Europe PMC Funders Group
Author Manuscript
Nat Genet. Author manuscript; available in PMC 2013 December 01.

Published in final edited form as:
Nat Genet. 2013 June ; 45(6): 670–675. doi:10.1038/ng.2616.

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

Jimmy Z. Liu[1,*], **Johannes Roksund Hov**[2,3,4,5,*], **Trine Folseraas**[2,3,4,*], **Eva Ellinghaus**[6,*], **Simon M. Rushbrook**[7], **Nadezhda T. Doncheva**[8], **Ole A. Andreassen**[4,9], **Rinse K. Weersma**[10], **Tobias J. Weismüller**[11,12,85], **Bertus Eksteen**[13], **Pietro Invernizzi**[14], **Gideon M. Hirschfield**[15,16], **Daniel Nils Gotthardt**[17], **Albert Pares**[18], **David Ellinghaus**[6], **Tejas Shah**[1], **Brian D. Juran**[19], **Piotr Milkiewicz**[20], **Christian Rust**[21], **Christoph Schramm**[22], **Tobias Müller**[23], **Brijesh Srivastava**[24], **Georgios Dalekos**[25,26], **Markus M. Nöthen**[27,28], **Stefan Herms**[27,28], **Juliane Winkelmann**[29,30,31], **Mitja Mitrovic**[32], **Felix Braun**[33], **Cyriel Y. Ponsioen**[34], **Peter J. P. Croucher**[35], **Martina Sterneck**[36], **Andreas Teufel**[37], **Andrew L. Mason**[38], **Janna Saarela**[39], **Virpi Leppa**[40], **Ruslan Dorfman**[41], **Domenico Alvaro**[42], **Annarosa Floreani**[43], **Suna Onengut-Gumuscu**[44,45], **Stephen S. Rich**[46,47], **Wesley K. Thompson**[48], **Andrew J. Schork**[49], **Sigrid Næss**[2,3,4], **Ingo Thomsen**[6], **Gabriele Mayr**[8], **Inke R. König**[50], **Kristian Hveem**[51], **Isabelle Cleynen**[1,52], **Javier Gutierrez-Achury**[32], **Isis Ricaño-Ponce**[32], **David van Heel**[53], **Einar Björnsson**[54], **Richard N. Sandford**[24], **Peter R. Durie**[55], **Espen Melum**[2,3,4], **Morten H Vatn**[4,5,56], **Mark S. Silverberg**[57], **Richard H. Duerr**[58,59], **Leonid Padyukov**[60], **Stephan Brand**[61], **Miquel Sans**[62], **Vito Annese**[63,64], **Jean-Paul Achkar**[65,66], **Kirsten Muri Boberg**[2,4,5], **Hanns-Ulrich Marschall**[67], **Olivier Chazouillères**[68], **Christopher L. Bowlus**[69], **Cisca Wijmenga**[32], **Erik Schrumpf**[2,4,5], **Severine Vermeire**[52,70], **Mario Albrecht**[8,71], **The UK-PSC Consortium**[72], **The International IBD Genetics Consortium**[72], **John D. Rioux**[73,74], **Graeme Alexander**[75], **Annika Bergquist**[76], **Judy Cho**[77], **Stefan Schreiber**[6,78,79], **Michael P. Manns**[11,12], **Martti Färkkilä**[80], **Anders M. Dale**[81,82], **Roger W. Chapman**[83], **Konstantinos N. Lazaridis**[19], **The International PSC Study Group**[72], **Andre Franke**[6,§], **Carl A. Anderson**[1,§], and **Tom H. Karlsen**[2,3,5,84,§]

[†]**Contact 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

J.Z.L., J.R.H., T.F., E.E., N.T.D., I.T., G.M., I.K., O.A.A., W.K.T, A.M.D., T.S. and C.A.A. performed data and statistical analysis. A.F., C.A.A. and T.H.K. coordinated the project and supervised the data-analysis. J.Z.L., J.R.H., T.F., E.E., A.F., C.A.A. and T.H.K. drafted the manuscript. S.M.R., R.K.W., T.J.W., B.E., P.I., G.M.H., D.N.G., A.P., D.E., B.D.J., P.M., C.R., C.S., T.M., B.S., G.D., M.M.N.,S.H., J.W., M.M., F.B., C.Y.P., P.J.P.C., M.S., A.T., A.L.M., J.S., V.L., V.P., R.D., D.A., A.Fl., S.O-G., S.S.R., A.J.S., S.N., K.H., I.C., J .G-A., I.R-P., D.H., E.B., R.N.S., P.R.D., E.M., M.H.V., M.S., R.H.D., L.P., S.B., M.S., V.A., J-P.A., K.M.B., H-U.M., O.C., C.L.B., C.W., E.S., S.V., M.A., J.D.R., G.A., A.B., J.C., S.S., M.P.M., M.F., R.W.C., K,N.L., The UK-PSC Consortium, The International IBD Consortium and The International PSC Study Group contributed to the patient ascertainment, sample collection and/or clinical data. All authors revised the manuscript for critical content and approved the final version.

Competing Interests Statement

The authors declare no competing interests.

URLs

1000 Genomes Omni2.5M genotype data: (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20110921_phase2_omni_genoty pes/Omni25_genotypes_1856_samples.b36.20110921.vcf.gz) Mangrove: (http://cran.r-project.org/web/packages/Mangrove/) DILGOM: (http://www.aka.fi/en-GB/A/Programmes-and-Cooperation/Research-programmes/Ongoing/ELVIRA/Projects/DILGOM/) The Norwegian PSC Research Center: (http://ous-research.no/nopsc/) PopGen biobank: (http://www.popgen.de)

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK [2]Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway [3]Research Institute of Internal Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway [4]Institute of Clinical Medicine, University of Oslo, Oslo, Norway [5]Section of Gastroenterology, Department of Transplantation Medicine, Division of Cancer, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway [6]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany [7]Department of Gastroenterology and Hepatology, Norfolk and Norwich, University Hospitals NHS Trust, Norwich, UK [8]Max Planck Institute for Informatics, Saarbrücken, Germany [9]KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital, Ulleval, Oslo, Norway [10]Department of Gastroenterology and Hepatology, University of Groningen and University Medical Centre Groningen, Groningen, the Netherlands [11]Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany [12]Integrated Research and Treatment Center-Transplantation (IFB-tx), Hannover Medical School, Hannover, Germany [13]Snyder Institute of Chronic Diseases, Department of Medicine, University of Calgary, Calgary, Canada [14]Center for Autoimmune Liver Diseases, Humanitas Clinical and Research Center, Rozzano (MI), Italy [15]Division of Gastroenterology, Department of Medicine, University of Toronto, Toronto, Canada [16]Centre for Liver Research, NIHR Biomedical Research Unit, Birmingham, UK [17]Department of Medicine, University Hospital of Heidelberg, Heidelberg, Germany [18]Liver Unit, Hospital Clínic, IDIBAPS, CIBERehd, University of Barcelona, Barcelona, Spain [19]Center for Basic Research in Digestive Diseases, Division of Gastroenterology and Hepatology, Mayo Clinic, College of Medicine, Rochester, Minnesota, USA [20]Liver Unit and Liver Research Laboratories, Pomeranian Medical University, Szczecin, Poland [21]Department of Medicine 2, Grosshadern, University of Munich, Munich, Germany [22]1st Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany [23]Department of Internal Medicine, Hepatology and Gastroenterology, Charité Universitätsmedizin Berlin, Berlin, Germany [24]Academic Department of Medical Genetics, University of Cambridge, Cambridge, UK [25]Department of Medicine, Medical School, University of Thessaly, Larissa, Greece [26]Research Laboratory of Internal Medicine, Medical School, University of Thessaly, Larissa, Greece [27]Institute of Human Genetics, University of Bonn, Bonn, Germany [28]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany [29]Institute of Human Genetics, Technische Universität München, Munich, Germany [30]Department of Neurology, Technische Universität München, Munich, Germany [31]Institute of Human Genetics, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany [32]Department of Genetics, University of Groningen and University Medical Centre Groningen, Groningen, The Netherlands [33]Department of General, Visceral, Thoracic, Transplantation and Pediatric Surgery, University Medical Centre Schleswig-Holstein, Campus Kiel, Germany [34]Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, the Netherlands [35]Department of Environmental Science, Policy, and Management, University of California, Berkeley, United States of America [36]Department of Hepatobiliary Surgery and Transplantation, University Medical Center Hamburg-Eppendorf, Hamburg, Germany [37]1st Department of Medicine, University of Mainz, Mainz, Germany [38]Division of Gastroenterology and Hepatology, University of Alberta, Edmonton, Alberta, Canada [39]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland [40]Public Health Genomics Unit, Institute for Molecular Medicine Finland FIMM, University of Helsinki and National Institute for Health and Welfare, Helsinki, Finland [41]Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Canada [42]Department of Clinical Medicine, Division of Gastroenterology, Sapienza University of Rome, Rome, Italy [43]Dept. of Surgical, Oncological and Gastroenterological Sciences, University of Padova, Padova, Italy [44]Center for Public Health Genomics, Division of Endocrinology & Metabolism, University of Virginia, Charlottesville, USA [45]Department of Internal Medicine, Division of Endocrinology & Metabolism, University of Virginia,

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Charlottesville, USA [46]Center for Public Health Genomics, University of Virginia, Charlottesville, USA [47]Department of Public Health Sciences, University of Virginia, Charlottesville, USA [48]Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA [49]Graduate Program in Cognitive Science, University of California, San Diego, La Jolla, CA, USA [50]Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany [51]Department of Public Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway [52]Department of Clinical and Experimental Medicine, KU Leuven, Leuven, Belgium [53]Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK [54]Department of Internal Medicine, Division of Gastroenterology and Hepatology, Landspitali University Hospital, Reykjavik, Iceland [55]Physiology and Experimental Medicine, Research Institute, Hospital for Sick Children, Toronto, Ontario, Canada [56]EpiGen, Campus AHUS, Akershus University Hospital, Nordbyhagen, Norway [57]Inflammatory Bowel Disease (IBD) Group, Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital Toronto, Ontario, Canada [58]Division of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA [59]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA [60]Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden [61]Department of Medicine II, University Hospital Munich-Grosshadern, Ludwig-Maximilians-University Munich, Germany [62]Department of Digestive Diseases, Centro Médico Teknon, Barcelona, Spain [63]Division of Gastroenterology, Istituto di Ricovero e Cura a Carattere Scientifico-Casa Sollievodella Sofferenza Hospital, San Giovanni Rotondo, Italy [64]Unit of Gastroenterology SOD2, Azienda Ospedaliero Universitaria Careggi, Florence, Italy [65]Department of Gastroenterology and Hepatology, Digestive Disease Institute, Cleveland Clinic, Cleveland, OH, USA [66]Department of Pathobiology, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA [67]Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy and University Hospital, Gothenburg, Sweden [68]AP-HP, Hôpital Saint Antoine, Department of Hepatology, UPMC Univ Paris 06, Paris, France [69]Division of Gastroenterology and Hepatology, University of California Davis, Davis, CA, USA [70]Department of Gastroenterology, University Hospitals Leuven, Leuven, Belgium [71]Department of Bioinformatics, Institute of Biometrics and Medical Informatics, University Medicine Greifswald, Greifswald, Germany [72]Further details on the consortium members are provided in the Supplementary Note [73]Université de Montréal, Research Center, Montreal, Quebec, Canada [74]Montreal Heart Institute, Research Center, Montreal, Quebec, Canada [75]Department of Medicine, Division of Hepatology, University of Cambridge, Cambridge, UK [76]Department of Gastroenterology and Hepatology, Karolinska University Hospital Huddinge, Karolinska Institutet, Stockholm, Sweden [77]Department of Medicine, Section of Digestive Diseases, Yale University, New Haven, Connecticut, USA [78]Department for General Internal Medicine, Christian-Albrechts-University, Kiel, Germany [79]Popgen Biobank, University Hospital Schleswig-Holstein, Christian-Albrechts-University, 24105 Kiel, Germany [80]Division of Gastroenterology, Department of Medicine, Helsinki University Hospital, Finland [81]Department of Radiology, University of California, San Diego, La Jolla, CA, USA [82]Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA [83]Department of Hepatology, John Radcliffe University Hospitals NHS Trust, Oxford, UK [84]Division of Gastroenterology, Institute of Medicine, University of Bergen, Bergen, Norway [85]Current affiliation: Department of Internal Medicine 1, University Hospital of Bonn, Bonn, Germany

Keywords

genetic association study; disease genetics; immunogenetics; liver

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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[1-3]. We compared 3,789 European ancestry PSC cases to 25,079 population controls across 130,422 single-nucleotide polymorphisms (SNPs) genotyped using Immunochip[4]. 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[5], 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%)[5,6] and autoimmune diseases (25%)[7]. Conversely, approximately only 5% of patients with IBD develop PSC[5,6]. A 9-39-fold sibling relative risk indicates a strong genetic component to PSC risk[8]. 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*)[9-13].

Several theories have been proposed to explain the development of PSC[5]. 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[11]. We genotyped 196,524 SNPs in 4,228 PSC cases and 27,077 population controls (see Online Methods and Supplementary Note) using the Immunochip[4,14], 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 (Supplementary Tables 1 and 2, Supplementary Figures 1 and 2). 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[15] to minimize the effect of population stratification and sample relatedness (λ_{GC} = 1.02, estimated using 2,544 “null” SNPs, see Online Methods).

We identified twelve non-HLA genome-wide significant ($P<5\times10^{-8}$) 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[16] within each locus did not identify additional independent genome-wide significant signals, nor did genotype-genotype or gender-genotype interaction analyses (Online Methods).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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=8.36×10^{-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=3.18×10^{-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=5.49×10^{-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=6.51×10^{-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

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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* 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=1.2x10^{-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 ($5\times10^{-5}\leq P<5\times10^{-8}$) 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.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Association analysis—Case-control association tests were performed using a linear mixed model as implemented in MMM[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 $\lambda_{GC,}$ estimated using the 2,544 "null" SNPS, from 1.24 to 1.02 (Supplementary Fig. 14), showing excellent control for population stratification. Stepwise conditional regression was used to identify possible independent associations at genome-wide significant loci. SNP×SNP interactions between all pairs of genome-wide significant SNPs were tested using the PLINK --epistasis command. Signal intensity plots of all non-HLA loci with association *P*-value<5×10^{-6} were visually inspected using Evoker[51]. SNPs that clustered poorly were removed (n=798).

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 rfunSim[53]. The rfunSim 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.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

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[56]. 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[41,57] (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[58] 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 Dahlen 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

1. Aadland E, et al. Primary sclerosing cholangitis: a long-term follow-up study. Scand J Gastroenterol. 1987; 22:655–64. [PubMed: 3659828]
2. Broome U, et al. Natural history and prognostic factors in 305 Swedish patients with primary sclerosing cholangitis. Gut. 1996; 38:610–5. [PubMed: 8707097]
3. Farrant JM, et al. Natural history and prognostic variables in primary sclerosing cholangitis. Gastroenterology. 1991; 100:1710–7. [PubMed: 1850376]
4. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011; 13:101. [PubMed: 21345260]
5. Karlsen TH, Schrumpf E, Boberg KM. Update on primary sclerosing cholangitis. Dig Liver Dis. 2010; 42:390–400. [PubMed: 20172772]
6. Karlsen TH, Kaser A. Deciphering the genetic predisposition to primary sclerosing cholangitis. Semin Liver Dis. 2011; 31:188–207. [PubMed: 21538284]
7. Saarinen S, Olerup O, Broome U. Increased frequency of autoimmune diseases in patients with primary sclerosing cholangitis. Am J Gastroenterol. 2000; 95:3195–9. [PubMed: 11095341]
8. Bergquist A, et al. Increased risk of primary sclerosing cholangitis and ulcerative colitis in first-degree relatives of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol. 2008; 6:939–43. [PubMed: 18674735]
9. Karlsen TH, et al. Genome-wide association analysis in primary sclerosing cholangitis. Gastroenterology. 2010; 138:1102–11. [PubMed: 19944697]
10. Srivastava B, et al. Fine mapping and replication of genetic risk loci in primary sclerosing cholangitis. Scand J Gastroenterol. 2012; 47:820–6. [PubMed: 22554193]
11. Folseraas T, et al. Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol. 2012; 57:366–75. [PubMed: 22521342]
12. Melum E, et al. Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. Nat Genet. 2011; 43:17–9. [PubMed: 21151127]
13. Ellinghaus D, et al. Genome-wide association analysis in sclerosing cholangitis and ulcerative colitis identifies risk loci at GPR35 and TCF4. Hepatology. 2012
14. Trynka G, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011; 43:1193–201. [PubMed: 22057235]
15. Pirinen M, Donnelly P, Spencer C. Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. Ann Appl Stat. 2012 In press.
16>. Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. Am J Hum Genet. 2002; 70:124–41. [PubMed: 11719900]
17. Peters U, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet. 2012; 131:217–34. [PubMed: 21761138]
18. Gerstein MB, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012; 489:91–100. [PubMed: 22955619]
19. Rossin EJ, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet. 2011; 7:e1001273. [PubMed: 21249183]

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

20. Raychaudhuri S, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 2009; 5:e1000534. [PubMed: 19557189]
21. Hanna RN, et al. The transcription factor NR4A1 (Nur77) controls bone marrow differentiation and the survival of Ly6C- monocytes. Nat Immunol. 2011; 12:778–85. [PubMed: 21725321]
22. Jostins L, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012; 491:119–24. [PubMed: 23128233]
23. Kasler HG, et al. Histone deacetylase 7 regulates cell survival and TCR signaling in CD4/CD8 double-positive thymocytes. J Immunol. 2011; 186:4782–93. [PubMed: 21398603]
24. Dequiedt F, et al. HDAC7, a thymus-specific class II histone deacetylase, regulates Nur77 transcription and TCR-mediated apoptosis. Immunity. 2003; 18:687–98. [PubMed: 12753745]
25. Dequiedt F, et al. Phosphorylation of histone deacetylase 7 by protein kinase D mediates T cell receptor-induced Nur77 expression and apoptosis. J Exp Med. 2005; 201:793–804. [PubMed: 15738054]
26. Clark K, et al. Phosphorylation of CRTC3 by the salt-inducible kinases controls the interconversion of classically activated and regulatory macrophages. Proc Natl Acad Sci U S A. 2012
27. Raychaudhuri S, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012; 44:291–6. [PubMed: 22286218]
28. Schrumpf E, et al. HLA antigens and immunoregulatory T cells in ulcerative colitis associated with hepatobiliary disease. Scand J Gastroenterol. 1982; 17:187–91. [PubMed: 6982501]
29. Spurkland A, et al. HLA class II haplotypes in primary sclerosing cholangitis patients from five European populations. Tissue Antigens. 1999; 53:459–69. [PubMed: 10372541]
30. Stokkers PC, Reitsma PH, Tytgat GN, van Deventer SJ. HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis. Gut. 1999; 45:395–401. [PubMed: 10446108]
31. Okada Y, et al. HLA-Cw*1202-B*5201-DRB1*1502 haplotype increases risk for ulcerative colitis but reduces risk for Crohn's disease. Gastroenterology. 2011; 141:864–871. e1–5. [PubMed: 21699788]
32. Horton R, et al. Gene map of the extended human MHC. Nature. 2004; 5:889–99.
33. Hov JR, et al. Electrostatic modifications of the human leukocyte antigen-DR P9 peptide-binding pocket and susceptibility to primary sclerosing cholangitis. Hepatology. 2011; 53:1967–76. [PubMed: 21413052]
34. Hovhannisyan Z, et al. The role of HLA-DQ8 beta57 polymorphism in the anti-gluten T-cell response in coeliac disease. Nature. 2008; 456:534–8. [PubMed: 19037317]
35. Broome U, Bergquist A. Primary sclerosing cholangitis, inflammatory bowel disease, and colon cancer. Semin Liver Dis. 2006; 26:31–41. [PubMed: 16496231]
36. The CDC, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet. 2012
37. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat Rev Genet. 2009; 10:43–55. [PubMed: 19092835]
38. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Roy Statist Soc Ser B. 1995; 57:289–300.
39. Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. Ann Statist. 2003; 31:2013–2035.
40. Efron B. Simultaneous Inference: When Should Hypothesis Testing Problems Be Combined? Ann Appl Statist. 2008; 2:197–223.
41. Sun L, Craiu RV, Paterson AD, Bull SB. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. Genet Epidemiol. 2006; 30:519–30. [PubMed: 16800000]
42. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]

Europe PMC Funders Author Manuscripts

43. Bellenguez C, Strange A, Freeman C, Donnelly P, Spencer CC. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. Bioinformatics. 2012; 28:134–5. [PubMed: 22057162]
44. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]
45. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–73. [PubMed: 20981092]
46. Liu JZ, et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nat Genet. 2012
47. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]
48. Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011; 476:214–9. [PubMed: 21833088]
49. Korte A, et al. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet. 2012; 44:1066–71. [PubMed: 22902788]
50. Tsoi LC, et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat Genet. 2012; 44:1341–8. [PubMed: 23143594]
51. Morris JA, Randall JC, Maller JB, Barrett JC. Evoker: a visualization tool for genotype intensity data. Bioinformatics. 2010; 26:1786–7. [PubMed: 20507892]
52. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988; 44:837–45. [PubMed: 3203132]
53. Schlicker A, Domingues FS, Rahnenfuhrer J, Lengauer T. A new measure for functional similarity of gene products based on Gene Ontology. BMC Bioinformatics. 2006; 7:302. [PubMed: 16776819]
54. Schlicker A, Albrecht M. FunSimMat update: new features for exploring functional similarity. Nucleic Acids Res. 2010; 38:D244–8. [PubMed: 19923227]
55. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504. [PubMed: 14597658]
56. Efron B. Size, power and false discovery rates. Ann Statist. 2007; 35:1351–77.
57. Yoo YJ, Pinnaduwage D, Waggott D, Bull SB, Sun L. Genome-wide association analyses of North American Rheumatoid Arthritis Consortium and Framingham Heart Study data utilizing genome-wide linkage results. BMC Proc. 2009; 3(Suppl 7):S103. [PubMed: 20017967]
58. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol. 2011; 35:310–7. [PubMed: 21374718]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

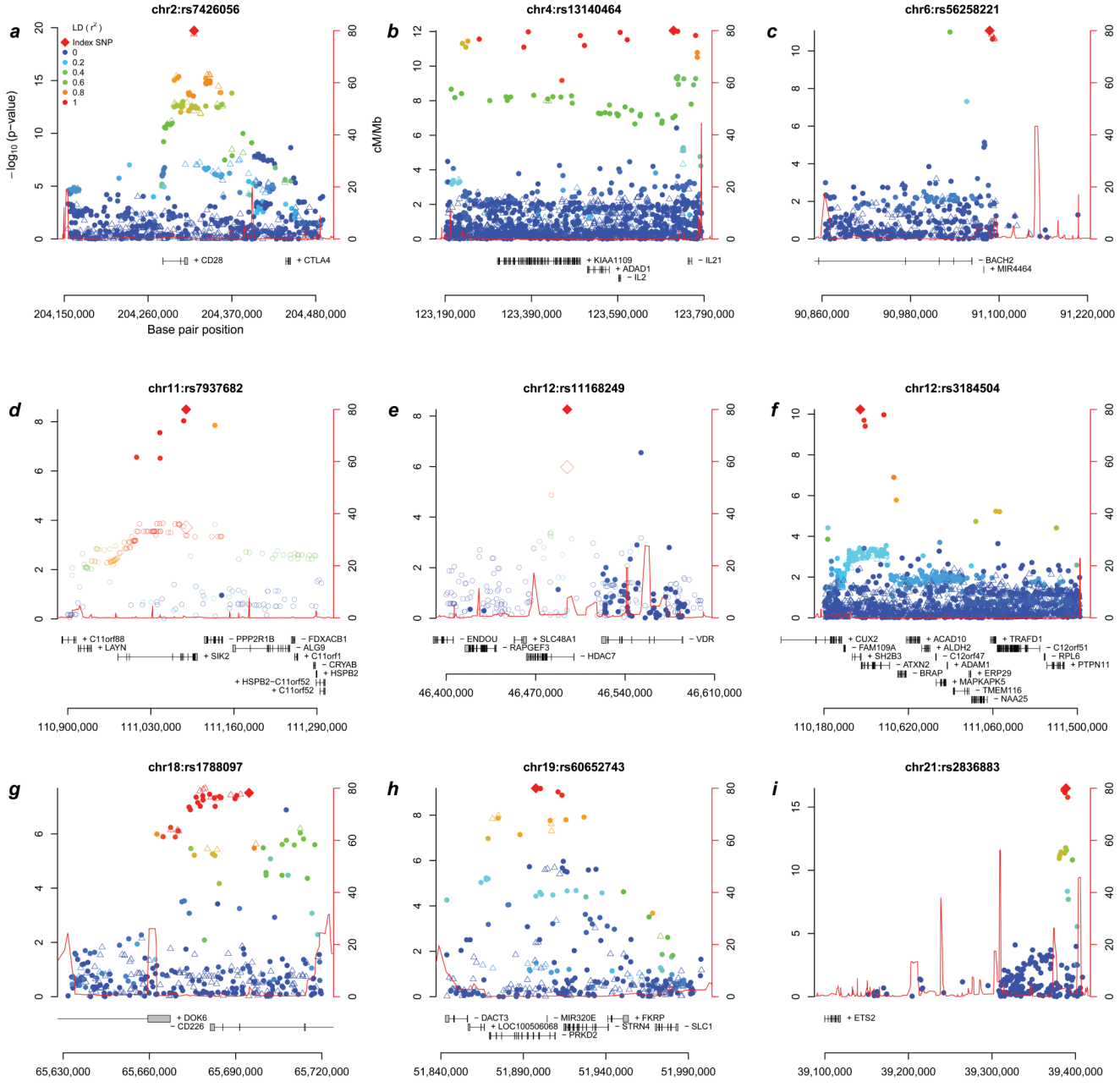
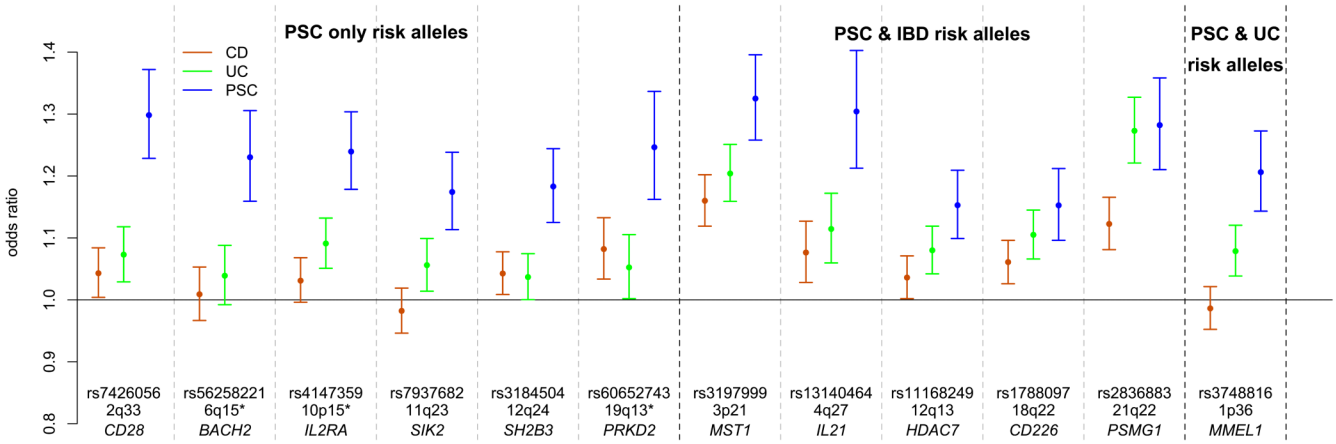


Figure 1. Association results across the nine newly associated primary sclerosing cholangitis (PSC) loci

Regional association plots of the 9 loci newly associated with PSC at genome-wide significance ($P<5\times10^{-8}$). 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).



Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

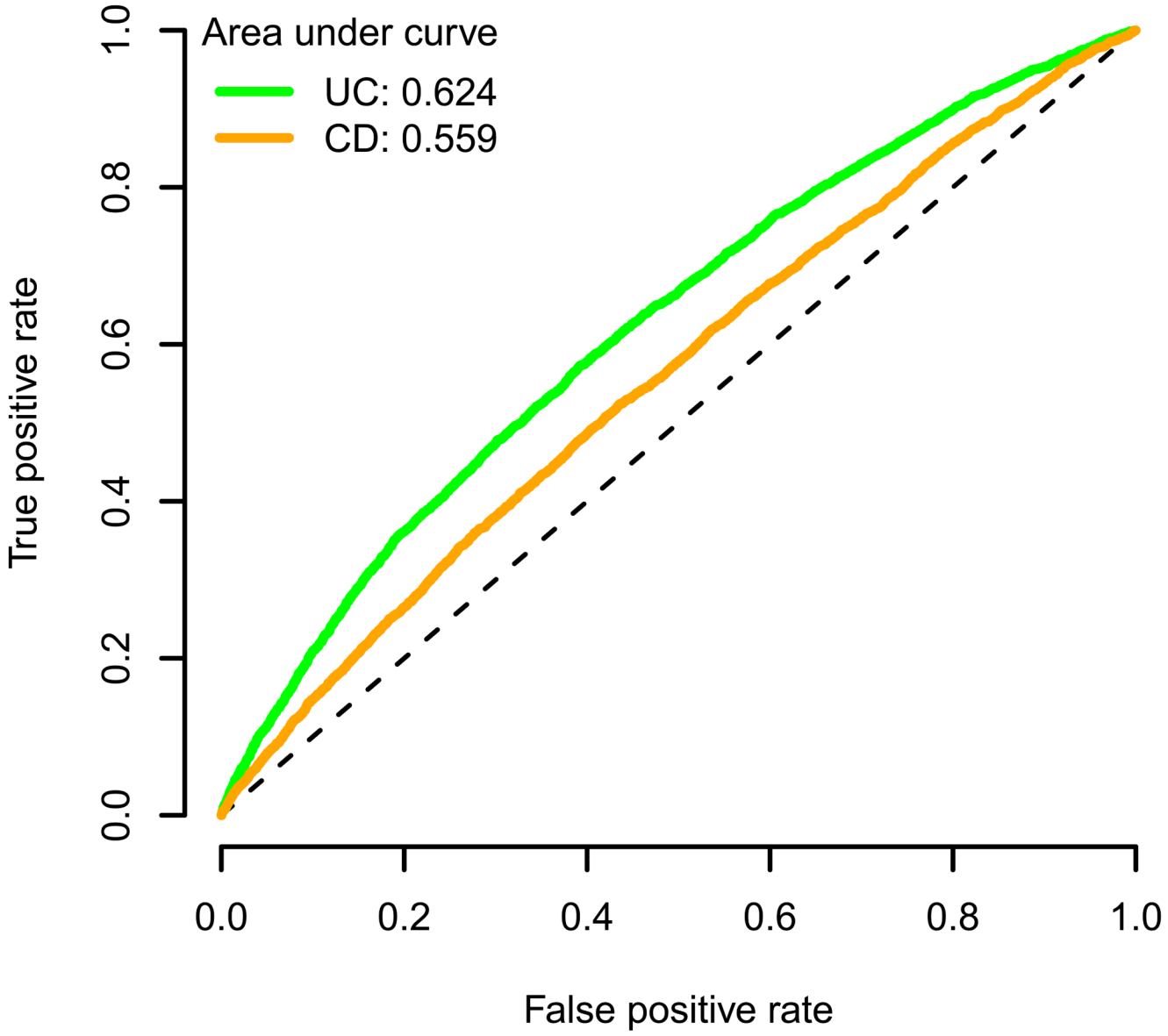


Figure 2. Genetic similarity of primary sclerosing cholangitis (PSC) and inflammatory bowel disease (IBD) associated loci

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[22]. The dashed diagonal line is $y=x$, and specifies the ROC curve of a random predictor.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

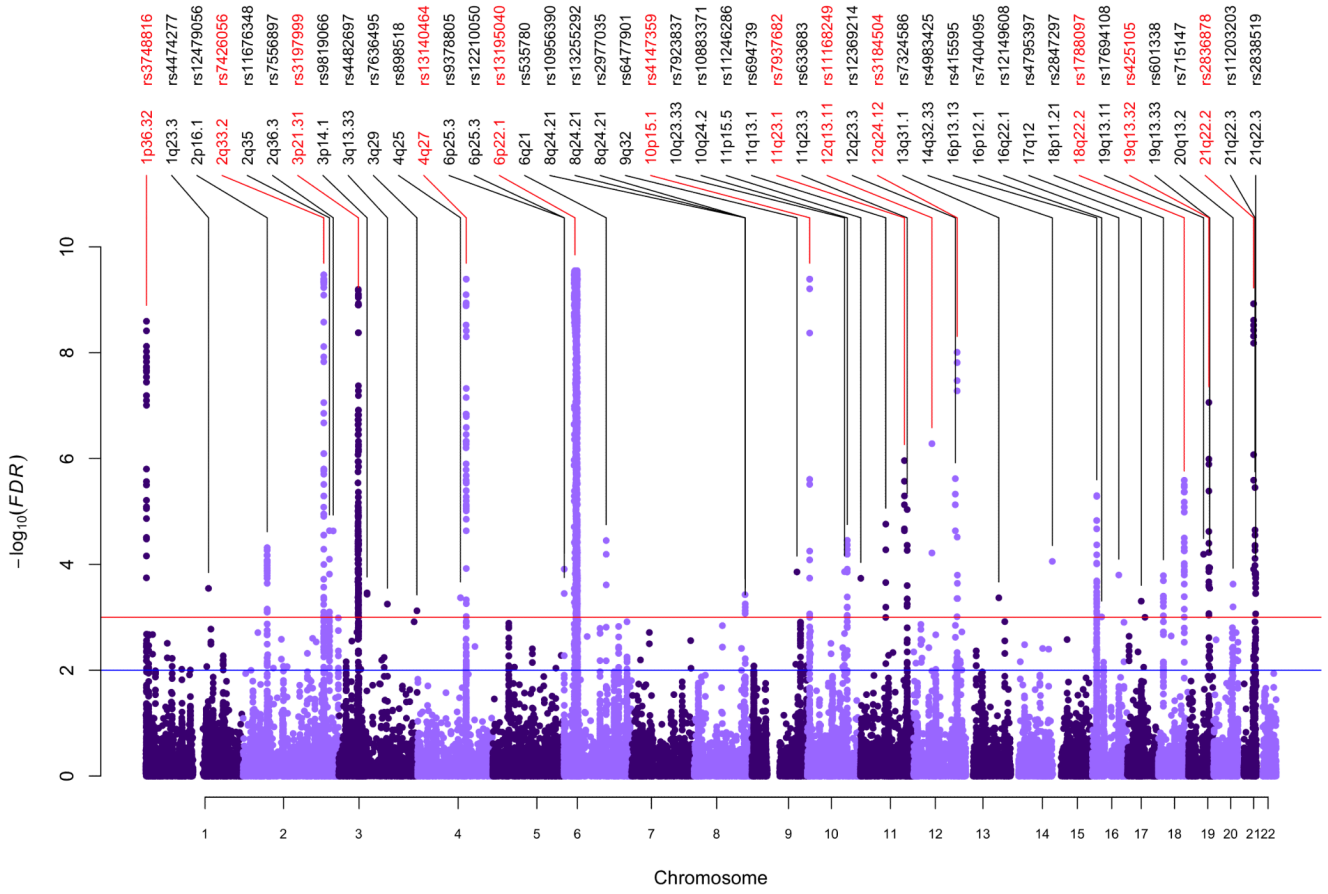


Figure 3. Pleiotropic primary sclerosing cholangitis (PSC) loci

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).

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Nat Genet. Author manuscript; available in PMC 2013 December 01.

Table 1

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

Chr	SNP[a]	RA	RAF cases	RAF controls	*P*-value	OR (95%CI)	LD region[b] (Kb)	RefSeq genes in LD region	Notable nearby gene(s)[c]	Functional annotation[d]
1p36	rs3748816	A	0.698	0.656	7.41×10^{-12}	1.21 (1.14-1.27)	2,398-2,775	9	*MMEL1, TNFRSF14*	eQTL,MS, OC, PB, HM
2q33	**rs7426056**	A	0.277	0.229	1.89×10^{-20}	1.3 (1.23-1.37)	204,155-204,397	1	*CD28*	HM, OC
3p21	rs3197999	A	0.352	0.285	2.45×10^{-26}	1.33 (1.26-1.4)	48,388-51,358	89	*MST1*	eQTL,MS, OC, PB HM
4q27	**rs13140464**	C	0.871	0.836	8.87×10^{-13}	1.3 (1.21-1.4)	123,204-123,784	4	*IL2, IL21*	OC, PB
6q15	**rs56258221**	G	0.213	0.183	8.36×10^{-12}	1.23 (1.16-1.31)	90,967-91,150	1	*BACH2*	OC, PB
10p15	rs4147359	A	0.401	0.349	8.19×10^{-17}	1.24 (1.18-1.3)	6,070-6,206	2	*IL2RA*	PB
11q23	**rs7937682**	G	0.298	0.265	3.17×10^{-09}	1.17 (1.11-1.24)	110,824-111,492	19	*SIK2*	OC, PB, HM
12q13	**rs11168249**	G	0.506	0.466	5.49×10^{-09}	1.15 (1.1-1.21)	46,442-46,534	3	*HDAC7*	OC, PB, HM
12q24	**rs3184504**	A	0.527	0.488	5.91×10^{-11}	1.18 (1.12-1.24)	110,186-111,512	16	*SH2B3, ATXN2*	MS, OC, HM
18q22	**rs1788097**	A	0.518	0.483	3.06×10^{-08}	1.15 (1.1-1.21)	65,633-65,721	2	*CD226*	MS, OC, PB, HM
19q13	**rs60652743**	A	0.864	0.836	6.51×10^{-10}	1.25 (1.16-1.34)	51,850-51,998	6	*PRKD2, STRN4*	OC, PB, HM
21q22	**rs2836883**	G	0.777	0.728	3.19×10^{-17}	1.28 (1.21-1.36)	39,374-39,404	-	*PSMG1*	OC, PB, HM

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

[a] SNPs from novel PSC-associated loci are shown in bold.

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

[c] Candidate gene(s) within same LD region as the associated SNPs.

[d] Denotes if there are SNPs with r^2>0.8 with the hit SNP that have functional annotations (Supplementary Tables 4-7).