Novel findings from family-based exome sequencing for children with biliary atresia

Kien Trung Tran1,6, Vinh Sy Le1,2,6, Lan Thi Mai Dao1, Huyen Khanh Nguyen3, Anh Kieu Mai4, Ha Thi Nguyen4, Minh Duy Ngo4, Quynh Anh Tran5 & Liem Thanh Nguyen1

Biliary atresia (BA) is a progressive inflammation and fibrosis of the biliary tree characterized by the obstruction of bile flow, which results in liver failure, scarring and cirrhosis. This study aimed to explore the elusive aetiology of BA by conducting whole exome sequencing for 41 children with BA and their parents (35 trios, including 1 family with 2 BA-diagnosed children and 5 child-mother cases). We exclusively identified and validated a total of 28 variants (17 X-linked, 6 de novo and 5 homozygous) in 25 candidate genes from our BA cohort. These variants were among the 10% most deleterious and had a low minor allele frequency against the employed databases: Kinh Vietnamese (KHV), GnomAD and 1000 Genome Project. Interestingly, AMER1, INVS and OCRL variants were found in unrelated probands and were first reported in a BA cohort. Liver specimens and blood samples showed identical variants, suggesting that somatic variants were unlikely to occur during morphogenesis. Consistent with earlier attempts, this study implicated genetic heterogeneity and non-Mendelian inheritance of BA.

Biliary atresia is a progressive inflammation and fibrosis of the biliary tree that consequently results in the development of cholestatic liver disease. BA was first described by surgeon John Thomson in 1892 and is among the most fatal diseases, leading to severe complications in infants. The disease occurs in the early stage of neonates and can be treated by hepatic portoenterostomy or Kasai operation. After surgical treatment, however, approximately 50% of affected infants require liver transplantation, while the rest would sustain their own liver up to the age of 5–10 years. A study on Vietnamese BA patients reported that 84% and 71% of Kasai-treated patients survived after 1–2 years, respectively. Additionally, the respective ratios were 52% and 28% for the group without Kasai treatment. It is estimated that after hepatic portoenterostomy operation, 70–80% of patients with BA still require liver transplantation by adulthood due to the progressive development of liver scarring, failure and cirrhosis.

Although BA has been extensively studied, its aetiology and pathogenesis remain elusive. Several hypotheses explaining the cause of the disease, including viral infection, autoimmune-mediated bile duct destruction, biliary toxin, and genetic abnormality, have been proposed. Regarding genetic aspects, debate over the Mendelian mechanism of the disease has been raised due to a lack of familial BA and a discordant presentation of BA in the monozygotic twin. Nevertheless, some cases with familial BA have been reported, suggesting that either a recessive autosomal inheritance or a combination of genetic and acquired factors might contribute to the disease’s aetiology. In addition, some studies have examined an association between BA and microchimerism, where the genetic trait is maternally transferred from the mother and later contributes to phenotypic heterogeneity and non-Mendelian inheritance. More specifically, a heterozygous transition CFC1:c.433G > A in 5 BA patients with polysplenia syndrome implies a genetic predisposition to BA splenic malformation. In a mouse model, inactivation of the hepatocyte nuclear factor 1 beta gene (Hnf1β) causes abnormalities of the gallbladder and intrahepatic bile ducts, resulting in severe jaundice. Observations of an increased incidence of BA in some groups, such as Asian and Polynesian populations, suggest that genetic and environmental factors might cause the disease. Recent genetic studies have revealed a linkage between cholestatic jaundice and genetic predispositions in both nuclear DNA and mitochondrial DNA.

1 Vinmec Research Institute of Stem Cell and Gene Technology, 458 Minh Khai, Hai Ba Trung District, Hanoi, Vietnam. 2 University of Engineering and Technology, Vietnam National University Hanoi, 144 Xuan Thuy, Cau Giay District, Hanoi, Vietnam. 3 Bioequivalence Center, National Institute of Drug Quality Control, 11/157 Bang B, Hoang Mai District, Hanoi, Vietnam. 4 Vinmec International Hospital, 458 Minh Khai, Hai Ba Trung District, Hanoi, Vietnam. 5 Vietnam National Children’s Hospital, 18/879 La Thanh, Dong Da District, Hanoi, Vietnam. 6 These authors contributed equally: Kien Trung Tran and Vinh Sy Le. *email: trantrungkien80@gmail.com
The prevalence of BA is 1 in 8000–18,000 live births and varies among countries and groups, with a dominance of females over males. The disease occurs more frequently in Southeast Asia and the Ocean Pacific. It is approximately 1 in 5000 live births in Taiwan compared to 1 in 14,000–20,000 in North America or Western Europe. To our knowledge, there is no epidemiological study of BA in Vietnamese, as is the prevalence of this fatal disease. The prevalence of BA in Vietnam is estimated to be as high as 1 in 2400 live births as equal to that of the Ocean Pacific regions. Although BA and BA-related liver diseases are often observed in Vietnamese infants and are life-threatening diseases, few studies have been reported thus far. To date, the Kasai portoenterostomy procedure has been introduced as a routine surgical practice and offers a better opportunity to patients. However, a number of patients still require liver transplantation after the operation or have a low quality of life due to the disease’s complications. Recently, next-generation sequencing (NGS), particularly whole exome sequencing (WES), has been increasingly applied for detecting variants in patients with cholestasis. It appears to be a powerful tool to aid diagnosis and to provide timely and accurate therapeutic treatments. Therefore, we aimed to investigate the genetic pattern of BA by conducting a family-based WES for children with BA in hope of exploring new and characterized causative variants, which would shed light on the aetiology of the deadly disease.

Materials and methods

Patient recruitment. BA diagnosis was based on intraoperative findings and liver biopsy. Patients with confirmed BA and their parents were recruited at Vinmec International Hospital and Vietnam National Hospital of Pediatrics in Hanoi from May 2019 to May 2020. A written informed consent form was provided to the parents for their participation. The study was approved by the Ethics Committee of the hospitals in accordance with the Declaration of Helsinki.

Sample collection and DNA extraction. Approximately 2 mL of peripheral blood from patients and their parents was collected in an EDTA anticoagulant tube and stored at −80 °C. Liver wedge specimens were collected from the Kasai operation, snap-frozen in liquid nitrogen and stored at −80 °C. Genomic DNA was extracted by using a DNA Mini Blood Isolation Kit based on the manufacturer’s protocol (Qiagen, Germany). DNA samples were quantified by fluorescence using a Qubit BR Quantification Kit (Invitrogen, USA). Extracted DNA samples were preserved at −80 °C for future uses.

Whole exome sequencing. Exome sequencing libraries were prepared by using a Nextera Rapid Capture Kit (Illumina, Calif, USA) based on the manufacturers’ protocol with slight modifications. The library concentration was quantified by a Qubit dsDNA Broad Range Assay Kit (Invitrogen, USA). Library size was measured by using a LabChip 3 K Hisense Kit (PerkinElmer, USA). Paired-end exome sequencing with 150 bp cycles was performed on a HiSeq 4000 (Illumina, Calif, USA), targeting an averaged depth of 100X.

Bioinformatics analysis. Variant calling and annotation were performed based on highly regarded tools. Reads with low quality, adapters and noise were removed prior to the downstream analysis by using FastQC and Trimmomatic. Reads were aligned with the reference genome GRCh38 version. Bowtie2, BWA and Qualimap were used for quality control. To minimize the false-positive rate, multiple variant calling tools, including the Genome Analysis Toolkit (GATK), SAMtools mpileup and Freebayes, were mutually used.

Variant classification, functional prediction and genotype-phenotype analysis. A stringent strategy was applied for variant classification, including (i) inclusion of rare and nonsynonymous variants with a minor allele frequency (MAF) <1% against three databases: the Kinh Vietnamese population (KHV) obtained from our previous study on the Vietnamese human genome database; GnomAD (https://gnomad.broadinstitute.org/) and 1000 Genome Project; (ii) inclusion of variants with 3 types of inheritance modes: X-linked, homozygous and putative de novo; and (iii) variants with a CADD Phred score of >10, indicating the 10% most deleterious variants in the genome. In silico prediction tools, including SIFT, PolyPhen-2, Mutation Taster, I-Mutant and HOPE, were employed to predict the impact of genetic changes. Molecular Signatures Databases (MSigDB v7.2) were used to compute the candidate genes with the gene sets of human phenotype ontology.

Validation of WES results. Selected variants were then confirmed by bidirectional Sanger sequencing. Proper primers were designed for these variants, followed by PCR amplification and sequencing on an ABI 3500 DX system using a BigDye Terminator v3.1 (Thermo Fisher, Massachusetts, USA).

Ethics approval. The study was approved by the Ethics Committee of Vinmec International General Hospital JSC (Decision No. 48/2019/QD-VMEC).

Consent to participate. A written informed consent form was provided to the parents prior to their participation.

Consent for publication. The participants provided consent for publication of all relevant data and this manuscript.
| Proband  | Birth year | Sex | Age diagnosed | Blood test (at the time of enrolment) | Family history, clinical description |
|----------|------------|-----|---------------|--------------------------------------|-------------------------------------|
| BA001    | 2016       | F   | 1 m/o         | N/A, N/A, 190.5, 211.9, 30.9, 73.59, 36.8 | 1st child; No family history of BA or other genetic disease |
| BA002_3  | 2014       | F   | 1.5 m/o       | N/A, 34, 57.8, 109.9, 168.4, 143.8, 83 | Her younger brother was diagnosed with BA; Currently, she developed signs of cirrhosis |
| BA002_4  | 2018       | M   | 2 m/o         | 439, 44.8, 163.3, 205.4, 1212.4, 289.6, 150.8 | His sister (BA002_3) showed similar CJ symptoms and diagnosed with BA |
| BA003    | 2018       | F   | 50 days       | N/A, N/A, 259.3, 289.9, N/A, 210.7, 111.5 | 2nd child; Her grandfather's daughter died at 1 m/o and showed pale stool |
| BA004    | 2010       | M   | 2 m/o         | 677, 43.3, 282.5, 301.5, 820, 20, 5.4 | 3rd child; No family history of BA or other genetic disease; Splenomegaly; stool with fresh blood |
| BA005    | 2011       | M   | 45 days       | 501, 36.8, 153.3, 166.7, 243.3, 15.7, 4.6 | 2nd child, full term, born via C-section with birthweight of 3.5 kg; No family history of BA or other genetic disease; Developed cirrhosis after Kasai operation |
| BA006    | 2015       | M   | 2 m/o         | N/A, 37.1, 107.3, 206.8, 200, 132.7, 73.3 | 1st child; No family history of BA or other genetic disease |
| BA007    | 2017       | M   | 2 m/o         | 275, 41.1, 66.1, 85.8, 58.8, 8.8, 2.2 | 1st child; No family history of BA or other genetic disease |
| BA009    | 2018       | M   | 1.5 m/o       | 777.2, 30.7, 167.6, 249.7, 410.5, 238.1, 131.7 | 3rd child; No family history of BA or other genetic disease |
| BA010    | 2010       | F   | 1 m/o         | 668, 40.1, 173.4, 129.2, 249.9, 25.7, 5.9 | 1st child, full term, C-section delivered with birthweight of 3.4 kg; No family history of BA or other genetic disease |
| BA011    | 2012       | M   | 1 m/o         | 310, 45, 50, 67.2, 66.2, 7.9, 1.3 | 2nd child; Vaginal delivered; No family history of BA or other genetic disease |
| BA012    | 2015       | F   | 2.5 m/o       | 748, 41.2, 876, 585.2, 624.4, 58.2, 30.6 | 2nd child; No family history of BA or other genetic disease; Cirrhosis developed; Splenomegaly |

Continued
| Proband | Birth year | Sex | Age diagnosed | Blood test (at the time of enrolment) | Family history; clinical description |
|---------|------------|-----|---------------|--------------------------------------|-------------------------------------|
|         |            |     |               | ALP (124–341 IU/L) | ALB (36–50 g/L) | ALT (< 50 IU/L) | AST (< 50 IU/L) | γ-GT (12–123 IU/L) | T-Bil (2–20 μmol/L) | D-Bil (< 8.6 μmol/L) |
| BA013   | 2010       | F   | 1 m/o         | 249                    | 40                    | 38.2            | 52.6            | 57.3            | 21.8          | 4.3               |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA014   | 2016       | F   | 2 m/o         | N/A                    | 42.2                  | 109.5           | 87.8            | 201.7          | 8.9           | 2.2               |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA015   | 2016       | F   | 1 m/o         | N/A                    | 39.5                  | 167.8           | 115.8           | 460            | 80.3          | 46.4              |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA016   | 2014       | M   | 3 m/o         | 353.8                  | 37.09                 | 110.4           | 196.1           | 424.1          | 16.8          | 5.7               |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA017   | 2016       | F   | 2 m/o         | 516                    | 34.6                  | 161.5           | 282.7           | 224.6          | 56.8          | 26.8              |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA018   | 2015       | M   | 2 m/o         | 33.5                   | 82.9                  | 191.2           | 371             | 180.8          | 107.9         |                  |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA019   | 2017       | F   | 3 m/o         | 1195                   | 29.7                  | 64.9            | 150.9           | 384.1          | 38.8          | 16.9              |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA020   | 2009       | M   | 2 m/o         | 386                    | 39.8                  | 87.4            | 80.7            | 176.7          | 16.7          | 5.9               |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA021   | 2018       | F   | 65 days       | 584.8                  | 37.6                  | 220.9           | 323.1           | 918.9          | 224.1         | 123               |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA023   | 2018       | M   | 3 m/o         | 635.7                  | 36.75                 | 163.9           | 258.5           | 404            | 153.9         | 85                |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA024   | 2017       | F   | 2 m/o         | 280.3                  | 33.2                  | 63.9            | 66.9            | 88             | 14.7          | 5                 |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA025   | 2018       | F   | 3 m/o         | 300                    | 41.2                  | 175.8           | 226.3           | 465.1          | 131.1         | 82.3              |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |
| BA026   | 2018       | M   | 2 m/o         | 498                    | 43.2                  | 178.7           | 240.5           | 781            | 76.7          | 52                |
|         |            |     |               |                        |                       |                 |                 |                 |               |                  |

Continued
| Proband | Birth year | Sex | Age diagnosed | Blood test (at the time of enrolment) | Family history, clinical description |
|---------|------------|-----|---------------|--------------------------------------|-------------------------------------|
|         |            |     |               | ALP (124–341 IU/L) | ALB (36–50 g/L) | ALT (< 50 IU/L) | AST (< 50 IU/L) | γ-GT (12–123 IU/L) | T-Bil (2–20 µmol/L) | D-Bil (< 8.6 µmol/L) |
| BA027   | 2018       | M   | 40 days       | 497        | 38.5    | 78     | 104.9    | 565.1               | 11.3               | 4.1                     |
|         |            |     |               |           |         |        |          |                      |                    |                         |
| BA028   | 2016       | M   | 28 days       | 421        | 36.5    | 63.7   | 80.2     | 171.6               | 9.2                | 2.4                     |
|         |            |     |               |           |         |        |          |                      |                    |                         |
| BA029   | 2014       | M   | 1 m/o         | N/A        | 39.5    | 221.6  | 227.8    | 527.7               | 89.9               | 51.7                    |
|         |            |     |               |           |         |        |          |                      |                    |                         |
| BA030   | 2018       | M   | 1 m/o         | 404        | 31.4    | 123.6  | 210.2    | 900.3               | 208                | 120.4                   |
|         |            |     |               |           |         |        |          |                      |                    |                         |
| BA031   | 2018       | M   | 15 days       | 556        | 35.8    | 56.9   | 142.6    | 855.5               | 143.7              | 80.3                    |
|         |            |     |               |           |         |        |          |                      |                    |                         |
| BA032   | 2018       | M   | 1 m/o         | 427.4      | 34.4    | 76.4   | 144.6    | 144.6               | 150.6              | 69.6                    |
|         |            |     |               |           |         |        |          |                      |                    |                         |
| BA033   | 2018       | F   | 2.5 m/o       | 648        | 32.8    | 134.7  | 255.2    | 131.8               | 368.8              | 188                     |
|         |            |     |               |           |         |        |          |                      |                    |                         |
| BA034   | 2018       | F   | 29 days       | 35.2        | 44.2    | 150.8  | N/A      | 104.9               | 60.9               |                         |
|         |            |     |               |           |         |        |          |                      |                    |                         |
| BA035   | 2015       | F   | 72 days       | 374        | 33.1    | 82.5   | 173.3    | 70                  | 278.8              | 142                     |
|         |            |     |               |           |         |        |          |                      |                    |                         |
| BA036   | 2018       | F   | 1.5 m/o       | 311        | 39.7    | 164.6  | 265.9    | 280.4               | 131.3              | 97.8                    |
|         |            |     |               |           |         |        |          |                      |                    |                         |

Continued
Results

Clinical features. We recruited a total of 42 children who had been diagnosed with BA based on intraoperative findings and liver biopsy. All patients showed typical BA symptoms, such as prolonged jaundice, acholic stool and abnormalities of the biliary tract at early infantile. The patients, including 23 males and 19 females born from 2009 to 2019, but the majority of patients were born in recent years. All patients underwent Kasai surgery immediately after birth (mostly after their 2 months of life), but the concentrations of bilirubin and serum enzymes indicating liver function, such as ALP, ALT and γ-GT, remained high at the time of enrolment (Table 1). Some patients have developed liver cirrhosis (BA002_3, BA005, BA012, BA013 and BA018). One BA patient has infected with CMV (BA037). Several probands whose siblings were reported to develop liver diseases or other genetic conditions, including BA (BA002_4), primary sclerosing cholangitis (BA032), choledochal cyst (BA042) and haemophilia (BA025). Four mothers experienced abnormal pregnancy (BA024, BA027, BA036, BA038). The remaining families did not show any significant concern during their pregnancy and had no family history of BA or other genetic conditions. Excluding one family who failed to come for blood drawing after the first health examination, we were finally able to collect blood samples from 41 BA-affected children and 42 healthy siblings.

| Proband | Birth year | Sex | Age diagnosed | ALP (124–341 IU/L) | ALB (36–50 g/L) | ALT (< 50 IU/L) | AST (< 50 IU/L) | γ-GT (12–123 IU/L) | T-Bil (2–20 μmol/L) | D-Bil (< 8.6 μmol/L) |
|----------|------------|-----|---------------|-------------------|----------------|----------------|----------------|------------------|-----------------|------------------|
| BA037    | 2018       | M   | 1 m/o         | 808.3             | 38.9           | 246.7          | 317.8          | 329.5            | 139.8           | 85.8             |
| BA038    | 2018       | M   | 66 days       | 240               | 39.8           | 113.8          | 87.2           | 833              | 81.4            | 47.3             |
| BA039    | 2019       | F   | 40 days       | 618               | 36.2           | 115.4          | 130.7          | 899.3            | 130.3           | 101.2            |
| BA040    | 2019       | M   | 2 m/o         | 629.2             | 37.7           | 93.4           | 220            | 604.1            | 256.7           | 117              |
| BA041    | 2019       | M   | 28 days       | 320               | 42.6           | 138.6          | 265.5          | 905.2            | 108             | 82.7             |
| BA042    | 2019       | F   | 1 m/o         | N/A               | N/A            | 211            | 663            | N/A              | 229             | 120              |

Table 1. Clinical features of children with biliary atresia. M male, F female, m/o month old, ALP alkaline phosphatase, ALB albumin, ALT alanine aminotransferase, AST aspartate aminotransferase, γ-GT gamma-glutamyl transferase, T-Bil total bilirubin, D-Bil direct bilirubin, HBV hepatitis B, CMV cytomegalovirus, CJ cholestatic jaundice, N/A not available, BA002_3 and BA002_4 were siblings.
their parents. Among these 41 children, we collected liver specimens from 18 children obtained from the Kasai operation.

**Genetic properties.** We applied a strict filtering strategy by removing variants with MAF > 1%, synonymous variants and variants with a CADD scaled score < 10. Finally, we identified a total of 28 variants in 25 genes from our BA-affected cohort (Table 2). All variants were subsequently confirmed by Sanger sequencing (Fig. S1). Among the 28 detected variants, 17 X-linked variants (61%) were detected in 17 different genes, 6 de novo variants (21%) were detected in 6 genes from 5 probands, including INV, ELPL2, TINAG, CEP63, CDC1536, and BCAR1, and 5 homozygous variants were identified in 5 genes (18%) (Fig. 1), including HACE1, VPS13C, RAPGEF4, FOCAD and INV (Table 2). Family #2 involved two siblings with similar phenotypes (early onset jaundice, BA diagnosed). Two X-linked and 1 homozygous variants were detected in the male sib of family #2, and none were detected in his sister (Table 2). Interestingly, several genes with genetic predisposition were observed in unrelated patients, including AMER1 (BA004 and BA007), INV (BA014 and BA041), and OCRL (BA032 and BA041). Noticeably, proband BA014 carried an INV de novo variant, while proband BA041 carried an INV homozygous variant (Table 2).

In addition to blood samples, we were able to collect 18 liver specimens from our BA cohort. Of these, blood and liver samples from 8 children shared identical variants (BA009, BA016, BA032, BA036, BA037, BA038, BA040 and BA041). Additionally, we did not detect any significant variants based on our rationales for variant classification (Table 3). In other words, this study did not detect any somatic variants from liver samples.

**Effect of genetic predisposition.** The detected variants showed extremely low MAFs against three employed databases: Kinh Vietnamese (KHV), GnomAD and 1000 Genome Project (Table 2). We noticed that the MAFs of the HACE1 and VPS13C variants were above 1% against the KHV database, while the rest were significantly below the thread hold of 1%. All variants with CADD scaled scores were above 10 and mostly above 20, indicating either the 10% or 1% most deleterious substitutions, respectively. Among these variants, INVS:c.C208>T was the most deleterious, with the highest scaled score of 37 (Table 2). The Polyphen-2 and SIFT tools showed a consensus on the damaging impact of HACE1, PHKA1, XIAP, and AMER1 (c.A1075>T), FOP1B, MAOA, BCAR1, FOCAD, ARSF and OCRL variant, while the rest varied from tools (Table S1). We used I-Mutant to predict the stability of amino acid substitution for 28 identified variants via the change of free energy change values (DDG). The results show that except OCRL:c.T2603 > A(p.Met876Lys), which increased the stability of the mutant compared to that of the wild-type variants, all variants showed decreased stability (Table S2). The HOPE tool was used to predict the structural effect of missense variants, showing changes in residue size and hydrophobic and structural stability (Fig. S2). Changes in amino acid size and charge resulted in a loss of interaction and disturbance of protein function. Several variants, for example HACE1:c.G1660>A(p.Ala651Thr) and PHKA1:c.G478>A(p.Asp160Asn), whose wild-type residues are located in important domains. Thus, any substitution in these regions was predicted to lead to a functional disturbance. In contrast to I-mutant prediction, HOPE showed that an alternation of methionine by lysine residue in the variant OCRL:c.T2603 > A(p.Met876Lys) can disturb the hydrophobic interaction of the altered residue with other molecules on the surface of the protein (Fig. S2).

**Analysis of biological function and human disease phenotype.** Compute overlaps of 25 candidate genes to the human phenotype ontology from the Molecular Signatures Database, involving 4,494 gene sets (FDR q value < 0.05), indicated that the candidate genes fell into various human phenotype gene sets, ranging from gonosomal inheritance and X-linked recessive inheritance to involuntary movements (Table 4). We also computed our gene set to find the association of these genes with the reported phenotypes available from the HPO and Monarch Initiative (Table S3). However, we did not find any overlapping phenotypes from these databases. The reason might be a lack of genes/pathways associated with the BA phenotype in these available databases, which are often dominated by studies on Caucasians, where the prevalence of BA in this group is much lower than that in Asians. By applying the same strategy to identify the potential contribution of ciliary dysgenesisa in BA diagnosis, we used a gene set containing 2016 genes of interest42. We found that some genes from our study, including BCO, INV and OCRL, were included in this gene set. This result suggested the novelty of BCO, INV and OCRL from our BA cohort.

Similar to a previous study41, we did not identify any variants in some genes that have been previously suggested to be associated with BA or BA-related diseases, such as PKD2 (polycystic kidney disease 2, polycystic kidney and hepatic disease 1), CFC1 (polysplenia), JAG1 (Alagille syndrome) and PKD1L1 (biliary atresia splenic malformation syndrome- BASM). We also did not find significant variants in the susceptibility loci of ADD3, XPNPEP1, GPC1, ARF6 and EFEMP1, as suggested by GWAS43.

**Discussion**

Similar to other previous studies, we attempted to reveal the genetic pattern of BA disorder by conducting trio-based exome sequencing for 40 families involving 41 children with BA. Going beyond this establishment in a genetic study for such a rare and complex disorder, we further tested our hypothesis of whether the detected variants occurred in somatic or germline cells by sequencing both blood and available liver specimens obtained from our BA cohort. Due to the complexity of BA, we applied a stringent bioinformatics pipeline and tight quality control to determine either the rarest variants or putative de novo events from our BA cohort, which would avoid a huge number of variants as often experienced from mass sequencing. Taking this straightforward principle enabled us to end up with a total of 28 variants in 25 respective genes. Identical variants detected from blood
and liver samples allowed us to rule out the occurrence of somatic variants in the development of the disease as previously hypothesized45.

In agreement with previous studies, our results showed an intriguing genetic aspect of BA, which was highly heterogeneous. It is worth noting that along with other variants, this study found 3 genes whose variants occurred

and liver samples allowed us to rule out the occurrence of somatic variants in the development of the disease as previously hypothesized45.

In agreement with previous studies, our results showed an intriguing genetic aspect of BA, which was highly heterogeneous. It is worth noting that along with other variants, this study found 3 genes whose variants occurred

Table 2. Genetic characteristics of Vietnamese children with biliary atresia. Chr chromosome, M male, F female, A.A amino acid, MOI mode of inheritance, UM unaffected mother, UF unaffected father, AS affected sibling, N/A not available, MAF minor allele frequency, KHV Kinh Vietnamese, 1 KG 1000 Genome Project, CADD scaled score.

| Proband | Sex (M/F) | Chr | Position | Gene | DNA change | MOI | Genotype | Proband | UM | UF | AS | Allele frequency | #homozygotes | CADD (Phred) |
|---------|-----------|-----|----------|------|------------|-----|----------|----------|-----|-----|----|-----------------|--------------|------------|
| BA002_4 | M         | chr6| 104,771,988 | HACE1 | NM_001350555:c.G1660>A | AR | T/T | T/C/T | C/T | 0.010/0.001731/0.001996 | 0 | 26.2 |
|         |           | chr7 | 72,684,557 | PHKA1 | NM_001122670:c.G478>A | X-linked | T | C/T | C/T | 0.008/0.001949/0.005298 | 0 | 23.2 |
|         |           | chrX | 123,655,767 | THOC2 | NM_001081550:c.G1261>A | X-linked | T | C/T | C/T | 0.008/0.001483/0.001589 | 0 | 21.6 |
| BA004   | M         | chrX | 123,888,703 | XIAP | NM_001167655:c.C962>G | X-linked | G | G/C | N/A | 0/0.003356/0.003958 | 0 | 29.5 |

Table 2. Genetic characteristics of Vietnamese children with biliary atresia. Chr chromosome, M male, F female, A.A amino acid, MOI mode of inheritance, UM unaffected mother, UF unaffected father, AS affected sibling, N/A not available, MAF minor allele frequency, KHV Kinh Vietnamese, 1 KG 1000 Genome Project, CADD scaled score.

and liver samples allowed us to rule out the occurrence of somatic variants in the development of the disease as previously hypothesized45.

In agreement with previous studies, our results showed an intriguing genetic aspect of BA, which was highly heterogeneous. It is worth noting that along with other variants, this study found 3 genes whose variants occurred
Figure 1. Mode of inheritance of identified variants from the biliary atresia cohort. X-linked variants are presented in blank; de novo variants are presented in grey; and autosomal recessive variants are presented in dots.

Table 3. Identical variants detected from blood and liver samples. *Chr* chromosome, *M* male, *F* female, *Ref* reference, *Alt* alternative, *A.A* amino acid, *MOI* mode of inheritance, *UM* unaffected mother, *UF* unaffected father.
in unrelated probands, including AMER1, INVS and OCRL. While the aetiology of BA remained unclear and was unlikely to follow the Mendelian model, our results implicated their role in the disease's development. Overlapping findings of BCOR, INVS and OCRL in the Vietnamese BA cohort with a large comprehensive ciliopathy and biliary genes of interest in the previous study further supported the possibility of the causative role of these genes in BA. AMER1 (MIM#300647) encodes APC membrane recruitment protein 1, which acts as an inhibitor of the canonical Wnt/beta-catenin signalling pathway and controls hepatobiliary development during embryogenesis. In mature healthy liver cells, it is mostly inactive, and the abnormal Wnt/beta-catenin signalling pathway can promote the development of liver diseases. AMER1 associates with osteopathia striata with cranial sclerosis and Wilms tumour development. The gene is involved in the activation of the Wnt/beta-catenin signalling pathway, which drives hepatocarcinoma and cholangiocarcinoma. In addition, analysis of the effect of genetic predispositions of AMER1 variants indicated that they were damaging because the alternated residues were located in highly conserved positions. The alternations might lead to destabilization of the local conformation and a loss of protein interaction (Table S1, S2, Fig. S2). Despite a lack of AMER1 to typical BA phenotypes, we inferred its indirect role in the development of BA as a result of activation of the Wnt/beta-catenin signalling pathway.

Our study highly suggested INVS as a BA candidate gene owing to INVS variant detection in 2 unrelated probands, their mode of inheritance and the effect of genetic predisposition. In particular, INVS: c.C208 > T (p.Arg396*) was de novo, and a loss-of-function variant with a CADD score of 37 and its allele frequency was absent from all employed databases. INVS encodes inversin protein, which plays a role in primary cilia function and is involved in the cell cycle. Intriguingly, inactivation of INVS in a mouse model shows a significant increase in bilirubin levels compared to that of the wild-type and pathogenic changes in ductal plate malformation in the intrahepatic biliary of the mutant mouse. The association of INVS with BA had not been previously established due to an absence of INVS variants detected in BA patients. However, INVS is associated with infantile nephropathitis type 2 and a homozygous variant from 2 BA unrelated patients (BA014 and BA041). To our knowledge, this novelty is first reported in BA patients, although future studies are needed to clearly explore the role of INVS in BA development. Similar to BCOR and INVS, OCRL encodes inositol polyphosphate 5-phosphatase, which might be involved in primary cilia assembly. OCRL has been widely reported to be linked to Lowe and Dent syndrome, where clinical manifestations often overlap with Zellweger spectrum disorders, characterized by low muscle tone, feeding difficulty, seizures and liver dysfunction. Likewise, a lack of an association of OCRL and BA or liver diseases remains a gap for future investigation.

As a result of a rapidly declining cost of DNA sequencing, dozens of rare and previously undiagnosed genetic disorders are currently detectable. For the last 10 years, NGS technology has revolutionized our understanding of human genetics with a high level of accuracy, cost effectiveness and high throughput capability. NGS is steadily becoming a standard in routine diagnostic practices. In BA studies, mitochondrial DNA has been found

| Gene set name | # genes in gene set (K) | Description | # genes in overlap (k) | p value | FDR q-value | Gene overlap |
|---------------|-------------------------|-------------|-----------------------|---------|-------------|--------------|
| HP_GONOSOMAL_INHERITANCE | 253 | Gonosomal inheritance | 12 | 0.0474 | 1.49E−20 | ATRX, OCRL, THOC2, MAOA, KIF4A, BCORL1, XIAE, POF1B, PHKA1, BCOR, AMER1, UBQLN2 |
| HP_X_LINKED_RECESSIVE_INHERITANCE | 173 | X-linked recessive inheritance | 9 | 0.0556 | 5.66E−11 | ATRX, OCRL, THOC2, MAOA, KIF4A, BCORL1, XIAE, POF1B, PHKA1 |
| HP_SELF_INJURIOUS_BEHAVIOR | 108 | Self-injurious behaviour | 6 | 0.0156 | 8.68E−09 | ATRX, OCRL, THOC2, MAOA, BCOR, ELP2 |
| HP_ABNORMAL_EMOTION_AFFECT_BEHAVIOR | 415 | Abnormal emotion/affect behaviour | 7 | 0.0169 | 4.97E−09 | ATRX, OCRL, THOC2, MAOA, BCOR, ELP2, VPS13C |
| HP_NEUROLOGICAL_SPEECH_IMPAIRMENT | 1022 | Neurological speech impairment | 9 | 0.0088 | 6.24E−09 | ATRX, OCRL, KIF4A, BCOR, AMER1, UBQLN2, ELP2, ZNF41, HACE1 |
| HP_DELAYED_SPEECH_AND_LANGUAGE DEVELOPMENT | 696 | Delayed speech and language development | 8 | 0.0115 | 6.63E−09 | ATRX, THOC2, BCORL1, AMER1, ELP2, ZNF41, HACE1, CEP63 |
| HP_AUTISTIC_BEHAVIOR | 450 | Autistic behaviour | 7 | 0.0156 | 8.68E−09 | ATRX, THOC2, MAOA, BCORL1, BCOR, VPS13C, ZNF41 |
| HP_ABNORMAL_AGGRESSIVE_IMPULSIVE_OR_VIOLENT_BEHAVIOR | 251 | Abnormal aggressive, impulsive or violent behaviour | 6 | 0.0239 | 9.12E−09 | ATRX, OCRL, THOC2, MAOA, BCOR, ELP2 |
| HP_SHORT_STATURE | 1152 | Short stature | 9 | 0.0078 | 1.75E−08 | ATRX, OCRL, THOC2, BCOR, AMER1, ELP2, HACE1, CEP63, INVS |
| HP_INVOLUNTARY_movements | 905 | Involuntary movements | 8 | 0.0088 | 5.05E−08 | ATRX, OCRL, THOC2, MAOA, BCORL1, UBQLN2, ELP2, VPS13C |

Table 4. Analysis of human phenotype ontology.
to associate with BA, suggesting the role of mitochondria in underling the pathogenic mechanism.\textsuperscript{17} WES has revealed dozen candidate genes either encode ATP-binding cassette transporters (the ABC superfamily)\textsuperscript{41,42} or are involved in the Notch signalling pathway, such as \textit{JAG1}\textsuperscript{43} and \textit{NOTCH2}\textsuperscript{44}. GWAS have highlighted a strong association between BA and some variants in the \textit{ADD3} gene found on 10q24.\textsuperscript{45} Another subsequent study on 171 BA patients and 1,630 controls of European descent found the strongest signal at rs7099604 in the \textit{ADD3} gene.\textsuperscript{45} A significant association was found between variant rs17095355 on the \textit{XPNPEP1} gene and the disease.\textsuperscript{45} Taken together, the aetiology of BA remains challenging due to the involvement of multiple genes and complex mechanisms. Being encouraged by the pioneers, we provided a concrete genetic aspect obtained from an exome trio-based study of a Vietnamese BA cohort. The findings add to our knowledge of the genetic heterogeneity and complexity of BA disorder.

Conclusion

The aetiology of BA remains challenging because there is a lack of conclusive evidence despite extensive research and medical practices for hundreds of years. However, the recent development of NGS technology and its application in studies of BA and liver diseases have gradually revealed the hidden genetic picture of BA aetiology, where dozens of BA-associated genes have been found. Our study identified 28 variants in 25 genes (all validated) from 41 children with BA. These variants were in the 10% most deleterious and were either rare or extremely rare in the population genome database. A combination of functional prediction and analysis of biological processes enabled us to suggest these candidate genes for the development of BA, particularly with those detected in unrelated BA individuals, including \textit{AMER1}, INV3 and \textit{OCRL}. Identical variants detected from blood and liver wedge specimens from each BA individual suggested that somatic variants in the liver cells were unlikely to occur during morphogenesis. Taken together, we highlighted the genetic heterogeneity of BA and ruled out the Mendelian model. Future studies are needed to further explore the roles of these genes in the development of BA.

Data availability

Data are available from this manuscript and supplementary information.

Received: 26 May 2021; Accepted: 21 October 2021
Published online: 08 November 2021

References

1. Thomson, J. On congenital obliteration of the bile-ducts. \textit{Edinb. Med. J.} \textbf{37}, 724–735 (1892).
2. Kasai, M. A new operation for “non-correctable” biliary atresia–Portoenterostomy. \textit{Shitijii} \textbf{13}, 733–739 (1959).
3. Davenport, M. \textit{et al.} The outcome of the older (≥ 100 days) infant with biliary atresia. \textit{J. Pediatr. Surg.} \textbf{39}, 575–581. https://doi.org/10.1016/j.jpedsurg.2003.12.014 (2004).
4. Liu, M. B. \textit{et al.} Biliary atresia in Vietnam: Management and the burden of disease. \textit{Surgery} \textbf{161}, 533–537. https://doi.org/10.1016/j.surg.2016.08.012 (2017).
5. Feldman, A. G. & Mack, C. L. Biliary atresia: Clinical lessons learned. \textit{J. Pediatr. Gastroenterol. Nutr.} \textbf{61}, 167–175. https://doi.org/10.1097/MPG.0000000000001075 (2015).
6. Chardot, C. \textit{et al.} Improving outcomes of biliary atresia: French national series 1986–2009. \textit{J. Hepatol.} \textbf{58}, 1209–1217. https://doi.org/10.1016/j.jhep.2013.01.040 (2013).
7. Mezina, A. & Karpen, S. J. Genetic contributors and modifiers of biliary atresia. \textit{Dig. Dis.} \textbf{33}, 408–414. https://doi.org/10.1159/000371694 (2015).
8. Fallon, S. C., Chang, S., Finegold, M. I., Karpen, S. J. & Brandt, M. L. Discordant presentation of biliary atresia in premature monozygotic twins. \textit{J. Pediatr. Gastroenterol. Nutr.} \textbf{57}, e22–e23. https://doi.org/10.1097/MGP.0b013e31826a1044 (2013).
9. Ando, K. \textit{et al.} Sibling occurrence of biliary atresia and biliary dilatation. \textit{J. Pediatr. Surg.} \textbf{31}, 1302–1304. https://doi.org/10.1016/s0022-3960(96)90259-6 (1996).
10. Gunasekaran, T. S., Hassall, E. G., Steinbrecher, U. P. & Yong, S. L. Recurrence of extrahepatic biliary atresia in two half sibs. \textit{Am. J. Med. Genet.} \textbf{43}, 592–594. https://doi.org/10.1002/ajmg.1320430317 (1992).
11. Lachaux, A. \textit{et al.} Familial extrahepatic biliary atresia. \textit{J. Pediatr. Gastroenterol. Nutr.} \textbf{7}, 280–283 (1988).
12. Smith, B. M., Laberge, J. M., Schreiber, R., Weber, A. M. & Blanchard, H. Familial biliary atresia in three siblings including twins. \textit{J. Pediatr. Surg.} \textbf{26}, 1331–1333. https://doi.org/10.1016/0022-3468(91)90613-X (1991).
13. Muraji, T. \textit{et al.} Maternal microchimerism in underlying pathogenesis of biliary atresia: quantification and phenotypes of maternal cells in the liver. \textit{Pediatrias} \textbf{121}, 517–521. https://doi.org/10.1542/peds.2007-0568 (2008).
14. Hayashida, M. \textit{et al.} The evidence of maternal microchimerism in biliary atresia using fluorescent in situ hybridization. \textit{J. Pediatr. Surg.} \textbf{42}, 2097–2101. https://doi.org/10.1016/j.jpedsurg.2007.08.039 (2007).
15. David-Spradl, A., Baussan, C., Hermezui, B., Bernard, O. & Jacquemin, E. CFTR gene involvement in biliary atresia with polysplenia syndrome. \textit{J. Pediatr. Gastroenterol. Nutr.} \textbf{46}, 111–112. https://doi.org/10.1097/mpg.0b013e3180465607884 (2008).
16. Coffinier, C. \textit{et al.} Bile system morphogenesis defects and liver dysfunction upon targeted deletion of HNF1beta. \textit{Development (Cambridge, England)} \textbf{129}, 1829–1838. https://doi.org/10.1242/dev.129.1829 (2002).
17. Koh, H. \textit{et al.} Mitochondrial mutations in cholestatic liver disease with biliary atresia. \textit{Sci. Rep.} \textbf{8}, 905. https://doi.org/10.1038/s41598-018-19558-8 (2018).
18. Mezina, A. \textit{et al.} 845 whole exome sequencing identifies ABCB4 gene variants as modifiers of biliary atresia outcomes. \textit{Gastroenterol.} \textbf{146}, S96. https://doi.org/10.1016/s0016-5085(14)33773-4 (2014).
19. Sangkhathat, S., Laochareonsuk, W., Maneeyach, W., Kayasut, K. & Chiangkiwrate, P. Variants associated with infantile cholestasis. \textit{Pediatr. Rep.} \textbf{11}, 8206. https://doi.org/10.4811/pr.2019.8206 (2019).
20. Bezzera, J. A. \textit{et al.} Biliary atresia: Clinical and research challenges for the twenty-first century. \textit{Hepatology} \textbf{68}, 1163–1173. https://doi.org/10.1002/hep.29905 (2018).
21. Wildhaber, B. E. Biliary atresia: 50 years after the first Kasai. \textit{ISRN Surg.} \textbf{2012}, 132089. https://doi.org/10.5402/2012/132089 (2012).
22. Hsiao, C. H. \textit{et al.} Universal screening for biliary atresia using an infant stool color card in Taiwan. \textit{Hepatology} \textbf{47}, 1233–1240. https://doi.org/10.1002/hep.22182 (2008).
64. Cheng, G. et al. Common genetic variants regulating ADD3 gene expression alter biliary atresia risk. J. Hepatol. 59, 1285–1291. https://doi.org/10.1016/j.jhep.2013.07.021 (2013).

65. Tsai, E. A. et al. Replication of a GWAS signal in a Caucasian population implicates ADD3 in susceptibility to biliary atresia. Hum. Genet. 133, 235–243. https://doi.org/10.1007/s00439-013-1368-2 (2014).

66. Garcia-Barcelo, M. M. et al. Genome-wide association study identifies a susceptibility locus for biliary atresia on 10q24.2. Hum. Mol. Genet. 19, 2917–2925. https://doi.org/10.1093/hmg/ddq196 (2010).

Author contributions
Conception of the study: K.T.T. and L.T.N. Patient recruitment and assessment: K.T.T.; H.K.N., L.T.N.; M.D.N.; H.T.N.; Q.A.T.; A.K.M. Genetic testing: K.T.T.; L.T.M.D.; H.K.N.; Bioinformatics analysis and data interpretation: V.S.L. and K.T.T. Manuscript preparation: all authors.

Funding
This study was funded by the Vinmec Healthcare System (Grant# ISC.18.13).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-01148-y.

Correspondence and requests for materials should be addressed to K.T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021