Research Article

The Bioinformatic Study Uncovers Probable Critical Genes Involved in the Pathophysiology of Biliary Atresia

Si Chen,¹ Songli Wu,² and Mingman Zhang ¹

¹Children’s Hospital of Chongqing Medical University, China
²University-Town of Chongqing Medical University, China

Correspondence should be addressed to Mingman Zhang; zhangmingman-a@163.com

Received 17 April 2022; Revised 1 June 2022; Accepted 3 June 2022; Published 21 June 2022

Academic Editor: Ahmed Faeq Hussein

Background. Biliary atresia (BA) is an uncommon illness that causes the bile ducts outside and within the liver to become clogged in babies. If left untreated, the cholestasis causes increasing conjugated hyperbilirubinemia, cirrhosis, and hepatic failure. BA has a complicated aetiology, and the mechanisms that drive its development are unknown. The objective of this study was to show the role of probable critical genes involved in the pathophysiology of biliary atresia. Methods. We utilised the public Gene Expression Omnibus (GEO) microarray expression profiling dataset GSE46960 to find differentially expressed genes (DEGs) in 64 biliary atresia newborns, 14 infants with various causes of intrahepatic cholestasis, and 7 deceased-donor children as control subjects in our study. The relevant information was looked into. The important modules were identified after functional enrichment, GO and KEGG pathway analyses, protein-protein interaction (PPI) network analyses, and GSEA analysis. Results. The differential expression analysis revealed a total of 22 elevated genes. To further understand the biological activities of the DEGs, we run functional enrichment analyses on them. Meanwhile, KEGG analysis has revealed significant enrichment of pathways involved in activating cross-talking with inflammation and fibrosis in BA. SERPINE1, THBS1, CCL2, MMP7, CXCL8, EPCAM, VCAN, ITGA2, AREG, and HAS2, which may play a significant regulatory role in the pathogenesis of BA, were identified by PPI studies. Conclusion. Our findings suggested 10 hub genes and probable mechanisms of BA in the current study through bioinformatic analysis.

1. Introduction

Biliary atresia (BA) is a rare disease in which the bile ducts outside and inside the liver become blocked in newborns. Increasing evidence showed that newborn screening with direct or conjugated bilirubin results in earlier diagnosis. The serum bilirubin level after Kasai portoenterostomy is still the most accurate clinical predictor of native liver survival. Cholestasis causes increasing conjugated hyperbilirubinemia, cirrhosis, and hepatic failure if not treated [1, 2]. It is the most likely cause for a liver transplant in a younger. BA has a tangled aetiology, with evidence pointing to viral, toxic, and genetic factors [3, 4]. The mechanisms that cause it are likewise unknown. We still do not know when BA starts or how to prevent the liver from deteriorating further [5]. A better knowledge of the aetiology of BA is required for novel therapy options other than liver transplantation to be developed. As a result, the goal of our research was to look at gene expression patterns in BA patients in order to look for potential biomarkers or pathological causes of the disease, as well as to find a better understanding and therapy for the condition.

2. Materials and Methods

2.1. Microarray Data. The gene expression profiling dataset GSE46960, which was deposited by Bessho et al. [6], was obtained using the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) [7]. The dataset was
Liver biopsy samples were obtained from 64 neonates with biliary atresia during an intraoperative cholangiogram, 14 age-matched babies with various kinds of intrahepatic cholestasis served as diseased controls, and 7 deceased-donor children served as normal controls. The age and sex of the participants, as well as their preoperative biochemical test data, were inaccessible because it was a public dataset, which looks to be a possible drawback.

GPL6244’s annotation file was also obtained from the GEO.

### 2.2. Differential Expression Analysis

Using the online analytic tool GEO2R, the expression profiles of BA, non-BA patients, and healthy controls were compared to find DEGs. P values and corrected P values were calculated using T-tests. The platform’s gene probes were translated into gene names by referencing the GPL6244 platform. The genes in each sample were preserved if they matched two criteria: (1) a \(|\log 2 (\text{fold change})| > 1\) and (2) an adjusted P \(\leq 0.05\). We identified the most important genes when the DEGs were repeated. The DEGs were found by the intersection of the two datasets, which were conducted independently for the BA versus NC and BA vs. non-BA groups.

The online tool E Venn [8] (http://www.ehbio.com/test/venn/#/) was used to construct a Venn diagram of DEGs, and the heat map for the DEGs was made using the online tool xiantao Xue shu (https://www.xiantao.love/).

### 2.3. Functional Enrichment Analysis of DEGs

To improve the identification of the biological activities of DEGs, we used the web tool DAVID (https://david.ncifcrf.gov/) to conduct Gene Ontology (GO) terms [9] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [10]. Based on the GO analysis description, the gene function annotations were categorised as biological processes (BP), cellular components (CC), or molecular activities (MF). Statistical significance was defined as adjusted P values of less than 0.05. ClueGo [11, 12], a Cytoscape application (Cytoscape v3.8.0) plug-in, was utilised to demonstrate the relationship between gene enrichment analysis terms.
2.4. PPI Network Creation and Hub Gene Identification. A PPI network of DEGs was created using the Search Tool for the Retrieval of Interacting Genes (STRING11.5; https://string-db.org/). [13], with an interaction score cut-off of >0.4. The hub genes were found using Cytohubba [14], a Cytoscape software plug-in, and the key modules in the PPI network were found using molecular complex identification (MCODE 1.5.1) [15], another Cytoscape software plug-in. The DEG clustering and scoring parameters were MCODE score = 4, degree cut-off = 2, node score cut-off = 0.2, max depth = 100, and k-score = 2.

2.5. GSEA Gene Set Enrichment Analysis. GSEA is a computer programme that determines if a set of genes that have been defined a priori demonstrate a statistically meaningful, congruent gap between different physiological situations (e.g., phenotypes). GSE46960 was submitted to Gene Set Enrichment Analysis with permutation = 1,000 using the GSEA tool (https://www.broadinstitute.org/gsea/) [16, 17]. A hypothetical P value was used to assess the statistically significant results of the enrichment score.

2.6. Statistical Analysis. Continuous normally distributed data are expressed as the means ± SDs. All statistical calculations were calculated through SPSS statistical software. P values < 0.05 were considered significant.

3. Results

3.1. Identification of DEGs in Biliary Atresia. The gene expression profile of the GSE46960 dataset comprised data from three separate groups (Table 1). Using a fold change (FC) value of $|\log 2FC| > 2$ and a P value of 0.05 as the cut-off, a total of 22 DEGs, all upregulated genes, were obtained from the notably regulated gene in the biliary atresia samples over both diseased control (non-BA) and normal
### Table 3: Top 8 of the most significantly enriched GO terms.

| Pathway ID | Terms                                         | Gene count | $P$ value  |
|------------|-----------------------------------------------|------------|------------|
| BP         |                                               |            |            |
| GO:0030198 | Extracellular matrix organization              | 7          | 5.60E-08   |
| GO:0071356 | Cellular response to tumor necrosis factor     | 6          | 1.19E-07   |
| GO:0071347 | Cellular response to interleukin-1             | 5          | 1.08E-06   |
| GO:0071222 | Cellular response to lipopolysaccharide        | 5          | 6.97E-06   |
| GO:007155  | Cell adhesion                                  | 6          | 1.26E-04   |
| GO:006954  | Inflammatory response                          | 5          | 7.57E-04   |
| GO:008284  | Positive regulation of cell proliferation      | 5          | 1.63E-03   |
| GO:006955  | Immune response                                | 4          | 1.12E-02   |
| CC         |                                               |            |            |
| GO:000615  | Extracellular space                            | 11         | 7.80E-07   |
| GO:005576  | Extracellular region                           | 10         | 3.53E-05   |
| GO:009986  | Cell surface                                   | 6          | 3.13E-04   |
| GO:0031012 | Extracellular matrix                           | 5          | 3.28E-04   |
| GO:009897  | External side of plasma membrane               | 3          | 2.47E-02   |
| GO:007062  | Extracellular exosome                          | 8          | 3.29E-02   |
| GO:004871  | Perinuclear region of cytoplasm                | 4          | 3.32E-02   |
| GO:005578  | Proteinaceous extracellular matrix             | 3          | 3.77E-02   |
| MF         |                                               |            |            |
| GO:005539  | Glycosaminoglycan binding                      | 5          | 9.29E-05   |
| GO:005125  | Cytokine activity                              | 5          | 1.20E-04   |
| GO:005102  | Receptor binding                               | 9          | 1.51E-04   |
| GO:008201  | Heparin binding                                | 4          | 8.46E-04   |
| GO:008009  | Chemokine activity                             | 3          | 1.52E-03   |
| GO:1901681 | Sulfur compound binding                        | 4          | 2.49E-03   |
| GO:0043236 | Laminin-binding                                | 2          | 2.92E-02   |
| GO:0050840 | Extracellular matrix binding                   | 2          | 3.04E-02   |

Notes. GO: Gene Ontology; BP: biological process; CC: cellular component; MF: molecular function.
control (NC) samples (Table 2). For the distribution of DEGs, an online tool was utilised to construct a Venn diagram and heat maps (Figures 1(a) and 1(b)).

3.2. GO and KEGG Pathway Analysis for Identifying the DEGs. DEGs were studied using the DAVID online tool for functional and pathway enrichment. The most three important processes revealed by GO analysis among the annotations of BP were extracellular matrix organization, cellular response to tumor necrosis factor, and cell adhesion. The three most important processes revealed among the CC annotations were extracellular space, extracellular area, and extracellular exosome. Finally, the three most significant processes among the MF annotations were receptor binding, glycosaminoglycan binding, and cytokine activity. Table 3 and Figure 2 show the number of genes and $P$ values of the top 8 enriched functional words based on the criteria.

The DEGs’ cell signaling pathway enrichment study yielded a total of eight relevant pathways that were investigated. ECM-receptor interaction, malaria, the PI3K-Akt signaling pathway, and others were among the cellular signaling pathways linked to biliary atresia. Table 4 and Figure 3 describe the specific enriched pathways discovered by DEG analysis. Figure 4 shows the relationship between the words of gene enrichment analysis.

3.3. Construction of the PPI Network and Identification of Hub Genes. STRING (https://string-db.org/) is a public

![Figure 2: Results of GO enrichment. The ordinate shows the number and ratio of differentially expressed genes, whereas the abscissa reflects the enriched GO. Biological process, cellular component, and molecular function are all represented by distinct colours. Abbreviation: GO: gene ontology.](image)

| Pathway ID | Term                                      | Gene count | $P$ value  | Genes                      |
|------------|-------------------------------------------|------------|------------|----------------------------|
| hsa04512   | ECM-receptor interaction                   | 4          | 5.10E-04   | THBS1, ITGA2, LAMC2, SPP1  |
| hsa05144   | Malaria                                   | 3          | 3.69E-03   | CCL2, THBS1, CXCL8         |
| hsa04510   | Focal adhesion                             | 4          | 6.06E-03   | THBS1, ITGA2, LAMC2, SPP1  |
| hsa05323   | Rheumatoid arthritis                       | 3          | 1.15E-02   | CCL20, CCL2, CXCL8         |
| hsa05142   | Chagas disease (American trypanosomiasis) | 3          | 1.58E-02   | CCL2, CXCL8, SERPINE1      |
| hsa04151   | PI3K-Akt signaling pathway                 | 4          | 2.46E-02   | THBS1, ITGA2, LAMC2, SPP1  |
| hsa04062   | Chemokine signaling pathway                | 3          | 4.66E-02   | CCL20, CCL2, CXCL8         |

Notes. KEGG: Kyoto Encyclopedia of Genes and Genomes.
A database that contains known and predicted protein interactions. PPI [18] is important for studying protein function since it can help elucidate the role of protein control. STRING’s official website was used to submit the 22 DEGs from the GSE46960 dataset in order to get protein interrelationships. The minimum required interaction score was set at 0.15 in order to see the interaction networks with Cytoscape (version v3.9.0) [19]. There were 22 nodes and 107 edges in the PPI network that resulted. The network visualisation created using STRING’s official website is shown in Figure 5(a). The degree of linkage between DEGs and genes was used to screen for hub genes, and the DEGs with the ten highest degrees were identified as hub genes (Table 5 and Figure 5(b)).

3.4. GSEA Analysis of All Detected Genes. GSEA was used to find gene sets with a statistically significant difference between BA and NC participants, and it revealed that the...
BA subjects had the highest enriched gene sets of all discovered genes. There are 4662 gene symbols in the dataset, 3651 of which are elevated in phenotype BA. ECM receptor interaction, integrin cell-surface interactions, and Andersen cholangiocarcinoma class1 were the top six most significant-enriched gene sets positively correlated with the BA subjects, followed by uterine fibroid up, ECM proteoglycans, and nonintegrin membrane ECM interactions (Figures 6(a)–6(f)).

4. Discussion
Biliary atresia (BA) is a fibroinflammatory disease of the intra- and extrahepatic biliary tree. In order to have a better
Table 5: Top 10 hub genes ranked by Cytobubba.

| Gene symbol | Score | Full name | Distribution |
|-------------|-------|-----------|--------------|
| SERPINE1    | 32    | Serpin family E member 1 | Serine protease inhibitor; plasminogen activator inhibitor 1. Tissue plasminogen activator, urokinase, protein C, and matriptase-3/TMPRSS7 are all attracted to this inhibitor. Its quick interaction with PLAT might be a crucial control point in fibrinolysis regulation. Cell-to-cell and cell-to-matrix adhesion glycoprotein thrombospondin-1 modulates cell-to-cell and cell-to-matrix connections. This chemical binds to heparin. Dentin and dental pulp dentinogenesis and/or maintenance may be affected (by similarity). A CD36 ligand is responsible for the antiangiogenic actions. It has a function in the ER stress response by interacting with the activating transcription factor 6 alpha (ATF6), which produces adaptive ER stress response factors (by similarity). |
| THBS1       | 32    | Thrombospondin 1 | Chemotactic factor with a C-C motif, which attracts monocytes and basophils but not neutrophils or eosinophils. Increases the antitumor activity of monocytes. Has been linked to the development of disorders with monocyte infiltrates, such as psoriasis, rheumatoid arthritis, and atherosclerosis. During the disease process of atherosclerosis, it may be implicated in the recruitment of monocytes into the artery wall. |
| CCL2        | 30    | C-C motif chemokine ligand 2 | Chemotactic factor with a C-C motif, which attracts monocytes and basophils but not neutrophils or eosinophils. Increases the antitumor activity of monocytes. Has been linked to the development of disorders with monocyte infiltrates, such as psoriasis, rheumatoid arthritis, and atherosclerosis. During the disease process of atherosclerosis, it may be implicated in the recruitment of monocytes into the artery wall. |
| MMP7        | 30    | Matrix metallopeptidase 7 | Matrilysin degrades casein, gelatins of types I, III, IV, and V, and fibronectin. Activates procollagenase; M10 matrix metallopeptidases |
| CXCL8       | 30    | C-X-C motif chemokine ligand 8 | Chemotactic factor with a C-C motif, which attracts monocytes and basophils but not neutrophils or eosinophils. Increases the antitumor activity of monocytes. Has been linked to the development of disorders with monocyte infiltrates, such as psoriasis, rheumatoid arthritis, and atherosclerosis. During the disease process of atherosclerosis, it may be implicated in the recruitment of monocytes into the artery wall. |
| EPCAM       | 28    | Epithelial cell adhesion molecule | Intercellular signaling and cell-extracellular matrix interactions may be mediated by the Versican core protein. It might play a part in cell motility, proliferation, and differentiation. Has a C-type lectin domain that binds hyaluronic acid. Integrin alpha-2 (integrin alpha-2/beta-1) is a receptor for laminin, collagen, collagen C-propeptides, fibronectin, and E-cadherin. It recognises the proline-hydroxylated sequence G-F-P-G-E-R in collagen. Platelet and other cell adhesion to collagens, collagen and collagenase gene expression control, force creation, and structuring of newly formed extracellular matrix are all mediated by CD molecules. |
| VCAN        | 24    | Versican | Hyaluronan synthase 2 adds GlcNAc or GlcUA monosaccharides to the nascent hyaluronan polymer. As a result, it is essential for the formation of hyaluronan, a major component of most extracellular matrices that regulates cell adhesion, migration, and differentiation and plays a structural role in tissue architecture. This is one of the isoforms that catalyses the process and is responsible for the production of high-molecular-mass hyaluronan. A key phase in the creation of the heart is the conversion of endocardial cushion cells to mesenchymal cells. |
| ITGA2       | 24    | Integrin subunit alpha 2 | Amphiregulin is a ligand for the EGF receptor/EGFR. Amphiregulin is an amitogen and autocrine growth factor for a range of target cells, including astrocytes, Schwann cells, and fibroblasts. |
| AREG        | 24    | Amphiregulin | Intercellular signaling and cell-extracellular matrix interactions may be mediated by the Versican core protein. It might play a part in cell motility, proliferation, and differentiation. Has a C-type lectin domain that binds hyaluronic acid. Integrin alpha-2 (integrin alpha-2/beta-1) is a receptor for laminin, collagen, collagen C-propeptides, fibronectin, and E-cadherin. It recognises the proline-hydroxylated sequence G-F-P-G-E-R in collagen. Platelet and other cell adhesion to collagens, collagen and collagenase gene expression control, force creation, and structuring of newly formed extracellular matrix are all mediated by CD molecules. |
| HAS2        | 22    | Hyaluronan synthase 2 | Intercellular signaling and cell-extracellular matrix interactions may be mediated by the Versican core protein. It might play a part in cell motility, proliferation, and differentiation. Has a C-type lectin domain that binds hyaluronic acid. Integrin alpha-2 (integrin alpha-2/beta-1) is a receptor for laminin, collagen, collagen C-propeptides, fibronectin, and E-cadherin. It recognises the proline-hydroxylated sequence G-F-P-G-E-R in collagen. Platelet and other cell adhesion to collagens, collagen and collagenase gene expression control, force creation, and structuring of newly formed extracellular matrix are all mediated by CD molecules. |
Figure 6: Continued.
understanding of the underlying cause(s) and pathogenesis of the disease, the National Institutes of Diabetes and Digestive and Kidney Diseases sponsored researches that study the promising and innovative approaches. In this investigation, we used the GEO database to screen for DEGs and acquire gene expression profiles from patients with BA, non-BA, and normal controls. There were a total of 22 DEGs confirmed.

The DEGs were considerably enriched in the cellular response to interleukin-1, according to BP in GO annotation, which was consistent with earlier evidence that inhibiting IL-1-mediated inflammation may be advantageous in selective liver fibrotic disease [20]. Other enhanced gene sets of DEGs in the BP of GO, such as immunological and inflammatory responses, have been linked to biliary atresia [21, 22]. The extracellular exosome was shown to be rich in CC. Exosomes have been explored as disease biomarkers [23, 24] or cell-cell communication factors because of their role in carrying a variety of proteins, noncoding RNA, and coding RNA from different cells. A new study suggests that serum exosomal H19 might be exploited as a noninvasive diagnostic biomarker and treatment target for BA [25]. According to KEGG enrichment analysis, DEGs are also detected in the ECM-receptor interaction, focal adhesion, PI3K-Akt signaling pathway, and chemokine signaling pathway. All of these results corroborated previous findings that BA interacts with inflammation and fibrosis [26].

SERPINE1, THBS1, CCL2, MMP7, CXCL8, EPCAM, VCAN, ITGA2, AREG, and HAS2 were among the 10 hub genes discovered in this study. Interleukin- (IL-) 8 (CXCL8) may mediate liver damage in BA by enhancing ductular response and related hepatic fibrogenesis, according to Godbole et al. [27]. The serum MMP-7 test, according to Yang et al., shows excellent sensitivity and specificity for distinguishing BA from other newborn cholestasis and may be a valid biomarker for BA [28]. SERPINE1 can be targeted to prevent biliary fibrosis, according to Aseem et al.

The most significant-enriched gene set connected with the BA individuals, according to GSEA, was ECM receptor interaction. Many studies have linked oxidative stress to liver fibrosis. It has been discovered that ROS can activate KCs (Kupffer cells) to trigger the inflammatory response, which subsequently leads to HSC (activated hepatic stellate cells) activation to create ECM proteins [29, 30] and fibrosis. It will offer a fresh look at the treatment strategy for BA’s fibrosis mechanism. The limit of this study is that there are only bioinformatic analysis and did not have cell and animal experiments. Therefore, many investigations need to be added to the article.

5. Conclusion

With bioinformatic analysis, we found 10 hub genes and probable mechanisms of BA in the current study. More research is needed to confirm the hub genes and identify relevant processes. All of the findings will pave the way for a possible treatment strategy for biliary atresia and associated fibrotic illnesses.

Data Availability

The data could be obtained from contacting the corresponding author.
Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. H. Hsiao, M. H. Chang, H. L. Chen et al., “Universal screening for biliary atresia using an infant stool color card in Taiwan,” Hepatology, vol. 47, no. 4, pp. 1233–1240, 2008.

[2] B. E. Wildhaber, P. Majno, J. Mayr et al., “Biliary atresia: Swiss national study, 1994-2004,” Journal of Pediatric Gastroenterology and Nutrition, vol. 46, no. 3, pp. 299–307, 2008.

[3] J. L. Hartley, M. Davenport, and D. A. Kelly, “Biliary atresia,” Lancet, vol. 374, no. 9702, pp. 1704–1713, 2009.

[4] K. Nakamura and A. Tanoue, “Etiology of biliary atresia as a developmental anomaly: recent advances,” Journal of Hepato-Biliary-Pancreatic Sciences, vol. 20, no. 5, pp. 459–464, 2013.

[5] A. Kilgore and C. L. Mack, “Update on investigations pertaining to the pathogenesis of biliary atresia,” Pediatric Surgery International, vol. 33, no. 12, pp. 1233–1241, 2017.

[6] B. Donnelly, H. Temple, G. Tiao, R. Bansal, and S. K. Mohanty, “Innate immunity and pathogenesis of biliary atresia,” Frontiers in Immunology, vol. 11, 2020.

[7] T. Chen, H. Zhang, Y. Liu, Y. X. Liu, and L. Huang, “EvVenn: easy to create repeatable and editable Venn diagrams and Venn networks online,” Journal of genomics and genomics = Yi chuan xue bao, vol. 48, no. 9, pp. 863–866, 2021.

[8] M. Ashburner, C. A. Ball, J. A. Blake et al., “Gene ontology: tool for the unification of biology,” Nature genetics, vol. 25, no. 1, pp. 25–29, 2000.

[9] T. Chen, H. Zhang, Y. Liu, Y. X. Liu, and L. Huang, “EvVenn: easy to create repeatable and editable Venn diagrams and Venn networks online,” Journal of genomics and genomics = Yi chuan xue bao, vol. 48, no. 9, pp. 863–866, 2021.

[10] M. Kanehsa and S. Goto, “KEGG: Kyoto encyclopedia of genes and genomes,” Nucleic Acids Research, vol. 28, no. 1, pp. 27–30, 2000.

[11] G. Bindea, B. Mlecnik, H. Hackl et al., “ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks,” Bioinformatics, vol. 25, no. 8, pp. 1091–1093, 2009.

[12] G. Bindea, J. Galon, and B. Mlecnik, “CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data,” Bioinformatics, vol. 29, no. 5, pp. 661–663, 2013.

[13] D. Szklarczyk, A. L. Gable, D. Lyon et al., “STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” Nucleic Acids Research, vol. 47, no. D1, pp. D607–D613, 2019.

[14] C. H. Chin, S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko, and C. Y. Lin, “cytoHubba: identifying hub objects and subnetworks from complex interactome,” BMC Systems Biology, vol. 8, no. S4, p. S11, 2014.

[15] W. P. Bandettini, P. Kellman, C. Mancini et al., “Multicontrast delayed enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study,” Journal of Cardiovascular Magnetic Reso-