Human Molecular Genetics, 2010, Vol. 19, No. 14 ***2917–2925***
doi:10.1093/hmg/ddq196
Advance Access published on May 11, 2010

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

Maria-Mercè Garcia-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[9], 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,*]

[1]Department of Surgery, [2]Department of Psychiatry, [3]Genome Research Centre, [4]Department of Biochemistry and [5]Department of Orthopedics and Traumatology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China, [6]Department of Paediatric Surgery and [7]Department of Surgery, Beijing Children's Hospital, Beijing, China, [8]Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [9]Department of Pediatric Surgery, The 2nd Affiliated Hospital, Shandong Medical University, Shandong, China, [10]Department of Surgery, Zhejiang Children's Hospital, Zhejiang, China, [11]Department of Surgery, The Chinese University of Hong Kong, Hong Kong SAR, China and [12]Department 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 6.94×10^{-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

*To whom correspondence should be addressed at: 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.)
†These authors contributed equally to this work.

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 etio-pathogenic 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, chemoattractant 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 only available after diagnosis, at which point primary and secondary effects are indistinguishable.

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 (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 χ^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 λ dropping to 1.032. The Q–Q plot (Fig. 1B) revealed that P-values $<5 \times 10^{-4}$ 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 1.5×10^{-9} and 7.71×10^{-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 ($P = 9.87 \times 10^{-7}$), located 65 kb upstream of the *HOXC13* gene on 12q13.3, and rs12571674 ($P = 3.86 \times 10^{-6}$) and rs17095355 ($P = 5.83 \times 10^{-6}$), 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

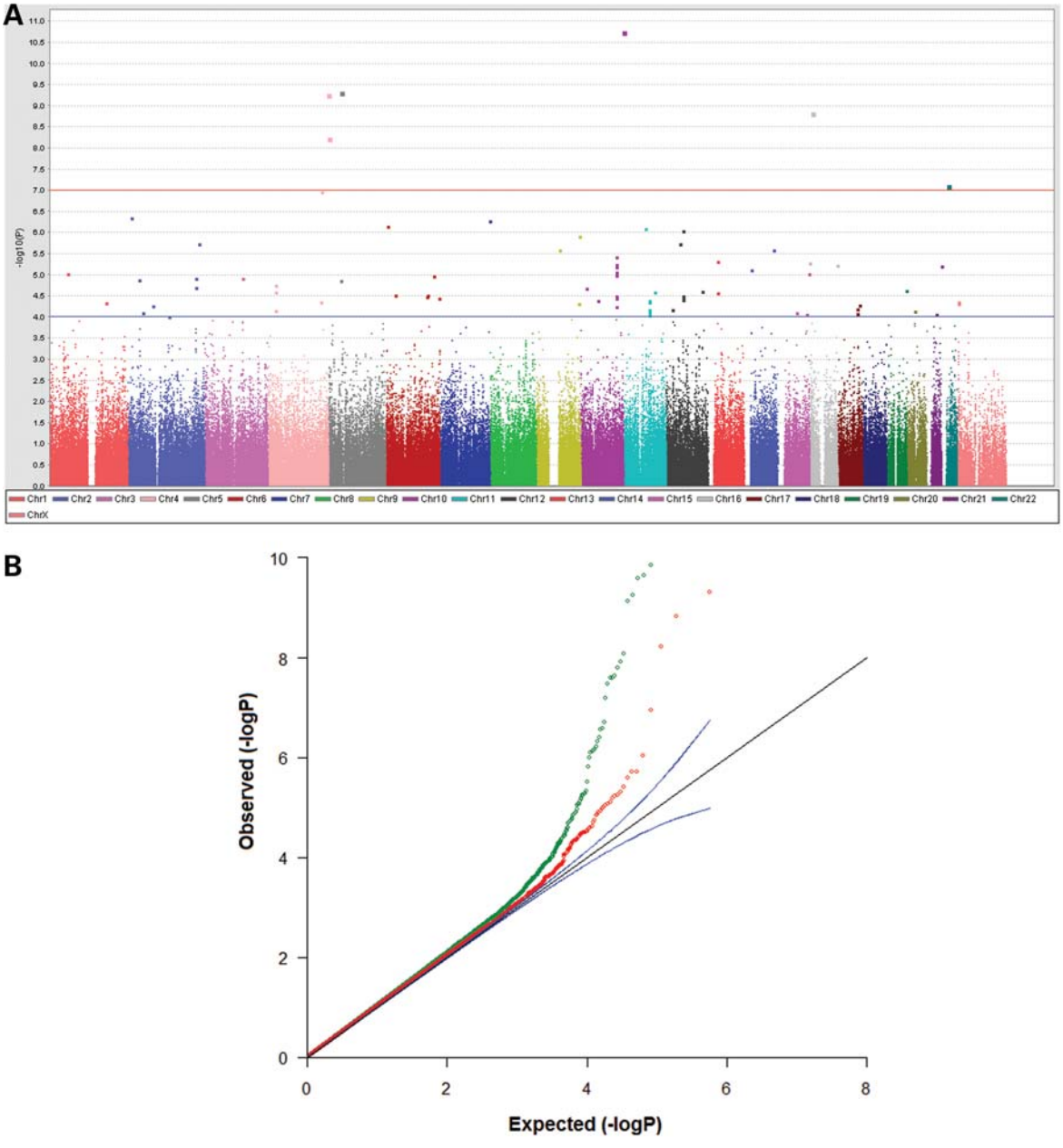

Figure 1. Results of the GWAS. (**A**) The Cochran–Armitage significance test results after EIGENSTRAT correction for population stratification. (**B**) Q–Q plot revealing deviation of association from expected, starting from $P = 5 \times 10^{-4}$ (green before and red after EIGENSTRAT correction) (standard error in blue).

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.

Table 1. List of BA-associated SNPs in contiguous chromosomal regions

Chromosome	rsID	Location	Nearby genes (kb away)	Alleles	MA	MA frequencies		*P*-values[a]			OR (95% CI)
						Patients	Controls	CA	EIGENSTRAT	Allelic[b]	
2q34-q35	rs4673859	215125553	*LOC402117*(0)	A/G	G	0.207	0.138	1.44E−03	3.62E−04	1.88E−03	1.64 (1.20–2.24)
	rs6724852	215134071	*LOC402117*(0)	A/G	A	0.208	0.136	9.74E−04	2.31E−04	1.28E−03	1.67 (1.22–2.28)
	rs6435833	215141809	*LOC402117*(0)	A/T	T	0.257	0.167	1.64E−04	1.24E−05	2.26E−04	1.72 (1.29–2.30)
	rs7425848	215166877	*LOC402117*(−17)	C/T	T	0.275	0.181	1.06E−04	1.98E−05	1.68E−04	1.72 (1.29–2.28)
	rs4430957	215177238	*LOC402117*(−28)	A/G	G	0.257	0.18	1.51E−03	1.89E−04	1.92E−03	1.56 (1.18–2.20)
4p15.1	**rs1038821**	23312744	*PPARGC1A*(93)	A/C	C	0.293	0.187	2.88E−05	1.76E−05	3.10E−05	1.80 (1.36–2.38)
	rs10938957	23318712	*PPARGC1A*(87)	C/T	C	0.293	0.193	1.38E−04	7.06E−05	9.99E−05	1.73 (1.31–2.28)
	rs7677217	23325736	*PPARGC1A*(80)	A/G	A	0.289	0.187	5.24E−05	2.57E−05	5.38E−05	1.77 (1.34–2.34)
10q24.2-q25.3	rs734165	111687778	*XPNPEP1*(−30)	T/C	T	0.533	0.399	1.71E−05	9.01E−06	1.18E−05	1.72 (1.35–2.19)
	rs11194926	111690911	*XPNPEP1*(−33)	C/T	T	0.536	0.403	1.45E−05	6.52E−06	1.40E−05	1.71 (1.34–2.18)
	rs921351	111694704	*XPNPEP1*(−37)	T/C	C	0.541	0.41	1.85E−05	8.68E−06	1.69E−05	1.70 (1.33–2.17)
	rs7080576	111700806	*XPNPEP1*(−43)	C/T	C	0.541	0.41	2.08E−05	1.01E−05	1.93E−05	1.69 (1.33–2.16)
	rs731793	111706656	*XPNPEP1*(−49)	C/G	C	0.541	0.408	1.34E−05	5.72E−06	1.23E−05	1.72 (1.35–2.19)
	rs12571674	111707597	*XPNPEP1*(−50)	A/G	G	0.539	0.399	7.40E−06	3.82E−06	5.14E−06	1.76 (1.38–2.24)
	rs17095355	111725739	*XPNPEP1*(−68)	C/T	T	0.551	0.409	5.49E−06	5.83E−06	4.75E−06	1.77 (1.38–2.26)
	rs3862006	111740753	*XPNPEP1*(−83)	A/G	A	0.522	0.398	6.22E−05	3.15E−05	4.93E−05	1.65 (1.29–2.11)
	rs2501577	111836676	*ADD3*(13)	G/A	G	0.539	0.417	9.52E−05	3.22E−05	7.14E−05	1.63 (1.28–2.08)
	rs6584970	111837809	*ADD3*(12)	C/T	C	0.536	0.418	1.58E−04	3.59E−05	1.18E−04	1.61 (1.26–2.06)
	rs2501575	111839853	*ADD3*(10)	G/A	A	0.539	0.422	1.68E−04	5.78E−05	1.38E−04	1.60 (1.26–2.04)
	rs11194981	111851043	*ADD3*(0)	C/T	C	0.536	0.417	1.42E−04	3.40E−05	1.04E−04	1.62 (1.27–2.06)
	rs12243027	111914806	*ADD3*(−31)	C/T	T	0.588	0.463	7.62E−05	1.14E−04	4.46E−05	1.66 (1.30–2.12)
11	rs1970734	80063971	Gene desert	G/A	A	0.258	0.372	1.62E−04	4.37E−05	1.02E−04	0.59 (0.45–0.77)
	rs11603258	80072209	Gene desert	A/G	G	0.254	0.368	1.64E−04	4.15E−05	9.76E−05	0.59 (0.48–0.77)
	rs12289502	80077060	Gene desert	A/C	C	0.251	0.36	3.01E−04	8.92E−05	1.88E−04	0.60 (0.46–0.78)
	rs1917858	80080034	Gene desert	A/T	A	0.249	0.357	2.96E−04	7.70E−05	1.91E−04	0.60 (0.45–0.78)
	rs1917857	80080289	Gene desert	T/C	C	0.247	0.358	2.33E−04	6.77E−05	1.41E−04	0.59 (0.45–0.78)
	rs1917855	80080705	Gene desert	T/C	C	0.261	0.358	1.34E−03	3.69E−04	9.70E−04	0.63 (0.48–0.83)
12q13.3	rs796720	52545875	*HOXC13*(73)	G/A	G	0.42	0.313	3.16E−04	4.02E−04	2.42E−04	1.59 (1.24–2.04)
	rs7306498	52545967	*HOXC13*(72)	A/G	A	0.323	0.206	1.95E−05	3.87E−05	7.79E−06	1.84 (1.41–2.41)
	rs6580967	52547422	*HOXC13*(71)	A/C	C	0.42	0.312	3.05E−04	3.61E−04	2.09E−04	1.60 (1.25–2.05)
	rs2120991	52556494	*HOXC13*(62)	A/C	A	0.249	0.134	7.50E−07	9.13E−07	5.77E−07	2.14 (1.58–2.89)
	rs2577864	52573427	*HOXC13*(45)	C/T	T	0.323	0.206	1.95E−05	3.36E−05	7.79E−06	1.84 (1.41–2.41)
	rs2590712	52573711	*HOXC13*(45)	C/G	G	0.323	0.205	1.64E−05	3.13E−05	6.32E−06	1.85 (1.41–2.43)
13q12-q13	rs1590501	29354306	*UBL3*(−35)	T/C	C	0.116	0.046	6.64E−06	2.74E−05	4.86E−06	2.71 (1.74–4.21)
	rs9314986	29356736	*UBL3*(−32)	G/T	G	0.119	0.044	9.92E−07	4.79E−06	6.67E−07	2.95 (1.89–4.60)
14	rs1569297	21842434	Gene desert	C/T	T	0.186	0.103	5.81E−05	2.40E−04	5.45E−05	1.98 (1.42–2.78)
	rs2331601	21844534	Gene desert	A/G	A	0.175	0.083	3.04E−06	7.74E−06	1.70E−06	2.34 (1.64–3.34)
15q26	rs12439738	90336554	*SLCO3A1*(0)	C/T	T	0.251	0.351	9.60E−04	2.14E−04	5.73E−04	0.62 (0.47–0.87)
	rs4932597	90338620	*SLCO3A1*(0)	A/G	G	0.249	0.355	3.92E−04	8.65E−05	2.49E−04	0.60 (0.46–0.79)
	rs4932598	90338848	*SLCO3A1*(0)	C/T	C	0.251	0.35	8.34E−04	2.19E−04	6.03E−04	0.62 (0.47–0.82)

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.

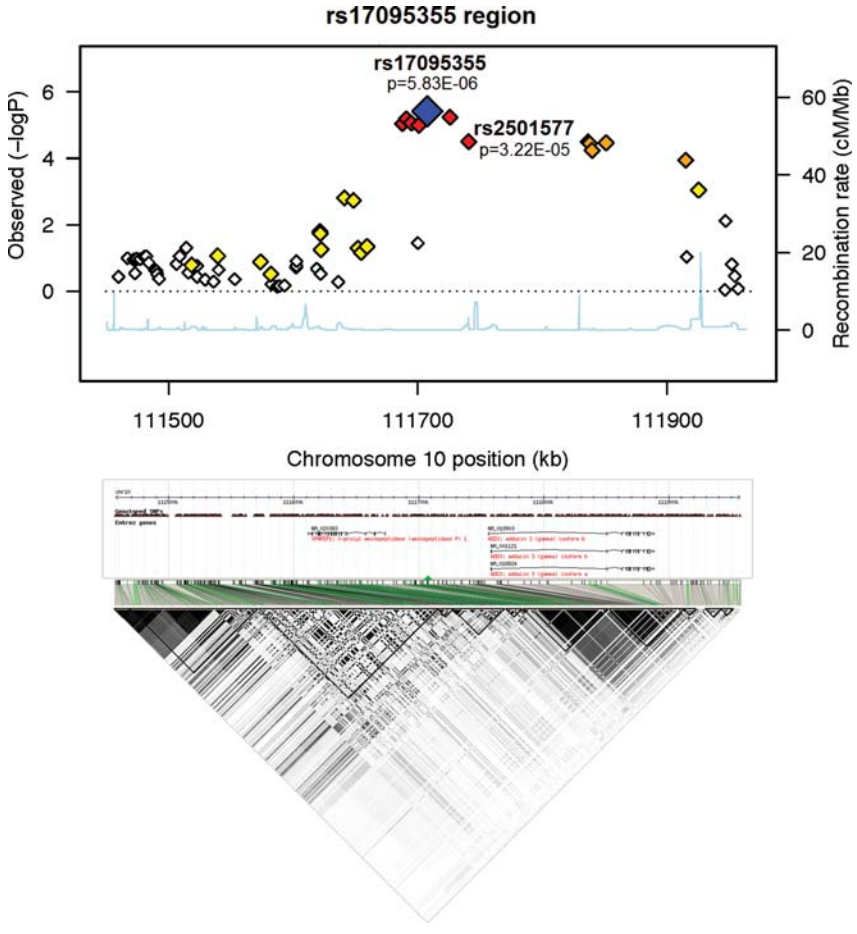

Figure 2. Genomic context for the BA-associated peak on 10q24.2. $-\log(P)$ value for the BA-associated SNPs rs17095355 (blue diamond) and rs2501577. Also shown is the position of recombination hotspots (light blue line at the bottom of top panel). Genes in the region are represented by green lines with arrow heads indicating the sense of transcription. The color of the nearby SNPs (diamond) depends on its r^2 with the associated SNP: red, ≥ 0.8; orange, ≥ 0.5 and < 0.8; yellow, ≥ 0.2 and < 0.5; white, < 0.2. Chromosome position is plotted with reference to the NCBI build 35 and gene names are plotted with reference to the University of California at Santa Cruz Genome Browser. Recombination rate estimated from the HapMap is plotted in light blue. In the bottom panel, LD structures based on pair-wise r^2 in the associated region in the CHB and JPT population are shown. The SNPs with the strongest association are highlighted in white and nearby SNPs also on the GWAS SNP panel in green. Other SNPs are in black.

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 EIGENTSRAT-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.

Table 2. List of SNPs followed-up with replication

Chromosome	rsID	Location	Nearby genes (kb away)	Alleles	MA	GWAS MAF				Replication MAF				Meta-analysis (*P*-value)
						Cases	Controls	*P*-value	OR (95% CI)	Cases	Controls	*P*-value	OR (95% CI)	
2q34-35	rs7425848	215166878	*LOC402117*(−17)	T/C	T	0.275	0.181	1.68E−04	1.72 (1.29–2.28)	0.221	0.167	1.68E−01	1.42 (0.86–2.33)	7.70E−05
4p15.1	rs1038821	23312745	*PPARGC1A*(93)	G/T	G	0.293	0.187	3.10E−05	1.80 (1.36–2.38)	0.219	0.196	6.49E−01	1.15 (0.63–2.09)	1.20E−04
10q24.2-q25.3	**rs17095355**	**111725740**	***XPNPEP1*(−68)**	**T/C**	**T**	**0.551**	**0.409**	**4.75E−06**	**1.77 (1.38–2.26)**	**0.539**	**0.355**	**2.41E−04**	**2.13 (1.42–3.2)**	**6.94E−09**
	rs2501577	111836677	*ADD3*(13)	G/A	G	0.539	0.417	7.14E−05	1.63 (1.28–2.08)	0.509	0.375	9.98E−03	1.72 (1.14–2.61)	2.29E−06
11	rs11603258	80072210	Gene desert	G/A	G	0.254	0.368	9.76E−05	0.59 (0.45–0.77)	0.35	0.306	3.38E−01	1.22 (0.81–1.85)	1.13E−04
12q13.3	rs2120991	52556495	*HOXC13*(62)	A/C	A	0.249	0.134	5.77E−07	2.14 (1.58–2.89)	0.196	0.188	8.61E−01	1.06 (0.58–1.93)	9.34E−06
	rs2577864	52573428	*HOXC13*(45)	T/C	T	0.323	0.206	7.79E−06	1.84 (1.41–2.41)	0.271	0.217	3.18E−01	1.34 (0.76–2.37)	1.19E−05
13q12-q13	rs9314986	29356737	*UBL3*(−32)	G/T	G	0.119	0.044	6.67E−07	2.95 (1.89–4.6)	0.055	0.078	4.12E−01	0.69 (0.29–1.67)	2.29E−06
15q26	rs4932597	90338621	*SLCO3A1*(0)	G/A	G	0.249	0.355	2.49E−04	0.60 (0.46–0.79)	0.306	0.299	8.76E−01	1.04 (0.67–1.59)	1.11E−03

MA, minor allele; MAF, minor allele frequency. *P*-values were calculated by the chi-square allelic test. Bold values are most associated SNP.

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). Contractions of the bile canalicular 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 α-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 innervations. 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 USCS 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 *Nsp*I and *Sty*I 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* Human1M 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 200 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 DNA from Northern Chinese BA patients and 90 from Northern Chinese controls were included in 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

1. Shim, W.K., Kasai, M. and Spence, M.A. (1974) Racial influence on the incidence of biliary atresia. *Prog. Pediatr. Surg.*, **6**, 53–62.
2. Petersen, C., Harder, D., Abola, Z., Alberti, D., Becker, T., Chardot, C., Davenport, M., Deutschmann, A., Khelif, K., Kobayashi, H. *et al.* (2008) European biliary atresia registries: summary of a symposium. *Eur. J. Pediatr. Surg.*, **18**, 111–116.
3. Hsiao, C.H., Chang, M.H., Chen, H.L., Lee, H.C., Wu, T.C., Lin, C.C., Yang, Y.J., Chen, A.C., Tiao, M.M., Lau, B.H. *et al.* (2008) Universal screening for biliary atresia using an infant stool color card in Taiwan. *Hepatology*, **47**, 1233–1240.
4. Omenetti, A., Porrello, A., Jung, Y., Yang, L., Popov, Y., Choi, S.S., Witek, R.P., Alpini, G., Venter, J., Vandongen, H.M. *et al.* (2008) Hedgehog signaling regulates epithelial-mesenchymal transition during biliary fibrosis in rodents and humans. *J. Clin. Invest.*, **118**, 3331–3342.
5. Petersen, C. (2007) Biliary atresia: interdisciplinary initiatives focus on a rare disease. *Pediatr. Surg. Int.*, **23**, 521–527.
6. Hartley, J.L., Davenport, M. and Kelly, D.A. (2009) Biliary atresia. *Lancet*, **374**, 1704–1713.
7. Lachaux, A., Descos, B., Plauchu, H., Wright, C., Louis, D., Raveau, J. and Hermier, M. (1988) Familial extrahepatic biliary atresia. *J. Pediatr. Gastroenterol. Nutr.*, **7**, 280–283.
8. Silveira, T.R., Salzano, F.M., Howard, E.R. and Mowat, A.P. (1991) Extrahepatic biliary atresia and twinning. *Braz. J. Med. Biol. Res.*, **24**, 67–71.
9. Poovorawan, Y., Chongsrisawat, V., Tanunytthawongse, C., Norapaksunthorn, T., Mutirangura, A. and Chandrakamol, B. (1996) Extrahepatic biliary atresia in twins: zygosity determination by short tandem repeat loci. *J. Med. Assoc. Thai.*, **79** (Suppl. 1), S119–S124.
10. Smith, B.M., Laberge, J.M., Schreiber, R., Weber, A.M. and Blanchard, H. (1991) Familial biliary atresia in three siblings including twins. *J. Pediatr. Surg.*, **26**, 1331–1333.
11. Cunningham, M.L. and Sybert, V.P. (1988) Idiopathic extrahepatic biliary atresia: recurrence in sibs in two families. *Am. J. Med. Genet.*, **31**, 421–426.
12. Arikan, C., Berdeli, A., Kilic, M., Tumgor, G., Yagci, R.V. and Aydogdu, S. (2008) Polymorphisms of the ICAM-1 gene are associated with biliary atresia. *Dig. Dis. Sci.*, **53**, 2000–2004.
13. Shih, H.H., Lin, T.M., Chuang, J.H., Eng, H.L., Juo, S.H., Huang, F.C., Chen, C.L. and Chen, H.L. (2005) Promoter polymorphism of the CD14 endotoxin receptor gene is associated with biliary atresia and idiopathic neonatal cholestasis. *Pediatrics*, **116**, 437–441.

14. Silveira, T.R., Salzano, F.M., Donaldson, P.T., Mieli-Vergani, G., Howard, E.R. and Mowat, A.P. (1993) Association between HLA and extrahepatic biliary atresia. *J. Pediatr. Gastroenterol. Nutr.*, **16**, 114–117.
15. Donaldson, P.T., Clare, M., Constantini, P.K., Hadzic, N., Mieli-Vergani, G., Howard, E. and Kelley, D. (2002) HLA and cytokine gene polymorphisms in biliary atresia. *Liver*, **22**, 213–219.
16. Davenport, M., Gonde, C., Redkar, R., Koukoulis, G., Tredger, M., Mieli-Vergani, G., Portmann, B. and Howard, E.R. (2001) Immunohistochemistry of the liver and biliary tree in extrahepatic biliary atresia. *J. Pediatr. Surg.*, **36**, 1017–1025.
17. Broome, U., Nemeth, A., Hultcrantz, R. and Scheynius, A. (1997) Different expression of HLA-DR and ICAM-1 in livers from patients with biliary atresia and Byler's disease. *J. Hepatol.*, **26**, 857–862.
18. Mack, C.L., Falta, M.T., Sullivan, A.K., Karrer, F., Sokol, R.J., Freed, B.M. and Fontenot, A.P. (2007) Oligoclonal expansions of CD4+ and CD8+ T-cells in the target organ of patients with biliary atresia. *Gastroenterology*, **133**, 278–287.
19. Carvalho, E., Liu, C., Shivakumar, P., Sabla, G., Aronow, B. and Bezerra, J.A. (2005) Analysis of the biliary transcriptome in experimental biliary atresia. *Gastroenterology*, **129**, 713–717.
20. Shivakumar, P., Sabla, G., Mohanty, S., McNeal, M., Ward, R., Stringer, K., Caldwell, C., Chougnet, C. and Bezerra, J.A. (2007) Effector role of neonatal hepatic CD8+ lymphocytes in epithelial injury and autoimmunity in experimental biliary atresia. *Gastroenterology*, **133**, 268–277.
21. Bezerra, J.A., Tiao, G., Ryckman, F.C., Alonso, M., Sabla, G.E., Shneider, B., Sokol, R.J. and Aronow, B.J. (2002) Genetic induction of proinflammatory immunity in children with biliary atresia. *Lancet*, **360**, 1653–1659.
22. Shivakumar, P., Campbell, K.M., Sabla, G.E., Miethke, A., Tiao, G., McNeal, M.M., Ward, R.L. and Bezerra, J.A. (2004) Obstruction of extrahepatic bile ducts by lymphocytes is regulated by IFN-gamma in experimental biliary atresia. *J. Clin. Invest.*, **114**, 322–329.
23. Chen, L., Goryachev, A., Sun, J., Kim, P., Zhang, H., Phillips, M.J., Macgregor, P., Lebel, S., Edwards, A.M., Cao, Q. *et al.* (2003) Altered expression of genes involved in hepatic morphogenesis and fibrogenesis are identified by cDNA microarray analysis in biliary atresia. *Hepatology*, **38**, 567–576.
24. Citterio, L., Tizzoni, L., Catalano, M., Zerbini, G., Bianchi, G. and Barlassina, C. (2003) Expression analysis of the human adducin gene family and evidence of ADD2 beta4 multiple splicing variants. *Biochem. Biophys. Res. Commun.*, **309**, 359–367.
25. Ku, N.O., Zhou, X., Toivola, D.M. and Omary, M.B. (1999) The cytoskeleton of digestive epithelia in health and disease. *Am. J. Physiol.*, **277**, G1108–G1137.
26. Oshio, C. and Phillips, M.J. (1981) Contractility of bile canaliculi: implications for liver function. *Science*, **212**, 1041–1042.
27. Segawa, O., Miyano, T., Fujimoto, T., Watanabe, S., Hirose, M. and Fujiwara, T. (1993) Actin and myosin deposition around bile canaliculi: a predictor of clinical outcome in biliary atresia. *J. Pediatr. Surg.*, **28**, 851–856.
28. Shteyer, E., Ramm, G.A., Xu, C., White, F.V. and Shepherd, R.W. (2006) Outcome after portoenterostomy in biliary atresia: pivotal role of degree of liver fibrosis and intensity of stellate cell activation. *J. Pediatr. Gastroenterol. Nutr.*, **42**, 93–99.
29. Nagasaka, T., Nomura, S., Okamura, M., Tsujimoto, M., Nakazato, H., Oiso, Y., Nakashima, N. and Mizutani, S. (1997) Immunohistochemical localization of placental leucine aminopeptidase/oxytocinase in normal human placental, fetal and adult tissues. *Reprod. Fertil. Dev.*, **9**, 747–753.
30. Cottrell, G.S., Coelho, A.M. and Bunnett, N.W. (2002) Protease-activated receptors: the role of cell-surface proteolysis in signalling. *Essays Biochem.*, **38**, 169–183.
31. Sharma, J.N. (2009) Hypertension and the bradykinin system. *Curr. Hypertens. Rep.*, **11**, 178–181.
32. Zhao, A., Lew, J.L., Huang, L., Yu, J., Zhang, T., Hrywna, Y., Thompson, J.R., De, P.N., Blevins, R.A., Pelaez, F. *et al.* (2003) Human kininogen gene is transactivated by the farnesoid X receptor. *J. Biol. Chem.*, **278**, 28765–28770.
33. Yang, H., Plosch, T., Lisman, T., Gouw, A.S., Porte, R.J., Verkade, H.J. and Hulscher, J.B. (2009) Inflammation mediated down-regulation of hepatobiliary transporters contributes to intrahepatic cholestasis and liver damage in murine biliary atresia. *Pediatr. Res.*, **66**, 380–385.
34. Tronche, F. and Yaniv, M. (1992) HNF1, a homeoprotein member of the hepatic transcription regulatory network. *Bioessays*, **14**, 579–587.
35. Gachon, F., Olela, F.F., Schaad, O., Descombes, P. and Schibler, U. (2006) The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification. *Cell Metab.*, **4**, 25–36.
36. Bodmer, W. and Bonilla, C. (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.*, **40**, 695–701.
37. Garcia-Barcelo, M.M., Tang, C.S., Ngan, E.S., Lui, V.C., Chen, Y., So, M.T., Leon, T.Y., Miao, X.P., Shum, C.K., Liu, F.Q. *et al.* (2009) Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease. *Proc. Natl Acad. Sci. USA*, **106**, 2694–2699.
38. Yeung, J.M., Sham, P.C., Chan, A.S. and Cherny, S.S. (2008) OpenADAM: an open source genome-wide association data management system for Affymetrix SNP arrays. *BMC Genomics*, **9**, 636.
39. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
40. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.