Journal of Pediatric Surgery 51 (2016) 1661–1664

Contents lists available at ScienceDirect

Journal of Pediatric Surgery

journal homepage: www.elsevier.com/locate/jpedsurg

Common genetic variants of GPC1 gene reduce risk of biliary atresia in a Chinese population

Juntao Ke [a,b], Shuaidan Zeng [a], Jianxiong Mao [a], Jianyao Wang [a], Jiao Lou [b], Jiaoyuan Li [b], Xueqin Chen [b], Cheng Liu [b], Liu-Ming Huang [c,*], Bin Wang [a,**], Lei Liu [a,**]

[a] Department of General Surgery, Shenzhen Children Hospital, Shenzhen, China
[b] State Key Laboratory of Environment Health (Incubation), MOE (Ministry of Education) Key Laboratory of Environment & Health, Ministry of Environmental Protection Key Laboratory of Environment and Health (Wuhan), and Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China
[c] Department of Pediatric Surgery, BaYi Children's hospital, The military general hospital of Beijing, Beijing, China

ARTICLE INFO

Article history:
Received 8 January 2016
Received in revised form 18 April 2016
Accepted 11 May 2016

Key words:
Biliary atresia
GPC1
Common genetic variants
Chinese population

ABSTRACT

Background: Biliary atresia (BA) is a major neonatal cholestatic disease and main indication for pediatric liver transplantation in the world. Recently, *GPC1* has been implicated as a risk gene for BA by genetic studies and follow-up functional experiments on zebrafish.
Methods: Two common genetic variants of *GPC1*, rs2292832 and rs3828336, were selected systematically through 'SNPinfo', and were examined using TaqMan Genotyping Assays for association studies in a Chinese population containing 134 cases and 618 controls.
Results: Of the two single nucleotide polymorphisms (SNPs), we found a significantly decreased BA risk associated with rs2292832 (additive model: OR = 0.638, 95% CI: 0.467–0.873, $P = 0.005$), and a marginal effect for rs3828336 (heterozygous model: OR = 0.564, 95% CI: 0.312–1.020, $P = 0.058$). The haplotype analysis indicated that either $C_{rs2292832}$-$C_{rs3828336}$&$T_{rs3828336}$ or $T_{rs2292832}$-$T_{rs3828336}$ conferred a protective effect from BA (OR = 0.569, 95% CI = 0.414–0.783, $P < 0.001$; OR = 0.528, 95% CI: 0.301–0.926, $P = 0.026$). Moreover, bioinformatics analysis suggested that rs2292832 altered *GPC1* expression via effect on transcription-factor-binding sites (TFBS) of upstream binding transcription factor (UBTF), as a regulatory DNA variation in Deoxyribonuclease I (DNase I) hypersensitive sites (DHSs).
Conclusion: Common variants of *GPC1* gene were genetically involved in BA risk.

Biliary atresia (BA) is a devastating neonatal disease characterized by progressive fibroinflammatory obstruction of the extrahepatic biliary tree, leading to cholestasis, fibrosis, and cirrhosis. Without medical and surgical intervention such as Kasai portoenterostomy, an operation routinely performed to reconstruct biliary system but still failed in more than half of children [1], BA results in liver failure and death by the age of two years [2,3]. It is a rare disorder, with an incidence of 1:15,000–19,000 live births of Caucasian [4–7], whereas the Asian incidence is obviously higher, ranging from 1:5400 to 5800 live births in Chinese [8].

The etiology of BA is largely unknown. But it is supposed to be caused by exposure of genetically susceptible infant to environmental factors, exampled by virus infections [9] or toxins [10]. Several genes, including *CFC1*[11], intercellular adhesion molecule-1 (*ICAM1*) [12], macrophage migration inhibitory factor gene (*MIF*) [13], CD14 endotoxin receptor gene [14], and upstream stimulatory factor 2 (*USF2*) [15], have been suggested to play a role in BA pathogenesis. A recent genetic study of 35 BA cases and 2026 controls identified a potential region of interest at 2q37.3 [16]. Cui et al. [17] extended the study of copy number variations (CNV) in 61 cases and 5088 controls, detected heterozygous deletions spanning the same locus in 6 subjects and narrowed the region to include only one gene *GPC1*. Moreover, they followed up with functional analysis in an animal model zebrafish, which showed that disruption of *GPC1* led to biliary defects involving overactivation of Hedgehog signaling.

Given that *GPC1* has been identified as a risk gene for BA by Cui et al., we hypothesized that not only rare genomic indels but also common variants (minor allele frequency, MAF > 0.05) of *GPC1* could constitute a genetic basis for susceptibility to BA. So we selected two single nucleotide polymorphisms (SNPs), rs2292832 and rs3828336, in *GPC1* gene through an integrated bioinformatics tool 'SNPinfo' [18] (http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm) where researchers can choose SNPs based on predicted functional characteristics. Afterwards, we investigated their association with BA via a case–control study of 134 cases and 618 controls, and attempted to explain the potential pathogenic role of positive SNP by further bioinformatics analysis. To the best of our knowledge,

* Correspondence to: L.M. Huang, Department of Pediatric Surgery, BaYi Children's hospital, The military general hospital of Beijing, 100700, Beijing, China.
** Corresponding authors at: Department of General Surgery, Shenzhen Children's Hospital, 518026 Shenzhen, Guangdong, China.
E-mail addresses: surhlm@126.com (L.-M. Huang), szwb1967@126.com (B. Wang), liulei3322@aliyun.com (L. Liu).

http://dx.doi.org/10.1016/j.jpedsurg.2016.05.009
0022-3468/

this is the first study on common genetic variants in *GPC1* and BA risk in a Chinese population.

1. Material and methods

1.1. Study population

This study contained 134 children (males = 65; females = 69) diagnosed with BA and without other associated congenital malformation by laparoscopic cholangiography and biopsy of liver and extrahepatic biliary tree, all of which were consecutively enrolled between 2010 and 2013 from Shenzhen Children's Hospital, China. As controls, 618 healthy individuals (males = 303; females = 315) of southern Chinese without diagnosis of BA, congenital disease or liver disease were included. All subjects were unrelated ethnic Han Chinese. Written informed consent has been obtained from each subject or their legal guardians during the enrollment. This study was approved by the institutional review board (IRB) of Shenzhen Children's Hospital.

1.2. SNP selection

We screened SNPs in the *GPC1* gene by a web-base SNP selection tool called 'SNPinfo' which integrates GWAS and candidate gene information into functional SNP selection for association studies [18], as the procedures described blow. Firstly, we extracted the range of the physical position of *GPC1* (chr 2: 241,023,788 ~ 241,056,165) and its upstream and downstream 2Kb range (chr 2: 241,021,788 ~ 241,058,165) from HapMap database (http://hapmap.ncbi.nlm.nih.gov/, HapMap Data Rel 24/phaseII Nov08, on NCBI B36 assembly, dbSNP b126). Secondly, we input the extensive range of the gene into 'SNPinfo' (http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm) and received a list of SNPs with possible functions. Thirdly, we sifted out the SNP whose minor allele frequency (MAF) of Han Chinese in Beijing, China (CHB) is greater than 5%, and got a total of 7 SNPs (Table S1 in the online version at http://dx.doi.org/10.1016/j.jpedsurg.2016.05.009). Fourthly, we prioritized 3 most possibly functional SNPs (rs3828336, rs2292832, rs2228331) which harbor more than one functional motifs (marked with a Y letter standing for 'Yes'), such as transcription-factor-binding sites (TFBS), splicing site, microRNA-binding site, non-synonymous SNP (nsSNP). Finally, after we deleted rs2228331, an nsSNP that was classified as 'benign' by Polyphen [19], rs2292832 and rs3828336 were left for genotyping in our case–control study. To avoid redundancy, we analyzed the LD between these two SNPs by Haploview v4.2 [20], and confirmed that rs2292832 and rs3828336 are not in LD ($r^2 = 0.000$).

1.3. Genotyping

Genomic DNA was extracted from peripheral blood leukocytes or liver tissues of BA children, and from peripheral blood of healthy controls, using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) by reference to the manufacturer's instructions. Both SNPs were genotyped with the TaqMan SNP Genotyping Assay (Applied Biosystems, Foster City, CA, USA) on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). 5% duplicated samples were randomly selected to assess the reproducibility for quality control, with a concordance rate of 100%.

1.4. Statistical analysis

The χ^2 test was applied to estimated differences in variables and distributions of genotypes between cases and controls. Hardy–Weinberg equilibrium was evaluated using the goodness-of-fit χ^2 test in controls and a value of $P < 0.05$ was considered as significant disequilibrium. The association between the case–control status and each SNP was measured by the odds ratio (OR) and its corresponding 95% confidence interval (95% CI). In order to avoid the assumption of genetic models, codominant, dominant and additive models were analyzed. The LD of candidate SNPs was analyzed using Haploview v4.2 [20] and haplotype were constructed by the PHASE v2.1 software [21]. The ORs and corresponding 95% CIs, adjusted by gender, were calculated by unconditional multivariate logistic regression. Statistical analyses were performed using SPSS Software v20.0 (SPSS, Chicago, Illinois, USA), $P < 0.05$ was considered statistically significant.

1.5. Bioinformatics analysis

To try to explain the biological possibilities underlying significant SNPs associated with BA in details, bioinformatics analysis was conducted on a online database 'UCSC' [22] (http://genome.ucsc.edu/) that has been updated with many new annotation data sets including 'Transcription Factor ChIP-seq Clusters' and the 'DNaseI Hypersensitivity Clusters' from ENCODE [23].

2. Results

2.1. Population characteristic

A total of 134 children with BA and 618 controls were enrolled in this study. 130 (97%) cases for rs2292832, 133 (99%) cases for rs3828336 and all controls were genotyped successfully. The genotype distributions of rs2292832 and rs3828336 in our controls are both in Hardy–Weinberg equilibrium (HWE, $P = 0.52$ and $P = 0.14$). And there was no statistically significant difference in gender distribution between cases and controls ($P = 0.913$, Pearson $\chi^2 = 0.012$).

2.2. Association analysis

The rs2292832 showed significant association in four genetic models, the other rs3828336 presented no obvious association with BA under any model. Detailed genotype frequencies of rs2292832 and rs3828336 are shown in Table 1. Under multivariate logistic regression model adjusted for sex, individuals with *CT* genotype of rs2292832 had a significantly decreased risk of BA (OR = 0.339, 95% CI = 0.214–0.535, $P < 0.001$) compared to those with *TT* homozygote. A dominant model was performed to increase statistical power by combining the *CT* with *CC* into a *C*-carrier group (*CT* plus CC), and it showed that the allele *C* carriers got an obviously lower risk (OR = 0.43, 95% CI = 0.289–0.641, $P < 0.001$). Likewise, the *C* allele presented protective effect for BA (OR = 0.629, 95% CI = 0.459–0.863, $P = 0.004$), while the *C* allele frequency in our controls was 31.3%, approximate to the MAF 26.7% of CHB from HapMap. Another positive result was found under additive model, with per-*C*-allele OR of 0.638 (95% CI = 0.467–0.873, $P = 0.005$).

As for rs3828336, we observed no statistically significant difference in genotype ($P = 0.125$, Pearson $\chi^2 = 4.162$) and allele ($P = 0.115$, Pearson $\chi^2 = 2.487$) distributions when comparing cases with controls, and found null associations with BA in all genetic models we studied (Table 1). Still, a marginal effect was observed in the heterozygous model (OR = 0.564, 95% CI: 0.312–1.020, $P = 0.058$).

2.3. Haplotypes analysis

As shown in Table 2, there is significant difference in the distribution of haplotype frequencies between cases and controls ($P < 0.001$, Pearson $\chi^2 = 15.423$). Compared with the $T_{rs2292832}$-$C_{rs3828336}$ that consists of the two wild alleles, haplotypes containing $C_{rs2292832}$ or $T_{rs3828336}$ allele were associated with decreased risk of BA. Of note, we combined $C_{rs2292832}$-$C_{rs3828336}$ and $C_{rs2292832}$-$T_{rs3828336}$ into a subgroup that containing the significant risk allele $C_{rs2292832}$ to increase statistical power, because of the extremely low frequencies of the $C_{rs2292832}$-$T_{rs3828336}$ in both cases and controls (0.4% and 0.2%). The sex-adjusted ORs calculated by logistic regression for the $C_{rs2292832}$-$C_{rs3828336}$&$T_{rs3828336}$ and