

ANKRD1 and SPP1 as diagnostic markers and correlated with immune infiltration in biliary atresia

Meng Kong, PhD^a, Teng Ma, MD^b, Bo Xiang, PhD^{c,*} 

Abstract

The diagnosis of biliary atresia (BA) remains a clinical challenge, reliable biomarkers that can easily distinguish BA and other forms of intrahepatic cholestasis (IC) are urgently needed.

Differentially expressed genes were identified by R software. The least absolute shrinkage and selection operator regression and support vector machine algorithms were used to filter the diagnostic biomarkers of BA. The candidate biomarkers were further validated in another independent cohort of patients with BA and IC. Then CIBERSORT was used for estimating the fractions of immune cell types in BA. Gene set enrichment analyses were conducted and the correlation between diagnostic genes and immune cells was analyzed.

A total of 419 differentially expressed genes in BA were detected and 2 genes (secreted phosphoprotein 1 [SPP1] and ankyrin repeat domain [ANKRD1]) among them were selected as diagnostic biomarkers. The SPP1 yielded an area under the curve (AUC) value of 0.798 (95% confidence interval [CI]: 0.742–0.854) to distinguish patients with BA from those with IC, and ANKRD1 exhibited AUC values of 0.686 (95% CI: 0.616–0.754) in discriminating BA patients and those with IC. Further integrating them into one variable resulted in a higher AUC of 0.830 (95% CI: 0.777–0.879). The regulatory T cells, M2 macrophages cells, CD4 memory T cells, and dendritic cells may be involved in the BA process. The ANKRD1 and SPP1 was negatively correlated with regulatory T cells.

In conclusion, the ANKRD1 and SPP1 could potentially provide extra guidance in discriminating BA and IC. The immune cell infiltration of BA gives us new insight to explore its pathogenesis.

Abbreviations: ANKRD1 = ankyrin repeat domain 1, AUC = area under the curve, BA = biliary atresia, CI = confidence interval, DEGs = differentially expressed genes, GSEA = gene set enrichment analyses, IC = intrahepatic cholestasis, LASSO = least absolute shrinkage and selection operator, SPP1 = secreted phosphoprotein 1, SVM = support vector machine.

Keywords: biliary atresia, CIBERSORT, diagnostic, immune cells

1. Introduction

Biliary atresia (BA) is a serious neonatal disease, which eventually leads to cholestatic liver cirrhosis due to the inflammation and

progressive fibrosis of intrahepatic and extrahepatic bile ducts.^[1] In a proportion of BA patients have no other option but to undergo liver transplantation.^[2] Differentiation between BA and other forms of intrahepatic cholestasis (IC) is difficult. The current gold standard for the diagnosis of BA is exploratory laparotomy or intraoperative laparoscopic cholangiography, but this is an invasive examination method for children.^[3] Abdominal ultrasound and hepatobiliary scintigraphy may be helpful in differentiating BA from other forms of IC. However, the accuracy of abdominal ultrasound is limited due to the experience of the operator.^[4] And the specificity of hepatobiliary scintigraphy is rather low.^[5] Some non-BA children have to undergo an exploratory laparotomy or laparoscopic surgery for definite diagnosis. Therefore, the simple, effective noninvasive biomarkers to provide clues for the early diagnosis of BA are urgently needed, which has great significance to simplify the diagnostic process.

The continuous development of gene chip and high-throughput sequencing technology provides important clues for uncovering diagnostic biomarker associated with BA. One study used a specially designed multi-gene panel that can correctly detect large indels simultaneously and made a potential genetic diagnosis in 22% patients with IC.^[6] Another previous study reported that gene sequence analysis assigned a molecular diagnosis in 27% patients with idiopathic cholestasis.^[7] A national research reported a molecular diagnosis in 26% neonatal and infantile IC.^[8] It is clear that the molecular diagnostics will become increasingly important and may become an early differential diagnostic method to BA.

Editor: Muhsin Jamal.

This work was supported by the Health and Family Planning Commission of Jinan (grant2017-2-21).

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

^a Department of Pediatric Surgery, Qilu Children's Hospital of Shandong University, Jinan, China, ^b Department of Internal Medicine, The Fifth People's Hospital of Jinan, Jinan, China, ^c Department of Pediatric Surgery, West China Hospital of Sichuan University, Chengdu, China.

* Correspondence: Bo Xiang, Department of Pediatric Surgery, West China Hospital of Sichuan University, No. 37 Guoxue Alley, Wuhou Dist, Chengdu 610041, China (e-mail: xb_scu_edu@163.com).

Copyright © 2021 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Kong M, Ma T, Xiang B. ANKRD1 and SPP1 as diagnostic markers and correlated with immune infiltration in biliary atresia. *Medicine* 2021;100:50(e28197).

Received: 18 February 2021 / Received in final form: 18 November 2021 / Accepted: 19 November 2021

<http://dx.doi.org/10.1097/MD.00000000000028197>

In recent years, more and more studies have shown that immune cell infiltration plays an important role in the occurrence and development of BA.^[9] Some scholars show that dendritic cells regulate natural killer cell activation and epithelial injury in experimental BA.^[10] A previous study show that infiltration of polarized macrophages associated with liver fibrosis in infants with BA.^[11] Therefore, a systematically discrimination of the liver immune infiltration in BA children, and a comprehensive evaluation of the liver immune microenvironment in BA, is helpful to clarify the reason of BA. It is also helpful to lay a theoretical foundation for preventing the occurrence of BA and improving the prognosis of BA, and provides new clues and new perspectives on the prevention or treatment of the disease.

In this study, gene expression data of BA from the public database were download. Then we performed bioinformatics analysis and used machine learning algorithms to further screen and determine^[12] the diagnostic biomarkers of BA. Subsequently, using CIBERSORT algorithm, we conduct a comprehensive evaluation of BA immune infiltration. In addition, the correlations between diagnostic biomarker and immune cells were analyzed to better understand the molecular immune mechanism during the development of BA.

2. Materials and methods

2.1. Data preprocessing and differentially expressed genes (DEGs) screening

The ethics committee of west China Hospital of Sichuan University approved the study. Gene expression data of BA or the other forms of IC were downloaded from the public data platform Gene Expression Omnibus, which included GSE46960 (64 BA samples, 14 diseased control samples, and 7 normal control samples), GSE15235 (samples from 47 patients with BA), and GSE119600 (90 primary biliary cholangitis samples and 45 primary sclerosing cholangitis samples). The inclusion and exclusion criteria or the age of the children included in the study can be seen in previous studies.^[13,14] Raw data were download through the “GEOquery” package of R software, R 3.6 (New Zealand), and the robust multi-array analysis algorithm was used for background correction and data normalization. We first explore the differentially expressed genes (DEGs) on the discovery cohort (GSE46960). DEGs were screened by the Linear Models for Microarray Data (limma) package in R software. DEGs with $P < .05$ and $|\log_2FC| > 2$ were considered statistically significant.

2.2. Screening and verification of diagnostic markers

In order to select the candidate diagnostic genes from DEGs, 2 kinds of feature selection methods, including least absolute shrinkage and selection operator (LASSO) logistic regression and support vector machine (SVM)-recursive feature elimination, were used to build diagnostic panels to discriminate BA and other forms of IC. The “glmnet” package in R software was applied to perform the LASSO algorithm.^[15] The diagnostic value of candidate genes in BA were classified and analyzed by SVM classifier of R package e1071.^[16] The GSE46960, GSE15235, and GSE119600 datasets were then combined as validation cohort, and The batch effect of the validation cohort was removed by the “Remove Batch Effect” function from limma package.^[17] Principal component analysis plot was generated after the inter-sample correction.^[18] The validation cohort consisted of 111 BA samples and 149 non-BA samples. The

diagnostic efficiency of the candidate gene markers (sensitivity, specificity) was analyzed using the receiver operating characteristics curve based on this validation cohort. Ultimately, the overlapped genes extracted from LASSO algorithm and SVM methods were regarded as the final diagnostic feature.

2.3. Gene set enrichment analysis

Enrichment analysis was performed on the GSE46960 gene sets using GSEA (<https://www.gsea-msigdb.org/gsea/login.jsp>).^[19] Annotated gene sets h.all.v7.1.entrez.gmt was downloaded from the MsigDB database as the reference gene sets. We carried out GSEA and identified differentially regulated hallmark gene sets between normal and BA samples. The random classification frequency was performed with 1000 permutations. The related pathways statistically enriched in each group were selected when the false discovery rate was < 0.25 and the nominal $P < .05$.

2.4. Evaluation of immune cell infiltration

For estimation of the abundance of immune cells in BA samples, we uploaded the gene expression matrix data to CIBERSORT web portal (<http://cibersort.stanford.edu/>). After deleting samples with $P > .05$, a reference set with 22 sorted immune cell subtypes (LM22) was used and finally obtained the immune cell infiltration matrix. In order to illustrate this correlation of 22 immune cell, a heatmap format generated by the corrplot package of R Studio software. A box plot was generated to illustrate immune cell infiltration difference using R Studio software. Next, Spearman correlation analysis was used to evaluate relationships between candidate diagnostic genes and different immune cell subsets. Correlations were calculated using “ggstatsplot” package in CRAN (<https://cran.r-project.org/web/packages/ggstatsplot/index.html>). The ggplot2 package in R platform was used to visualize.

3. Results

3.1. Data preprocessing and DEGs screening

Based on the cutoff criteria in the methods, total 419 DEGs were identified when the BA compared with normal control, including 225 upregulated genes and 194 downregulated genes. When the BA compared with diseased control, there were 18 DEGs, including 15 upregulated genes and 3 downregulated genes. A total of 16 genes overlapped among the 2 set, and the overlapped DEGs was shown in the heatmap (Fig. 1).

3.2. Screening and verification of diagnostic markers

In our study, we obtained 8 potential diagnostic genes (ankyrin repeat domain [ANKRD1], VCAN, LUM, ITGA2, MMP7, EPCAM, secreted phosphoprotein 1 [SPP1], CCL2) that were screen out through LASSO algorithm in the overlapped DEGs (Fig. 2A). By SVM-recursive feature elimination analysis, a total of 3 genes (SERPINE1, ANKRD1, SPP1) were screen out as the most important variable to provide diagnosis method for BA (Fig. 2B). Finally, a total of 2 genes (ANKRD1, SPP1) simultaneously recommended by the 2 algorithms were defined as the final diagnostic related genes. The mRNA expression of ANKRD1 (Fig. 2C) and SPP1 (Fig. 2D) was increased in BA tissues when compared with other forms of IC and normal



Figure 1. Heat maps show the expression of 16 genes in GSE46960. (BA=biliary atresia, DC=diseased control, other forms of intrahepatic cholestasis, NC=normal control; levels of gene expression are shown as color variation from red [high] to blue [low]).

control. In order to explore the diagnostic efficiency of ANKRD1 and SPP1 for BA, a total of 111 cases of BA and 149 non-BA specimens from 3 merged datasets (GSE46960, GSE15235, and GSE119600) were included in our study as test cohort. Additionally, principal component analysis plot was generated after the batch effects among those 3 eligible datasets were adjusted by “Remove Batch Effect” function (Fig. 2E). The area under the curve (AUC) of ANKRD1 for the diagnosis of BA was 0.686 (95% confidence interval [CI]: 0.616–0.754) with a sensitivity of 86.58%, a specificity of 60.36%. The AUC of SPP1 for the diagnosis of BA was 0.798 (95% CI: 0.742–0.854) with a sensitivity of 69.80%, a specificity of 77.48%. Additionally, when ANKRD1 and SPP1 were fitted into 1 variable, the combined diagnostic genes also achieved a well diagnosis performance in BA from the validation cohort (Fig. 2F), the AUC was 0.830 (95% CI: 0.777–0.879) with a sensitivity of 71.81%, a specificity of 80.18%.

3.3. Gene set enrichment analysis

GSEA was used to conduct biological function and key pathway enrichment analysis, we compared levels of signaling pathway enrichment between the BA and normal samples in GSE46960. Gene sets were found to be activated in BA were mainly involved in immunological processes such as TNF- α signaling via NF- κ B,

inflammatory response, and IL2-STAT5 signaling. In addition, signaling pathways related to the bile acid metabolism, cholesterol homeostasis and fatty acid metabolism were suppressed in the BA group (Fig. 3). This finding initiate our hypothesis that the inflammatory response plays a pivotal role in BA and whether ANKRD1 and SPP1 are involved in immune-related biological processes.

3.4. Immune cell infiltration results

The barplot showed that the infiltration rates of M2 macrophages cells (green), resting CD4 memory T cells (yellow) and resting dendritic cells (blue) in each sample were relatively high (Fig. 4A). The correlation heatmap indicated that there existed difference in immune cell infiltration pattern between BA and DC groups (Fig. 4B). For instance, Regulatory T cells (Tregs) had a significant negative correlation with naïve CD4 T cells ($r = -0.66$), while the correlation between regulatory T cells (Tregs) and resting CD4 memory T cells in BA was very high ($r = 0.68$). The box plot showed that the infiltration ratio of activated Mast cells, resting CD4 memory T cells and gamma delta T cells in BA samples was significantly higher than that in diseased control samples ($P < .05$), while the infiltration ratio of naïve B cells and resting Mast cells in BA samples was lower than that in DC samples ($P < .05$) (Fig. 4C). Besides, the immune correlation

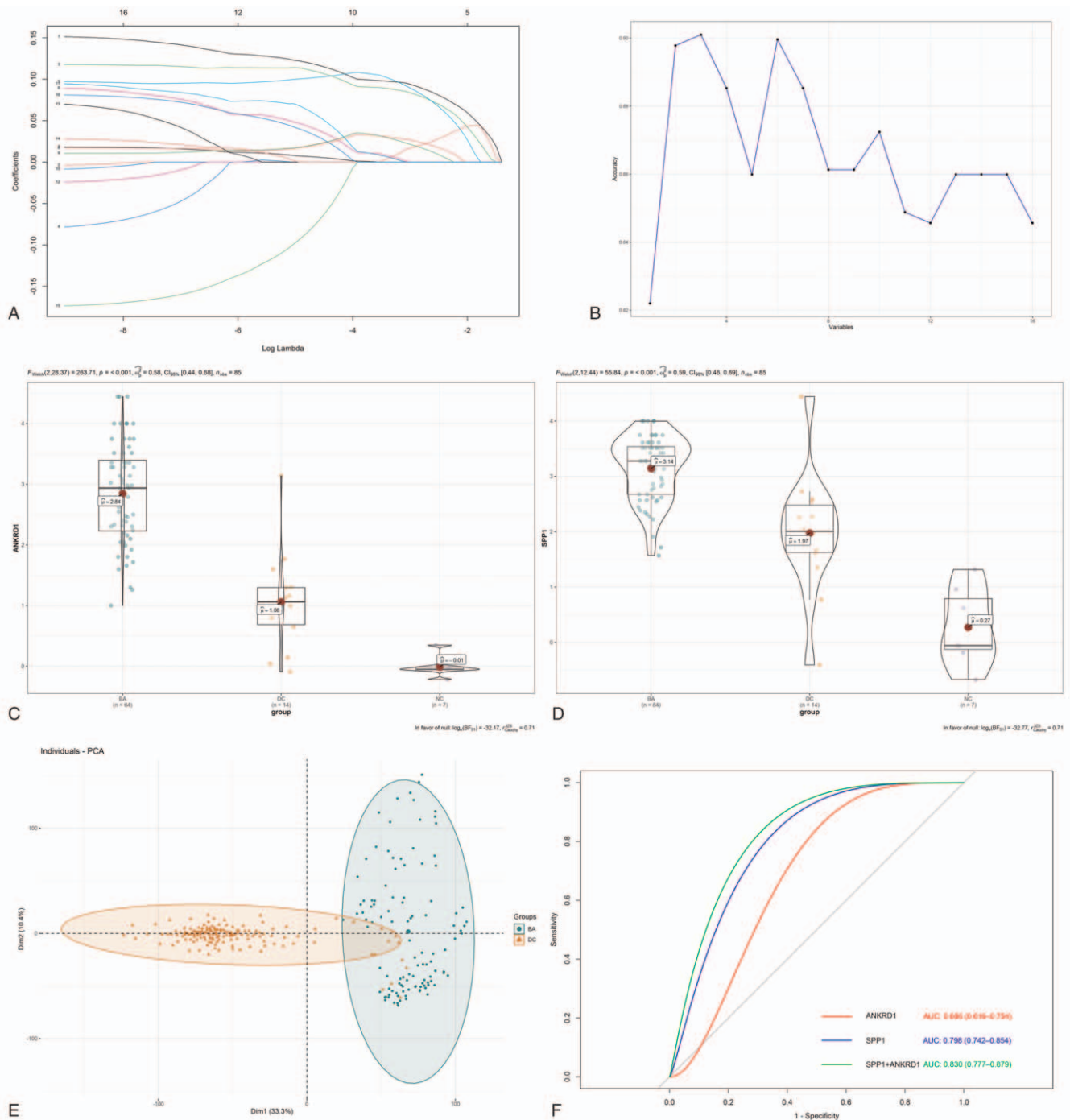


Figure 2. Screening and verification of diagnostic markers. (A) Least absolute shrinkage and selection operator (LASSO) regression to screen diagnostic markers. Different colors represent different genes. (B) Support vector machine (SVM) to screen diagnostic markers. (C) The difference of ankyrin repeat domain 1 (ANKRD1) expression in biliary atresia (BA), other forms of intrahepatic cholestasis (DC) and normal control (NC). (D) The difference of SPP1 expression in biliary atresia (BA), other forms of intrahepatic cholestasis (DC) and normal control (NC). (E) PCA cluster plot after batch effects correction. (F) Receive operating characteristic (ROC) plots.

analysis showed that SPP1 was positively correlated with resting CD4 memory T cells ($r = 0.501, P = .02$) and negatively correlated with regulatory T cells ($r = -0.611, P = .02$) (Fig. 5A); ANKRD1 was positively correlated with resting CD4 memory T cells ($r = 0.512, P = .03$) and negatively correlated with regulatory T cells ($r = -0.362, P = .03$) (Fig. 5B).

4. Discussion

Early diagnosis and intervention of BA in infants are of great significance. Preoperative diagnostic methods include liver biopsy or cholangiography is complicated and invasive, which will bring more potential harm to suspected infants. Thus, the development of simple and effective diagnostic biomarkers, has the potential to

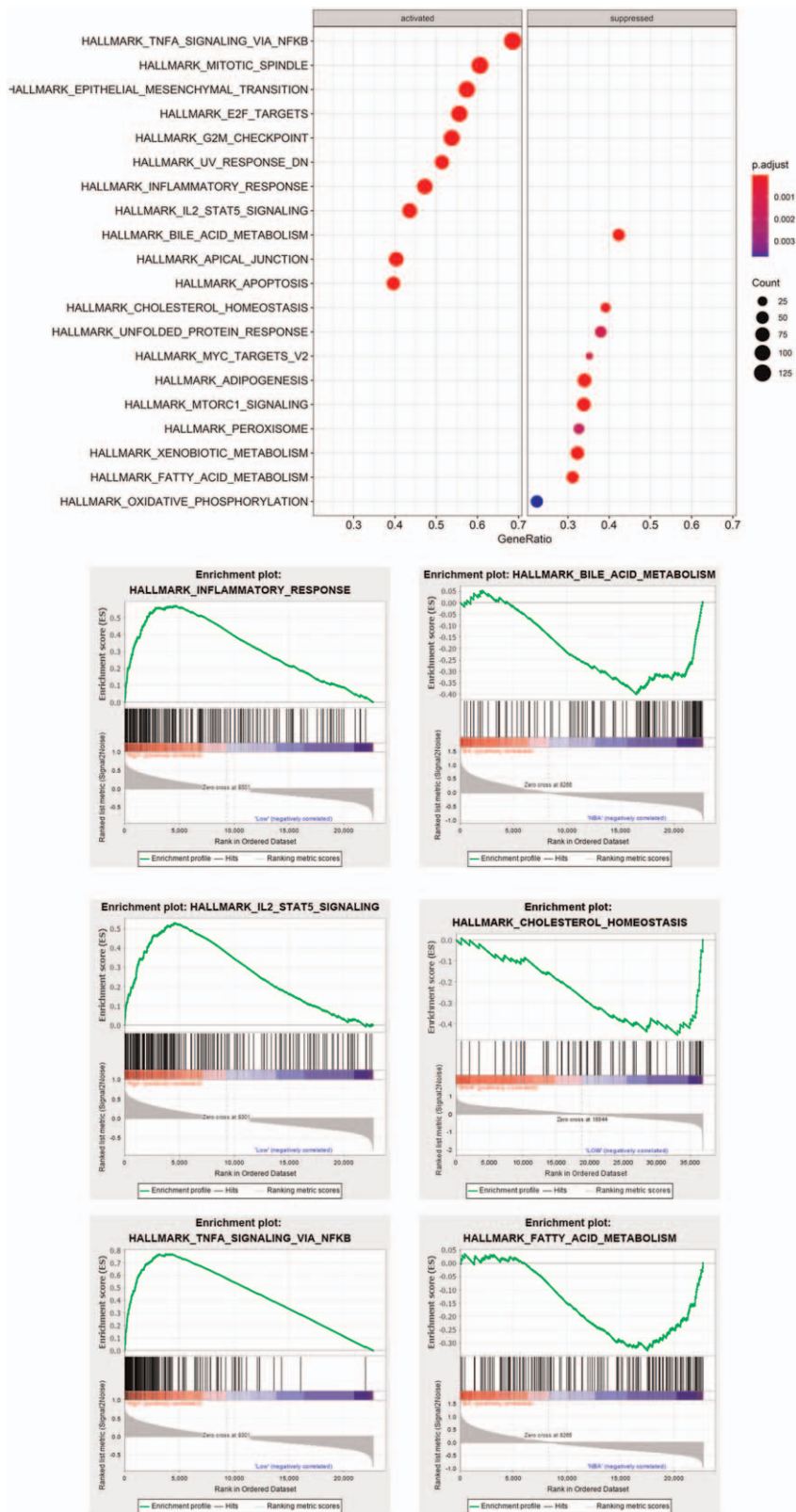


Figure 3. Gene set enrichment analysis.

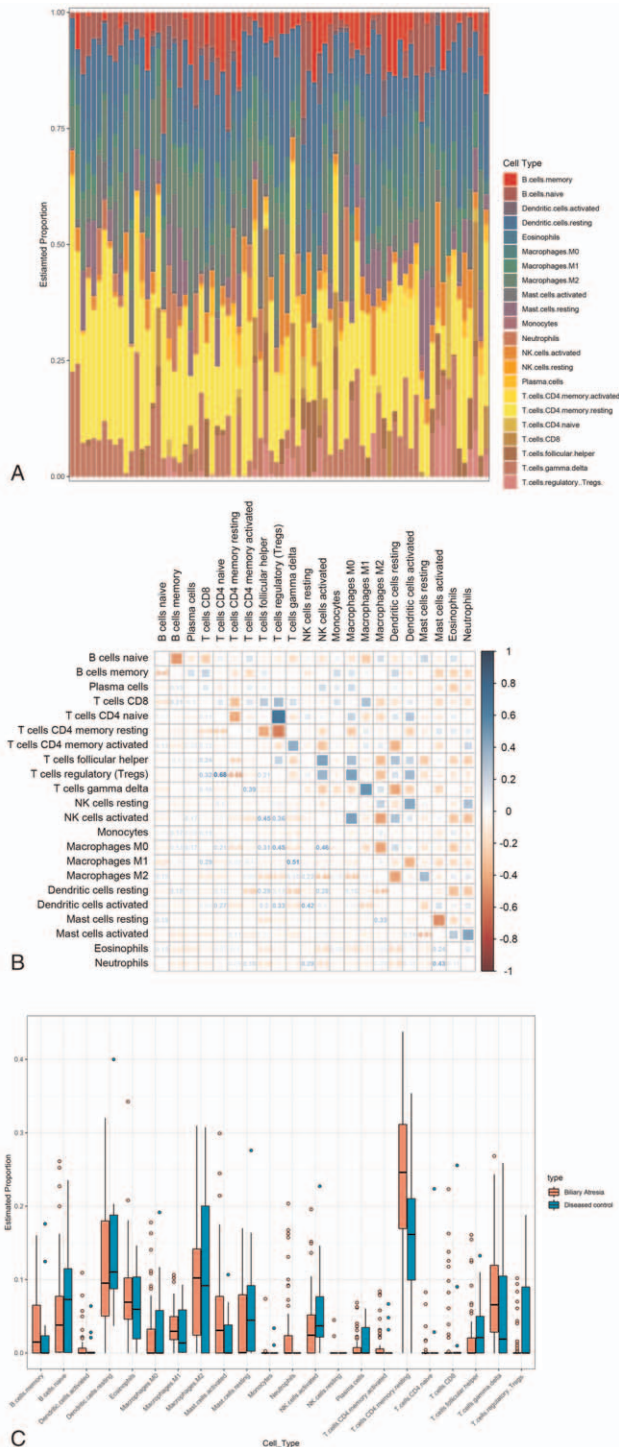


Figure 4. Evaluation and visualization of immune cell infiltration. (A) The bar plot of immune cell infiltration in biliary atresia. (B) Correlation heatmap of 22 types of immune cells (blue represents a positive correlation, red represents a negative correlation, the darker the color, the stronger the correlation). (C) The box plot of the proportion of 22 types of immune cells (blue represents other forms of intrahepatic cholestasis group, red represents biliary atresia group).

significantly shorten and simplify BA diagnosis process. In this study, we find and validated a diagnostic biomarker to discriminate between BA and other forms of IC, which might provide a novel and promising field in the diagnosis of BA.

In the present study, BA datasets were downloaded from Gene Expression Omnibus and screened for DEGs. Total 16 overlap DEGs were screened from GSE46960 datasets. Then, SVM and LASSO regression methods were applied to filter the DEGs. Finally, SPP1 and ANKRD1 that may be related to the diagnosis of BA were found by the 2 machine learning method. Additionally, when ANKRD1 and SPP1 were fitted into 1 variable, the combined diagnostic genes achieved a well diagnosis performance in BA, the AUC was 0.830 with a sensitivity of 71.81%, a specificity of 80.18%. The SPP1, also called osteopontin, is a protein coding gene. The protein encoded by this gene is secreted, acts as a cytokine involved in enhancing production of IFN- γ and IL-12 and reducing production of IL-10.^[20] SPP1 can induce Th1 cytokine activity and promote fibrogenesis, both of which has previously been linked to the pathogenesis of human BA.^[21] A previous study showed that a modest upregulation of SPP1 expression in rotavirus infected cholangiocytes, and the SPP1 high-expressing cells in the liver are cholangiocytes.^[22] Similar with our findings, we also find SPP1 was upregulated in BA. Another previous research showed that SPP1 mRNA and protein expression were significantly increased in BA when compared with other cholestatic diseases, which leads to biliary proliferation and hepatic portal fibrosis.^[23] A research group had reported that SPP1 protein levels were increased in BA, with the most striking increases observed in Kupffer cell and hepatic stellate cells in the necrotic areas, which correlates with liver fibrosis.^[24] Given that SPP1 encoded proteins often interact with BA, we believe that SPP1 is likely to be involved in the pathological process of BA and is a promising serum diagnostic biomarker. The ANKRD1 is a protein coding gene, localized to the nucleus of endothelial cells and is induced by IL-1 and TNF- α stimulation.^[25] In hepatitis C virus (HCV)-infected cells, the upregulation of ANKRD1 expression is mediated through altered calcium homeostasis and endoplasmic reticulum stress. The protein expression level of ANKRD1 could be increased by various stresses, which may be associated with virus infection in the liver.^[12] ANKRD1 dysregulation is associated with various viral replications and ANKRD1 is one of the 50 highly upregulated genes during acute HCV infection.^[26] In B cells, ANKRD1 is involved in HIV replication through the interaction with gp120.^[27] In addition, ANKRD1 is upregulated in all 3 common types of human papillomavirus infections.^[28] The ANKRD1 gene expression in activated hepatic stellate cells was completely dependent on YAP, has been directly associated with liver fibrosis and regenerative processes in other tissues.^[29] Therefore, we hypothesize that ANKRD1 might play an important role in the progression of BA. The results suggested that SPP1 and ANKRD1 probably be a potential diagnostic biomarker in BA, further studies are still needed to evaluate the diagnostic accuracy in clinical applications.

Gene set enrichment analysis was used to perform pathway enrichment analysis. Gene sets were found to be activated in BA were mainly involved in immunological processes such as TNF- α signaling via NF- κ B, inflammatory response, and IL2-STAT5 signaling. Previous work showed that the expression of TNF α are up-regulated in the liver tissue of BA model mice and the over expression of TNF α eventually lead to bile duct epithelial cell damage and bile duct obstruction, but the molecular mechanism is unclear.^[30] One laboratory reported that the IL2 plays a role in the pathogenesis of BA and is associated with increased infiltration of CD8+ T cells.^[31] Accumulating evidence show

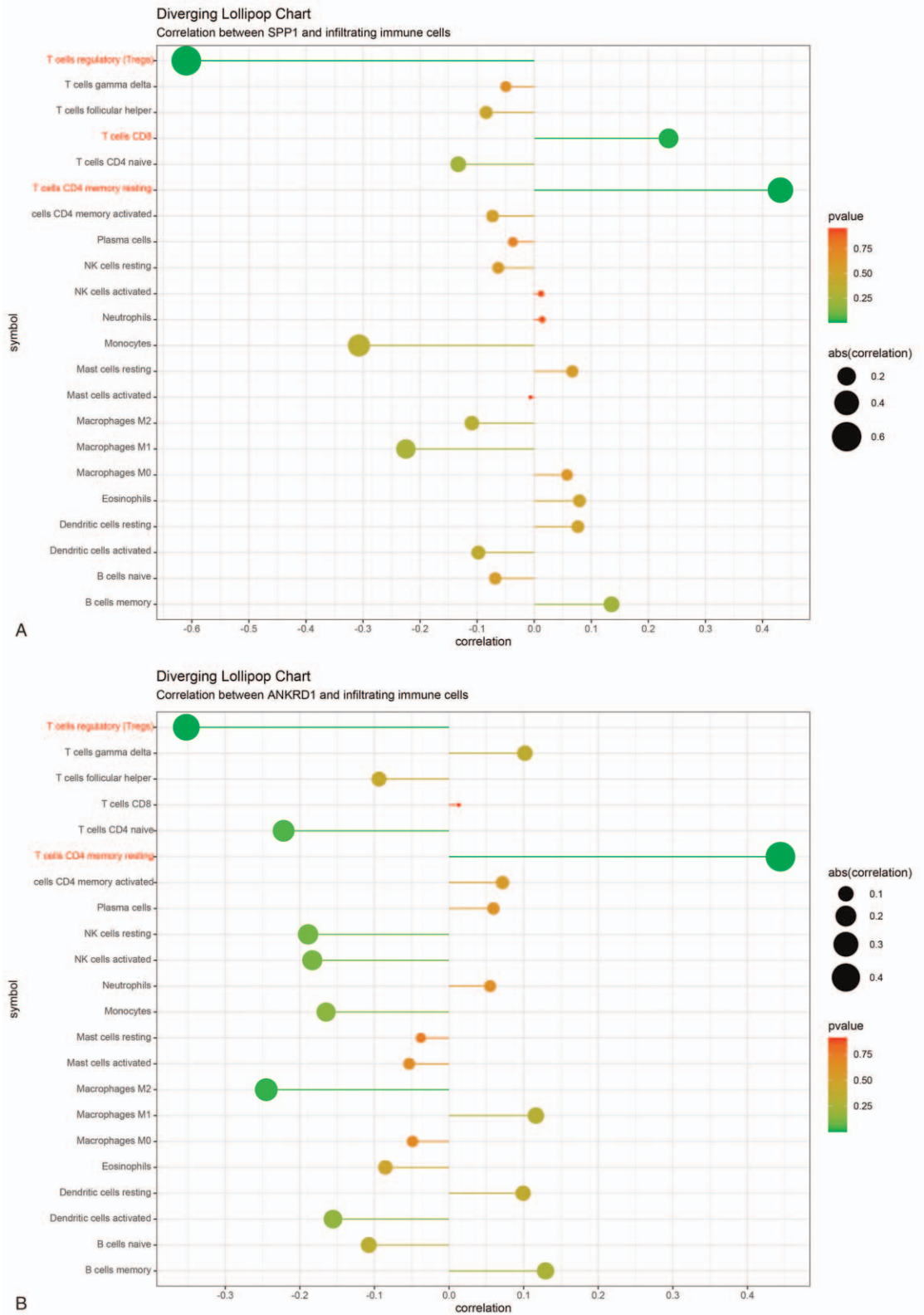


Figure 5. Correlation between secreted phosphoprotein 1 (SPP1), ankyrin repeat domain 1 (ANKRD1), and infiltrating immune cells. (A) Correlation between SPP1 and infiltrating immune cells. (B) Correlation between ANKRD1 and infiltrating immune cells. The size of the dots represents the strength of the correlation between genes and immune cells; the larger the dots.

that NF- κ B activation plays an important role in the pathogenesis of BA.^[32] The results of previous literature are consistent with our gene set enrichment analysis results, suggesting that the bioinformatics analysis results of our research are relatively accurate. The above results suggest that the immune response plays an important role in BA, so CIBERSORT algorithms were used to explore the relationship between BA and immune cell infiltration. We found that the infiltration rates of M2 macrophages cells, resting CD4 memory T cells and resting dendritic cells in BA were relatively high. M2 macrophages are thought to associate with increased fibrogenesis and a previous study demonstrated that M2 macrophages are significantly increased in BA livers.^[11] The dendritic cells secreted IL-6 and paralyze the Treg cells, which lead to the increase of Th17 cells and contributed to bile duct injury.^[33] However, there is no research conducted on the role of resting CD4 memory T cells in BA. In addition, in our study, regulatory T cells had a significant negative correlation with naïve CD4 T cells, while the correlation between regulatory T cells and resting CD4 memory T cells in BA was positive. The specific mechanisms underlying these associations require further exploration. Besides, the immune correlation analysis showed that SPP1 was negatively correlated with regulatory T cells. The ANKRD1 was negatively correlated with regulatory T cells. Studies have shown that regulatory T cells play important roles in BA through inhibiting Th1 cell-mediated bile duct injury in murine.^[34] Therefore, we speculate that SPP1 and ANKRD1 reduces regulatory T cells, participating in the occurrence and progress of BA. The functional consequences of these complex interactions between genes and immune cells require further research.

In conclusion, the present study identified potential diagnostic genes (SPP1 and ANKRD1) of BA. Furthermore, we screened out immune cells that may be closely related to the occurrence and progress of BA. In addition, SPP1 and ANKRD1 was negatively correlated with regulatory T cells. The SPP1 and ANKRD1 exhibited great promise and might serve as a noninvasive biomarker in discriminating BA and other forms of IC.

Author contributions

Conceptualization: Meng Kong, Bo Xiang.

Data curation: Meng Kong, Teng Ma, Bo Xiang.

Formal analysis: Meng Kong, Teng Ma.

Methodology: Meng Kong.

Supervision: Bo Xiang.

Validation: Meng Kong.

Visualization: Meng Kong, Teng Ma.

Writing – original draft: Meng Kong.

Writing – review & editing: Bo Xiang.

References

- Bezerra JA, Wells RG, Mack CL, et al. Biliary atresia: clinical and research challenges for the twenty-first century. *Hepatology* 2018;68:1163–73.
- Gunadi , Sirait DN, Budiarti LR, et al. Histopathological findings for prediction of liver cirrhosis and survival in biliary atresia patients after Kasai procedure. *Diagn Pathol* 2020;15:79.
- Shreef K, Alhelal A. Evaluation of the use of laparoscopic-guided cholecystocholangiography and liver biopsy in definitive diagnosis of neonatal cholestatic jaundice. *Afr J Paediatr Surg* 2016;13:181–4.
- Yoon HM, Suh CH, Kim JR, Lee JS, Jung AY, Cho YA. Diagnostic performance of sonographic features in patients with biliary atresia: a systematic review and meta-analysis. *J Ultrasound Med* 2017;36:2027–38.
- Kianifar HR, Tehranian S, Shojaei P, et al. Accuracy of hepatobiliary scintigraphy for differentiation of neonatal hepatitis from biliary atresia: systematic review and meta-analysis of the literature. *Pediatr Radiol* 2013;43:905–19.
- Wang NL, Lu YL, Zhang P, et al. A Specially designed multi-gene panel facilitates genetic diagnosis in children with intrahepatic cholestasis: simultaneous test of known large insertions/deletions. *PLoS One* 2016;11:e0164058.
- Matte U, Mourya R, Miethke A, et al. Analysis of gene mutations in children with cholestasis of undefined etiology. *J Pediatr Gastroenterol Nutr* 2010;51:488–93.
- Togawa T, Sugiura T, Ito K, et al. Molecular genetic dissection and neonatal/infantile intrahepatic cholestasis using targeted next-generation sequencing. *J Pediatr* 2016;171:171–7. e1–e4.
- Wang J, Xu Y, Chen Z, et al. Liver immune profiling reveals pathogenesis and therapeutics for biliary atresia. *Cell* 2020;183:1867–83. e1826.
- Saxena V, Shivakumar P, Sabla G, Mourya R, Chougnet C, Bezerra JA. Dendritic cells regulate natural killer cell activation and epithelial injury in experimental biliary atresia. *Sci Transl Med* 2011;3:102ra194.
- Yang Y, Dong R, Zheng C, Zheng S, Chen G. Infiltration of polarized macrophages associated with liver fibrosis in infants with biliary atresia. *J Pediatr Surg* 2017;52:1984–8.
- Than TT, Tran GV, Son K, et al. Ankyrin repeat domain 1 is up-regulated during hepatitis C virus infection and regulates hepatitis C virus entry. *Sci Rep* 2016;6:20819.
- Bessho K, Mourya R, Shivakumar P, et al. Gene expression signature for biliary atresia and a role for interleukin-8 in pathogenesis of experimental disease. *Hepatology* 2014;60:211–23.
- Moyer K, Kaimal V, Pacheco C, et al. Staging of biliary atresia at diagnosis by molecular profiling of the liver. *Genome Med* 2010;2:33.
- Reid S, Tibshirani R. Regularization paths for conditional logistic regression: the clogitL1 package. *J Stat Softw* 2014;58:12.
- Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* 2018;15:41–51.
- Zhou W, Koudijs KKM, Bohringer S. Influence of batch effect correction methods on drug induced differential gene expression profiles. *BMC Bioinform* 2019;20:437.
- Tom JA, Reeder J, Forrest WF, et al. Identifying and mitigating batch effects in whole genome sequencing data. *BMC Bioinform* 2017;18:351.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
- Song Z, Chen W, Athavale D, et al. Osteopontin takes center stage in chronic liver disease. *Hepatology* 2020;73:1594–608.
- Cantor H, Shinohara ML. Regulation of T-helper-cell lineage development by osteopontin: the inside story. *Nat Rev Immunol* 2009;9:137–41.
- Hertel PM, Crawford SE, Finegold MJ, Estes MK. Osteopontin upregulation in rotavirus-induced murine biliary atresia requires replicating virus but is not necessary for development of biliary atresia. *Virology* 2011;417:281–92.
- Whittington PF, Malladi P, Melin-Aldana H, Azzam R, Mack CL, Sahai A. Expression of osteopontin correlates with portal biliary proliferation and fibrosis in biliary atresia. *Pediatr Res* 2005;57:837–44.
- Lee SH, Seo GS, Park YN, Yoo TM, Sohn DH. Effects and regulation of osteopontin in rat hepatic stellate cells. *Biochem Pharmacol* 2004;68:2367–78.
- Mikhailov AT, Torrado M. The enigmatic role of the ankyrin repeat domain 1 gene in heart development and disease. *Int J Dev Biol* 2008;52:811–21.
- Papic N, Maxwell CI, Delker DA, Liu S, Heale BS, Hagedorn CH. RNA-sequencing analysis of 5' capped RNAs identifies many new differentially expressed genes in acute hepatitis C virus infection. *Viruses* 2012;4:581–612.
- Jelicic K, Cimbri R, Nawaz F, et al. The HIV-1 envelope protein gp120 impairs B cell proliferation by inducing TGF- β 1 production and FcRL4 expression. *Nat Immunol* 2013;14:1256–65.
- Kaczowski B, Rossing M, Andersen DK, et al. Integrative analyses reveal novel strategies in HPV11, -16 and -45 early infection. *Sci Rep* 2012;2:515.
- Mannaerts I, Leite SB, Verhulst S, et al. The Hippo pathway effector YAP controls mouse hepatic stellate cell activation. *J Hepatol* 2015;63:679–88.
- Shivakumar P, Mizuochi T, Mourya R, et al. Preferential TNF α signaling via TNFR2 regulates epithelial injury and duct obstruction in experimental biliary atresia. *JCI Insight* 2017;2:e88747.

- [31] Shinkai M, Shinkai T, Puri P, Stringer MD. Elevated expression of IL2 is associated with increased infiltration of CD8+ T cells in biliary atresia. *J Pediatr Surg* 2006;41:300–5.
- [32] Song Z, Dong R, Zhao R, Zheng S. Overexpression of LDOC1 in human biliary epithelial cells inhibits apoptosis through NF- κ B signaling. *J Pediatr Gastroenterol Nutr* 2013;57:713–7.
- [33] Liu YJ, Li K, Yang L, et al. Dendritic cells regulate Treg-Th17 axis in obstructive phase of bile duct injury in murine biliary atresia. *PLoS One* 2015;10:e0136214.
- [34] Tucker RM, Feldman AG, Fenner EK, Mack CL. Regulatory T cells inhibit Th1 cell-mediated bile duct injury in murine biliary atresia. *J Hepatol* 2013;59:790–6.