

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

Primary sclerosing cholangitis (PSC) is a severe liver disease of unknown etiology leading to fibrotic destruction of the bile ducts and ultimately to the need for liver transplantation<sup>1–3</sup>. We compared 3,789 PSC cases of European ancestry to 25,079 population controls across 130,422 SNPs genotyped using the ImmunoChip<sup>4</sup>. We identified 12 genome-wide significant associations outside the human leukocyte antigen (HLA) complex, 9 of which were new, increasing the number of known PSC risk loci to 16. Despite comorbidity with inflammatory bowel disease (IBD) in 72% of the cases, 6 of the 12 loci showed significantly stronger association with PSC than with IBD, suggesting overlapping yet distinct genetic architectures for these two diseases. We incorporated association statistics from 7 diseases clinically occurring with PSC in the analysis and found suggestive evidence for 33 additional pleiotropic PSC risk loci. Together with network analyses, these findings add to the genetic risk map of PSC and expand on the relationship between PSC and other immune-mediated diseases.

The pathogenesis of PSC is poorly understood, and, owing to the lack of effective medical therapy, PSC remains a leading indicator for liver transplantation in northern Europe and the United States<sup>5</sup>, despite its relatively low prevalence (1 in 10,000). Affected

individuals are diagnosed at a median age of 30–40 years and suffer from an increased frequency of IBD (60–80%)<sup>5,6</sup> and autoimmune diseases (25%)<sup>7</sup>. Conversely, approximately only 5% of individuals with IBD develop PSC<sup>5,6</sup>. Sibling relative risk of 9- to 39-fold indicates a strong genetic component to PSC risk<sup>8</sup>. 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*)<sup>9–13</sup>.

Several theories have been proposed to explain the development of PSC<sup>5</sup>. The strong HLA associations and the clinical occurrence of PSC with immune-mediated diseases suggest that autoimmunity has a role in pathogenesis. To further characterize the genetic etiology of PSC, we recruited individuals with PSC throughout Europe and North America, more than doubling the number of ascertained cases included in previous genetic studies<sup>11</sup>. We genotyped 196,524 SNPs in 4,228 PSC cases and 27,077 population controls (Online Methods and **Supplementary Note**) using the ImmunoChip<sup>4,14</sup>, a targeted genotyping array with dense marker coverage across 186 known disease loci from 12 immune-mediated diseases. Outside these 186 loci, the ImmunoChip also assays thousands of SNPs of intermediate significance from multiple meta-analyses of immune-mediated diseases.

**Table 1 Association results of 12 non-HLA genome-wide significant risk loci for PSC**

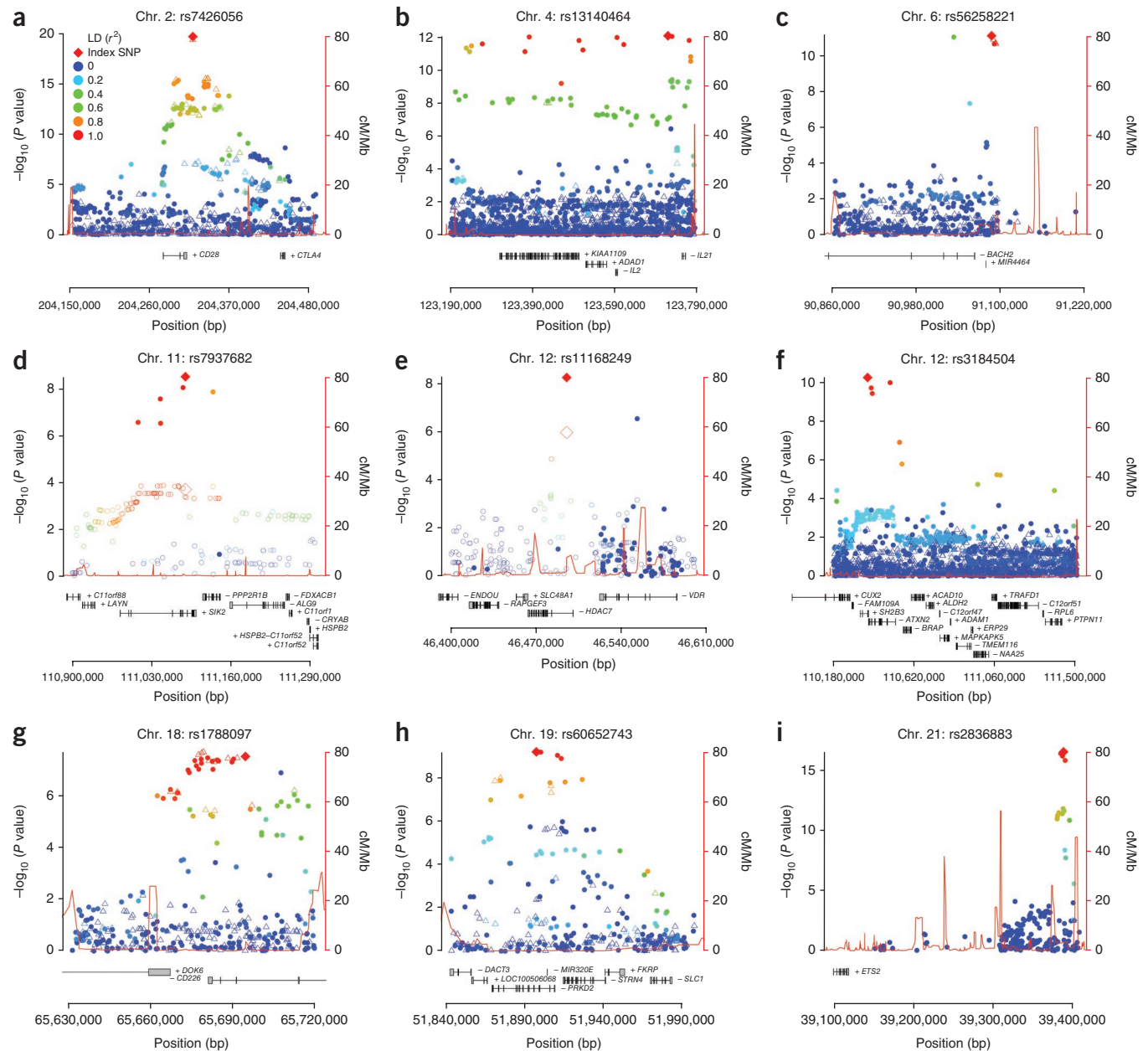
Chr.	SNP <sup>a</sup>	Risk allele	RAF cases	RAF controls	<i>P</i> value	OR (95% CI)	LD region <sup>b</sup> (kb)	RefSeq genes in LD region	Notable nearby genes <sup>c</sup>	Functional annotation <sup>d</sup>
1p36	rs3748816	A	0.698	0.656	$7.41 \times 10^{-12}$	1.21 (1.14–1.27)	2,398–2,775	9	<i>MMEL1-TNFRSF14</i>	eQTL, MS, OC, PB, HM
2q33	<b>rs7426056</b>	A	0.277	0.229	$1.89 \times 10^{-20}$	1.30 (1.23–1.37)	204,155–204,397	1	<i>CD28</i>	HM, OC
3p21	rs3197999	A	0.352	0.285	$2.45 \times 10^{-26}$	1.33 (1.26–1.40)	48,388–51,358	89	<i>MST1</i>	eQTL, MS, OC, PB, HM
4q27	<b>rs13140464</b>	C	0.871	0.836	$8.87 \times 10^{-13}$	1.30 (1.21–1.40)	123,204–123,784	4	<i>IL2-IL21</i>	OC, PB
6q15	<b>rs56258221</b>	G	0.213	0.183	$8.36 \times 10^{-12}$	1.23 (1.16–1.31)	90,967–91,150	1	<i>BACH2</i>	OC, PB
10p15	rs4147359	A	0.401	0.349	$8.19 \times 10^{-17}$	1.24 (1.18–1.30)	6,070–6,206	2	<i>IL2RA</i>	PB
11q23	<b>rs7937682</b>	G	0.298	0.265	$3.17 \times 10^{-09}$	1.17 (1.11–1.24)	110,824–111,492	19	<i>SIK2</i>	OC, PB, HM
12q13	<b>rs11168249</b>	G	0.506	0.466	$5.49 \times 10^{-09}$	1.15 (1.10–1.21)	46,442–46,534	3	<i>HDAC7</i>	OC, PB, HM
12q24	<b>rs3184504</b>	A	0.527	0.488	$5.91 \times 10^{-11}$	1.18 (1.12–1.24)	110,186–111,512	16	<i>SH2B3-ATXN2</i>	MS, OC, HM
18q22	<b>rs1788097</b>	A	0.518	0.483	$3.06 \times 10^{-08}$	1.15 (1.10–1.21)	65,633–65,721	2	<i>CD226</i>	MS, OC, PB, HM
19q13	<b>rs60652743</b>	A	0.864	0.836	$6.51 \times 10^{-10}$	1.25 (1.16–1.34)	51,850–51,998	6	<i>PRKD2-STRN4</i>	OC, PB, HM
21q22	<b>rs2836883</b>	G	0.777	0.728	$3.19 \times 10^{-17}$	1.28 (1.21–1.36)	39,374–39,404	–	<i>PSMG1</i>	OC, PB, HM

Chr., chromosome; RAF, risk allele frequency; OR, odds ratio; CI, confidence interval; HM, overlaps a region of histone modification; MS, missense mutation; OC, overlaps a known region of open chromatin; PB, overlaps a region of protein binding.

<sup>a</sup>SNPs from new PSC-associated loci are shown in bold. <sup>b</sup>LD regions around lead SNPs were calculated by extending a distance of 0.1 cM in both directions as defined by the HapMap recombination map. <sup>c</sup>Candidate genes within the same LD region as the associated SNPs. <sup>d</sup>Denotes whether there are SNPs with  $r^2 > 0.8$  with the top SNP that have functional annotations (**Supplementary Tables 4–7**).

A full list of authors and affiliations appears at the end of the paper.

Received 9 October 2012; accepted 29 March 2013; published online 21 April 2013; doi:10.1038/ng.2616

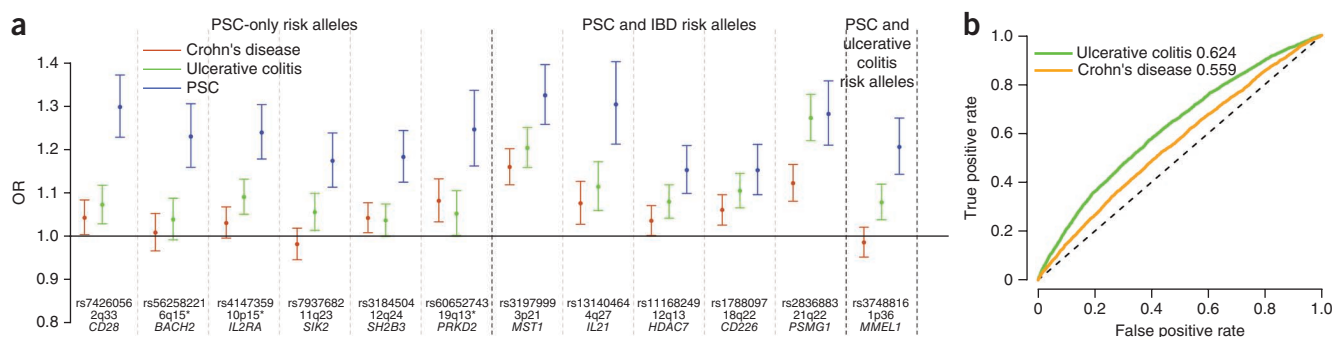


**Figure 1** Regional association plots of the nine loci newly associated with PSC at genome-wide significance ( $P < 5 \times 10^{-8}$ ). (a–i) Filled circles represent directly genotyped SNPs, and unfilled triangles represent imputed SNPs (Online Methods). The color of the markers shows LD with the most strongly associated SNP. Results are shown for SNPs rs7426056 (a), rs13140464 (b), rs56258221 (c), rs7937682 (d), rs11168249 (e), rs3184504 (f), rs1788097 (g), rs60652743 (h) and rs2836883 (i). Because the most strongly associated SNPs in d and e are located outside Immunochip fine-mapping regions, association results from the discovery panel of the largest PSC genome-wide association study (GWAS) so far<sup>12</sup> are shown as unfilled circles, and the most associated SNP at each locus is shown as an unfilled diamond (genotyped and imputed HapMap release 22 SNPs; cases overlap with those in the current study). Plus and minus signs by the gene symbols refer to the direction of gene transcription.

After quality control (Online Methods), 130,422 SNPs from 3,789 PSC cases and 25,079 population controls were available for analysis (Supplementary Figs. 1 and 2 and Supplementary Tables 1 and 2). We imputed a further 80,183 SNPs located in the Immunochip fine-mapping regions using the 1000 Genomes Project reference panel (Online Methods). We performed case-control association tests using a linear mixed model as implemented in MMM<sup>15</sup> to minimize the effect of population stratification and sample relatedness (genomic inflation factor  $\lambda_{GC} = 1.02$ , estimated using 2,544 ‘null’ SNPs, that is, SNPs located outside of known risk loci for immune-mediated diseases; Online Methods).

We identified 12 non-HLA genome-wide significant ( $P < 5 \times 10^{-8}$ ) susceptibility loci (Table 1), 9 of which were new (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 1.33) (Table 1). Genotype imputation and stepwise conditional regression analyses<sup>16</sup> within each locus did not identify additional independent genome-wide significant signals, nor did genotype-genotype or sex-genotype interaction analyses (Online Methods).

For seven of the nine new loci, the most significantly associated SNP in the locus was the same SNP or was in strong linkage



**Figure 2** Genetic similarity of loci associated with PSC and IBD. **(a)** Comparison of ORs for 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 designation of IBD loci as ulcerative colitis, Crohn's disease or both (IBD) were obtained from Jostins *et al.*<sup>22</sup>. Error bars represent 95% confidence intervals of 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 are located in the same broad genetic region as the IBD-associated SNPs; 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 ROC curves for discriminating PSC cases from population controls using ulcerative colitis or Crohn's disease ORs from 163 loci associated with IBD, respectively<sup>22</sup>. The dashed diagonal line is  $y = x$ , and specifies the ROC curve of a random predictor.

disequilibrium (LD;  $r^2 > 0.8$ ) with the original association reported 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<sup>17</sup>, and 6q15, where the most significantly associated PSC variant, rs56258221 (OR = 1.23;  $P = 8.36 \times 10^{-12}$ ), was 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 quality control and were confirmed in our analysis (1p36, 3p21 and 10p15; **Supplementary Fig. 3** and **Supplementary Note**).

To prioritize candidate genes within the non-HLA genome-wide significant loci, we searched for the functional consequences of the most associated SNPs or the SNPs in high LD with these ( $r^2 > 0.8$ ), focusing on missense SNPs (**Supplementary Fig. 4** and **Supplementary Table 4**) and expression quantitative trait loci (eQTLs) (**Supplementary Table 5**), and we functionally annotated risk loci using data from the Encyclopedia of DNA Elements (ENCODE) Project (**Supplementary Table 6** and **Supplementary Note**)<sup>18</sup>. We also constructed networks on the basis of functional similarity measures (Online Methods and **Supplementary Fig. 5**), known protein-protein interactions (DAPPLE<sup>19</sup>; **Supplementary Table 7** and **Supplementary Note**) and the published literature (GRAIL<sup>20</sup>; **Supplementary Fig. 6** and **Supplementary Note**) to identify disease-relevant genes. For 6 of the 12 genome-wide significant loci, the same gene (*MMEL1*, *CD28*, *MST1*, *SH2B3*, *CD226* and *SIK2*) was annotated by more than 1 method (**Supplementary Table 7**), suggesting these as candidates for further investigation at these loci.

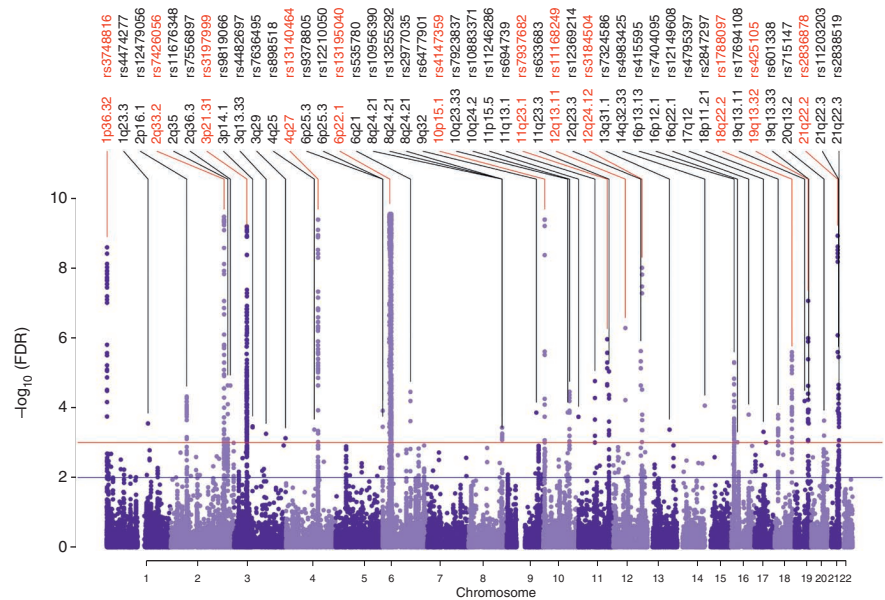
Two newly associated loci are located outside of the ImmunoChip fine-mapping regions (**Fig. 1d,e**). At 11q23, the most strongly associated SNP, rs7937682 (OR = 1.17;  $P = 3.17 \times 10^{-9}$ ), is located in an intron of *SIK2* (encoding salt-inducible kinase 2), which influences the expression of both interleukin (IL)-10 in macrophages and Nur77, an important transcription factor in leukocytes<sup>21</sup>. The association at 12q13 was with an intronic SNP, rs11168249 (OR = 1.15;  $P = 5.49 \times 10^{-9}$ ), within the *HDAC7* gene (encoding histone deacetylase 7), which has also been associated with IBD<sup>22</sup>. *HDAC7* has been implicated in the negative selection of T cells in the thymus<sup>23</sup>, a key process in the development of immune tolerance. A role for *HDAC7* in PSC etiology is supported by the new association

at 19q13, where the most associated SNP, rs60652743 (OR = 1.25;  $P = 6.51 \times 10^{-10}$ ), is located within an intron of *PRKD2* (encoding serine-threonine protein kinase D2). 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<sup>24,25</sup>. Notably, this negative selection takes place owing to a loss of *HDAC7*-mediated repression of *Nur77* (regulated by *SIK2*)<sup>26</sup>, linking three new 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-DQA1* and *HLA-DPB1* (**Supplementary Note**)<sup>27</sup>. The top associated SNP (rs4143332) was in almost perfect LD ( $r^2 = 0.996$ ) with *HLA-B\*08:01* (**Supplementary Note**). In 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**) explained most of the HLA association signal in PSC. When performing stepwise regression of the HLA alleles only, the class II associations were consistent with previous reports, apart from a new association with *HLA-DQA1\*01:01* (**Supplementary Tables 8–10** and **Supplementary Note**)<sup>9,28,29</sup>. The *HLA-DRB1\*15:01* association overlapped with the allele for ulcerative colitis (risk increasing) and Crohn's disease (risk decreasing)<sup>30,31</sup>. Because imputed genotypes in the class II region were only available for 4 (*HLA-DRB1*, *HLA-DQB1*, *HLA-DQA1* and *HLA-DPB1*) out of 20 loci<sup>32</sup>, further studies involving direct sequencing of all HLA class II loci along with assessments of protein structure and peptide binding are required to causally resolve the link between this HLA subregion and PSC development<sup>33,34</sup>.

Although 72% of the PSC cases 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 Fig. 8** and **Supplementary Table 3**)<sup>22</sup>, despite the greater sample size of that study (25,683 cases and 15,977 controls). Across the 12 non-HLA PSC-associated loci, we observed greater similarity between the OR estimates for PSC and ulcerative colitis than for those of PSC and Crohn's disease. We used the Crohn's disease and ulcerative colitis OR estimates for the 163 IBD-associated

**Figure 3** Pleiotropic PSC loci. Manhattan plot of conditional associations in PSC calculated as stratified FDRs on the basis of 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) (Online Methods and **Supplementary Figs. 10–12**). SNPs in red represent genome-wide significant findings from the main association analysis (**Table 1**), and SNPs 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**), and the blue horizontal line represents a threshold of  $FDR < 0.01$  (see **Supplementary Table 14** for a full listing of identified loci).



loci to predict PSC case-control status in our sample (Online Methods)<sup>22</sup> and found a significantly greater area under the receiver operating characteristic (ROC) curve (AUC) when applying ulcerative colitis ORs compared to Crohn's disease ORs (ulcerative colitis AUC = 0.62, Crohn's disease AUC = 0.56;  $P = 1.2 \times 10^{-57}$ ; **Fig. 2b**). This suggests that PSC is genetically more similar to ulcerative colitis than to Crohn's disease and is consistent with clinical observations of greater comorbidity of PSC with ulcerative colitis than with Crohn's disease<sup>35</sup>. To further compare the genetic profiles of PSC and IBD, we combined our genome-wide significant PSC-associated loci with the 163 confirmed IBD-associated loci<sup>22</sup> in a functional similarity network (**Supplementary Fig. 9** and **Supplementary Table 12**). We found that the PSC loci were distributed throughout the IBD loci (**Supplementary Fig. 9**), suggesting that there is no particular functional subcluster of IBD susceptibility genes associated with PSC and vice versa.

Although we consider only those loci reaching a stringent significance threshold ( $P < 5 \times 10^{-8}$ ) to be conclusively associated with 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 control is well suited to focused genotyping platforms such as the MetaboChip<sup>36</sup> and the ImmunoChip because it implicitly accounts for the expected enrichment in association. To further increase this enrichment, we exploited the known pleiotropy between related immune-mediated traits<sup>37</sup> and calculated the FDR<sup>38–40</sup> for association with PSC conditional on previously published summary statistics from each of the related phenotypes (yielding a per-SNP conditional FDR)<sup>41</sup> (Online Methods). We identified 33 non-HLA loci with conditional  $FDR < 0.001$  in this analysis (**Fig. 3**), all of which showed suggestive levels of significance ( $5 \times 10^{-8} < P < 5 \times 10^{-5}$ ) in the standard association analysis (**Supplementary Figs. 10–12** and **Supplementary Tables 13** and **14**). These loci were integrated into the functional similarity network analysis (**Supplementary Fig. 13** and **Supplementary Table 15**), highlighting potential candidate susceptibility genes.

In conclusion, the present study increases the number of genome-wide significant loci in PSC from 7 to 16 (including the HLA complex). The 9 new 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 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 pathogenesis.

**URLs.** 1000 Genomes HumanOmni2.5M genotype data, [ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20110921\\_phase2\\_omni\\_genotypes/Omni25\\_genotypes\\_1856\\_samples.b36.20110921.vcf.gz](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20110921_phase2_omni_genotypes/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 (in German), <http://www.popgen.de/>.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

*Note: Supplementary information is available in the online version of the paper.*

## ACKNOWLEDGMENTS

We thank all individuals with PSC and healthy controls for their participation, and we are indebted to all physicians and nursing staff who recruited subjects. We thank T. Wesse, T. Henke, S. Sedghpour Sabet, R. Vogler, G. Jacobs, I. Urbach, W. Albrecht, V. Pelkonen, V. Barbu, K. Holm, H. Dahlen Sollid, B. Woldseth, J.A. Anmarkrud and L.W. Torbjørnsen for expert help. U. Beuers, F. Braun, W. Kreisel, T. Berg and R. Günther are acknowledged for contributing German individuals with PSC. B.A. Lie and The Norwegian Bone Marrow Donor Registry at Oslo University Hospital, Rikshospitalet (Oslo, Norway) and the Nord-Trøndelag Health Study (HUNT) are acknowledged for sharing 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 the Dietary, Life style and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) consortium (see URLs), the Cooperative Research in the Region of Augsburg (KORA) study and the Heinz Nixdorf Recall (Risk Factors, Evaluation of Coronary Calcification, and Lifestyle) study. We acknowledge the members of the International PSC Study Group, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Inflammatory Bowel Disease Genetics Consortium (IBDGC), the UK-PSC Consortium and the Alberta IBD Consortium for their participation. J. Barrett is acknowledged for contributions to the design of the

ImmunoChip experiment. Individuals who have shared 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 by the PopGen Biobank (see URLs). J.Z.L., T.S. and C.A.A. are supported by a grant from the Wellcome Trust (098051). Additional financial support of the study and the coauthors is listed in the **Supplementary Note**.

#### AUTHOR CONTRIBUTIONS

J.Z.L., J.R.H., T.F., E.E., N.T.D., I.T., G.M., I.R.K., O.A.A., W.K.T., A.M.D., T.S. and C.A.A. performed data and statistical analyses. A. Franke, 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. Franke, C.A.A. and T.H.K. drafted the manuscript. S.M.R., R.K.W., T.J.W., B.E., P.L., 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. Sterneck, A.T., A.L.M., J.S., V.L., R.D., D.A., A. Floreani, S.O.-G., S.S.R., A.J.S., S.N., K.H., I.C., J.G.-A., I.R.-P., D.v.H., E.B., R.N.S., P.R.D., E.M., M.H.V., M.S.S., R.H.D., L.P., S.B., M. Sans, 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 ascertainment of affected individuals and/or sample and clinical data collection. All authors revised the manuscript for critical content and approved the final version.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Aadland, E. *et al.* Primary sclerosing cholangitis: a long-term follow-up study. *Scand. J. Gastroenterol.* **22**, 655–664 (1987).
- Broomé, U. *et al.* Natural history and prognostic factors in 305 Swedish patients with primary sclerosing cholangitis. *Gut* **38**, 610–615 (1996).
- Farrant, J.M. *et al.* Natural history and prognostic variables in primary sclerosing cholangitis. *Gastroenterology* **100**, 1710–1717 (1991).
- Cortes, A. & Brown, M.A. Promise and pitfalls of the ImmunoChip. *Arthritis Res. Ther.* **13**, 101 (2011).
- Karlsen, T.H., Schruppf, E. & Boberg, K.M. Update on primary sclerosing cholangitis. *Dig. Liver Dis.* **42**, 390–400 (2010).
- Karlsen, T.H. & Kaser, A. Deciphering the genetic predisposition to primary sclerosing cholangitis. *Semin. Liver Dis.* **31**, 188–207 (2011).
- Saarinen, S., Olerup, O. & Broome, U. Increased frequency of autoimmune diseases in patients with primary sclerosing cholangitis. *Am. J. Gastroenterol.* **95**, 3195–3199 (2000).
- 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.* **6**, 939–943 (2008).
- Karlsen, T.H. *et al.* Genome-wide association analysis in primary sclerosing cholangitis. *Gastroenterology* **138**, 1102–1111 (2010).
- Srivastava, B. *et al.* Fine mapping and replication of genetic risk loci in primary sclerosing cholangitis. *Scand. J. Gastroenterol.* **47**, 820–826 (2012).
- Folseraas, T. *et al.* Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. *J. Hepatol.* **57**, 366–375 (2012).
- Melum, E. *et al.* Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. *Nat. Genet.* **43**, 17–19 (2011).
- Ellinghaus, D. *et al.* Genome-wide association analysis in sclerosing cholangitis and ulcerative colitis identifies risk loci at *GPR35* and *TCF4*. *Hepatology* published online; doi:10.1002/hep.25977 (23 July 2012).
- Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* **43**, 1193–1201 (2011).
- 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.* **7**, 369–390 (2013).
- Cordell, H.J. & Clayton, D.G. 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.* **70**, 124–141 (2002).
- Peters, U. *et al.* Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum. Genet.* **131**, 217–234 (2012).
- Gerstein, M.B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
- Rossin, E.J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* **7**, e1001273 (2011).
- Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
- Hanna, R.N. *et al.* The transcription factor NR4A1 (Nur77) controls bone marrow differentiation and the survival of Ly6C<sup>+</sup> monocytes. *Nat. Immunol.* **12**, 778–785 (2011).
- Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
- Kasler, H.G. *et al.* Histone deacetylase 7 regulates cell survival and TCR signaling in CD4/CD8 double-positive thymocytes. *J. Immunol.* **186**, 4782–4793 (2011).
- Dequiedt, F. *et al.* HDAC7, a thymus-specific class II histone deacetylase, regulates Nur77 transcription and TCR-mediated apoptosis. *Immunity* **18**, 687–698 (2003).
- 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.* **201**, 793–804 (2005).
- Clark, K. *et al.* Phosphorylation of CRT3 by the salt-inducible kinases controls the interconversion of classically activated and regulatory macrophages. *Proc. Natl. Acad. Sci. USA* **109**, 16986–16991 (2012).
- Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44**, 291–296 (2012).
- Schrumpf, E. *et al.* HLA antigens and immunoregulatory T cells in ulcerative colitis associated with hepatobiliary disease. *Scand. J. Gastroenterol.* **17**, 187–191 (1982).
- Spurkland, A. *et al.* HLA class II haplotypes in primary sclerosing cholangitis patients from five European populations. *Tissue Antigens* **53**, 459–469 (1999).
- Stokkers, P.C., Reitsma, P.H., Tytgat, G.N. & van Deventer, S.J. HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis. *Gut* **45**, 395–401 (1999).
- 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* **141**, 864–871. e1–5 (2011).
- Horton, R. *et al.* Gene map of the extended human MHC. *Nat. Rev. Genet.* **5**, 889–899 (2004).
- Hov, J.R. *et al.* Electrostatic modifications of the human leukocyte antigen-DR P9 peptide-binding pocket and susceptibility to primary sclerosing cholangitis. *Hepatology* **53**, 1967–1976 (2011).
- Hovhannisyan, Z. *et al.* The role of HLA-DQ8 β57 polymorphism in the anti-gluten T-cell response in coeliac disease. *Nature* **456**, 534–538 (2008).
- Broomé, U. & Bergquist, A. Primary sclerosing cholangitis, inflammatory bowel disease, and colon cancer. *Semin. Liver Dis.* **26**, 31–41 (2006).
- CARDIOGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25–33 (2013).
- Zhernakova, A., van Diemen, C.C. & Wijmenga, C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat. Rev. Genet.* **10**, 43–55 (2009).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
- Storey, J.D. The positive false discovery rate: a Bayesian interpretation and the *q*-value. *Ann. Stat.* **31**, 2013–2035 (2003).
- Efron, B. Simultaneous inference: when should hypothesis testing problems be combined? *Ann. Appl. Statist.* **2**, 197–223 (2008).
- Sun, L., Craiu, R.V., Paterson, A.D. & Bull, S.B. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genet. Epidemiol.* **30**, 519–530 (2006).

Jimmy Z Liu<sup>1,86</sup>, Johannes Roksdund Hov<sup>2–5,86</sup>, Trine Folseraas<sup>2–4,86</sup>, Eva Ellinghaus<sup>6,86</sup>, Simon M Rushbrook<sup>7</sup>, Nadezhda T Doncheva<sup>8</sup>, Ole A Andreassen<sup>4,9</sup>, Rinse K Weersma<sup>10</sup>, Tobias J Weismüller<sup>11,12,85</sup>, Bertus Eksteen<sup>13</sup>, Pietro Invernizzi<sup>14</sup>, Gideon M Hirschfield<sup>15,16</sup>, Daniel Nils Gotthardt<sup>17</sup>, Albert Pares<sup>18</sup>, David Ellinghaus<sup>6</sup>, Tejas Shah<sup>1</sup>, Brian D Juran<sup>19</sup>, Piotr Milkiewicz<sup>20</sup>, Christian Rust<sup>21</sup>, Christoph Schramm<sup>22</sup>, Tobias Müller<sup>23</sup>, Brijesh Srivastava<sup>24</sup>, Georgios Dalekos<sup>25,26</sup>, Markus M Nöthen<sup>27,28</sup>, Stefan Herms<sup>27,28</sup>, Juliane Winkelmann<sup>29–31</sup>, Mitja Mitrovic<sup>32</sup>, Felix Braun<sup>33</sup>, Cyriel Y Ponsioen<sup>34</sup>, Peter J P Croucher<sup>35</sup>, Martina Sterneck<sup>36</sup>, Andreas Teufel<sup>37</sup>, Andrew L Mason<sup>38</sup>, Janna Saarela<sup>39</sup>, Virpi Leppä<sup>40</sup>, Ruslan Dorfman<sup>41</sup>, Domenico Alvaro<sup>42</sup>, Annarosa Floreani<sup>43</sup>, Suna Onengut-Gumuscu<sup>44,45</sup>, Stephen S Rich<sup>46,47</sup>, Wesley K Thompson<sup>48</sup>,

Andrew J Schork<sup>49</sup>, Sigrid Næss<sup>2-4</sup>, Ingo Thomsen<sup>6</sup>, Gabriele Mayr<sup>8</sup>, Inke R König<sup>50</sup>, Kristian Hveem<sup>51</sup>, Isabelle Cleynen<sup>1,52</sup>, Javier Gutierrez-Achury<sup>32</sup>, Isis Ricaño-Ponce<sup>32</sup>, David van Heel<sup>53</sup>, Einar Björnsson<sup>54</sup>, Richard N Sandford<sup>24</sup>, Peter R Durie<sup>55</sup>, Espen Melum<sup>2-4</sup>, Morten H Vatn<sup>4,5,56</sup>, Mark S Silverberg<sup>57</sup>, Richard H Duerr<sup>58,59</sup>, Leonid Padyukov<sup>60</sup>, Stephan Brand<sup>61</sup>, Miquel Sans<sup>62</sup>, Vito Annese<sup>63,64</sup>, Jean-Paul Achkar<sup>65,66</sup>, Kirsten Muri Boberg<sup>2,4,5</sup>, Hanns-Ulrich Marschall<sup>67</sup>, Olivier Chazouillères<sup>68</sup>, Christopher L Bowlus<sup>69</sup>, Cisca Wijmenga<sup>32</sup>, Erik Schrumpf<sup>2,4,5</sup>, Severine Vermeire<sup>52,70</sup>, Mario Albrecht<sup>8,71</sup>, The UK-PSC Consortium<sup>72</sup>, The International IBD Genetics Consortium<sup>72</sup>, John D Rioux<sup>73,74</sup>, Graeme Alexander<sup>75</sup>, Annika Bergquist<sup>76</sup>, Judy Cho<sup>77</sup>, Stefan Schreiber<sup>6,78,79</sup>, Michael P Manns<sup>11,12</sup>, Martti Färkkilä<sup>80</sup>, Anders M Dale<sup>81,82</sup>, Roger W Chapman<sup>83</sup>, Konstantinos N Lazaridis<sup>19</sup>, The International PSC Study Group<sup>72</sup>, Andre Franke<sup>6,87</sup>, Carl A Anderson<sup>1,87</sup> & Tom H Karlsen<sup>2,3,5,84,87</sup>

<sup>1</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. <sup>2</sup>Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway. <sup>3</sup>K.G. Jebsen Inflammation Research Centre, Research Institute of Internal Medicine, Oslo University Hospital, Rikshospitalet, Oslo, Norway. <sup>4</sup>Institute of Clinical Medicine, University of Oslo, Oslo, Norway. <sup>5</sup>Section of Gastroenterology, Department of Transplantation Medicine, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway. <sup>6</sup>Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany. <sup>7</sup>Department of Gastroenterology and Hepatology, Norfolk and Norwich, University Hospitals National Health Service (NHS) Trust, Norwich, UK. <sup>8</sup>Max Planck Institute for Informatics, Saarbrücken, Germany. <sup>9</sup>K.G. Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital, Ullevål, Oslo, Norway. <sup>10</sup>Department of Gastroenterology and Hepatology, University of Groningen and University Medical Centre Groningen, Groningen, The Netherlands. <sup>11</sup>Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany. <sup>12</sup>Integrated Research and Treatment Center-Transplantation (IFB-tx), Hannover Medical School, Hannover, Germany. <sup>13</sup>Snyder Institute of Chronic Diseases, Department of Medicine, University of Calgary, Calgary, Alberta, Canada. <sup>14</sup>Center for Autoimmune Liver Diseases, Humanitas Clinical and Research Center, Rozzano, Italy. <sup>15</sup>Division of Gastroenterology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada. <sup>16</sup>Centre for Liver Research, National Institute for Health Research (NIHR) Biomedical Research Unit, Birmingham, UK. <sup>17</sup>Department of Medicine, University Hospital of Heidelberg, Heidelberg, Germany. <sup>18</sup>Liver Unit, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red en el Área Temática de Enfermedades Hepáticas y Digestivas (CIBERehd), University of Barcelona, Barcelona, Spain. <sup>19</sup>Center for Basic Research in Digestive Diseases, Division of Gastroenterology and Hepatology, Mayo Clinic, College of Medicine, Rochester, Minnesota, USA. <sup>20</sup>Liver Unit and Liver Research Laboratories, Pomeranian Medical University, Szczecin, Poland. <sup>21</sup>Department of Medicine 2, Grosshadern, University of Munich, Munich, Germany. <sup>22</sup>1st Department of Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. <sup>23</sup>Department of Internal Medicine, Hepatology and Gastroenterology, Charité Universitätsmedizin Berlin, Berlin, Germany. <sup>24</sup>Academic Department of Medical Genetics, University of Cambridge, Cambridge, UK. <sup>25</sup>Department of Medicine, Medical School, University of Thessaly, Larissa, Greece. <sup>26</sup>Research Laboratory of Internal Medicine, Medical School, University of Thessaly, Larissa, Greece. <sup>27</sup>Institute of Human Genetics, University of Bonn, Bonn, Germany. <sup>28</sup>Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany. <sup>29</sup>Institute of Human Genetics, Technische Universität München, Munich, Germany. <sup>30</sup>Department of Neurology, Technische Universität München, Munich, Germany. <sup>31</sup>Institute of Human Genetics, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany. <sup>32</sup>Department of Genetics, University of Groningen and University Medical Centre Groningen, Groningen, The Netherlands. <sup>33</sup>Department of General, Visceral, Thoracic, Transplantation and Pediatric Surgery, University Medical Centre Schleswig-Holstein, Campus Kiel, Kiel, Germany. <sup>34</sup>Department of Gastroenterology and Hepatology, Academic Medical Center, Amsterdam, The Netherlands. <sup>35</sup>Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, California, USA. <sup>36</sup>Department of Hepatobiliary Surgery and Transplantation, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. <sup>37</sup>1st Department of Medicine, University of Mainz, Mainz, Germany. <sup>38</sup>Division of Gastroenterology and Hepatology, University of Alberta, Edmonton, Alberta, Canada. <sup>39</sup>Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. <sup>40</sup>Public Health Genomics Unit, FIMM, University of Helsinki and National Institute for Health and Welfare, Helsinki, Finland. <sup>41</sup>Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada. <sup>42</sup>Division of Gastroenterology, Department of Clinical Medicine, Sapienza University of Rome, Rome, Italy. <sup>43</sup>Department of Surgical, Oncological and Gastroenterological Sciences, University of Padova, Padova, Italy. <sup>44</sup>Center for Public Health Genomics, Division of Endocrinology & Metabolism, University of Virginia, Charlottesville, Virginia, USA. <sup>45</sup>Department of Internal Medicine, Division of Endocrinology & Metabolism, University of Virginia, Charlottesville, Virginia, USA. <sup>46</sup>Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA. <sup>47</sup>Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA. <sup>48</sup>Department of Psychiatry, University of California, San Diego, La Jolla, California, USA. <sup>49</sup>Graduate Program in Cognitive Science, University of California, San Diego, La Jolla, California, USA. <sup>50</sup>Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany. <sup>51</sup>Department of Public Health, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway. <sup>52</sup>Department of Clinical and Experimental Medicine, KU Leuven, Leuven, Belgium. <sup>53</sup>Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. <sup>54</sup>Division of Gastroenterology and Hepatology, Department of Internal Medicine, Landspítali University Hospital, Reykjavík, Iceland. <sup>55</sup>Physiology and Experimental Medicine, Research Institute, Hospital for Sick Children, Toronto, Ontario, Canada. <sup>56</sup>EpiGen, Campus Akershus University Hospital (AHUS), Akershus University Hospital, Nordbyhagen, Norway. <sup>57</sup>Inflammatory Bowel Disease (IBD) Group, Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital, Toronto, Ontario, Canada. <sup>58</sup>Division of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA. <sup>59</sup>Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. <sup>60</sup>Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden. <sup>61</sup>Department of Medicine II, University Hospital Munich-Grosshadern, Ludwig-Maximilians-University Munich, Germany. <sup>62</sup>Department of Digestive Diseases, Centro Médico Teknon, Barcelona, Spain. <sup>63</sup>Division of Gastroenterology, Istituto di Ricovero e Cura a Carattere Scientifico, Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy. <sup>64</sup>Unit of Gastroenterology SOD2, Azienda Ospedaliero Universitaria Careggi, Florence, Italy. <sup>65</sup>Department of Gastroenterology and Hepatology, Digestive Disease Institute, Cleveland Clinic, Cleveland, Ohio, USA. <sup>66</sup>Department of Pathobiology, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA. <sup>67</sup>Department of Internal Medicine, Institute of Medicine, Sahlgrenska Academy and University Hospital, Gothenburg, Sweden. <sup>68</sup>Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Saint Antoine, Department of Hepatology, University Pierre et Marie Curie (UPMC) Université de Paris 6, Paris, France. <sup>69</sup>Division of Gastroenterology and Hepatology, University of California, Davis, Davis, California, USA. <sup>70</sup>Department of Gastroenterology, University Hospitals Leuven, Leuven, Belgium. <sup>71</sup>Department of Bioinformatics, Institute of Biometrics and Medical Informatics, University Medicine Greifswald, Greifswald, Germany. <sup>72</sup>Further details appear in the **Supplementary Note**. <sup>73</sup>Université de Montréal, Research Center, Montreal, Quebec, Canada. <sup>74</sup>Montreal Heart Institute, Research Center, Montreal, Quebec, Canada. <sup>75</sup>Division of Hepatology, Department of Medicine, University of Cambridge, Cambridge, UK. <sup>76</sup>Department of Gastroenterology and Hepatology, Karolinska University Hospital Huddinge, Karolinska Institutet, Stockholm, Sweden. <sup>77</sup>Section of Digestive Diseases, Department of Medicine, Yale University, New Haven, Connecticut, USA. <sup>78</sup>Department for General Internal Medicine, Christian-Albrechts-University of Kiel, Kiel, Germany. <sup>79</sup>PopGen Biobank, University Hospital Schleswig-Holstein, Christian-Albrechts-University of Kiel, Kiel, Germany. <sup>80</sup>Division of Gastroenterology, Department of Medicine, Helsinki University Hospital, Helsinki, Finland. <sup>81</sup>Department of Radiology, University of California, San Diego, La Jolla, California, USA. <sup>82</sup>Department of Neurosciences, University of California, San Diego, La Jolla, California, USA. <sup>83</sup>Department of Hepatology, John Radcliffe University Hospitals NHS Trust, Oxford, UK. <sup>84</sup>Division of Gastroenterology, Department of Clinical Medicine, University of Bergen, Bergen, Norway. <sup>85</sup>Present address: Department of Internal Medicine 1, University Hospital of Bonn, Bonn, Germany. <sup>86</sup>These authors contributed equally to this work. <sup>87</sup>These authors jointly directed this work. Correspondence should be addressed to T.H.K. (t.h.karlsen@medisin.uio.no) or C.A.A. (carl.anderson@sanger.ac.uk).

## ONLINE METHODS

**Study subjects.** The study participants are described in **Supplementary Table 16** and the **Supplementary Note**.

**Ethical approval.** Subject 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 call rate < 80% were removed before commencing sample quality control ( $n = 235$ ). Per-individual genotype call rate and heterozygosity rate were calculated using PLINK<sup>42</sup>, and outlying samples were identified using Aberrant<sup>43</sup>, which automatically identifies outliers in 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 HumanOmni2.5-8 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  $\geq 0.9$ , the sample with the 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 ( $0.1875 < \text{identity by descent} < 0.9$ ) remained in the analysis to maximize power because mixed model association analysis can correctly account for relatedness. Principal-components analysis, implemented in SMARTPCA (Eigenstrat)<sup>44</sup>, was used to identify samples of non-European ancestry. Principal components were defined using population samples from the 1000 Genomes Project<sup>45</sup> genotyped using the Illumina HumanOmni2.5-8 genotyping array (see URLs) and then projected into cases and controls (**Supplementary Fig. 2**)<sup>14,22,46</sup>. After sample quality control, 3,789 PSC cases and 25,079 controls remained. SNPs with MAF of less than 0.1%, Hardy-Weinberg equilibrium  $P < 1 \times 10^{-5}$  or call rate lower than 98% and those that failed the PLINK v1.07 non-random differential missing data rate test between cases and controls ( $P < 1 \times 10^{-5}$ ) were excluded. After the completion of marker quality control (**Supplementary Table 2**), 131,220 SNPs were available for analysis, further reduced to 130,422 after cluster plot inspection.

**Statistical methods. Genomic inflation factor.** The ImmunoChip contains 3,120 SNPs that were part of a bipolar disease replication effort and other studies that were not immune related. After quality control, 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 SNPs after quality control located in the ImmunoChip fine-mapping regions, additional genotypes were imputed using IMPUTE2 with the 1000 Genomes Project Phase 1 (March 2012) reference panel of 1,092 individuals<sup>47</sup> 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 ( $P < 1 \times 10^{-5}$ ) in the ten batches were removed, as were SNPs that did not pass the exclusion thresholds used for genotyped SNP quality control. 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<sup>15</sup>. 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<sup>48–50</sup>.  $R$  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 it is typically estimated using genome-wide SNP data<sup>15</sup>. 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 were in the HLA region or had MAF < 10%. Finally, we excluded SNPs that showed modest association ( $P < 0.005$ ) with PSC in a linear regression model fitting the first ten principal components as covariates. A total of 17,260 SNPs were used to estimate  $R$ .

Owing 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-weighted) meta-analysis, as carried out previously<sup>48</sup>. This reduced the  $\lambda_{GC}$  inflation factor, 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  $\times$  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  $< 5 \times 10^{-6}$  were visually inspected using Evoker<sup>51</sup>. SNPs that clustered poorly were removed ( $n = 798$ ).

**Prediction of PSC using IBD SNPs.** OR estimates for Crohn's disease and ulcerative colitis in 163 IBD-associated SNPs were obtained from Jostins *et al.*<sup>22</sup>. We used the R package Mangrove (see URLs) to estimate each individual's probability of developing PSC in 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 powers of the ulcerative colitis and Crohn's disease ORs. The DeLong method was used to test whether the AUC was significantly greater using ulcerative colitis ORs compared to Crohn's disease ORs<sup>52</sup>.

**Functional similarity networks.** In functional similarity networks, each edge represents the strong functional similarity of two genes based on annotated Gene Ontology (GO) terms as determined by the functional similarity measure rfunSim<sup>53</sup>. rfunSim similarity values above the recommended cutoff of 0.8 were retrieved using the FunSimMat web service<sup>54</sup>. Resulting networks were visualized and analyzed using Cytoscape<sup>55</sup>.

To construct PSC-specific networks from functional similarity networks that contained more than one gene per locus (**Supplementary Figs. 5 and 13**), the connectivity of each gene was assessed by computing different topology measures for the corresponding node: (i) degree (number of direct edges to other nodes), (ii) shortest path closeness (inverted average shortest path distance to other nodes) and (iii) 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. Genes were first ranked according to each measure and were then assigned the best of the three ranks. The PSC-specific network was generated from the top ranked genes in their respective loci.

**Pleiotropy analysis.** We included summary statistics from GWAS of seven PSC-associated diseases (Crohn's disease, celiac disease, psoriasis, rheumatoid arthritis, sarcoidosis, type 1 diabetes and ulcerative colitis; **Supplementary Table 17**). For all diseases, we constructed conditional stratified quantile-quantile plots of the empirical quantiles of nominal  $-\log_{10}(P$  values) for SNP association with PSC for all SNPs (**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 by  $P < 1$ ,  $P < 0.1$ ,  $P < 0.01$  and  $P < 0.001$  in the pleiotropic phenotype). For a given PSC-associated phenotype, enrichment for pleiotropic signals in PSC can be observed as an increasing leftward deflection from the expected null distribution with lower  $P$ -value thresholds in the second phenotype (**Supplementary Note**). Enrichment in the stratified quantile-quantile plots is directly interpretable in terms of the true discovery rate (TDR), equivalent to  $1 - \text{FDR}$ <sup>56</sup>. Specifically, it can be shown that a conservative estimate of FDR can be calculated from the horizontal shift of the quantile-quantile curve from the expected line  $y = x$ , 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**).

To assess the significance of the association with PSC, we assigned a pleiotropic (conditional) FDR value for PSC for each SNP. The pleiotropic FDR value for each SNP was 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 were defined by the pleiotropic association (lowest  $P$  value among associated diseases) of the SNP. Notably, the conditioning procedure was blind to the  $P$  value of the SNP with respect to PSC. The pleiotropic FDR was then interpolated from conditional FDR curves using established stratified FDR methods<sup>41,57</sup> (**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 the lowest FDR within each LD block (as defined by 1000 Genomes Project data) was considered the lead SNP of a new pleiotropic PSC locus, if below a threshold of 0.001 (loci defined by  $FDR < 0.001$  and  $FDR < 0.01$  are shown in **Supplementary Tables 13 and 14**). All test statistics were adjusted for population stratification by genomic control (**Supplementary Fig. 16 and Supplementary Note**).

**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<sup>58</sup> assuming a disease prevalence of 10 in 100,000 and multiplicative risk.

42. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
43. Bellenguez, C., Strange, A., Freeman, C., Donnelly, P. & Spencer, C.C. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. *Bioinformatics* **28**, 134–135 (2012).
44. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
45. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
46. Liu, J.Z. *et al.* Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* **44**, 1137–1141 (2012).
47. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
48. Sawcer, S. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
49. Korte, A. *et al.* A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* **44**, 1066–1071 (2012).
50. Tsoi, L.C. *et al.* Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat. Genet.* **44**, 1341–1348 (2012).
51. Morris, J.A., Randall, J.C., Maller, J.B. & Barrett, J.C. Evoker: a visualization tool for genotype intensity data. *Bioinformatics* **26**, 1786–1787 (2010).
52. DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
53. Schlicker, A., Domingues, F.S., Rahnenfuhrer, J. & Lengauer, T. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics* **7**, 302 (2006).
54. Schlicker, A. & Albrecht, M. FunSimMat update: new features for exploring functional similarity. *Nucleic Acids Res.* **38**, D244–D248 (2010).
55. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
56. Efron, B. Size, power and false discovery rates. *Ann. Stat.* **35**, 1351–1377 (2007).
57. Yoo, Y.J., Pinnaduwa, D., Waggott, D., Bull, S.B. & Sun, L. Genome-wide association analyses of North American Rheumatoid Arthritis Consortium and Framingham Heart Study data utilizing genome-wide linkage results. *BMC Proc.* **3** (suppl. 7), S103 (2009).
58. So, H.C., Gui, A.H., Cherny, S.S. & Sham, P.C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.* **35**, 310–317 (2011).