Novel Findings From Family-based Exome Sequencing for Children With Biliary Atresia

Kien Tran (trantrungkien80@gmail.com)
Vinmec Research Institute of Stem Cell and Gene Technology
Vinh Le
University of Engineering and Technology, Vietnam National University Hanoi
Lan Dao
Vinmec Research Institute of Stem Cell and Gene Technology
Huyen Nguyen
Bioequivalence Center, National Institute of Drug Quality Control
Anh Mai
Vinmec International Hospital
Ha Nguyen
Vinmec International Hospital
Minh Ngo
Vinmec International Hospital
Quynh Tran
Vietnam National Children's Hospital
Liem Nguyen
Vinmec Research Institute of Stem Cell and Gene Technology

Research Article

Keywords: Biliary atresia, cholestatic jaundice, exome sequencing, Vietnamese

Posted Date: June 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-561902/v1

Abstract

Biliary atresia (BA) is a progressive inflammation and fibrosis of the biliary tree, characterized by the obstruction of bile flow led to liver failure, scarring and cirrhosis. This study aimed to explore the elusive etiology of BA by conducting whole exome sequencing (WES) for 41 children with BA and their parents (35 trios, including one family with two BA diagnosed children and five child-mother cases). We exclusively identified and validated a total of 28 variants (17 X-linked, six *de novo* and five homozygous) in 25 candidate genes from our BA cohort. These variants were among the 10% most deleterious and having a low minor allele frequency against three employed databases: Kinh Vietnamese (KHV), gnomad and 1000 Genome project. Interestingly, *AMER1*, *INVS* and *OCRL* variants were repeatedly found in unrelated probands, and were firstly reported in a BA cohort. Liver specimens and blood samples showed identical variants, suggesting that somatic mutations were unlikely to occur during the morphogenesis. In agreement with earlier attempts, this study implicated a genetical heterogeneity and non-Mendelian inheritance of BA.

Introduction

Biliary atresia (BA) is a progressive inflammation and fibrosis of the biliary tree, characterized by the obstruction of bile flow resulted in a development of the cholestatic liver disease. It was first described by surgeon John Thomson in 1892[1], and thus far is among the most fatal diseases leading to severe complications in infants. BA occurs in early stage of neonates and can be treated by hepatic portoenterostomy or Kasai procedure[2]. After the surgical treatment, however, about 50% of affected infants require a liver transplantation, while the rests would sustain their own liver up to the age of five to 10 years[3]. A study on the Vietnamese BA patients reported that 84% and 71% of Kasai treated patients would survive after one and two years, respectively. Meanwhile, respective ratios are 52% and 28% for the group without Kasai treatment[4]. It is estimated that after hepatic portoenterostomy surgery, from 70-80% of patients with BA still require liver transplantation by their adulthood due to a progressive development of liver scarring, failure and cirrhosis[5,6].

Despite of an early discovery, extensive studies and treatment exercises, the etiology and pathogenesis of BA remain elusive. There are several hypotheses, including viral infection, autoimmune mediated bile duct destruction, biliary toxin, and genetic abnormality[7]. Regarding genetic aspect, debate over Mendelian mechanism of the disease has been raised due to a lack of familial BA, discordant presentation of BA among twins, including monozygotic twin[8]. Nevertheless, some cases of familial BA have been reported, suggesting that either a recessive autosomal inheritance or a combination of genetic and acquired factors contributes to the disease's etiology[9-12]. Besides, some studies have inferred that BA is related to microchimerism, where the genetic trait is maternally transferred from mother contributing to the phenotypic heterogeneity and non-Mendelian inheritance[13,14]. More specifically, a heterozygous transition of c.433G>A in five BA patient with polysplenia syndrome implicating the genetic predisposition to BA splenic malformation[15]. In the mouse model, inactivation of the hepatocyte nuclear factor 1 beta gene (Hnf1b) causes abnormalities of the gallbladder and intrahepatic bile ducts resulted in an exhibition of severe jaundice[16]. Observations of an increased incidence of BA in some areas and groups, such as in the Asian and Polynesian suggest that genetic and environmental factors might cause the disease. Recent genetic studies have revealed a linkage between cholestatic jaundice and some mutations in both nuclear DNA and mitochondrial DNA[17-20].

The prevalence of BA is 1 in 8,000 to 18,000 live births and varies from countries and groups with a dominance of females over males[21]. The disease occurs more frequently in Southeast Asia and the Ocean Pacific[22]. The prevalence of BA in Taiwan is around 1:5,000 live births in comparison to that of 1:14,000-20,000 in North America and Western Europe[6,23,24]. To our knowledge there is unavailable epidemiology study of BA in the Vietnamese, so is the prevalence of this fatal disease. The prevalence of BA in Vietnam is estimated as high as 1:2,400 live births as equal to that of the Ocean Pacific regions[22]. Although BA and BA related liver disease are rather common disorders in Vietnamese infants and is life threatening disease, few studies have been previously reported[4,25,26]. To date, Kasai portoenterostomy procedure has been introduced as a routine surgical practices and it apparently offers a better opportunity to the patients[25]. However, there is a big portion of patients still require a liver transplantation or having a low quality of life due to the disease's complications after surgical treatment. Recently, next generation sequencing (NGS), particularly exome sequencing (WES) has been increasingly applied for detecting mutations in patients with cholestasis[27]. It appears as a powerful tool to aid the diagnosis, 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 diagnosed in hope of exploring new and characterized causative variants, which would shed a light to underline the etiology of the deadly disease.

Materials And Methods

Patient recruitment

Patients with BA firmly diagnosed and their parents were recruited at Vinmec International Hospital and the 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 and in accordance with the Declaration of Helsinki.

Sample collection and DNA extraction

About 2mL of peripheral blood of patients and their parents were collected in a EDTA anticoagulant tube and stored at -80^{o}C. Liver wedge specimens were collected from the Kasai surgery and were snapped frozen in liquid nitrogen and stored at -80^{o}C. Genomic DNA were extracted by using DNA Mini Blood Isolation Kit based on the manufacture's protocol (Qiagen, Germany). DNA samples were quantified by fluorescence method using a Qubit BR Quantification Kit (Invitrogen, USA). Extracted DNA samples were preserved at -80^{o}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 manufactures' protocol with slight modifications. The library concentration was quantified by a Qubit dsDNA Broad Range Assay Kit (Invitrogen, USA). Library's size was measured by using a Lab chip 3K Hisense Kit (Perkin Elmer, USA). A paired-end exome sequencing with 150bp 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 the highly regarded tools as mentioned previous study[28]. 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[29]. Common tools, including Bowtie2, BWA and Qualimap were used for quality control[30]. To minimize the false positive rate, multiple variant calling tools, including Genome Analysis Toolkit (GATK)[31], SAMtools mpileup[32], Freebayes[33] 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 non-synonymous variants with a Minor Allele Frequency (MAF) <1% against three database: the Kinh Vietnamese population (KHV) obtained from our previous study on the Vietnamese human genome database[34], gnomad (https://gnomad.broadinstitute.org/) and 1000 Genome Project[29]; (*ii*) Inclusion of variants with three types of inheritance mode: X-linked, homozygous and putative *de novo*; (*iii*) Variants with CADD Phred score of >10, indicating the 10% most deleterious variants in the genome[35]. *In silico* prediction tools, including SnpEff [36], PolyPhen-2[37], Mutation Taster[38] were employed to predict the impact of genetic changes. Molecular Signatures Databases (MSigDB v7.2) was used to compute the candidate genes with the gene sets of Human phenotype ontology[39,40].

Validation of WES results

Selected variants were then confirmed by bi-directional 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 (Thermofisher, Massachusetts, USA).

Results

Clinical features

We have recruited a total of 42 children with firmly diagnosed BA and their parents. All patients showed typical BA symptoms, such as prolonged jaundice, acholic stool and abnormality of biliary tract at early infantile and have been under Kasai operation either before or when enrolling in this study. The patients, including 23 males and 19 females born from 2009-2019, but the majority was in recent years. Despite all patients have undergone Kasai operation right after birth (mostly after their 2 months of life), bilirubin levels as well as serum enzymes indicating liver function such as ALP, ALT, AST and γ-GT remained high at the time of enrolment (**Table 1**). Some patients developed liver cirrhosis. Only one BA patient had CMV infection (BA037). Regarding the family history, one family had two children with BA (BA002); primary sclerosing cholangitis (BA032); hemophilia (BA025); and choledochal cyst (BA042). Four mothers experienced abnormal pregnancy (BA024, BA027; BA036, BA038). The rests did not show any significant concern during their pregnancy and having no family history of BA or other genetic diseases. Except a patient from a family, who did not come for blood drawling after our first health check, we were able to collect blood samples of 41 BA affected children and their parents. Among these 41 children, we could collect the liver specimens of 18 children obtained from Kasai operation.

Genetic properties

We applied a strict filtering strategy by removing variants with MAF>1%, synonymous variants and variants with CADD scaled score <10. Finally, we identified a total of 28 variants in 25 genes from our BA affected cohort (**Table 2**). All the variants were subsequently confirmed by Sanger sequencing (**Fig.S1**). Among 28 detected variants, 17 were X-linked (accounting for 61%), 6 were *de novo* (21%) and 5 were homozygous (18%) **(Fig.1)**. Homozygous variants were found in five genes, including *HACE1*, *VPS13C*, *RAPGEF4*, *FOCAD* and *INVS* (**Table 2**). We found 6 *de novo* variants in 6 respective genes from 5 probands, including *INVS* (proband #BA014), *ELP2* (proband#BA028), *TINAG* (proband#BA033), *CEP63* and *CCDC136* (proband#BA035), *BCAR1* (proband#BA036). The family #2 involved two siblings with similar phenotypes (early onset jaundice, BA diagnosed), we did not detect any variant from the female patient, but two X-linked and one homozygous variants from her younger brother (**Table 2**).

Interestingly, several unrelated probands carried variants occurred in the same genes, including *AMER1* (proband #BA004 and BA007), *INVS* (proband #BA014 and BA041), and *OCRL* (proband #BA032 and BA041). Noticeably, proband BA014 carried a *de novo* variant in the *INVS* gene, while proband BA041 carried a *INVS* homozygous variant. This highly suggested the causative role of *INVS* in the presentation of BA (**Table 2**).

In addition to blood samples, we were able to collect 18 liver specimens from our BA cohort. Among that blood and liver samples from 8 children shared identical variants (BA009, BA016, BA032, BA036, BA037, BA038, BA040 and BA041). Meanwhile, we did not detect any significant variants based on our rationales for variant classification (**Table 3**). In other word, this study did not detect any somatic variants from liver samples. The findings suggested that the occurrence of somatic mutations during the liver cell differentiation as previously hypothesized[41] was unlikely.

Effect of genetic predisposition

The detected variants showed their extremely low MAF against three employed databases: Kinh Vietnamese (KHV), gnomad and 1000 Genome Project (**Table 4**). We noticed only two variants in the *HACE1* and *VPS13C* gene had an MAF of 1% against the KHV database. All variants had their CADD Phred scaled score

above 10, and mostly above 20, indicating their either the 10% or 1% most deleterious substitutions. Among that, INVS:c.C208T was the most deleterious with the highest scaled score of 37, and Mutation Taster predicted this variant as disease causing (**Table 4**).

We used I-mutant tool 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.T2603A(p.Met876Lys), which increased the stability of the mutant compared to that of the wild-type variants, all variants showed to decrease their stability (**Table S1**). HOPE tool was used to predict the structural effect of missense mutations showing the changes in residue size, hydrophobic and structural stability (**Fig.S2**). Change in amino acid size and charge resulted in a loss of interaction and disturbance of the protein function. Several variants, for example HACE1:c.G1660A(pAla651Thr), PHKA1: c.G478A(p.Asp160Asn), whose wild-type residues located in the important domains, thus any substitution in these regions were 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 OCR:c.T2603A(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 & 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 felt into various human phenotype gene sets, ranging from gonosomal inheritance, X-linked recessive inheritance to involuntary movements (**Table 5**). We also computed our gene set to find the association of these genes to reported phenotypes available from HPO and Monarch Initiative (**Table S2**). However, we did not find any of that from these databases. The reason might be a lack of associated genes/pathway to BA's phenotype in these available database, which are often dominated by researches on the Caucasian where BA's prevalence is rare and not well studied. By apply the same strategy to identify the potential contribution of ciliary dysgenesis underlying BA phenotype, we used a gene set containing 2,016 genes of interest, which is derived from four large datasets[42]. We found that *BCOR*, *INVS* and *OCRL* were included in this gene set. This results suggested the novelty of *BCOR*, *INVS* and *OCRL* from our BA affected cohort.

Similar to the previous study[42], we did not identify any proposed associated genes for 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* gene (biliary atresia splenic malformation syndrome-BASM). We also did not find any significant variant in either susceptibility locus of *ADD3*, *XPNPEP1*, *GPC1*, *ARF6* and *EFEMP1*, which are suggested from several GWAS studies or associated SNPs in BA patients[43].

Discussion

Like other previous studies, we attempted to reveal the genetic pattern of BA disorder by conducting a trio-based exome sequencing for 40 families, involving 41 children with confirmed BA diagnosed. Going beyond this establishment in genetic study for such a rare and complex disorder, we further tested our hypothesis whether the detected variants were 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 find out either the rarest variants or putative *de novo* events from our BA cohort, that 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 occurred in 25 respective genes and allowed us to rule out the occurrence of somatic variants in the development of the disease as we could not find any differences in terms of genetics between blood and liver specimens.

In agreement with previous studies, our results showed an intriguing genetic aspect of BA, which was highly heterogenous. It is worthy to note that along with other variants detected from unrelated individuals, this study found three genes whose variants occurred in the unrelated probands, including *AMER1*, *INVS* and *OCRL*. While the etiology of BA remained unclear and was unlikely to follow the Mendelian model, our results implicated their role in the disease's development. Overlapped findings of *BCOR*, *INVS* and *OCRL* gene in our BA cohort with a large comprehensive ciliopathy and biliary genes of interest in the previous study[42] further supported the possibility of the causative role of these genes to BA. To that, *AMER1* (MIM#300647) encodes the APC membrane recruitment protein 1 acting as an inhibitor of the canonical Wnt/beta-catenin signaling pathway[44], which controls the hepatobiliary development during the embryogenesis. In the mature healthy liver, it is mostly inactive and the abnormal Wnt/beta-catenin signaling pathway can promote the development of liver diseases[45]. *AMER1* has been found to be associated with osteopathia striata with cranial sclerosis[46] or its debating role in Wilms tumor development[47-49], activation of the Wnt/beta-catenin signaling pathway drives to hapatocarcinoma and cholangiocarcinoma[50]. Albeit a lack of reported association of *AMER1* to typical BA phenotypes, we here inferred its indirect role in the development of cholestatic jaundice, including BA as a result of the activation of the Wnt/beta-catenin signaling pathway.

We here suggested *INVS* as a BA candidate gene owing to *INVS* variants detected in the unrelated probands and they were predicted as highly damaging. This gene encodes inversin protein, which plays a role in the primary cilia function and involves in the cell cycle. Intriguingly, inactivation of *INVS* in mouse model shows a significant increase in bilirubin level compared to that of the wild-type, and the pathogenic changes in ductal plate malformation in the intrahepatic biliary of the mutant mouse[51]. However, since then the association of *INVS* with BA had not been firmly established due to an absence of *INVS* mutations in patients with BA[52,53]. In fact, *INVS* mutations have been observed in patients with infantile nephronophthisis type 2 regardless of populations[54-56]. In our study, we detected a *INVS* heterozygous *de novo* and a homozygous variant from unrelated patients with BA (BA014 and BA041). To our knowledge, this novelty would be firstly reported in BA patients and further studies are needed to taking *INVS* into account. Similar to the *BCOR* and *INVS*, *OCRL* encodes inositol polyphosphate 5-phosphatase that might involve in the primary cilia assembly. *OCRL* mutations have been widely reported to link 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[57-59]. Due to a lack of *OCRL* in relation to BA or BA related diseases, we suggested this gap for future investigation.

As a result of an emerging DNA sequencing technology and its rapid decline cost in recent years, dozens of rare and previously undiagnosed genetic disorders are currently detectable. Numbers of causative variants have been increased over years that tremendously help supporting and explaining the pathogenesis of previously undetermined genetic disorders. For the last 10 years, next generation sequencing (NGS) technology has revolutionized our understanding on human genetics with a high level of accuracy, cost effectiveness and high throughput capability. NGS is firmly becoming a standard for routine diagnostic practices[60]. In BA studies, mitochondrial DNA mutations (mtDNA) have been found to associate with BA, suggesting the role of mitochondrial in underling the pathogenic mechanism[17]. Whole exome sequencing (WES) has revealed dozen candidate genes either encode the ATP-binding cassette transporters (the ABC superfamily)[18,19] or involve in the Notch signaling pathway, such as *JAG1*[19,61] and *NOTHC2*[20]. Genome wide association study (GWAS) highlighted a strong association between BA and some variants in the *ADD3* gene located on the 10q24.2[62]. Other following study on 171 BA patients and 1,630 controls of European descent found a strongest signal at rs7099604 in the *ADD3* gene[63]. A significant association was found between variant rs17095355 on the *XPNPEP1* gene located on the 10q24 with the disease[64]. Taken together, the etiology of BA remains challenging due to an involvement of multiple genes and complex mechanisms. Being encouraged from the pioneers, we provided a concrete genetic aspect obtained from a Vietnamese BA cohort by applying a stringent participant selection, trio-based approach, variant classification and confirmed validation. The findings would add up to our knowledge of the genetic heterogeneity and complexity of BA disorder.

Conclusions

The etiology of BA remains challenging as there is a lack of conclusive evidences despite of extensive researches and medical practices for hundred years. However, recent development of NGS technology and its application in BA and BA related diseases research have gradually revealed the hidden genetic picture of BA's etiology where dozens of genes have been found to be associated with the disease. Our study using family-based exome sequencing with an involvement of 40 families having BA diagnosed children identified 28 variants in 25 genes (all validated). These variants were in the 10% most deleterious and were either rare or extreme rare in the population genome database. A combination of functional prediction and analysis of biological process enabled us to suggest these candidate genes to the development of BA, particularly with those detected in unrelated BA individuals, including *AMER1*, *INVS* and *OCRL*. Similar variants detected from blood and liver wedge specimens from the same BA affected individuals suggested that somatic mutations in liver cells were unlikely to occur during the morphogenesis. Taken together, we highlighted the genetic heterogeneity of BA and ruled out the Mendelian model, and encouraged future studies to further address the roles of these genes in the development of BA.

Declarations

Funding

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

Competing interests

The authors declare there is no competition of interest. We deeply thank to the BA affected families for their participation in this study.

Availability of data and material

Data were available from this manuscript and supplementary information.

Authors' 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 & K.T.T. Manuscript preparation: all authors.

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 were consent for publication of all relevant data and this manuscript.

References

1 Thomson, J. On Congenital Obliteration of the Bile-Ducts. *Edinb Med J***37**, 724-735 (1892).

2 Kasai M & SA, S. A new operation for "non-correctable" biliary atresia-Portoenterostomy. *Shijitsu***13:733–739** (1959).

3 Davenport, M. *et al.* The outcome of the older (≥100 days) infant with biliary atresia. *J Pediatr Surg***39**, 575-581, doi:10.1016/j.jpedsurg.2003.12.014 (2004).

4 Liu, M. B. *et al.* Biliary atresia in Vietnam: Management and the burden of disease. *Surgery***161**, 533-537, doi:10.1016/j.surg.2016.08.012 (2017).

5 Feldman, A. G. & Mack, C. L. Biliary Atresia: Clinical Lessons Learned. *J Pediatr Gastroenterol Nutr* **61**, 167-175, doi:10.1097/MPG.0000000000000755 (2015).

6 Chardot, C. *et al.* Improving outcomes of biliary atresia: French national series 1986-2009. *J Hepatol* **58**, 1209-1217, doi:10.1016/j.jhep.2013.01.040 (2013).

7 Mezina, A. & Karpen, S. J. Genetic contributors and modifiers of biliary atresia. *Dig Dis* **33**, 408-414, doi:10.1159/000371694 (2015).

8 Fallon, S. C., Chang, S., Finegold, M. J., Karpen, S. J. & Brandt, M. L. Discordant presentation of biliary atresia in premature monozygotic twins. *J Pediatr Gastroenterol Nutr* **57**, e22-23, doi:10.1097/MPG.0b013e31826a1044 (2013).

9 Ando, K. *et al.* Sibling occurrence of biliary atresia and biliary dilatation. *J Pediatr Surg* **31**, 1302-1304, doi:https://doi.org/10.1016/S0022-3468(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. *American journal of medical genetics* **43**, 592-594, doi:10.1002/ajmg.1320430317 (1992).

11 Lachaux, A. *et al.* Familial extrahepatic biliary atresia. *J Pediatr Gastroenterol Nutr* **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. *J Pediatr Surg* **26**, 1331-1333, doi:https://doi.org/10.1016/0022-3468(91)90613-X (1991).

13 Muraji, T. *et al.* Maternal microchimerism in underlying pathogenesis of biliary atresia: quantification and phenotypes of maternal cells in the liver. *Pediatrics* **121**, 517-521, doi:10.1542/peds.2007-0568 (2008).

14 Hayashida, M. *et al.* The evidence of maternal microchimerism in biliary atresia using fluorescent in situ hybridization. *J Pediatr Surg* **42**, 2097-2101, doi:10.1016/j.jpedsurg.2007.08.039 (2007).

15 Davit-Spraul, A., Baussan, C., Hermeziu, B., Bernard, O. & Jacquemin, E. CFC1 Gene Involvement in Biliary Atresia With Polysplenia Syndrome. *J Pediatr Gastroenterol Nutr* **46**, 111-112, doi:10.1097/01.mpg.0000304465.60788.f4 (2008).

16 Coffinier, C. *et al.* Bile system morphogenesis defects and liver dysfunction upon targeted deletion of HNF1beta. *Development (Cambridge, England)* **129**, 1829-1838, doi:https://doi.org/10.1242/dev.129.8.1829 (2002).

17 Koh, H. *et al.* Mitochondrial Mutations in Cholestatic Liver Disease with Biliary Atresia. *Sci Rep* **8**, 905, doi:10.1038/s41598-017-18958-8 (2018).

18 Mezina, A. *et al.* 845 Whole Exome Sequencing Identifies ABCB4 Gene Variants As Modifiers of Biliary Atresia Outcomes. *Gastroenterology* **146**, S-928, doi:10.1016/s0016-5085(14)63373-4 (2014).

19 Sangkhathat, S., Laochareonsuk, 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. *Journal of pediatric genetics* **7**, 67-73, doi:10.1055/s-0038-1632395 (2018).

20 Shaul, E. *et al.* Novel mutations in NOTCH2 gene in infants with neonatal cholestasis. *Pediatr Rep* **11**, 8206, doi:10.4081/pr.2019.8206 (2019).

21 Bezerra, J. A. *et al.* Biliary Atresia: Clinical and Research Challenges for the Twenty-First Century. *Hepatology* **68**, 1163-1173, doi:10.1002/hep.29905 (2018).

22 Wildhaber, B. E. Biliary atresia: 50 years after the first kasai. *ISRN Surg* **2012**, 132089, doi:10.5402/2012/132089 (2012).

23 Hsiao, C. H. *et al.* Universal screening for biliary atresia using an infant stool color card in Taiwan. *Hepatology* **47**, 1233-1240, doi:10.1002/hep.22182 (2008).

24 Fischler, B., Haglund, B. & Hjern, A. A population-based study on the incidence and possible pre- and perinatal etiologic risk factors of biliary atresia. *J Pediatr* **141**, 217-222, doi:10.1067/mpd.2002.126001 (2002).

25 Liem, N. T., Son, T. N., Quynh, T. A. & Hoa, N. P. Early outcomes of laparoscopic surgery for biliary atresia. *J Pediatr Surg* **45**, 1665-1667, doi:10.1016/j.jpedsurg.2010.01.019 (2010).

26 Luan, P. C., An, P. L., Phong, N. H. & Ngoc, N. M. Characteristics of infants with cholestatic jaundice at gastrointestinal department, children's hospital 2. *Nghiên cứu Y học* **18**, 402-407 (2014).

27 Sambrotta, M. *et al.* Mutations in TJP2 cause progressive cholestatic liver disease. *Nat Genet* **46**, 326-328, doi:10.1038/ng.2918 (2014).

28 Tran, K. T. *et al.* Genetic landscape of autism spectrum disorder in Vietnamese children. *Sci Rep* **10**, 5034, doi:10.1038/s41598-020-61695-8 (2020).

29 Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

30 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997v2 [q-bio.GN]* **1303** (2013).

31 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res***20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

32 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics***25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

33 Garrison, E. & Marth, G. *Haplotype-based variant detection from short-read sequencing*. Vol. 1207 (2012).

34 Le, V. S. *et al.* A Vietnamese human genetic variation database. *Hum Mutat***40**, 1664-1675, doi:10.1002/humu.23835 (2019).

35 Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res***47**, D886-D894, doi:10.1093/nar/gky1016 (2019).

36 Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nature protocols***11**, 1-9, doi:10.1038/nprot.2015.123 (2016).

37 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods***7**, 248-249, doi:10.1038/nmeth0410-248 (2010).

38 Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods***11**, 361, doi:10.1038/nmeth.2890 (2014).

39 Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA***102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

40 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics***27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).

41 Fabre, A., Roman, C. & Roquelaure, B. Somatic mutation, a cause of biliary atresia: A hypothesis. *Med Hypotheses***102**, 91-93, doi:10.1016/j.mehy.2017.03.015 (2017).

42 Berauer, J. P. *et al.* Identification of Polycystic Kidney Disease 1 Like 1 Gene Variants in Children With Biliary Atresia Splenic Malformation Syndrome. *Hepatology***70**, 899-910, doi:10.1002/hep.30515 (2019).

43 Girard, M. & Panasyuk, G. Genetics in biliary atresia. *Curr Opin Gastroenterol***35**, 73-81, doi:10.1097/MOG.0000000000000509 (2019).

44 Tanneberger, K. *et al.* Structural and functional characterization of the Wnt inhibitor APC membrane recruitment 1 (Amer1). *The Journal of biological chemistry***286**, 19204-19214, doi:10.1074/jbc.M111.224881 (2011).

45 Perugorria, M. J. *et al.* Wnt-beta-catenin signalling in liver development, health and disease. *Nat Rev Gastroenterol Hepatol***16**, 121-136, doi:10.1038/s41575-018-0075-9 (2019).

46 Holman, S. K. *et al.* Osteopathia striata congenita with cranial sclerosis and intellectual disability due to contiguous gene deletions involving the WTX locus. *Clin Genet***83**, 251-256, doi:10.1111/j.1399-0004.2012.01905.x (2013).

47 Rivera, M. N. *et al.* An X Chromosome Gene, WTX, is Commonly Inactivated in Wilms Tumor. *Science***315**, 642, doi:10.1126/science.1137509 (2007).

48 Ruteshouser, E. C., Robinson, S. M. & Huff, V. Wilms tumor genetics: mutations in WT1, WTX, and CTNNB1 account for only about one-third of tumors. *Genes Chromosomes Cancer***47**, 461-470, doi:10.1002/gcc.20553 (2008).

49 Akhavanfard, S. *et al.* Inactivation of the tumor suppressor WTX in a subset of pediatric tumors. *Genes Chromosomes Cancer***53**, 67-77, doi:10.1002/gcc.22118 (2014).

50 Carotenuto, P. *et al.* Wnt signalling modulates transcribed-ultraconserved regions in hepatobiliary cancers. *Gut***66**, 1268-1277, doi:10.1136/gutjnl-2016-312278 (2017).

51 Shimadera, S. *et al.* The inv mouse as an experimental model of biliary atresia. *J Pediatr Surg***42**, 1555-1560, doi:10.1016/j.jpedsurg.2007.04.018 (2007).

52 Rajagopalan, R. *et al.* Exome Sequencing in Individuals with Isolated Biliary Atresia. *Sci Rep***10**, 2709, doi:10.1038/s41598-020-59379-4 (2020).

53 Schon, P. *et al.* Identification, genomic organization, chromosomal mapping and mutation analysis of the human INV gene, the ortholog of a murine gene implicated in left-right axis development and biliary atresia. *Human genetics***110**, 157-165, doi:10.1007/s00439-001-0655-5 (2002).

54 Halbritter, J. *et al.* Identification of 99 novel mutations in a worldwide cohort of 1,056 patients with a nephronophthisis-related ciliopathy. *Human genetics***132**, 865-884, doi:10.1007/s00439-013-1297-0 (2013).

55 Kang, H. G. *et al.* Targeted exome sequencing resolves allelic and the genetic heterogeneity in the genetic diagnosis of nephronophthisis-related ciliopathy. *Exp Mol Med***48**, e251, doi:10.1038/emm.2016.63 (2016).

56 Schueler, M. *et al.* Large-scale targeted sequencing comparison highlights extreme genetic heterogeneity in nephronophthisis-related ciliopathies. *J Med Genet***53**, 208-214, doi:10.1136/jmedgenet-2015-103304 (2016).

57 Coon, B. G. *et al.* The Lowe syndrome protein OCRL1 is involved in primary cilia assembly. *Hum Mol Genet***21**, 1835-1847, doi:10.1093/hmg/ddr615 (2012).

58 Luo, N. *et al.* OCRL localizes to the primary cilium: a new role for cilia in Lowe syndrome. *Hum Mol Genet***21**, 3333-3344, doi:10.1093/hmg/dds163 (2012).

59 Zhang, X., Jefferson, A. B., Auethavekiat, V. & Majerus, P. W. The protein deficient in Lowe syndrome is a phosphatidylinositol-4,5-bisphosphate 5-phosphatase. *Proc Natl Acad Sci U S A***92**, 4853-4856, doi:10.1073/pnas.92.11.4853 (1995).

60 Nicastro, E. & D'Antiga, L. Next generation sequencing in pediatric hepatology and liver transplantation. *Liver Transpl***24**, 282-293, doi:10.1002/lt.24964 (2018).

61 Fischetto, R. *et al.* Alagille Syndrome: A Novel Mutation in JAG1 Gene. *Front Pediatr***7**, 199, doi:10.3389/fped.2019.00199 (2019).

62 Cheng, G. *et al.* Common genetic variants regulating ADD3 gene expression alter biliary atresia risk. *J Hepatol***59**, 1285-1291, doi:10.1016/j.jhep.2013.07.021 (2013).

63 Tsai, E. A. *et al.* Replication of a GWAS signal in a Caucasian population implicates ADD3 in susceptibility to biliary atresia. *Human genetics***133**, 235-243, doi:10.1007/s00439-013-1368-2 (2014).

64 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, doi:10.1093/hmg/ddq196 (2010).

Tables

Table 1. Clinical features of children with biliary atresia.

Proband	Birth Year	Sex	Age Diagnosed	Blood test (at the time of enrollment)							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)	
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.
BA013	2010	F	1 m/o	249	40	38.2	52.6	57.3	21.8	4.3	1st child; No family history of BA or other genetic disease; Cirrhosis developed.
BA014	2016	F	2 m/o	N/A	42.2	109.5	87.8	201.7	8.9	2.2	1st child; C-section delivered with birthweight of 3.4 Kg; No family history of BA or other genetic disease.
BA015	2016	F	1 m/o	N/A	39.5	167.8	115.8	460	80.3	46.4	2nd child; No family history of BA or other genetic disease.
BA016	2014	M	3 m/o	353.8	37.09	110.4	196.1	424.1	16.8	5.7	1st child; No family history of BA or other genetic disease.
BA017	2016	F	2 m/o	516	34.6	161.5	282.7	224.6	56.8	26.8	2nd child; No family history of BA or other genetic disease.
BA018	2015	M	2 m/o		33.5	82.9	191.2	371	180.8	107.9	2nd child; No family history of BA or other genetic disease; Prolonged jaundice, acholic stool; cirrhosis after Kasai operation.
BA019	2017	F	3 m/o	1195	29.7	64.9	150.9	384.1	38.8	16.9	1st child; No family history of BA or other genetic disease.
BA020	2009	M	2 m/o	386	39.8	87.4	80.7	176.7	16.7	5.9	1st child; No family history of BA or other genetic disease.
BA021	2018	F	65 days	584.8	37.6	220.9	323.1	918.9	224.1	123	1st child; No family history of BA or other genetic disease.
BA023	2018	M	3 m/o	635.7	36.75	163.9	258.5	404	153.9	85	2nd child; No family history of BA or other genetic disease.
BA024	2017	F	2 m/o	280.3	33.2	63.9	66.9	88	14.7	5	A child from 2nd pregnancy; C-section delivered, full term; 1st pregnancy was a boy, stillbirth at 5 m/o of gestation due to a low level of amniotic fluid. No family history of BA or other genetic disease.

BA025	2018	F	3 m/o	300	41.2	175.8	226.3	465.1	131.1	82.3	3nd child; her older brother was with hemophilia; her older sister was healthy.
BA026	2018	M	2 m/o	498	43.2	178.7	240.5	781	76.7	52	1st child; No family history of BA or other genetic disease.
BA027	2018	M	40 days	497	38.5	78	104.9	565.1	11.3	4.1	He was a child from his mother's 3rd pregnancy; the 1st pregnancy was stillbirth at 7 weeks of gestation due to no heartbeat; the 2nd was a molar pregnancy discovered at 8 weeks of gestation.
BA028	2016	M	28 days	421	36.5	63.7	80.2	171.6	9.2	2.4	1st child of healthy parents. His paternal grandfather developed liver cirrhosis at age of 50.
BA029	2014	M	1 m/o	N/A	39.5	221.6	227.8	527.7	89.9	51.7	1st child; Full term, C-section delivered with birthweight of 3.2 Kg; No family history of BA or other genetic disease. Prolonged jaundice, acholic stool.
BA030	2018	M	1 m/o	404	31.4	123.6	210.2	900.3	208	120.4	2nd child; No family history of BA or other genetic disease.
BA031	2018	M	15 days	556	35.8	56.9	142.6	855.5	143.7	80.3	2nd child; No family history of BA or other genetic disease.
BA032	2018	M	1 m/o	427.4	34.4	76.4	144.6	144.6	150.6	69.6	He was the 2nd child. The first child was diagnosed with primary sclerosing cholangitis and died at 28 m/o.
BA033	2018	F	2.5 m/o	648	32.8	134.7	255.2	131.8	368.8	188	2nd child; No family history of BA or other genetic disease.
BA034	2018	F	29 days		35.2	44.2	150.8	N/A	104.9	60.9	2nd child; No family history of BA or other genetic disease.
BA035	2015	F	72 days	374	33.1	82.5	173.3	70	278.8	142	1st child; No family history of BA or other genetic disease.
BA036	2018	F	1.5 m/o	311	39.7	164.6	265.9	280.4	131.3	97.8	She was a child from her mother's 2nd pregnancy. The first pregnancy was miscarriage.
BA037	2018	M	1 m/o	808.3	38.9	246.7	317.8	329.5	139.8	85.8	1st child; He was infected with CMV. His father was infected with HBV. No family history of BA or other genetic disease.
BA038	2018	M	66 days	240	39.8	113.8	87.2	833	81.4	47.3	Full term, vaginal delivered with birthweight of 3.1 Kg. He was a child from his mother's 3rd pregnancy; The 1st and 2nd pregnancy were stillbirth at 8 weeks of gestation. He had an inguinal hernia.
BA039	2019	F	40 days	618	36.2	115.4	130.7	899.3	130.3	101.2	She was a child of her mother's 2nd pregnancy; the first was aborted. Her maternal grandfather was with hepatitis.
BA040	2019	M	2 m/o	629.2	37.7	93.4	220	604.1	256.7	137	1st child; No family history of BA or other genetic disease.
BA041	2019	M	28 days	320	42.6	138.6	265.5	905.2	108	82.7	1st child; No family history of BA or other genetic disease. His prenatal grandmother was infected with HBV.
BA042	2019	F	1 m/o	N/A	N/A	211	663	N/A	229	120	2nd child; the first child was a healthy boy. Her mother was diagnosed with choledochal cyst at age of 13.

Abbreviation: 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 the siblings.

Table 2. Genetic characteristics of Vietnamese children with biliary atresia.

Proband	Sex (M/F)	Chr	Position	SNP	Ref	Alt	Gene	DNA change	A.A change	MOI	Genotype
											Proband
BA002	M	chr6	104771988	rs199554586	C	T	*HACE1*	NM_001350555:c.G1660A	p.Ala651Thr	AR	T/T
		chrX	72684557	rs201601894	C	T	*PHKA1*	NM_001122670:c.G478A	p.Asp160Asn	X-linked	T
		chrX	123665767	rs145905162	C	T	*THOC2*	NM_001081550:c.G1261A	p.Ala421Thr	X-linked	T
BA004	M	chrX	123888703	rs182340753	C	G	*XIAP*	NM_001167:c.C962G	p.Ala321Gly	X-linked	G
BA007	M	chr15	61929659	rs149882066	G	C	*VPS13C*	NM_017684:c.C5999G	p.Ala2043Gly	AR	C/C
		chrX	64192212	rs771015941	T	A	*AMER1*	NM_152424:c.A1075T	p.Ser359Cys	X-linked	A
		chrX	77508398	rs199543136	G	C	*ATRX*	NM_138270:c.C7318G	p.Pro2478Ala	X-linked	C
		chrX	85367724	rs764400863	T	G	*POF1B*	NM_001307940:c.A325C	p.Ser109Arg	X-linked	G
BA009	M	chrX	130015441	rs201843717	G	A	*BCORL1*	NM_001184772:c.G2669A	p.Arg890Gln	X-linked	A
BA014	F	chr9	100252390	rs773049314	C	T	*INVS*	NM_001318382:c.C208T	p.Arg396*	*De novo*	C/T
		chrX	40073898	rs587778096	G	A	*BCOR*	NM_001123383:c.C1448T	p.Pro483Leu	X-linked	A/A
BA016	M	chrX	56565305	.	C	G	*UBQLN2*	NM_013444:c.C1432G	p.Pro478Ala	X-linked	G
BA020	M	chrX	43693330	.	G	A	*MAOA*	NM_000240:c.G208A	p.V70M	X-linked	A
		chrX	108733510	.	C	G	*IRS4*	NM_003604:c.G2835C	p.Trp945Cys	X-linked	G
BA028	M	chr18	36156524	rs1206216028	C	T	*ELP2*	NM_001242879:c.C1124T	p.Ala445Val	*De novo*	C/T
BA032	M	chr2	173016403	rs773527960	C	A	*RAPGEF4*	NM_001282901:c.C1204A	p.Gln622Lys	AR	A/A
		chrX	129590191	rs752439587	T	A	*OCRL*	NM_001587:c.T2603A	p.Met876Lys	X-linked	A
BA033	F	chr6	54308777	rs757092324	C	T	*TINAG*	NM_014464:c.C227T	p.Ala76Val	*De novo*	C/T
BA035	F	chr3	134561515	.	C	A	*CEP63*	NM_001042383:c.C1468A	p.Gln490Lys	*De novo*	C/A
		chr7	128812751	.	C	A	*CCDC136*	NM_022742:c.C2585A	p.Ala862Glu	*De novo*	C/A
BA036	F	chr16	75243074	rs1327850193	G	A	*BCAR1*	NM_001170715:c.C83T	p.Ala10Val	*De novo*	G/A
BA037	M	chr9	20948857	rs544335294	C	A	*FOCAD*	NM_017794:c.C3805A	p.Pro1269Thr	AR	A/A
		chrX	70341839	rs371383515	A	C	*KIF4A*	NM_012310:c.A1174C	p.Asn392His	X-linked	C
BA038	M	chrX	47448875	rs758443040	G	A	*ZNF41*	NM_001324139:c.C637T	p.Arg299Cys	X-linked	A
BA040	M	chrX	3112294	rs1426850924	C	T	*ARSF*	NM_001201538:c.C1511T	p.Pro504Leu	X-linked	T
		chrX	64191164	rs764261510	G	T	*AMER1*	NM_152424:c.C2123A	p.Thr708Asn	X-linked	T
BA041	M	chr9	100126394	rs148219510	C	G	*INVS*	NM_014425:c.C118G	p.Leu40Val	AR	G/G
		chrX	129557351	rs753369725	G	C	*OCRL*	NM_000276:c.G265C	p.Asp89His	X-linked	C/C

Abbreviations: 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.

Table 3. Identical variants detected from blood and liver samples.

Proband	Sex (M/F)	Chr	Position	SNP ID	Ref	Alt	Gene	DNA change	A.A change	MOI	Genotype	
											Proband	U
BA003	F	Undetected										
BA005	M	Undetected										
BA009	M	X	130015441	rs201843717	G	A	*BCORL1*	NM_001184772:c.G2669A	p.Arg325Gln	X-linked	A	G
BA016	M	X	56565305		C	G	*UBQLN2*	NM_013444:c.C1432G	p.Pro478Ala	X-linked	G	C
BA021	F	Undetected										
BA023	M	Undetected										
BA025	F	Undetected										
BA030	M	Undetected										
BA031	M	Undetected										
BA032	M	2	173016403	rs773527960	C	A	*RAPGEF4*	NM_001282901:c.C1204A	p.Gln622Lys	AR	A/A	C
		X	129590191	rs752439587	T	A	*OCRL*	NM_001587:c.T2603A	p.Met876Lys	X-linked	A	T
BA035	F	Undetected										
BA036	F	16	75243074	rs1327850193	G	A	*BCAR1*	NM_001170715:c.C83T	p.Ala10Val	*De novo*	G/A	G
BA037	M	9	20948857	rs544335294	C	A	*FOCAD*	NM_017794:c.C3805A	p.Pro1269Thr	AR	A/A	C
		X	70341839	rs371383515	A	C	*KIF4A*	NM_012310:c.A1174C	p.Asn392His	X-linked	C	A
BA038	M	X	47448875	rs758443040	G	A	*ZNF41*	NM_001324139:c.C637T	p.Arg299Cys	X-linked	A	G
BA039	F	Undetected										
BA040	M	X	3112294	rs1426850924	C	T	*ARSF*	NM_001201538:c.C1511T	p.Pro504Leu	X-linked	T	C
		X	64191164	rs764261510	G	T	*AMER1*	NM_152424:c.C2123A	p.Thr708Asn	X-linked	T	G
BA041	M	9	100126394	rs148219510	C	G	*INVS*	NM_014425:c.C118G	p.Leu40Val	AR	G/G	C
		X	129557351	rs753369725	G	C	*OCRL*	NM_000276:c.G265C	p.Asp89His	X-linked	C/C	G

Table 4. Variant characterization and functional prediction.

Gene	DNA change	A.A change	MAF	Intolerance	*In silico* prediction			
			KHV/gnomad/1000G	pLI/Z score	SIFT	Polyphen_2	Mutation Taster	CADD (Phred)
HACE1	NM_001350555:c.G1660A	p.Ala651Thr	0.01/0/0	0/3.59	D:0.001	D:0.96	D:0.810	26.2
PHKA1	NM_001122670:c.G478A	p.Asp160Asn	0.008/0.001/0	0/1.88	D:0.012	D:0.999	D:0.810	23.2
THOC2	NM_001081550:c.G1261A	p.Ala421Thr	0.008/0.004/0.001	1/5.53	T:0.176	B:0.002	D:0.548	21.6
XIAP	NM_001167:c.C962G	p.Ala321Gly	0/0.001/0	0.92/1.49	D:0.0	D:0.99	D:0.810	29.5
VPS13C	NM_017684:c.C5999G	p.Ala2043Gly	0.01/0.002/0	0/-1.29	D:0.0	B:0.283	D:0.537	24.3
AMER1	NM_152424:c.A1075T	p.Ser359Cys	0/0/0	0.85/-0.57	D:0.002	D:0.995	D:0.380	25
ATRX	NM_138270:c.C7318G	p.Pro2478Ala	0/0/0	1/3.1	D:0.005	B:0.164	N:0.293	22.3
POF1B	NM_001307940:c.A325C	p.Ser109Arg	0.003/0/0	0/0.78	D:0.013	B:0.12	N:0.231	19.47
BCORL1	NM_001184772:c.G2669A	p.Arg890Gln	0.008/0.001/0	1/2.06	D:0.022	D:0.99	N:0.236	23.3
INVS	NM_001318382:c.C208T	p.Arg396*	0/0/0	0/1.07	.:.	.:.	A:0.810	37
BCOR	NM_001123383:c.C1448T	p.Pro483Leu	0.003/0/0	1/1.88	T:1.0	P:0.861	D:0.442	23.1
UBQLN2	NM_013444:c.C1432G	p.Pro478Ala	0/0/0	0.85/1.5	D:0.021	P:0.519	D:0.504	23.7
MAOA	NM_000240:c.G208A	p.V70M	0/0/0	1/2.38	D:0.002	D:0.921	D:0.810	22.6
IRS4	NM_003604:c.G2835C	p.Trp945Cys	0.003/0/0	0.58/0.09	D:0.004	B:0.43	D:0.466	22.8
ELP2	NM_001242879:c.C1124T	p.Ala445Val	0/0/0	0/0.58	D:0.004	P:0.493	D:0.810	26.3
RAPGEF4	NM_001282901:c.C1204A	p.Gln622Lys	0/0/0	0.99/1.83	T:0.153	B:0.098	D:0.588	23.2
OCRL	NM_001587:c.T2603A	p.Met876Lys	0.003/0/0	1/2.96	T:0.448	B:0.001	D:0.346	18.03
TINAG	NM_014464:c.C227T	p.Ala76Val	0/0/0	0/-1.17	T:0.154	B:0.074	D:0.810	22.2
CEP63	NM_001042383:c.C1468A	p.Gln490Lys	0/0/0	0/-0.1	T:0.704	P:0.879	D:0.810	23.5
CCDC136	NM_022742:c.C2585A	p.Ala862Glu	0/0/0	0/1.13	D:0.002	P:0.756	D:0.810	14.62
BCAR1	NM_001170715:c.C83T	p.Ala10Val	0/0/0	1/-0.59	D:0.0	D:0.999	D:0.810	29.4
FOCAD	NM_017794:c.C3805A	p.Pro1269Thr	0.005/0/0	0/-2.93	D:0.003	D:0.989	D:0.537	23.2
KIF4A	NM_012310:c.A1174C	p.Asn392His	0/0/0	1/2.56	T:0.073	D:0.999	D:0.588	25.1
ZNF41	NM_001324139:c.C637T	p.Arg299Cys	0.003/0/0	0.04/1.3	D:0.042	B:0.01	N:0.090	15.37
ARSF	NM_001201538:c.C1511T	p.Pro504Leu	0.003/0/0	0/-0.39	D:0.001	D:0.968	D:0.810	22.8
AMER1	NM_152424:c.C2123A	p.Thr708Asn	0.003/0/0	0.85/-0.57	D:0.032	B:0.002	N:0.252	14.16
INVS	NM_014425:c.C118G	p.Leu40Val	0.007/0.001/0	0/1.07	T:0.204	P:0.56	D:0.810	22.7
OCRL	NM_000276:c.G265C	p.Asp89His	0.007/0.001/0	1/2.96	D:0.003	D:0.936	D:0.407	24.1

Abbreviations: A.A: amino acid; MAF: Minor Allele Frequency; KHV: Kinh Vietnamese; 1000G: 1000 Genome project; SIFT (T: tolerated; D: Damaging; score value <0.05 is likely damaging/deleterious); Polyphen2 (B: Benign; P: Possibly Damaging; D: Damaging; core value close to 1 indicates likely damaging/deleterious); Mutation Taster (A: Disease causing automatic; D: Polymorphism, probably harmless; N: Polymorphism, known to be harmless; score value close to 1 shows a high security of the prediction). CADD (scaled score; the higher the value, the more deleterious the mutation is).

Table 5. Analysis of Human Phenotype Ontology

Gene Set Name	# Genes in Gene Set (K)	Description	# Genes in Overlap (k)	k/K	p-value	FDR q-value	Gene overlap
HP_GONOSOMAL_INHERITANCE	253	Gonosomal inheritance	12	0.0474	1.49E-20	6.70E-17	*ATRX, OCRL, THOC2, MAOA, KIF4A, BCORL1, XIAP, POF1B, PHKA1, BCOR, AMER1, UBQLN2*
HP_X_LINKED_RECESSIVE_INHERITANCE	173	X-linked recessive inheritance	9	0.052	8.13E-16	1.83E-12	*ATRX, OCRL, THOC2, MAOA, KIF4A, BCORL1, XIAP, POF1B, PHKA1*
HP_SELF_INJURIOUS_BEHAVIOR	108	Self-injurious behavior	6	0.0556	5.66E-11	8.47E-08	*ATRX, OCRL, THOC2, MAOA, BCOR, ELP2*
HP_ABNORMAL_EMOTION_AFFECT_BEHAVIOR	415	Abnormal emotion/affect behavior	7	0.0169	4.97E-09	4.97E-06	*ATRX, OCRL, THOC2, MAOA, BCOR, ELP2, VPS13C*
HP_NEUROLOGICAL_SPEECH_IMPAIRMENT	1022	Neurological speech impairment	9	0.0088	6.24E-09	4.97E-06	*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	4.97E-06	*ATRX, THOC2, BCORL1, AMER1, ELP2, ZNF41, HACE1, CEP63*
HP_AUTISTIC_BEHAVIOR	450	Autistic behavior	7	0.0156	8.68E-09	5.12E-06	*ATRX, THOC2, MAOA, BCORL1, BCOR, VPS13C, ZNF41*
HP_ABNORMAL_AGGRESSIVE_IMPULSIVE_OR_VIOLENT_BEHAVIOR	251	Abnormal aggressive, impulsive or violent behavior	6	0.0239	9.12E-09	5.12E-06	*ATRX, OCRL, THOC2, MAOA, BCOR, ELP2*
HP_SHORT_STATURE	1152	Short stature	9	0.0078	1.75E-08	8.76E-06	*ATRX, OCRL, THOC2, BCOR, AMER1, ELP2, HACE1, CEP63, INVS*
HP_INVOLUNTARY_MOVEMENTS	905	Involuntary movements	8	0.0088	5.05E-08	2.19E-05	*ATRX, OCRL, THOC2, MAOA, BCORL1, UBQLN2, ELP2, VPS13C*

Figures

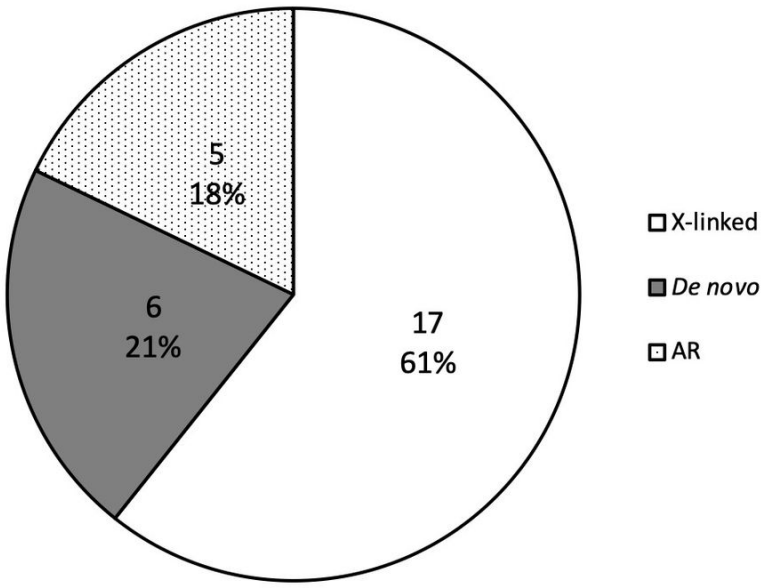

Figure 1

Mode of inheritance of identified variants from biliary atresia cohort. X-linked variants were presented in blank; De novo variants were presented in grey; and autosomal recessive were presented in dot.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf