Identification of key genes and pathways associated with cholangiocarcinoma development based on weighted gene correlation network analysis

Jingwei Liu 1, Weixin Liu Corresp. 1, Hao Li 1, Qiuping Deng 1, Meiqi Yang 1, Xuemei Li 1, Zeng Liang 1

1 Department of Gastroenterology, the First Affiliated Hospital of China Medical University, Shenyang, Liaoning, China

Corresponding Author: Weixin Liu
Email address: chunchang88@sohu.com

Background: As the most frequently occurred tumor in biliary tract, cholangiocarcinoma (CCA) is mainly characterized by its late diagnosis and poor outcome. It is therefore urgent to identify specific genes and pathways associated with its progression and prognosis.

Materials and Methods: The differentially expressed genes in The Cancer Genome Atlas (TCGA) were analysed to build the co-expression network by Weighted gene co-expression network analysis (WGCNA). Gene ontology (GO) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were conducted for the selected genes. Module-clinical trait relationships were analyzed to explore the association with clinicopathological parameters. Log-rank tests and cox regression were used to identify the prognosis-related genes.

Results: The most related modules with CCA development were tan module containing 181 genes and salmon module with 148 genes. GO analysis suggested enrichment terms of digestion, hormone transport and secretion, epithelial cell proliferation, signal release, fibroblast activation, response to acid chemical, wnt, NADP metabolism. KEGG analysis demonstrated 15 significantly altered pathways including glutathione metabolism, wnt, central carbon metabolism, mTOR, pancreatic secretion, protein digestion, axon guidance, retinol metabolism, insulin secretion, salivary secretion, fat digestion. Key genes of SOX2, KIT, PRSS56, WNT9A, SLC4A4, PRRG4, PANX2, PIR, RASSF8, MFSD4A, INS, RNF39, IL1R2, CST1 and PPP3CA might be potential prognostic markers for CCA, of which RNF39 and PRSS56 also showed significant correlation with clinical stage.

Discussion: Differentially expressed genes and key modules contributing to CCA development were identified by WGCNA. Our results offer novel insights into the characteristics in the etiology, prognosis and treatment of CCA.
Identification of key genes and pathways associated with cholangiocarcinoma development based on weighted gene correlation network analysis

Jingwei Liu¹, Weixin Liu¹*, Hao Li¹, Qiuping Deng¹, Meiqi Yang¹, Xuemei Li¹, Zeng Liang¹

¹, Department of Gastroenterology, the First Affiliated Hospital of China Medical University, 155# North Nanjing Street, Heping District, Shenyang City, 110001, Liaoning Province, China.

* Corresponding author: Prof. Weixin Liu, Department of Gastroenterology, the First Affiliated Hospital of China Medical University, 155# North Nanjing Street, Heping District, Shenyang City, 110001, Liaoning Province, China.

E-mail address: chunchang88@sohu.com

Running title: WGCNA analysis of cholangiocarcinoma
Abstract

Background: As the most frequently occurred tumor in biliary tract, cholangiocarcinoma (CCA) is mainly characterized by its late diagnosis and poor outcome. It is therefore urgent to identify specific genes and pathways associated with its progression and prognosis.

Materials and Methods: The differentially expressed genes in The Cancer Genome Atlas (TCGA) were analysed to build the co-expression network by Weighted gene co-expression network analysis (WGCNA). Gene ontology (GO) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were conducted for the selected genes. Module–clinical trait relationships were analyzed to explore the association with clinicopathological parameters. Log-rank tests and cox regression were used to identify the prognosis-related genes.

Results: The most related modules with CCA development were tan module containing 181 genes and salmon module with 148 genes. GO analysis suggested enrichment terms of digestion, hormone transport and secretion, epithelial cell proliferation, signal release, fibroblast activation, response to acid chemical, wnt, NADP metabolism. KEGG analysis demonstrated 15 significantly altered pathways including glutathione metabolism, wnt, central carbon metabolism, mTOR, pancreatic secretion, protein digestion, axon guidance, retinol metabolism, insulin secretion, salivary secretion, fat digestion. Key genes of SOX2, KIT, PRSS56, WNT9A, SLC4A4, PRRG4, PANX2, PIR, RASSF8, MFSD4A, INS, RNF39, IL1R2, CST1 and PPP3CA might be potential prognostic markers.
for CCA, of which RNF39 and PRSS56 also showed significant correlation with clinical stage.

**Discussion:** Differentially expressed genes and key modules contributing to CCA development were identified by WGCNA. Our results offer novel insights into the characteristics in the etiology, prognosis and treatment of CCA.

**Key words:** cholangiocarcinoma, weighted gene correlation network analysis, progression, prognosis.
Introduction

Cholangiocarcinoma (CCA), characterized by cholangiocyte differentiation, represents a type of aggressive tumor which derive from lesions in different parts of the biliary tree (Bergquist & von Seth 2015). As the most frequently occurred biliary tumor and the second after liver cancer, cholangiocarcinoma is mainly characterized by its late diagnosis and poor outcome (Blechacz 2017). Currently, it is estimated that overall 5-year survival rate of CCA is less than 10% and only approximately one third of the CCA patients are suitable for curative treatment at the time of diagnosis (Doherty et al. 2017). Therefore, it is urgently required to identify specific genes and pathways associated with its initiation, progression and prognosis.

The common category of CCA are on basis of different locations of intrahepatic, perihilar or distal cholangiocarcinoma (Oliveira et al. 2017). Most patients are diagnosed with non-resectable status and suffer an extremely poor prognosis (Rizvi & Gores 2017). A number of diseases have been reported to be related with the development of CCA including viral hepatitis B and C, choledocholithiasis, obesity, exposure to toxic agents, liver cirrhosis, diabetes mellitus, congenital hepatic fibrosis and primary sclerosing cholangitis (Erichsen et al. 2009). Every year, a larger number of cases occur in Southeast Asia, in which hepatobiliary flukes Opisthorchis viverrini infection or Clonorchis sinensis infection were considered to be implicated in the development of CCA (Razumilava & Gores 2014). Interleukin-6, a pro-inflammatory cytokine associated with downstream
activation of oncogenic pathways, has been linked with cholangiocarcinoma development (Park et al. 1999). Frequent mutations of oncogenes such as KRAS, as well as cancer suppressor genes of TP53 and SMAD4 were identified by next generation sequencing in CCA (Chan-On et al. 2013). In addition, research from several case-controlled studies has demonstrated multiple genetic polymorphisms that might be implicated in cholangiocarcinoma carcinogenesis (Bridgewater et al. 2014).

Although a number of genes and mechanisms have been proved to be closely implicated in the development of CCA, the comprehensive picture of the whole genes and regulations of CCA is still unclear. In recent years, bioinformatic methods become increasingly effective in exploration and analysis of multiple genes or proteins of complicated diseases. Weighted gene co-expression network analysis (WGCNA), a new gene co-expression analysis method, has been successfully used to screen biomarkers and pathways that could be applied in susceptibility genes, diagnose and treatment of cancer. In this study, WGCNA was conducted to analyse data of TCGA data repository of CCA to screen modules and core genes in pathogenesis, progression and survival of CCA.

**Materials and Methods**

**Publically available data sets**

RNA expression as well as clinical parameters of CCA patients were obtained from The Cancer Genome Atlas (TCGA) database (cancergenome.nih.gov). The level of gene expression was tested as Transcripts Per Kilobase of exon model per Million mapped...
reads (TPM). Clinical characteristics had the sample type, histology grade, recurrence, histologic grade, and prognosis. Each sample must have complete pathology stage and histology records. If the expression of genes showed limited variation, we regraded them as noise and discard these ones because the results of these genes might come from systematic error and have limited significance.

**Construction of co-expression network of genes**

In this study, we adopted WGCNA method to build a co-expression network for certain genes using R language (Langfelder & Horvath 2008). We used WGCNA method to calculate power number in order to construct modules through co-expression. WGCNA method was also performed for construction of the co-expression network and extraction of the genetic information in the most relevant module. “Heatmap tool” package of R software was selected to analyze the correlation degree among modules. As a representative of the gene expression profiles of a module, module eigengene (ME) was used to evaluate the relationship between module and overall survival.

**Identification of clinical traits-related modules**

We then performed Pearson’s correlation examination to explore the relation of MEs with clinical traits of histologic grade, recurrence, pathologic M, pathologic N, pathologic T. A p value less than 0.05 indicated statistical significance. The two most significant modules are selected as hub modules.

**GO and KEGG analysis**

In the present study, we adopted the clusterprofiler package of R language to
investigate the possible biological procedures and signaling pathways of selected genes in different module. Gene ontology (GO) analysis (Ashburner et al. 2000) as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were conducted (Kanehisa et al. 2004), which enable computational prediction of higher-level complexity of cellular processes and organism behavior from genomic information.

Prognostic value of genes of hub modules

Prognosis investigation was conducted using the ‘survival’ package within R software. Cox regression test was used to calculate the hazard ratio and its 95% confidence interval while Kaplan-Meier method was adopted to draw survival curve. The increased and decreased expression level of each gene in the candidate module was decided according to median value.

Screening for candidate hub genes

We chose genes associated with prognosis as candidate genes. Next, we analyzed the association between candidate genes and stage parameter. The genes associated with both overall survival and clinical stage were selected as hub genes.

Results

Gene expression profile of CCA

Clinical characteristics and RNA expression value for 33 Cholangiocarcinoma individuals were downloaded at TCGA. Altogether 6417 relatively more variant genes on the basis of median absolute deviation value were considered in our subsequent investigation. When soft thresholding power $\beta$ came to 7, the connectivity among genes
accord with a scale-free network distribution (Supplementary Figure 1). A total of 21 modules were screened using hierarchical clustering along with Dynamic branch Cutting by WGCNA analysis. And each different color represent a sole module. Interaction relationship analysis of co-expression genes was shown in Figure 1. The gene counts within modules were from 37 to 909. In this study, the grey module collected a series of genes which did not belong to any module.

**Association of module with clinical traits**

Identification of genes and modules which associate with clinical characteristic would greatly help the understanding the underlying mechanisms of certain traits. In our analysis, the clinical parameters of CCA of histologic grade, recurrence, pathologic M, pathologic N, pathologic T were selected in the module–trait relationship (MTR) analysis. As was suggested in Figure 2, the two relatively most significant modules of tan and salmon were selected as hub modules because their significant relation with clinical parameters including grade and T, N, M of CCA.

**Enrichment investigation of the hub modules**

We next proceeded GO and KEGG investigation of the genes within tan module and salmon module. Altogether 50 terms showed significant difference in GO enrichment analysis (Supplementary Table 1) and the most important terms were summarized in Table 1. As was illustrated in Figure 3, the salmon module in GO analysis was related with digestion, hormone transport, hormone secretion, epithelial cell proliferation, regulation of hormone levels, signal release while the tan module was associated with
fibroblast activation, response to acid chemical, wnt signaling pathway, cell-cell signaling
by wnt, NADP metabolic process, negative regulation of neuron differentiation, negative
regulation of cell development. According to the KEGG analysis, 15 pathways were
significantly altered in hub modules including glutathione metabolism, wnt signaling
pathway, central carbon metabolism in cancer, mTOR signal, pancreas secretion, protein
digestion or absorption, axon guidance, retinol metabolism, insulin secretion, salivary
secretion, and fat digestion and absorption (Figure 4 and Table 2).

Prognosis analysis

Since the tan and salmon module were selected from as the hub modules, we are
interested in whether the genes with in the modules might predict the survival of CCA
patients. Of the 358 genes, a total of 15 genes (SOX2, KIT, PRSS56, WNT9A, SLC4A4,
PRRG4, PANX2, PIR, RASSF8, MFSD4A, INS, RNF39, IL1R2, CST1 and PPP3CA)
demonstrated significant association with the prognosis of CCA (Table 3). The survival
curve for the top 4 genes in salmon module (INS: HR=4.670, P= 0.006; RNF39:
HR=2.811, P= 0.032; CST1: HR=2.846, P=0.017; PPP3CA: HR=3.151, P=0.010) and the
top 4 gene in tan module (PRSS56: HR = 2.876, P = 0.018); SLC4A4: HR = 0.352, P =
0.025; PRRG4: HR = 0.339, P = 0.023; PIR: HR = 2.701, P = 0.027) were illustrated in
Figure 5. In addition, key genes of DNA repair and immune regulation were analysed in
relation to CCA prognosis, the results of which suggested that DNA repair gene XPC (HR
= 0.369, P = 0.048) and immune regulator CD28 (HR = 0.364, P = 0.045) were related
with favourable survival of CCA (Supplementary Table 2). We also analyzed the
prognosis of genes more frequently implicated in CCA including KRAS, TP53, BAP1. The mutation frequency of KRAS, TP53, BAP1 were 5.71%(2), 8.57%(3), 5.71%(2) out of 33 CCA patients. And no significant association with prognosis was observed.

Identification of hub genes associated with both stage and survival

We then analyse the association between clinical stage and gene which showed significant relation with prognosis. As was shown in Figure 6, two genes (RNF39 and PRSS56) showed difference between two groups (Table 4).

Discussion

Although surgical therapy and liver transplantation might be possible options for certain patients who suffer CCA, the five-year survival ratio of CCA remain extremely low. Comprehensive exploration of the genetic and epigenetic profiles as well as their complicated interactions with environment would greatly help the treatment and survival for CCA patients. Until now, our understanding of the complex mechanisms of CCA is still limited. One study using CCA data also analysed the genes implicated in CCA (Tian et al. 2019). They dissected the genome network by other bioinformatic tool to explore the protein-protein interaction. In this study, we provided our own gene lists with significance by WGCNA. Besides, we adopted Kyoto Encyclopedia of Genes and Genomes analysis to explore the biological functions for the tumor progression. A differentially expression analysis followed by WGCNA to identify genes and pathways related were performed to the clinical parameters and prognosis of CCA. In addition, enrichment analysis of the genes in core modules suggested significant involvement of pathways such as
glutathione metabolism, wnt signaling, mTOR signaling, pancreatic secretion, protein digestion and absorption, insulin secretion and salivary secretion.

In this present study, RNA sequencing data for CCA samples from TCGA were systematically analyzed. According to the WGCNA analysis of most variant genes, altogether 21 modules were identified and the two most significant modules of tan and salmon were selected as hub modules. GO analysis of salmon module demonstrated enrichment in digestion, hormone transport, hormone secretion, epithelial cell proliferation, regulation of hormone levels, signal release. It can be concluded from our analysis that abnormal hormone regulation might be closely implicated in CCA progression. Previously, autocrine parathyroid hormone-like hormone was indicated to promote intrahepatic CCA cell growth through enhanced ERK/JNK-ATF2-cyclinD1 signalling (Tang et al. 2017). There were also reports of CCA patients with aberrant levels of parathyroid hormone-related protein (Matsumoto et al. 2014; Ozawa et al. 2017). From this point of view, future investigations concerning the synthesis, transport and secretion of relevant hormones might provide novel insight into CCA development.

In addition, the tan module was associated with fibroblast activation, response to acid chemical, wnt signalling pathway, NADP metabolic process, negative modulation of neuron differentiation, negative modulation of cell development. Wnt signalling pathway has been proved to be involved in the growth and progression of various types of cancer (Clevers & Nusse 2012). Previously, enhanced wnt signalling has been found to be a characteristic of CCA (Boulter et al. 2015) which related with metastasis or
chemoresistance (Wang et al. 2015). Besides, IncRNA PCAT1 regulated the development of extrahepatic CCA progression by Wnt and β-catenin signalling pathway(Zhang et al. 2017a). It is worth noting that response to acid chemical was also a hallmark in CCA development according to our study. Indeed, glycocholic acid and taurochenodeoxycholic acid were found to be potential phenotypic biomarkers in CCA(Song et al. 2018). Conjugated bile acids were suggested to promote invasive growth of cholangiocarcinoma cells via activation of sphingosine 1-phosphate receptor 2(Liu et al. 2014). The role of other biological processes such as NADP metabolic process and neuron differentiation required further studies to elucidate.

According to the KEGG analysis, 15 pathways were significantly altered in hub modules including glutathione metabolism, wnt signaling pathway, central carbon metabolism in cancer, mTOR signaling pathway, pancreatic secretion, protein digestion and absorption, axon guidance, retinol metabolism, insulin secretion, salivary secretion, and fat digestion and absorption. Some of the identified pathways have been suggested to play critical role in CCA progression and therapy. In intrahepatic cholangiocarcinoma (ICC) mouse model, investigators proved that mTOR kinase inhibitors inducing strong ICC cell apoptosis and may be beneficial for the treatment of ICC(Zhang et al. 2017b). It has been reported that the contact inhibition of cholangiocarcinoma cells could be overcome by c-Myc via the mTOR pathway(Luo et al. 2017). The identified significant pathways of secretion and absorbance of multiple digestive juice remind us the importance of digestive balance in the development of CCA, which required further
investigations to follow. The significant altered pathways of insulin secretion, salivary secretion, and fat digestion and absorption confirmed the key role of digestive regulation in CCA, as was indicated by a number of previous investigations (Chaiyarit et al. 2011; Chen et al. 2013; Rizvi et al. 2014).

The five-year survival rate of CCA currently remains unsatisfactory despite recent improvements in treatments of surgical resection and chemotherapy (Squadroni et al. 2017). The selected genes hold great possibilities to become potential biomarkers for the prediction of survival of CCA patients. We eventually screened 15 prognosis-related genes in CAA after analyzing the hub modules of tan and salmon. Increased expression of SOX2 in CCA cells lead to cell proliferation, decreased cell apoptosis, and increased cell invasion and metastasis (Sun et al. 2014). Besides, SOX2 expression correlated with aggressive behavior and worse overall survival in intrahepatic cholangiocarcinoma (Gu & Jang 2014). Among the identified prognosis-related genes, RNF39 and PRSS56 also showed significant correlation with clinical stage. Until now, little is known about the functions of these two genes in cancer. RNF39 encoded a member of ring finger proteins, which was studied in Behcet's disease, a chronic inflammatory autoimmune disease (Kurata et al. 2010). In addition, RNF39 protein was suggested to participate in the replication process according to both studies on HIV-1 individuals and cell-lines (Lin et al. 2014). PRSS56 has been established as an effective marker of adult neurogenesis in the brain of mouse (Jourdon et al. 2016). Variants in PRSS56 may lead to primary angle-closure glaucoma and high hyperopia (Jiang et al. 2013). Whether RNF39 and
PRSS56 could be a robust biomarker for CCA and their underlying mechanisms still required further studies to explore. In addition, DNA repair gene XPC and immune regulator CD28 were related with favourable survival of CCA, which indicate probable implication of DNA repair and immune response in CCA development. Because no distant metastasis information was available in TCGA data of CCA, we did not perform association analysis of genes with CCA metastasis, which require future studies to clarify.

Conclusion

In summary, differentially expressed genes and key modules contributing to progression of cholangiocarcinoma were identified by means of WGCNA. Pathways including glutathione metabolism, wnt signaling, central carbon metabolism in cancer, mTOR signaling, pancreatic secretion, protein and fat digestion, insulin secretion, and salivary secretion might be closely related with CCA development. Key genes such as RNF39 and PRSS56 might be potential prognostic markers for CCA. The results offer novel insights into the understanding of the development and prognosis of CCA.

Conflict of interest: All of the authors declare that there is no conflict of interest.

Funding: This study is supported by grants from the Nature Science Foundation of Liaoning Province (No. 20180530043).
Reference

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matise JC, Richardson JE, Ringwald M, Rubin GM, and Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25-29. 10.1038/75556

Bergquist A, and von Seth E. 2015. Epidemiology of cholangiocarcinoma. *Best Pract Res Clin Gastroenterol* 29:221-232. 10.1016/j.bpg.2015.02.003

Blechacz B. 2017. Cholangiocarcinoma: Current Knowledge and New Developments. *Gut Liver* 11:13-26. 10.5009/gnl15568

Boulter L, Guest RV, Kendall TJ, Wilson DH, Wojtacha D, Robson AJ, Ridgway RA, Samuel K, Van Rooijen N, Barry ST, Wigmore SJ, Sansom OJ, and Forbes SJ. 2015. WNT signaling drives cholangiocarcinoma growth and can be pharmacologically inhibited. *J Clin Invest* 125:1269-1285. 10.1172/jci76452

Bridgewater J, Galle PR, Khan SA, Llovet JM, Park JW, Patel T, Pawlik TM, and Gores GJ. 2014. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol* 60:1268-1289. 10.1016/j.jhep.2014.01.021

Chaiyarit P, Sithithaworn P, Thuwajit C, and Yongvanit P. 2011. Detection of salivary antibodies to crude antigens of Opisthorchis viverrini in opisthorchiasis and cholangiocarcinoma patients. *Clin Oral Investig* 15:477-483. 10.1007/s00784-010-0421-y

Chan-On W, Nairismagi ML, Ong CK, Lim WK, Dima S, Pairojkul C, Lim KH, McPherson JR, Cutcutache I, Heng HL, Ooi L, Chung A, Chow P, Cheow PC, Lee SY, Choo SP, Tan IB, Duda D, Nastase A, Myint SS, Wong BH, Gan A, Rajasegaran V, Ng CC, Nagarajan S, Usakul A, Zhang S, Vohra P, Yu W, Huang D, Sithithaworn P, Yongvanit P, Wongkham S, Khuntikeo N, Bhudhisawasdi V, Popescu I, Rozen SG, Tan P, and Teh BT. 2013. Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. *Nat Genet* 45:1474-1478. 10.1038/ng.2806
Chen YL, Jeng YM, Hsu HC, Lee PH, Lai PL, and Yuan RH. 2013. Expression of insulin-like growth factor II mRNA-binding protein 3 predicts early recurrence and poor prognosis in intrahepatic cholangiocarcinoma. *Int J Surg* 11:85-91. 10.1016/j.ijsu.2012.11.021

Clevers H, and Nusse R. 2012. Wnt/beta-catenin signaling and disease. *Cell* 149:1192-1205. 10.1016/j.cell.2012.05.012

Doherty B, Nambudiri VE, and Palmer WC. 2017. Update on the Diagnosis and Treatment of Cholangiocarcinoma. *Curr Gastroenterol Rep* 19:2. 10.1007/s11894-017-0542-4

Erichsen R, Jepsen P, Vilstrup H, Ekbom A, and Sorensen HT. 2009. Incidence and prognosis of cholangiocarcinoma in Danish patients with and without inflammatory bowel disease: a national cohort study, 1978-2003. *Eur J Epidemiol* 24:513-520. 10.1007/s10654-009-9365-4

Gu MJ, and Jang BI. 2014. Clinicopathologic significance of Sox2, CD44 and CD44v6 expression in intrahepatic cholangiocarcinoma. *Pathol Oncol Res* 20:655-660. 10.1007/s12253-014-9745-2

Jiang D, Yang Z, Li S, Xiao X, Jia X, Wang P, Guo X, Liu X, and Zhang Q. 2013. Evaluation of PRSS56 in Chinese subjects with hyperopia or primary angle-closure glaucoma. *Mol Vis* 19:2217-2226.

Jourdon A, Gresset A, Spassky N, Charnay P, Topilko P, and Santos R. 2016. Prss56, a novel marker of adult neurogenesis in the mouse brain. *Brain Struct Funct* 221:4411-4427. 10.1007/s00429-015-1171-z

Kanehisa M, Goto S, Kawashima S, Okuno Y, and Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277-280. 10.1093/nar/gkh063

Kurata R, Nakaoka H, Tajima A, Hosomichi K, Shiina T, Meguro A, Mizuki N, Ohono S, Inoue I, and Inoko H. 2010. TRIM39 and RNF39 are associated with Behcet's disease independently of HLA-B *51 and -A *26. *Biochem Biophys Res Commun* 401:533-537. 10.1016/j.bbrc.2010.09.088

Langfelder P, and Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. 10.1186/1471-2105-9-559

Lin YJ, Chen CY, Jeang KT, Liu X, Wang JH, Hung CH, Tsang H, Lin TH, Liao CC, Huang SM, Lin CW, Ho MW, Chien WK, Chen JH, Ho TJ, and Tsai FJ. 2014. Ring finger protein 39 genetic variants associate with HIV-1 plasma viral loads and its replication in cell culture. *Cell Biosci* 4:40. 10.1186/2045-3701-4-40

Liu R, Zhao R, Zhou X, Liang X, Campbell DJ, Zhang X, Zhang L, Shi R, Wang G, Pandak WM, Sirica AE, Hylemon PB, and Zhou H. 2014. Conjugated bile acids promote cholangiocarcinoma cell invasive growth through activation of sphingosine 1-phosphate receptor 2. *Hepatology* 60:908-918. 10.1002/hep.27085

Luo G, Li B, Duan C, Cheng Y, Xiao B, Yao F, Wei M, Tao Q, Feng C, Xia X, Zhou H, Zhao X, and Dai R. 2017. cMyc promotes cholangiocarcinoma cells to overcome contact inhibition via the mTOR pathway. *Oncol Rep* 38:2498-2506. 10.3892/or.2017.5913

Matsumoto M, Wakiyama S, Shiba H, Gocho T, Misawa T, Ishida Y, Itsubo M, Suzuki M, and Yanaga K. 2014. Combined hepatocellular-cholangiocarcinoma producing parathyroid hormone-related protein: report of a case. *Surg Today* 44:1577-1583. 10.1007/s00595-013-0714-2

Oliveira IS, Kilcoyne A, Everett JM, Mino-Kenudson M, Harisinghani MG, and Ganesan K. 2017. Cholangiocarcinoma: classification, diagnosis, staging, imaging features, and management. *Abdom Radiol (NY)* 42:1637-1649. 10.1007/s00261-017-1094-7

Ozawa N, Doi S, Tsujikawa T, Mabuchi M, Kajiyma Y, Sato K, Kikuchi K, Takahashi M, Kawamoto M, and Yasuda I. 2017. Intrahepatic cholangiocarcinoma producing granulocyte colony-stimulating factor and parathyroid hormone-related protein. *Nihon Shokakibyo Gakkai Zasshi* 114:1285-1292. 10.11405/nissshoshi.114.1285
Park J, Tadlock L, Gores GJ, and Patel T. 1999. Inhibition of interleukin 6-mediated mitogen-activated protein kinase activation attenuates growth of a cholangiocarcinoma cell line. *Hepatology* 30:1128-1133. 10.1002/hep.510300522

Razmilava N, and Gores GJ. 2014. Cholangiocarcinoma. *Lancet* 383:2168-2179. 10.1016/s0140-6736(13)61903-0

Rizvi S, Borad MJ, Patel T, and Gores GJ. 2014. Cholangiocarcinoma: molecular pathways and therapeutic opportunities. *Semin Liver Dis* 34:456-464. 10.1055/s-0034-1394144

Rizvi S, and Gores GJ. 2017. Emerging molecular therapeutic targets for cholangiocarcinoma. *J Hepatol* 67:632-644. 10.1016/j.jhep.2017.03.026

Song WS, Park HM, Ha JM, Shin SG, Park HG, Kim J, Zhang T, Ahn DH, Kim SM, Yang YH, Jeong JH, Theberge AB, Kim BG, Lee JK, and Kim YG. 2018. Discovery of glycocholic acid and taurochenodeoxycholic acid as phenotypic biomarkers in cholangiocarcinoma. *Sci Rep* 8:11088. 10.1038/s41598-018-29445-z

Squadroni M, Tondulli L, Gatta G, Mosconi S, Beretta G, and Labianca R. 2017. Cholangiocarcinoma. *Crit Rev Oncol Hematol* 116:11-31. 10.1016/j.critrevonc.2016.11.012

Sun Q, Li J, Wang G, and Xie Y. 2014. Role of the embryonic protein SOX2 in cholangiocarcinoma. *Cell Biochem Biophys* 70:1311-1316. 10.1007/s12013-014-0056-8

Tang J, Liao Y, He S, Shi J, Peng L, Xu X, Xie F, Diao N, Huang J, Xie Q, Lin C, Luo X, Liao K, Ma J, Li J, Zhou D, Li Z, Xu J, Zhong C, Wang G, and Bai L. 2017. Autocrine parathyroid hormone-like hormone promotes intrahepatic cholangiocarcinoma cell proliferation via increased ERK/JNK-ATF2-cyclinD1 signaling. *J Transl Med* 15:238. 10.1186/s12967-017-1342-1

Tian A, Pu K, Li B, Li M, Liu X, Gao L, Mao X. 2019. Weighted gene coexpression network analysis reveals hub genes involved in cholangiocarcinoma progression and prognosis. *Hepatol Res.* Jun 8. doi: 10.1111/hepr.13386.

Wang W, Zhong W, Yuan J, Yan C, Hu S, Tong Y, Mao Y, Hu T, Zhang B, and Song G. 2015. Involvement of Wnt/beta-catenin signaling in the mesenchymal stem cells promote metastatic growth and chemoresistance of cholangiocarcinoma. *Oncotarget* 6:42276-42289. 10.18632/oncotarget.5514

Zhang F, Wan M, Xu Y, Li Z, Leng K, Kang P, Cui Y, and Jiang X. 2017a. Long noncoding RNA PCAT1 regulates extrahepatic cholangiocarcinoma progression via the Wnt/beta-catenin-signaling pathway. *Biomed Pharmacother* 94:55-62. 10.1016/j.biopha.2017.07.025

Zhang S, Song X, Cao D, Xu Z, Fan B, Che L, Hu J, Chen B, Dong M, Pilo MG, Cigliano A, Evert K, Ribback S, Dombrowski F, Pascale RM, Cossu A, Vidili G, Porcu A, Simile MM, Pes GM, Giannelli G, Gordan J, Wei L, Evert M, Cong W, Calvisi DF, and Chen X. 2017b. Pan-mTOR inhibitor MLN0128 is effective against intrahepatic cholangiocarcinoma in mice. *J Hepatol* 67:1194-1203. 10.1016/j.jhep.2017.07.006
Figure legends

Figure 1. Interaction relationship analysis of co-expression genes in cholangiocarcinoma.

Figure 2. The module–clinical trait relationships of genes involved in clinicopathological parameters of cholangiocarcinoma patients.

Figure 3. Gene Ontology analysis for genes in the hub modules of tan module and salmon module in cholangiocarcinoma.

Figure 4. KEGG pathway enrichment analysis for genes in the hub modules of tan module and salmon module in cholangiocarcinoma.

Figure 5. The correlation between the expression levels of key genes of hub modules and the survival of CAA patients. A, INS; B, PPP3CA; C, CST1; D, RNF39; E, PRSS56; F, PRRG4; G, SLC4A4; H, PIR.

Figure 6. The differential expression of potential hub genes in different stages of CCA patients. A, differential expression of INS, PPP3CA, CST1 and RNF39; B, differential expression of PRSS56, PRRG4, SLC4A4 and PIR.

Supplementary Figure 1. Threshold selection of WGCNA analysis.
**Table 1** (on next page)

Gene ontology (GO) analysis of the genes involved in hub modules.

Gene ontology (GO) analysis of the genes involved in hub modules.
Table 1. Gene ontology (GO) analysis of the genes involved in hub modules.

| Module | ID       | Description                                | GeneRatio | P value     | Count |
|--------|----------|--------------------------------------------|-----------|-------------|-------|
| tan    | GO:0003179 | heart valve morphogenesis                  | 5/181     | 3.36E-05    | 5     |
| tan    | GO:0006739 | NADP metabolic process                     | 5/181     | 3.36E-05    | 5     |
| tan    | GO:0003170 | heart valve development                    | 5/181     | 5.67E-05    | 5     |
| tan    | GO:0003281 | ventricular septum development             | 6/181     | 5.83E-05    | 6     |
| tan    | GO:0060411 | cardiac septum morphogenesis               | 6/181     | 6.92E-05    | 6     |
| tan    | GO:0003279 | cardiac septum development                 | 7/181     | 7.37E-05    | 7     |
| tan    | GO:0072537 | fibroblast activation                      | 3/181     | 0.00013     | 3     |
| tan    | GO:0016055 | Wnt signaling pathway                      | 15/181    | 0.00016     | 15    |
| tan    | GO:0198738 | cell-cell signaling by wnt                  | 15/181    | 0.00017     | 15    |
| salmon | GO:0007586 | digestion                                  | 16/148    | 1.02E-12    | 16    |
| salmon | GO:0010817 | regulation of hormone levels                | 22/148    | 1.53E-10    | 22    |
| salmon | GO:0009914 | hormone transport                          | 15/148    | 7.10E-08    | 15    |
| salmon | GO:0046879 | hormone secretion                          | 14/148    | 2.85E-07    | 14    |
| salmon | GO:0023061 | signal release                             | 16/148    | 6.22E-07    | 16    |
| salmon | GO:0030072 | peptide hormone secretion                  | 12/148    | 1.16E-06    | 12    |
| salmon | GO:0046883 | regulation of hormone secretion             | 12/148    | 1.84E-06    | 12    |
| salmon | GO:0050679 | positive regulation of epithelial cell proliferation | 10/148 | 2.07E-06 | 10 |
| salmon | GO:0090276 | regulation of peptide hormone secretion     | 10/148    | 8.73E-06    | 10    |
| salmon | GO:0050673 | epithelial cell proliferation               | 13/148    | 1.32E-05    | 13    |
Table 2 (on next page)

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis of the genes involved in hub modules.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis of the genes involved in hub modules.
Table 2. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis of the genes involved in hub modules.

| Module | ID       | Description                                      | GeneRatio | P value     | Count |
|--------|----------|--------------------------------------------------|-----------|-------------|-------|
| tan    | hsa05224 | Breast cancer                                    | 10/81     | 3.65E-06    | 10    |
| tan    | hsa00480 | Glutathione metabolism                           | 6/81      | 2.84E-05    | 6     |
| tan    | hsa04310 | Wnt signaling pathway                            | 8/81      | 0.00017     | 8     |
| tan    | hsa05230 | Central carbon metabolism in cancer              | 5/81      | 0.00066     | 5     |
| tan    | hsa05226 | Gastric cancer                                   | 7/81      | 0.00111     | 7     |
| tan    | hsa04150 | mTOR signaling pathway                           | 7/81      | 0.0012      | 7     |
| salmon | hsa04972 | Pancreatic secretion                             | 11/64     | 3.74E-10    | 11    |
| salmon | hsa04974 | Protein digestion and absorption                 | 10/64     | 3.45E-09    | 10    |
| salmon | hsa05217 | Basal cell carcinoma                             | 3/64      | 0.01654     | 3     |
| salmon | hsa04360 | Axon guidance                                    | 5/64      | 0.0167      | 5     |
| salmon | hsa00830 | Retinol metabolism                               | 3/64      | 0.01948     | 3     |
| salmon | hsa04380 | Osteoclast differentiation                       | 4/64      | 0.02372     | 4     |
| salmon | hsa04911 | Insulin secretion                                | 3/64      | 0.03605     | 3     |
| salmon | hsa04970 | Salivary secretion                               | 3/64      | 0.04162     | 3     |
| salmon | hsa04975 | Fat digestion and absorption                     | 2/64      | 0.04786     | 2     |
Table 3 (on next page)

Key genes associated with prognosis of cholangiocarcinoma patients in hub modules.
Table 3. Key genes associated with prognosis of cholangiocarcinoma patients in hub modules.

| Gene       | Full name                                      | Module | HR   | 95%CI     | P value |
|------------|------------------------------------------------|--------|------|-----------|---------|
| SOX2       | SRY-box 2                                      | tan    | 0.382| 0.145-1.002| 0.043   |
| KIT        | KIT proto-oncogene receptor tyrosine kinase    | tan    | 0.341| 0.132-0.883| 0.048   |
| PRSS56     | serine protease 56                             | tan    | 2.876| 1.048-7.892| 0.018   |
| WNT9A      | Wnt family member 9A                           | tan    | 0.389| 0.144-1.050| 0.041   |
| SLC4A4     | solute carrier family 4 member 4               | tan    | 0.352| 0.133-0.931| 0.025   |
| PRRG4      | proline rich and Gla domain 4                  | tan    | 0.339| 0.127-0.900| 0.023   |
| PANX2      | pannexin 2                                     | tan    | 2.600| 0.963-7.022| 0.037   |
| PIR        | pirin                                          | tan    | 2.701| 0.994-7.337| 0.027   |
| RASSF8     | Ras association domain family member 8         | tan    | 0.396| 0.147-1.065| 0.040   |
| MFS4D4A    | major facilitator superfamily domain containing 4A | salmon | 0.371| 0.143-0.962| 0.043   |
| INS        | insulin                                        | salmon | 4.670| 0.458-47.564| 0.006   |
| RNF39      | ring finger protein 39                         | salmon | 2.811| 1.063-7.434| 0.032   |
| IL1R2      | interleukin 1 receptor type 2                  | salmon | 0.379| 0.144-0.997| 0.041   |
| CST1       | cystatin SN                                    | salmon | 2.846| 1.039-7.797| 0.017   |
| PPP3CA     | protein phosphatase 3 catalytic subunit alpha   | salmon | 3.151| 1.129-8.793| 0.010   |
Table 4 (on next page)

The association of prognosis-related genes with clinical of stage of cholangiocarcinoma patients.

The association of prognosis-related genes with clinical of stage of cholangiocarcinoma patients.
Table 4. The association of prognosis-related genes with clinical of stage of cholangiocarcinoma patients.

| Gene symbol | Gene title                                           | Position | OR         | P         |
|-------------|------------------------------------------------------|----------|------------|-----------|
| MFSD4A      | Major Facilitator Superfamily Domain Containing 4A   | 1q32.1   | 1.036(0.988-1.093) | 0.155     |
| SOX2        | SRY-Box 2                                            | 3q26.33  | 1.037(0.986-1.115) | 0.184     |
| INS         | Insulin                                              | 11p15.5  | 1.125(0.996-1.320) | 0.059     |
| KIT         | KIT Proto-Oncogene Receptor Tyrosine Kinase          | 4q12     | 1.033(0.991-1.091) | 0.146     |
| PRSS56      | Serine Protease 56                                   | 2q37.1   | 1.162(1.002-1.431) | 0.013     |
| RNF39       | Ring Finger Protein 39                               | 6p22.1   | 1.069(1.006-1.167) | 0.019     |
| WNT9A       | Wnt Family Member 9A                                 | 1q42.13  | 0.965(0.904-1.018) | 0.235     |
| SLC4A4      | Solute Carrier Family 4 Member 4                     | 4q13.3   | 1.001(0.973-1.029) | 0.945     |
| IL1R2       | Interleukin 1 Receptor Type 2                        | 2q11.2   | 1.027(0.993-1.068) | 0.131     |
| CST1        | Cystatin SN                                          | 20p11.21 | 1.007(0.988-1.026) | 0.490     |
| PRRG4       | Proline Rich And Gla Domain 4                        | 11p13    | 0.983(0.922-1.041) | 0.574     |
| PANX2       | Pannexin 2                                           | 22q13.33 | 0.987(0.940-1.024) | 0.517     |
| PIR         | Pirin                                                | Xp22.2   | 0.983(0.930-1.027) | 0.489     |
| RASSF8      | Serine Protease 8                                    | 12p12.1  | 0.942(0.863-1.002) | 0.094     |
| PPP3CA      | Protein Phosphatase 3 Catalytic Subunit Alpha        | 4q24     | 1.021(0.960-1.086) | 0.492     |
Figure 1

Interaction relationship analysis of co-expression genes in cholangiocarcinoma.
Figure 2

The module–clinical trait relationships of genes involved in clinicopathological parameters of cholangiocarcinoma patients.

The module–clinical trait relationships of genes involved in clinicopathological parameters of cholangiocarcinoma patients.
### Module–trait relationships

| Heatmap | Value 1 | Value 2 | Value 3 | Value 4 | Value 5 | Value 6 |
|---------|---------|---------|---------|---------|---------|---------|
| MEroyalblue | -0.099 (0.6) | 0.074 (0.7) | -0.13 (0.5) | -0.27 (0.1) | -0.09 (0.6) | -0.12 (0.5) |
| MEmagenta | -0.15 (0.4) | 0.14 (0.5) | 0.17 (0.3) | -0.15 (0.4) | -0.021 (0.9) | -0.014 (0.8) |
| MEBlue | -0.19 (0.3) | 0.083 (0.8) | 0.056 (0.8) | -0.33 (0.06) | -0.1 (0.8) | -0.14 (0.4) |
| METurquoise | 0.025 (0.9) | 0.14 (0.4) | 0.13 (0.5) | -0.26 (0.1) | -0.063 (0.7) | -0.012 (0.9) |
| MEGreen | -0.045 (0.8) | 0.092 (0.6) | -0.28 (0.1) | -0.35 (0.04) | -0.00038 (1) | -0.1 (0.6) |
| MELightgreen | 0.00095 (1) | 0.12 (0.5) | 0.14 (0.4) | 0.16 (0.4) | 0.31 (0.08) | 0.31 (0.08) |
| MElightyellow | -0.099 (0.6) | 0.22 (0.2) | -0.12 (0.5) | -0.057 (0.8) | -0.18 (0.3) | -0.17 (0.3) |
| MEBlack | 0.032 (0.9) | -0.19 (0.3) | -0.096 (0.6) | -0.12 (0.5) | -0.13 (0.5) | -0.13 (0.5) |
| MEGreenyellow | 0.42 (0.01) | 0.23 (0.2) | 0.091 (0.6) | 0.1 (0.6) | -0.13 (0.5) | -0.0041 (1) |
| MEBrown | 0.35 (0.04) | -0.069 (0.7) | 0.37 (0.03) | 0.19 (0.3) | 0.076 (0.7) | 0.26 (0.1) |
| MELightbrown | 0.098 (0.6) | -0.23 (0.2) | -0.052 (0.6) | -0.28 (0.1) | -0.27 (0.1) | -0.2 (0.3) |
| MEpurple | -0.11 (0.6) | 0.14 (0.4) | 0.17 (0.3) | 0.19 (0.3) | -0.064 (0.7) | -0.074 (0.7) |
| MECyan | -0.27 (0.1) | 0.038 (0.8) | 0.13 (0.5) | 0.15 (0.4) | -0.041 (0.1) | -0.0996 (1) |
| MESalmon | -0.45 (0.009) | -0.11 (0.5) | 0.38 (0.03) | 0.41 (0.02) | 0.31 (0.06) | 0.27 (0.1) |
| MEgrey60 | -0.1 (0.6) | -0.072 (0.7) | 0.015 (0.9) | -0.23 (0.2) | -0.026 (0.9) | -0.096 (0.8) |
| MEmidnightblue | -0.32 (0.07) | -0.21 (0.2) | -0.0083 (1) | -0.0056 (1) | 0.022 (0.9) | -0.064 (0.7) |
| MEred | -0.16 (0.4) | 0.11 (0.6) | 0.17 (0.4) | 0.25 (0.2) | 0.38 (0.1) | 0.19 (0.3) |
| MElightcyan | -0.26 (0.1) | -0.22 (0.2) | -0.23 (0.2) | -0.24 (0.2) | -0.18 (0.3) | -0.25 (0.2) |
| MEBrown | -0.23 (0.2) | 0.028 (0.9) | -0.23 (0.2) | -0.24 (0.2) | -0.51 (0.08) | -0.35 (0.05) |
| MEPink | -0.16 (0.4) | -0.17 (0.2) | -0.24 (0.2) | -0.25 (0.2) | -0.31 (0.09) | -0.34 (0.05) |
| MEgrey | 0.24 (0.2) | 0.17 (0.3) | -0.2 (0.3) | 0.026 (0.9) | -0.11 (0.5) | -0.056 (0.8) |
Figure 3

Gene Ontology analysis for genes in the hub modules of tan module and salmon module in cholangiocarcinoma.
Figure 4

The correlation between the expression levels of key genes of hub modules and the survival of CAA patients.

The correlation between the expression levels of key genes of hub modules and the survival of CAA patients.
Figure 5

The correlation between the expression levels of key genes of hub modules and the survival of CAA patients.

The correlation between the expression levels of key genes of hub modules and the survival of CAA patients.
Figure 6

The differential expression of potential hub genes in different stages of CCA patients.