KIF4A as a novel prognostic biomarker in cholangiocarcinoma

Deng Yong Zhang, MD, Shuo Shuo Ma, MD, Wan-liang Sun, MD, Xue Chen Huang Lv, MD, Zheng Lu, PHD

Abstract
Cholangiocarcinoma (CCA) is one of the most common malignant tumors. Although gene-targeted therapies have significantly improved the outcome of many cancers, the results are still not satisfactory for patients with CCA. Owing to the lack of an effective biomarker for guiding clinical treatment and monitoring prognosis in patients with CCA, the purpose of this study was to identify a new biomarker that could help predict the outcome of patients with CCA using bioinformatics tools. Gene expression data were collected from three publicly available datasets, comprising 263 patients with CCA and 22 healthy controls. Differentially expressed genes were obtained using the limma package (FDR < 0.05, |Log2 FC| > 1), and the respective protein–protein interaction revealed five relevant genes in the STRING dataset (TOP2A, BUB1, RRM2, TYMS, and KIF4A). The immunohistochemistry and PCR were used to analyze the difference in KIF4A expression in CCA. Kinesin Family Member 4A (KIF4A) was the only gene significantly associated with overall patient survival (P = 0.035), with higher KIF4A expression being associated with poor survival rates. Moreover, KIF4A was significantly correlated with the infiltration of activated memory T cells (P = 0.0198) and activated mast cells (P = 0.008) in the tumor microenvironment. Increase in KIF4A expression affected the infiltration degree of the immune cells, which may be involved in the regulation of immune tolerance by CCA cells. The results indicated that the expression of KIF4A in CCA was higher than that in paracancerous tissues. Taken together, these findings suggest that KIF4A could be a potential new biomarker in CCA for predicting the response of patients to targeted immunotherapies.

Abbreviations: Cholangiocarcinoma = CCA, DEGs = differently expressed genes, GEO = Gene Expression Omnibus, GEPIA = Gene Expression Profiling Interactive Analysis, GO = gene ontology, GSEA = Gene set enrichment analysis, KEGG = Kyoto Encyclopedia of Genes and Genomes, KIF4A = Kinesin Family Member 4A, PBS = phosphate buffer saline, PPI = Protein–protein interaction, qRT-PCR = Quantitative real-time PCR, TCGA = The Cancer Genome Atlas.

Keywords: biomarker, cholangiocarcinoma, immunotherapies, prognosis

1. Introduction
Cholangiocarcinoma (CCA) is a heterogeneous group of malignancies that can emerge at every point of the biliary tree. It is the second most common primary liver tumor and accounts for approximately 10–15% of all hepatobiliary malignancies. CCA can be classified as intrahepatic, perihilar, and distal CCA according to their anatomical location, and these subtypes have not only particular similarities but also important inter- and intra-tumor differences that can affect the CCA pathogenesis and outcome. These subtypes differ in their epidemiology, etiology, pathology, and treatment. CCA is difficult to diagnose at an early stage, with most cases being already at an advanced middle or late stage at the time of diagnosis. Surgical treatment is preferred for all subtypes; however, vascular structure and lymph node involvement need to be considered. The high fibrinolytic quality of CCA, extensive support of a rich tumor microenvironment, and profound genetic heterogeneity, are all important contributing factors to therapeutic resistance in CCA. Most CCA patients lack an obvious correlation with known etiological factors, complicating the monitoring and early diagnosis of patients at risk for tumor development. CA19-9 is currently used to diagnose CCA, but owing to its low specificity and sensitivity, early diagnosis and prognostic detection of CCA are unreliable. Therefore, biomarkers have a critical role in the diagnosis, prognosis, and treatment of patients with CCA. A reliable biomarker can be used for early diagnosis of CCA,
evaluation of therapeutic efficacy, prediction of the likelihood of recurrence, and targeted drug development.

KIF4A is a member of the kinesin superfamily, with a molecular weight of approximately 140 kDa. It is mainly located in the nucleus, and its localization and function in the nucleus vary with the cell cycle stage. In addition to the involvement of KIF4A in the transport of intracellular materials, it is closely related to the formation and dynamics of spindles during mitosis, concentration and arrangement of chromosomes, and completion of cytokinesis. Some studies have revealed that KIF4A is involved in DNA damage and repair and that abnormal expression of KIF4A may affect the expression of the homologous recombination enzyme Rad51 and its regulator BRCA2, leading to the failure of damaged DNA repair. DNA damage may lead to abnormal cell proliferation and differentiation, ultimately promoting tumor formation. Therefore, KIF4A is closely related to the occurrence and prognosis of various human tumors.

In this study, we used bioinformatics techniques to identify and verify the prognostic value of differentially expressed genes (DEGs) in CCA and conducted in-depth studies on the biological functions and involvement of potential biomarkers to better understand the biology of CCA. Improved knowledge of the carcinogenic background of this disease and its complex interactions with the tumor microenvironment can lead to optimal treatment and improved patient survival.

2. Materials and methods

2.1. Data and tissue collection

The gene expression data of bile duct cancer were collected from three publicly available datasets stored in the Gene Expression Omnibus (GEO) database. The datasets were selected based on the sample size (greater than 20) and presence of control samples. The GSE26366 dataset was based on the Illumina GPL6104 platform (Illumina humanRef-8 v2.0 expression beadchip) and comprised information on 111 patients with bile duct carcinoma. The GSE32225 dataset was based on the Illumina GPL4342 platform (HumanRef-8 WG-DASL v3.0) and included information on 149 tissues from patients with CCA and expression spectrum data from six adjacent tissues. Lastly, the GSE45001 dataset was based on the Agilent GPL4550 platform (Agilent-028004 SurePrint G3 Human GE 8x60K Microarray) and consisted data on 10 patients with CCA and 10 adjacent expression spectrum data. The GEOquery toolkit was used for downloading and processing the data. Considering the differences among data platforms, we used the SVA toolkit for removing the batch effect. Tissue specimens from the First Affiliated Hospital of Bengbu Medical College (No. 2019035).

2.2. Differentially expressed gene analysis and gene set enrichment analysis

We used the limma R package to obtain 588 DEGs (FDR < 0.05, |log2FC| > 1) between healthy controls and patients with CCA (216 upregulated genes and 372 downregulated genes). Next, we used the clusterProfiler R package to carry out the gene ontology (GO) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses on DEGs. A P-value of < 0.05 was considered statistically significant and used as a threshold to filter the data.

2.3. Protein–protein interaction (PPI) network and selection of hub genes

We further constructed a PPI network of DEGs based on the STRING database. The network was visualized using Cytoscape, and the five most relevant genes were extracted using the CytoHubba plug-in.

2.4. Survival and mutation analyses

Owing to the lack of prognostic information on patients with bile duct carcinoma in the selected GEO datasets, we used Gene Expression Profiling Interactive Analysis (GEPIA), which is based on information from The Cancer Genome Atlas (TCGA) database, to assess the predictive potential of the hub genes with regard to patient survival. The mutational spectrum of the five identified hub genes was explored using the cBioPortal website.

2.5. Analysis of immune cell infiltration and its correlation with KIF4A expression

The Cibersort algorithm was used to analyze immune cell infiltration in the cancer samples (N = 263), based on the combined gene chip data. The correlation analysis of 0.05 samples was generally considered as statistically significant. The calculated results of the 0.05 samples were more accurate and reliable, and thus, in total, 93 samples were selected for further analysis. Then, we calculated the correlation between KIF4A expression and the level of immune cell infiltration.

2.6. Gene set enrichment analysis (GSEA)

GSEA is a statistical method for evaluating whether a congenitally defined set of genes shows statistically significant differences in consistency between two different biological states. A dataset comprising 270 patients with intrahepatic CCA was classified based on the median value of KIF4A expression, and the clusterProfiler R package was used to identify the main cellular pathways affected by KIF4A.

2.7. Quantitative real-time PCR

Total RNA in CCA tissues was extracted using a Trizol kit (Invitrogen, USA) according to the manufacturer’s instructions. Quantitative real-time PCR (qRT-PCR) was performed using an Applied Biosystems 7500 Real-Time PCR system (Applied Biosystems, Foster City, CA, USA) and SYBR Premix Ex Taq Tli RNaseH Plus (Takara Bio; RR820A).

The primers used were as follows: KIF4A – Forward primer 5'-GGGATGACGGGAATGGA-3', reverse primer 5'-CAGCAACAGTCCACAC-3'; GAPDH – forward primer 5'-CACCAGGGCTGCTTTTAACTCTG-3' and reverse primer 5'-GATTTTGGAGGGATCTCGCTCCTG-3'. Data are presented as mean ± SD in at least three independent experiments.

2.8. Immunohistochemistry

After the CCA and adjacent normal tissues were embedded into paraffin and sectioned, the tissue sections were dewaxed with
xylene, hydrated with gradient ethanol, repaired with EDTA antigen retrieval solution, and soaked in hydrogen peroxide solution at 25°C for 30 min. Next, the sections were incubated with 3% bovine serum albumin at 37°C for 30 min. Later, the sections were washed with phosphate buffer saline (PBS) and incubated with a primary antibody (rabbit anti-human KIF4A, 1:50 dilution, Abcam, Cambridge, MA, USA; ab124903) overnight at 4°C. Following incubation, the sections were washed three times with PBS for 5 min each. After discarding the PBS, the sections were incubated with an HRP-labeled rabbit secondary antibody (31460; Waltham, MA, USA) 37°C for 30 min. Then, the sections were washed with PBS and incubated with DAB chromogenic agent. The sections were again washed with PBS, counterstained with hematoxylin, placed in cold water.
for 10 min, dehydrated with gradient ethanol, air dried, and sealed with neutral gum.

2.9. Statistical analysis

Differentially expressed gene analysis, GO, KEGG, and Pearson correlation analysis were conducted using Rstudio (v.3.6.1). We obtained the outcomes of Kaplan Meier survival analysis in the GEPIA database. The selection criteria were as follows: Methods (Overall Survival); Group Cutoff (Median); 95% Confidence Interval (YES); Datasets Selection (CHOL).

3. Results

3.1. DEGs may contribute to important biological functions in CCA

The genetic landscape of 263 cancer samples was compared with that of 22 healthy control samples. In total, 588 DEGs were identified (FDR<0.05, |Log2FC|>1) (Fig. 1).

Further, GO and KEGG analyses were performed on these DEGs to predict some possible biological functions (Fig. 2). KEGG analysis revealed that the complement and coagulation cascades, metabolism of xenobiotics by cytochrome P450, and PPAR signaling pathway, among other pathways, were significantly enriched. Meanwhile, GO enrichment analysis indicated that the

Figure 3. Protein–protein interaction network of the differentially expressed genes in cholangiocarcinoma. The node size represents the degree value of each protein in the network.

Figure 4. The top five proteins identified in the protein–protein interaction network of cholangiocarcinoma. The key subnetwork was selected according to the node degree of each protein.
DEGs were also involved in extracellular structure organization, biological processes, and the carboxylic acid catabolic process.

3.2. Five hub genes were identified from the PPI network

The interaction between different proteins in CCA was analyzed using the STRING database (Fig. 3), and the CytoHubba algorithm was used to calculate the node degrees of all the genes in the PPI network. Among the 588 DEGs, five genes stood out – TOP2A, BUB1, RRM2, TYMS, and KIF4A – as the most relevant interaction nodes (Fig. 4).

3.3. KIF4A overexpression reduces patient survival in CCA

Next, we used the GEPIA database to analyze the effect of the expression patterns of TOP2A, BUB1, RRM2, TYMS, and

Figure 5. Survival analysis diagrams of the top five proteins identified in the critical cholangiocarcinoma subnetwork.
KIF4A on the survival of patients with CCA. Only the expression of KIF4A had a significant effect on the survival outcome of patients with CCA, with KIF4A overexpression being associated with significantly lower survival (P = .035) (Fig. 5).

### 3.4. Mutational status of the five hub genes

To evaluate the mutational status of RRM2, TYMS, TOP2A, KIF4A, and BUB1, the genes were analyzed using the cBioPortal database. The mutation frequency of these genes is low in CCA (Fig. 6).

The analysis of KIF4A expression in other carcinomas revealed that KIF4A is overexpressed in several malignant tumors compared to its expression in the corresponding normal cells/tissues (Fig. 7). Therefore, KIF4A may play an important role in a variety of cancers, although its underlying mechanism and function are not well understood.

### 3.5. Association of KIF4A with immune infiltrating cells

Next, the relationship between KIF4A expression and immune cell infiltration was further explored. We used the CIBERSORT
algorithm to calculate the immune infiltration level in 263 CCA samples and selected 93 samples for subsequent analysis ($P < .05$). KIF4A expression significantly correlated with the infiltration of various immune cells within the tumors (Fig. 8). In particular, higher KIF4A expression significantly increased the infiltration of CD4+ memory T cells ($P = .0198$), whereas the number of resting CD4+ memory T cells decreased ($P = .0184$). Similarly, increased KIF4A expression was associated with
significantly increased mast cell activation ($P = .008$) and reduced number of resting mast cells ($P = .0086$). Altogether, these results suggest that KIF4A is involved in the regulation of immunity within CCA tumors.

3.6. GSEA of KIF4A

Lastly, the impact of KIF4A expression on cell behavior and biological function was explored. According to the median expression value of KIF4A, the 263 CCA samples were divided into two groups: high expression and low expression groups. The GSEA algorithm was used to enrich and analyze the genes of the high and low expression groups. The results revealed that the GO functions (Fig. 9A) were significantly enriched in the cell cycle, cytokine receptor interaction, and complement and coagulation cascades, showing a positive correlation with KIF4A. In contrast, KEGG pathway analysis (Fig. 9B) was mainly concentrated in the inflammatory response and regulation of cell activation. These results show that KIF4A plays a key role in the immune status of CCA.

3.7. Expression of KIF4A in CCA tissues

We conducted qRT-PCR and immunohistochemistry to verify the expression of KIF4A in five CCA samples. GAPDH was used as a reference gene for PCR. The results of immunohistochemistry showed significantly high KIF4A expression in CCA tissues than that in paracarcinoma tissues (Fig. 10 (A)).
qRT-PCR results showed that the expression of KIF4A mRNA was significantly high in CCA samples than that in para-carcinoma samples (average of five patients), with an average increase in expression by approximately 4.5-fold \((P < .001)\) (Fig. 10(B)).

4. Discussion
In this study, we evaluated and compared the genetic landscape of 263 cases of CCA and 22 healthy counterparts. Based on 588 DEGs, a PPI network was constructed, and five genes – RRM2, TYMS, TOP2A, KIF4A, and BUB1 – were identified as being...
most relevant. Of note, overexpression of KIF4A was found to be associated with poorer prognosis in CCA, resulting in significantly shorter survival time. KIF4A expression was significantly correlated with immune cell infiltration at the tumor site, further predicting the GO and KEGG enrichment of KIF4A. To our knowledge, the role of KIF4A in CCA is still unknown.

CCA has received increasing attention due to its high malignancy potential and poor outcomes. In recent decades, the global incidence of CCA, in particular intrahepatic CCA, has increased. Recent studies have shown that immunotherapy has made some initial achievements in the treatment of many malignant tumors. However, not all patients with advanced cancer are eligible for such treatments and may not have a good prognosis, as immunotherapies require a long time to show their therapeutic effects. Therefore, biomarkers, such as KIF4A identified in this study, can be used to estimate treatment response and survival outcomes in patients with CCA.

In this study, the analysis of the most relevant DEGs (the PPI nodal genes) revealed their potential involvement in CCA. We found that RRM2, TYMS, TOP2A, KIF4A, and BUB1 were mainly enriched in pathways involved in the complement and coagulation cascades, metabolism of xenobiotics by cytochrome P450, and PPAR signaling pathway. In addition, GO function analysis revealed that these five genes were implicated in the organization of biological processes and the carboxylic acid catabolic process. Analysis of the mutational status of the five genes revealed a low frequency of mutations associated with CCA.

As mentioned above, we selected the genes with the top five node degrees in the PPI network; however, only KIF4A demonstrated predictive potential for poor prognosis in CCA. The survival time of patients with overexpression of KIF4A was significantly lower than that of patients with low KIF4A expression. Therefore, inhibition of the KIF4A expression may affect the biological behavior of bile duct cancer cells, leading to the prevention of cancer cell growth, spread, and overall CCA progression. KIF4A expression was significantly altered in a variety of malignant tumors in comparison with that in healthy counterparts. Therefore, we chose KIF4A for the next step of molecular research.

We investigated the relationship between KIF4A expression and immune infiltration and found that KIF4A expression was significantly correlated with the activation of CD4+ memory T cells and mast cells. These results suggest that KIF4A is involved in immune regulation and immune escape mechanisms of CCA cells through promoting the infiltration of activated immune cells. In particular, KIF4A may contribute to enhanced immune tolerance of bile duct cancer cells, allowing them to evade the host immune cancer-targeted activity. Furthermore, we found that KIF4A was mainly involved in cytokine–receptor interaction, complement and coagulation cascades, natural killer cell-mediated cytotoxicity, cell cycle, ribosome, and calcium signaling pathways. Moreover, we conducted qRT-PCR and immunohistochemistry to verify the expression of KIF4A in CCA and paracarcinoma tissues. As shown in Figure 10, the results of both immunohistochemistry and qRT-PCR indicated an increased expression of KIF4A in CCA tissues compared with that in paracarcinoma tissues. These results suggest that KIF4A has a broad cellular impact, which warrants further studies. Nevertheless, based on our findings, we speculate that KIF4A may hold predictive potential and represent a useful biomarker in CCA.

There were some limitations to this study. First, our survival and KIF4A expression association analysis was derived from the GEPIA database because of the lack of clinical information on patients with CCA and the low number of samples available. Second, no laboratory experiments were conducted to verify the biological activity of KIF4A in CCA. The functional aspects of KIF4A will be the focus of our future studies.

5. Conclusion
In this study, we screened and validated the expression, survival prognosis, and mutational status of five potential biomarkers of CCA. Particularly, KIF4A showed potential as a representative novel biomarker in CCA, with the ability to evaluate the tumor microenvironment immunity and identify patients who could benefit from gene-targeted therapy or immunotherapy. These findings will provide new insights into the biology of CCA and potential new treatment strategies.
Author contributions

Conceptualization: Lu zheng.
Data curation: Deng Yong Zhang.
Methodology: Xue Chen Huang Lv.
Visualization: Shuo Shuo Ma, Wan-liang Sun.
Writing – review & editing: Lu zheng.

References

[1] Bergquist A, von Seth E. Epidemiology of cholangiocarcinoma. Best Pract Res Clin Gastroenterol 2015;29:221–32.
[2] Banales JM, Cardinale V, Carpino G, et al. Expert consensus document: Cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). Nat Rev Gastroenterol Hepatol 2016;13:261–80.
[3] Razumilava N, Gores GJ. Cholangiocarcinoma. Lancet 2014;383:2168–79.
[4] Rizvi S, Gores GJ. Pathogenesis, diagnosis and management of cholangiocarcinoma. Gastroenterology 2013;145:1215–29.
[5] Rizvi S, Khan SA, Hallemeyer CL, Kelley RK, Gores GJ. Cholangiocarcinoma—evolving concepts and therapeutic strategies. Nat Rev Clin Oncol 2018;15:95–111.
[6] Macias RIR, Kornek M, Rodrigues PM, et al. Diagnostic and prognostic biomarkers in cholangiocarcinoma. Liver Int 2019;39(Suppl 1):108–22.
[7] Macias RIR, Banales JM, Sangro B, et al. The search for novel diagnostic and prognostic biomarkers in cholangiocarcinoma. Biochim Biophys Acta Mol Basis Dis 2018;1864(4 Pt B):1468–77.
[8] Kurasawa Y, Earnshaw WC, Mochizuki Y, Dohmae N, Todokoro K. Essential roles of KIF4 and its binding partner PRC1 in organized central spindle midzone formation. EMBO J 2004;23:3237–48.
[9] Mazumdar M, Sundareshan S, Misteli T. Human chromokinesin KIF4A functions in chromosome condensation and segregation. J Cell Biol 2004;166:613–20.
[10] Zhu C, Jiang W. Cell cycle-dependent translocation of PRC1 on the spindle by Kif4 is essential for midzone formation and cytokinesis. Proc Natl Acad Sci U S A 2005;102:343–8.
[11] Lee YM, Kim W. Association of human kinesin superfamily protein member 4 with BRCA2-associated factor 35. Biochem J 2003;374(Pt 2):497–503.
[12] Wu G, Zhou L, Khidr L, et al. A novel role of the chromokinesin Kif4A in DNA damage response. Cell Cycle 2008;7:2013–20.
[13] Taniwaki M, Takano A, Ishikawa N, et al. Activation of KIF4A as a prognostic biomarker and therapeutic target for lung cancer. Clin Cancer Res 2007;13(22 Pt 1):6624–31.
[14] Narayan G, Bourdon V, Chaganti S, et al. Gene dosage alterations revealed by cDNA microarray analysis in cervical cancer: identification of candidate amplified and overexpressed genes. Genes Chromosomes Cancer 2007;46:373–84.
[15] Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846–7.
[16] Leek JT, Johnson WE, Parker HS, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012;28:882–3.
[17] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7.
[18] Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res 2017;45(D1):D362–8.
[19] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
[20] 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;4:511.
[21] Tang Z, Li C, Kang B, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2017;45(W1):W98–102.
[22] Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6:11.
[23] Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453–7.
[24] Bridgewater J, Galle PR, Khan SA, et al. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. J Hepatol 2014;60:1268–89.
[25] O’Donnell JS, Hoefsmite EP, Smyth MJ, Blank CU, Teng MWL. The promise of neoadjuvant immunotherapy and surgery for cancer treatment. Clin Cancer Res 2019;25:5743–51.
[26] Sharma P, Hu-Lieskovan S, Wargo JA, et al. Primary, adaptive, and acquired resistance to cancer immunotherapy. Cell 2017;168:707–23.