A novel pathogenesis concept of biliary atresia approached by combined molecular strategies

Wison Laochareonsuk, Komwit Surachat, Piyawan Chiengkriwate, Surasak Sangkhathat

1 Faculty of Medicine, Department of Surgery, Prince of Songkla University, Hat Yai, Songkhla, Thailand,
2 Faculty of Medicine, Translational Medicine Research Center, Prince of Songkla University, Hat Yai, Songkhla, Thailand,
3 Faculty of Medicine, Department of Biomedical Sciences and Biomedical Engineering, Prince of Songkla University, Hat Yai, Songkhla, Thailand

*surasak.sa@psu.ac.th

Abstract

Cholestatic jaundice is one of the most common neonatal conditions. BA, a correctable cholangiopathy, presents with cholestatic jaundice within the first weeks of life. The inflammation of bile ducts leads to progressive fibrosclerosis involving biliary trees, followed by cirrhosis and liver failure. With the use of modern molecular studies, this research aimed to define a novel pathogenesis by exploring variations. We performed genetic discovery by using supervised and unsupervised approaches. Ultimately, a combination of genetic variations and survival data was analyzed to strengthen the novel concept in this study. In this study, coding regions were explored to identify rare deleterious variants within genes from the first analysis together with gene sets reported in PFIC, and diseases with hyperbilirubinemia. Our unsupervised prioritization was primarily designed to identify novel causal genes from nonsynonymo us variants derived by three biostatistical algorithms: enrichment analysis, burden test, and trio study. Survival analysis was integrativel y evaluated with a combination of identified causal genes. The individuals with identified variants from the supervised approach were frequently related to the severity of cirrhosis and poor postoperative outcome. In the unsupervised approach, nonsynonymous variants were enriched. Cilium and muscle related pathways had a significant correlation. CCDC8 was statistically significant gene in which six cases carried mutations identified through burden analysis. Individuals who carried variants in corresponding genes and significant pathways had significantly lower native-liver survival than individuals in whom none of these variants were identified (log-rank p value 0.016). This study explored genetic variations by multiple strategies. Different pathways of cholestatic diseases have been found to be associated with BA. Therefore, BA may be characterized as a shared sequela of many cholestatic disorders. Susceptibility in those pathways suggested an association with BA and strengthened this proposed novel hypothesis. The results emphasized the consequences of many disruptive pathophysologies.
**Introduction**

Biliary atresia (BA) is an inflammatory cholangiopathy manifesting as gradual obstructive jaundice within the first two weeks of life. The primary pathology, severe periductal inflammation, leads to progressive fibrosclerosis involving both extrahepatic and intrahepatic bile ducts, followed by biliary cirrhosis, portal hypertension and fulminant liver failure [1]. Surgical bypass of the extrahepatic biliary system, hepatic portoenterostomy (HPE), is the standard treatment in patients diagnosed with BA. Without appropriate treatment, infants with BA are unlikely to survive with their native liver for longer than two years [2,3]. The highest incidence of BA is reported in Asia-Pacific regions, ranging from 1:320–1:2000 live births. It remains unclear if the pathogenesis of inflammation in BA is secondary to viral infection, an immunological process or a developmental mishap [1]. A recent genetic study by Sangkhathat et al. [4] proposed that BA might be a common pathological process of various entities, as genetic variants reported in other infantile cholestatic conditions were also detected in clinically diagnosed BA. As the molecular pathology regarding the pathophysiology of cholestatic jaundice has been more elucidated, novel diseases have been characterized, and some of those diseases may share similar clinical presentations and histopathology with BA and BA with germline variants may have different prognosis from sporadic one.

In recent decades, the spotlight on the pathogenetic theories of BA has moved to genetic susceptibility following the introduction of high-throughput genetic studies, led by genome-wide association studies (GWASs). Single-nucleotide polymorphisms (SNPs) within the locus 10q25, near the region of Adducin3 (ADD3), were identified as having a significant association with the occurrence of BA in various independent populations, including Han Chinese, Thai, and European populations [5–8]. Furthermore, the GWAS in Europeans localized a second SNP on locus 2p16, an intronic region of EGF, situating the Fibulin Extracellular Matrix Protein 1 (EFEMP1) gene as a novel candidate gene [9]. Next-generation sequencing (NGS), an evolutionary high-throughput platform, enables faster and more robust variant discovery than conventional sequencing technologies [10]. Recently, whole-exome sequencing (WES) has identified novel deleterious variants in various genes that play a role in ciliary functions (KIF3B, PCNT, and TTC17) in BA patients.

This study aimed to investigate a novel perspective concept in the pathogenesis of BA by exploring genetic variations and biostatistical appraisal. The molecular genetic analyses were performed using the WES platform, followed by bioinformatic analysis. Our genetic discovery algorithm consisted of supervised variant prioritization looking for previously reported genes in infantile cholestatic jaundice diseases/syndromes and an unsupervised approach which aimed to identify novel variants, followed by biostatistical analysis of all identified variants. To gain more insight into the phenotype correlation, genetic information was categorized and collectively analyzed with the clinical manifestations, the outcomes of surgical correction, and associated syndromic features.

**Materials and methods**

**Patients and biomaterials**

Infants with BA who underwent HPE at our institute from 2003 to 2020 were recruited into the study through informed consent from their parents. The study was approved by the Human Research Ethics Committee of the Faculty of Medicine, Prince of Songkla University (REC-61-351-10-1). The diagnosis of BA was made through intraoperative cholangiography and intraoperative findings together with additional conformation by pathological report and clinical follow-up data. Other syndromes which clinical features similar to biliary atresia e.g.,
Alagille syndrome, were excluded when there was strong diagnostic evidence such as histopathology that revealed biliary hypoplasia or paucity of bile duct. During the same operations, liver biopsy specimens were retrieved and cryo-preserved at the Human Biobank (Translational Medicine Research Center, Faculty of Medicine, Prince of Songkla University). Peripheral blood samples of the parents were obtained during the follow-up periods. Genome data from unrelated controls from the same geographic region aged between 1 and 60 years were provided by the South Thai Genome Data Bank.

Genomic DNA was extracted from the liver biopsy specimens or peripheral blood samples using a QIAGEN DNA Mini kit (QIAGEN, Hilden, Germany) and a High Pure PCR template preparation kit (Roche, Berlin, Germany). The extracted DNA was quantified for concentration and purity using a nanodrop (Thermo Scientific, Delaware, United States). The qualified DNA was subjected to archived fluorescent electrophoresis to evaluate DNA fragmentation using the Agilent Genomic DNA ScreenTape assay (Agilent Technologies, Santa Clara, California, United States).

**Clinical follow-up and data retrieval**

Postoperative care after HPE consisted of systemic steroids for 2 weeks, ursodeoxycholic acid for one month, and oral antibiotics for one month unless there was an indication for extended prophylaxis. Clinical follow-up and liver function studies were scheduled every six months until five years and then annually thereafter. In addition, active surveillance of liver cirrhosis was performed once a year with ultrasonography and esophagoscopy in cases with cirrhosis and clinical signs of portal hypertension. Patients with remarkable liver deterioration were put on a waiting list for the liver transplantation program at Siriraj Hospital, Bangkok.

Clinical data were retrieved from the electronic medical records, including date of birth, date of surgery, anatomical subtype based on Kasai classification, liver function test at diagnosis (total serum bilirubin, direct bilirubin, AST, ALT, ALP, GGT and serum albumin), liver function test 1 month after surgery, operative procedure and findings, adjunctive treatment after surgery (oral antibiotics, ursodeoxycholic acid, oral-systemic steroids), last follow-up date and status. The definition of ‘jaundice-clear’ was decreased serum bilirubin to less than 2 mg/dL within 1–3 month after surgery, with ‘jaundice-persistent’ defined in those who had a persistently high level of serum bilirubin. Within the jaundice-persistent group, there were 2 subcategories: (1) jaundice improved (total bilirubin decreases more than 80% of total) and (2) jaundice not improved (total bilirubin decreases less than 20% of total).

**Whole-exome sequencing**

Exome sequencing libraries were prepared, and coding regions were captured and enriched using Agilent SureSelect XT Human All Exon v6 following the manufacturer’s protocol (Agilent Technologies, Santa Clara, California, United States). The library was quantified with a Qubit dsDNA High Sense Assay Kit (Invitrogen, USA). The library size was measured by an Agilent D1000 ScreenTape assay. Paired-end sequencing with a 150-bp platform was carried out on an Illumina NovaSeq-6000 (Illumina, San Diego, California, United States) at an average targeted coverage of 200x depth.

**Bioinformatic analysis**

The paired-end sequence files were qualified by FastQC (version 0.11.9) and trimmed using Trimmomatic (version 0.39) with a paired-end trimming function. The optimally prepared FASTQ files were aligned with the human reference genome (version GRCh38.13) using BWA (version 0.7.17) [11]. Subsequently, the SAM (Sequence Alignment Map) was converted to a
BAM (Binary Alignment Map) and sorted by SAMtools (version 1.10) [12]. Then, the sorted BAM was regrouped and had identical sequences marked using Picard (version 2.18.26). The unduplicated BAM was adjusted for base quality score by GATK (version 4.2.0) [13] BQSR. The variants were subtracted from the exome using GATK HaplotypeCaller and stored as intermediate genotype information in GVCF (variant calling format genome VCF). When variant calling was finished for every sample, the GVCFs of those cases were merged with the disease control data using GATK CombineGCVFs and calculated for true possible variants using GATK GenotypeGVCFs. The combined variants were filtered out using a Gaussian mixture model with 99% truth sensitivity to select true variants. Finally, those variants were appended to functional annotation with Variant Effect Predictor (VEP) version 102 [14] based on dbSNP 151.

**Variant selection**

All identified variants were crudely filtered using the criteria of having a reading depth more than 20x and allele frequency of individual sample more than 25%. A variant was considered to have a high or moderate impact when the minor allele frequencies were less than 0.01 in the East Asian population. A variant was included in downstream analysis when its clinical correlation was available in the Human Gene Mutation Database (University of Cardiff, United States), ClinVar [15] and OMIM (Online Mendelian Inheritance in Man) [16]. If there was no report available, the effect on protein was evaluated by Sorting Tolerant From Intolerant (SIFT) [17], Polymorphism Phenotyping data collection 2 (PolyPhen2) [18], Mutation Taster – AA change score and probability value (Mutation Taster2) [19], and Combined Annotation Dependent Depletion (CADD) [20]. The criteria for defining pathogenic consequences of a variant were prediction as damaging (SIFT), possibly deleterious or deleterious (PolyPhen2), disease-causing (Mutation Taster), and CADD Phred score over 20.

**Supervised variant prioritization**

With a hypothesis that variants associated with congenital cholestatic diseases might be overlapping with variants found in those with BA phenotypes. Supervised annotation and prioritization were performed based on known candidate genes involved in cholestasis jaundice conditions: 1. progressive familial intrahepatic cholestasis (PFIC), 2. syndromes in which BA is reported as one of the phenotypes, and 3. syndromes with hyperbilirubinemia as one of the phenotypes (shown in S1–S3 Tables). The positions and nucleotide alterations of the variants were confirmed by demonstrating their differential alignment in Integrative Genomic Viewer (IGV). All confirmed variants were revalidated using PCR and Sanger sequencing.

**Unsupervised variant prioritization**

Selected variants from the cases and controls were submitted for three downstream variant prioritization algorithms: case-only, case–control, and trio analysis. In the case-only study, the variants were manipulated from VCF to tab delimited format using BCFtools (version 1.10.2). The genes were enriched by the EnrichR package (version 3.0) following eight databases: Gene Ontology (GO) molecular function, GO cellular component, GO biological process, Kyoto Encyclopedia of Genes and Genomes (KEGG), Human Reactome, Panther, Human WikiPathway, and BioCarta. A pathway was considered as significant when the false discovery rate (FDR) was below the threshold of 0.05 as calculated by Benjamini and Hochberg [21].

For the case–control analysis, genotyped VCF files from both groups were converted to the PED and MAP formats using PLINK (version 1.90) [22]. The variants were qualified and filtered out high missing variants following the criterion of call rate more than 95% for each
sample using PLINK. Hardy Weinberg Equilibriums (HWE) were calculated, and the variants were considered as passed when the p value was less than 0.05. Identical-by-descent (IBD) relatedness was assessed for the relationships between the samples in both groups in which the estimated pi-hat score should be lower than 0.9. The remaining variants were evaluated for homogeneity using principal component analysis (PCA) from the IPCAPS package (version 1.1.8) [23]. Nonhomogenous samples were excluded from the study. The qualified variants were selected from the raw merged VCF depending on the variant selection criteria and reconverted to PED and MAP, which were approved to be compatible based on rvtests (rare variant test, version 20190205) software. The variants were grouped into gene sets based on refFlat gene definitions (version GRCh38.13) and burden analysis of the different conditions. The gene-set was performed using the collapsing and combine (CMC) model with the exact test. A gene was considered as significant when the FDR was below the threshold of 0.05.

Trio analysis was performed by selecting variants belonging to a mode of inheritance by fitting either autosomal dominant or recessive models. The dominant model was designed to discover de novo mutations, which are only found in probands, while the recessive model assessed variants inherited from the parents. All variants identified from the three analysis methods demonstrated differential alignments in IGV and were revalidated using PCR and Sanger sequencing.

Statistical analysis
The statistical analysis of continuous clinical data was executed by a T-test for normally distributive data and a Mann-Whitney U test for non-normally distributive data. On the other hand, categorical data are analyzed by the chi-square (X²) test. The statistical analysis is achieved on the R base package and a p-value <0.05 is taken to indicate statistical significance. The survival analysis was performed by using duration after surgery to last follow-up with native liver together with combination of clinical and genetic factors. In addition, survival analysis used a log-rank test and displayed in Kaplan-Meier curves. Cox proportional hazards model was applied to identify the correlation between survival rate and factors using survival package (version 3.2–13) and survminer package (version 0.4.9). Statistical significance was considered when P value <0.05. Boxplot and bar graph were created by using ggplot2 package (version 3.3.5) and Kaplan-Meier graph was initiated by using sur package (version 1.0.4).

Results
Clinical characteristics
The clinical manifestations of the 75 BA patients in the study are shown in Table 1 and Fig 1. There were 39 males (52%) and 36 females (48%). The mean age at surgery was 83 days (SD 32 days). The average total and direct bilirubin levels before surgery were 11.89 mg/dL (SD 3.84 mg/dL) and 10.38 mg/dL (SD 3.56 mg/dL), respectively. At one month following the operations, the average total and direct bilirubin levels had gradually decreased in the jaundice-improved cases to total bilirubin 4.13 mg/dL (SD 3.54 mg/dL) and direct bilirubin 3.39 mg/dL (SD 2.98 mg/dL). However, the jaundice unimproved cases had severe cholestatic jaundice with average serum total bilirubin at 17.36 mg/dL (SD 7.64 mg/dL) and direct bilirubin at 15.23 mg/dL (SD 5.81 mg/dL). In addition, serum aspartate transaminase was significantly different between the two groups (P < 0.0001), which was inferred to mean continuity of hepatitis in the jaundice not improved cases. Twenty-three of the jaundice improved patients (76.0%) survived, while 23 of the jaundice not improved (53.4%) patients survived.
Supervised variant prioritization

The variants in the related cholestatic jaundice genes were filtered and prioritized from a combined genotype VCF by selecting the Entrez gene symbol. The twenty-six variants identified in 22 of the BA cases were ABCB4, ABCC2, ATP8B1, JAG1, KCNH1, KMT2D, MYO5B, SLCO1B1, TJP2, and UGT1A1 (shown in Table 2). All prioritized variants were consistent with the criteria of moderate or high impact variants, no reported benign or likely benign in the ClinVar database, low allele frequency in the EAS population, or predicted deleterious function by SIFT, PolyPhen2, MutationTaster, or CADD. KMT2D was the most frequently prioritized variant, identified in 6 cases. Half of the cases with variants in this group had decompensated cirrhosis and sequelae of portal hypertension, e.g., hypersplenism, esophageal varices, or ascites. Five cases of BA eventually underwent liver transplantation after developing liver failure (shown in Table 3). All identified variants were validated by visualization alignment on IGV software and confirmed by Sanger sequencing.

Unsupervised variant prioritization

Case-only analysis (enrichment analysis). In the case-only analysis, 254,990 variants from the cases recorded in VCF format were transformed to a tab-delimited format and imported to the R program by which the Entrez gene symbol was appended. Enrichment analysis was performed for 106 coding genes using the EnrichR package together with 8 pathway databases. The remaining three databases with significant pathway enrichment were GO

https://doi.org/10.1371/journal.pone.0277334.t001

Table 1. Clinical and demographic data of the studied biliary atresia patients.

|                      | Jaundice-clear (n = 32) | Jaundice-persistent (n = 43) | Total (n = 75) | P value |
|----------------------|------------------------|----------------------------|----------------|---------|
| Sex                  |                        |                           |                |         |
| Male                 | 17 (47.2%)             | 19 (38.5%)                | 39 (52%)       | 0.443a  |
| Female               | 15 (52.8%)             | 24 (61.5%)                | 36 (48%)       |         |
| Age at surgery (Mean ± S.D.) | 86 ± 38               | 80 ± 25                   | 83 ± 32        | 0.043b  |
| Serum bilirubin before surgery |                |                           |                |         |
| Total Bilirubin (mg/dL) (Mean ± S.D.) | 12.13 ± 3.81           | 11.57 ± 3.93              | 11.89 ± 3.84   | 0.301b  |
| Directed Bilirubin (mg/dL) (Mean ± SD.) | 10.75 ± 3.44           | 9.89 ± 3.63               | 10.38 ± 3.56   | 0.535b  |
| Serum bilirubin 1 month after surgery |                |                           |                |         |
| Total Bilirubin (mg/dL) (Mean ± S.D.) | 4.13 ± 3.54            | 17.36 ± 7.64              | 9.78 ± 8.65    | <0.001b |
| Directed Bilirubin (mg/dL) (Mean ± SD.) | 3.39 ± 2.98            | 15.23 ± 5.81              | 8.43 ± 7.34    | <0.001b |
| Liver function test 1 month after surgery |                |                           |                |         |
| AST (Unit/L) (Mean ± S.D.) | 146.19 ± 71.81         | 327.59 ± 165.59           | 223.63 ± 142.12| <0.001b |
| ALT (Unit/L) (Mean ± S.D.) | 138.95 ± 126.76        | 202.84 ± 147.08           | 169.55 ± 151.15| 0.052b  |
| ALP (Unit/L) (Mean ± S.D.) | 594.05 ± 305.42        | 683.66 ± 533.19           | 642.42 ± 435.73| 0.361b  |
| Anatomical subtype |                        |                           |                |         |
| Type I               | 1 (3.13%)              | 2 (4.65%)                 | 3 (4.00%)      |         |
| Type II              | 2 (6.25%)              | 2 (4.65%)                 | 4 (5.33%)      |         |
| Type III             | 29 (90.62%)            | 39 (90.69%)               | 68 (90.67%)    | 0.907a  |
| Alive                | 23 (50.0%)             | 23 (50.0%)                | 46 (61.3%)     | 0.106a  |
| Died                 | 9 (31.0%)              | 20 (69.0%)                | 29 (38.7%)     |         |

SD, standard deviation; AST, serum aspartate transaminase; ALT, serum alanine aminotransferase; ALP, serum alkaline phosphatase.

a Chi-square test or Fisher exact test.
b T-test or Wilcoxon signed-rank test.

https://doi.org/10.1371/journal.pone.0277334.t001
biological process, WikiPathway, and Reactome. There were 18 pathways enriched from the GO biological process database: skin development (GO:0043588), cilium assembly (GO:0060271), cilium organization (GO:0044782), actin-myosin filament sliding (GO:0033275), muscle filament sliding (GO:0030049), sensory perception (GO:0007600),

Fig 1. Total BA Cases Stratified by Jaundice Status Together with Sex (A) and Day of Surgery (B).

https://doi.org/10.1371/journal.pone.0277334.g001
sensory perception of light stimulus (GO:0050953), inner ear morphogenesis (GO:0042472), retinal cone cell development (GO:0046549), retinal cone cell differentiation (GO:0042670), complement activation, alternative pathway (GO:0006957), parallel actin filament bundle assembly (GO:0030046), positive regulation of cardiac epithelial to mesenchymal transition (GO:0062043), peptide cross-linking (GO:0018149), and organelle assembly (GO:0070925), as shown in S4 Table. With WikiPathway, four pathways, the striated muscle contraction pathway (WP383), Joubert syndrome (WP4656), ciliopathies (WP4803), and arrhythmogenic right ventricular

| Case | Chr position | SNP ID | Gene | Variant Effect | Nucleotide change (Zygosity) | Amino acid change | MAF | SIFT | PolyPhen2 | Mutation Taster | CADD |
|------|--------------|--------|------|----------------|-------------------------------|-------------------|-----|------|-----------|----------------|------|
| B007 | chr20:1065450 | rs876660980 | JAG1 | High | c.703C>T (het) | p.Arg235Ter | 0.00 | 0.12 (T) | 0.906 (P) | 1 (D) | 23.4 |
| B019 | chr18:5765252 | rs754583409 | ATP8B1 | Moderate | c.3220G>A (het) | p.Val1074Ile | 0.00 | 0.02 (T) | 1 (P) | 0.999992 (D) | 22.7 |
| B021 | chr12:21205930 | rs756813138 | MYO5B | Moderate | c.4922G>A (het) | p.Arg1641His | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B031 | chr10:99819250 | rs749441611 | KMT2D | High | c.2601_2602 insTA (het) | p.Pro868TyrfsTer21 | 0.00 | 0.001 (D) | 0.981213 (P) | 23.3 |
| B040 | chr20:10641521 | rs754583409 | JAG1 | Moderate | c.2855C>A (het) | p.Ser952Ter | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B044 | chr12:49040922 | rs34946978 | UGT1A1 | Moderate | c.1091C>T (het) | p.Pro364Leu | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B053 | chr9:69249400 | rs777739913 | TJP2 | Moderate | c.2999C>T (het) | p.Pro1000Leu | 0.00 | 0.001 (D) | 0.98527 (D) | 23.4 |
| B067 | chr20:10641492 | rs527420845 | JAG1 | Moderate | c.2884A>G (het) | p.Thr962Leu | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B080 | chr12:49026295 | rs3782356 | KMT2D | Moderate | c.15671G>A (het) | p.Arg524His | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B085 | chr1:210683976 | rs772578710 | KCNHD1 | Moderate | c.2194C>T (het) | p.Arg736Trp | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B087 | chr7:87422149 | rs540946679 | ABCB4 | Moderate | c.2288T>C (het) | p.Ile763Thr | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B089 | chr12:49034450 | rs535351117 | KMT2D | Moderate | c.10467G>T (het) | p.Gln348His | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B093 | chr12:20901442 | rs535351117 | KMT2D | Moderate | c.1756A>G (het) | p.Arg586Gly | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B097 | chr2:233772309 | rs114982090 | UGT1A1 | Moderate | c.1552C>T (het) | p.Pro517Leu | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B099 | chr12:20875331 | rs34946978 | UGT1A1 | Moderate | c.194C>T (het) | p.Arg65Leu | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B105 | chr12:49034540 | rs535351117 | KMT2D | Moderate | c.10467G>T (het) | p.Gln348His | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |
| B107 | chr18:49847198 | rs764072135 | SLCO1B3 | Moderate | c.4407G>A (het) | p.Met1469Ile | 0.00 | 0.001 (D) | 0.972 (P) | 0.999983 (D) | 27.5 |

Chr, chromosome; SNP, single nucleotide polymorphism; HET, heterozygous; HOM, homozygous; MAF, minor allele frequency in East Asian populations; SIFT, scale-invariant feature transform (T, tolerated; D, damaging); PolyPhen, Polymorphism Phenotyping data collection (B, benign; P, possibly deleterious; D, deleterious); Mutation Taster, Mutation Taster–AA change score and probability value (D, disease-causing; P, polymorphism); CADD, Combined Annotation Dependent Depletion recorded in Phred score.

https://doi.org/10.1371/journal.pone.0277334.1002

Table 2. Rare variants identified by whole-exome studies from related cholestatic jaundice genes.

Chr: chromosome; SNP: single nucleotide polymorphism; Het: heterozygous; Hom: homozygous; MAF: minor allele frequency in East Asian populations; SIFT: scale-invariant feature transform (T, tolerated; D, damaging); PolyPhen: Polymorphism Phenotyping data collection (B, benign; P, possibly deleterious; D, deleterious); Mutation Taster: Mutation Taster–AA change score and probability value (D, disease-causing; P, polymorphism); CADD: Combined Annotation Dependent Depletion recorded in Phred score.

sensory perception of light stimulus (GO:0050953), inner ear morphogenesis (GO:0042472), retinal cone cell development (GO:0046549), retinal cone cell differentiation (GO:0042670), complement activation, alternative pathway (GO:0006957), parallel actin filament bundle assembly (GO:0030046), positive regulation of cardiac epithelial to mesenchymal transition (GO:0062043), peptide cross-linking (GO:0018149), and organelle assembly (GO:0070925), as shown in S4 Table.

With WikiPathway, four pathways, the striated muscle contraction pathway (WP383), Joubert syndrome (WP4656), ciliopathies (WP4803), and arrhythmogenic right ventricular

https://doi.org/10.1371/journal.pone.0277334.1002

PLOS ONE | https://doi.org/10.1371/journal.pone.0277334 November 9, 2022 8 / 17
cardiomyopathy (WP2118), were statistically and significantly enriched from the genes identified in the BA cases, as shown in S5 Table.

For Reactome, Striated Muscle Contraction Homo sapiens (R-HSA-390522) and Muscle Contraction Homo sapiens (R-HSA-397014) were considered to have a statistically significant association with the variants from the cases. The enriched genes from all databases are displayed in Fig 2 and shown in S6 Table.
Case–control analysis (Burden analysis). Seventy-five BA cases and 358 controls were included in the case–control analysis. The 254,990 filtered variants in the coding regions were stored in a single aggregated VCF file. The VCF was converted into PED and MAP files by PLINK software. The phenotypes and sexes of both groups were corrected using R software. The variants were assessed for SNP call rate and IBD relatedness by PLINK software. Sixteen disease control samples had SNP call rates lower than 95% and were excluded from the study; however, the estimated pi-hat score was lower than 0.9 for every sample. The allelotypes of each variant were calculated for Hardy-Weinberg equilibrium, and 15,002 variants were not in equilibrium. A PCA of those samples was created using IPCAP software (S1 Fig). Three BA case outliers were excluded from the analysis.

After filtering, 239,988 variants of 72 BA cases and 342 disease controls remained for burden analysis, which was performed based on gene region (refFlat version hg38) using rvtests software with the CMC exact test. Coiled-coil domains containing 8 (CCDC8) were significantly different among the cases and disease controls with FDR 0.005 (S6 Table).

Trio analysis. Seven BA families were included in the trio analysis based on the mode of inheritance. Variant prioritization was achieved following a dominant and recessive model. In the dominant model, two different de novo variants in KMT2D were detected in unrelated families. However, those variants were later reviewed during the first approach. The result in this section was confirmed as potentially causal variants. There were no nonsynonymous variants inherited from any parents in the recessive model.

Survival analysis of clinical and genomic data

Survival analysis showed that the 5-year median native liver survival for all 75 patients was 1.8 years. Survival in patients who underwent HPE before 90 days of life was significantly better than in those who underwent surgery later than this cutoff age (log-rank p value 0.017), as shown in Fig 3.

After identifying the associated genes by unsupervised and supervised variant prioritization, the cases who carried the variants in any PFIC gene set (as shown in S1 Table), ciliu-
associated genes (GO:0060271, GO:0044782, WP4803), muscle contraction-associated genes (GO:0033275, GO:0030049, WP383, R-HSA-397014), EFEMP1, Sox9, or CCDC8 had significantly lower 5-year native liver survival (25.0%) than those without any variants (50.0%), log-rank p value 0.016 (Fig 4).

A Cox proportional hazards model was used to assess differences between survival rates and the identified variables. When assessed with a multivariate model, patients who underwent surgery before 90 days of life and who carried variants in any of the gene sets independently had a significantly increased risk of death, with hazard ratios of 1.8 (1.02–3.30) and 2.1 (1.19–3.80), respectively (Fig 5).

**Discussion**

Genetic predisposition has been a focus of studies on the pathogenesis of several congenital diseases after high-throughput genome technology became available. GWAS was the first high-throughput technology to explore the genetic associations of BA in various populations. Validated in Thai and other ethnic groups, polymorphisms in the ADD3 and Glypican 1 (GPC1) regions have been reported to be influentially associated with the disease and have significant gene–gene interactions [5–8]. In addition, a single GWAS in Caucasians identified a genetic association with SNPs in the EFEMP1 region [9]. However, follow-on fine mapping studies failed to find a single gene responsible for the BA phenotype. Studies have found that in BA cases, variants known to be associated with other cholestatic diseases could be annotated, and it has been proposed that the inflammatory cholangiopathy found in BA might be a common pathological pathway of multiple disease entities [4].

In this study, nonsynonymous variants were identified through WES in 75 BA cases by two strategic approaches: supervised and unsupervised variant prioritizations. Initially, the coding
sequence variants were filtered according to criteria concerning the depth of coverage, allele depth, impact, minor allele frequency, and predicted deleterious function. For the supervised approach, the study has expanded the discovery of associated variants by prioritizing nonsynonymous variants located in reported gene sets of PFIC, genetic syndromes with BA as a phenotype, and diseases with hyperbilirubinemia features. The cases with candidate genes tended to have mimic phenotypes of BA and poor clinical outcomes as determined by poor postoperative jaundice clearance and early development of biliary cirrhosis.

The unsupervised approach was primarily designed to identify novel causal genes that could help to explain the pathogenesis of BA by exploring rare and deleterious variants by biostatistical analysis. In this approach, the deleterious variants were derived by three biostatistical methods: enrichment analysis, burden testing, and trio studies. The 31 identified coding genes were statistically enriched from 254,990 nonsynonymous variants of the cases.

Cilia and muscle contraction-related functional pathways were surprisingly identified to have the most significant correlation among the three databases. Seven ciliary-associated genes were extracted from this analysis: RPI, TTC8, FAM161A, DNAHS, ARMC9, PCNT, and SDCCAG8. Cholangiocyte ciliopathy has been suspected to be involved in the pathogenesis of BA since the genes identified by WES (KIF3B, PCNT, and TTC17) are involved in ciliary biological processes [24]. Ciliogenesis is required for the optimal development and regular function of cholangiocytes [25]. Cilia are located on the apical membrane of the inner lumen of the biliary epithelium and have a physiologic function as promoters of bile flow and sensors of bile components and osmolarity [26]. Deleterious mutations of ciliary genes have been reported to disrupt ciliary structure and function, leading to ciliopathy [27]. Stagnation of bile flow may lead to destructive inflammation and fibrosclerosis. In addition to the group of ciliopathy genes, the Hedgehog (Hh) signaling pathway plays an essential regulatory role in hepatobiliary embryogenesis. Hh signaling is transduced by binding of the Hh ligand on primary
cilia, by which sequential phosphorylation of kinases is suppressed. Because of the inhibition of downstream pathways, the transcription factor translocates to the nucleus and promotes gene transcription. The Hh pathway is physiologically inactive due to the absence of Hh ligand expression in healthy individuals [28]. Hh signaling is triggered during liver damage, leading to hepatocyte regeneration [29].

**Sox9**, identified in our study, is one of the downstream molecules of the Hh signaling pathway, which is essential for hepatobiliary embryogenesis, microtubule remodeling, and biliary tubulogenesis [30]. In particular, **Sox9** may have an association with BA according to the shared pathogenesis of biliary development.

Five muscle contraction-related genes (**RYR1**, **MYOM1**, **TNNT1**, **NEB**, and **TTN**) were found to be statistically and significantly enriched. A burden test was performed for the second analysis to explore deleterious and rare causative genes between the cases and controls. **CCDC8** was the only statistically significant gene in which six cases carried heterozygous missense mutations. **CCDC8** encodes microtubule regulatory proteins and stabilized proteins that cause 3-M syndrome [31]. In addition, a cholangiocyte primary cilium consists of a microtubule-based axoneme and microtubule-organizing center, which influences normal biliary physiology. Trio-based analysis was performed in seven BA families. Two de novo missense variants in **KMT2D** were identified in two unrelated BA families. The results of this section are the putative loss-of-function of **KMT2D**, which was discovered in the supervised prioritization approach.

---

**Fig 5.** Forest plot of the multivariate cox proportional hazards model of survival data and genetic variations.

https://doi.org/10.1371/journal.pone.0277334.g005
A multivariate Cox proportional hazards model was applied to adjust the time-to-event covariation of clinical and genetic factors. However, anatomic features of biliary atresia were one of statistical interference of the model according to previous studies [32,33]. After adjusted analysis, the analysis showed an independently increasing hazard risk in BA cases whose age at surgery was more than 90 days and those with identified variant status. To correlate with survival data, the individuals who carried nonsynonymous variants in *CCDC8*, *Sox9*, and the gene sets of *PFIC*, ciliary, and muscle contraction biological pathways had significantly lower native liver survival compared to those who did not have any variants in those genes.

Genetic susceptibility has been one of the most studied potential pathogeneses of BA. Many studies have successfully identified several different genes involved in biliary development and function in BA. Moreover, BA has been proposed as a complementary feature in various syndromes that frequently have heterogeneity in clinical manifestation. We suggest that BA could be a shared sequela of a number of obstructive biliary diseases. Biliary tract malformation, insufficiency of physiologic bile clearance, and alteration of bile components are all possibly involved in the initial pathophysiology of biliary obstruction leading to accumulation of bile components in the biliary system and activation of inflammatory cascades. Periduct inflammation progressively damages cholangiocytes and the surrounding mesenchyme and transforms the typical duct structure to irreversible fibrosclerosis found in both the operative findings and histopathology of BA.

In conclusion, coding sequences of the cases were explored in this study to discover rare and nonsynonymous variants in the gene sets described in biliary obstructive diseases or syndromes. Together with unsupervised variant prioritization, cilia and biological muscle pathways were statistically computed to have significant enrichment. A recent study reported that nonsynonymous variants in primary cilia genes were involved in the pathogenesis of BA. The burden test, a genetic association study of rare variants, contributed to the discovery of novel rare associated genes in BA. *CCDC8*, a microtubule coding gene, was the recently reported gene with a significant association with BA as identified by various computational algorithms. Ultimately, the combination of susceptibility identified by many approaches was correlated with poor native liver survival and independently associated with an increased hazard risk of survival with age at surgery. The results of survival analysis emphasized the hypothesis concerning the associations of identified variants with BA, which may be a consequence of various form of congenital cholestatic diseases.

**Supporting information**

S1 Fig. Principal component analysis of snps in the cases and disease controls. (TIF)

S1 Table. Types of progressive familial intrahepatic cholestasis and reported candidate genes. (DOCX)

S2 Table. Syndromes in which BA is one of the phenotypes. (DOCX)

S3 Table. Syndromes with hyperbilirubinemia. (DOCX)

S4 Table. Significant enrichment pathways from the GO biological process database. (DOCX)
S5 Table. Significant enrichment pathways from human WikiPathway database.

(DOCX)

S6 Table. Significant enrichment pathways from human Reactome database.

(DOCX)

S7 Table. Results of burden analysis computed by rvtests software.

(DOCX)

Acknowledgments

Dave Patterson edited the English in the manuscript.

Author Contributions

Conceptualization: Wison Laochareonsuk, Surasak Sangkhathat.

Data curation: Wison Laochareonsuk, Piyawan Chiengkriwate.

Formal analysis: Wison Laochareonsuk, Surasak Sangkhathat.

Investigation: Wison Laochareonsuk.

Methodology: Wison Laochareonsuk, Surasak Sangkhathat.

Project administration: Surasak Sangkhathat.

Software: Komwit Surachat.

Supervision: Surasak Sangkhathat.

Validation: Wison Laochareonsuk.

Visualization: Wison Laochareonsuk, Komwit Surachat.

Writing – original draft: Wison Laochareonsuk.

Writing – review & editing: Surasak Sangkhathat.

References

1. Orlowska E, Czubkowski P, Socha P. Biliary atresia—signs and symptoms, diagnosis, clinical management. Wiad Lek. 2017; 70: 112–117. PMID: 28343205

2. Bill AH, Brennnom WS, Huseby TL. Biliary atresia. New concepts of pathology, diagnosis, and management. Arch Surg. 1974; 109: 367–369. https://doi.org/10.1001/archsurg.1974.01360030019005 PMID: 4850749

3. Wildhaber BE. Biliary atresia: 50 years after the first kasai. ISRN Surg. 2012; 2012: 132089. https://doi.org/10.5402/2012/132089 PMID: 23304557

4. Sangkhathat S, Loachareonsuk W, Maneechay W, Kayasut K, Chiengkriwate P. Variants associated with infantile cholestatic syndromes detected in extrahepatic biliary atresia by whole exome studies: a 20-case series from Thailand. J Pediatr Genet. 2018; 7: 67–73. https://doi.org/10.1055/s-0038-1632395 PMID: 29707407

5. Garcia-Barcelo MM, Yeung MY, Miao XP, Tang CS, Cheng G, So MT, et al. Genome-wide association study identifies a susceptibility locus for biliary atresia on 10q24.2. Hum Mol Genet. 2010; 19: 2917–2925. https://doi.org/10.1093/hmg/ddq196 PMID: 20460270

6. Kaewkiattiyot S, Honsawek S, Vejchhippat P, Chongsrisawat V, Poovorawan Y. Association of X-prolyl aminopeptidase 1 rs17085355 polymorphism with biliary atresia in Thai children. Hepatol Res. 2011; 41: 1249–1252. https://doi.org/10.1111/j.1872-034X.2011.00870.x PMID: 22118303

7. Tsai EA, Grochowski CM, Loones KM, Bessho K, Hakonarson H, Bezerra JA, et al. Replication of a GWAS signal in a Caucasian population implicates ADD3 in susceptibility to biliary atresia. Hum Genet. 2014; 133: 235–243. https://doi.org/10.1007/s00439-013-1368-2 PMID: 24104524
8. Laohareonsuk W, Chiengkriwate P, Sangkhathat S. Single nucleotide polymorphisms within Adducin 3 and Adducin 3 antisense RNA1 genes are associated with biliary atresia in Thai infants. Pediatr Surg Int. 2018; 34: 515–520. https://doi.org/10.1007/s00383-018-4243-3 PMID: 29508064
9. Chen Y, Gilbert MA, Grochoski CM, McEldrew D, Llewellyn J, Waisbourd-Zinman O, et al. A genome-wide association study identifies a susceptibility locus for biliary atresia on 2p16.1 within the gene EFEMP1. PLoS Genet. 2018; 14: e1007532. https://doi.org/10.1371/journal.pgen.1007532 PMID: 30102696
10. Farnaes L, Hildreth A, Sweeney NM, Clark MM, Chowdhury S, Nahas S, et al. Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. NPJ Genom Med. 2018; 3: 10. https://doi.org/10.1038/s41525-018-0049-4 PMID: 29644095
11. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. [cited 20 May 2023]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/MAP format and SAMtools. Bioinformatics. 2009; 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943
13. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20: 1297–1303. https://doi.org/10.1101/gr.107524.110 PMID: 20644199
14. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thorerman KK, et al. The ensembl variant effect predictor. Genome Biol. 2016; 17: 122. https://doi.org/10.1186/s13059-016-0974-4 PMID: 27268795
15. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014; 42: D980–D985. https://doi.org/10.1093/nar/gkt1113 PMID: 24234437
16. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015; 43: D789–D798.
17. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31: 3812–3814. https://doi.org/10.1093/nar/gkg509 PMID: 12824425
18. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013; Chapter 7: Unit 7.20. https://doi.org/10.1002/0471142905.hg0720s76 PMID: 23315928
19. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014; 11: 361–362. https://doi.org/10.1038/nmeth.2890 PMID: 24681721
20. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019; 47: D886–D894. https://doi.org/10.1093/nar/gky1016 PMID: 30371827
21. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B. 1995; 57: 289–300.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81: 559–575. https://doi.org/10.1086/519795 PMID: 17701901
23. Chaichompu K, Abegaz F, Tongsima S, Shaw PJ, Sakuntabhai A, Pereira L, et al. IPCAPS: an R package for iterative pruning to capture population structure. Source Code Biol Med. 2019; 14: 2. https://doi.org/10.1186/s13029-019-0072-6 PMID: 30936940
24. Lam WY, Tang CS, So MT, Yue H, Hsu JS, Chung PH, et al. Identification of a wide spectrum of ciliary gene mutations in nonsyndromic biliary atresia patients implicates ciliary dysfunction as a novel disease mechanism. EBioMedicine. 2021; 71: 103530. https://doi.org/10.1016/j.ebiom.2021.103530 PMID: 34455394
25. Mansini AP, Peixoto E, Thelen KM, Gaspari C, Jia J, Gargiulo SA. The cholangiocyte primary cilium in health and disease. Biochim Biophys Acta Mol Basis Dis. 2018; 1864: 1245–1253. https://doi.org/10.1016/j.bbadis.2017.06.006 PMID: 28625917
26. Larusso NF, Masyuk TV. The role of cilia in the regulation of bile flow. Dig Dis. 2011; 29: 6–12. https://doi.org/10.1159/000324121 PMID: 21691098
27. Waters AM, Beales PL. Ciliopathies: an expanding disease spectrum. Pediatr Nephrol. 2011; 26: 1039–1056. https://doi.org/10.1007/s00467-010-1731-7 PMID: 21210154
28. Omenetti A, Choi S, Michielti G, Diehl AM. Hedgehog signaling in the liver. J Hepatol. 2011; 54: 366–373. https://doi.org/10.1016/j.jhep.2010.03.003 PMID: 21093090
29. Sicklick JK, Li YX, Melhem A, Schmelzer E, Zdanowicz M, Huang J, et al. Hedgehog signaling maintains resident hepatic progenitors throughout life. Am J Physiol Gastrointest Liver Physiol. 2006; 290: G859–G870. https://doi.org/10.1152/ajpgi.00456.2005 PMID: 16322088

30. Antoniou A, Raynaud P, Cordi S, Zong Y, Tronche F, Stanger BZ, et al. Intrahepatic bile ducts develop according to a new mode of tubulogenesis regulated by the transcription factor SOX9. Gastroenterology. 2009; 136: 2325–2333. https://doi.org/10.1053/j.gastro.2009.02.051 PMID: 19403103

31. Hanson D, Murray PG, O’Sullivan J, Urquhart J, Daly S, Bhaskar SS, et al. Exome sequencing identifies CCDC8 mutations in 3-M syndrome, suggesting that CCDC8 contributes in a pathway with CUL7 and OBSL1 to control human growth. Am J Hum Genet. 2011; 89: 148–153. https://doi.org/10.1016/j.ajhg.2011.05.028 PMID: 21737058

32. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015; 526: 68–74. https://doi.org/10.1038/nature15393 PMID: 26432245

33. Superina R, Magee JC, Brandt ML, Healey PJ, Tiao G, Ryckman F, et al. The anatomic pattern of biliary atresia identified at time of Kasai hepatopancreaticoenterostomy and early postoperative clearance of jaundice are significant predictors of transplant-free survival. Ann Surg. 2011; 254: 577–585. https://doi.org/10.1097/SLA.0b013e3182300950 PMID: 21869674