nature genetics

Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis

George F Mells[1,2,11], James A B Floyd[3,11], Katherine I Morley[3,4,11], Heather J Cordell[5], Christopher S Franklin[3], So-Youn Shin[3], Michael A Heneghan[6], James M Neuberger[7], Peter T Donaldson[8], Darren B Day[1], Samantha J Ducker[8], Agnes W Muriithi[1], Elizabeth F Wheater[1], Christopher J Hammond[9], Muhammad F Dawwas[2], The UK PBC Consortium[10], The Wellcome Trust Case Control Consortium 3[10], David E Jones[8], Leena Peltonen[3], Graeme J Alexander[2], Richard N Sandford[1,12] & Carl A Anderson[3,12]

In addition to the *HLA* locus, six genetic risk factors for primary biliary cirrhosis (PBC) have been identified in recent genome-wide association studies (GWAS). To identify additional loci, we carried out a GWAS using 1,840 cases from the UK PBC Consortium and 5,163 UK population controls as part of the Wellcome Trust Case Control Consortium 3 (WTCCC3). We followed up 28 loci in an additional UK cohort of 620 PBC cases and 2,514 population controls. We identified 12 new susceptibility loci (at a genome-wide significance level of $P < 5 \times 10^{-8}$) and replicated all previously associated loci. We identified three further new loci in a meta-analysis of data from our study and previously published GWAS results. New candidate genes include *STAT4*, *DENND1B*, *CD80*, *IL7R*, *CXCR5*, *TNFRSF1A*, *CLEC16A* and *NFKB1*. This study has considerably expanded our knowledge of the genetic architecture of PBC.

Primary biliary cirrhosis is a chronic, autoimmune liver disease characterized by nonsuppurative destructive cholangitis and highly specific autoantibodies to pyruvate dehydrogenase complex (PDC)[1]. This disease is an important cause of chronic liver disease and a well-established indication for liver transplantation. In the UK, the prevalence of PBC is approximately 35 per 100,000 adults and 94 per 100,000 women aged 40 years or older[2]. PBC has a sibling relative risk of ~10, suggesting a substantial genetic contribution to PBC[3].

Previous studies have established that PBC is associated with *HLA-DR8*, with odds ratios ranging from 2.4 to 3.3, depending on the population examined[4]. To date, six non-*HLA* loci have been associated with PBC at a genome-wide level of significance ($P < 5 \times 10^{-8}$): *IL12A* (3q25)[5], *IL12RB2* (1p31)[5], *IRF5/TNPO3* (7q32)[6,7], *ORMDL3/IKZF3* (17q12)[6,7], *MMEL1* (1p36)[6] and *SPIB* (19q13)[7]. To identify additional PBC risk loci, we conducted a GWAS in a large cohort of UK PBC cases and population controls as part of the WTCCC3.

We drew PBC cases from the UK PBC Consortium, which consists of 142 National Health Service (NHS) Trusts in the UK, including all liver transplant centers. All cases were of self-declared British or Irish ancestry. PBC cases were genotyped using the Illumina 660W-Quad array. UK population controls were genotyped using the Illumina Human1M-Duo by the Wellcome Trust Case Control Consortium 2 and consisted of individuals from the 1958 British Birth Cohort and the National Blood Service[8]. Following stringent quality control measures (Online Methods), 507,467 SNPs were available across 1,840 cases and 5,163 population controls (**Supplementary Tables 1,2**). The quantile-quantile plot of the case-control χ^2 test statistics showed a substantial excess of significant associations in the tail of the distribution, even after removal of known loci, which cannot be ascribed to overall inflation of the distribution (genomic control $\lambda = 1.09$; **Supplementary Fig. 1**).

We identified 34 loci where one or more SNPs showed at least suggestive evidence for association ($P < 1 \times 10^{-5}$), including six of the seven previously associated loci (**Table 1**). We found weaker evidence for replication at the seventh previously associated locus, 1p36 (containing *MMEL1*; $P = 4 \times 10^{-3}$). We followed up 28 loci achieving at least suggestive significance and not previously associated with PBC at genome-wide significance by genotyping 46 SNPs in an independent panel of 620 cases from the UK PBC Consortium (**Supplementary Table 3**). We obtained a comparison set of 2,514 UK population controls previously genotyped in the TwinsUK study using the Illumina HumanHap610 array (Online Methods). Twelve of these 28 loci were significant at $P < 0.05$ in the replication study and at $P < 5 \times 10^{-8}$ in the combined analysis of the discovery and replication cohorts and thus represent new PBC associations (**Table 2**). Two of these loci (2q32 and

[1]Academic Department of Medical Genetics, Cambridge University, Cambridge, UK. [2]Department of Hepatology, Cambridge University Hospitals National Health Service (NHS) Foundation Trust, Cambridge, UK. [3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [4]Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, University of Melbourne, Melbourne, Australia. [5]Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, UK. [6]Institute of Liver Studies, King's College Hospital NHS Foundation Trust, Denmark Hill, London, UK. [7]The Liver Unit, Queen Elizabeth Hospital, Birmingham, UK. [8]Institute of Cellular Medicine, Medical School, Newcastle University, Newcastle upon Tyne, UK. [9]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [10]A full list of members is provided in the **Supplementary Note**. [11]These authors contributed equally to this work. [12]These authors jointly directed this work. Correspondence should be addressed to R.N.S. (rns13@cam.ac.uk) or C.A.A. (carl.anderson@sanger.ac.uk).

Received 19 August 2010; accepted 15 February 2011; published online 13 March 2011; doi:10.1038/ng.789

Table 1 Association results and *in silico* analyses for the seven previously confirmed primary biliary cirrhosis risk loci

Chr.	SNP	Risk allele	Left–right region (Mb)	Candidate gene[a] (number of genes in region)	AI[b]	nsSNP[c]	GWAS cohort			
							Control RAF	Case RAF	*P*	OR (95% CI)
1p36	rs10752747	T	2.39–2.78	*MMEL1* (6)	○		0.339	0.367	2.65×10^{-3}	1.13 (1.04–1.22)
1p31	rs17129789	C	67.53–67.71	*IL12RB2* (2)	○		0.177	0.247	9.48×10^{-20}	1.52 (1.39–1.67)
3q25	rs485499	T	160.96–161.3	*IL12A* (2)	•		0.574	0.651	2.29×10^{-16}	1.38 (1.28–1.50)
6p21	rs7774434	C	26.21–33.74	Many (MHC)	○	•	0.379	0.494	3.86×10^{-34}	1.60 (1.48–1.73)
7q32	rs12531711	G	128.33–128.57	*IRF5* (2)	•		0.107	0.159	8.90×10^{-17}	1.58 (1.41–1.76)
17q12	rs7208487	T	34.61–35.49	*ORMDL3* (23)	•		0.840	0.874	7.89×10^{-7}	1.32 (1.18–1.48)
19q13	rs3745516	A	55.52–55.73	*SPIB* (11)		•	0.226	0.287	1.63×10^{-13}	1.38 (1.32–1.44)

Chr., chromosome; RAF, risk allele frequency; OR, odds ratio; 95% CI, 95% confidence interval.
[a]The putative candidate gene represents the strongest candidate within the region based on available evidence but does not preclude the existence of other plausible candidate genes within the region. The number of genes is based upon the RefSeq gene track. [b]Open circles indicate the locus has been previously associated with another autoimmune (AI) disease; a filled circle indicates the same candidate gene has also been suggested. [c]Filled circles indicate that a non-synonymous SNP (nsSNP) in LD with our top SNP was identified in the candidate gene.

1q31) showed suggestive evidence of association in a previous PBC GWAS study[5]. Follow-up genotyping in large independent panels of cases and controls from a range of ethnicities is needed to further characterize these loci.

To identify additional risk loci, we combined summary statistics from our discovery cohort with those from the two datasets included in the previously published meta-analysis of PBC GWAS[7]. Three further new loci reached genome-wide significance (**Table 3**). This included one locus (14q32) that just failed to achieve genome-wide significance in our combined analysis of the discovery and replication cohorts ($P = 1.69 \times 10^{-7}$) but did achieve significance with the addition of data from the study by Liu *et al.*[7] ($P = 2.61 \times 10^{-13}$). As SNPs at the two other loci (3p24 and 11q13) were not genotyped in our replication cohort and the loci were identified based on summary statistics alone, genotyping using an independent technology in additional cohorts is needed to fully validate these associations. A combined GWAS meta-analysis is still warranted because we were only able to meta-analyze the top 100 SNPs from the Liu *et al.*[7] study. Genome-wide imputation using HapMap3 reference panels did not identify any further genome-wide significant loci (Online Methods, **Supplementary Fig. 2** and **Supplementary Table 4**), although for some loci, imputed SNPs provided stronger evidence of association than genotyped SNPs. We detected no statistically significant gene-gene interactions between associated loci or after fitting an *HLA*-risk model (Online Methods and **Supplementary Table 5**).

We found evidence for a second independent association at the 3q25 locus containing *IL12A* and *SCHIP1*, as did Liu *et al.*[7] (**Supplementary Table 6**). Three SNPs, located between *IL12A* and *SCHIP1*, remained genome-wide significant after conditional logistic regression adjusting for the most significant SNP in the region (rs485499). Conducting the same analysis using the imputed data identified a further two SNPs reaching genome-wide significance. These five SNPs are all in linkage disequilibrium (LD; $r^2 > 0.2$) with each other, but none are in LD with

Table 2 Association results and *in silico* analyses for 12 newly confirmed primary biliary cirrhosis risk loci

Chr.	SNP	Risk allele	Left–right region (Mb)	Candidate gene[a] (number of genes in region)	AI[b]	GRAIL[c]	nsSNP[d]	GWAS cohort				Replication cohort				Combined samples	
								Control RAF	Case RAF	*P*	OR (95% CI)	Control RAF	Case RAF	*P*	OR (95% CI)	*P*	OR (95% CI)
1q31	rs12134279	T	195.58–196.21	*DENND1B* (4)	○	•		0.202	0.250	1.07×10^{-9}	1.32 (1.21–1.44)	0.201	0.263	2.57×10^{-6}	1.42 (1.23–1.65)	2.06×10^{-14}	1.34 (1.25–1.45)
2q32	rs10931468	A	190.77–191.61	*STAT4* (7)	•	•		0.119	0.164	2.55×10^{-12}	1.46 (1.31–1.62)	0.120	0.183	2.64×10^{-9}	1.64 (1.39–1.94)	2.35×10^{-19}	1.50 (1.37–1.64)
3q13	rs2293370	G	120.58–120.79	*CD80* (6)	•	•		0.804	0.853	7.70×10^{-11}	1.41 (1.27–1.56)	0.809	0.835	0.036	1.19 (1.01–1.41)	2.53×10^{-11}	1.35 (1.23–1.47)
4q24	rs7665090	C	103.61–104.24	*NFKB1* (7)				0.524	0.572	5.33×10^{-7}	1.21 (1.13–1.31)	0.513	0.593	5.50×10^{-7}	1.38 (1.22–1.57)	4.06×10^{-12}	1.26 (1.18–1.34)
5p13	rs860413	A	35.74–36.08	*IL7R* (5)	•	•	•	0.719	0.773	3.09×10^{-10}	1.33 (1.22–1.45)	0.729	0.769	4.50×10^{-3}	1.24 (1.07–1.43)	1.02×10^{-11}	1.30 (1.21–1.40)
7p14	rs6974491	A	37.32–37.41	(0)				0.170	0.205	3.39×10^{-6}	1.25 (1.14–1.38)	0.177	0.215	2.40×10^{-3}	1.27 (1.09–1.49)	4.44×10^{-8}	1.25 (1.16–1.36)
11q23	rs6421571	C	117.82–118.30	*CXCR5* (10)	○			0.809	0.855	3.53×10^{-10}	1.40 (1.26–1.55)	0.810	0.847	2.10×10^{-3}	1.30 (1.10–1.55)	2.69×10^{-12}	1.37 (1.25–1.50)
12p13	rs1800693	C	6.29–6.33	*TNFRSF1A* (3)	•	•		0.401	0.452	5.51×10^{-8}	1.23 (1.14–1.33)	0.403	0.445	8.70×10^{-3}	1.18 (1.04–1.34)	1.80×10^{-9}	1.22 (1.14–1.30)
14q24	rs911263	T	67.34–67.98	*RAD51L1* (2)				0.712	0.764	1.68×10^{-9}	1.31 (1.20–1.43)	0.717	0.760	2.30×10^{-3}	1.25 (1.08–1.45)	1.76×10^{-11}	1.29 (1.20–1.39)
16p13	rs12924729	G	10.92–11.22	*CLEC16A* (3)	•			0.679	0.737	7.68×10^{-11}	1.32 (1.21–1.44)	0.680	0.718	8.80×10^{-3}	1.20 (1.05–1.38)	2.95×10^{-12}	1.29 (1.20–1.38)
16q24	rs11117432	G	84.55–84.58	(0)	○			0.760	0.808	1.20×10^{-6}	1.26 (1.15–1.39)	0.774	0.838	9.52×10^{-7}	1.52 (1.28–1.79)	4.66×10^{-11}	1.31 (1.21–1.43)
22q13	rs968451	T	37.87–38.19	*MAP3K7IP1* (3)				0.194	0.233	4.31×10^{-7}	1.27 (1.16–1.39)	0.193	0.237	6.45×10^{-4}	1.30 (1.12–1.51)	1.08×10^{-9}	1.27 (1.18–1.38)

PBC loci that meet genome-wide significance ($P < 5 \times 10^{-8}$) in the combined analysis and $P < 0.05$ in the replication cohort. GWAS and replication cohort data for the replicated SNPs were merged using PLINK. Chr., chromosome; RAF, risk allele frequency.
[a]The putative candidate gene represents the strongest candidate within the region based on available evidence but does not preclude the existence of other plausible candidate genes within the region. The number of genes is based upon the RefSeq gene track. [b]Open circles indicate the locus as been previously associated with another autoimmune (AI) disease; a filled circle indicates the same candidate gene has also been suggested. [c]Filled circles indicate the gene was identified by GRAIL as the most plausible functional candidate in the region ($P_{text} < 0.01$). GRAIL results are not provided for previously confirmed loci as these were used as seeds in the analysis. [d]Filled circles indicate a non-synonymous SNP (nsSNP) in LD ($r^2 > 0.8$) with our top SNP identified in the candidate gene.

Table 3 Genomic regions reaching genome-wide significance after meta-analysis with Liu *et al.*[7] data

Chr.	SNP	Risk allele	Left–right region (Mb)	Candidate gene[a] (number of genes in region)	Discovery sample				Liu *et al.*[7]		Meta-analysis	
					Control RAF	Case RAF	*P*	OR (95% CI)	*P*	OR[b]	*P*	OR (95% CI)
3p24	rs1372072	A	16.82–17.13	*PLCL2* (1)	0.365	0.400	1.38×10^{-4}	1.16 (1.08–1.25)	1.52×10^{-5}	1.27	2.28×10^{-8}	1.20 (1.12–1.27)
11q13	rs538147	G	63.60–64.04	*RPS6KA4* (20)	0.606	0.647	1.01×10^{-5}	1.19 (1.10–1.29)	7.72×10^{-6}	1.28	2.06×10^{-10}	1.23 (1.15–1.31)
14q32	rs8017161	A	102.54–102.68	*TNFAIP2* (3)	0.396	0.439	4.71×10^{-6}	1.20 (1.11–1.29)	4.86×10^{-7}	1.31	2.61×10^{-13}	1.22 (1.16–1.27)

Chr., chromosome; RAF, risk allele frequency.
[a]The putative candidate gene is the strongest candidate within the region based on available evidence but does not preclude the existence of other plausible candidate genes within the region. The number of genes is based upon the RefSeq gene track. [b]Liu *et al.*[7] do not provide confidence intervals for the odds ratios (ORs) estimated from their meta-analysis for these SNPs. None of these genomic regions have previously been associated at genome-wide significance with another autoimmune disease. GRAIL failed to identify any strong candidate genes within these regions and no non-synonymous SNPs were identified in high LD ($r^2 > 0.8$) with the most associated SNP at each locus.

rs485499. They are located downstream of *SCHIP1* but upstream of *IL12A* (**Supplementary Fig. 3**). Fine mapping of this locus is needed to determine whether these association signals implicate independent variants affecting the same gene or two different genes.

We identified plausible candidate genes within associated loci through manual curation, supported by evidence from: (i) previous GWAS findings for other autoimmune diseases; (ii) GRAIL[9], a literature-mining tool that identifies non-random, evidence-based links between genes; (iii) identification of non-synonymous SNPs in 1000 Genomes Project data that are in LD ($r^2 > 0.8$) with the most associated genotyped SNP in each locus; (iv) identification of expression quantitative trait loci (eQTL) within associated loci that are in LD ($r^2 > 0.8$) with the most associated SNP at that locus using data from Dixon *et al.*[10] (Online Methods and **Supplementary Tables 7–10**). Even in aggregate, these analyses do not confirm which gene(s) contain causal variants, but they allow us to identify potential candidate genes for future follow-up studies. **Supplementary Figure 4** shows all genes within each of the associated loci.

The results from the GWAS of PBC conducted to date provide additional support for the involvement of three pathways previously implicated in the pathogenesis of PBC: NF-κB signaling, T-cell differentiation, and toll-like receptor (TLR) and tumor necrosis factor (TNF) signaling.

We identified several loci containing genes involved in activation of NF-κB, a transcription factor which regulates expression of many genes involved in the immune response and that is highly activated in other autoimmune disorders, such as rheumatoid arthritis, multiple sclerosis and asthma[11]. Its importance in PBC is suggested by evidence that NF-κB modulates the balance of survival and apoptosis in activated hepatic stellate cells[12], and NF-κB $p50^{-/-}$ mice show aggressive hepatic inflammation and fibrosis[13]. The locus we identified at 4q24 contains the *NFKB1* gene itself, and we identified genes in pathways leading to NF-κB activation at four other loci: 22q13 (*TAB1*), 12p13 (*TNFRSF1A*), 3q13 (*CD80*) and 11q13 (*RPS6KA4*).

Loci identified to date suggest a role for T-lymphocyte differentiation in the development of PBC. T_H1 immune responses have been implicated in many autoimmune diseases[14] and may be involved in the development of autoreactive T-cells, consistent with the putative role of PDC-specific autoreactive T_H1 cells in the pathogenesis of human PBC and animal disease models[15]. IL-12 signaling promotes T_H1-type immune responses by driving differentiation of activated, naïve T-cells into T_H1 cells[16], and three loci containing genes involved in IL-12 signaling have been identified for PBC: 3q25 (*IL12A*) and 1p31 (*IL12RB2*) by Hirschfield *et al.*[5] and 2q32 (*STAT4*) in this study. These results provide further support for the T_H1 hypothesis regarding PBC development.

Activation of TLR signaling and its downstream effectors, such as TNFα, are well described in PBC[17]. The 7q32 locus, identified by Liu *et al.*[7], contains *IRF5*, which is activated in response to TLR signaling and leads to selective expression of TNFα. We identified a locus at 11q13 containing *RPS6KA4*, which suppresses TLR-dependent cytokine production[18]. TNFα is an activating factor for a number of intracellular pathways that determine the fate of hepatocytes and thus plays a key role in liver homeostasis[19]. We identified three loci containing genes in TNFα signaling pathways: 12p13 (*TNFRSF1A*), 1q31 (*DENND1B*) and 14q32 (*TNFAIP2*). *TNFRSF1A* is one of two receptors for TNFα, and *Tnfrsf1a*$^{-/-}$ mice show attenuated liver fibrosis when compared to wild-type mice after administration of a potent hepatotoxin[20]. *DENND1B* interacts directly with *TNFRSF1A* (ref. 21) and has previously been associated with asthma[22]. TNFα signaling also directly induces *TNFAIP2* expression[23].

In summary, this is the first report in a new series of GWAS undertaken by the WTCCC3. Twelve new PBC susceptibility loci have been identified in this study of more than 7,000 European samples, making this the largest GWAS of PBC to date. In addition, a further three loci achieved genome-wide significance following a meta-analysis of published data. For many of the associated loci, we identified plausible candidate genes that support the involvement of the innate and adaptive immune systems in PBC etiology, particularly signaling via the NF-κB, TLR and TNF pathways, although these findings require confirmation through fine-mapping, gene-expression and functional studies.

URLs. WTCCC3, http://www.wtccc.org.uk/; Oragene, http://www.dnagenotek.com/; Source Bioscience Healthcare, http://healthcare.sourcebioscience.com/; TwinsUK, http://www.twinsuk.ac.uk/; data access committee, http://www.ebi.ac.uk/ega/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; HapMap, http://hapmap.ncbi.nlm.nih.gov/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The PBC sample collection was funded by the Isaac Newton Trust, the PBC Foundation and the Wellcome Trust (085925/Z/08/Z). The PBC Genetics Study is a portfolio study of the National Institute for Health Research Comprehensive Clinical Research Network (NIHR CRN, portfolio reference 5630). The WTCCC3 project is also supported by the Wellcome Trust (WT090355/A/09/Z, WT090355/B/09/Z). G.F.M. is a Clinical Research Training Fellow of the Medical Research Council (G0800460). G.F.M. is also supported by a Raymond and Beverly Sackler Studentship. C.J.H. is an NIHR Senior Research Fellow. C.A.A. is funded by the Wellcome Trust (WT91745/Z/10/Z).

We are grateful to the PBC Foundation for helping us to establish the PBC Genetics Study, for endorsing it, and for encouraging members of the Foundation to contribute samples. We thank all of the research nurses who assisted with participant recruitment in collaborating centers. We thank the staff in the NIHR CRN and Clinical Research Collaboration (CRC) Cymru for providing invaluable

support. We are grateful to K. Chittock and his colleagues at Source Bioscience for their efforts. We thank O. Burren for designing the participant database and for providing information technology support. We acknowledge the Department of Twin Research and Genetic Epidemiology, King's College London for additional control genotype data, which was part-funded by a US National Institutes of Health National Eye Institute grant 1RO1EY018246 (PI, T. Young) and genotyping by the NIH Center for Inherited Disease Research (CIDR), with additional genotyping performed at the Wellcome Trust Sanger Institute. The study also received support from the NIHR comprehensive Biomedical Research Centre award to Guy's & St. Thomas' NHS Foundation Trust partnering King's College London. Finally, we thank the individuals who contributed samples used in this study.

AUTHOR CONTRIBUTIONS

Study concept and design: G.F.M., H.J.C., M.A.H., J.M.N., P.T.D., the WTCCC3 management committee (see **Supplementary Note**), L.P., D.E.J., G.J.A., R.N.S., C.A.A.

PBC Genetics Study management: G.F.M., D.B.D., S.J.D., A.W.M., E.F.W., R.N.S.

Case ascertainment and phenotyping: G.F.M., D.B.D., S.J.D., A.W.M., E.F.W., M.F.D., The UK PBC Consortium (see **Supplementary Note**), D.E.J., G.J.A., R.N.S.

Control sample ascertainment: The UK Blood Service Controls group (see **Supplementary Note**), The 1958 Birth Cohort Controls group (see **Supplementary Note**), C.J.H., C.A.A.

Genotyping: The WTCCC3 DNA, Genotyping and Informatics group (see **Supplementary Note**).

Statistical analysis: J.A.B.F., K.I.M., H.J.C., C.S.F., S.-Y.S., The WTCCC3 data analysis group (see **Supplementary Note**), C.A.A.

Manuscript preparation: G.F.M., J.A.B.F., K.I.M., H.J.C., D.E.J., G.J.A., R.N.S., C.A.A.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Kaplan, M.M. & Gershwin, M.E. Primary biliary cirrhosis. *N. Engl. J. Med.* **353**, 1261–1273 (2005).
2. James, O.F. *et al.* Primary biliary cirrhosis once rare, now common in the United Kingdom? *Hepatology* **30**, 390–394 (1999).
3. Jones, D.E., Watt, F.E., Metcalf, J.V., Bassendine, M.F. & James, O.F. Familial primary biliary cirrhosis reassessed: a geographically-based population study. *J. Hepatol.* **30**, 402–407 (1999).
4. Donaldson, P.T. *et al.* HLA class II alleles, genotypes, haplotypes, and amino acids in primary biliary cirrhosis: a large-scale study. *Hepatology* **44**, 667–674 (2006).
5. Hirschfield, G.M. *et al.* Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N. Engl. J. Med.* **360**, 2544–2555 (2009).
6. Hirschfield, G.M. *et al.* Variants at *IRF5-TNPO3*, 17q12–21 and *MMEL1* are associated with primary biliary cirrhosis. *Nat. Genet.* **42**, 655–657 (2010).
7. Liu, X. *et al.* Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. *Nat. Genet.* **42**, 658–660 (2010).
8. Barrett, J.C. *et al.* Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the *HNF4A* region. *Nat. Genet.* **41**, 1330–1334 (2009).
9. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
10. Dixon, A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
11. Li, Q. & Verma, I.M. NF-κB regulation in the immune system. *Nat. Rev. Immunol.* **2**, 725–734 (2002).
12. Elsharkawy, A.M. *et al.* The NF-κB p50:p50:HDAC-1 repressor complex orchestrates transcriptional inhibition of multiple pro-inflammatory genes. *J. Hepatol.* **53**, 519–527 (2010).
13. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
14. Zhernakova, A., van Diemen, C.C. & Wijmenga, C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat. Rev. Genet.* **10**, 43–55 (2009).
15. Jones, D.E. Pathogenesis of primary biliary cirrhosis. *Gut* **56**, 1615–1624 (2007).
16. Trinchieri, G. Interleukin-12 and the regulation of innate resistance and adaptive immunity. *Nat. Rev. Immunol.* **3**, 133–146 (2003).
17. Mao, T.K. *et al.* Altered monocyte responses to defined TLR ligands in patients with primary biliary cirrhosis. *Hepatology* **42**, 802–808 (2005).
18. Vermeulen, L., Berghe, W.V., Beck, I.M., De Bosscher, K. & Haegeman, G. The versatile role of MSKs in transcriptional regulation. *Trends Biochem. Sci.* **34**, 311–318 (2009).
19. Tacke, F., Luedde, T. & Trautwein, C. Inflammatory pathways in liver homeostasis and liver injury. *Clin. Rev. Allergy Immunol.* **36**, 4–12 (2009).
20. Kitamura, K. *et al.* Pathogenic roles of tumor necrosis factor receptor p55-mediated signals in dimethylnitrosamine-induced murine liver fibrosis. *Lab. Invest.* **82**, 571–583 (2002).
21. Del Villar, K. & Miller, C.A. Down-regulation of DENN/MADD, a TNF receptor binding protein, correlates with neuronal cell death in Alzheimer's disease brain and hippocampal neurons. *Proc. Natl. Acad. Sci. USA* **101**, 4210–4215 (2004).
22. Sleiman, P.M. *et al.* Variants of *DENND1B* associated with asthma in children. *N. Engl. J. Med.* **362**, 36–44 (2010).
23. Sarma, V., Wolf, F.W., Marks, R.M., Shows, T.B. & Dixit, V.M. Cloning of a novel tumor necrosis factor-alpha-inducible primary response gene that is differentially expressed in development and capillary tube-like formation *in vitro*. *J. Immunol.* **148**, 3302–3312 (1992).

ONLINE METHODS

Ethics approval. The study was approved by the Oxford Research Ethics Committee C (Oxford REC C reference 07/H0606/96) and by the Research and Development Departments of all NHS Trusts collaborating in the study.

Subjects. Cases were recruited from the UK PBC Consortium, a group of 142 NHS Trusts. Cases were adults (>18 years old) with probable or definite PBC. The diagnosis of PBC was based on the following three criteria[2]: (i) abnormal liver biochemistry (one or more of bilirubin, alanine transaminase, aspartate transaminase, alkaline phosphatase or gamma-glutamyl transferase above the upper reference level); (ii) a positive test for antimitochondrial antibodies (titer 1:40 or greater); or (iii) a liver biopsy with histological abnormalities consistent with PBC. We recorded a diagnosis of probable PBC if two criteria had been fulfilled and definite PBC if all three criteria had been fulfilled[2].

In the discovery cohort, PBC cases had a mean age of 64.3 years (with a range of 30–94 years); 91.4% of cases were female, 93.5% were AMA positive and 19.1% had received a liver transplant. In the replication cohort, PBC cases had a mean age of 63.4 years (range 32–99 years); 89.8% of cases were female, 93.7% were AMA positive and 12% had received a liver transplant.

DNA sample preparation. DNA was extracted from blood or saliva. Saliva samples were collected using the Oragene kit, and DNA was prepared according to the manufacturer's protocols at Source BioScience Healthcare. Blood samples were extracted by the East Anglian Medical Genetics Service. All DNA samples were plated, normalized and shipped to the Sanger Centre for sample quality control.

Genotyping. *GWAS samples.* Both cases and controls were genotyped at the Wellcome Trust Sanger Institute Genotyping Facility on the Illumina Human-660W Quad and Illumina 1M-Duo, respectively. For all samples passing standard laboratory quality control measures, normalized intensities were extracted and genotypes were called using the Illuminus algorithm. Individual genotypes with a posterior probability <0.95 were set as missing. Controls comprised 2,930 1958 British Birth Cohort samples and 2,737 National Blood Service samples genotyped as part of the Wellcome Trust Case Control Consortium 2 project.

Replication sample. Genotyping was carried out using the Sequenom iPLEX Gold assay. Fifty-one SNPs and three gender informative markers were typed across 661 samples. Twenty-eight samples were excluded as duplicates and 13 were excluded due to gender inconsistencies, leaving a replication dataset of 620 samples. A set of 2,514 unrelated individuals genotyped on the Illumina HumanHap610 array by the TwinsUK project were used as controls.

Quality control. Samples with a call rate <98% or heterozygosity >3 standard deviations from the mean were removed, in addition to one individual from each related pair (identity by descent >10%). Principal components analysis[13] was used to identify (for subsequent removal) individuals of non-European ancestry (**Supplementary Fig. 5**). After sample quality control, 1,840 cases and 5,163 controls remained (**Supplementary Table 1**). SNPs with a call rate <98%, a Hardy-Weinberg equilibrium $P < 1 \times 10^{-6}$, a minor allele frequency <0.05 or that failed PLINK's non-random missingness test were removed (see URLs, **Supplementary Table 2** and **Supplementary Fig. 6**). Of the 594,398 SNPs on the case genotyping panel, after applying quality control thresholds in each of the case and control cohorts, 507,467 were taken forward for analysis (**Supplementary Table 2**).

Statistical methods. *Association analysis.* Case-control association tests were conducted using 1-degree-of-freedom Cochran-Armitage trend tests as implemented in PLINK v1.07 (ref. 24). Cluster plots for all SNPs with $P < 1 \times 10^{-5}$ were examined using Evoker[25], and 32 SNPs with poor genotype calling were removed (cluster plots for the index SNP for each genome-wide significant region are shown in **Supplementary Fig. 7**). Left and right boundaries for each associated region were defined by finding a 0.1 cM interval on either side of the most associated SNP where no SNP has $P < 1 \times 10^{-5}$.

We searched for secondary associations within all loci showing at least suggestive evidence of association ($P < 1 \times 10^{-5}$) by fitting linear regression models conditioning on the allelic dosage of the index SNP. Results (for SNPs with conditional $P < 5 \times 10^{-8}$) are shown in **Supplementary Table 6**. PLINK was also used to test for pair-wise interactions between the most associated SNPs from each confirmed PBC locus.

Meta-analysis. Liu *et al.*[7] provide summary statistics for the two datasets analyzed in their meta-analysis for all SNPs with $P < 1 \times 10^{-4}$. We used the software GWAMA[26] to conduct a new fixed-effects meta-analysis including our GWAS samples. For all associations reaching genome-wide significance, a Cochran's test for between-study heterogeneity was non-significant ($P > 0.05$), and the direction of the risk allele was the same in all three cohorts.

Imputation. Imputation was conducted using BEAGLE version 3.0.2 (ref. 27) with the HapMap3 Indo-European data (phase 2 CEU, MEX, TSI and GIH populations) as the reference genotype set (see URLs, **Supplementary Fig. 2** and **Supplementary Table 4**).

Interaction with HLA genotypes. We examined our dataset for evidence of a statistical interaction between genome-wide significant loci and the HLA region. We fitted a lasso regression model (using the lasso2 library within R v2.9.0 (ref. 28)) to genotype data from the *HLA* region on chromosome 6 to identify the subset of SNPs most predictive of PBC status. Coefficients from this regression model were used to generate an estimate of HLA-associated risk for PBC for cases and controls, assuming an additive effect on disease risk. We then fitted logistic regression models for each hit SNP, including SNP genotype and HLA-risk score as main effects. We tested for SNP × HLA risk interactions by comparing the fit of models with and without an interaction term (**Supplementary Table 5**).

Identification of candidate genes. *Selection of genes in regions.* Regions around SNPs were defined as described above. The list of potential genes within each region was based upon the RefSeq track of the UCSC Genome Browser (**Supplementary Fig. 4**).

Autoimmune disease overlap. We used a manually curated list of loci for other autoimmune disorders that had reached genome-wide significance with replication in a primary study or in a meta-analysis. A region around the most significant SNP for each locus was defined (in the manner previously described for the PBC loci) and its overlap with PBC regions was evaluated (**Supplementary Table 7**).

GRAIL analysis. GRAIL[9] was used to identify the most plausible candidate gene for each new locus reaching genome-wide significance. We used the implementation of the GRAIL algorithm that permits the user to specify 'seed' and 'query' regions using our top SNP from the seven loci (including the *HLA*) previously associated with PBC to specify 'seed' regions, with the remaining 15 new loci as 'query' regions (**Supplementary Table 8**).

Non-synonymous SNP analysis. Data from the 1000 Genomes Project were used to find non-synonymous, splice or stop-encoding SNPs that are in LD ($r^2 > 0.8$) with the most associated directly genotyped SNP in each confirmed PBC locus. LD was calculated for all pairwise SNPs within 1 Mb for the HapMap CEU individuals from the 1000 Genomes Project data, March 2010 release (**Supplementary Table 9**).

eQTL analysis. Data from the University of Michigan Center for Statistical Genetics[10] were used to identify eQTL within PBC-associated loci that are in LD ($r^2 > 0.8$) with the most associated SNP at that locus (**Supplementary Table 10**).

24. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
25. Morris, J.A., Randall, J.C., Maller, J.B. & Barrett, J.C. Evoker: a visualization tool for genotype intensity data. *Bioinformatics* **26**, 1786–1787 (2010).
26. Magi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
27. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223 (2009).
28. R Development Core Team. R: a language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria, 2006.