A genome-wide association study identifies six novel risk loci for primary biliary cholangitis

Fang Qiu1, Ruqi Tang2, Xianbo Zuo3, Xingjuan Shi1, Yiran Wei2, Xiaodong Zheng3, Yaping Dai4, Yuhua Gong5, Lan Wang6, Ping Xu7, Xiang Zhu7, Jian Wu8, Chongxu Han9, Yueqiu Gao10, Kui Zhang11, Yuzhang Jiang12, Jianbo Zhou13, Youlin Shao14, Zhigang Hu15, Ye Tian16, Haiyan Zhang2, Na Dai17, Lei Liu18, Xudong Wu19, Weifeng Zhao3, Xiaomin Zhang3, Zhidong Zang21, Jinshie Nee22, Weihao Sun23, Yi Zhao24, Yuan Mao25, Po Jiang26, Hualiang Ji27, Qing Dong28, Junming Li29, Zhenzhong Li30, Xinli Bai30, Li Li31, Maosong Lin32, Ming Dong3, Jinxin Li1, Ping Zhu1, Chan Wang1, Yanqi Zhang1, Peng Jiang3, Yujue Wang1, Rohil Jawed1, Jing Xu1, Yu Zhang1, Qixia Wang2, Yue Yang2, Fan Yang2, Min Lian2, Xiang Jiang2, Xiao Xiao2, Yanmei Li2, Jingyuan Fang2, Dekai Qiu2, Zhen Zhu14, Hong Qiu6, Jianqiong Zhang1, Wenyang Tian8, Sufang Chen7, Ling Jiang8, Bing Ji8, Ping Li9, Guochang Chen18, Tianxue Wu9, Yan Sun9, Jianxiang Yu13, Huijuan Tang13, Michun He8, Min Xia15, Hao Pei4, Lihua Huang4, Zhuye Qing25, Jianfang Wu23, Qinghai Huang1, Junhai Han1, Wei Xie1, Zhongsheng Sun3, Jian Guo35, Gengsheng He26, M. Eric Gershwin37, Zhexiong Lian38, Xiang Liu39, Michael F. Seldin37, Xiangdong Liu1, Weichang Chen8 & Xiong Ma2

Primary biliary cholangitis (PBC) is an autoimmune liver disease with a strong hereditary component. Here, we report a genome-wide association study that included 1,122 PBC cases and 4,036 controls of Han Chinese descent, with subsequent replication in a separate cohort of 907 PBC cases and 2,127 controls. Our results show genome-wide association of 14 PBC risk loci including previously identified 6p21 (HLA-RA and DPB1), 17q12 (ORMDL3), 3q13.33 (CD80), 2q32.3 (STAT1/STAT4), 3q25.33 (IL2), 4q24, 14q11.2 (HLA-DRA and DPB1), and 17q12 (ORMDL3). We also identified variants in IL21, IL2R, CD28/CTLA4/ICOS, CD58, ARID3A and IL6 as novel PBC risk loci. These new findings and histochemical studies showing enhanced expression of IL21 and IL2R in PBC livers (particularly in the hepatic portal tracts) support a disease mechanism in which the deregulation of the IL21 signaling pathway, in addition to CD4 T-cell activation and T-cell co-stimulation are critical components in the development of PBC.
Primary biliary cholangitis (PBC) is an autoimmune disease characterized by chronic inflammation and destruction of intrahepatic bile ducts. It is often considered a model autoimmune disease because of its specific autoantibodies (the anti-mitochondrial antibodies) and distinctive bile duct pathology. PBC shows a strong genetic predisposition, with the highest concordance rate of all autoimmune diseases among identical twins. Several genome-wide association studies (GWAS) and Immunochip studies have revealed a nearly uniform level of genetic susceptibility among people (or populations) of different European ancestries, showing significant association of more than 20 loci with PBC. However, a PBC GWAS in Japanese (including 487 PBC cases and 476 controls and replication in 787 cases and 615 controls) showed very different results identifying only TNFSF15 and POU2AF1 as the main susceptibility loci, neither of which was described in the European studies.

Recently, we performed a replication study of the 14 most significant single nucleotide polymorphisms (SNPs) identified in either or both of the European and Japanese cohorts, using 1,070 Han Chinese PBC cases and 1,198 healthy controls, and confirmed strong association of the TNFSF15, CD80 and 17q12 loci with the disease in Chinese patients, with P-value reaching GWA significance ($P < 5 \times 10^{-8}$) at TNFSF15 and CD80 loci. These early results also ensured required samples for our subsequent GWAS design.

To identify additional genetic risk factors for PBC and further define differences in susceptibility to the disease between European and East Asian populations, we conducted a large case–control GWAS of 1,127 PBC subjects and 4,074 controls of self-declared Han Chinese origin with population-specific HumanOmniZhongHua-8 beadchip and verified the results by genotyping top SNPs in a separate cohort of 907 PBC cases and 2,127 controls. We confirmed eight previously identified PBC susceptibility loci and also identified variants in IL21, IL21R, CD28/CTLA4/ICOS, CD58, ARID3A and IL16 as novel PBC risk loci. Subsequent histochemical studies show enhanced expression of IL21 and IL21R in PBC livers. These findings significantly expand our understanding of the disease susceptibility and suggest IL21 signalling pathway, in addition to CD4 T-cell activation and T-cell co-stimulation, are critical components in the development of PBC.

**Results**

**Genome-wide discovery analysis.** The HumanOmniZhongHua-8 beadchip (v1.1) contains 894,956 markers covering common, intermediate and rare variations found within Chinese populations. Following stringent quality control measures, 776,516 autosomal SNPs were available across 1,122 cases and 4,036 geographically matched controls (Supplementary Tables 1–3; Supplementary Fig. 1). Principal-component analysis (PCA) did not demonstrate substantial stratification in our study population ($\lambda_{gc} = 1.029$, and see Supplementary Fig. 2). The quantile–quantile plot of the case–control test showed a substantial excess of significant associations in the tail of the distribution even after removal of the major histocompatibility complex (MHC) region that includes human leukocyte antigen (HLA) genes (Supplementary Fig. 3).

We identified 179 SNPs in the MHC region and 39 SNPs in 8 non-MHC loci exceeding genome-wide significance in the discovery stage (Fig. 1; Supplementary Tables 4–7). The most significant association was found in the HLA-DRA locus. A SNP within HLA-DRPB1 was the second most significant locus in the MHC region. Two SNPs from the MHC region and 32 SNPs from 24 non-MHC loci with at least suggestive association ($P < 5 \times 10^{-8}$) were selected for replication studies (Supplementary Table 6). There were 907 PBC cases and 2,127 controls in the replication panels that confirmed association of the MHC region and 15 non-MHC loci. Each of these loci in the combined analysis (2,029 PBC cases and 6,163 controls) also met genome-wide significance. These included 8 non-MHC risk loci previously identified in Caucasian and Japanese cohorts (Supplementary Fig. 4). In addition, a previously unidentified risk locus, DDX6-CXCR5, did not replicate but still achieved genome-wide significance ($P = 2.56 \times 10^{-13}$) in the combined analysis (Table 1; Supplementary Fig. 4).

Among the non-MHC risk loci, the strongest association was observed in rs4979467, a SNP in the TNFSF15-TNFSF8 locus, the same as reported in the Japanese cohort. This locus has been implicated in several autoimmune diseases. Two previous studies have reported that two functional SNPs, rs6478108 and rs4979462, in the TNFSF15 promoter region are associated with TNFSF15 expression. The rs4979467 SNP is in strong linkage disequilibrium (LD) with rs6478108 ($r^2 = 1$) and rs4979462 ($r^2 = 0.68$) in the Han Chinese population, implying that the...
functional variation in the TNFSF15 promoter can affect the genetic susceptibility to PBC.

IL21A was found to be strongly associated with Han Chinese PBC cohorts, in contrast to both the Japanese GWAS and our previous candidate gene study.\textsuperscript{6,9} On the basis of the dense fine-mapping study with the Immunochip, four independent associations were observed at the IL21A-SCHPI1 locus in European ancestry PBC cases.\textsuperscript{6} The significant association (rs582537) in our GWAS corresponded to the third signal rs669987 in the European studies (Supplementary Table 8). This association also explained the discrepancy with the previous candidate gene study, in which selected SNPs represented only the first signal at the IL21A locus.

**Discovery and validation of six novel PBC risk loci.** Six novel risk loci associated with PBC were identified and validated in the current study (Table 2; Fig. 2). Both IL21R and IL21 were strongly associated in the Han Chinese PBC cohorts. Notably, several recent studies have found increased production of IL21 from T follicular helper (Tfh) cells in PBC patients, which might mediate B-cell maturation and autoantibody secretion.\textsuperscript{13,14} Multiple SNPs in the IL21 and IL21R loci were significantly associated with PBC (Supplementary Table 5). We identified two independent signals represented by variants rs925550 and rs17005934 in the IL21 locus using stepwise conditional regression analysis with the two SNPs (Supplementary Table 9). These two signals were separated in the Han Chinese population as a result of a major recombination event (Fig. 2c). A much weaker association (rs108507092) observed in the nearby IL2 region was presumably due to linkage disequilibrium ($r^2 = 0.64$) with rs17005934.

The most significant SNP, rs2189521, at the IL21R locus is located in the 5' UTR of IL21R, indicating its potential involvement in regulating differential IL21R expression. IL4R is located 27 kb upstream from IL21R. To analyse if the significance came from the contribution of IL4R, we performed LDL analysis with the Han Chinese 1000 genome sequencing data. The results indicate that IL21R and IL4R are in two separate LD blocks in the Han population (Fig. 2a). None of the significant SNPs observed at the IL21R locus are linked to IL4R, excluding the association of IL4R with Han PBC cohorts. The variant rs2189521 is an expression quantitative trait locus (eQTL), in strong LD ($r^2 = 0.98$) with multiple SNPs in the IL21R region, suggesting

### Table 1 | Known PBC loci reaching genome-wide significance.

| Chr | SNP    | MA  | Position (bp) | Genes in the region | GWAS cohort | Replication cohort | Combined samples |
|-----|--------|-----|---------------|---------------------|-------------|-------------------|-----------------|
|     |        | 98  |               |                     | Case        | Control           | MAF             |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |

### Table 2 | Novel PBC loci reaching genome-wide significance.

| Chr | SNP    | MA  | Position (bp) | Genes in the region | GWAS cohort | Replication cohort | Combined samples |
|-----|--------|-----|---------------|---------------------|-------------|-------------------|-----------------|
|     |        | 98  |               |                     | Case        | Control           | MAF             |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | P-value     | MAF (95% CI)      | MAF (95% CI)    |
|     |        |     |               |                     | OR           | MAF (95% CI)      | MAF (95% CI)    |

Chr, chromosome; MA, minor allele; MAF, minor allele frequency. $P$-value is based on additive model; OR, odds ratio, is calculated for minor allele.

*Sequence position was annotated based on the GRCh37/hg19 assembly.
that variation in *IL21R* expression may explain this association (Supplementary Tables 10 and 11). We also note that in a recent study of European PBC the genomic region containing *IL21R* while not achieving significant association did show suggestive association ($P = 1.6 \times 10^{-7}$) (ref. 5).

Multiple SNPs across the *CD28-CTLA4-ICOS* locus reaching genome-wide significance were found at the discovery stage (Supplementary Table 5) and further confirmed by the replication study. On the basis of the functional role of all the three genes in T-cell costimulatory pathways, the *CD28-CTLA4-ICOS* locus has long been suggested by many candidate gene studies as a PBC risk locus, but large-scale GWAS and dense fine-mapping studies in European PBC cohorts have not revealed this association$^{2–7}$. An earlier study in a small Chinese PBC cohort also suggested association of the *CTLA4* variant rs231775 with PBC$^{15}$. Our GWAS study is the first to report the genome-wide significance of the *CD28-CTLA4-ICOS* locus with PBC. The most significant signal was obtained for the intergenic variant rs4675369. Haplotype analysis using the Han Chinese genome sequencing data identified a major recombination breakpoint between *CD28* and *CTLA4*. In addition, rs4675369, together with five other highly linked SNPs ($r^2 > 0.9$), was located in the LD block.
containing CD28, indicating that the observed significance may be associated with CD28, rather than CTLA4 or ICOS (Fig. 2b).

Association of the CD28-CTLA4-ICOS locus with primary sclerosing cholangitis (PSC) was also reported previously in a Immunochip analysis16. However, the variant (rs7426056) associated with PSC is not in LD with any significant variants identified in our study.

We identified significant association of the CD58 locus with PBC. The SNP showing the strongest association, rs2300747 has been previously associated with multiple sclerosis (MS) and rheumatoid arthritis (RA) in large-scale GWAS and replication studies17,18. The 1000 genome sequencing data from the Han Chinese population indicated two major haplotype blocks in the CD58 gene. The first haplotype block includes part of CD58 intron-1, exon-1 and the 5’ regulatory region and the nearby IGSF3 gene. The second haplotype block includes most of intron-1 and the remaining CD58 region (Fig. 2d). The variant rs2300747 and multiple SNPs in high LD ($r^2 > 0.9$) in the block are eQTLs in lymphoblastoid cell lines (LCLs) (Supplementary Tables 8 and 9). A previous study found that the A allele of rs2300747 was associated with lower expression of CD58 (ref. 19).

In our PBC cohort, the A allele of rs2300747 is the risk allele, suggesting that decreased expression of CD58 levels may be related to PBC susceptibility.

Another significant association found in our study, rs10415976, is located in the intronic region of ARID3A, a B-cell-restricted transcription factor that activates immunoglobulin heavy chain gene transcription. During the discovery stage of GWAS, five SNPs at the ARID3A locus showed significant association, two of them reaching genome-wide significance. These two SNPs were further confirmed in the replication study. All of the significant SNPs, including SNPs in LD ($r^2 > 0.8$), were located in the intronic region and identified as cis eQTL loci (Supplementary Tables 10 and 11). ARID3A is the first B-cell-restricted
transcription factor demonstrated to induce autoimmunity\textsuperscript{20}. Transgenic mice constitutively expressing ARID3A in B lineage cells produce anti-nuclear antibodies. Systemic lupus erythematosus (SLE) patients show dramatically increased numbers of ARID3A\textsuperscript{+} B cells compared to healthy controls\textsuperscript{21}. Our study is the first to report that the ARID3A locus is associated with human autoimmune of inflammatory disease. Further studies should be aimed at investigating its functional role in the pathogenesis of PBC.

In our GWAS discovery stage, seven SNPs in the IL16 region showed strong association ($P = 9.42 \times 10^{-6}$) with PBC. One of the variants, rs4778636, was selected in the replication study and reached genome-wide significance ($P = 8.99 \times 10^{-10}$). IL16 is a T-cell chemo-attractant factor and has been linked to RA, MS and Crohn’s disease\textsuperscript{22}. However, the pathophysiological role of IL16 in PBC remains unclear. IL16 has been found to be constitutively expressed at the mRNA and protein level in monocytes, T lymphocytes, mast cells, eosinophils and dendritic cells. Several of the associated IL16 variants are cis eQTL in peripheral monocytes (Supplementary Tables 10 and 11). This is the first report that IL16 is a disease susceptibility locus with genome-wide significance.

There was also a novel locus CSNK2A2/CCDC113 that did not meet significance in the replication study after correction for multiple comparison, but was significant by genome-wide criteria ($P = 1.51 \times 10^{-8}$) in the combined analysis (Table 2).

**Aberrant expression of IL21/IL21R in PBC liver lesion.** To further explore whether deregulated IL21 and IL21R contribute to the pathogenesis of PBC, we performed histochemical analysis with liver biopsy samples. Significantly elevated levels of expression of IL21 and IL21R was observed in the livers of PBC patients, compared with those of chronic hepatitis B (CHB) patients, autoimmune liver hepatitis (AIH) patients or healthy controls (HC). IL21 is extensively expressed by inflamed hepatocytes and infiltrating inflammatory cells around the portal tracts in PBC patients (Fig. 3a–d), while IL21R cells were generally aggregated in the inflamed portal tracts, especially around the damaged interlobular bile ducts in PBC (Fig. 4a–d). Notably, the numbers of IL21 cells and IL21R cells were positively correlated with inflammation severity ($r = 0.45$, $P < 0.05$; $r = 0.43$, $P < 0.05$; respectively) and hepatic fibrosis stages ($r = 0.66$, $P < 0.01$; $r = 0.60$, $P < 0.01$; respectively) in PBC (Figs 3e–g and 4e–g). Therefore, there is marked activation of IL21/IL21R signalling in PBC patients, suggesting that IL21/IL21R interaction is an important component in the pathogenesis of PBC.

**Discussion**

Our PBC GWAS included 2,029 cases and 6,163 controls in a Han Chinese population and has identified six new PBC susceptibility loci. Our study supports the hypothesis that the IL21 signalling pathway and Tfh cells are involved in PBC pathogenesis. Our results also draw attention to the importance of T-cell costimulatory signalling in PBC and advance our understanding of PBC as a disease of immune deregulation.

**Methods**

**Patient and control selection and diagnosis.** PBC patients were recruited with the approval of the research ethics boards of Southeast University and Jiaotong University and in accordance with the guidelines of the Declaration of Helsinki (2008). Informed consent was obtained from all subjects recruited. PBC diagnosis in this study was based on the criteria recommended by AASLD and the European Association for the Study of the Liver (EAS)\textsuperscript{23,24}. PBC patients were recruited from two main sources, the Jiangsu Province PBC Collaboration Group (JSPPCG) and Renji Hospital of Jiaotong University. Patient blood samples from Renji Hospital were tested for PBC-specific autoantibodies by EUROLINE Profile Autoimmune Liver Diseases (DL1300-1601-5G). JSPPCG samples were collected from participating hospitals and PBC-specific autoantibodies were tested by different clinical laboratories. Samples from the JSPPCG collection were verified for autoantibody status with an in-house ELISA assay for anti-PDC-E2 and a 3E assay.

**GWAS analysis.** GWAS analysis has achieved great success in the understanding of PBC pathogenesis. Multiple PBC risk loci identified in this study were also shared risk loci in other autoimmune diseases (Supplementary Table 13). But several unique features were observed in our studies. Our IL21 results in particular have important implications and this gene might be worth exploring as a therapeutic target. IL21 is predominantly produced by Tfh cells although high levels are also noted in Th17 and natural killer (NK) cells. Th17 are characterized by expression of the cell surface marker CXCR5, encoded by the gene located at the PBC loci previously identified and strongly suggested by the current study\textsuperscript{4}. Binding of IL21 with its ligand IL21R leads to activation of multiple downstream signalling molecules, including STAT1 and STAT3, which is important for proliferation and differentiation of T-cells, B cells and NK cells. We also note that IL12 (both IL12A and IL12RB2 are associated with PBC) has been shown to enhance IL21 production and the frequency of IL21\textsuperscript{+} CD4\textsuperscript{+} T cells\textsuperscript{27}.

Two genes (CD80 and TNFSF15) encoding T-cell costimulatory molecules had been previously identified as PBC risk loci and were confirmed in our study. Identification of two more genetic loci encoding CD28/CTLA4/ICOS and CD58 provided further evidence that deregulated costimulatory signalling confers predisposition to the disease.

Association of IL16 locus with PBC may have significant application in PBC treatment. A previous animal study found that neutralizing anti-IL16 mAb can prevent insulitis and type 1 diabetes in mice\textsuperscript{28}. Neutralizing IL16 in PBC patients may represent a potential therapeutic approach.

In summary, this is the first report of a large scale GWAS of PBC in a Han Chinese population and has identified six new PBC susceptibility loci. Our study supports the hypothesis that the IL21 signalling pathway and Tfh cells are involved in PBC pathogenesis. Our results also draw attention to the importance of T-cell costimulatory signalling in PBC and advance our understanding of PBC as a disease of immune deregulation.
with a commercial ELISA kit (Shanghai Kexin Biotech, Shanghai, China). We confirm our study is compliant with the 'Guidance of the Ministry of Science and Technology (MOST) for the Review and Approval of Human Investigative Research', which requires formal approval for the export of human genetic material or data from China.

The control samples used in the discovery stage were recruited from previous studies. Healthy control samples used in replication stage were recruited in Nanjing, from the annual health check of staff members of Southeast University. All the control samples used in the replication study have been tested for AMA-M2 autoantibody using 3E-based ELISA assay. Only AMA-M2 negative samples were retained for study. To avoid potential population stratification, the vast majority of PBC patients and controls for GWA and replication studies originally came from Central China (Supplementary Fig. 1). All samples from both the discovery and replication stages were unrelated individuals of self-claimed Chinese Han descent. The demographic information of patients and controls were summarized in Supplementary Tables 1 and 2.

Quality control in the GWA discovery stage. The Illumina Human- OmniZhongHua-8 Beadchip (v1.1), a population-specific beadchip targeting common, intermediate, and rare variations found within Chinese populations, was used for both PBC and control samples. The Illumina beadscan data of 1,127 PBC samples was performed by Genenenergy Bio-Technology, an Illumina CSpro-certified service provider. Scanning of 4,074 control samples was conducted at the Key Laboratory of Dermatology at Anhui Medical University (Ministry of Education). PBC and control samples with overall call rates of <98% were excluded from further analysis. The discovery stage population outlier and stratification was performed on all the samples from Han Chinese origin with 30% minor allele frequency (MAF) based on pairwise identity-by-state comparisons using the ‘PI_HAT’ value in PLINK (all PI_HAT > 0.25)35. The remaining samples were subsequently assessed for population outlier and stratification using a PCA-based approach. PCA was performed using the smartpca software36. After PCA filtering, 1,122 cases and 1,026 controls were retained for further analysis. In the GWA stage, single-marker association analyses were performed using logistic regression with gender as a covariate. The Manhattan plot of $\log_{10} P$ was generated using Haplovie v4.2 (ref. 34). The Quantile–quantile (Q–Q) plot was used to assess the number and magnitude of observed associations between genotyped SNPs and PBC, compared with the association statistics expected under the null hypothesis of no association. The Q–Q plot was created using the R qqplot function.

We performed systematic quality control on the raw genotyping data to filter out both unqualified samples and SNPs. After excluding mitochondrial SNPs and SNPs on sex chromosomes, the following criteria were applied to disqualify additional SNPs: SNPs with a call rate < 98%, or with minor allele frequency (MAF) <0.01 in all samples or SNPs with genotype distributions that deviated from those expected by the Hardy–Weinberg equilibrium ($P < 1 \times 10^{-4}$) in the controls. For all the SNPs with $P < 1 \times 10^{-5}$ in the discovery stage, manual checking of SNP genotyping clustering in each dataset was performed and 54 SNPs were further excluded for the analysis. After quality control filtering, there were 776,516 SNPs remained in the discovery stage (Supplementary Table 3).

SNP selection and genotyping in the replication stage. SNPs for the replication stage were selected using the following criteria: each locus with one SNP reaching $P < 1 \times 10^{-6}$ or with minimum three SNPs reaching $P < 5 \times 10^{-8}$ for discovery samples. A total of 22 non-MHC loci matched these criteria, except TNFSF1A locus, which has three SNPs reaching $P < 6 \times 10^{-5}$. The most significant SNPs in HLA-DRA locus (rs9268644) and DPB1 locus (rs9051251) were selected for validation. For all the known PBC loci published previously, one of the most significant SNPs was selected for validation. We also intentionally selected one SNP (rs1944918) in POU2AF1 locus, since this locus was reported in Japanese GWAS as the second most significant locus for PBC. For each locus newly discovered, two of most significant SNPs were selected for validation. Genotyping analyses of replication were conducted by the Sequenom MassARRAY system using the iPLEX method35. All SNPs achieved a call rate >98% and had no deviation from HWE ($P > 0.05$) in the controls.

Statistical analysis in the replication and combined stage. For the replication studies, 34 SNPs that passed quality control were analysed using logistic regression with gender as a covariate, assuming an additive genetic effect. The $P$-values adjusted by gender were reported without correction for multiple testing. Joint analysis of all combined samples of Han Chinese origin was conducted by both the random effects model ($\tau^2 > 25\%$) and by the fixed-effect model ($\tau^2 \leq 25\%$). The $P$-value reported in Tables 1 and 2; Supplementary Table 6 was based the results by the fixed-effect model. The regional association plot was created using LocusZoom38. Imputation of selected loci was conducted using IMPUTE2 software version 2.2.2 and version 3 of 1000 Genomes Project data as the reference set37.

We used seeQTL (http://www.bios.uic.edu/research/genomic_software/seeQTL/) and the University of Chicago eQTL browser (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/?) to identify eQTLs amongst significant variants38–40. eQTLs that are in high linkage disequilibrium ($r^2 > 0.8$) with the most strongly associated SNP at the locus were extracted. Gene regulatory elements from the Encyclopedia of DNA Elements (ENCODE) database were annotated using HaploReg41. For each locus, SNPs in high linkage disequilibrium ($r^2 > 0.8$) with the most significantly associated SNP was assessed as to whether they lie within regions with promoter and enhancer histone marks, DNase-I hypersensitivity, protein binding or regulatory motifs in one or more of 147 cell types.

Immunohistochemistry. Paraffin-embedded liver tissues, derived from ultrasound-guided needle liver biopsies of 30 PBC patients, 30 AIH patients, 25 CHB patients and 6 controls, were studied. Sections were stained with either hematoxylin and eosin (H&E) or Masson and were independently reviewed 'blindly'. Immunofluorescence and fibrotic stages were graded according to the Scheuer scoring system. The six healthy liver tissues were collected from donors whose livers were subsequently used for liver transplantation. Immunohistochemistry imaging of IL21 and IL21R staining of human liver tissues was performed on a Leica Bond system (Leica, Germany) using the standard protocol. The liver sections were pre-treated using heat-mediated antigen retrieval with sodium citrate buffer (pH 6, epitope retrieval solution 1) for 20 min followed by incubation with an anti-IL21 antibody (1:200 dilution, ab5978, Abcam, Cambridge, MA, USA) or an anti-IL21R antibody (1:1350 dilution, ab13268, Abcam) for 20 min at room temperature and then detected using a horse radish peroxidase-conjugated compact polymer system. 3,3'-diaminobenzidine was used as the chromogen42. The liver sections were then counterstained with hematoxylin. All the sections were visualized using light microscopy (Olympus, Japan), and five fields were randomly selected for each section. The numbers of IL21 and IL21R-positive cells were quantified at 40 × 10 magnification. All the samples were analysed by a hepatic pathologist. All data are reported as mean ± standard error (s.e.). The Mann–Whitney U-test was used to evaluate differences in continuous variables. A $P$-value of <0.05 was considered statistically significant. Correlations were determined using the Spearman’s correlation coefficient. All analyses were two-tailed and performed using Prism software Version 6.0 (Graphpad Software, La Jolla, CA, USA).

Data availability. The GWA SNP results are available upon request by contacting X.L. at xiangdongliu@seu.edu.cn. Any additional data (beyond those included in the main text and Supplementary Information) that support the findings of this study are also available from the corresponding author upon request.

References

1. Selmi, C. et al. Primary biliary cirrhosis in monozygotic and dizygotic twins: genetics, epigenetics, and environment. Gastroenterology 127, 485–492 (2004).

2. Hirschfield, G. M. et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. N. Engl. J. Med. 360, 2544–2555 (2009).

3. Liu, X. et al. Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. Nat. Genet. 42, 658–660 (2010).

4. Mells, G. F. et al. Genome-wide association study identifies 12 new primary biliary cirrhosis loci and targetable pathogenic pathways. Nat. Commun. 6, 8019 (2015).

5. Liu, J. Z. et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nat. Genet. 44, 1137–1141 (2012).

6. Juran, R. D. et al. Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q41, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. Hum. Mol. Genet. 21, 5209–5221 (2012).

7. Nakamura, M. et al. Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. Am. J. Hum. Genet. 91, 721–728 (2012).

8. Dong, M. et al. Multiple genetic variants associated with primary biliary cirrhosis in a Han Chinese population. Clin. Rev. Allergy Immunol. 48, 316–321 (2015).

9. Alba, Y. & Nakamura, M. The role of TL1A and DR3 in autoimmune and inflammatory diseases. Mediators Inflamm. 2013, 258164 (2013).

10. Hitomi, Y. et al. Human primary biliary cirrhosis-susceptible allele of rs4979462 enhances TNFSF15 expression by binding NF-1. Hum. Genet. 134, 737–747 (2015).

11. Hedli, M. & Abraham, C. A. TNFSF15 disease-risk polymorphism increases pattern-recognition receptor-induced signaling through caspase-8-induced IL-1, Proc. Natl Acad. Sci. USA 111, 13451–13456 (2014).

12. Li, Y. et al. Chemokine (C-X-C motif) ligand 13 promotes intrahepatic chemokine (C-X-C motif) receptor 5+ lymphocyte homing and aberrant B-cell immune responses in primary biliary cirrhosis. Hepatology 61, 1998–2007 (2015).

13. Wu, J. et al. CXCR5 + CD4+ T follicular helper cells participate in the pathogenesis of primary biliary cirrhosis. Hepatology 61, 627–638 (2015).

14. Fan, L. Y. et al. Cytotoxic T lymphocyte associated antigen-4 gene polymorphisms confer susceptibility to primary biliary cirrhosis and...
autoimmune hepatitis in Chinese population. World J. Gastroenterol. 10, 3056–3059 (2004).
16. J. Z. et al. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. Nat. Genet. 45, 670–675 (2013).
17. Raychaudhuri, S. et al. Genetic variants at CD28, PRDM1 and CD2/CD8 are associated with rheumatoid arthritis risk. Nat. Genet. 41, 1313–1318 (2009).
18. Australia & New Zealand Multiple Sclerosis Genetics Consortium. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat. Genet. 41, 824–828 (2009).
19. De Jager, P. L. et al. The role of the CD58 locus in multiple sclerosis. Proc. Natl Acad. Sci. USA 106, 5264–5269 (2009).
20. Webb, C. F. et al. The ARID family transcription factor bright is required for both hematopoietic stem cell and B lineage development. Mol. Cell Biol. 31, 1041–1053 (2011).
21. Ward, J. M. et al. Disease activity in systemic lupus erythematosus correlates with expression of the transcription factor AT-rich-interactive domain 3A. Arthritis Rheumatol. 66, 3404–3412 (2014).
22. Richmond, J., Tuzova, M., Cruikshank, W. & Center, D. Regulation of cellular processes by interleukin-16 in homeostasis and cancer. J. Cell Physiol. 229, 139–147 (2014).
23. Jones, D. E. & Mells, G. The genetics of primary biliary cirrhosis: the revolution moves on. Hepatology 53, 362–364 (2011).
24. Yamaguchi-Kabata, Y. et al. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. Am. J. Hum. Genet. 83, 445–456 (2008).
25. Tian, C. et al. Analysis and application of European genetic substructure using 300 K SNP information. PLoS Genet. 4, e1000280 (2008).
26. Chen, J. et al. Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. Am. J. Hum. Genet. 85, 775–785 (2009).
27. Xiao L. et al. Human IL-21 + IFN-γ+ CD4+ T cells in nasal polyps are regulated by IL-12. Sci. Rep. 5, 12781 (2015).
28. Meagher, C. et al. Neutralization of interleukin-16 protects nonobese diabetic mice from autoimmune type 1 diabetes by a CCL4-dependent mechanism. Diabetes 59, 2862–2871 (2010).
29. Lindor, K. D. et al. Primary biliary cirrhosis. Hepatology 50, 291–308 (2009).
30. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. J. Hepatol. 51, 237–267 (2009).
31. Zhang, X. J. et al. Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21. Nat. Genet. 41, 205–210 (2009).
32. Zhao, X. et al. Whole-exome SNP array identifies 15 new susceptibility loci for psoriasis. Nat. Commun. 6, 6793 (2015).
33. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).
34. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265 (2005).
35. Gabriel, S., Ziaugra, L. & Tabbaa, D. SNP genotyping using the Sequenom MassARRAY iPLEX platform. Curr. Protoc. Hum. Genet. 2.12.1–2.12.18 (2009).
36. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).
37. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
38. Stranger, B. E. et al. Population genomics of human gene expression. Nat. Genet. 39, 1217–1224 (2007).
39. Westra, H. J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243 (2013).
40. Xia, K. et al. seeQTL: a searchable database for human eQTLs. Bioinformatics 28, 451–452 (2012).
41. Ward, L. D. & Kellis, M. HaplOReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40, D930–D934 (2012).
42. Tzartos, J. S. et al. IL-21 and IL-21 receptor expression in lymphocytes and neurons in multiple sclerosis brain. Am. J. Pathol. 178, 794–802 (2011).

Acknowledgements
We thank all participating members of The Jiangsu Provincial PBC Collaboration Group for providing patient samples and clinical information. We would like to thank the patients for their participation in this study. This work was supported in part by grants from the National Natural Science Foundation of China (No. 81270295 to Xiangdong L.; 81325002 and 81620108002 to X.M.; No. 81272737 to W.C.; No. 81400608 and 81570469 to R.T.; Nos. 81130058, 81430034 and 91542123 to Z.L.; and No. 31010113 to X.S.), by grants from Jiangsu provincial Healthy fund (No. BL2014066 to W.C.), by grants from Jiangsu Natural Science Foundation (SBJ2015020697 to Xiangdong L. and BK20130607 to X.S.) and by grant from the Innovation Program of Shanghai Jiao Tong University (YG2014MS43 to R.T.). This work was also supported by NIH Grant R01DK091823 to M.F.S. and M.E.G. Xiangdong L. is a fellow at the Collaborative Innovation Center for Cardiovascular Disease Translational Medicine and the recipient of the Jiangsu Provincial Innovation fund.

Author contributions
This study was initially conceived and designed by Xiangdong L., X.M., W.C., Z.L., X. Zuo; the collection of samples and clinical information for the study were performed by F.Q., R.T., X.S., Y.D., L.W., P.X., X. Zhu, Jian W., Jianfang W., Y. Gong, C.H., Y. Gao, K.Z., Y.J., J.Z., Y.S., Z.H., Y.T., H.Z., N.D., L. Liu, X.W., W.Z., Z.Z., J.N., W.S., Y.Z., Y.M., X. Zha, P.J., H.J., Q.D., Jun. L., Z.L., X.B., L. Li, M.L., Q.W., Y.Y., F.Y., M.L., X.I., X.X., Y.L., I.F., D.Q., Z.Z., H.Q., S.C., B.J., P.L., G.C., T.W., Y.S., J.Y., H.T., M.H., M.X., H.P., L.H., Z.Q., J.G., L.J., W.T., Xiangdong L.; Processing of samples for the study and experimental work were supervised by Xiangdong L., X.S., X.M., R.T., W.X., J.Z. and M.E.G. Xiangdong L. is a fellow at the Collaborative Innovation Center for Cardiovascular Disease Translational Medicine and the recipient of the Jiangsu Provincial Innovation fund.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Qiu, F. et al. A genome-wide association study identifies six novel risk loci for primary biliary cholangitis. Nat. Commun. 8, 14828 doi: 10.1038/ncomms14828 (2017).

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

© The Author(s) 2017