Genome-wide association study identifies a susceptibility locus for biliary atresia on 10q24.2

Maria-Merce García-Barceló 1,†, Ming-Yiu Yeung 2,†, Xiao-Ping Miao 8,†, Clara Sze-Man Tang 2,†, Guo Chen 1, Man-Ting So 1, Elly Sau-Wai Ngan 1, Vincent Chi-Hang Lui 1, Yan Chen 1, Xue-Lai Liu 1, Kenneth-Jeremy W.S. Hui 1, Long Li 6, Wei-Hong Guo 7, Xiao-Bin Sun 8, Jin-Fa Tou 10, Kin-Wai Chan 11, Xuan-Zhao Wu 12, You-Qiang Song 4, Danny Chan 4, Kenneth Cheung 5, Patrick Ho-Yu Chung 1, Kenneth Kak-Yuen Wong 1, Pak-Chung Sham 2,3, Stacey S. Cherny 2,3,* and Paul Kwong-Hang Tam 1,*

1Department of Surgery, 2Department of Psychiatry, 3Genome Research Centre, 4Department of Biochemistry and 5Department of Orthopedics and Traumatology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China, 6Department of Paediatric Surgery and 7Department of Surgery, Beijing Children’s Hospital, Beijing, China, 8Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, 9Department of Pediatric Surgery, The 2nd Affiliated Hospital, Shandong Medical University, Shandong, China, 10Department of Surgery, Zhejiang Children’s Hospital, Zhejiang, China, 11Department of Surgery, The Chinese University of Hong Kong, Hong Kong SAR, China and 12Department of Surgery, Guiyang Medical College Affiliated Hospital, Guiyang, China

Received January 28, 2010; Revised April 27, 2010; Accepted May 6, 2010

Biliary atresia (BA) is characterized by the progressive fibrosclerosing obliteration of the extrahepatic biliary system during the first few weeks of life. Despite early diagnosis and prompt surgical intervention, the disease progresses to cirrhosis in many patients. The current theory for the pathogenesis of BA proposes that during the perinatal period, a still unknown exogenous factor meets the innate immune system of a genetically predisposed individual and induces an uncontrollable and potentially self-limiting immune response, which becomes manifest in liver fibrosis and atresia of the extrahepatic bile ducts. Genetic factors that could account for the disease, let alone for its high incidence in Chinese, are to be investigated. To identify BA susceptibility loci, we carried out a genome-wide association study (GWAS) using the Affymetrix 5.0 and 500 K marker sets. We genotyped nearly 500,000 single-nucleotide polymorphisms (SNPs) in 200 Chinese BA patients and 481 ethnically matched control subjects. The 10 most BA-associated SNPs from the GWAS were genotyped in an independent set of 124 BA and 90 control subjects. The strongest overall association was found for rs17095355 on 10q24, downstream XPNPEP1, a gene involved in the metabolism of inflammatory mediators. Allelic chi-square test P-value for the meta-analysis of the GWAS and replication results was $6.94 \times 10^{-9}$. The identification of putative BA susceptibility loci not only opens new fields of investigation into the mechanisms underlying BA but may also provide new clues for the development of preventive and curative strategies.

INTRODUCTION

Biliary atresia [BA (OMIM 210500)] is a major cause of neonatal cholestasis (arrest of the normal flow of bile) and is characterized by progressive fibrosclerosing and inflammatory obliteration of the extrahepatic biliary system during the first few weeks of life. The only current treatment is surgery: typically portoenterostomy (the Kasai operation) is performed in

†These authors contributed equally to this work.

*To whom correspondence should be addressed: Division of Paediatric Surgery, Department of Surgery, University of Hong Kong Medical Center, Queen Mary Hospital, Hong Kong SAR, China. Tel: +86 85228554850; Fax: +86 85228173155; Email: paultam@hkucc.hku.hk (P.K.-H.T.), cherny@hku.hk (S.S.C.)
early infancy, but the clearance of jaundice is only achieved in 40–60%. Post-surgical complications include cholangitis (50%) and portal hypertension (>60%). Even if bile flow is successfully restored, many patients still develop progressive inflammation and sclerosis of the biliary tree leading to secondary biliary cirrhosis. For these patients and those with failed portoenterostomy, liver transplantation is the only treatment option. The incidence of the disease varies among populations and ranges from ~1 in 5000 in Asians to 1 in 18 000 in Caucasians (1–3). BA can be classified into two groups according to the origin of the etiopathogenic process: (i) embryonic form (5–10% of the cases in Caucasians, <3% in Asians), which is attributed to defective morphogenesis of the bile duct during development; and (ii) perinatal form (most of the cases), in which bile ducts are presumed to be normal at the time of birth but undergo obliteration in the perinatal period. The inflammatory injury and subsequent obliteration of extrahepatic bile ducts occurs uniformly in all BA patients in the early post-natal period and is consistent with a common biological process that possibly reflects the response of the infant’s biliary tree to a variety of insults, either pre- or post-natal. Cholangiocytes (epithelial cells of the bile duct) initially respond to cellular injury through proliferation and with the secretion of inflammatory cytokines, chemotactant proteins and inhibitors of apoptosis. These epithelial cells acquire mesenchymal properties including migratory and invasive behavior which is essential for tissue remodeling under normal conditions. This cellular transformation (epithelial-mesenchymal transition) is thought to be involved in the fibrosis progression and obliteration of the biliary tree when imbalanced (4). Yet, the quest for the BA precipitating factors has proven futile as patient’s material is meets the innate immune system of a genetically predisposed individual and induces an uncontrollable and potentially self-limiting immune response, which becomes manifest in liver fibrosis and atresia of the extrahepatic bile ducts (5,6). BA is likely to result from the interplay between genetic and environmental factors and the immunogenetic vulnerability of the patients is likely to be critical to the progressive inflammation and sclerosing processes. Although familial cases of BA have been reported, the search for genetic susceptibility factors has been hampered by the lack of systematic twin or family studies mainly due to the low survival rate of the patients and the relative low incidence of the disease, especially in Western countries (7–11).

Several candidate-gene association studies have been conducted thus far on genes selected mainly for their role in inflammatory, immune and autoimmune responses and gene-expression analyses (12–15). Indeed, gene-expression analyses on bile duct or liver tissues of mice and humans indicated that genes involved immune and autoimmune response may play a role in BA development (16–23). However, these data are not conclusive. Thus far, genetic factors that could account for the disease, let alone for its high incidence in Chinese, are yet to be identified. BA remains the most serious liver disease in children, especially in the Chinese population. Despite the current surgical efforts to correct BA, the outcome and prognosis are still inadequate. A better understanding of the etiology and pathogenesis is needed in order to develop new strategies for early diagnosis, treatment and prevention of BA including surgery-related complications. We therefore conducted a genome-wide association study (GWAS) to discover genes involved in BA on 200 patients and 516 control individuals.

RESULTS

After stringent quality control (see Materials and Methods), association analysis was performed on 289 118 single-nucleotide polymorphism (SNPs) in 181 BA patients (94 females and 87 males) and 481 controls (183 males and 298 females). The Cochran–Armitage trend test was used to assess the association of SNPs with disease status after stratification correction. The Cochran–Armitage trend test of $\chi^2$ distribution showed that our sample was somehow inflated (median $\lambda$ = 1.067). Yet, most of the initially associated peaks persisted after the EIGENSTRAT correction (Fig. 1A), with $\lambda$ dropping to 1.032. The Q–Q plot (Fig. 1B) revealed that $P$-values $<5 \times 10^{-8}$ had deviated from those expected by chance and corresponded to 206 SNPs (Supplementary Material, Table S1).

Although some of the SNPs showing strong association resided in the same contiguous genomic region due to strong linkage disequilibrium (LD) forming ‘association peaks’, other associated SNPs were scattered throughout the genome. To verify whether those ‘isolated BA-associated SNPs’ arose from a genotyping error or from a true association in a region in weak LD, we counted the number of marker pairs having $r^2 > 0.8$ within a window of 1 Mb. If there were no other SNPs within the window, that SNP was regarded as a singleton SNP and represented a potentially true association. We identified 65 BA-associated singleton SNPs. The most associated were rs17139085 in chromosome 16p13.3 and rs9612574 in chromosome 22q11.21, with adjusted $P$-values of $1.5 \times 10^{-9}$ and $7.71 \times 10^{-8}$, respectively. No genes mapped to or near (<20 kb) these SNPs nor were they located in regions with regulatory potential as indicated by bioinformatics analysis (see Supplementary Material, Table S2, for details).

As for the associations of SNPs residing in contiguous genomic regions (Table 1), the strongest were for rs2120991 ($P = 9.87 \times 10^{-5}$), located 65 kb upstream of the HOXC13 gene on 12q13.3, and rs12571674 $(P = 3.86 \times 10^{-5})$ and rs17095355 ($P = 5.83 \times 10^{-5}$), between the XPNPEP1 and ADD3 genes on 10q24.2 (Fig. 2). Neighboring SNPs also exhibited associations, showing that the values obtained were unlikely due to genotyping artifacts. Potential bias due to systematic platform differences (see Materials and Methods) between cases and controls was investigated for the SNPs listed in Table 1. No difference between cases and controls in missing data rate and call rate was observed (Supplementary Material, Table S3).

Other BA-associated peaks were identified in 2q34–q35, 4p15.1, 11q14.1 and 15q26 (Supplementary Material, Fig. S2a–d, respectively). All BA-associated SNPs in contiguous chromosomal regions map to potentially highly conserved gene
regulatory sequences or gene transcripts, with exception of those in the chromosome 11 and 14 regions (Table 1).

Expression of genes encoding the members of the immunoglobulin superfamily and genes regulating lymphocyte function were found to be suppressed or activated in a transcriptional profile study of liver in infants with BA (20). We compared these genome-wide gene-expression results with our genome-wide association data. For each of the transcripts found to be differentially expressed between infants with BA and controls, we extracted SNPs from our GWA panel that were located within 250 kb upstream and downstream of the transcription start and end sites, respectively.

Then, the $P$-values for the extracted SNPs were investigated for any deviations from the null distribution by inspection of the Q–Q plot of these $P$-values. Unfortunately, no deviation from the distribution of $P$-values expected under the null hypothesis was observed. This may simply be a result of our particular SNPs not having a functional effect, but perhaps other SNPs in these genes would still be found to influence susceptibility to BA. Also, there may be an indirect relationship between genetic composition and gene expression. No epistasis was detected when pair-wise SNP × SNP interaction between the top 206 SNPs with each of the other SNPs in the genome-wide data set.

Figure 1. Results of the GWAS. (A) The Cochran–Armitage significance test results after EIGENSTRAT correction for population stratification. (B) Q–Q plot revealing deviation of association from expected, starting from $P = 5 \times 10^{-4}$ (green before and red after EIGENSTRAT correction) (standard error in blue).
<table>
<thead>
<tr>
<th>Chromosome</th>
<th>rsID</th>
<th>Location</th>
<th>Nearby genes (kb away)</th>
<th>Alleles</th>
<th>MA frequencies</th>
<th>P-values(^a)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Patients</td>
<td>Controls</td>
<td></td>
</tr>
<tr>
<td>2q34-q35</td>
<td>rs4673859</td>
<td>215125553</td>
<td>LOC402117(0)</td>
<td>A/G</td>
<td>0.207</td>
<td>0.138</td>
<td>1.44E-03</td>
</tr>
<tr>
<td></td>
<td>rs6724852</td>
<td>215134071</td>
<td>LOC402117(0)</td>
<td>A/G</td>
<td>0.208</td>
<td>0.136</td>
<td>9.74E-04</td>
</tr>
<tr>
<td></td>
<td>rs6435833</td>
<td>215141809</td>
<td>A/T</td>
<td>T</td>
<td>0.257</td>
<td>0.167</td>
<td>1.64E-04</td>
</tr>
<tr>
<td></td>
<td>rs7425848</td>
<td>215168677</td>
<td>LOC402117 (–17)</td>
<td>C/T</td>
<td>0.275</td>
<td>0.181</td>
<td>1.06E-04</td>
</tr>
<tr>
<td></td>
<td>rs4430957</td>
<td>215177238</td>
<td>LOC402117 (–28)</td>
<td>A/G</td>
<td>0.257</td>
<td>0.181</td>
<td>1.98E-05</td>
</tr>
<tr>
<td>4p15.1</td>
<td>rs1038821</td>
<td>23312744</td>
<td>PPARGC1A(93)</td>
<td>A/C</td>
<td>0.293</td>
<td>0.187</td>
<td>2.88E-05</td>
</tr>
<tr>
<td></td>
<td>rs3538975</td>
<td>23318712</td>
<td>PPARGC1A(87)</td>
<td>C/T</td>
<td>0.293</td>
<td>0.193</td>
<td>7.06E-05</td>
</tr>
<tr>
<td></td>
<td>rs7677217</td>
<td>23327536</td>
<td>PPARGC1A(80)</td>
<td>A/G</td>
<td>0.289</td>
<td>0.187</td>
<td>5.24E-05</td>
</tr>
<tr>
<td>10q24.2-q25.3</td>
<td>rs734165</td>
<td>111687778</td>
<td>XPNPEP1(–30)</td>
<td>T/C</td>
<td>0.533</td>
<td>0.399</td>
<td>1.71E-05</td>
</tr>
<tr>
<td></td>
<td>rs11194926</td>
<td>111690911</td>
<td>XPNPEP1 (–33)</td>
<td>C/T</td>
<td>0.536</td>
<td>0.403</td>
<td>1.45E-05</td>
</tr>
<tr>
<td></td>
<td>rs921351</td>
<td>111694704</td>
<td>XPNPEP1 (–37)</td>
<td>T/C</td>
<td>0.541</td>
<td>0.411</td>
<td>1.85E-05</td>
</tr>
<tr>
<td></td>
<td>rs7080576</td>
<td>111700806</td>
<td>XPNPEP1 (–43)</td>
<td>C/T</td>
<td>0.541</td>
<td>0.411</td>
<td>2.08E-05</td>
</tr>
<tr>
<td></td>
<td>rs731793</td>
<td>111706656</td>
<td>XPNPEP1 (–49)</td>
<td>C/G</td>
<td>0.541</td>
<td>0.408</td>
<td>1.34E-05</td>
</tr>
<tr>
<td></td>
<td>rs1257164</td>
<td>111707597</td>
<td>XPNPEP1 (–50)</td>
<td>A/G</td>
<td>0.539</td>
<td>0.399</td>
<td>7.40E-06</td>
</tr>
<tr>
<td></td>
<td>rs1709535</td>
<td>111725739</td>
<td>XPNPEP1 (–68)</td>
<td>C/T</td>
<td>0.551</td>
<td>0.409</td>
<td>5.94E-06</td>
</tr>
<tr>
<td></td>
<td>rs862066</td>
<td>111740753</td>
<td>XPNPEP1 (–83)</td>
<td>A/G</td>
<td>0.522</td>
<td>0.398</td>
<td>6.22E-05</td>
</tr>
<tr>
<td></td>
<td>rs2501577</td>
<td>111836676</td>
<td>ADD3(15)</td>
<td>G/A</td>
<td>0.539</td>
<td>0.417</td>
<td>9.52E-05</td>
</tr>
<tr>
<td></td>
<td>rs5654970</td>
<td>111837809</td>
<td>ADD3(12)</td>
<td>C/T</td>
<td>0.536</td>
<td>0.418</td>
<td>1.58E-04</td>
</tr>
<tr>
<td></td>
<td>rs2501575</td>
<td>111839853</td>
<td>ADD3(10)</td>
<td>G/A</td>
<td>0.539</td>
<td>0.422</td>
<td>1.68E-04</td>
</tr>
<tr>
<td></td>
<td>rs11194981</td>
<td>111851043</td>
<td>ADD3(0)</td>
<td>C/T</td>
<td>0.536</td>
<td>0.417</td>
<td>1.42E-04</td>
</tr>
<tr>
<td></td>
<td>rs12243027</td>
<td>111914806</td>
<td>ADD3 (–31)</td>
<td>C/T</td>
<td>0.588</td>
<td>0.463</td>
<td>7.62E-05</td>
</tr>
<tr>
<td>11</td>
<td>rs1970734</td>
<td>80063971</td>
<td>Gene desert</td>
<td>A/G</td>
<td>0.258</td>
<td>0.372</td>
<td>1.62E-04</td>
</tr>
<tr>
<td></td>
<td>rs11603258</td>
<td>80072209</td>
<td>Gene desert</td>
<td>A/G</td>
<td>0.254</td>
<td>0.368</td>
<td>1.64E-04</td>
</tr>
<tr>
<td></td>
<td>rs12289502</td>
<td>80077060</td>
<td>Gene desert</td>
<td>A/C</td>
<td>0.251</td>
<td>0.361</td>
<td>3.01E-04</td>
</tr>
<tr>
<td></td>
<td>rs1917858</td>
<td>80080034</td>
<td>Gene desert</td>
<td>A/T</td>
<td>0.249</td>
<td>0.357</td>
<td>2.96E-04</td>
</tr>
<tr>
<td></td>
<td>rs1917857</td>
<td>80080289</td>
<td>Gene desert</td>
<td>T/C</td>
<td>0.247</td>
<td>0.358</td>
<td>2.33E-04</td>
</tr>
<tr>
<td></td>
<td>rs1917855</td>
<td>80080705</td>
<td>Gene desert</td>
<td>T/C</td>
<td>0.261</td>
<td>0.358</td>
<td>1.34E-03</td>
</tr>
<tr>
<td>12q13.3</td>
<td>rs796720</td>
<td>52545875</td>
<td>HOX13 (73)</td>
<td>G/A</td>
<td>0.42</td>
<td>0.313</td>
<td>3.16E-04</td>
</tr>
<tr>
<td></td>
<td>rs7306498</td>
<td>52545967</td>
<td>HOX13 (72)</td>
<td>A/G</td>
<td>0.323</td>
<td>0.206</td>
<td>1.95E-05</td>
</tr>
<tr>
<td></td>
<td>rs6580967</td>
<td>52547422</td>
<td>HOX13 (71)</td>
<td>A/C</td>
<td>0.42</td>
<td>0.312</td>
<td>3.05E-04</td>
</tr>
<tr>
<td></td>
<td>rs2120991</td>
<td>52556494</td>
<td>HOX13 (62)</td>
<td>A/C</td>
<td>0.249</td>
<td>0.134</td>
<td>7.50E-07</td>
</tr>
<tr>
<td></td>
<td>rs2577864</td>
<td>52573427</td>
<td>HOX13 (45)</td>
<td>C/T</td>
<td>0.323</td>
<td>0.206</td>
<td>1.95E-05</td>
</tr>
<tr>
<td></td>
<td>rs2590712</td>
<td>25273711</td>
<td>HOX13 (45)</td>
<td>C/G</td>
<td>0.323</td>
<td>0.205</td>
<td>1.64E-05</td>
</tr>
<tr>
<td>13q12-q13</td>
<td>rs1590501</td>
<td>29354306</td>
<td>UBLN (–35)</td>
<td>T/C</td>
<td>0.116</td>
<td>0.046</td>
<td>6.64E-06</td>
</tr>
<tr>
<td></td>
<td>rs3149886</td>
<td>29356736</td>
<td>UBLN (–32)</td>
<td>G/T</td>
<td>0.119</td>
<td>0.044</td>
<td>9.92E-07</td>
</tr>
<tr>
<td>14</td>
<td>rs1569297</td>
<td>21842434</td>
<td>Gene desert</td>
<td>C/T</td>
<td>0.186</td>
<td>0.103</td>
<td>5.81E-05</td>
</tr>
<tr>
<td></td>
<td>rs2331601</td>
<td>21844534</td>
<td>Gene desert</td>
<td>A/G</td>
<td>0.175</td>
<td>0.083</td>
<td>3.04E-06</td>
</tr>
<tr>
<td>15q26</td>
<td>rs12439738</td>
<td>90336554</td>
<td>SLC03A4(0)</td>
<td>C/T</td>
<td>0.251</td>
<td>0.351</td>
<td>9.60E-04</td>
</tr>
<tr>
<td></td>
<td>rs4932597</td>
<td>90338620</td>
<td>SLC03A4(0)</td>
<td>A/G</td>
<td>0.249</td>
<td>0.355</td>
<td>3.92E-04</td>
</tr>
<tr>
<td></td>
<td>rs4932598</td>
<td>90338848</td>
<td>SLC03A4(0)</td>
<td>C/T</td>
<td>0.251</td>
<td>0.351</td>
<td>8.34E-04</td>
</tr>
</tbody>
</table>

MA, minor allele; CA, Cochran–Armitage test. The embolden SNPs are chosen for the second stage.

\(^{a}\)P-values were calculated by the Cochran–Armitage test and adjusted by EIGENSTRAT.

\(^{b}\)p-values were calculated by the chi-square allelic test.

SNP associations are confirmed in an independent sample and in combined analysis

To validate the finding from the whole-genome scan, we genotyped 10 SNPs in 124 additional BA cases and in 90 additional controls. The 10 SNPs (Table 1, boldface) were selected not only according to their EIGENSRAT-adjusted Cochran–Armitage test $P$-values but also to the $P$-values of the associated SNPs in the same contiguous genomic region. The SNPs selected for replication were not subjected to population stratification as observed from the unadjusted and adjusted $P$-values depicted in Table 1. All but rs2331601 were successfully genotyped in all individuals. In Table 2, we present the results of the association tests obtained using the simple allelic chi-square test for both GWAS and follow-up stages together with the results of the meta-analyses of the GWAS and replication results. As for the meta-analysis of $P$-values using the METAL software, data sets were weighted to the square-root of the number of individuals examined in each data set such that the test might be more powerful.

The data strongly suggest that the initial finding for rs17095355, upstream XPNPEP1, was genuine as this SNP showed strong and significant association with BA in the independent sample ($P = 2.4 \times 10^{-4}$). More importantly, the association value of the meta-analysis of follow-up and genome-wide scan samples was $P = 6.94 \times 10^{-9}$. This association value is still significant after a conservative Bonferroni genome-wide correction for multiple testing of 500,000 markers, which would require $P < 0.05/500,000$, or $P < 10^{-7}$.

As for the other SNPs genotyped on the second stage, none reached a genome-wide significant association $P$-values. Yet, we would like to emphasize rs2501577, which maps to the same contiguous chromosomal region as rs17095355. Even if the association values are not small enough to survive the conservative Bonferroni correction, the data show that there is an association trend encompassing the 10q24.2 association peak. Moreover, this region is just between the XPNPEP1 [X-prolyl aminopeptidase (aminopeptidase P)1, soluble] and ADD3 (adducin 3) genes, both relevant liver functions.
**DISCUSSION**

This study shows that the likelihood for developing BA is influenced by DNA variants in a region spanning 129 kb and encompassing the \(XPNPEP1\) and \(ADD3\) genes (Fig. 2). As the most associated SNP identified in this study, rs17095355, falls on the intergenic region, we assessed (i) the biological plausibility of these ‘flanking’ genes in regard to the pathology of BA and (ii) the possible functional consequences on gene regulation of the associated SNP allele.

\(ADD3\) encodes adducin 3, which belongs to a family of membrane skeletal proteins involved in the assembly of spectrin–actin network in erythrocytes and at sites of cell–cell contact in epithelial tissues, including that of the digestive tract, liver and biliary tract (24,25). \(ADD3\) is expressed in liver and biliary epithelia, and noticeably, is more abundantly expressed in fetal liver than in adult liver (25). Contraction of the bile canicular membrane (facilitate the bile flow) are controlled by actin–myosin interactions. Importantly, impairment of these interaction mechanisms with experimental drugs causes severe cholestasis (26). Increased actin and myosin deposition around bile canaliculi has been observed in BA patients who did not exhibit bile flow after surgery (27). Also, the expression intensity of \(\alpha\)-smooth-muscle actin is correlated with the degree of fibrosis in patients with BA (28).

\(XPNPEP1\) is expressed in epithelial cells of the hepatobiliary system (29). It encodes soluble X-prolyl aminopeptidase P or soluble aminopeptidase P (APP1). APP1 is a near-ubiquitous enzyme, involved in the metabolism of bradykinin (BK) and substance P (SP) (30). BK is involved in the regulation of vasodilatation and capillary permeability and its expression is directly regulated by bile acid nuclear receptor, the farnesoid X receptor (FXR) known to play a role in the regulation of bile acid metabolism and secretion and inflammatory processes (31,32). SP functions as a neurotransmitter and inflammatory mediator and is also involved in the regulation of bile secretion, biliary dynamics and liver involvements. A role for hepatobiliary transporters (particularly Fxr) in murine BA has recently been reported (33).

We believe that there is evidence substantiating a role for \(ADD3\) and/or \(XPNPEP1\) in BA development, possibly by disturbing the networks involved in the control of inflammatory processes. Yet, even if our genetic data points at a regulatory region comprised between these two genes, a more refined search for the DNA variants affecting gene regulation is needed.

Thus, in order to find out whether the most BA-associated SNP could exert an effect on the regulation of \(ADD3\) or \(XPNPEP1\) genes, we exhaustively examined the region in which rs17095355 is located using bioinformatics tools (Genomatix suite). Analysis revealed that the C > T transition (‘T’ BA-associated allele) created two and disrupted three transcription factor binding sites (TFBSs). Yet, comparative genomics analysis indicated that the sequence encompassing and surrounding the TFBSs overlapped by rs17095355 was hardly conserved among species making rs17095355 unlikely to have a functional effect.

As association of an SNP with a disease indicates that the associated SNP is either itself functional or that it is merely
a proxy (due to LD) for other functional variants and because the finding of association in non-coding sequence (CDS) implies the involvement of regulatory elements; we resorted to further comparative genomics analyses within the region.

We then used the UCSC browser to identify multispecies conserved non-CDS that could comprise SNPs in LD ($r^2 > 0.8$) with rs17095355 and that overlapped or were in the vicinity of TFBS implicated in the regulation of liver genes. This revealed that rs921348 and rs9630101 were flanking the TFBS sequences of the hepatic leukemia factor (HLF) and homeobox containing protein 1 (HMBOX1), respectively. Although the SNPs do not create or disrupt any of the TFBS, given the conservation of the region, and their location, these SNPs might introduce new sites that interfere with neighboring regulatory sequences. These transcription factors (TFs) belong to the PAR/bZIP and hepatic nuclear factor 1 (HNF1) families, known to be expressed in liver (34). In particular, HLF expression is relevant in hepatic metabolism and liver responses to xenobiotic agents (35).

However, an exhaustive literature search failed to find evidence linking these TFs expressed in liver and overlapping BA-associated SNPs with the regulation of ADD3 and XPNPEP1, although it is worth mentioning that the existing literature regarding the regulation of these two genes is scarce.

The GWAS presented here has its limitations, namely relatively small sample size. However, the fact that the association of rs17095355 was confirmed in an independent sample, the strong evidence provided by the meta- and combined analysis and the association trend showed by the SNPs within the 10q24.2–q25.3 region, indicates that our data are reliable and worthy of follow-up. Given all of the above and the biological plausibility of the genes pointed out, we believe that our data may help evaluate the potential etiology of BA.

The ‘common disease–common variant’ hypothesis has been the basis of GWAS under the assumption that the risk of common disease is conferred by a small number of common variants. This is in contrast to the recently proposed, ‘common disease–rare variants’ hypothesis, which suggest that complex diseases may be due many highly penetrant but rare predisposing variants in numerous genes (36). Although rare variants involved in BA may only be detected by large-scale resequencing efforts, common variants may be uncovered by GWAS may also provide new insight into possible candidate genes for subsequent search for rare variants. Thus, sequencing of ADD3 and/or XPNPEP1 after fine-mapping of the intergenic regions should be the next step. In fact, BA may be explained by either ‘rare disease–common variant’ and ‘rare disease–rare variant’ models indicating that susceptibility genes may contain both rare highly penetrant mutations and common variants that exhibit low penetrance.

This GWAS, to our knowledge, is unprecedented in BA research being the largest in any genetic analysis of this disease to date and represents a major step forward in elucidating genetic susceptibility to BA. We believe that the data generated opens the door to not only SNP-based analysis but also to gene-based and pathway-based analysis, which may prove to be more powerful in dissecting the genetic component for this multifactorial disorder.

### MATERIALS AND METHODS

#### Subjects

The study was approved by the Institutional Review Board of the University of Hong Kong Hospital Authority Hong Kong West Cluster (UW 05-282 T/945). Blood samples were drawn from all participants after obtaining informed parental consent.

**BA patients.** BA was diagnosed by hepatobiliary scintigraphy and confirmed by operative cholangiography. All patients included in this study were affected with the perinatal form of BA. A total of 324 BA patients of Northern or Southern Chinese origin (1:4) were included in the study. Patients were initially grouped into genome-wide scan (200 BA; 107 females and 93 males) and replication series (124 patients).

**Controls.** Three hundred and sixty-four individuals (125 male and 239 female) of Northern or Southern Chinese origin (1:4) were included in the genome-wide scan group. These individuals had been initially recruited for a GWA aimed at the genetic discovery of quantitative trait loci for degeneration disc disease (DDD, see details in Supplementary material and in 37). To strengthen the statistical power, genotypes of 152 Chinese (82 from Beijing and 70 from Denver; 72 males and 80 females) participants of the International HapMap Project Phase 3 were included as additional controls. Thus, 516 individuals were initially included in the genome-scan control group, however, after quality control (see below), only 481 individuals (333 DDD, 114 males and 219 females, and 148 HapMap, 69 males and 79 females) were analyzed. For replication, we used DNA of 90 non-affected individuals with no familial history of BA (35 females and 55 males).

#### Whole-genome scan genotyping and analysis

DNAs from BA patients and control individuals were genotyped using the *Affymetrix* Genome-Wide Human SNP Array 5.0 and the *Affymetrix* GeneChip Human Mapping 500 K Array Set (250 K *NspI* and *StyI* arrays), respectively, according to the manufacturer’s protocol.

Genotypes were called from the raw cell intensity data by the *Affymetrix* Power Tools command line programs that implement several algorithms for genotype calling. As for genotype data of the HapMap Phase 3 CHB samples included in the control data set, we used the concordant genotypes from the merged Illumina HumanM1 and the *Affymetrix* SNP 6.0 platforms.

Although only the BRLMM (Bayesian Robust Linear Model with Mahalanobis Distance) algorithm could be used for genotype calling of the controls’ 500 K chips, two different algorithms, namely BRLMM-P and Birdseed, were available for genotype calling of the patients’ 5.0 chips. In order to minimize the false-positive associations due to genotyping artifacts, cases’ genotypes that were concordant between the BRLMM-P and Birdseed calls were included in the subsequent GWA analysis. The genotype data were stored in OpenADAM for data manipulation and export for downstream analysis (38).
PLINK was employed to perform data quality control and case–control allelic association analysis using the chi-square test (39). Detection for population stratification and calculation of adjusted association statistics were done by the EIGENSTRAT software version 2.0 based on principle component analysis and the Armitage trend test, respectively. Haploview 4.1 was used for visualization of association data and inspection of LD patterns (40).

Q–Q plots were used to investigate if the interaction deviated from the null distribution. A SNP was assigned to a gene when it was located within 250 kb upstream and downstream of the transcription start and end sites, respectively. When there was no gene within a 250 kb region of an SNP, then the SNP was defined to be in a gene desert.

Meta-analyses of P-values from the GWAS and replication results were performed using the Metal software, based on the weighted inverse variance method (http://www.sph.umich.edu/csg/abecasis/metal/index.html).

Criteria for exclusion of individuals. BA samples with call rate lower than 95% were excluded. As a preliminary control for population stratification, samples were clustered using the multidimensional scaling (Supplementary Material, Fig. S1) and the nearest neighbor analysis in PLINK. In addition, samples were subjected to biological relationship and duplication verification based on the genome-wide identity-by-state information computed by PLINK. A total of 19 BA patients and 35 controls (31 DD and 4 HapMap) were excluded from subsequent analysis. The final set of samples in the genome-wide scan was therefore composed of 181 cases and 481 controls (including 333 DD and 148 HapMap Phase 3 Chinese samples) with a mean call rate of 99.629%.

Criteria for exclusion of SNPs. SNPs satisfying the following criteria were excluded from the analysis: (i) having a call rate lower than 95% in the GWA data set (n = 29 769), or (ii) having a minor allele frequency (MAF) lower than 5% in the GWA data set (n = 149 207) or (iii) having a genotype frequency deviated from the Hardy–Weinberg equilibrium (P < 0.001) in the controls (n = 1574). A total of 289 172 survived the quality-control criteria. The threshold of MAF < 5% was chosen because the observed frequencies for those SNPs may have excessive stochastic variation when assessed in a relatively small sample size leading to false-positive association, partially a result of difficulties in calling such genotypes.

Replication. PCR followed by direct sequencing were used for genotyping 10 SNPs selected from the GWAS on the basis of P-association values of not only the selected SNPs but also of those in vicinity (same contiguous genomic region) and biological plausibility. One hundred and twenty-four samples of BA patients and 35 controls (31 DD and 4 HapMap) were excluded from the replication. Only 10 SNPs were chosen since, due to the poor quality of the DNA of the replication sample set, sequencing was the only method available for genotyping as high-throughput genotyping techniques require high-quality DNA.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGEMENTS**

We extend our gratitude to all subjects who participated in the study. We are also grateful to Dr Agnes Chan and Prof. Kathryn S.E. Cheah for their assistance and to the Genomics Strategic Research Theme and the Centre for Reproduction, Development and Growth of the University of Hong Kong.

**Conflict of Interest statement.** None declared.

**FUNDING**

This work was supported by the Hong Kong Research Grants Council (HKU 7628/06M to P.K.-H.T.); and the Seed Funding Programme for Basic Research (200511159030 to P.K.-H.T.). Support was also received from the University Grants Committee of Hong Kong (AoE/M-04/04); the National Institutes of Health (EY-12562 to S.S.C. and P.C.S.), and AOSPINE (AOSBRC-07-02 to D.C.).

**REFERENCES**


