STRING. This network-based approach was used to explore molecular interaction patterns relevant to cholangiocarcinoma

Identification of Hub genes using Cytohubba

Hub genes were identified from the STRING-derived protein–protein interaction network using the CytoHubba plugin in Cytoscape. CytoHubba ranks nodes based on topological properties of the network to identify key regulatory genes. In this study, hub genes were ranked using the Degree, DMNC (Density of Maximum Neighborhood Component), and Bottleneck algorithms. Genes consistently ranked among the top candidates were considered hub genes and selected for subsequent functional enrichment analyses.

ClueGo and BINGO

ClueGo is a tool that generates the first binary gene term matrix with their selected associated terms and genes. By looking at this matrix, there is a similarity between the terms to determine the association between the pathways, using the kappa statistics to assess the strength of the terms. Finally, the created networks represent the network using the kappa statistical score level. Using the custom method, the kappa score will automatically adjust based on the positive scale from 0 to 1, which reflects the network connectivity. By gene clustering, ClueGO allows the visualization of the groups based on their network to cluster distribution over the terms. There are two main modes for selecting the set of genes that can be functionally profiled in the default mode, like choosing the nodes from the Cytoscape network or using the other plugins using the MCODE or BINGO in Cytoscape. They are compiled from different sources for a set of upregulated genes in microarray experiments displayed in STRING network analysis. BINGO shows the relevant GO annotations, propagates them through the GO hierarchy, and shows a similar set of parental categories.

Software and Reproducibility

Differential gene expression analysis was performed using GEO2R, a publicly accessible web-based tool provided by the NCBI Gene Expression Omnibus, which applies the limma statistical framework. Protein–protein interaction networks were generated using the STRING database (version 12.0). Network visualization and analysis were conducted using Cytoscape (version 3.10.4). Hub gene identification was carried out using the CytoHubba plugin, and functional enrichment analyses were performed using the ClueGO and BiNGO plugins with default parameters. All datasets used in this study are publicly available, and the analysis can be reproduced by applying the same sample grouping and threshold criteria described above.

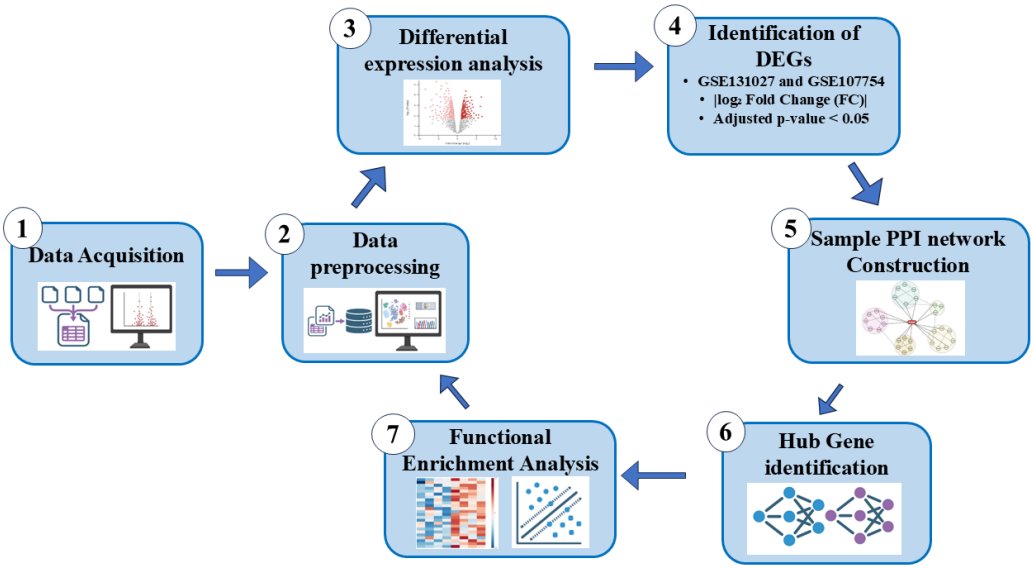

Figure 1 – Integrated Bioinformatics workflow for Cholangiocarcinoma analysis

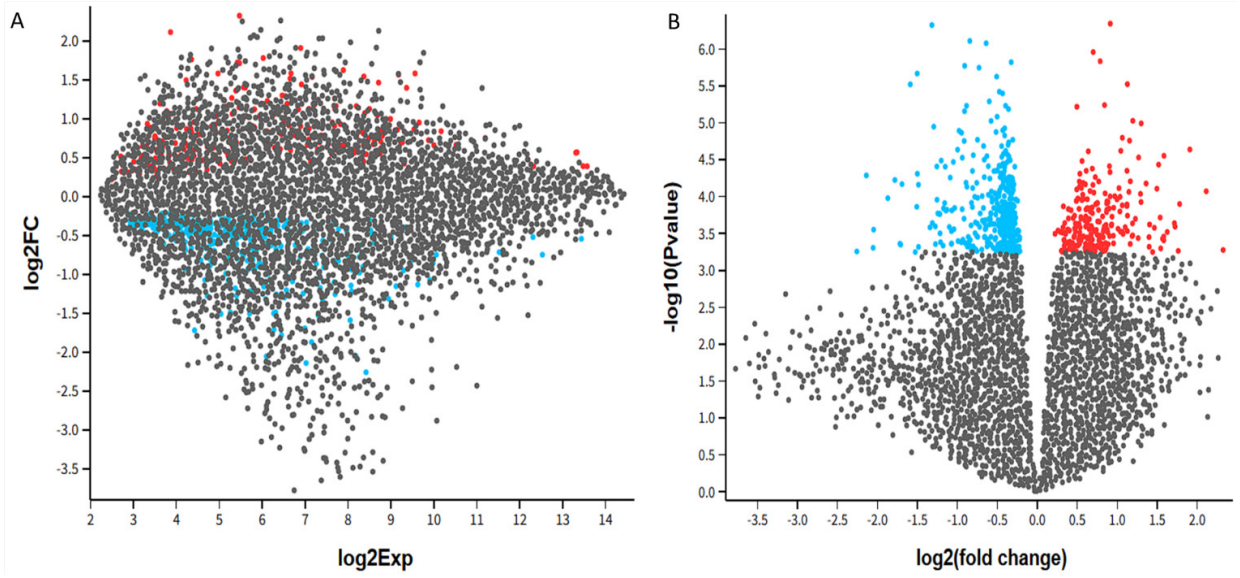

Figure 2 – (a) Mean-difference (MD) plot showing differentially expressed genes (DEGs) between bile duct cancer and control samples. Upregulated genes are indicated in red, while downregulated genes are shown in blue. (b) Volcano plot illustrating the distribution of DEGs based on $\log_2$ fold change and statistical significance. Red dots represent significantly upregulated genes, and blue dots represent significantly downregulated genes, as defined by $|\log_2 FC| > 1.0$ and adjusted p-value < 0.05.

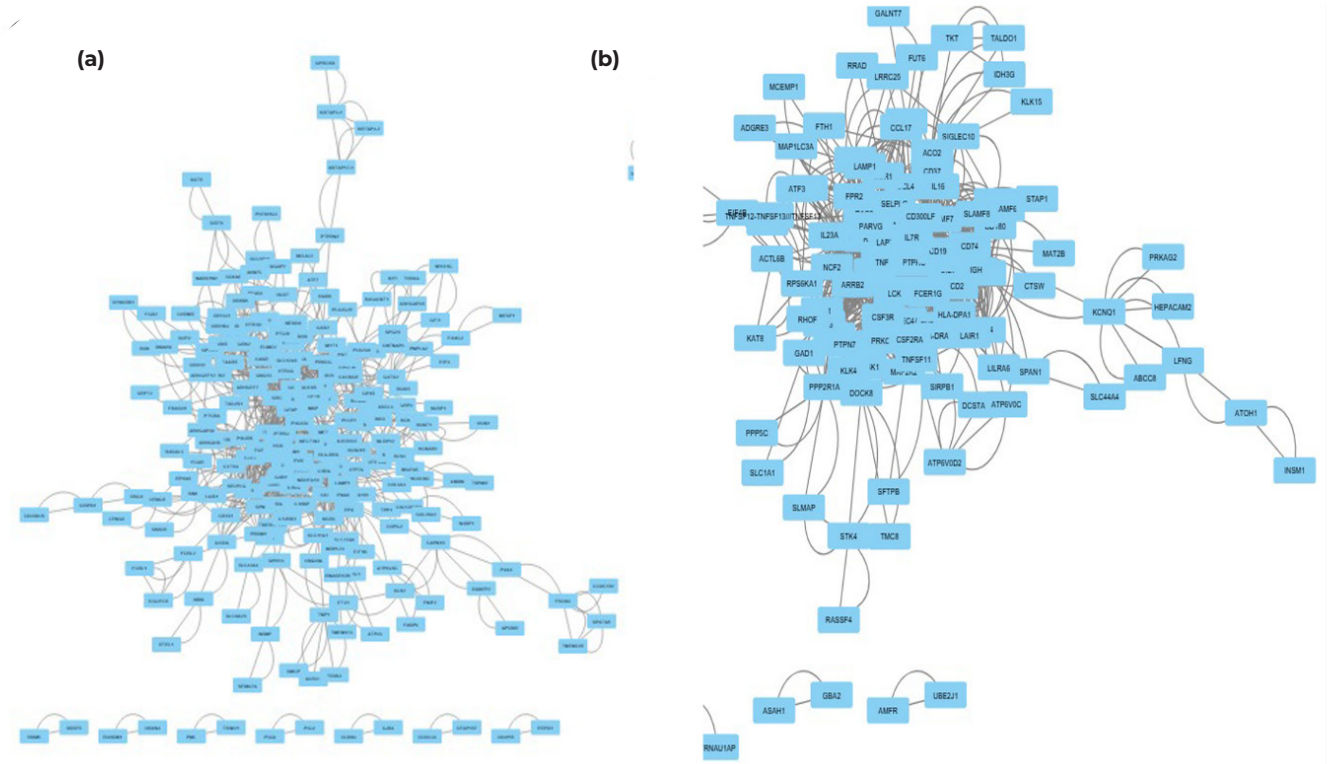

Figure 3 – (a) Protein–protein interaction (PPI) network of upregulated differentially expressed genes (DEGs) from Set 1 (GSE131027) constructed using the STRING database and visualized in Cytoscape. Nodes represent proteins encoded by upregulated genes, while edges indicate predicted or experimentally validated interactions based on STRING confidence scores. (b) Protein–protein interaction (PPI) network of downregulated DEGs from Set 1 (GSE131027) generated using the same STRING-based approach. The network highlights functional connectivity and interaction patterns among downregulated genes associated with bile duct cancer.

Schematic overview of the integrative bioinformatics workflow used to identify differentially expressed genes, hub genes, and enriched pathways in cholangiocarcinoma was shown in Figure 1.

Results

Selection of samples from the selected GEO datasets

Through the NCBI GEO database, they selected two different datasets from the bile duct cancer. After analyzing the STRING data and the samples, they collected the number of samples from sets 1 and 2 from the database.

Identification of DEG

Differential expression analysis revealed substantial transcriptomic alterations across the two analyzed GEO datasets. In both datasets, more than 1,000 genes were identified as significantly differentially expressed, including both upregulated and downregulated genes.In dataset set 1, the mean-difference plot (Figure 2a) illustrates a wide distribution of gene expression

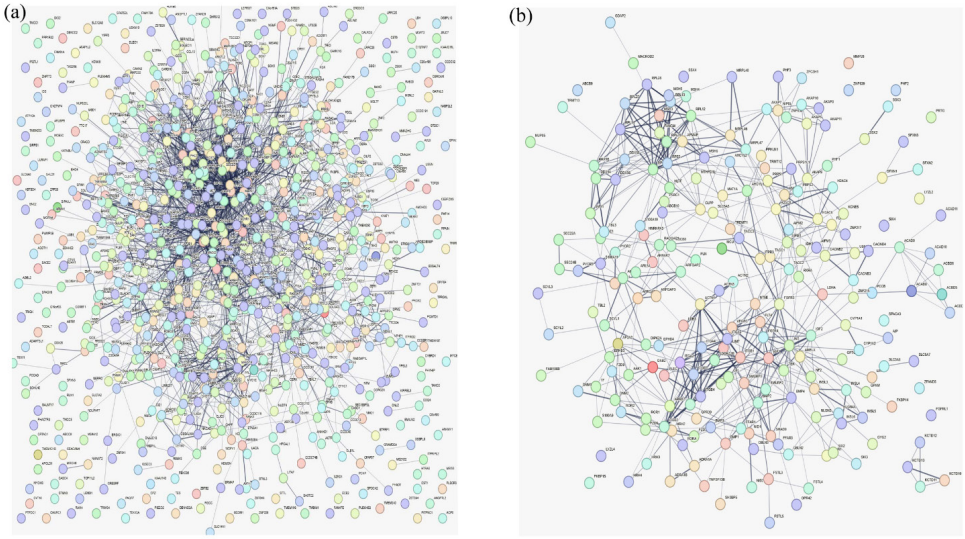

Figure 4 – (a) Protein–protein interaction (PPI) network of upregulated differentially expressed genes (DEGs) from Set 2 (GSE107754) constructed using the STRING database and visualized in Cytoscape. Nodes represent proteins encoded by upregulated genes, while edges denote known and predicted protein–protein interactions based on STRING confidence scores. (b) Protein–protein interaction (PPI) network of downregulated DEGs from Set 2 (GSE107754) generated using the same STRING-based approach. The network illustrates interaction density and functional organization among downregulated genes associated with bile duct cancer.

Table 1 Top 15 Upregulated and Downregulated Genes in Set 1 and Set 2 Based on DEG Analysis

Set 1 (Upregulated)				**Set 1 (Down-regulated)**			
S.NO	NAME	RANK	SCORE	S.NO	NAME	RANK	SCORE
1	TP53	1	138	1	PTPRC	1	88
2	HIST1H3F	2	84	2	TNF	2	76
3	H2AFZ	3	70	3	CD19	3	60
4	FOS	4	68	4	ACTB	4	58
5	POLR2B	5	52	5	FCER1G	4	58
6	CAV1	6	48	6	LCK	6	46
7	CS	7	46	7	IL7R	7	44
8	FEN1	8	44	8	CD2	7	44
9	NUP98	9	42	9	CD74	9	42
10	SMAD3	10	40	10	RAC2	10	38
11	RAN	10	40	11	CCL4	11	34
12	TPI1	12	38	12	HLA-DRA	11	34
13	RAC1	12	38	13	CCRL2	13	30
14	RAE1	14	36	14	CD53	13	30
15	EIF4A3	15	34	15	LAPTM5	15	26
Set 2 (Upregulated)				Set 2 (Down-regulated)			
S.NO	NAME	RANK	SCORE	S.NO	NAME	RANK	SCORE
1	INS	1	94	1	VCAM1	1	10735
2	PTEN	2	84	2	HGF	2	10443
3	CDKN2A	3	54	3	INS	3	9652
4	IGF1	3	54	4	CCR2	4	8478
5	VCAM1	5	50	5	IGF1	5	8307
6	CDK4	5	50	6	ITCB1	6	8119
7	CCR2	7	48	7	NT5E	7	7383
8	HGF	7	48	8	TEK	8	5906
9	RPS2	9	44	9	CD80	9	3425
10	RPL26	9	44	10	PTEN	10	2680
11	RARA	9	44	11	BDNF	11	1745
12	BDNF	9	44	12	FLT4	12	1563
13	NT5E	9	44	13	CCR5	13	1139
14	CD80	14	42	14	TNFSF13B	14	930
15	H2AC8	15	40	15	TNFRSF1B	15	928

The table lists the top 15 upregulated and downregulated genes identified in Set 1 and Set 2, ranked by their respective scores. The scores reflect the significance or impact of each gene based on expression analysis. Notably, genes such as TP53, INS, PTEN, and VCAM1 appear prominently across the datasets, indicating their potential biological relevance in the studied conditions.

changes, with a clear separation between significantly regulated genes and genes showing no significant variation. Dataset set 2 is represented using a volcano plot (Figure 2b), which highlights genes exhibiting both high statistical significance and large magnitude changes in expression.

Overall, the DEG profiles of the two datasets showed comparable patterns of differential regulation, indicating consistent transcriptomic dysregulation associated with cholangiocarcinoma across independent datasets.

Identification of genes from the STRING interactions

To identify key interaction patterns among differentially expressed genes, protein–protein interaction (PPI) networks were constructed using the STRING database for both upregulated and downregulated gene sets from the two datasets. The STRING-derived networks for Set 1 are shown in Figures 3a and3b, while the corresponding upregulated and downregulated networks for Set 2 are illustrated in (Figures 4a and 4b).

For Set 2, the upregulated PPI network exhibited an average node degree of 5.58 and an average local clustering coefficient of 0.3444, whereas the downregulated network showed an average node degree of 3.87 and an average local clustering coefficient of 0.339. In Set 1, the upregulated network had an average node degree of 4.32 and an average local clustering coefficient of 0.385, while the downregulated network demonstrated an average node degree of 3.12 and an average local clustering coefficient of 0.358.

In these networks, nodes represent proteins encoded by individual gene loci, with splice isoforms and post-translationally modified forms combined into a single node. Edges represent functional protein–protein associations curated by STRING, with confidence scores categorized as low (0.150), medium (0.400), high (0.700), and highest (0.900).

Identification of Top hub genes from the STRING networks

Table 1 summarizes the top 15 hub genes identified from the two gene expression datasets (Set 1 and Set 2) based on CytoHubba ranking scores derived from the STRING protein–protein interaction networks. These hub genes represent highly connected nodes within the networks, indicating their central positions in the interaction architecture.

In Set 1, genes such as TP53, HIST1H3F, H2AFZ, FOS, and POLR2B consistently ranked among the top hub genes in both upregulated and downregulated networks, reflecting their high network connectivity across different regulatory contexts. The ranking patterns suggest differential expression dynamics rather than binary on–off regulation.

In Set 2, the distribution of hub genes showed greater separation between upregulated and downregulated categories. Genes including INS, PTEN, CDKN2A, and IGF1 ranked highly in the upregulated network, while VCAM1, HGF, and CCR2 were among the top-ranked genes in the downregulated network.

The STRING-derived interaction networks for Set 1 and Set 2 are illustrated in Figures 4 and 5. The upregulated and downregulated hub gene networks for Set 1 consisted of 15 nodes with 50 and 74 edges, respectively (Figure 5a–b). For Set 2, the upregulated and downregulated networks comprised 15 nodes with 43 and 33 edges, respectively (Figure 6a–b). These hub genes were subsequently used for functional enrichment analysis using ClueGO.

Exploration of pathway and GO functional correlation using Cluego analysis

We have used the ranks of hub genes and placed them in the Cluego analysis to perform the pathways and GO functional correlation. Cytoscape has five databases for pathways analysis: Bio Carta, Elsevier pathway, KEGG, Reactome, and Wikipathway. Where set 1 upregulated and down-regulated clue consists of 34 nodes and 109 edges, as shown in Fig. 7(a). Set 2 upregulated clue consists of 34 nodes and 109 edges, shown in Fig. 7(b). Set 2 down-regulated clue contains 107 nodes and 236 edges, as shown in Fig. 8 (a, b). Fig. 7 (a, b) illustrates functional enrichment networks derived from Gene

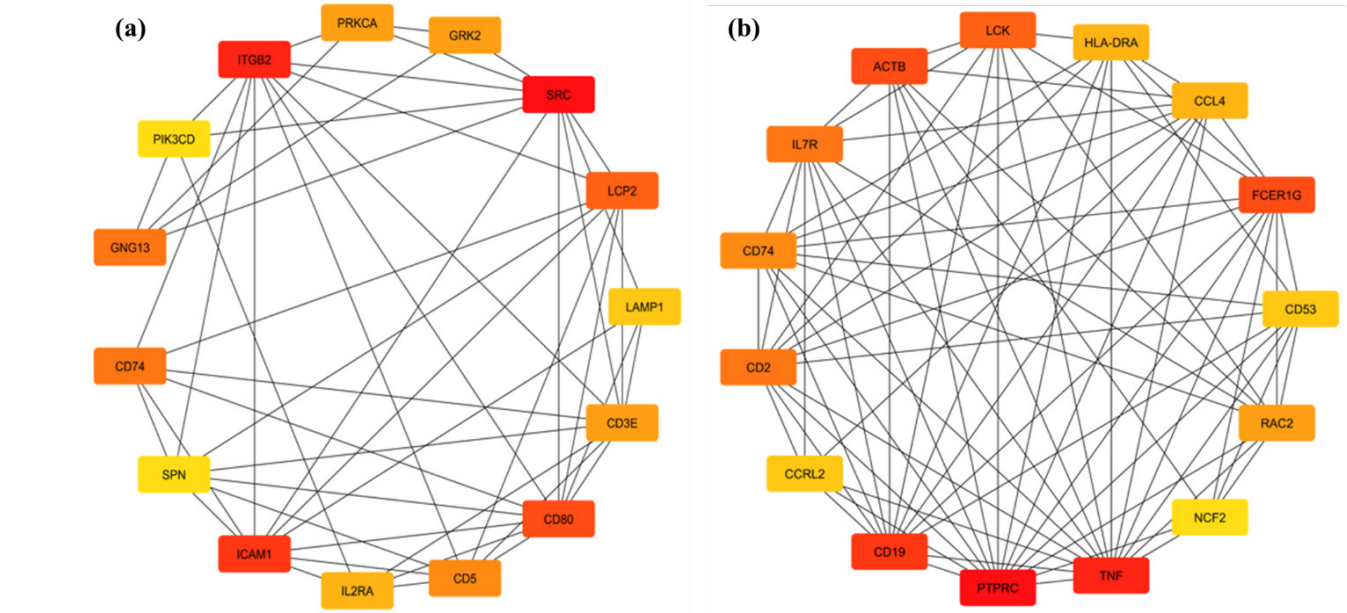

Figure 5 – (a) Hub gene interaction network of upregulated differentially expressed genes (DEGs) from Set 1 (GSE131027) identified using the CytoHubba plugin in Cytoscape. (b) Hub gene interaction network of downregulated DEGs from Set 1 (GSE131027) identified using the same CytoHubba criteria. Highly connected nodes indicate key regulatory genes potentially involved in immune signaling, inflammation, and tumor-related pathways in bile duct cancer.

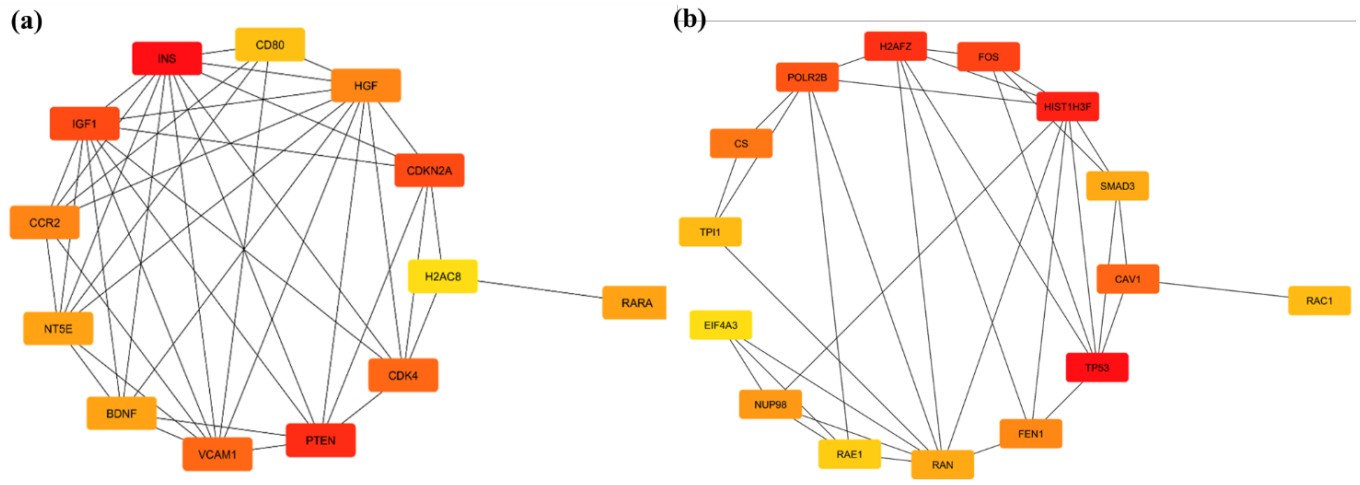

Figure 6 – (a) Hub gene interaction network of upregulated differentially expressed genes (DEGs) from Set 2 (GSE107754) identified using the CytoHubba plugin in Cytoscape. Hub genes were ranked based on network topological parameters, including Degree and BottleNeck algorithms. (b) Hub gene interaction network of downregulated DEGs from Set 2 (GSE107754) identified using the same CytoHubba criteria.

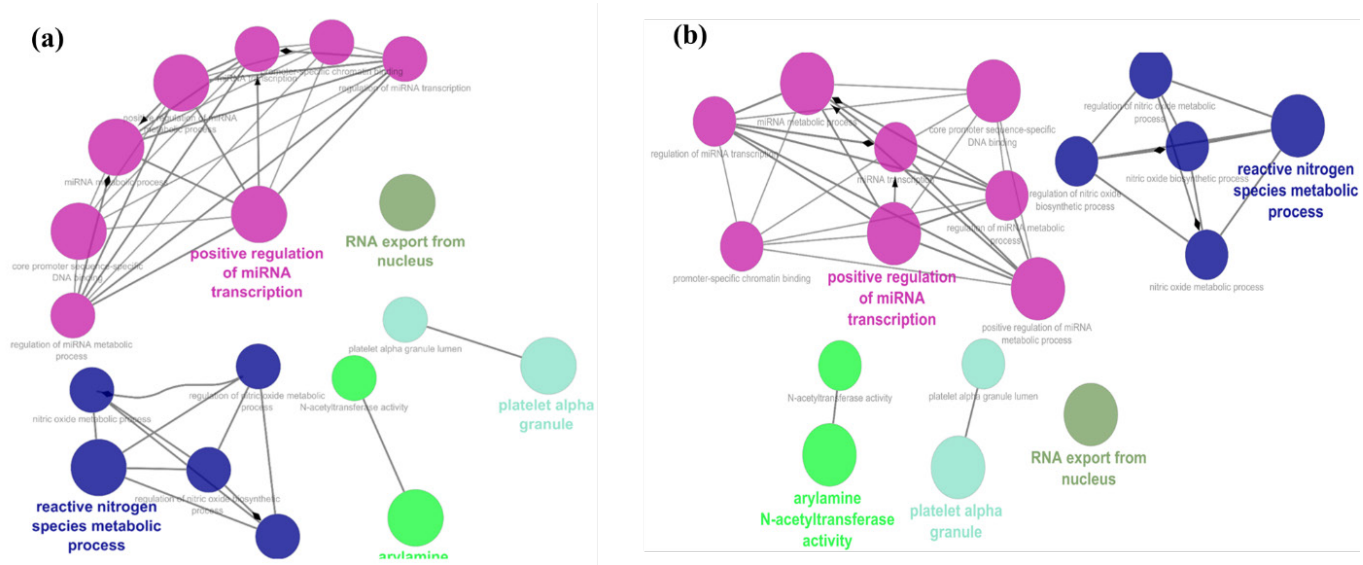

Figure 7 – (a) Set 1 top upregulated hub genes visualized as a functionally grouped network, highlighting significant Gene Ontology (GO) biological processes and molecular functions. (b) Set 1 top downregulated hub genes represented as a ClueGO network showing enriched GO terms predominantly associated metabolic process and cellular components.

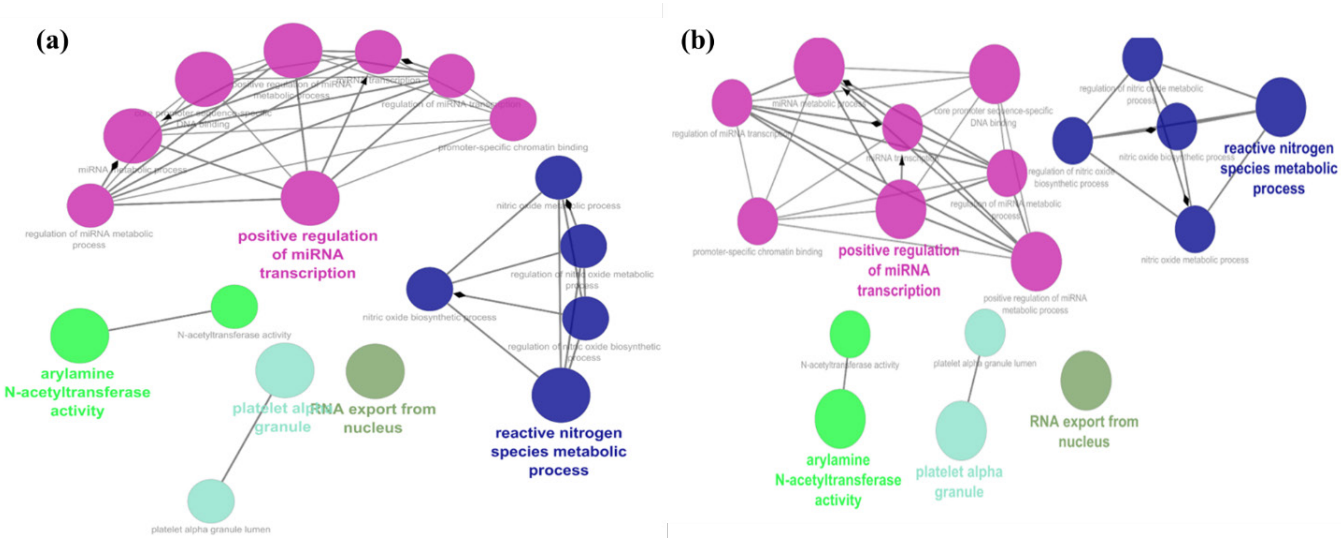

Figure 8 – (a) Set 2 top upregulated hub genes showing significantly enriched Gene Ontology (GO) terms, with dominant functional clusters related to metabolic process. (b) Set 2 top down-regulated hub genes illustrating enriched GO biological processes, molecular functions and cellular components.

Ontology analysis, comparing biological processes for Set 1 in Panel A and Set 2 in Panel B. Variations in gene cluster frequencies are represented by colored nodes categorized by functional features, including miRNA transcription (magenta),

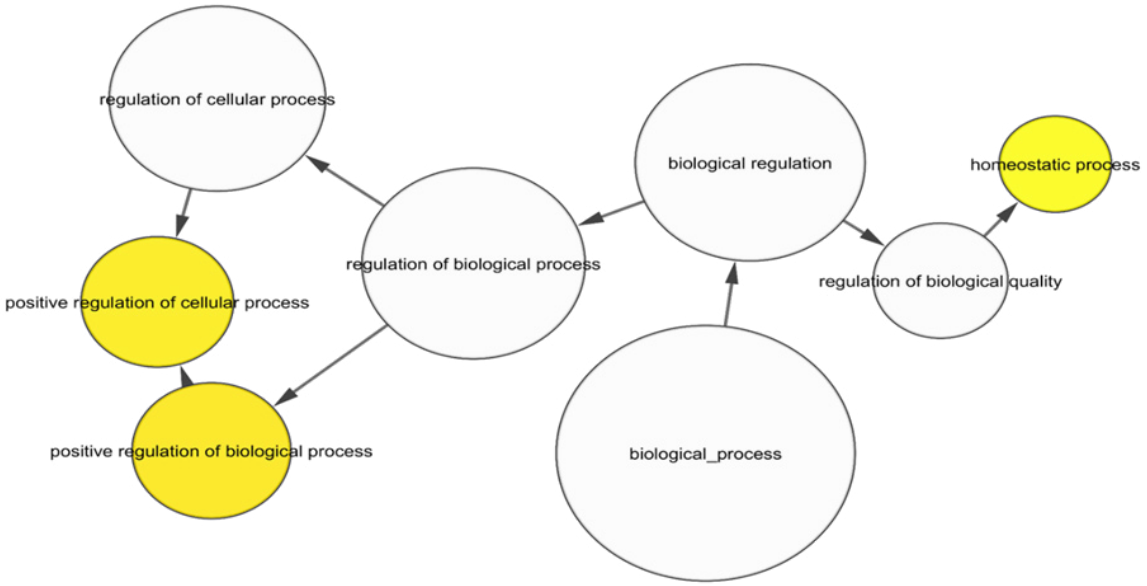

Figure 9 – The network depicts significantly overrepresented Gene Ontology (GO) biological process terms associated with Set 2 upregulated hub genes. Major enriched categories include biological process, biological regulation, regulation of biological process, regulation of cellular process, and homeostatic process.

reactive nitrogen species metabolism (dark blue), RNA export (gray-green), platelet granules (aqua), and N-acetyltransferase activity (green). The network lets users discern that nodes expand proportionately to word relevance or gene count while edges connect functional and hierarchical relationships among GO terms. Panel A's primary cluster (Set 1) centers on "positive regulation of miRNA transcription," illustrating a significant amplification of miRNA regulatory mechanisms. The procedures activate gene silencing post-transcription and regulate gene expression in response to cellular stimuli. A considerable quantity of "reactive nitrogen species" molecular constituents arises alongside "positive regulation of miRNA transcription" as primary clusters indicative of active oxidative/nitrosative stress responses. Despite the limited connections between RNA export from the nucleus and the actions of platelet alpha granules and arylamine N-acetyltransferase, they signify distinct biological processes regulated within this dataset.

The "positive regulation of miRNA transcription" hub is central in the results of Set 2, mirroring its position in Set 1, indicating the continued significance of miRNA-mediated gene regulation across datasets. The "reactive nitrogen species metabolic process" cluster in Set 2 functions as an integrated system with numerous interrelated processes, in contrast to Set 1, indicating its potential as a chronic or systemic oxidative stress response. The "arylamine N-acetyltransferase activity" node is more prominent in Set 2, suggesting its role in detoxification or drug metabolism under specific testing settings. The "platelet alpha granule" pathway persists in analysis, exhibiting varying degrees of significance relative to the initial configuration, while "RNA export from nucleus" functions similarly with negligible alterations.

Fig. 8 (a, b) illustrates GO term-based functional enrichment network maps for Set 1 (Panel A) and Set 2 (Panel B) gene groups, highlighting the biological processes and molecular activities associated with differentially expressed genes. Colored nodes depict distinct GO terms, while edges illustrate functional linkages and hierarchical connections between these terms. The illustration depicts various biological activities represented by colored pathways: magenta for miRNA transcription, dark blue for reactive nitrogen species metabolism, green for N-acetyltransferase activity, cyan for platelet alpha granule, and olive for RNA export from the nucleus.

The primary cluster identified in Panel A (Set 1) emphasizes "positive regulation of miRNA transcription," highlighting the importance of miRNA synthesis in post-transcriptional regulation for the examined biological situation. Set 1 genes significantly influence RNA-mediated gene expression modulation, which appears to be triggered by cellular stress signals and differentiation cues. The phrases "arylamine N-acetyltransferase activity," "platelet alpha granule," and "RNA export from nucleus" represent disparate categories, signifying varied physiological activities encompassing metabolic detoxification, vesicle formation, and mRNA transport.

This panel illustrates analogous functional groupings that exhibit changes in network architectural relationships relative to the initial set. A prominent cluster of "positive regulation of miRNA transcription" links both Set 1 and Set 2. However, the "reactive nitrogen species metabolic process" exhibits more significant divergence from the central network in this panel. The setup of Set 2 displays indicators for specific oxidative processes. The N-acetyltransferase activity and platelet-related granule pathways are present in Set 2, although their connection patterns vary, indicating distinct functional regulators that modify their activation patterns.

Determination of gene ontology using BINGO

The GO enrichment analysis via BiNGO reveals biological processes associated with the upregulated genes in set 2, as illustrated in Fig. 9. The GO network, utilizing STRING network data, illustrates functional associations among differentially expressed genes throughout biological pathways. Each network node displays a GO word that connects to several associated genes, with node size and yellow hue indicating statistically significant relationships. The hierarchical structure in the network is illustrated by arrows indicating that subordinate parts transmit data to overarching parent portions. Several significant GO keywords are prominent in the network, including "positive regulation of the biological process," "positive regulation of the

cellular process," and "homeostatic process." The GO keywords suggest that the genes in set 2 facilitate biological functions and sustain operational stability within cellular systems. The relationships between the nodes "regulation of biological process" and "biological regulation" culminate in the broader category "biological process," signifying their significant regulatory mechanisms inside cellular contexts that encompass the genes from set 2. The enrichment analysis failed to uncover any GO pathways associated with the downregulated genes in set 2. The deficiency of genetic annotations suggests that many down-regulated genes and their related proteins remain unrecognized, as no existing studies link them to biological processes. These genetic components suggest potential discoveries as they exist outside traditional databases despite their possible significance in context-specific biological responses. The Functional analysis pathway table was given10 in Supplementary tables 1 and 2.

Identification of Hub Genes from DEG and PPI Network Analysis

Differential expression analysis of the two GEO datasets (Set 1 and Set 2) identified more than 1,000 significantly differentially expressed genes (DEGs) in each dataset based on the applied statistical thresholds. To further prioritize genes with potential regulatory importance, protein–protein interaction (PPI) networks were constructed using the STRING database, and hub genes were identified using the CytoHubba plugin in Cytoscape. Table 1 summarizes the top 15 upregulated and downregulated hub genes identified from each dataset based on network topological scoring. In Set 1, TP53 emerged as the highest-ranked hub gene among upregulated genes, followed by HIST1H3F, H2AFZ, FOS, POLR2B, and CAV1, indicating their strong connectivity and central roles within the interaction network. These genes are primarily associated with transcriptional regulation, chromatin organization, stress response, and signal transduction. The downregulated hub genes in Set 1 included immune- and signaling-related genes such as PTPRC, TNF, CD19, and LCK, suggesting alterations in immune-associated pathways in bile duct cancer.

In Set 2, hub gene analysis revealed prominent enrichment of genes involved in metabolic regulation and growth factor signaling. INS, PTEN, CDKN2A, and IGF1 were identified among the top upregulated hub genes, while VCAM1, HGF, CCR2, and NT5E were among the most strongly downregulated genes. Notably, some genes such as INS, IGF1, PTEN, and VCAM1 appeared in both upregulated and downregulated lists, reflecting context-dependent regulation across different sample groups within the dataset.

Overall, the hub gene profiles differed between Set 1 and Set 2, likely reflecting differences in dataset composition and biological context. Nevertheless, several genes with known roles in tumor suppression, epigenetic regulation, and oncogenic signaling consistently ranked highly, supporting their potential relevance in cholangiocarcinoma biology and justifying their further functional interpretation.

Discussion

In the present study, an integrative bioinformatics framework was applied to identify key hub genes and functionally enriched biological pathways associated with cholangiocarcinoma using publicly available Gene Expression Omnibus (GEO) datasets. By integrating differential gene expression analysis with protein–protein interaction (PPI) network construction and hub gene prioritization, this study aimed to uncover genes with potential regulatory significance in bile duct cancer pathogenesis.

Among the identified hub genes, TP53 consistently ranked as the most central node across both datasets. TP53 is a well-established tumor suppressor gene involved in cell cycle regulation, apoptosis, and maintenance of genomic stability, and its dysregulation has been frequently reported

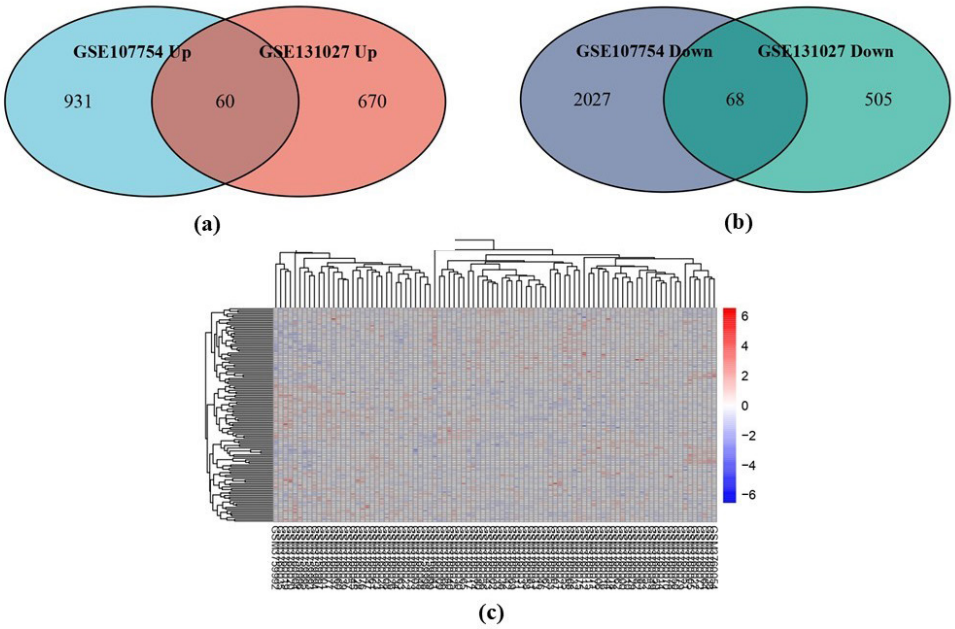

Figure 10 – (a) Venn diagram showing the overlap of upregulated DEGs between GSE107754 and GSE131027, with 60 genes commonly upregulated in both datasets. (b) Venn diagram illustrating the overlap of downregulated DEGs between GSE107754 and GSE131027, identifying 68 commonly downregulated genes. (c) Hierarchical clustering heatmap of the common DEGs, displaying their expression patterns across samples; red indicates higher expression levels and blue indicates lower expression levels, as shown by the color scale.

in cholangiocarcinoma. The prominent network centrality of TP53 observed in this study further supports its pivotal role in bile duct cancer pathogenesis. Previous studies have reported that alterations in TP53 are associated with aggressive tumor behavior, poor prognosis, and increased tumor mutation burden in intrahepatic cholangiocarcinoma, further highlighting its relevance in bile duct cancer biology [31]

In addition to TP53, SMAD3 emerged as a highly connected hub gene, underscoring the importance of the transforming growth factor-β (TGF-β) signaling pathway in cholangiocarcinoma. SMAD3 functions as a key intracellular mediator of TGF-β signaling and has been shown to promote epithelial–mesenchymal transition (EMT), tumor invasiveness, and metastatic dissemination in bile duct cancer. Experimental studies have demonstrated that suppression of SMAD3 phosphorylation attenuates EMT and inhibits tumor progression, highlighting the oncogenic role of aberrant TGF-β1/SMAD3 activation in cholangiocarcinoma [32]. The identification of SMAD3 as a hub gene in our network analysis further supports its biological relevance in disease progression.

We propose that the HSP module maintains proteostasis by improving the structural integrity of proteins necessary to oncogenic proteins. Many genes are upregulated and downregulated. In this study, HSP significantly contributes to oncogenesis via homologous recombination repair. In our study, several key genes implicated in thermoresistance were marked overexpressed in tumors of different tissues. Moreover, the expression patterns of the thermogenesis like TP53, HIST1H3F, H2AFZ, FOS, POLR2B, CAV1, CS, FEN1, NUP98, SMAD3, RAN, TPI1, RAC1, RAE1, EIF4A3 these are the top upregulated genes which shows the positively correlated in diverse types of the bile duct cancer [33–42]. Histone-related genes such as HIST1H3F and H2AFZ were also identified among the hub genes, indicating that alterations in chromatin organization and transcriptional regulation may contribute to the abnormal gene expression landscape observed in cholangiocarcinoma. These findings are consistent with previous reports highlighting the importance of epigenetic dysregulation in biliary tract cancers.

Caveolin-1 (CAV1) was another prominent hub gene identified in this study, highlighting the contribution of tumor–stromal interactions to cholangiocarcinoma progression. Previous investigations have demonstrated that elevated CAV1 expression in cancer-associated fibroblasts correlates with poor overall and recurrence-free survival in intrahepatic cholangiocarcinoma. Mechanistically, CAV1-positive fibroblasts promote an immunosuppressive tumor microenvironment by enhancing regulatory T-cell infiltration and impairing effective anti-tumor immune responses. These findings suggest that CAV1 contributes to disease aggressiveness through modulation of stromal signaling and immune regulation, supporting its potential value as both a prognostic biomarker and a therapeutic target.

The transcription factor FOS, a core component of the activator protein-1 (AP-1) complex, was also identified as a central hub gene. AP-1 regulates gene expression programs involved in cell proliferation, apoptosis, stress responses, and oncogenic transformation. Integrative epigenomic profiling studies have demonstrated that AP-1 binding sites are enriched in regions exhibiting altered histone modifications in intrahepatic cholangiocarcinoma, indicating a critical role for FOS-mediated transcriptional regulation. Dysregulation of AP-1 components, including FOS, has been associated with tumor progression and unfavorable clinical outcomes, supporting its biological and prognostic relevance in bile duct cancer [43].

In one dataset, genes involved in metabolic and growth factor signaling, including INS, IGF1, and PTEN, were identified as hub genes. However, as this dataset represents a broader pan-cancer cohort, the involvement of these genes may reflect tissue-contextual or systemic regulatory effects rather than cholangiocarcinoma-specific oncogenic drivers. Consequently, these findings should be interpreted with caution and require further experimental validation to establish their disease-specific relevance.

Functional enrichment analysis using ClueGO and BiNGO revealed that the identified hub genes were significantly associated with biological processes related to transcriptional regulation, oxidative and nitrosative stress responses, RNA processing, and intracellular transport. These processes have been previously implicated in cancer initiation and progression and support the biological plausibility of the identified gene networks. Overall, the integration of network-based hub gene identification with functional enrichment analysis provides a systems-level perspective on the molecular mechanisms underlying cholangiocarcinoma and highlights potential targets for future therapeutic exploration. Hierarchical clustering heatmap analysis of the common differentially expressed genes (DEGs) demonstrated distinct expression profiles between the sample groups. The consistent clustering pattern indicates robust differential expression and supports the reliability of the identified DEGs across datasets was shown in Fig. 10.

The hub genes identified in this study provide important biological insights into the molecular mechanisms underlying bile duct cancer progression. TP53 emerged as the most significant hub gene, reaffirming its critical role in regulating cell cycle control, DNA repair, and apoptosis, with its dysregulation contributing to genomic instability in cholangiocarcinoma [10,11]. Epigenetic regulators such as HIST1H3F and H2AFZ suggest that chromatin remodeling and transcriptional reprogramming are central to tumor development [44,45]. FOS links chronic inflammation to oncogenic signaling, while POLR2B reflects increased transcriptional activity in malignant cells. Additionally, CAV1 and SMAD3 implicate tumor–stroma interactions, TGF-β signaling, and epithelial–mesenchymal transition in disease progression [46–49].

Our study highlights key differentially expressed and hub genes potentially involved in bile duct cancer progression, providing insights into its molecular mechanisms. By integrating two independent GEO datasets, we enhanced the robustness of our findings. However, the sample size remains limited, which may reduce statistical power and limit representation of molecular heterogeneity across diverse patient populations and clinical subtypes. The analyses were entirely computational, and experimental validation of the identified genes was not performed. Therefore, the biological functions and clinical relevance of these hub genes require confirmation through in vitro and in vivo experiments, including quantitative PCR, protein expression studies, and functional assays.

Protein–protein interaction networks were constructed using the STRING database, which combines experimentally validated and predicted interactions. While this approach offers a systems-level perspective, it may introduce biases due to incomplete annotations, potential false positives, and lack of tissue- or disease-specific context, warranting cautious interpretation of network relationships. Furthermore, clinical

variables such as tumor stage, histological subtype, treatment status, and patient outcomes were unavailable, limiting correlations between gene expression and clinical features. Future studies incorporating larger sample sizes, detailed clinical data, and experimental validation are essential to confirm the diagnostic and therapeutic potential of the identified hub genes and facilitate their translation into clinical applications.

Limitations

This study is limited by the use of publicly available datasets with relatively small sample sizes and by the absence of experimental validation. Additionally, GSE131027 represents a pan-cancer cohort; although bile duct cancer samples were extracted for analysis, residual biological heterogeneity may influence the results. Furthermore, protein–protein interaction networks derived from STRING represent predicted and curated interactions that may not fully capture tissue-specific biological contexts. Future studies incorporating independent cohorts and experimental validation are required to confirm these findings.

Conclusion

In conclusion, this study applied an integrated bioinformatics framework to investigate the molecular landscape of cholangiocarcinoma using publicly available GEO transcriptomic datasets. By combining differential gene expression analysis, protein–protein interaction network construction, hub gene ranking, and functional enrichment analysis, we identified several highly connected genes and biologically relevant pathways associated with bile duct cancer. Network-based analysis consistently highlighted TP53 as a central hub gene, along with other key regulators such as SMAD3, CAV1, FOS, HIST1H3F, H2AFZ, and POLR2B, suggesting coordinated dysregulation of pathways related to transcriptional control, signaling, and chromatin organization. Functional enrichment further supported the involvement of processes linked to gene regulation, oxidative and nitrosative stress responses, and RNA processing in cholangiocarcinoma. Although the findings are derived from in silico analyses and require experimental validation, this study provides a systems-level perspective on the molecular mechanisms underlying bile duct cancer. The identified hub genes and pathways offer a valuable resource for future experimental studies and may contribute to improved understanding of cholangiocarcinoma biology.

Supplementary materials

The Supplementary information includes tables:

• Supplementary Table 1. KEGG pathway enrichment analysis of differentially expressed genes (DEGs);

• Supplementary Table 2. Gene Ontology (GO) Biological Process enrichment analysis of differentially expressed genes (DEGs).

This supplemental materials have been provided by the authors to give readers additional information about their work.

The file can be accessed using: https://www.editorialpark.com/download/article-supp/758/Supplementary-data.docx.

Author Contributions: Conceptualization, M. K.; methodology, R. N., P. H. V. and M. K; validation, M. K. and P. H. V; formal analysis, R. N.; investigation, P. H. V.; resources, P. H. V.; data curation, R. N., P. H. V.; writing – original draft preparation, P. H. V. and R. N.; writing – review and editing, R. N. and P. H. V.; visualization, R. N. and P. H. V.; supervision, M. K.; project administration, M. K. All authors have read and agreed to the published version of the manuscript.

Disclosures: The authors have no conflicts of interest.

Acknowledgments: Authors are thankful to Saveetha School of Engineering, SIMATS, India for providing support to carry out the work.

Funding: None.

Data availability statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Artificial Intelligence (AI) Disclosure Statement: The authors declare no AI Tools used for preparation of this work.

References

1. Chun K. Recent Classifications of the Common Bile Duct Injury. Korean J. Hepato-Biliary-Pancreatic Surg. 2014; 18: 69, https://doi.org/10.14701/kjhbps.2014.18.3.69.
2. Joo I, Lee JM. Imaging Bile Duct Tumors: Pathologic Concepts, Classification, and Early Tumor Detection. Abdom. Imaging. 2013; 38: 1334–1350. https://doi.org/10.1007/s00261-013-0027-3.
3. Lee SH, Song SY. Recent Advancement in Diagnosis of Biliary Tract Cancer through Pathological and Molecular Classifications. Cancers (Basel). 2024; 16: 1761. https://doi.org/10.3390/cancers16091761.
4. Oba A, Del Chiaro M, Satoi S, Kim S, Takahashi H, Yu J, Hioki M, Tanaka M, Kato Y, Ariake K, Wu YHA, Inoue Y, Takahashi Y, Hackert T, Wolfgang CL, Besselink MG, Schulick RD, Nagakawa Y, Isaji S, Tsuchida A, Endo I. New Criteria of Resectability for Pancreatic Cancer: A Position Paper by the Japanese Society of Hepato-Biliary-Pancreatic Surgery (JSHBPS). J. Hepatobiliary. Pancreat. Sci. 2022; 29: 725–731. https://doi.org/10.1002/jhbp.1049.
5. Otsubo T, Kobayashi S, Sano K, Misawa T, Ota T, Katagiri S, Yanaga K, Yamaue H, Kokudo N, Unno M, Fujimoto J, Miura F, Miyazaki M, Yamamoto M. Safety-Related Outcomes of the Japanese Society of Hepato-Biliary-Pancreatic Surgery Board Certification System for Expert Surgeons. J. Hepatobiliary. Pancreat. Sci. 2017; 24: 252–261. https://doi.org/10.1002/jhbp.444.
6. Kuroda S, Kobayashi T, Hatano E, Kubo S, Endo I, Ohdan H. Questionnaire on the Surgical Indications for Intrahepatic Cholangiocarcinoma Administered to Japanese Board-certified Expert Hepatobiliary and Pancreatic Surgeons and Instructors. J. Hepatobiliary. Pancreat. Sci.

2025; 32: 179–193. https://doi.org/10.1002/jhbp.12108.
7. Baison GN, Bonds MM, Helton WS, Kozarek RA. Choledochal Cysts: Similarities and Differences between Asian and Western Countries. World J. Gastroenterol. 2019; 25: 3334–3343. https://doi.org/10.3748/wjg.v25.i26.3334.
8. Oze I, Ito H, Koyanagi YN, Abe SK, Rahman MS, Islam MR, Saito E, Gupta PC, Sawada N, Tamakoshi A, et al. Obesity Is Associated with Biliary Tract Cancer Mortality and Incidence: A Pooled Analysis of 21 Cohort Studies in the Asia Cohort Consortium. Int. J. Cancer. 2024; 154: 1174–1190. https://doi.org/10.1002/ijc.34794.
9. Keane MG, Horsfall L, Rait G, Pereira SP. A Case–Control Study Comparing the Incidence of Early Symptoms in Pancreatic and Biliary Tract Cancer. BMJ Open. 2014; 4: e005720. https://doi.org/10.1136/bmjopen-2014-005720.
10. Guo L, Zhou F, Liu H, Kou X, Zhang H, Chen X, Qiu J. Genomic Mutation Characteristics and Prognosis of Biliary Tract Cancer. Am. J. Transl. Res. 2022; 14: 4990–5002.
11. Wardell CP, Fujita M, Yamada T, Simbolo M, Fassan M, Karlic R, Polak P, Kim J, Hatanaka Y, Maejima K et al. Genomic Characterization of Biliary Tract Cancers Identifies Driver Genes and Predisposing Mutations. J. Hepatol. 2018; 68: 959–969. https://doi.org/10.1016/j.jhep.2018.01.009.
12. Tompkins RK, Saunders K, Roslyn JJ, Longmire WP. Changing Patterns in Diagnosis and Management of Bile Duct Cancer. Ann. Surg. 1990; 211: 614–620, discussion 620–621.
13. Tanaka K, Kida M. Role of endoscopy in screening of early pancreatic cancer and bile duct cancer. Dig. Endosc. 2009; 21. https://doi.org/10.1111/j.1443-1661.2009.00856.x.
14. Ryan ME. Cytologic Brushings of Ductal Lesions during ERCP. Gastrointest. Endosc. 1991; 37: 139–142. https://doi.org/10.1016/S0016-5107(91)70671-8.
15. Seyama Y, Makuuchi M. Current Surgical Treatment for Bile Duct Cancer. World J. Gastroenterol. 2007; 13: 1505–1515. https://doi.org/10.3748/wjg.v13.i10.1505.
16. Cowzer D, Harding, JJ. Advanced Bile Duct Cancers: A Focused Review on Current and Emerging Systemic Treatments. Cancers (Basel). 2022; 14: 1800. https://doi.org/10.3390/cancers14071800.
17. Huguet JM, Lobo M, Labrador JM, Boix C, Albert C, Ferrer-Barceló L, Durá AB, Suárez P, Iranzo I, Gil-Raga M, Burgos CB, Sempere J. Diagnostic-Therapeutic Management of Bile Duct Cancer. World J. Clin. Cases. 2019; 7: 1732–1752. https://doi.org/10.12998/wjcc.v7.i14.1732.
18. Min L, Ziyu D, Xiaofei Z, Shunhe X, Bolin W. Analysis of Levels and Clinical Value of CA19-9, NLR and SIRI in Patients with Pancreatic Cancer with Different Clinical Features. Cell. Mol. Biol. 2022; 67: 302–308. https://doi.org/10.14715/cmb/2021.67.4.41.
19. Liu J, Gao J, Du Y, Li Z, Ren Y, Gu J, Wang X, Gong Y, Wang W, Kong X. Combination of Plasma MicroRNAs with Serum CA19-9 for Early Detection of Pancreatic Cancer. Int. J. Cancer. 2012; 131: 683–691. https://doi.org/10.1002/ijc.26422.
20. Liao B. Research on the Factors That Affecting the Occurrence of Gastric Cancer Based on NCBI Gene Expression Omnibus Database. AIP Conference Proceedings. 2020; 2208 (1): 020007. https://doi.org/10.1063/5.0000016.
21. Yu-jing T, Wen-jing T, Biao T. Integrated Analysis of Hub Genes and Pathways In Esophageal Carcinoma Based on NCBI's Gene Expression Omnibus (GEO) Database: A Bioinformatics Analysis. Med. Sci. Monit. 2020; 26. https://doi.org/10.12659/MSM.923934.
22. Kumar SU, Kumar DT, Siva R, Doss CGP, Zayed H. Integrative Bioinformatics Approaches to Map Potential Novel Genes and Pathways Involved in Ovarian Cancer. Front. Bioeng. Biotechnol. 2019; 7. https://doi.org/10.3389/fbioe.2019.00391.
23. Yang Y, Qi S, Shi C, Han X, Yu J, Zhang L, Qin S, Gao Y. Identification of Metastasis and Prognosis-Associated Genes for Serous Ovarian Cancer. Biosci. Rep. 2020; 40. https://doi.org/10.1042/BSR20194324.
24. Barrett T, Edgar R. Mining microarray data at NCBI's Gene Expression Omnibus (GEO). Methods Mol Biol. 2006; 338: 175–190. https://doi.org/10.1385/1-59745-097-9:175.
25. Clough E, Barrett T. The Gene Expression Omnibus Database. Methods Mol Biol. 2016; 1418: 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5.
26. Vella D, Marini S, Vitali F, Di Silvestre D, Mauri G, Bellazzi R. MTGO: PPI Network Analysis Via Topological and Functional Module Identification. Sci. Rep. 2018; 8: 5499, https://doi.org/10.1038/s41598-018-23672-0.
27. Murakami Y, Tripathi LP, Prathipati P, Mizuguchi K. Network Analysis and in Silico Prediction of Protein–Protein Interactions with Applications in Drug Discovery. Curr. Opin. Struct. Biol. 2017; 44: 134–142. https://doi.org/10.1016/j.sbi.2017.02.005.
28. Tomkins JE, Manzoni C. Advances in Protein-Protein Interaction Network Analysis for Parkinson's Disease. Neurobiol. Dis. 2021; 155: 105395. https://doi.org/10.1016/j.nbd.2021.105395.
29. Li T, Gao X, Han L, Yu J, Li H. Identification of Hub Genes with Prognostic Values in Gastric Cancer by Bioinformatics Analysis. World J. Surg. Oncol. 2018; 16: 114. https://doi.org/10.1186/s12957-018-1409-3.
30. Lv J, Li L. Hub Genes and Key Pathway Identification in Colorectal Cancer Based on Bioinformatic Analysis. Biomed Res. Int. 2019; 2019: 1–13. https://doi.org/10.1155/2019/1545680.
31. Guo C, Liu Z, Yu Y, Chen Y, Liu H, Guo Y, Peng Z, Cai G, Hua Z, Han X, Li Z. TP53 /KRAS Co-Mutations Create Divergent Prognosis Signatures in Intrahepatic Cholangiocarcinoma. Front. Genet. 2022; 13. https://doi.org/10.3389/fgene.2022.844800.
32. Deng L, Bao W, Zhang B, Zhang S, Chen Z, Zhu X, He B, Wu L, Chen X, Deng T, Chen B, Yu Z, Wang Y, Chen G. AZGP1 Activation by Lenvatinib Suppresses Intrahepatic Cholangiocarcinoma Epithelial-Mesenchymal Transition through the TGF-B1/Smad3 Pathway. Cell Death Dis. 2023; 14: 590. https://doi.org/10.1038/s41419-023-06092-5.
33. Zou W, Zhang Q, Sun R, Li X, He S. Study on TFF1 and PALB2 Gene Variants Associated with Gastric Carcinoma Risk in the Chinese Han Population. Cancer Epidemiol. 2023; 83: 102333. https://doi.org/10.1016/j.canep.2023.102333.
34. Ghojazadeh M, Somi MH, Naseri A, Salehi-Pourmehr H, Hassannezhad S, Hajikamanaj Olia A, Kafshdouz L, Nikniaz Z. Systematic Review and Meta-Analysis of TP53, HER2/ERBB2, KRAS, APC, and PIK3CA Genes Expression Pattern in Gastric Cancer. Middle East J. Dig. Dis. 2022; 14: 335–345, https://doi.org/10.34172/mejdd.2022.292.
35. Wu Y, Zhao H. Circ_0074027 Binds to EIF4A3 and Promotes Gastric Cancer Progression. Oncol. Lett. 2021; 22: 704, https://doi.org/10.3892/ol.2021.12965.
36. Liang X, Chen W, Shi H, Gu X, Li Y, Qi Y, Xu K, Zhao A, Liu J. PTBP3 Contributes to the Metastasis of Gastric Cancer by Mediating CAV1 Alternative Splicing. Cell Death Dis. 2018; 9: 569, https://doi.org/10.1038/s41419-018-0608-8.

37. Peng J, Liang S, Li L. SFRP1 Exerts Effects on Gastric Cancer Cells through GSK3β/Rac1-mediated Restraint of TGFβ/Smad3 Signaling. Oncol. Rep. 2018. https://doi.org/10.3892/or.2018.6838.
38. Fenoglio-Preiser CM, Wang J, Stemmermann GN, Noffsinger A. TP53 and Gastric Carcinoma: A Review. Hum. Mutat. 2003; 21: 258–270, https://doi.org/10.1002/humu.10180.
39. Zhou Q, Zheng X, Chen L, Xu B, Yang X, Jiang J, Wu C. Smad2/3/4 Pathway Contributes to TGF-β-Induced MiRNA-181b Expression to Promote Gastric Cancer Metastasis by Targeting Timp3. Cell. Physiol. Biochem. 2016; 39: 453–466, https://doi.org/10.1159/000445638.
40. Liu L, Zhou C, Zhou L, Peng L, Li D, Zhang X, Zhou M, Kuang P, Yuan Q, Song X, Yang M. Functional FEN1 Genetic Variants Contribute to Risk of Hepatocellular Carcinoma, Esophageal Cancer, Gastric Cancer and Colorectal Cancer. Carcinogenesis. 2012; 33: 119–123, https://doi.org/10.1093/carcin/bgr250.
41. Zhu Z, Peng R, Cai H. The Value of Nucleoporin 188 in Diagnosis, Prognosis and Immunoregulation: From Pan-Cancer Analysis to Gastric Cancer Verification. Front. Immunol. 2025; 16. https://doi.org/10.3389/fimmu.2025.1586784.
42. Muste Sadurni, M, Saponaro, M. Deregulations of RNA Pol II Subunits in Cancer. Appl. Biosci. 2023; 2: 459–476, https://doi.org/10.3390/applbiosci2030029.
43. He K, Feng Y, An S, Liu F, Xiang G. Integrative Epigenomic Profiling Reveal AP-1 Is a Key Regulator in Intrahepatic Cholangiocarcinoma. Genomics. 2022; 114: 241–252, https://doi.org/10.1016/j.ygeno.2021.12.008.
44. Subramanian V, Fields PA, Boyer LA. H2A.Z: A Molecular Rheostat for Transcriptional Control. F1000Prime Rep. 2015; 7. https://doi.org/10.12703/P7-01.
45. Vardabasso C, Hasson D, Ratnakumar K, Chung C-Y, Duarte LF, Bernstein E. Histone Variants: Emerging Players in Cancer Biology. Cell. Mol. Life Sci. 2014; 71: 379–404, https://doi.org/10.1007/s00018-013-1343-z.
46. Jafri Z, Li Y, Zhang J, O'Meara CH, Khachigian LM. Jun, an Oncological Foe or Friend? Int. J. Mol. Sci. 2025; 26: 555, https://doi.org/10.3390/ijms26020555.
47. Bradner JE, Hnisz D, Young RA. Transcriptional Addiction in Cancer. Cell. 2017; 168: 629–643, https://doi.org/10.1016/j.cell.2016.12.013.
48. Goetz JG, Lajoie P, Wiseman SM, Nabi IR. Caveolin-1 in Tumor Progression: The Good, the Bad and the Ugly. Cancer Metastasis Rev. 2008; 27: 715–735, https://doi.org/10.1007/s10555-008-9160-9.
49. Xu J, Lamouille S, Derynck R. TGF-β-Induced Epithelial to Mesenchymal Transition. Cell Res. 2009; 19: 156–172, https://doi.org/10.1038/cr.2009.5.