Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis


1Contact Information: Prof. Tom H. Karlsen, MD, PhD, Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital Rikshospitalet, Postboks 4950 Nydalen, N-0424 Oslo, Norway, Tel.: +47 23 07 2469; Fax: +47 2307 3928; t.h.karlsen@medisin.uio.no; Dr Carl. A. Anderson, PhD., Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom, Tel.: ++44 1223 492371; Fax: ++44 1223 496826; carl.anderson@sanger.ac.uk.

*These authors contributed equally to this work
§Shared senior authorship

Author contributions

Competing Interests Statement
The authors declare no competing interests.

URLs
1Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
2Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway
3Research Institute of Internal Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway
4Institute of Clinical Medicine, University of Oslo, Oslo, Norway
5Section of Gastroenterology, Department of Transplantation Medicine, Division of Cancer, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway
6Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
7Department of Gastroenterology and Hepatology, Norfolk and Norwich, University Hospitals NHS Trust, Norwich, UK
8Max Planck Institute for Informatics, Saarbrücken, Germany
9KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital, Ulleval, Oslo, Norway
10Department of Gastroenterology and Hepatology, University of Groningen and University Medical Centre Groningen, Groningen, the Netherlands
11Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany
12Integrated Research and Treatment Center-Transplantation (IFB-tx), Hannover Medical School, Hannover, Germany
13Snyder Institute of Chronic Diseases, Department of Medicine, University of Calgary, Calgary, Canada
14Center for Autoimmune Liver Diseases, Humanitas Clinical and Research Center, Rozzano (MI), Italy
15Division of Gastroenterology, Department of Medicine, University of Toronto, Toronto, Canada
16Centre for Liver Research, NIHR Biomedical Research Unit, Birmingham, UK
17Department of Medicine, University Hospital of Heidelberg, Heidelberg, Germany
18Liver Unit, Hospital Clinic, IDIBAPS, CIBERehd, University of Barcelona, Barcelona, Spain
19Center for Basic Research in Digestive Diseases, Division of Gastroenterology and Hepatology, Mayo Clinic, College of Medicine, Rochester, Minnesota, USA
20Liver Unit and Liver Research Laboratories, Pomeranian Medical University, Szczecin, Poland
21Department of Medicine 2, Grosshadern, University of Munich, Munich, Germany
221st Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
23Department of Internal Medicine, Hepatology and Gastroenterology, Charité Universitätsmedizin Berlin, Berlin, Germany
24Academic Department of Medical Genetics, University of Cambridge, Cambridge, UK
25Department of Medicine, Medical School, University of Thessaly, Larissa, Greece
26Research Laboratory of Internal Medicine, Medical School, University of Thessaly, Larissa, Greece
27Institute of Human Genetics, University of Bonn, Bonn, Germany
28Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany
29Institute of Human Genetics, Technische Universität München, Munich, Germany
30Department of Neurology, Technische Universität München, Munich, Germany
31Institute of Human Genetics, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany
32Department of Genetics, University of Groningen and University Medical Centre Groningen, Groningen, The Netherlands
33Department of General, Visceral, Thoracic, Transplantation and Pediatric Surgery, University Medical Centre Schleswig-Holstein, Campus Kiel, Germany
34Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, the Netherlands
35Department of Environmental Science, Policy, and Management, University of California, Berkeley, United States of America
36Department of Hepatobiliary Surgery and Transplantation, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
371st Department of Medicine, University of Mainz, Mainz, Germany
38Division of Gastroenterology and Hepatology, University of Alberta, Edmonton, Alberta, Canada
39Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
40Public Health Genomics Unit, Institute for Molecular Medicine Finland FIMM, University of Helsinki and National Institute for Health and Welfare, Helsinki, Finland
41Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Canada
42Department of Clinical Medicine, Division of Gastroenterology, Sapienza University of Rome, Rome, Italy
43Dept. of Surgical, Oncological and Gastroenterological Sciences, University of Padova, Padova, Italy
44Center for Public Health Genomics, Division of Endocrinology & Metabolism, University of Virginia, Charlottesville, USA
45Department of Internal Medicine, Division of Endocrinology & Metabolism, University of Virginia,
Keywords

- genetic association study
- disease genetics
- immunogenetics
- liver
Primary sclerosing cholangitis (PSC) is a severe liver disease of unknown etiology leading to fibrotic destruction of the bile ducts and ultimately the need for liver transplantation. We compared 3,789 European ancestry PSC cases to 25,079 population controls across 130,422 single-nucleotide polymorphisms (SNPs) genotyped using Immunochip. We identified 12 genome-wide significant associations outside the human leukocyte antigen (HLA) complex, nine of which were novel, thereby increasing the number of known PSC risk loci to 16. Despite comorbidity with inflammatory bowel disease (IBD) in 72% of the patients, six of the 12 loci showed significantly stronger association with PSC than IBD, suggesting an overlapping yet distinct genetic architecture. We incorporated pleiotropy with seven diseases clinically co-occurring with PSC and found suggestive evidence for 33 additional PSC risk loci. Together with network analyses, these findings further complete the genetic risk map of PSC and considerably expand on the relationship between PSC and other immune-mediated diseases.

The pathogenesis of PSC is poorly understood, and due to lack of effective medical therapy, PSC remains a leading indication for liver transplantation in Northern Europe and the US, despite the relatively low prevalence (1/10,000). Affected individuals are diagnosed at a median age of 30-40 years and suffer from an increased frequency of IBD (60-80%) and autoimmune diseases (25%). Conversely, approximately only 5% of patients with IBD develop PSC. A 9-39-fold sibling relative risk indicates a strong genetic component to PSC risk. In addition to multiple strong associations within the HLA complex, recent association studies have identified genome-wide significant loci at 1p36 (MMEL1/TNFRSF14), 2q13 (BCL2L11), 2q37 (GPR35), 3p21 (MST1), 10p15 (IL2RA) and 18q21 (TCF4).

Several theories have been proposed to explain the development of PSC. The strong HLA associations and the clinical co-occurrence of immune-mediated diseases suggest that autoimmunity plays a role. To further characterize the genetic etiology of PSC, we recruited PSC patients throughout Europe and North America, more than doubling the number of ascertained cases included in previous genetic studies. We genotyped 196,524 SNPs in 4,228 PSC cases and 27,077 population controls using the Immunochip, a targeted genotyping array with dense marker coverage across 186 known disease loci from 12 immune-mediated diseases. Outside these 186 loci, Immunochip also assays thousands of SNPs of intermediate significance from multiple meta-analyses of immune-mediated diseases.

Following quality control (QC; see Online Methods), 130,422 SNPs from 3,789 PSC cases and 25,079 population controls were available for analysis. We imputed a further 80,183 SNPs located in the Immunochip fine mapping regions using the 1000 Genomes reference panel (Online Methods). We performed case-control association tests using a linear mixed model as implemented in [MMM](http://www.mmm.com) to minimize the effect of population stratification and sample relatedness (QC = 1.02, estimated using 2,544 “null” SNPs, see Online Methods).

We identified twelve non-HLA genome-wide significant (P<5x10^-8) susceptibility loci (Table 1), nine of which were novel (Fig. 1). The most associated SNP within each locus was a common variant (all risk allele frequencies >0.18) of moderate effect (odds ratios (ORs) between 1.15 and 1.4) (Table 1). Genotype imputation and stepwise conditional regressions within each locus did not identify additional independent genome-wide significant signals, nor did genotype-genotype or gender-genotype interaction analyses (Online Methods).
For seven of the nine novel loci, the most significantly associated SNP in the locus was the same SNP or was in strong linkage disequilibrium (LD; $r^2>0.8$) with the original association reports for another disease (Supplementary Table 3). The two exceptions were 11q23, where only independent disease associations ($r^2<0.01$) have so far been reported\(^ {17}\), and 6q15, where the most significantly associated PSC variant, rs56258221 (OR=1.23, $P=8.36\times10^{-12}$), is in low-to-moderate LD with the previously reported $BACH2$ variants in Crohn’s disease ($r^2=0.23$) and type 1 diabetes ($r^2=0.12$). Three out of four known non-HLA PSC risk loci present on the Immunochip passed genotyping QC and were confirmed in our analysis (1p36, 3p21 and 10p15; see Supplementary Note and Supplementary Fig. 3).

To prioritize candidate genes within the non-HLA genome-wide significant loci, we searched for functional consequences of the most associated SNPs or SNPs in high LD ($r^2>0.8$), i.e. missense SNPs (Supplementary Table 4 and Supplementary Fig. 4) and expression quantitative trait loci (eQTLs) (Supplementary Table 5), and we functionally annotated risk loci using data from the ENCODE project (Supplementary Table 6 and Supplementary Note)\(^ {18}\). We also constructed networks based on functional similarity measures (Supplementary Fig. 5 and Online Methods), known protein-protein interactions (DAPPLE\(^ {19}\), Supplementary Table 7 and Supplementary Note), and the published literature (GRAIL\(^ {20}\), Supplementary Fig. 6 and Supplementary Note) to identify important disease-relevant genes. For six of the 12 genome-wide significant loci, the same gene (MMEL1, CD28, MST1, SH2B3, CD226 and SIK2) was annotated by more than one method (Supplementary Table 7), suggesting these as candidates for further investigation at these loci.

Two newly associated loci are located outside of the Immunochip fine mapping regions (Figures 1d and 1e). At 11q23, the most strongly associated SNP, rs7937682 (OR=1.17, $P=3.18\times10^{-9}$), is located in an intron of salt-inducible kinase 2 ($SIK2$), which both influences the expression of interleukin-10 in macrophages and Nur77, an important transcription factor in leukocytes\(^ {21}\). The association at 12q13 is with an intronic SNP (rs11168249, OR=1.15, $P=5.49\times10^{-9}$) within the histone deacetylase 7 ($HDAC7$) gene, which has also been associated with IBD\(^ {22}\). $HDAC7$ has been implicated in negative selection of T cells in the thymus\(^ {23}\), a key factor in the development of immune tolerance. A role for $HDAC7$ in PSC etiology is supported by the novel association at 19q13, where the most associated SNP, rs60652743 (OR=1.25, $P=6.51\times10^{-10}$) is located within an intron of serine-threonine protein kinase D2 ($PRKD2$). When T cell receptors of thymocytes are engaged, PRKD2 phosphorylates $HDAC7$, leading to nuclear exclusion of $HDAC7$ and loss of its gene regulatory functions, ultimately resulting in apoptosis and negative selection of immature T cells\(^ {24,25}\). Interestingly, this negative selection takes place due to a loss of $HDAC7$-mediated repression of Nur77 (regulated by SIK2)\(^ {26}\), linking three novel PSC loci to this pathway.

The associations at the HLA complex at 6p21 were refined by imputing alleles at $HLA-A$, $HLA-B$, $HLA-C$, $HLA-DRB1$, $HLA-DQB1$, $HLA-DQA$ and $HLA-DPB1$ (see Supplementary Note)\(^ {27}\). The top associated SNP (rs4143332) was in almost perfect LD ($r^2=0.996$) with HLA-B*08:01 (Supplementary Note). In a stepwise conditional analysis including both SNP and HLA allele genotypes, rs4143332 (tagging HLA-B*08:01) and a complex HLA class II association signal determined by $HLA-DQA1*01:03$ and SNPs rs532098, rs1794282 and rs9263964 (Supplementary Fig. 7) explain most of the HLA association signal in PSC.

When performing a stepwise regression of the HLA alleles only, the class II associations are coherent with previous reports, apart from a novel association with $HLA-DQA1*01:01$ (see Supplementary Note and Supplementary Tables 8, 9, 10)\(^ {9,28,29}\). The $HLA-DRB1*15:01$
association overlaps with that of ulcerative colitis (risk increasing) and Crohn’s disease (risk decreasing)\(^{30,31}\). Since imputed genotypes at the class II region were only available for four \((HLA-DRB1, HLA-DQB1, HLA-DQA1 \text{ and } HLA-DPB1)\) out of 20 loci\(^{32}\), further studies involving direct sequencing of all HLA class II loci along with assessments of their protein structure and peptide binding are required to causally resolve the link between this HLA subregion and PSC development\(^{33,34}\).

Although 72% of the PSC patients in this study have a diagnosis of concomitant IBD (Supplementary Table 11), only half of our genome-wide significant loci were associated with IBD in the recent International IBD Genetics Consortium (IIBDGC) Immunochip analysis (Fig. 2a, Supplementary Table 3 and Supplementary Fig. 8)\(^{22}\), despite the greater sample size of that study (25,683 cases and 15,977 controls). Across the 12 non-HLA PSC loci we observed greater similarity between the OR estimates for PSC and ulcerative colitis than for PSC and Crohn’s disease. We used the Crohn’s disease and ulcerative colitis OR estimates for the 163 IBD-associated loci to predict PSC case/control status in our sample (Online Methods)\(^{22}\), and found a significantly greater area under the receiver operating characteristic curve (AUC) when applying ulcerative colitis ORs compared to Crohn’s disease ORs (ulcerative colitis AUC=0.62, Crohn’s disease AUC=0.56, \(P=1.2\times10^{-57}\), Fig. 2b). This suggests that PSC is genetically more similar to ulcerative colitis than Crohn’s disease and is consistent with clinical observations of greater comorbidity between PSC with ulcerative colitis than Crohn’s disease\(^{35}\). To further compare the genetic profile of PSC and IBD, we combined our genome-wide significant PSC loci with the 163 confirmed IBD loci\(^{22}\) in a functional similarity network (Supplementary Fig. 9 and Supplementary Table 12). The figure shows that the PSC loci are distributed throughout the IBD loci, suggesting that there is no particular functional subcluster of IBD susceptibility genes associated with PSC and vice versa.

While we consider only those loci reaching a stringent significance threshold \((P<5\times10^{-8})\) to be conclusively associated to PSC, it is likely that additional true associations lie among SNPs with weaker associations. An alternative approach for controlling for multiple hypothesis testing is false discovery rate (FDR) control, which regulates the expected proportion of incorrectly rejected null hypotheses. FDR is well suited to focused genotyping platforms such as Metabochip\(^{36}\) and Immunochip because it implicitly accounts for the expected enrichment of association. To further increase this enrichment, we exploited the known pleiotropy between related immune-mediated traits\(^{37}\), and calculated the FDR\(^{38-40}\) for association with PSC conditional on previously published summary statistics from each of the related phenotypes (yielding a per SNP conditional FDR)\(^{41}\) (Online Methods). We identified 33 non-HLA loci with a conditional FDR<0.001 in this analysis (Fig. 3), all of which showed suggestive significance \((5\times10^{-5} < P \leq 5\times10^{-8})\) in the standard association analysis (Supplementary Table 13 and Supplementary Figures 10-12). These loci were integrated in the functional similarity network analysis (Supplementary Fig. 13), highlighting potential candidate susceptibility genes.

In conclusion, the present study increases the number of genome-wide significant loci in PSC from seven to 16 (including the HLA complex). The nine novel variants together explain 0.9% of variance in PSC liability, increasing the total amount of variance explained by the 16 known loci to 7.3% (Online Methods). The data convincingly show that genetic susceptibility to PSC extends considerably beyond the risk factors involved in the closely related IBD phenotype and into autoimmune pathophysiology. Furthermore, analysis of pleiotropic immune-related genetic variants highlights 33 additional suggestive loci in PSC, overall representing major new avenues for research into disease pathogenesis.
Online Methods

Study Subjects

The study participants are described in the Supplementary Note and Supplementary Table 15.

Ethical approval

The patient recruitment was approved by the ethics committees or institutional review boards of all participating centers. Written informed consent was obtained from all participants.

Quality control

SNPs with a call rate <80% were removed prior to commencing sample QC (n=235). Per individual genotype call rate and heterozygosity rate were calculated using PLINK\(^42\) and outlying samples were identified using Aberrant\(^43\), which automatically identifies outliers from otherwise Gaussian distributions (Supplementary Fig. 1). A set of 20,837 LD-pruned ($r^2$<0.1) SNPs with MAF>10% present on both the Immunochip and the Illumina Omni2.5-8 array used in the 1000 Genomes project (see URLs) were used to estimate identity by descent and ancestry. For each pair of individuals with estimated identity by descent 0.9, the sample with lower call rate was removed (unless case/control status was discordant between the pair, in which case both samples were removed, n=92). Related individuals (0.1875<identity by descent<0.9) remained in the analysis to maximize power because the mixed model association analysis can correctly account for the relatedness. Principal components analysis, implemented in SMARTPCA (Eigenstrat)\(^44\), was used to identify samples of non-European ancestry. Principal components were defined using population samples from the 1000 Genomes project\(^45\) genotyped using the Illumina Omni2.5-8 genotyping array (see URLs) and then projected into cases and controls (Supplementary Fig. 2)\(^14,22,46\). Following sample QC, 3,789 PSC cases and 25,079 controls remained. SNPs with a minor allele frequency less than 0.1%, Hardy-Weinberg equilibrium $P<10^{-5}$, call rate lower than 98%, or failing the PLINK v1.07 non-random differential missing data rate test between cases and controls ($P<10^{-5}$) were excluded. After completion of marker QC (Supplementary Table 2), 131,220 SNPs were available for analysis, further reduced to 130,422 after cluster plot inspection (see below).

Statistical methods

Genomic inflation factor—The Immunochip contains 3,120 SNPs that were part of a bipolar disease replication effort and other non-immune-related studies. After QC, 2,544 of these were used as null markers to estimate the overall inflation of the distribution of association test statistics.

Imputation—Using 85,747 post-QC SNPs located in the Immunochip fine mapping regions, additional genotypes were imputed using IMPUTE2 with the 1000 Genomes Phase 1 (March, 2012) reference panel of 1,092 individuals\(^47\) and 744,740 SNPs. Imputation was performed separately in ten batches, with the case:control and country of origin ratios constant across batches. SNPs with a posterior probability less than 0.9 and those with differential missingness ($P<10^{-5}$) between the 10 batches were removed, as were SNPs failing the exclusion thresholds used for genotyped SNP QC. After imputation, a total of 163,379 SNPs in the Immunochip fine mapping regions, including 153,857 SNPs from the reference panel, were available for analysis.
**Association analysis**—Case-control association tests were performed using a linear mixed model as implemented in MMM\(^\text{15}\). A covariance matrix, \(R\), of a random effects component was included in the model to explicitly account for confounding due to population stratification and cryptic relatedness between individuals. This method has been shown to better control for population stratification than correction for principal components or meta-analyses of matched subgroups of cases and controls\(^{48-50}\). \(R\) is a symmetric \(n \times n\) matrix with each entry representing the relative sharing of alleles between two individuals compared to the average in the sample, and is typically estimated using genome-wide SNP data\(^{15}\). To avoid biases in the estimation of \(R\) due to the design of the Immunochip, SNPs were first pruned for LD (\(r^2 < 0.1\)). Of the remaining SNPs, we then removed those that lie in the HLA region or have a minor allele frequency <10%. Finally, we excluded SNPs that showed modest association (\(P < 0.005\)) with PSC in a linear regression model fitting the first 10 principal components as covariates. A total of 17,260 SNPs were used to estimate \(R\).

Due to computational limitations, we estimated the \(R\) matrix and performed all association analyses applying \(R\) separately for UK (n=9,696) and non-UK (n=19,172) samples, and then combined the results using a fixed-effects (inverse-variance weighting) meta-analysis, as done previously\(^{48}\). This reduced the \(\lambda_{\text{GC}}\), estimated using the 2,544 “null” SNPs, from 1.24 to 1.02 (Supplementary Fig. 14), showing excellent control for population stratification.

**Prediction of PSC using IBD SNPs**—ORs for Crohn’s disease and ulcerative colitis in 163 IBD-associated SNPs were obtained from Jostins et al.\(^{22}\). We used the R package Mangrove (see URLs) to estimate each individual’s probability of developing PSC among our 3,789 PSC cases and 25,079 controls assuming additive risk (log-additive OR). The performance of our predictor using either Crohn’s disease or ulcerative colitis ORs, was assessed by constructing a ROC curve, showing the proportion of true and false positives at each probability threshold. The AUC was calculated to compare the predictive power of the ulcerative colitis and Crohn’s disease ORs. The DeLong method was used to test if the AUC was significantly greater using ulcerative colitis than Crohn’s disease ORs\(^{52}\).

**Functional similarity networks**—In functional similarity networks, each edge represents strong functional similarity of two genes based on annotated Gene Ontology terms as determined by the functional similarity measure \(r_{\text{funSim}}\)\(^{53}\). The \(r_{\text{funSim}}\) similarity values above the recommended cutoff 0.8 were retrieved using the FunSimMat web service\(^{54}\). The resulting networks were visualized and analyzed using Cytoscape\(^{55}\).

To construct PSC-specific networks from functional similarity networks that contained more than one gene per locus (Supplementary Figures 5 and 13), the connectivity of each gene was assessed by computing different topology measures for the corresponding node: (1) degree (number of direct edges to other nodes), (2) shortest path closeness (inverted average shortest path distance to other nodes) and (3) shortest path betweenness (fraction of shortest paths passing through the node). Similarity edges between genes in the same locus and gene nodes that were not contained in the resulting largest connected subnetworks were ignored. The genes were first ranked according to each measure and then assigned the best of the three ranks. The PSC-specific network was generated from the top ranked genes in their respective locus.
Pleiotropy analysis—We included summary statistics from genome-wide association studies in seven PSC-associated diseases (Crohn’s disease, celiac disease, psoriasis, rheumatoid arthritis, sarcoidosis, type 1 diabetes, ulcerative colitis, see Supplementary Table 16). For all diseases we constructed conditional stratified Q-Q plots of empirical quantiles of nominal −log_{10}(P) values for SNP association with PSC for all SNPs (see Supplementary Fig. 10), and for different overlapping subsets of SNPs determined by the significance of their association with the PSC-associated autoimmune disorder (SNP subsets defined as P<1, P<0.1, P<0.01 and P<0.001 in the pleiotropic phenotype, respectively). For a given PSC associated phenotype, ‘enrichment’ for pleiotropic signals in PSC can be observed as an increasing leftward deflection from the expected null line with lower P-value thresholds in the second phenotype (Supplementary Note). The ‘enrichment’ in the stratified Q-Q plots is directly interpretable in terms of the true discovery rate (TDR), equivalent to one minus the FDR. Specifically, it can be shown that a conservative estimate of FDR can be calculated from the horizontal shift of the Q-Q curve from the expected line x=y, with a larger shift corresponding to a smaller FDR for a given nominal P-value (Supplementary Note). We calculated the conditional TDR as a function of P-value in PSC across a series of P-value thresholds in the pleiotropic trait (Supplementary Fig. 10).

In order to assess significance of the association with PSC, we assigned a pleiotropic (conditional) FDR value for PSC per SNP. The pleiotropic FDR value for each SNP is based on the P-value of the SNP in PSC relative to the P-value distribution of other SNPs in the same conditioning subset, where subsets are defined by the pleiotropic association (lowest P-value among associated diseases) of the SNP. Importantly, the conditioning procedure is blind to the P-value of the SNP with respect to PSC. The pleiotropic FDR is then interpolated from conditional FDR curves using established stratified FDR methods (see Supplementary Note). The increase in power from using pleiotropic FDR is demonstrated by dividing the total sample in half and observing that empirical replication rates between the training and test halves increase with decreasing P-value in the pleiotropic disease (Supplementary Fig. 15). The SNP with lowest FDR within each LD block (as defined by 1000 Genomes) was considered the lead SNP of a new pleiotropic PSC locus, if below a 0.001 threshold (loci defined by FDR<0.001 and FDR<0.01 shown in Supplementary Tables 13 and 14). All test statistics were adjusted for population stratification by genomic control (see Supplementary Note and Supplementary Fig. 16).

Variance explained and heritability—The proportion of variance explained by the genome-wide significant loci and HLA alleles was calculated using a liability threshold model assuming a disease prevalence of 10/100,000 and multiplicative risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all PSC patients and healthy controls for their participation, and we are indebted to all physicians and nursing staff who recruited patients. We thank Tanja Wesse, Tanja Henke, Sanaz Sedghpour Sabet, Rainer Vogler, Gunnar Jacobs, Ilona Urbach, Wolfgang Albrecht, Virpi Pelkonen, Kristian Holm, Hege Dahlén Sollid, Bente Woldseth, Jarl Andreas Anmarkrud and Liv Wenche Torbjørnsen for expert help. Prof. Ulrich Beuers, Dr. Felix Braun, Dr. Wolfgang Kreisel, Dr. Thomas Berg and Dr. Rainer Günther are acknowledged for contributing German PSC patients. Benedicte A. Lie and The Norwegian Bone Marrow Donor Registry at Oslo University Hospital, Rikshospitalet in Oslo and the Nord-Trøndelag Health Study (HUNT) are acknowledged for sharing the healthy Norwegian controls. Banco Nacional de ADN, Salamanca, Spain is acknowledged for providing Spanish control samples. This study makes use of genotyping data generated by DILGOM (see URLs), the Cooperative Research in the Region of Augsburg (KORA) study and by the Heinz Nixdorf Recall (Risk Factors, Evaluation of Coronary Calcification, and Lifestyle) study. We acknowledge the members of the International PSC Study Group, the
NIDDK Inflammatory Bowel Disease Genetics Consortium (IBDGC), the UK-PSC Consortium and the Alberta IBD Consortium for their participation. Individuals who have been sharing summary statistics and statistical software are acknowledged in the Supplementary Note.

The study was supported by The Norwegian PSC Research Center (see URLs), by the German Ministry of Education and Research through the National Genome Research Network (01GS0809-GP7), by the Deutsche Forschungsgemeinschaft (FR 2821/2-1), by the EU Seventh Framework Programme FP7/2007-2013 (262055) ESGI, by the Integrated Research and Treatment Center - Transplantation (01EO0802) and the PopGen biobank (see URLs).

J.Z.L., T.S., C.A.A. are supported by a grant from the Wellcome Trust (098051). Additional financial support to the study and the co-authors is listed in the Supplementary Note.

References


Figure 1. Association results across the nine newly associated primary sclerosing cholangitis (PSC) loci
Regional association plots of the 9 loci newly associated with PSC at genome-wide significance ($P<5 \times 10^{-8}$). Filled-in circles are directly genotyped and hollow triangles are imputed (see Online Methods) single-nucleotide polymorphisms (SNPs). The color of the marker (see legend in panel a) illustrates linkage disequilibrium with the most associated SNP. Since the most associated SNPs in panels d and e are located outside Immunochip fine mapping regions, association results from the discovery panel of the largest PSC genome-wide association study to date$^{12}$ are shown as hollow circles and the most associated SNP is a hollow diamond (genotyped and imputed HapMap release 22 SNPs, cases overlap with the current study).
Liu et al. Page 14

Nat Genet. Author manuscript; available in PMC 2013 December 01.
Figure 2. Genetic similarity of primary sclerosing cholangitis (PSC) and inflammatory bowel disease (IBD) associated loci

a) Comparison of odds ratios (ORs) between the most associated risk allele in PSC and the same allele in Crohn’s disease and ulcerative colitis across the 12 genome-wide significant PSC-associated loci. Ulcerative colitis and Crohn’s disease ORs, and the denomination of IBD loci as ulcerative colitis, Crohn’s disease or both (IBD), were obtained from Jostins et al.22. The error bars represent 95% confidence intervals around the ORs. *The PSC associated alleles at 6q15 (BACH2), 10p15 (IL2RA) and 19q13 (PRKD2) are independent of the reported IBD associations ($r^2<0.3$), but locate to the same broad genetic region; for this reason these loci are defined as PSC-IBD loci in Supplementary Figure 9, but as PSC only allelic associations here.

b) Predicting PSC using OR estimates across 163 IBD-associated loci. The green and orange lines represent the receiver operating characteristic (ROC) curves for discriminating PSC cases from population controls using ulcerative colitis or Crohn’s disease ORs from 163 loci.
associated with IBD, respectively\textsuperscript{22}. The dashed diagonal line is $y=x$, and specifies the ROC curve of a random predictor.
Figure 3. Pleiotropic primary sclerosing cholangitis (PSC) loci Manhattan plot of conditional associations in PSC calculated as stratified false discovery rates (FDRs) based on the results of the present PSC analysis and genetic associations previously reported in seven immune-mediated diseases (Crohn’s disease, celiac disease, psoriasis, rheumatoid arthritis, sarcoidosis; type 1 diabetes and ulcerative colitis) (see Online Methods and Supplementary Figures 10-12). Single-nucleotide polymorphisms (SNPs) in red represent genome-wide significant findings from the regular association analysis, while SNPs listed in black are significantly associated with PSC conditional on their pleiotropic effects across the related immune-mediated diseases. The horizontal red line represents a threshold of FDR<0.001 (Supplementary Table 13), while the blue line represents a threshold of FDR<0.01 (see Supplementary Table 14 for full listing).
Table 1

Association results of twelve non-HLA genome-wide significant risk loci for primary sclerosing cholangitis (PSC).

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>RA</th>
<th>RAF cases</th>
<th>RAF controls</th>
<th>P-value</th>
<th>LD region (Kb)</th>
<th>RefSeq genes in LD region</th>
<th>Notable nearby gene(s)</th>
<th>Functional annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1p36</td>
<td>rs3748816</td>
<td>A</td>
<td>0.698</td>
<td>0.656</td>
<td>1.21 (1.14-1.27)</td>
<td>2,398-2,775</td>
<td>9 MMEL1, TNFRSF14</td>
<td>eQTL, MS, OC, PB, HM</td>
<td></td>
</tr>
<tr>
<td>2q33</td>
<td>rs742056</td>
<td>A</td>
<td>0.277</td>
<td>0.229</td>
<td>1.89 (1.23-1.37)</td>
<td>204,155-204,397</td>
<td>1 CD28</td>
<td>HM, OC</td>
<td></td>
</tr>
<tr>
<td>3p21</td>
<td>rs3197999</td>
<td>A</td>
<td>0.352</td>
<td>0.285</td>
<td>2.45 (1.26-1.4)</td>
<td>48,388-51,358</td>
<td>89 MST1</td>
<td>eQTL, MS, OC, PB</td>
<td></td>
</tr>
<tr>
<td>4q27</td>
<td>rs13140464</td>
<td>C</td>
<td>0.871</td>
<td>0.836</td>
<td>8.87 (1.21-1.4)</td>
<td>123,204-123,784</td>
<td>4 IL2, IL21</td>
<td>OC, PB</td>
<td></td>
</tr>
<tr>
<td>6q15</td>
<td>rs56258221</td>
<td>G</td>
<td>0.213</td>
<td>0.183</td>
<td>8.36 (1.16-1.31)</td>
<td>90,967-91,150</td>
<td>1 BACH2</td>
<td>OC, PB</td>
<td></td>
</tr>
<tr>
<td>10p15</td>
<td>rs4147359</td>
<td>A</td>
<td>0.401</td>
<td>0.349</td>
<td>8.19 (1.18-1.3)</td>
<td>6,070-6,206</td>
<td>2 IL2RA</td>
<td>PB</td>
<td></td>
</tr>
<tr>
<td>11q23</td>
<td>rs7937682</td>
<td>G</td>
<td>0.298</td>
<td>0.265</td>
<td>3.17 (1.11-1.24)</td>
<td>110,824-111,492</td>
<td>19 SIK2</td>
<td>OC, PB, HM</td>
<td></td>
</tr>
<tr>
<td>12q13</td>
<td>rs11168249</td>
<td>G</td>
<td>0.506</td>
<td>0.466</td>
<td>5.49 (1.11-1.21)</td>
<td>46,442-46,534</td>
<td>3 HDAC7</td>
<td>OC, PB, HM</td>
<td></td>
</tr>
<tr>
<td>12q24</td>
<td>rs3184504</td>
<td>A</td>
<td>0.527</td>
<td>0.488</td>
<td>5.91 (1.12-1.24)</td>
<td>110,186-111,512</td>
<td>16 SH2B3, ATXN2</td>
<td>MS, OC, HM</td>
<td></td>
</tr>
<tr>
<td>18q22</td>
<td>rs1788097</td>
<td>A</td>
<td>0.518</td>
<td>0.483</td>
<td>3.06 (1.11-1.21)</td>
<td>65,633-65,721</td>
<td>2 CD26</td>
<td>MS, OC, PB, HM</td>
<td></td>
</tr>
<tr>
<td>19q13</td>
<td>rs60652743</td>
<td>A</td>
<td>0.864</td>
<td>0.836</td>
<td>6.51 (1.16-1.34)</td>
<td>51,850-51,998</td>
<td>6 PRKD2, STRN4</td>
<td>OC, PB, HM</td>
<td></td>
</tr>
<tr>
<td>21q22</td>
<td>rs2836883</td>
<td>G</td>
<td>0.777</td>
<td>0.728</td>
<td>3.19 (1.21-1.36)</td>
<td>39,374-39,404</td>
<td>4 PSMG1</td>
<td>OC, PB, HM</td>
<td></td>
</tr>
</tbody>
</table>

Chr: chromosome; CI: confidence interval; eQTL: expression quantitative trait locus; HM: overlaps a region of histone modification; Kb: kilobasepairs; LD: linkage disequilibrium; MS: missense mutation; OC: overlaps known region of open chromatin; OR: odds ratio; PB: overlaps a region of protein binding; RA: risk allele; RAF: risk allele frequency

aSNPs from novel PSC-associated loci are shown in bold.

bLD regions around lead SNPs were calculated by extending in both directions a distance of 0.1 centimorgans as defined by the HapMap recombination map.

cCandidate gene(s) within same LD region as the associated SNPs.

Denotes if there are SNPs with r²>0.8 with the hit SNP that have functional annotations (Supplementary Tables 4-7).